



**Dipartimento di Statistica**  
**"Giuseppe Parenti"**

Dipartimento di Statistica "G. Parenti" – Viale Morgagni 59 – 50134 Firenze – [www.ds.unifi.it](http://www.ds.unifi.it)

Q U A D E R N I 2 0 0 5 / 0 2

PROCESSI A MEMORIA LUNGA E  
FRACTIONALLY INTEGRATED  
GARCH

Marco J. Lombardi,  
Giampiero M. Gallo



Università degli Studi  
di Firenze

*Marco J. Lombardi e Giampiero M. Gallo*

**PROCESSI A MEMORIA LUNGA  
E FRACTIONALLY INTEGRATED  
GARCH**



*«Doing econometrics is like trying to learn  
the laws of electricity by playing the radio.»*

*Guy Orcutt*

$$e^{\pi i} + 1 = 0$$

*Leonhard Euler*

# Premessa

Questo lavoro nasce da una revisione critica della tesi di laurea di Marco J. Lombardi. Rispetto all'originale, ha un taglio più generale, essendo più focalizzato sulla teoria generale dei processi a memoria lunga. In particolare, si è rivista e ampliata la parte relativa alla stima e alle metodologie di test per modelli ARFIMA, corredandola anche di un'esemplificazione pratica.

Molte persone hanno contribuito alla stesura di questo lavoro ed ad esse vanno i nostri più vivi ringraziamenti. Umberto Cherubini, Giovanni Della Lunga e Marianna Marziliani hanno avuto la pazienza di insegnare al primo autore i rudimenti della programmazione C/C++ e dei metodi numerici di ottimizzazione, senza i quali egli non avrebbe potuto realizzare buona parte di questo lavoro.

Pur trattandosi di lavoro congiunto, Giampiero M. Gallo ha curato in particolare il primo capitolo e Marco J. Lombardi i restanti.

# Sommario

<b>1. Introduzione.....</b>	<b>1</b>
1.1 Lo studio delle serie storiche.....	2
1.1.1 Cenni preliminari.....	2
1.1.2 Il ruolo delle previsioni.....	4
1.1.3 I rapporti con le scienze sperimentali.....	7
1.1.4 La struttura della disciplina.....	8
1.2 La nascita dell'integrazione frazionaria.....	10
1.2.1 La modellistica tradizionale.....	10
1.2.2 Gli studi sulla densità spettrale.....	14
1.2.3 Nuovi modelli e nuove teorie.....	16
1.3 Applicazioni empiriche.....	18
1.3.1 Applicazioni macroeconomiche.....	18
1.3.2 Applicazioni a serie finanziarie ad alta frequenza.....	18
1.3.3 Introduzione alla modellistica ARCH a memoria lunga.....	19
1.4 Piano e obiettivi del lavoro.....	21

<b>2. Richiami matematici e statistici</b> .....	24
2.1 Funzioni utili.....	26
2.1.1 <i>La funzione Gamma</i> .....	26
2.1.2 <i>La funzione Beta</i> .....	27
2.1.3 <i>I polinomi di Legendre</i> .....	29
2.2 Numeri complessi.....	31
2.2.1 <i>Definizioni</i> .....	31
2.2.2 <i>La rappresentazione grafica</i> .....	31
2.2.3 <i>Il teorema di De Moivre</i> .....	32
2.3 I fondamenti dell'analisi delle serie storiche.....	34
2.3.1 <i>Il concetto di processo stocastico</i> .....	34
2.3.2 <i>I processi ARMA</i> .....	35
2.3.3 <i>Stazionarietà e invertibilità</i> .....	36
2.3.4 <i>Processi integrati</i> .....	38
2.3.5 <i>I test per radice unitaria</i> .....	39
2.4 Analisi spettrale.....	41
2.4.1 <i>La funzione generatrice delle autocovarianze</i> .....	41
2.4.2 <i>Lo spettro del processo</i> .....	42
2.4.3 <i>Il significato dello spettro</i> .....	43
2.4.4 <i>Il periodogramma</i> .....	46
2.4.5 <i>Il teorema di rappresentazione spettrale</i> .....	47
2.4.6 <i>La stima dello spettro</i> .....	48

<b>3. L'integrazione frazionaria</b>	50
3.1 Il concetto di memoria lunga	51
3.2 Il modello ARFIMA: derivazione e proprietà	53
3.2.1 <i>Il fractional white noise</i>	53
3.2.2 <i>ARFIMA (p, d, q)</i>	57
3.2.3 <i>Le distribuzioni asintotiche</i>	58
3.3 L'individuazione dell'integrazione frazionaria	61
3.3.1 <i>Il test R/S</i>	61
3.3.2 <i>Il test KPSS</i>	62
3.3.3 <i>Il test Durbin-Watson adattato</i>	65
3.4 Modelli teorici	68
3.4.1 <i>L'aggregazione di modelli dinamici</i>	68
3.4.2 <i>Il modello Error-Duration</i>	72
3.5 Risultati empirici	74
3.5.1 <i>L'integrazione frazionaria nel PNL</i>	75
3.5.2 <i>L'integrazione frazionaria nell'inflazione</i>	76
3.5.3 <i>L'integrazione frazionaria nei tassi di interesse</i>	77
3.6 I processi di Gegenbauer	78
3.6.1 <i>Il processo di Gegenbauer</i>	79
3.6.2 <i>Le proprietà salienti</i>	79
3.6.3 <i>I processi GARMA</i>	81



<b>4. Problemi di stima econometrica.....</b>	<b>84</b>
4.1 La stima di massima verosimiglianza.....	85
4.1.1 <i>La funzione di verosimiglianza per serie storiche.....</i>	86
4.1.2 <i>Le proprietà asintotiche.....</i>	92
4.1.3 <i>La quasi-massima verosimiglianza.....</i>	93
4.2 I test basati sulla verosimiglianza.....	97
4.2.1 <i>Il test LR.....</i>	97
4.2.2 <i>Il test LM.....</i>	98
4.3 La stima dei modelli a integrazione frazionaria.....	99
4.3.1 <i>Massima verosimiglianza approssimata.....</i>	99
4.3.2 <i>Massima verosimiglianza esatta.....</i>	102
4.3.3 <i>La verosimiglianza di Fox e Taqqu.....</i>	106
4.3.4 <i>Il metodo GPH.....</i>	108
4.3.5 <i>MDE.....</i>	110
4.3.6 <i>Una tabella riepilogativa.....</i>	111
4.4 La stima dei modelli GARMA.....	112
4.5 Un esempio pratico.....	115

<b>5. Il modello FIGARCH.....</b>	<b>122</b>
5.1 Introduzione alla modellistica ARCH.....	123
5.1.1 Il modello ARCH.....	125
5.1.2 Il modello GARCH.....	128
5.1.3 Il modello IGARCH.....	130
5.1.4 I modelli asimmetrici.....	132
5.1.5 La stima dei modelli ARCH.....	133
5.1.6 I test ARCH.....	134
5.1.7 Le previsioni sulla volatilità.....	136
5.1.8 Il calcolo del Value at Risk.....	138
5.2 La memoria lunga nei rendimenti ad alta frequenza.....	141
5.2.1 Le ricerche empiriche.....	141
5.2.2 Il modello LM-ARCH.....	144
5.2.3 Il modello FIGARCH.....	147
5.2.4 La stima del modello FIGARCH.....	152
5.2.5 Le applicazioni ai tassi di cambio.....	156
5.2.6 Il modello FIEGARCH e i mercati azionari.....	158
5.2.7 Il prezzaggio delle opzioni.....	160
5.2.8 Le estensioni più recenti.....	161
5.3 Il modello HYGARCH.....	162

<b>6. Un'applicazione empirica.....</b>	<b>164</b>
6.1 Un confronto su dati reali.....	165
6.1.1 <i>Analisi preliminare dei dati</i> .....	165
6.1.2 <i>Il modello FIGARCH originale</i> .....	174
6.1.3 <i>La parametrizzazione alternativa</i> .....	183
6.1.4 <i>Previsioni sulla volatilità</i> .....	184
6.2 Conclusioni.....	189
6.2.1 <i>Le previsioni nei mercati finanziari</i> .....	190
6.2.2 <i>L'utilizzo dei modelli a memoria lunga</i> .....	194
6.2.3 <i>Gli sviluppi futuri</i> .....	196

# Appendici

<b>A. Altri strumenti utili.....</b>	<b>198</b>
A.1 Alcune diagnostiche utili.....	199
A.1.1 I criteri informativi.....	199
A.1.2 I test sui residui.....	199
A.2 Metodi di ottimizzazione.....	201
A.2.1 Steepest ascent.....	201
A.2.2 Newton-Raphson.....	203
A.2.3 DFP-BFGS.....	203
A.2.4 BHHH.....	204
A.2.5 Marquardt.....	205
A.2.6 Il controllo del passo.....	206
A.2.7 Il problema dei valori iniziali.....	207

<b>B. Codice sorgente</b> .....	209
B.1 Una panoramica introduttiva.....	210
<i>B.1.1 La programmazione per l'econometria</i> .....	210
<i>B.1.2 Gli ambienti di sviluppo principali</i> .....	210
<i>B.1.3 Il pacchetto G@RCH 1.1</i> .....	211
<i>B.1.4 Il linguaggio C++</i> .....	212
B.2 Codice C++.....	215
<i>B.2.1 La serie della varianza condizionata</i> .....	215
<i>B.2.2 La funzione di verosimiglianza</i> .....	217
<i>B.2.3 Il gradiente analitico</i> .....	218
<i>B.2.4 Il gradiente numerico</i> .....	222
<i>B.2.5 L'Hessiana</i> .....	223
<i>B.2.6 Il main</i> .....	224
<i>B.2.7 Le altre funzioni sviluppate</i> .....	232
<i>B.2.8 Prototipi delle funzioni Numerical Recipes</i> .....	235
B.3 La forma dell'output.....	239
 <b>Bibliografia</b> .....	 241

# **Capitolo 1**

## **Introduzione**

## 1.1 Lo studio delle serie storiche<sup>1</sup>

### *1.1.1 Cenni preliminari*

Da un punto di vista tradizionale l'approccio econometrico si identifica con lo studio dell'insieme di strumenti che l'economista teorico ha a disposizione per misurare, sottoporre a verifica e sviluppare le proprie congetture sulle relazioni tra le variabili economiche. In quest'ottica, l'econometria si configura come una disciplina ancillare dell'economia, nel senso che fornisce a quest'ultima risultati utili e dati importanti, che devono essere però interpretati e inseriti nel contesto della teoria economica. L'econometrico quindi non deve perdere mai di vista gli aspetti economici dei modelli che utilizza, in modo da produrre risultati suscettibili di essere interpretati in maniera soddisfacente dal punto di vista teorico. Tale approccio all'analisi econometrica è noto come "approccio strutturale" e si contrappone al più recente "approccio ateorico", sulla base del quale l'analisi econometrica si slega dalla sottostante teoria economica e si configura come analisi empirica dei fenomeni economici.

Lo studio delle serie storiche fa parte di quest'ultima branca dell'econometria. Una serie storica è costituita da un'insieme ordinato di osservazioni di una certa variabile (economica, nel nostro caso) a diversi intervalli temporali. L'individuazione di regolarità nell'andamento di una serie, e la possibilità di riprodurre le caratteristiche attraverso la formulazione di una relazione che leghi valori passati della serie al valore corrente, stanno alla base dell'analisi delle serie storiche. La metodologia ad essa relativa è stata inizialmente sviluppata in ambito ingegneristico per applicazioni nelle quali era importante separare un segnale da un rumore che si sovrapponeva ad esso. In questa falsariga, l'obiettivo teorico che abbiamo richiamato

---

<sup>1</sup> È difficile riuscire a identificare e citare con precisione i contributi che ci hanno condotto alla stesura di questa prima sezione dell'introduzione; ci limitiamo a dire che si sono lette con attenzione le introduzioni ai testi di Greene (2000), Enders (1995) e Mills (1990), nonché i lavori curati da Granger (1990) e da Gabbi (1999).

poc'anzi sfuma, mentre diventa sempre più rilevante la capacità che il modello ha di prevedere l'andamento futuro della serie.

Per fare un semplice esempio, limitandoci al caso in cui la variabile di interesse sia una sola, consideriamo la produzione industriale italiana nel periodo 1945 – 2000. La teoria economica fornisce alcune indicazioni generali sulle possibili determinanti dell'andamento del fenomeno quali l'andamento dell'occupazione, le condizioni di investimento e della capacità di utilizzo del capitale, le aspettative sulle vendite future, le condizioni del tasso di cambio e della domanda internazionale, e così via. Dal punto di vista del comportamento della serie, tradizionalmente si ipotizza che esso sia la risultante di varie componenti che sono collegabili ad un sentiero di crescita di lungo periodo, alla presenza di congiunture economiche cicliche e stagionali, viste esse stesse come sintesi di comportamenti di altre serie che non sono necessariamente considerate in maniera esplicita.

Ovviamente queste informazioni non sono sufficienti a rappresentare perfettamente l'andamento della produzione, specialmente quando si considerino esigenze di proiezione in avanti di tale andamento: per quante altre variabili si decidano di considerare, ci sarà sempre qualcosa che non si riesce a spiegare e che dipende da eventi assolutamente aleatori e imprevedibili. Nell'approccio econometrico i fattori non esplicitamente considerabili sulla base dell'informazione a disposizione si dicono termini di disturbo o innovazioni.

L'economista potrebbe essere interessato ad utilizzare il modello econometrico per valutare l'impatto e l'importanza relativa di ciascuna delle sue determinanti nello spiegare la variabile o le variabili di interesse, o inserirne di nuove; lo studioso di serie storiche si occupa invece di qualcosa di parzialmente diverso: cerca di scomporre la serie nelle sue componenti e di prevedere, sulla base di questa scomposizione, l'andamento futuro della variabile considerata. In entrambi i casi il termine di disturbo viene identificato attraverso una componente residuale della stima attraverso la quale si possono trarre elementi di giudizio sulla bontà della



specificazione prescelta: l'informazione contenuta nelle variabili deve essere trasferita nella capacità esplicativa del modello e non deve essere possibile isolare informazione rilevante nella componente residuale.

### *1.1.2 Il ruolo delle previsioni*

Al di là della capacità di un modello di spiegare o dar conto del comportamento di una variabile, la sua qualità può essere giudicata sulla base della possibilità di estrapolare l'informazione relativa al passato in una diagnosi sul comportamento futuro della serie. Questa differente impostazione assume particolare rilevanza poiché si può osservare che non necessariamente la modellazione migliore dal punto di vista teorico (cioè, in ultima analisi, quella che considera l'insieme più ampio di variabili) dà le previsioni più attendibili. Ad esempio, il rischio che si corre nell'utilizzare un ampio numero di variabili è quello della *sovraparametrizzazione*: il modello descrive in maniera molto precisa l'andamento della serie all'interno del campione considerato, ma fornisce previsioni poco attendibili.

Nel testo di Box e Jenkins (1976), che segna un punto di svolta fondamentale per lo studio delle serie storiche, si osserva che, ai fini della previsione, non è rilevante considerare un numero elevato di variabili esogene, ma è sufficiente osservare l'andamento passato della serie che si sta analizzando.

L'ipotesi fondamentale a questo proposito è che la serie, una volta liberata dal trend di lungo periodo e dalla ciclicità, dipenda dai propri valori passati (*AutoRegressive*) e dai disturbi che si sono osservati in passato (*Moving Average*). In questa maniera si riescono a costruire modelli estremamente parsimoniosi, e quindi semplici da trattare e da stimare, che ciononostante hanno un comportamento molto buono per quanto riguarda le previsioni. Questa classe di modelli prende il nome di ARMA (*AutoRegressive – Moving Average*).

Vale la pena soffermarci meglio sul concetto di utilizzo di un modello a fini previsivi attraverso un esempio. Supponiamo che la teoria economica ci dica che una certa variabile, ad esempio la produzione, dipenda, al tempo  $t$ , dal tasso di interesse in  $t$ . Se volessimo prevedere il livello della produzione in  $t + 1$  ci troveremmo in difficoltà, poiché ovviamente in  $t$  non ci è noto quanto varrà il tasso di interesse in  $t + 1$ .

Ipotizziamo invece che il valore della produzione in  $t + 1$  sia uguale a quello della produzione in  $t$  a meno di un'innovazione casuale nel processo che genera la produzione che si verifica fra  $t$  e  $t + 1$ : in questo caso si potrà avere una previsione più sensata, perché sulla base dell'informazione disponibile in  $t$ , la produzione attesa in  $t + 1$  sarà data dalla produzione attuale più il valore atteso del termine di disturbo. Questo modello è estremamente semplice, ma è assai poco interessante dal punto di vista teorico/strutturale: la macroeconomia che studia interdipendenze non è arricchita dalla verifica di questa ipotesi se non nella constatazione che le innovazioni o *shock* aggregati alla produzione si accumulano successivamente ed hanno un impatto permanente, dato che, a partire da un valore iniziale della produzione, valori futuri sono semplicemente accumulazioni di *shock* occorsi a tempi successivi. Se questo è il meccanismo che regola il comportamento della serie, risulta complicato rintracciare relazioni tra variabili economiche o la spiegazione di un certo fenomeno economico: da un punto di vista previsivo, poi, questo modello ci dice semplicemente di attenderci un valore futuro della produzione pari a quello attuale.

Evidentemente, per un economista potrebbe essere molto più interessante sapere che esiste, e che ha una certa forza, una relazione che lega tra loro diverse variabili economiche, soprattutto se l'economista in questione è un *policy-maker* che desidera sapere come manipolare il valore della variabile di interesse a partire da alcuni strumenti sui quali egli ha potere di intervento. È ovvio quindi osservare che sarà più importante sapere se esiste una relazione fra tasso di interesse e livello della produzione e quale sarà il cambiamento da

attendersi nella produzione a fronte di una variazione di un punto percentuale nel tasso di interesse: poiché sarà lui stesso a fissare il tasso di interesse futuro, il *policy-maker* quanto meno sarà in grado di formulare degli scenari di previsione sull'andamento futuro della produzione a seconda dei valori prescelti dello strumento<sup>2</sup>.

Ci sono però molte situazioni nelle quali:

- Le previsioni sull'andamento futuro di una certa grandezza sono assai più rilevanti delle implicazioni sulla teoria economica del modello;
- Non si ha alcun modo di influire, attraverso il nostro comportamento, su detto andamento.

Queste situazioni si verificano ad esempio sui mercati finanziari cosiddetti efficienti<sup>3</sup>: tutta l'informazione disponibile è contenuta nei prezzi, per cui essi sono l'unica grandezza rilevante, e nessun operatore è in grado di influire sull'andamento del mercato. Si pensi, per fare un esempio, al gestore di un certo portafoglio azionario: egli non è senz'altro in grado di influenzare, con le proprie operazioni, l'andamento del mercato, ma deve comunque cercare di prevederlo, al fine di massimizzare l'incremento di valore del proprio portafoglio, utilizzando al meglio l'informazione attualmente disponibile. Dati questi obiettivi, è evidente come egli possa essere assai poco interessato agli aspetti teorici del modello che utilizza ai fini previsivi.

In situazioni come questa si rivela l'importanza di un modello che, anche se poco appetibile dal punto di vista teorico, sia suscettibile di fornire le previsioni più accurate sfruttando al meglio l'informazione disponibile, cioè l'andamento passato della serie. In questo senso la teoria passa in secondo piano, e diventa invece molto più rilevante lo studio dei dati, tant'è

---

<sup>2</sup> Non ci si addentrerà in questa discussione nella cosiddetta critica di Lucas, vale a dire nelle questioni relative all'impatto della manipolazione degli strumenti di politica economica sulle aspettative degli agenti e quindi sulla costanza dei parametri di un modello utilizzato per analisi di politiche.

<sup>3</sup> In questo caso è sufficiente ipotizzare efficienza debole; per una trattazione più approfondita di questo tema si rimanda a Caparrelli (1995).

che alcuni indicano lo studio delle serie storiche economiche col termine “macroeconomia empirica”.

I passi che si percorrono se si segue tale approccio sono i seguenti:

1. Si esegue l’analisi esplorativa dei dati per individuare *fatti stilizzati*;
2. Si procede quindi alla specificazione di un opportuno modello statistico che riesca a rappresentare tali caratteristiche empiriche;
3. Si stimano poi i parametri, eseguendo l’analisi dei residui di stima;
4. Si completa la diagnostica sul modello mettendone alla prova la capacità previsiva su osservazioni non appartenenti al periodo campionario su cui si è effettuata la stima (*out-of-sample*).

### *1.1.3 I rapporti con le scienze sperimentali*

Questa prospettiva empirica avvicina di molto lo studio delle serie storiche alle scienze sperimentali, in particolare alla fisica e all’ingegneria<sup>4</sup>. Così come il fisico osserva un fenomeno naturale e cerca di descriverne il comportamento attraverso una relazione matematica, lo studioso di serie storiche, a partire da un insieme di dati, li interpreta come realizzazioni di variabili casuali e cerca una formulazione matematica che leghi fra loro le variabili riferite a istanti temporali diversi (processo stocastico) che può averli generati.

Questa prende il nome di processo generatore dei dati, o DGP (*Data Generating Process*).

Ovviamente, così come avviene per la fisica, una rappresentazione assolutamente perfetta è impensabile, e il modello che costruiremo sarà al più una *proxy* del vero processo generatore dei dati; dovrà comunque riuscire a catturare le dinamiche principali della serie che studiamo.

---

<sup>4</sup> Esiste un intero filone di ricerca in questa direzione, che prende il nome di “econofisica” (Mantegna e Stanley, 2000; per ulteriori approfondimenti si rimanda alla consultazione dell’URL <http://www.econophysics.org>). L’ovvia differenza è rappresentata dal fatto che i fenomeni studiati dalle scienze naturali sono riproducibili in laboratorio, mentre i fenomeni economici, per ovvie ragioni, non sono “sperimentabili” ma solo osservabili nella realtà. Si parla pertanto di “scienze osservative”.

Abbiamo già rilevato che i modelli di serie storiche si rivolgono, oltre che agli accademici che ne studiano le caratteristiche e le proprietà, a chi ha intenzione di utilizzarli per fini previsivi, ad esempio gli analisti finanziari. Chi desidera previsioni su una certa grandezza è interessato a che il modello possieda due qualità fondamentali: la semplicità e la precisione (non sempre raggiungibili contemporaneamente, come spesso succede).

Vale la pena avere chiaro cosa si intende per semplicità: non ci si riferisce semplicemente all'indubbio vantaggio di avere un modello con poche variabili e formulato in modo da essere intelligibile anche a un "profano", quanto piuttosto ai problemi applicativi che una rappresentazione più complessa porta con sé: si pensi alla ricerca di uno stimatore adeguato, o alla lentezza computazionale che può provocare una specificazione complessa.

#### *1.1.4 La struttura della disciplina*

In generale, lo studio delle serie storiche può essere pensato come una costruzione in cui si parte da modelli estremamente semplici e facili da trattare (ad esempio la famiglia ARMA) e si aggiungono complicazioni, casi particolari e, soprattutto, generalizzazioni. Per fare un esempio<sup>5</sup>, si è arrivati alla definizione dei modelli GARMA, che tratteremo nel paragrafo 3.6.3, attraverso:

- Il modello ARMA tradizionale;
- L'introduzione dell'integrazione frazionaria (generalizzazione ARFIMA);
- La sostituzione<sup>6</sup> dell'operatore  $(1 - L)^d$  col polinomio di Gegenbauer  $(1 - 2\xi L + L^2)^d$ .

Dall'esempio basti cogliere non solo la generalizzazione che corrisponde alla necessità di riprodurre fatti stilizzati più complessi, ma anche il fatto che dalla formulazione più generale è possibile risalire a quella più semplice introducendo opportune restrizioni. L'interesse di tali

---

<sup>5</sup> Il seguente esempio ha funzioni puramente illustrative, ci si perdonerà perciò l'utilizzo di concetti, come l'integrazione frazionaria e il polinomio di Gegenbauer, che non sono stati ancora presentati.

vincoli dal generale al particolare è che essi possono essere sottoposti a verifica empirica con procedure statistiche formali. Una rappresentazione più generale è quindi più varia e più flessibile, per cui può descrivere il processo generatore dei dati in maniera più precisa; perde però in semplicità. Per valutare se la generalizzazione sia opportuna, dobbiamo perciò ragionare in termini di costi e benefici, e chiederci in particolare:

- la generalizzazione che si è introdotta dà risultati sensibilmente diversi rispetto a quella tradizionale? Riesce a cogliere delle dinamiche che quest'ultima non coglieva? In ultima analisi, è più precisa? Per fare un esempio, se una volta definiti i processi GARMA si fosse osservato che, sui dati reali, il parametro  $\xi$  valeva sistematicamente uno, non ci sarebbe stato un gran bisogno di generalizzare il modello ARFIMA (lo stesso si dica se l'esponente  $d$  non risulti statisticamente diverso da 0 o 1).
- Quanto costa la maggior precisione in termini di semplicità applicativa, nel senso che abbiamo visto sopra? Avere un buon modello sulla carta serve a poco se, per varie ragioni, si hanno problemi insormontabili per quanto riguarda l'applicazione ai dati reali.

Siamo convinti che, sulla base di questi concetti, si possa comprendere più chiaramente la filosofia di lavoro che ha ispirato questa ricerca.

---

<sup>6</sup> Si osserva facilmente, infatti, che per  $\xi = 1$ , il polinomio di Gegenbauer vale  $(1 - L)^{2d}$ .

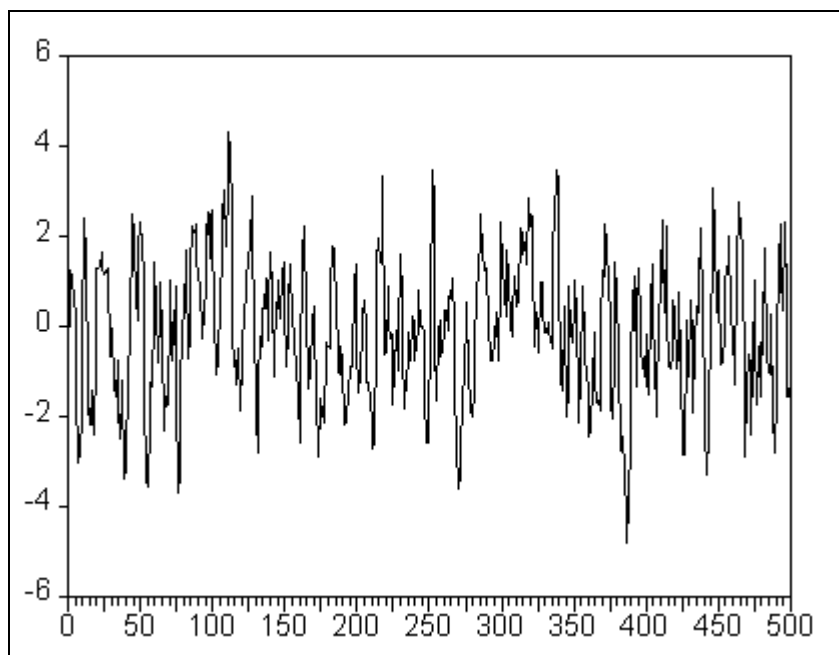
## 1.2 La nascita dell'integrazione frazionaria

L'interesse per i processi a integrazione frazionaria è nato negli anni '50 nel campo dell'ingegneria, in particolare nell'idrologia: il primo studio fu proposto da Hurst (1951) a proposito delle proprietà di riempimento dei bacini idrici. Anche se il primo studio applicato a dati finanziari è dovuto a Mandelbrot (1971), questo tipo di modellistica ha iniziato a destare vero interesse negli econometrici solo all'inizio degli anni '80.

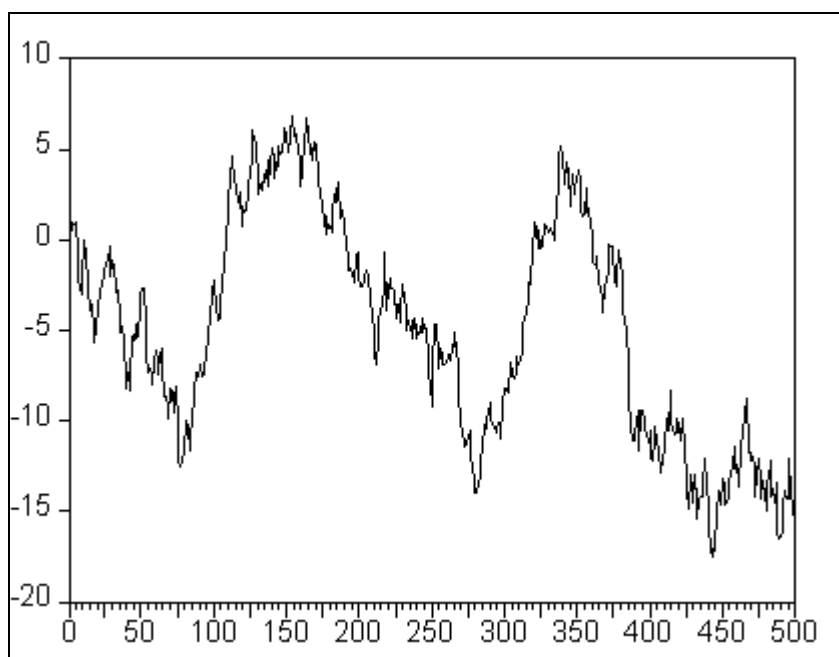
### *1.2.1 La modellistica tradizionale*

Fino al 1980 lo strumento d'elezione per gli studi di serie storiche economiche era la modellistica ARMA; tale classe di modelli presenta, insieme all'indubbio pregio della semplicità, alcune grosse rigidità.

Prima di procedere oltre, vale la pena introdurre alcuni concetti. Un processo si dice stazionario se le proprietà delle variabili casuali che lo compongono (in particolare la loro distribuzione di probabilità marginale o congiunta) non dipendono dall'istante temporale a cui le si analizzano. Si possono avere le idee più chiare confrontando la figura 1.1 con la figura 1.2:



**Figura 1.1**  
*Una serie stazionaria*



**Figura 1.2**  
*Una serie non stazionaria*



I processi stazionari godono di alcune proprietà particolarmente interessanti che vedremo in seguito; tuttavia in buona parte delle serie economiche (si pensi ad esempio al livello del PNL) la presenza di un sentiero di crescita di lungo periodo valutabile visivamente suggerisce la presenza di non stazionarietà. Tradizionalmente, la presenza di una componente tendenziale è stata interpretata in modo deterministico e la rimozione del trend effettuata regredendo i valori della variabile rilevati a diversi istanti temporali su un polinomio di una variabile ottenuta come sequenza crescente di numeri naturali in corrispondenza dei diversi istanti e studiandone i residui di stima (componente de-trendizzata). Più recentemente, lo studio dei processi non stazionari di tipo passeggiata aleatoria (*random walk*) ha mostrato come l'evidenza empirica sia coerente anche con la presenza di un *trend* stocastico ottenuto, come menzionato prima, dalla somma di innovazioni (o *shock*) realizzatesi ai diversi istanti temporali. La rimozione della componente tendenziale in questo caso si opera lavorando su variabili casuali che sono ottenute come differenze successive delle variabili casuali originarie. Le conseguenze della presenza dell'uno o dell'altro tipo di trend non sono di poco conto, soprattutto se si adotta un modo di eliminare la componente tendenziale che non è adeguato a quella presente nel DGP. Infatti, se il DGP contiene un trend deterministico, applicare le differenze prime per eliminare il trend corre il rischio di inserire una componente ciclica spuria nel termine di disturbo, mentre se si utilizza il metodo della regressione per rimuovere un trend stocastico presente nel DGP, le proprietà degli stimatori e quindi la qualità delle valutazioni di tipo inferenziale seguono caratteristiche non standard che se ignorate portano a risultati non corretti.

Il caso che si vuole analizzare più in dettaglio in quel che segue è quello della presenza di uno o più trend stocastici. Si ricorre al concetto di integrazione di un certo ordine per un processo per indicare il numero di volte che si deve applicare l'operatore differenze prime per ottenere un processo stazionario. Formalmente, un processo si dice integrato di ordine  $n$ , e si indica

con la notazione  $I(n)$ , se, differenziato  $n$  volte, genera un processo stazionario. Per estensione si attribuisce l'aggettivo di “integrato” o “stazionario” anche alla serie osservata di realizzazioni del processo. La serie della figura 1.2, ad esempio è  $I(1)$ : infatti è stata generata dall'equazione  $y_t = y_{t-1} + \varepsilon_t$ , dove  $\varepsilon_t$  è distribuito normalmente con media nulla, varianza unitaria e incorrelato. Se calcoliamo le differenze prime di questa equazione (in seguito si utilizzerà il termine “differenziare la serie”) otteniamo  $y_t - y_{t-1} = \varepsilon_t$ , e questi residui, essendo indipendenti e identicamente distribuiti, sono chiaramente stazionari. Qualora gli  $\varepsilon_t$  abbiano struttura più complessa (ad esempio, siano un processo ARMA<sup>7</sup>), la presenza di integrazione rende il processo più generale definito come ARIMA (*AutoRegressive – Integrated – Moving Average*), e sono stati introdotti pressoché simultaneamente agli ARMA ordinari.

Come si è menzionato in precedenza, tuttavia, non tutti i fenomeni possono essere descritti da modelli semplici, e, in effetti, alcune serie presentano difficoltà di adattamento all'ipotesi che siano generate da un processo ARMA. In particolare, in alcuni casi a fronte di autocorrelazioni<sup>8</sup> teoriche, che in un modello ARMA decadono esponenzialmente<sup>9</sup>, quelle empiriche calcolate a partire dai valori di una serie seguono un sentiero di discesa molto più “morbido”, suggerendo l'inadeguatezza della semplicità del modello adottato.

In sostanza, la conseguenza dell'adattare un modello ARMA a dati che presentano tale *pattern* nell'autocorrelazione empirica è che il processo “dimentica il proprio passato” in modo più lento rispetto a quello che ci si sarebbe aspettato.

---

<sup>7</sup> Le definizioni più formali dei processi introdotti discorsivamente in questa introduzione saranno proposte nel corpo del lavoro.

<sup>8</sup> L'autocorrelazione misura, per diversi valori di ritardo  $k$ , la dipendenza delle osservazioni della serie da quelle ad esse precedenti di  $k$  periodi.

<sup>9</sup> Si rimanda, per la dimostrazione, al paragrafo 2.3.3.

### 1.2.2 Gli studi sulla densità spettrale

Altre incongruenze tra i risultati ottenuti dai dati reali e quelli che ci si sarebbero attesi sulla base della modellistica esistente provenivano dagli studi sulla densità spettrale, che gli analisti di serie storiche avevano iniziato a utilizzare all'inizio degli anni '60.

In prima approssimazione<sup>10</sup> si può dire che l'analisi spettrale si occupa di scomporre la serie in cicli di diversa frequenza, e di analizzarne la dipendenza sulla base dell'andamento di una funzione che descrive il contributo di ciascuno di questi cicli all'andamento complessivo della serie; tale funzione prende il nome di densità spettrale. L'area racchiusa sotto la funzione di densità spettrale misura la varianza della serie, mentre le aree "parziali" indicano la parte della varianza complessiva che è imputabile a cicli di frequenza inferiore a quella considerata.

Sulla base di tale risultato possiamo osservare che:

- un processo in cui le osservazioni sono indipendenti (ad esempio un *white noise*) deve avere spettro piatto;
- la varianza di un processo che ha densità spettrale alta a frequenza basse è dovuta in larga parte a cicli di bassa frequenza, per cui la dipendenza di lungo termine del processo è più forte di quella di breve termine;
- la densità spettrale di un processo ciclico ha un picco in corrispondenza della frequenza del ciclo stesso;
- lo spettro di processi non stazionari è esplosivo per frequenze tendenti a zero.

Nel corso del tempo erano stati stimati gli spettri di diverse serie economiche; in particolare, Granger e Morgenstern (1963) avevano focalizzato la loro analisi sui prezzi azionari.

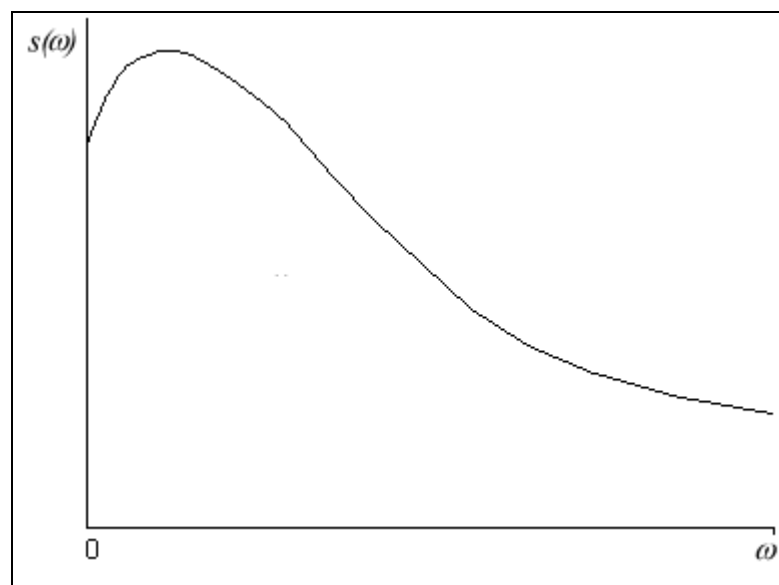
Granger (1966) cercò di fare il punto della situazione, confrontando i risultati di queste stime e cercando caratteristiche comuni. Ciò che venne ricavato da tale confronto fu che in tutte le

---

<sup>10</sup> L'argomento è piuttosto complesso, e una trattazione non formalizzata rischia di risultare ellittica e poco chiara; si è deciso comunque di cimentarsi con l'argomento per dare al lettore almeno un'idea di massima della questione. Per una trattazione più approfondita si rimanda comunque alla sezione 2.4.

serie analizzate la maggior parte della varianza è spiegata da cicli di frequenza bassa, il che indica che si ha una forte dipendenza delle osservazioni da quelle passate. Non si tratta, a dire il vero, di un risultato sorprendente in sé, visto che si ritiene che buona parte delle serie economiche possiedano dipendenza di questo tipo.

Il risultato insolito è che utilizzando gli strumenti ARMA tradizionali non si riusciva a liberare la serie dalla dipendenza (e a restituire quindi uno spettro dei residui piatto). La densità spettrale dei residui appariva infatti del tipo disegnato nella figura 1.3.



**Figura 1.3**

*Lo spettro dei residui di una generica variabile economica  
filtrata al meglio con la modellistica tradizionale  
(da Granger, 1966)*

Si può osservare che la densità spettrale resta forte alle basse frequenze, segno che permangono relazioni di lungo periodo che la modellistica ARMA non è stata in grado di cogliere. Si rendono quindi necessari modelli più adatti a cogliere tale dinamica di lungo termine.

### 1.2.3 Nuovi modelli e nuove teorie

Sulla scia di queste problematiche appariva chiara la necessità di una modellistica più flessibile e più adeguata ai risultati empirici. Serviva in particolare una modellazione che, pur preservando proprietà importanti come la stazionarietà, offrisse maggior aderenza ai dati reali, sia per quello che riguarda la memoria lunga, sia di conseguenza per la forma dello spettro.

Come abbiamo detto, i modelli di questi tipo già esistevano; si trattava quindi di adattarli alle necessità dell'analisi delle serie storiche: in particolare, trasformare il *fractional Brownian motion* di Mandelbrot e Van Ness (1968), definito in tempo continuo, in un processo discreto.

Tale lavoro fu realizzato indipendentemente da Granger e Joyeux (1980) e Hosking (1981).

Si considera dunque la possibilità che l'indice di integrazione dei processi ARIMA assuma valori non interi e si introduce la nuova classe di processi ARFIMA ( $p, d, q$ ), i quali costituiscono quindi una generalizzazione degli ARIMA.

Granger e Joyeux (1980) e Hosking (1981) derivarono le proprietà principali di questa classe di processi, cioè la memoria lunga e la stazionarietà per  $d < 1/2$ . Tali processi si definiscono  $I(d)$ , cioè integrati di ordine  $d$ . Poiché  $d$ , come abbiamo detto, può assumere valori non interi, si parla anche di integrazione frazionaria.

Contemporaneamente, Granger (1980) mostrava come un processo ARFIMA potesse sorgere dall'aggregazione di più processi AR: la comunità scientifica si trovò a disporre, quasi contemporaneamente, di un modello adeguato ai dati e di una spiegazione teorica sull'occorrenza della memoria lunga.

Vale la pena rilevare l'importanza che assume questo risultato: se i processi a integrazione frazionaria sorgono aggregando altri processi, se ne comprende l'enorme importanza non solo ai fini della modellazione (e quindi nell'interesse dello studio delle serie storiche), ma dell'analisi macroeconomica in generale. La quasi totalità delle grandezze analizzate dai macroeconomisti sono difatti frutto di aggregazioni: si pensi al PNL (generato aggregando i

singoli contributi alla produzione nazionale di ciascuna impresa), o all'inflazione (generata aggregando l'incremento dei prezzi di più beni di consumo).

## 1.3 Applicazioni empiriche

A questo punto rimane da chiedersi se effettivamente i dati presentino evidenza di integrazione frazionaria. Dopo aver superato gli ostacoli posti dalla questione della stima del coefficiente  $d$  (si veda in proposito il capitolo 4), l'attenzione si è concentrata sulla stima ARFIMA di varie serie<sup>11</sup>.

### *1.3.1 Applicazioni macroeconomiche*

Le prime applicazioni dei modelli a integrazione frazionaria a dati macroeconomici riguardano l'evoluzione nel tempo del PNL reale, in particolare il dibattito se tale variabile sia  $I(1)$  oppure  $I(0)$  con trend deterministico. Studi in questa direzione sono stati portati avanti da Diebold e Rudebusch (1989) e Sowell (1992b).

Ancora più interessanti sono i risultati ottenuti relativamente agli indici dei prezzi al consumo da Baillie, Chung e Tieslau (1996). Si è infatti osservato che i dati di tutti i paesi osservati, meno quelli del Giappone, presentavano evidenza di comportamento  $I(d)$ .

### *1.3.2 Applicazioni a serie finanziarie ad alta frequenza*

Molto più interessanti per questo lavoro sono però i risultati che si sono ottenuti nella modellazione a integrazione frazionaria delle serie finanziarie ad alta frequenza. Studi sono stati condotti nel campo dei tassi di interesse (Crato e Rothman, 1994, Tsay, 2000), ma soprattutto in quello dei rendimenti giornalieri assoluti o quadrati dei mercati azionari e dei cambi. Le numerose ricerche in questo campo (si vedano ad esempio Crato e de Lima, 1994, Baillie e Bollerslev, 1994, Granger e Ding, 1996a e 1996b, Baillie, Bollerslev e Mikkelsen, 1996, Bollerslev e Mikkelsen, 1996, Cheung, 1996, Andersen e Bollerslev, 1997a) hanno

messo in evidenza la presenza di integrazione frazionaria nei rendimenti assoluti e nella varianza condizionata.

### *1.3.3 Introduzione alla modellistica ARCH a memoria lunga*

I modelli ARCH sono stati introdotti da Engle (1982) per spiegare la persistenza che si era osservata in numerose serie finanziarie sui rendimenti assoluti o sui rendimenti al quadrato.

Poiché in una serie a media nulla il valore atteso del quadrato del termine di disturbo è uguale alla varianza condizionata, si ipotizza che la varianza condizionata segua un processo di tipo autoregressivo, e che quindi sia prevedibile. Ciò significa, in termini più concreti, che utilizzando questo tipo di strumenti si possono fare previsioni sulla volatilità futura dei rendimenti di un certo attivo finanziario. È evidente l'importanza che questo risultato ricopre, specie per tutti gli operatori professionali (analisti finanziari, cambisti, gestori di portafogli) che si trovano a dover stimare o gestire il rischio di determinati attivi finanziari.

Le ricerche che hanno evidenziato la presenza di integrazione frazionaria nei rendimenti al quadrato dei mercati finanziari, menzionate nel paragrafo precedente, hanno portato all'applicazione dell'integrazione frazionaria alla modellistica ARCH di Engle (1982) e GARCH<sup>12</sup> di Bollerslev (1986): sono stati proposti i modelli LM-ARCH (*Long Memory ARCH* – Ding e Granger, 1996) e FIGARCH (*Fractionally Integrated GARCH* – Baillie, Bollerslev e Mikkelsen, 1996).

Questo filone rappresenta a tutt'oggi l'applicazione più interessante e più promettente dell'integrazione frazionaria. Difatti l'utilizzo di un modello a integrazione frazionaria ci consente non solo di avere una modellazione più precisa e più aderente alla realtà, ma

---

<sup>11</sup> Ci concentreremo qui su dati di tipo economico, anche se abbiamo già avuto modo di rilevare che la modellistica a memoria lunga origina nelle scienze ideologiche.

<sup>12</sup> I modelli GARCH sono una semplice estensione degli ARCH; si rimanda per ulteriori approfondimenti al paragrafo 5.1.2.



soprattutto, grazie alla memoria lunga, di avere stime a più lungo termine sulla volatilità di una certa serie. I vantaggi operativi che tale risultato può portare sono assai rilevanti.

Per rendersene conto basta passare in rassegna tutte quelle situazioni in cui la modellistica ARCH tradizionale viene utilizzata per stimare la volatilità. Per fare alcuni esempi particolarmente significativi: calcolo del *Value at Risk* di un certo portafoglio, prezzaggio delle opzioni secondo Black e Scholes (1973) o Hull e White (1987). L'importanza della modellistica a memoria lunga per la varianza condizionata è evidente: se cerchiamo di prevedere la volatilità su un orizzonte lungo utilizzando un tradizionale ARCH otterremo, a causa della memoria corta, una previsione poco distante o addirittura coincidente con la varianza non condizionata; ciò significa, all'atto pratico, trascurare l'eteroschedasticità. Al contrario, sfruttando la proprietà di memoria lunga dei modelli che proporremo, si potrà ottenere una previsione più significativa e più sensibile alle osservazioni passate.

Quella delle previsioni sulla volatilità è una tematica particolarmente interessante sia per gli accademici che per gli operatori professionali. Se con l'introduzione della modellistica ARCH si è potuto gettare uno sguardo sulla dinamica di breve periodo della varianza condizionata, con l'utilizzo dell'integrazione frazionaria possiamo fare un passo avanti e volgere l'attenzione al medio e lungo periodo.

## 1.4 Piano e obiettivi del lavoro

I modelli a integrazione frazionaria hanno ormai raggiunto una diffusione considerevole. La letteratura che si è sviluppata sull'argomento si sta ingrandendo a ritmo sostenuto e si avvia a raggiungere una dimensione imponente. La prima parte di questo lavoro sarà pertanto volta a tracciare un quadro dell'ambiente entro al quale ci muoviamo.

Nel prossimo capitolo<sup>13</sup> richiameremo alcuni strumenti matematici e statistici che ci saranno d'ausilio nella presentazione della materia: introdurremo brevemente i concetti di processo stocastico, stazionarietà, integrazione e analisi spettrale, richiamando, quando è necessario, gli strumenti matematici non elementari che sono necessari per affrontare ciascuno di questi argomenti.

Nel terzo capitolo entreremo nel vivo della materia: ci occuperemo in primo luogo di chiarire in maniera più formale quello che si intende per memoria lunga, e definiremo il processo a integrazione frazionaria più semplice, cioè il *fractional white noise*, di cui saranno presentate dettagliatamente le proprietà (Hosking, 1981 e Granger e Joyeux, 1980). In seguito ci occuperemo dell'applicazione del *fractional white noise* ai processi ARMA, e definiremo la nuova classe di processi ARFIMA, per la quale presenteremo brevemente le proprietà; infine si introdurrà una generalizzazione del *fractional white noise* nota come processo di Gegenbauer (Gray, Zhang e Woodward, 1989) e, di conseguenza, la classe generalizzata GARMA.

Presenteremo quindi alcune metodologie di test per la presenza di integrazione frazionaria, e introdurremo infine i modelli teorici che sono stati proposti per spiegare l'insorgenza di integrazione frazionaria e i risultati ottenuti tramite le più rilevanti ricerche empiriche.

---

<sup>13</sup> A piè di pagina, sul frontespizio di ciascun capitolo, si è ritenuto opportuno, per chiarezza espositiva, riportare un breve sommario del suo contenuto.

Nel capitolo 4 presenteremo in primo luogo alcuni richiami che riteniamo utili sul concetto di stima di massima e quasi-massima verosimiglianza, rimandando alla sezione A.2 in appendice la descrizione sommaria del funzionamento dei principali metodi numerici di massimizzazione per la funzione di verosimiglianza. Una volta chiariti questi concetti saremo pronti per affrontare le problematiche relative alla stima dei processi a integrazione frazionaria. Mostreremo in primo luogo l'applicazione alle serie  $I(d)$  della massima verosimiglianza condizionata (Hosking, 1984a) e la forma della funzione di verosimiglianza esatta (Sowell, 1992a). Presenteremo poi le metodologie di stima più specifiche, ricavate appositamente per le serie a integrazione frazionaria: la massima verosimiglianza sul dominio della frequenza (Fox e Taquq, 1986) e la stima semiparametrica del parametro  $d$  (Geweke e Porter-Hudak, 1983); per finire tratteremo brevemente la stima dei processi GARMA (Chung, 1996a). Sarà infine presentata un'applicazione delle metodologie di stima sopra indicate su dati reali.

Nel capitolo quinto, dopo aver richiamato i risultati e le proprietà principali dei modelli ARCH (Engle, 1982), procederemo presentando le ricerche che hanno mostrato evidenza di integrazione frazionaria nei residui assoluti o nei residui al quadrato. Presenteremo quindi il modello *Long Memory ARCH* (LM-ARCH) di Ding e Granger (1996) e il *Fractionally Integrated GARCH* (FIGARCH) di Baillie, Bollerslev e Mikkelsen (1996), nonché il FIGARCH modificato di Chung (1999). Si affronteranno le questioni relative alla stima del modello FIGARCH e si presenteranno le prime ricerche che hanno applicato tale modello a dati reali.

Il passo finale di questo lavoro sarà l'applicazione del modello FIGARCH al tasso di cambio dollaro/euro: si presenteranno i risultati ottenuti tramite le stime GARCH tradizionali e le si confronteranno con quelle ottenute utilizzando il modello FIGARCH. In particolare, vedremo gli effetti che ha l'utilizzo del modello FIGARCH piuttosto che il GARCH

tradizionale sulle previsioni della volatilità a breve e lungo termine e sul calcolo del *Value at Risk*<sup>14</sup> di un ipotetico portafoglio.

Gli obiettivi di questo lavoro sono, come si sarà già capito, in larga parte metodologici. Quello che ci è sembrato prioritario è stato tracciare un quadro il più possibile chiaro e completo dell'integrazione frazionaria e mostrare come procedere operativamente per condurre una stima FIGARCH. In questo senso le applicazioni a dati reali che si proporranno si configureranno come utili strumenti illustrativi.

---

<sup>14</sup> Il *Value at Risk*, o *VaR*, è un indicatore utilizzato per la misura e la copertura del rischio dei portafogli finanziari. Per una descrizione più approfondita si rimanda al paragrafo 5.1.8.

## Capitolo 2

# Richiami matematici e statistici

In questo capitolo richiameremo in primo luogo alcuni risultati matematici assolutamente indispensabili per la comprensione di ciò che presenteremo in seguito: primi su tutti la funzione  $\Gamma$  e i numeri complessi.

Richiameremo quindi brevemente i principi dell'analisi delle serie storiche: introdurremo il concetto di processo stocastico e suo valore atteso, varianza e autocovarianza, per poi presentare gli strumenti più semplici per l'analisi delle serie storiche, cioè la famiglia di processi ARIMA. Introdurremo il concetto di stazionarietà forte e in covarianza e quello di integrazione; si tratteranno infine le tecniche di analisi spettrale.

Come abbiamo già detto, in questo capitolo ci occuperemo di introdurre alcuni strumenti, sia matematici che statistici, che saranno richiamati e utilizzati nel seguito del nostro lavoro. Si tratta prevalentemente di strumenti standard e ben noti per cui si è seguita un'impostazione di tipo manualistico.

Nonostante la lettura di questo capitolo sia del tutto superflua per un "addetto ai lavori", abbiamo ritenuto comunque utile presentarla. Considerando un lettore che sia uso ai principi fondamentali della statistica inferenziale, dell'econometria classica, dell'analisi e dell'algebra lineare, esso potrà giungere alla piena comprensione di questo lavoro semplicemente assimilando i pochi strumenti che saranno presentati in questo capitolo: funzioni speciali, numeri complessi, fondamenti sulle serie storiche e analisi spettrale.

Il lettore che è già uso a questi strumenti potrà evitare la lettura di questo capitolo, utilizzandolo come riferimento immediato quando, nel corso del lavoro, ci riferiremo a formule in esso presentate.

## 2.1 Funzioni utili<sup>1</sup>

### 2.1.1 La funzione Gamma

Definiamo la funzione  $\Gamma$ , applicazione di  $\mathbf{R}$  in  $\mathbf{R}$ , come:

$$\Gamma(z) = \int_0^{+\infty} t^{z-1} e^{-t} dt, \quad (2.1.1)$$

che per  $n$  naturale verifica

$$\begin{aligned} \Gamma(n+1) &= n!, \\ \Gamma\left(n + \frac{1}{2}\right) &= \frac{1 \cdot 3 \cdot 5 \cdots (2n-1)}{2^n} \sqrt{\pi}. \end{aligned}$$

Si noti che, nella (2.1.1),  $\Gamma$  è limitata solo per  $z > 0$ . Per estendere la funzione a tutto  $\mathbf{R}$ , utilizziamo la riflessione:

$$\Gamma(1-z) = \frac{\pi}{\Gamma(z) \operatorname{sen}(\pi z)} = \frac{\pi z}{\Gamma(1+z) \operatorname{sen}(\pi z)}, \quad (2.1.2)$$

Vale la pena osservare che la funzione  $\Gamma$  è un mattone fondamentale per l'inferenza statistica.

La funzione di densità di probabilità per una distribuzione  $\chi_r^2$  è infatti

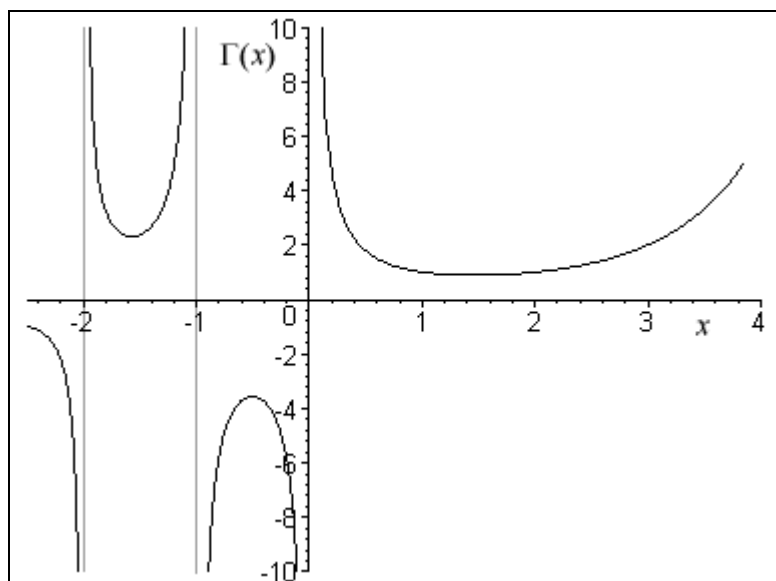
$$f_x(r) = \frac{x^{\frac{r}{2}-1} e^{-\frac{x}{2}}}{2^{\frac{r}{2}} \Gamma\left(\frac{r}{2}\right)},$$

e similmente troviamo la  $\Gamma$  nelle funzioni di densità di probabilità per le distribuzioni  $t$  e  $F$ .

Nella figura 2.1 è disegnata  $\Gamma(x)$  in un intervallo ridotto: possiamo osservare le discontinuità per ogni  $z$  intero minore di 1, compreso lo zero.

---

<sup>1</sup> Si è seguito prevalentemente Mood, Graybill e Boes (1974).



**Figura 2.1**

*La funzione Gamma*

Si noti che i valori di  $\Gamma$  possono essere ottenuti per via ricursiva, integrando per parti:

$$\Gamma(z+1) = z\Gamma(z). \quad (2.1.3)$$

Un'altra proprietà interessante di  $\Gamma$  è la seguente, nota come approssimazione di Stirling:

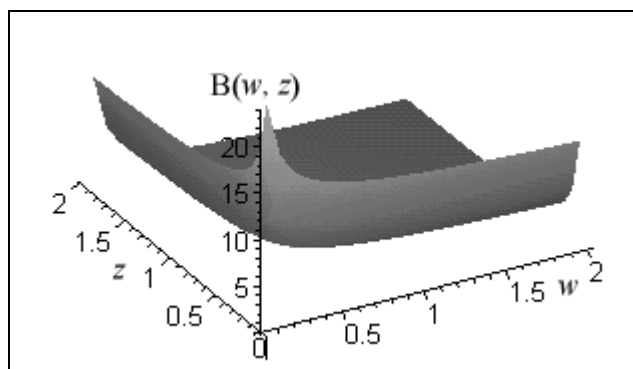
$$\lim_{k \rightarrow +\infty} \frac{\Gamma(k+a)}{\Gamma(k+b)} = k^{a-b}. \quad (2.1.4)$$

### 2.1.2 La funzione Beta

Collegata alla  $\Gamma$  è la funzione Beta, tracciata nella figura 2.2, che utilizzeremo in seguito:

$$B(z, w) = B(w, z) = \int_0^1 t^{z-1} (1-t)^{w-1} dt = \frac{\Gamma(z)\Gamma(w)}{\Gamma(z+w)}. \quad (2.1.5)$$





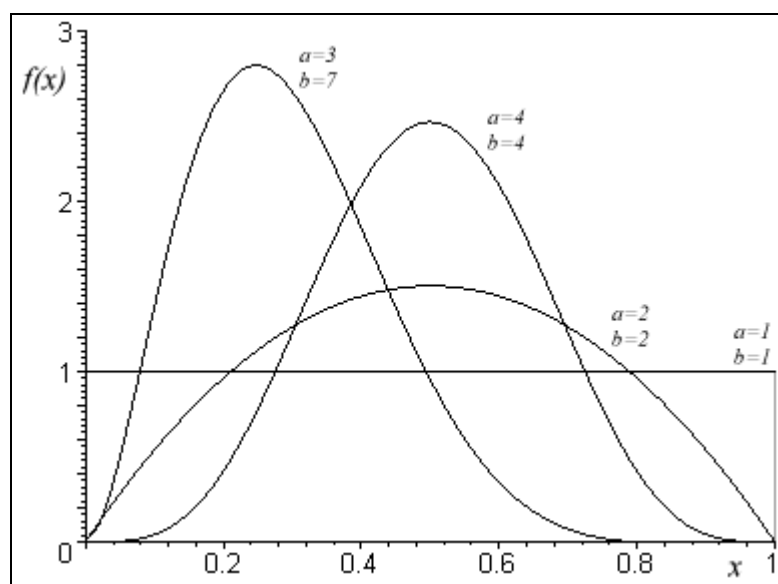
**Figura 2.2**

*La funzione Beta*

A partire dalla (2.1.5) possiamo definire la distribuzione Beta, che ha densità di probabilità:

$$f_x(a, b) = \frac{x^{a-1} (1-x)^{b-1}}{B(a, b)} \text{ per } x \in [0, 1]. \quad (2.1.6)$$

La distribuzione Beta risulta particolarmente utile perché può assumere, come si vede dalla figura 2.3, forme molto diverse al variare dei parametri  $a$  e  $b$ . Ciò ne rende appetibile l'utilizzo ogniqualvolta si debbano fare ipotesi il più generali possibili sulla distribuzione di una certa variabile aleatoria definita su un supporto limitato.



**Figura 2.3**

*La distribuzione Beta per diversi valori di  $a$  e  $b$*

Si dimostra (si veda Mood, Graybill e Boes, 1974 – pagina 116) che, se  $x$  ha distribuzione Beta:

$$E(x) = \frac{a}{a+b}, \quad (2.1.7)$$

e:

$$\text{var}(x) = \frac{ab}{(a+b+1)(a+b)^2}. \quad (2.1.8)$$

Notiamo infine che anche la funzione ipergeometrica può essere definita tramite la  $\Gamma$ :

$$F_x(a, b, c) = \sum_{n=0}^{+\infty} \frac{\Gamma(a+n)\Gamma(b+n)\Gamma(c)}{\Gamma(a)\Gamma(b)\Gamma(c+n)\Gamma(n+1)} x^n. \quad (2.1.9)$$

### 2.1.3 I polinomi di Legendre

Si dice polinomio ortogonale un polinomio ottenuto, a partire da dati polinomi  $p_i$  e  $p_j$ , imponendo condizioni del tipo:

$$\int_a^b w(x) p_j(x) p_i(x) dx = 0, \quad (2.1.10)$$

con  $i \neq j$  e  $w(x)$  funzione di pesi. Se  $w(x)=1$ ,  $a=-1$  e  $b=1$  si ottengono i polinomi di Legendre:

$$P_l(x) = \frac{1}{2^n \cdot n!} \frac{\partial^l (x^2 - 1)^n}{\partial x^l}, \quad (2.1.11)$$

che costituiscono la soluzione dell'equazione differenziale:

$$(1-x^2)y'' - 2xy' + l(l+1)y = 0.$$

La (2.1.11) può essere espressa anche in via ricorsiva:

$$\begin{cases} P_0(x)=1, & P_1(x)=x, \\ P_{l+1}(x)=\frac{2l+1}{l+1}xP_l(x)-\frac{l}{l+1}P_{l-1}(x). \end{cases} \quad (2.1.12)$$

A partire dalla (2.1.11) definiamo i polinomi di Legendre associati:

$$P_l^m(x) = (-1)^m (1-x^2)^{\frac{m}{2}} \frac{\partial^m P_l(x)}{\partial x^m}. \quad (2.1.13)$$

## 2.2 Numeri complessi

### 2.2.1 Definizioni

Si definisce numero immaginario, e si indica col simbolo  $i$ , un oggetto (immaginario, appunto) che verifica:

$$i = \sqrt{-1}. \quad (2.2.1)$$

Assumiamo che  $i$  possa essere trattato come una costante e che, in tale forma, soddisfi le regole algebriche comuni, ad esempio  $2i + 3i = 5i$ , o  $(2i)(3i) = 6i^2 = -6$ . Per due generici numeri reali  $a$  e  $b$  possiamo costruire la quantità:

$$x = a + bi. \quad (2.2.2)$$

Un numero rappresentato nella forma (2.2.2) prende il nome di numero complesso e le sue componenti  $a$  e  $bi$  prendono il nome rispettivamente di parte reale e parte immaginaria; l'insieme dei numeri complessi si indica con la lettera  $C$ . Si noti infine che un modo elegante per rappresentare il numero complesso  $x$  è quello di considerarlo un vettore di due dimensioni  $(a, b)$ .

### 2.2.2 La rappresentazione grafica

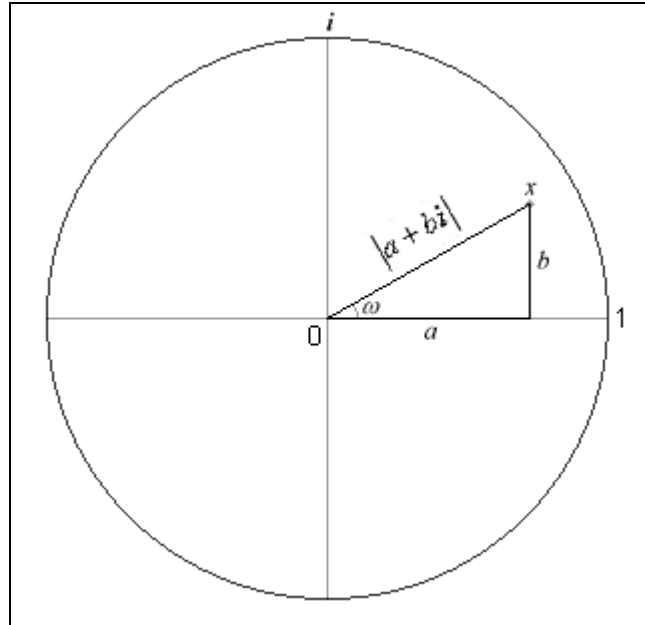
Sulla base della forma vettoriale che abbiamo introdotto poc'anzi è naturale passare alla rappresentazione grafica di un numero complesso.

Se  $b = 0$ ,  $x$  è un numero reale, se  $a = 0$   $x$  è immaginario; negli altri casi,  $x$  giace in un piano formato da un asse reale e un asse immaginario e ha coordinate  $a$  (sull'asse reale) e  $b$  (sull'asse immaginario). Questa rappresentazione è nota come diagramma di Argand, ed è schematizzata nella figura 2.4, riportata alla pagina seguente.

Il modulo di un numero complesso è calcolato come distanza dall'origine degli assi:

$$|a + bi| = \sqrt{a^2 + b^2} ; \quad (2.2.3)$$

il cerchio unitario è il luogo geometrico di tutti i numeri complessi il cui modulo è 1. Ad esempio giacciono sul cerchio unitario 1, -1,  $i$  e  $-i$ .



**Figura 2.4**

*Il diagramma di Argand e il cerchio unitario*

I numeri complessi  $a + bi$  e  $a - bi$  prendono il nome di coppia coniugata e godono delle seguenti proprietà:

$$(a + bi) + (a - bi) = 2a , \quad (2.2.4)$$

$$(a + bi)(a - bi) = a^2 + b^2 , \quad (2.2.5)$$

per cui la (2.2.3) può essere scritta anche come:

$$|a + bi| = |a - bi| = \sqrt{(a + bi)(a - bi)} . \quad (2.2.6)$$

### 2.2.3 Il teorema di De Moivre

L'ultimo risultato importante ai fini della nostra ricerca è il teorema di De Moivre, di cui omettiamo la dimostrazione (si può vedere Hamilton, 1994 – pagina 716):

$$e^{i\omega} = \cos \omega + i \sin \omega . \quad (2.2.7)$$

Un utilizzo particolarmente interessante del teorema di De Moivre è il seguente. Supponiamo di avere un numero complesso  $z = a + bi$  giacente sul cerchio unitario, e di volerlo esprimere in funzione dell'angolo ( $\omega$ ) al centro del diagramma di Argand.

Se  $z$  ha modulo uno si ha, per i risultati trigonometrici fondamentali  $a = \cos \omega$  e  $b = \sin \omega$ , per cui  $z = \cos \omega + i \sin \omega$ . Si ha quindi, per la (2.2.7) e per la (2.2.6):

$$z = e^{i\omega} = e^{-i\omega} . \quad (2.2.8)$$

## 2.3 I fondamenti dell'analisi delle serie storiche<sup>2</sup>

### 2.3.1 Il concetto di processo stocastico

Supponiamo di voler ripetere  $T$  volte<sup>3</sup> un generico esperimento aleatorio nello spazio degli eventi  $\Omega$ , e di indicare con un indice inferiore le relative variabili aleatorie  $\{x_t\}$ . Se ipotizziamo che questo esperimento sia effettuato a intervalli temporali precisi e definiti, potremmo voler raccogliere le realizzazioni di  $\{x_t\}$  in un vettore, in modo da mantenerle ordinate in ordine temporale. Potremmo scrivere allora i risultati ottenuti come:

$$\{x_t\} = (x_1, x_2, x_3, \dots, x_T). \quad (2.3.1)$$

La (2.3.1) rappresenta una realizzazione del processo stocastico  $\{x_t\}$ . Volendo utilizzare una terminologia meno intuitiva e più formale, diciamo che un processo stocastico è una famiglia di variabili aleatorie (applicazioni di  $\Omega$  in  $\mathbf{R}$ ) definite sullo stesso spazio di probabilità e ordinate secondo un parametro  $t \in T$  (normalmente il tempo); un processo stocastico si configura quindi come una variabile aleatoria  $\Omega \rightarrow \mathbf{R}^T$ .

Abbiamo visto che un processo stocastico è formato da variabili aleatorie: è pertanto possibile analizzarlo con gli stessi strumenti di esse; in particolare potremo calcolarne il valore atteso e la varianza per ognuno degli istanti  $t$ . Siccome però le variabili aleatorie sono ordinate in senso temporale, possiamo procedere oltre e calcolare il valore atteso (o la varianza) di  $x_t$  utilizzando l'informazione disponibile in  $t - 1$ , ovvero la “storia” del processo che, se ci troviamo in  $t$ , già conosciamo; si parla in questo caso di valore atteso (o varianza) condizionato, e si utilizza la notazione  $E(x_t | I_{t-1})$ , dove  $I_{t-1}$  indica l'insieme informativo in  $t - 1$ .

---

<sup>2</sup> In questa sezione si è seguita l'impostazione di Gallo e Pacini (2002).

<sup>3</sup> Si noti che  $T$  può essere infinito.

### 2.3.2 I processi ARMA

Una volta identificato un processo stocastico, quello che ci interessa è trovarne la formulazione matematica, ovvero la relazione (se esiste) che lega tra loro le realizzazioni delle variabili aleatorie (cioè il “sentiero evolutivo” del processo). Questa relazione, che rappresenta la formula matematica che “produce” le osservazioni del processo, prende il nome di processo generatore dei dati, o DGP (*Data Generating Process*). Un processo  $\{x_t\}$  tale che:

$$\begin{cases} x_t = \varepsilon_t, & t = 1, \dots, T \\ \varepsilon_t \sim N(\mu, \sigma^2) \end{cases}$$

prende il nome di *white noise* (rumore bianco).

La maggior parte dei DGP che si utilizzano sono generati a partire da un rumore  $\varepsilon_t$  che si sovrappone a un segnale individuabile a priori. Un processo AR ( $p$ ) è ad esempio:

$$y_t = \mu + \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \dots + \varphi_p y_{t-p} + \varepsilon_t,$$

ovvero un processo autoregressivo che considera  $p$  ritardi.

Un processo a media mobile MA ( $q$ ) è invece:

$$y_t = \mu + \psi_1 \varepsilon_{t-1} + \psi_2 \varepsilon_{t-2} + \dots + \psi_q \varepsilon_{t-q} + \varepsilon_t.$$

Sovrapponendo processi AR e MA si ottiene la famiglia ARMA. Ad esempio un ARMA ( $p, q$ ) è dato da:

$$\begin{aligned} y_t = & \mu + \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \dots + \varphi_p y_{t-p} + \\ & + \psi_1 \varepsilon_{t-1} + \psi_2 \varepsilon_{t-2} + \dots + \psi_q \varepsilon_{t-q} + \varepsilon_t. \end{aligned} \quad (2.3.2)$$

Se definiamo un operatore *lag*, o di ritardo, tale che:  $L^k x_t = x_{t-k}$ , possiamo scrivere il processo ARMA ( $p, q$ ) della (2.3.2) in forma più compatta utilizzando i polinomi

$$\varphi(L) = 1 - \varphi_1 L + \varphi_2 L^2 + \dots + \varphi_p L^p$$

e

$$\psi(L) = 1 + \psi_1 L + \psi_2 L^2 + \dots + \psi_q L^q.$$



La (2.3.2) diventa allora:

$$\varphi(L)(y_t - \mu) = \psi(L)\varepsilon_t. \quad (2.3.3)$$

### 2.3.3 Stazionarietà e invertibilità

Un processo si dice strettamente stazionario (o stazionario forte) se le sue proprietà non dipendono dall'origine temporale che si considera; in termini più formali se la distribuzione di probabilità congiunta in  $t_1, t_2, \dots, t_m$  è identica a quella in  $t_{k+1}, t_{k+2}, \dots, t_{k+m}$  per ogni  $k$  e per ogni  $t_1, t_2, \dots, t_m$ . Se consideriamo  $m = 1$  osserviamo che assumere stazionarietà forte significa assumere che la distribuzione in  $t$  è uguale a quella in  $t + k$  per ogni valore di  $k$ .

La covarianza di due variabili aleatorie che compongono il processo prende il nome di autocovarianza (ugualmente si definisce l'autocorrelazione). Formalmente:

$$\begin{aligned} \text{cov}(x_t, x_{t-k}) &= \gamma_k = E[(x_t - \mu)(x_{t-k} - \mu)], \\ \text{cor}(x_t, x_{t-k}) &= \rho_k = \frac{\gamma_k}{\gamma_0}, \end{aligned} \quad (2.3.4)$$

dove  $\mu$  indica il valore atteso del processo.

Un processo si dice stazionario in covarianza (o stazionario debole) se la (2.3.4) è verificata, ovvero se l'autocovarianza dipende solo dal ritardo  $k$  e non dal tempo  $t$  a cui la si calcola. In genere è più semplice dimostrare la stazionarietà debole rispetto a quella forte per cui, nel seguito di questo lavoro, quando parleremo di stazionarietà *tout court*, intenderemo (a meno che non si dica diversamente) riferirci alla stazionarietà debole.

Si può dimostrare (si rimanda a Hamilton, 1994 – pagina 59) che un processo ARMA è stazionario se e solo se tutte le radici del polinomio AR  $\varphi(z)$  giacciono al di fuori del cerchio unitario.

Se un processo ARMA è stazionario, vale un'importante proprietà, nota come scomposizione di Wold (per la dimostrazione si rimanda a Hamilton, 1994 – pagina 109): cioè che ogni

processo ARMA stazionario può essere rappresentato come MA ( $\infty$ ). La prova di questo risultato può essere ricavata anche per via intuitiva: basta infatti sostituire a ciascun termine  $y_{t-j}$  la sua espressione secondo la (2.3.2).

Se consideriamo, ad esempio, un semplice AR (1) si avrà:

$$y_t = \frac{\mu}{1-\varphi} + \sum_{j=0}^{+\infty} \varphi^j \varepsilon_{t-j}. \quad (2.3.5)$$

Osservando la (2.3.5) possiamo ricavare altresì il valore atteso non condizionato, che risulta essere ovviamente  $\frac{\mu}{1-\varphi}$ , e la varianza:

$$\begin{aligned} \gamma_0 &= E(y_t - \mu)^2 = E\left(\sum_{j=0}^{+\infty} \varphi^j \varepsilon_{t-j}\right)^2 \Rightarrow \\ \gamma_0 &= \sigma^2 \sum_{j=0}^{+\infty} \varphi^{2j} = \frac{\sigma^2}{1-\varphi^2}. \end{aligned} \quad (2.3.6)$$

La funzione di autocorrelazione vale invece<sup>4</sup>:

$$\begin{aligned} \gamma_k &= \sigma^2 \frac{\varphi^k}{1-\varphi^2} \Rightarrow \\ \rho_k &= \frac{\gamma_k}{\gamma_0} = \varphi^k. \end{aligned} \quad (2.3.7)$$

Similmente si può dimostrare<sup>5</sup> che, per un processo MA (1), valgono:

$$\begin{cases} E(y_t) = \mu, \\ \gamma_0 = \sigma^2(1 + \psi^2), \\ \gamma_1 = \psi\sigma^2, \\ \gamma_k = 0 \quad \forall k > 1, \\ \rho_1 = \frac{\psi}{1 + \psi^2}, \\ \rho_k = 0 \quad \forall k > 1. \end{cases} \quad (2.3.8)$$

<sup>4</sup> Si veda Hamilton (1994 – pagina 53) per la semplice dimostrazione.

<sup>5</sup> Si veda anche in questo caso Hamilton (1994 – pagina 48) per la dimostrazione.

Un processo si dice invertibile se, data una forma come quella della (2.3.2), può essere esplicitato il termine di disturbo:

$$\begin{aligned}\varepsilon_t = y_t - \mu - \varphi_1 y_{t-1} - \varphi_2 y_{t-2} - \cdots - \varphi_p y_{t-p} + \\ - \psi_1 \varepsilon_{t-1} - \psi_2 \varepsilon_{t-2} + \cdots - \psi_q \varepsilon_{t-q}.\end{aligned}\quad (2.3.9)$$

Un processo invertibile può essere rappresentato (anche in questo caso è piuttosto facile rendersene conto, poiché basta sostituire all'infinito tutti gli  $\varepsilon_t$  che compaiono nella (2.3.9)), come AR ( $\infty$ ).

Si può dimostrare (di nuovo si rimanda a Hamilton, 1994 – pagina 67) che un processo ARMA è invertibile se tutte le radici del polinomio MA  $\psi(z)$  giacciono al di fuori del cerchio unitario.

#### 2.3.4 Processi integrati

Abbiamo visto che i processi stazionari godono di importanti proprietà, una su tutte la scomposizione di Wold. Sembrerebbe quindi utile trovare degli strumenti che consentano di ricondurre un processo non stazionario alla stazionarietà.

Si dice integrato di ordine  $n$  e si indica con la notazione  $I(n)$  un processo che, differenziato  $n$  volte<sup>6</sup>, diviene stazionario in covarianza. Il più semplice processo  $I(1)$  è il *random walk*:

$$y_t = y_{t-1} + \varepsilon_t,$$

che può essere scritto, utilizzando l'operatore di ritardo, come:

$$(1 - L)y_t = \varepsilon_t. \quad (2.3.10)$$

Un *random walk* è un processo AR (1) in cui la radice del polinomio  $\varphi(z)$  vale uno: per questo si parla anche di processo a radice unitaria. Utilizzando questo processo come esempio, proviamo a vedere cosa succede se cerchiamo di calcolare la media e la varianza

secondo la formula (2.3.6): il denominatore vale zero, per cui sia la media che la varianza sono infinite. In generale, se la radice del polinomio AR giace all'interno del cerchio unitario otterremo valori della media non sensati e varianze negative. Chiaramente questa situazione pone dei problemi: ecco perché è conveniente avere degli strumenti che ci consentano di ricondursi alla stazionarietà.

Osserviamo infine che, utilizzando la notazione della (2.3.10), possiamo pervenire a una rappresentazione più generica per i processi integrati:

$$(1 - L)^n y_t = \varepsilon_t, \quad (2.3.11)$$

che indica un processo  $I(n)$ .

### 2.3.5 I test per radice unitaria<sup>7</sup>

Testare la presenza di radici unitarie in una serie può sembrare assai semplice: l'idea che si può avere è di regredire la serie sui propri ritardi e vedere se il coefficiente di regressione è statisticamente uguale a uno. Purtroppo però, quando il processo generatore dei dati è un *random walk*, la distribuzione asintotica dello stimatore del coefficiente di regressione non è una *t* di Student, ma ha una forma assai più complessa, variabile a seconda del numero dei ritardi che si utilizzano e del tipo di modello utilizzato (*random walk* semplice, con drift o con trend deterministico).

Tali distribuzioni sono state tabulate da Dickey e Fuller (1979), per cui è possibile utilizzarle per sottoporre a verifica empirica la presenza di radici unitarie.

Formalmente, si conduce la regressione lineare:

$$\Delta y_t = \mu + \beta y_{t-1} + \varepsilon_t, \quad (2.3.12)$$

---

<sup>6</sup> Come abbiamo già detto nell'introduzione, differenziare un processo  $\{y_t\}$  significa generare la serie  $w_t = y_t - y_{t-1}$ .

<sup>7</sup> Per una trattazione più approfondita e per le distribuzioni asintotiche si rimanda a Hamilton (1994 – capitolo 17).

e si sottopone a verifica l'ipotesi nulla  $\beta = 0$  contro l'alternativa  $\beta < 0$ .

Nel caso in cui vi sia autocorrelazione nei disturbi, Dickey e Fuller (1979) inseriscono ulteriori termini AR nella regressione; in questo caso il test prende il nome di *Augmented Dickey-Fuller* (ADF). La regressione che si utilizza è, considerando  $p$  ritardi:

$$\Delta y_t = \mu + \beta y_{t-1} + \zeta_1 \Delta y_{t-1} + \dots + \zeta_{p-1} \Delta y_{t-p+1} + \varepsilon_t, \quad (2.3.13)$$

e il test è di nuovo  $\beta = 0$  contro  $\beta < 0$ .

Un'ulteriore possibilità di trattare i disturbi autocorrelati è infine considerata da Phillips e Perron (1988), che invece di aggiungere termini alla regressione modificano la statistica test. Dato lo stimatore delle covarianze di Newey e West (1987):

$$s^2(l) = \frac{\sum_{t=1}^T e_t^2}{T} + 2 \frac{\sum_{s=1}^l \left(1 - \frac{s}{l+1}\right) \sum_{t=s+1}^T e_t e_{t-s}}{T}, \quad (2.3.14)$$

la statistica test che si utilizza è:

$$PP = \frac{t_{\hat{\beta}} \sqrt{\gamma_0}}{s(p)} - \frac{[s^2(p) - \gamma_0] T \hat{\sigma}_{\hat{\beta}}}{2s(p) \hat{\sigma}}, \quad (2.3.15)$$

dove  $\hat{\sigma}$  è lo *standard error* della regressione (2.3.13) e  $t_{\hat{\beta}}$  e  $\hat{\sigma}_{\hat{\beta}}$  rappresentano rispettivamente la statistica t e lo *standard error* relativi al coefficiente  $\beta$ .

## 2.4 Analisi spettrale<sup>8</sup>

L'idea naturale nello studio delle serie storiche è quella di analizzare una serie in funzione delle realizzazioni che assume a diversi istanti di tempo (si parla anche di analisi nel dominio del tempo). Questa impostazione non deve essere però presa per assoluta. Possiamo infatti rappresentare una serie anche come somma di funzioni periodiche<sup>9</sup> con l'obiettivo di individuare dei cicli rilevanti nel comportamento della serie. Questo filone di ricerca prende il nome di analisi spettrale, o analisi nel dominio della frequenza.

### 2.4.1 La funzione generatrice delle autocovarianze

Sia  $\{y_t\}$  un processo stazionario in covarianza. Definiamo la funzione generatrice delle autocovarianze come:

$$g(z) = \sum_{j=-\infty}^{+\infty} \gamma_j z^j, \quad (2.4.1)$$

dove  $z$  indica uno scalare complesso. La (2.4.1) rappresenta un modo elegante per raccogliere in un polinomio le autocovarianze di un generico processo: si ha infatti una somma di termini in  $z$  in cui il coefficiente indica il valore dell'autocovarianza, mentre l'esponente di  $z$  indica il ritardo a cui tale autocovarianza si riferisce.

Supponiamo ad esempio che  $\{y_t\}$  sia un processo MA (1): sappiamo che  $\gamma_0 = (1 + \psi^2)\sigma^2$ ,

$\gamma_1 = \gamma_{-1} = \psi\sigma^2$ . Si avrà allora:

$$\begin{aligned} g(z) &= (\psi\sigma^2)z^{-1} + [(1 + \psi^2)\sigma^2]z^0 + (\psi\sigma^2)z^1 = \\ &= (1 + \psi z)(1 + \psi z^{-1})\sigma^2. \end{aligned}$$

---

<sup>8</sup> In questa sezione si è seguita la presentazione di Hamilton (1994).

<sup>9</sup> Per chi ha dimestichezza con strumenti matematici più avanzati, ciò equivale ad applicare al processo generatore dei dati la trasformazione di Fourier. Per una trattazione dell'argomento si rimanda a Giusti (1988) e a Brockwell e Davis (1991).

### 2.4.2 Lo spettro del processo

Limitiamo  $z$  ai valori che giacciono sul cerchio unitario:

$$z = e^{-i\omega}, \text{ con } i = \sqrt{-1},$$

cosicché la  $g$  dipenda esclusivamente da  $\omega$ , dividiamo per  $2\pi$  e otteniamo lo spettro del processo:

$$s(\omega) = \frac{g(e^{-i\omega})}{2\pi} = \frac{1}{2\pi} \sum_{j=-\infty}^{+\infty} \gamma_j e^{-i\omega j}. \quad (2.4.2)$$

Per la (2.2.7) abbiamo:

$$e^{-i\omega j} = \cos(\omega j) - i \sin(\omega j), \quad (2.4.3)$$

inoltre  $\{y_t\}$  è stazionario in covarianza, per cui:

$$\gamma_{-j} = \gamma_j. \quad (2.4.4)$$

Sostituendo la (2.4.3) e la (2.4.4) nella (2.4.2) otteniamo:

$$\begin{aligned} s(\omega) &= \frac{\gamma_0}{2\pi} [\cos 0 - i \sin 0] + \\ &+ \frac{1}{2\pi} \left\{ \sum_{j=1}^{+\infty} \gamma_j [\cos(\omega j) + \cos(-\omega j) - i \sin(\omega j) - i \sin(-\omega j)] \right\}, \end{aligned} \quad (2.4.5)$$

che utilizzando semplici proprietà trigonometriche diventa:

$$s(\omega) = \frac{1}{2\pi} \left[ \gamma_0 + 2 \sum_{j=1}^{+\infty} \gamma_j \cos(\omega j) \right]. \quad (2.4.6)$$

Dalla (2.4.6) si apprezzano le proprietà principali dello spettro: la simmetria attorno allo zero (poiché  $\cos(\omega)$  è simmetrico) e la periodicità, anch'essa dovuta alla periodicità del coseno.

Pertanto, conoscendo lo spettro nell'intervallo  $[0, \pi]$ , possiamo ricavarlo su tutto  $\mathbf{R}$ .

Riprendiamo l'esempio di cui sopra e calcoliamo lo spettro di un processo MA (1):

$$\begin{aligned}
s(\omega) &= \frac{\sigma^2}{2\pi} (1 + \psi e^{-i\omega}) (1 + \psi e^{i\omega}) = \\
&= \frac{\sigma^2}{2\pi} [1 + \psi(e^{-i\omega} + e^{i\omega}) + \psi^2],
\end{aligned}$$

ricordiamo che, per la (2.2.7),

$$e^{-i\omega} + e^{i\omega} = \cos \omega - i \sin \omega + \cos \omega + i \sin \omega = 2 \cos \omega$$

e otteniamo:

$$s(\omega) = \frac{\sigma^2}{2\pi} [1 + 2\psi \cos \omega + \psi^2].$$

Per un generico processo ARMA  $(p, q)$  vale invece (si veda per la dimostrazione Hamilton, 1994 – pagina 155):

$$\begin{aligned}
s(\omega) &= \frac{\sigma^2}{2\pi} \frac{|\psi(z)|^2}{|\varphi(z)|^2} = \\
&= \frac{\sigma^2}{2\pi} \frac{(1 + \psi_1 e^{-i\omega} + \psi_2 e^{-i2\omega} + \dots + \psi_q e^{-iq\omega})}{(1 + \varphi_1 e^{-i\omega} + \varphi_2 e^{-i2\omega} + \dots + \varphi_p e^{-ip\omega})} \times \\
&\quad \times \frac{(1 + \psi_1 e^{i\omega} + \psi_2 e^{i2\omega} + \dots + \psi_q e^{iq\omega})}{(1 + \varphi_1 e^{i\omega} + \varphi_2 e^{i2\omega} + \dots + \varphi_p e^{ip\omega})}.
\end{aligned} \tag{2.4.7}$$

Abbiamo costruito lo spettro a partire dalle autocovarianze: possiamo fare anche l'operazione inversa, cioè ricavare le autocovarianze a partire dallo spettro. Si può dimostrare (Hamilton, 1994 – pagina 155) che:

$$\gamma_j = \int_{-\pi}^{+\pi} s(\omega) e^{i\omega j} d\omega = \int_{-\pi}^{+\pi} s(\omega) \cos(\omega j) d\omega. \tag{2.4.8}$$

### 2.4.3 Il significato dello spettro

Proviamo a ricavare  $\gamma_0$ , cioè la varianza del processo, dalla (2.4.8):

$$\gamma_0 = \int_{-\pi}^{+\pi} s(\omega) d\omega. \tag{2.4.9}$$



Ciò significa, in termini più intuitivi, che l'area delimitata dallo spettro tra  $-\pi$  e  $\pi$  corrisponde alla varianza. Questo risultato può essere esteso a un generico  $\omega_1$  compreso in  $[-\pi, \pi]$ : la

quantità  $\int_{-\omega_1}^{+\omega_1} s(\omega) d\omega = 2 \int_0^{\omega_1} s(\omega) d\omega$  rappresenterà la porzione di varianza del processo che è

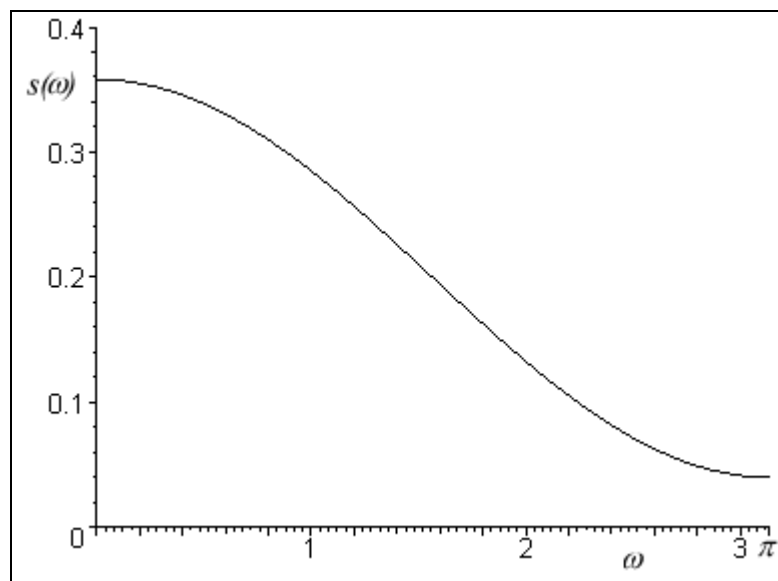
spiegata da cicli con frequenza inferiore a  $\omega_1$ .

Cerchiamo di precisare meglio questo concetto. Un processo con una densità spettrale alta a frequenze basse ha, per quanto detto sopra, la maggior parte della varianza addensata in cicli di bassa frequenza, il che equivale a dire che la maggior parte delle deviazioni dal valore atteso che si osservano sono dovute a effetti di breve termine. Questo significa che, all'atto pratico, la dipendenza a breve termine del processo è bassa (perché più alta è la parte di varianza ad esso imputabile), mentre è più forte quella di lungo periodo.

La situazione sarà più chiara confrontando alcuni spettri. Nella figura 2.5 si è tracciato lo spettro di un processo MA (1) con  $\psi = 1/2$  e varianza, per semplicità, unitaria. Ci aspettiamo che un processo AR (1), a parità del parametro, abbia una dipendenza di lungo periodo più forte: ciò risalta subito confrontando la figura 2.5 con la figura 2.6, dove si è tracciato lo spettro di un AR (1) con  $\phi = 1/2$  e varianza unitaria. Nella figura 2.7 vediamo invece lo spettro di un processo *random walk*, cioè un AR (1) con radice unitaria<sup>10</sup>: in questo caso, con una dipendenza di lungo termine fortissima, lo spettro è addirittura esplosivo per  $\omega \rightarrow 0$ .

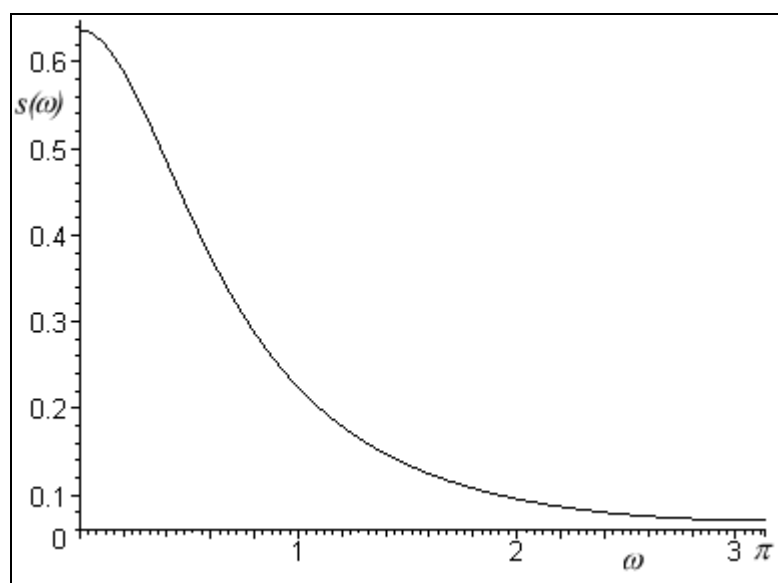
---

<sup>10</sup> In realtà, nel caso di processi non stazionari, lo spettro non è definito e si può fare riferimento solo al periodogramma. Ci si perdonerà la semplificazione che ha scopi unicamente illustrativi.



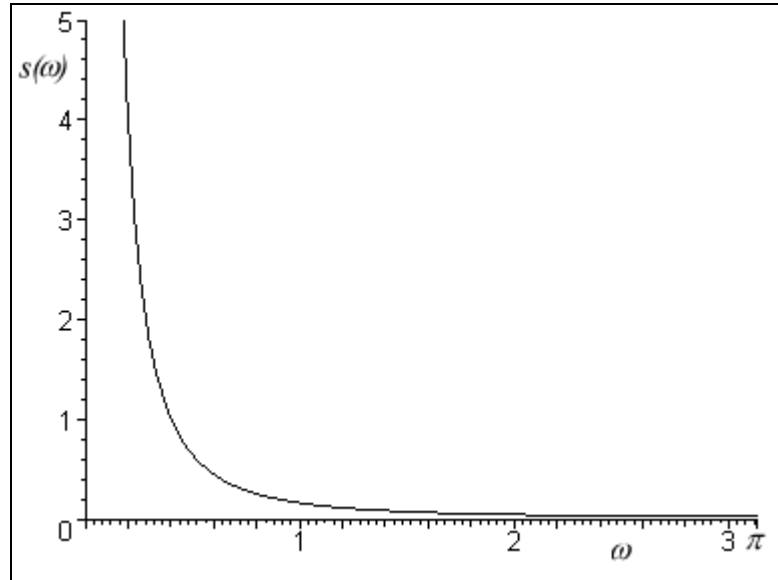
**Figura 2.5**

*Lo spettro di un processo MA (1) con  $\psi = 1/2$*



**Figura 2.6**

*Lo spettro di un processo AR (1) con  $\phi = 1/2$*



**Figura 2.7**

*Lo spettro di un random walk*

#### 2.4.4 Il periodogramma

Nel paragrafo 2.4.2 abbiamo visto come costruire lo spettro a partire dalla forma analitica del processo generatore dei dati. All'atto pratico, però, ci troviamo il più delle volte nella situazione opposta: abbiamo un insieme di dati ma non conosciamo la forma esatta del processo che li ha generati. Vediamo adesso come ricavare lo spettro della popolazione a partire dai dati campionari. In questo caso, ovviamente, non potremo ricavare lo spettro "vero", ma solo una sua *proxy*, che prende il nome di periodogramma.

Supponiamo di avere una serie di  $T$  osservazioni per un processo  $\{y_t\}$ , stazionario e con processo generatore dei dati ignoto. Possiamo stimare le autocovarianze a partire dai dati:

$$\hat{\gamma}_j = \hat{\gamma}_{-j} = T^{-1} \sum_{t=j+1}^T (y_t - \bar{y})(y_{t-j} - \bar{y}), \quad (2.4.10)$$

dove  $\bar{y}$  indica la media campionaria delle osservazioni.

A questo punto basta inserire la (2.4.10) nella (2.4.2):

$$\hat{s}(\omega) = \frac{1}{2\pi} \sum_{j=-T+1}^{T+1} \hat{\gamma}_j e^{-i\omega j} . \quad (2.4.11)$$

Ovviamente anche per il periodogramma vale quanto derivato nei paragrafi precedenti per lo spettro, in particolare la periodicit , la simmetria e l'espressione alternativa per la (2.4.11):

$$\hat{s}(\omega) = \frac{1}{2\pi} \left[ \hat{\gamma}_0 + 2 \sum_{j=1}^{T-1} \hat{\gamma}_j \cos(\omega j) \right] . \quad (2.4.12)$$

### 2.4.5 Il teorema di rappresentazione spettrale

Uno dei risultati pi  interessanti dell'analisi spettrale riguarda la rappresentazione di un generico processo  $\{y_t\}$ , in particolare la possibilit  di passare da una rappresentazione classica in funzione di  $t$  a una rappresentazione che esprima il processo come somma di funzioni periodiche.

Consideriamo una serie di  $M = \frac{T-1}{2}$  frequenze tali che  $\omega_j = \frac{2\pi j}{T}$ ,  $j = 1, 2, \dots, M$ .

Regrediamo  $y_t$  secondo il modello:

$$y_t = \mu + \sum_{j=1}^M \left\{ \alpha_j \cos[\omega_j(t-1)] + \delta_j \sin[\omega_j(t-1)] \right\} + \varepsilon_t , \quad (2.4.13)$$

che pu  essere espresso in forma matriciale come:

$$y_t = \mathbf{\beta}' \mathbf{x}_t + \varepsilon_t , \quad (2.4.14)$$

con:

$$\mathbf{x}_t = \begin{bmatrix} 1 & \cos[\omega_1(t-1)] & \sin[\omega_1(t-1)] & \cdots \\ \cdots & \cos[\omega_M(t-1)] & \sin[\omega_M(t-1)] \end{bmatrix}' \quad (2.4.15)$$

e:

$$\mathbf{\beta}' = [\mu \quad \alpha_1 \quad \delta_1 \quad \alpha_2 \quad \delta_2 \quad \cdots \quad \alpha_M \quad \delta_M] . \quad (2.4.16)$$

Si dimostra (si veda Hamilton, 1994 – pagina 173) che:

$$\sum_{t=1}^T \mathbf{x}_t \mathbf{x}_t' = \begin{bmatrix} T & \mathbf{0}' \\ \mathbf{0} & (T/2) \mathbf{I}_{T-1} \end{bmatrix}. \quad (2.4.17)$$

Siano ora:

$$\begin{aligned} \hat{\alpha}_j &= (2/T) \sum_{t=1}^T y_t \cos[\omega_j(t-1)], \\ \hat{\delta}_j &= (2/T) \sum_{t=1}^T y_t \sin[\omega_j(t-1)], \\ \hat{\mu} &= \bar{y}. \end{aligned} \quad (2.4.18)$$

Si ha allora:

$$y_t = \hat{\mu} + \sum_{j=1}^M \left\{ \hat{\alpha}_j \cos[\omega_j(t-1)] + \hat{\delta}_j \sin[\omega_j(t-1)] \right\}, \quad (2.4.19)$$

che, generalizzando al di fuori dei dati campionari (Cramér e Leadbetter, 1967), diventa:

$$y_t = \mu + \int_0^\pi \alpha(\omega) \cos(\omega t) + \delta(\omega) \sin(\omega t) d\omega, \quad (2.4.20)$$

cioè quanto volevamo dimostrare.

#### 2.4.6 La stima dello spettro

L'idea immediata che si può avere a proposito della stima dello spettro è quella di utilizzare il periodogramma. Questa procedura è però molto limitante. Si può infatti dimostrare che la quantità  $\frac{2\hat{s}(\omega)}{s(\omega)}$  si distribuisce come un  $\chi^2(2)$ , che ha valore atteso 2. Però l'intervallo di confidenza al 95% risulta particolarmente ampio:  $\hat{s}(\omega) \in [0.025s(\omega), 3.7s(\omega)]$ , e ciò è decisamente insoddisfacente. Inoltre, la quantità  $\frac{2\hat{s}(\omega)}{s(\omega)}$  non dipende da  $T$ : ciò significa che la precisione dello stimatore non migliora all'aumentare delle dimensioni del campione.

Per ovviare a questi inconvenienti, possiamo utilizzare la stima di massima verosimiglianza (si veda il capitolo 4): specifichiamo un modello ARMA  $(p, q)$  per i dati, ricaviamo i coefficienti e li inseriamo nella (2.4.7).

Un'altra possibilità è quella delle cosiddette *stime kernel*, che muovono dall'ipotesi che, se  $\omega$  è vicino a  $\lambda$ , allora anche  $s(\omega)$  sarà vicino a  $s(\lambda)$ . Lo spettro in  $\omega$  potrà essere così stimato come somma pesata dei valori  $s(\lambda)$  in un intorno di  $\omega$ . Dato  $\omega_j = \frac{2\pi j}{T}$ :

$$\hat{s}(\omega_j) = \sum_{m=-h}^{+h} \kappa(\omega_{j+m}, \omega_j) \hat{s}(\omega_{j+m}). \quad (2.4.21)$$

Il parametro  $h$ , che prende il nome di ampiezza di banda, indica quante frequenze attorno a  $\omega$  si ritengono indicative, mentre il kernel  $\kappa(\omega_{j+m}, \omega_j)$  indica i pesi da dare a ogni frequenza; ovviamente le somme in  $m$  dei kernel valgono 1.

Poiché si può dimostrare che  $\sum_{m=-h}^{+h} [h+1-|m|] = (h+1)^2$ , una possibilità è quella di utilizzare:

$$\kappa(\omega_{j+m}, \omega_j) = \frac{h+1-|m|}{(h+1)^2}, \quad (2.4.22)$$

per cui la (2.4.21) diventa:

$$\hat{s}(\omega_j) = \sum_{m=-h}^{+h} \left[ \frac{h+1-|m|}{(h+1)^2} \right] \hat{s}(\omega_{j+m}). \quad (2.4.23)$$

Si osservi che il periodogramma è una stima kernel del tutto particolare: l'ampiezza di banda è zero, e tutto il peso (cioè 1), va sull'osservazione al punto  $\omega$ . Da una considerazione di questo tipo si può visualizzare meglio cosa succede applicando una stima kernel: si smussa il periodogramma ponderando ogni osservazione con quelle ad essa vicine. Ovviamente, se per i pesi decidiamo di utilizzare la (2.4.22), lo spettro stimato sarà tanto più liscio quanto più larga è la banda che decidiamo di utilizzare.

# Capitolo 3

## L'integrazione frazionaria

Dopo aver definito formalmente la proprietà di memoria lunga saranno introdotti il *fractional white noise*, il processo ARFIMA  $(p, d, q)$  e il processo GARMA  $(p, \xi, d, q)$  e ne saranno ricavate le proprietà salienti, con particolare attenzione alla funzione di autocorrelazione e alla densità spettrale. Accenneremo alle metodologie di test per l'integrazione frazionaria, presenteremo alcuni modelli teorici che spiegano l'insorgenza di comportamento  $I(d)$  in una serie storica e infine le ricerche empiriche che hanno mostrato evidenza di integrazione frazionaria in dati reali.

### 3.1 Il concetto di memoria lunga

Abbiamo già parlato nell'introduzione di processi a memoria lunga, che abbiamo definito in maniera tanto semplice quanto rozza a partire dal decadimento più lento delle autocorrelazioni; introduciamo adesso delle definizioni più formali.

La prima definizione che presentiamo è dovuta a McLeod e Hipel (1978). Dato un processo  $\{y_t\}$  in tempo discreto, con autocorrelazione  $\rho_j$  al ritardo  $j$ , diciamo che il processo possiede memoria lunga se la quantità

$$\sum_{j=-\infty}^{+\infty} |\rho_j| \quad (3.1.1)$$

non è finita.

Ovviamente un comportamento di questo tipo ha i suoi effetti anche sullo spettro.

Richiamiamo la (2.4.6):

$$s(\omega) = \frac{1}{2\pi} \left[ \gamma_0 + 2 \sum_{j=1}^{+\infty} \gamma_j \cos(\omega j) \right], \quad (2.4.6)$$

e osserviamo che, se la (3.1.1) non converge, è anche:

$$\sum_{j=-\infty}^{+\infty} \gamma_j = 2 \sum_{j=0}^{+\infty} \gamma_j = +\infty. \quad (3.1.2)$$

Per  $\omega \rightarrow 0$ ,  $\cos(\omega j) \rightarrow 1$ , per cui il secondo addendo della (2.4.6) non converge. Ciò implica che:

$$\lim_{\omega \rightarrow 0} s(\omega) = +\infty. \quad (3.1.3)$$

Un'altra possibilità è quella di definire la “memoria corta”. Seguendo Rosenblatt (1956)

definiamo  $S_T = \sum_{t=1}^T y_t$ : il processo si dice a memoria corta se

$$\sigma^2 = \lim_{T \rightarrow +\infty} E(T^{-1} S_T^2) \quad (3.1.4)$$

esiste ed è diverso da zero e se



$$\frac{S_{[rT]}}{\sigma\sqrt{T}} \xrightarrow{d} B(r) \quad \forall r \in [0,1], \quad (3.1.5)$$

dove  $[rT]$  è la parte intera di  $rT$  e  $B(r)$  è un moto Browniano<sup>1</sup>.

Un'ultima possibilità è considerata da Helson e Sarason (1967). Un processo si dice a memoria lunga se la sua funzione di autocorrelazione è, per  $k$  sufficientemente grande e  $d > 0$ :

$$\gamma_k \cong \Xi(k)k^{2d-2}. \quad (3.1.6)$$

Le caratteristiche della funzione  $\Xi(k)$ , a variazioni lente<sup>2</sup> a infinito, sono studiate da Resnick (1987).

Tra queste definizioni, quella che si rivela di più semplice utilizzo e più utile per la nostra analisi è senz'altro quella di McLeod e Hipel (1978) della (3.1.1), per cui in seguito ci si riferirà esclusivamente a questa.

È semplice dimostrare che i processi ARMA “tradizionali” sono a memoria corta: l'autocorrelazione di un processo MA ( $q$ ) decade esponenzialmente fino al  $q$ -esimo ritardo e poi vale zero (si veda Hamilton, 1994 – pagina 50); poiché, come abbiamo visto nel paragrafo 2.3.3, ogni processo ARMA stazionario può essere trasformato in MA ( $\infty$ ) attraverso la scomposizione di Wold, se ne conclude che in ogni caso la funzione di autocorrelazione ha un decadimento esponenziale. Ciò significa che la (3.1.1) converge, per cui il processo non ha memoria lunga.

---

<sup>1</sup> Il moto Browniano può essere introdotto intuitivamente come la trasposizione in tempo continuo di un *random walk* con origine zero. Scrivendo  $B(r)$  si intende il valore, al tempo  $r$ , di un processo il cui DGP è un moto Browniano; trattandosi di una variabile aleatoria essa possiede una propria funzione di densità di probabilità. Dati due istanti  $s < t$ , la distribuzione dell'incremento  $B(t) - B(s)$  è normale con media zero e varianza  $t - s$ ;  $B(1)$  è pertanto distribuito normalmente con media nulla e varianza unitaria.

Un integrale in cui l'integratore è un moto browniano si indica con la notazione  $I(t) = \int_0^t x(t)dB(t)$  e prende il nome di integrale stocastico, o integrale di Ito.

Si è inteso con questa nota semplicemente introdurre intuitivamente concetti che saranno talvolta richiamati nel corso di questo lavoro, ma la cui padronanza non è necessaria per la comprensione dello stesso. Una trattazione più approfondita e più formale richiederebbe l'utilizzo di nozioni probabilistiche e soprattutto matematiche molto approfondite; si rimanda pertanto a Geronazzo (1998) o Neftci (1996).

<sup>2</sup> Una funzione si dice a variazioni lente a infinito se  $\lim_{t \rightarrow +\infty} f(tx)/f(t) = 1$  per tutti gli  $x > 0$ .

## 3.2 Il modello ARFIMA: derivazione e proprietà

Le proprietà dei processi ARFIMA sono state derivate indipendentemente da Granger e Joyeux (1980) e da Hosking (1981); per la nostra presentazione, attingeremo ad entrambe le impostazioni.

### 3.2.1 *Il fractional white noise*

Richiamiamo in primo luogo, seguendo Hosking (1981), l'operatore *lag*, o di ritardo:

$$L^k x_t = x_{t-k};$$

e la forma generica di un processo integrato di ordine  $d$  che abbiamo introdotto nella (2.3.11):

$$(1-L)^d y_t = \varepsilon_t.$$

Trattiamo  $L$  come uno scalare<sup>3</sup> e scriviamo l'espansione in serie di McLaurin di  $(1-L)^d$ :

$$\begin{aligned} (1-L)^d &= \sum_{k=0}^d \binom{d}{k} (-L)^k = \\ &= 1 - dL - \frac{1}{2}d(d-1)L^2 + \frac{1}{6}d(d-1)(d-2)L^3 + \dots. \end{aligned}$$

Consideriamo ora la possibilità che  $d$  assuma un valore non intero. Poiché la funzione fattoriale è definita solo per valori naturali, dobbiamo in primo luogo ridefinire i coefficienti binomiali utilizzando la funzione  $\Gamma$  che abbiamo introdotto nella (2.1.1):

$$\binom{d}{k} = \frac{d!}{k!(d-k)!} = \frac{\Gamma(d+1)}{\Gamma(k+1)\Gamma(d-k+1)}.$$

---

<sup>3</sup> In realtà la questione è assai più complessa; si può dimostrare comunque (si rimanda a Gouriéroux e Monfort, 1997 – pagine 137 e seguenti) che un generico polinomio  $\varphi(L)$  può essere trattato come  $\varphi(z)$ , con  $z$  numero complesso.

L'espansione in serie di McLaurin vale allora:

$$(1-L)^d = \sum_{k=0}^{+\infty} \frac{\Gamma(d+1)}{\Gamma(k+1)\Gamma(d-k+1)} (-L)^k \quad (3.2.1)$$

Se ipotizziamo che  $d$  possa assumere valori compresi tra zero e uno, il processo  $x_t = (1-L)^{-d} \varepsilon_t$  prende il nome di *fractional white noise*.

Torniamo per un attimo al caso in cui  $d$  è naturale ed espandendo  $(1-L)^{-d}$  secondo la (3.2.1).

Si ottiene la serie di coefficienti di  $L^k$ :

$$\psi_k = \frac{(k+d-1)!}{k!(d-1)!}, \quad (3.2.2)$$

o, se si preferisce, in notazione ricorsiva:

$$\psi_k = \psi_{k-1} \frac{k+d-1}{k}, \quad \text{con } \psi_0 = 1. \quad (3.2.3)$$

Consideriamo ora il caso del *fractional white noise*, e scriviamo la (3.2.2) utilizzando la funzione  $\Gamma$ :

$$\psi_k = \frac{\Gamma(k+d)}{\Gamma(k+1)\Gamma(d)}. \quad (3.2.4)$$

Se applichiamo l'approssimazione di Stirling, per la quale si rimanda alla (2.1.4), alla (3.2.4) otteniamo:

$$\psi_k \cong \frac{k^{d-1}}{\Gamma(d)}. \quad (3.2.5)$$

La rappresentazione MA ( $\infty$ ) per un generico  $\{x_t\}$  è:  $x_t = \psi(L)\varepsilon_t$ . Riferendoci al caso di cui sopra, e ricordando di trattare  $L$  come scalare, ricaviamo che  $\psi(z) = (1-z)^{-d}$ . Se  $d < 1/2$  e  $|z| \leq 1$ , l'espansione di  $\psi(L)$  converge, per cui  $\{x_t\}$  è stazionario.

---

<sup>4</sup> Si osservi che in questo caso la sommatoria è composta da un numero infinito di termini. Difatti, nel prodotto del coefficiente di  $L^k$ , per  $k > d$ , non si incontra mai il termine  $d-d=0$ , per cui tali coefficienti non si annullano come avveniva nel caso di  $d$  naturale.

Ugualmente, sostituendo  $d$  a  $-d$ , e osservando che la condizione di invertibilità è  $d > -\frac{1}{2}$ , si

ottiene la rappresentazione AR ( $\infty$ ) di  $\{x_t\}$ , cioè  $\varphi(L)x_t = \varepsilon_t$ :

$$\varphi_k = \frac{(k-d-1)!}{k!(-d-1)!},$$

ovvero, per  $d$  reale:

$$\varphi_k = \frac{\Gamma(k-d)}{\Gamma(k+1)\Gamma(-d)}, \quad (3.2.6)$$

che asintoticamente diventa:

$$\varphi_k \cong \frac{k^{-d-1}}{\Gamma(-d)}. \quad (3.2.7)$$

Supponiamo ora, per semplicità, che la varianza del *white noise* valga 1, e limitiamo la nostra analisi all'intervallo  $0 < d < \frac{1}{2}$ . Lo spettro di un generico processo MA ( $\infty$ ) è:

$$s(\omega) = (2\pi)^{-1} \sigma^2 \psi(e^{i\omega}) \psi(e^{-i\omega}). \quad (3.2.8)$$

Sia  $z = e^{-i\omega}$ : richiamando che  $\psi(z) = (1-z)^{-d}$  otteniamo, seguendo Granger e Joyeux (1980):

$$s(\omega) = (2\pi)^{-1} (1-z)^{-d} (1-z^{-1})^{-d}, \quad (3.2.9)$$

indi

$$\begin{aligned} s(\omega) &= (2\pi)^{-1} [(1-z)(1-z^{-1})]^{-d} = \\ &= (2\pi)^{-1} [2 - z - z^{-1}]^{-d} = \\ &= (2\pi)^{-1} [2 - (e^{-i\omega} + e^{i\omega})]^{-d}, \end{aligned}$$

da cui, per la (2.2.7):

$$s(\omega) = (2\pi)^{-1} [2 - 2\cos(\omega)]^{-d},$$

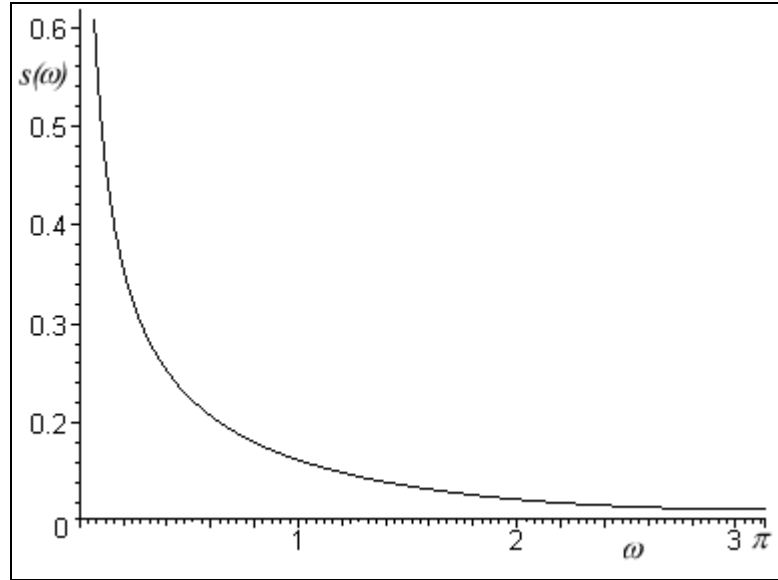
e infine, utilizzando semplici proprietà trigonometriche,

$$s(\omega) = (2\pi)^{-1} [4\sin^2(\frac{1}{2}\omega)]^{-d}, \quad (3.2.10)$$

da cui si ricava immediatamente che

$$\lim_{\omega \rightarrow 0} s(\omega) = \omega^{-2d} . \quad (3.2.11)$$

La (3.2.11) diverge per  $d > 0$ : confrontandola con la (3.1.3) si conclude che il *fractional white noise* ha memoria lunga.



**Figura 3.1**

*Lo spettro di un processo ARFIMA (0, 1/4, 0)*

Passando all'analisi delle autocovarianze, richiamiamo la (2.4.8) e inseriamoci la (3.2.10):

$$\begin{aligned} \gamma_k &= \int_{-\pi}^{+\pi} s(\omega) \cos(\omega k) d\omega = \\ &= \int_0^{\pi} \frac{2^{-d}}{2\pi} \left[ 2 \sin^2\left(\frac{1}{2}\omega\right) \right]^{-d} \cos(\omega k) d\omega , \end{aligned}$$

da cui, utilizzando le soluzioni presentate da Gradshteyn e Ryzhik (1965):

$$\begin{aligned} \gamma_k &= \frac{2^{-d} \cdot 2^{1+d}}{2\pi} \sin(\pi d) \cdot \Gamma(1-2d) \frac{\Gamma(k+d)}{\Gamma(k+1-d)} = \\ &= \frac{\sin(\pi d)}{\pi} \cdot \Gamma(1-2d) \frac{\Gamma(k+d)}{\Gamma(k+1-d)} , \end{aligned}$$

e infine, utilizzando la (2.1.2):

$$\gamma_k = \frac{\Gamma(1-2d)}{\Gamma(d)\Gamma(1-d)} \frac{\Gamma(k+d)}{\Gamma(k+1-d)} . \quad (3.2.12)$$

La (3.2.12) può essere anche scritta come (si veda Hosking, 1981 – pagina 168):

$$\gamma_k = \frac{(-1)^k \Gamma(-2d+1)}{\Gamma(k-d+1)\Gamma(-k-d+1)}. \quad (3.2.13)$$

Sulla base di (3.2.12) si ha  $\gamma_0 = \frac{\Gamma(1-2d)}{\Gamma^2(1-d)}$  e quindi, passando alle autocorrelazioni:

$$\begin{aligned} \rho_k &= \frac{\gamma_k}{\gamma_0} = \frac{\Gamma(1-2d)}{\Gamma(d)\Gamma(1-d)} \frac{\Gamma(k+d)}{\Gamma(k+1-d)} \frac{\Gamma^2(1-d)}{\Gamma(1-2d)} = \\ &= \frac{\Gamma(1-d)\Gamma(k+d)}{\Gamma(d)\Gamma(k-d+1)} \cong \end{aligned} \quad (3.2.14)$$

$$\cong \frac{\Gamma(1-d)}{\Gamma(d)} k^{2d-1}. \quad (3.2.15)$$

Dalla (3.2.15) risalta l'altra importante caratteristica del *fractional white noise*, cioè l'andamento delle autocorrelazioni. Infatti si nota immediatamente che, per  $0 < d < \frac{1}{2}$ , le autocorrelazioni decadono iperbolicamente (soddisfacendo, di nuovo, la condizione di memoria lunga enunciata in (3.1.1)).

Per quanto riguarda il caso  $-\frac{1}{2} < d < 0$ , non rilevante ai fini della nostra analisi, limitiamoci ad osservare che ciò indica un processo a memoria corta e “antipersistente”, secondo la definizione di Mandelbrot (1977). Le autocorrelazioni sono tutte negative (a parte, ovviamente,  $\rho_0$ , che vale 1) e tendono a zero al crescere di  $k$ , mentre la densità spettrale è dominata da frequenze alte e tende a zero al decrescere di  $\omega$ .

### 3.2.2 ARFIMA ( $p, d, q$ )

Dopo aver identificato le proprietà del più semplice processo a integrazione frazionaria, è naturale procedere oltre e cercare di complicare le cose. In particolare, possiamo combinare a una rappresentazione a integrazione frazionaria la consueta modellistica ARMA. Così facendo otterremo un processo “a due velocità”: la dipendenza di breve periodo sarà catturata dai

coefficienti ARMA, quella di lungo periodo dalla presenza del coefficiente di integrazione frazionaria  $d$ .

Sulla base della consueta notazione di Box e Jenkins (1976) definiamo il processo  $\{y_t\}$  come ARFIMA  $(p, d, q)$  se il suo DGP è:

$$\varphi(L)(1-L)^d(y_t - \mu) = \psi(L)\varepsilon_t, \quad (3.2.16)$$

con  $\varepsilon_t$ , come di consueto, *white noise*, per semplicità a varianza unitaria. Possiamo aspettarci che un processo ARFIMA  $(p, d, q)$  goda delle stesse proprietà asintotiche del *fractional white noise*, visto che la differenza sta solo nella dipendenza di breve periodo. In effetti è dimostrato sempre da Hosking (1981) che:

- $\{y_t\}$  è stazionario se  $d < 1/2$  e tutte le radici di  $\varphi(z)$  giacciono fuori dal cerchio unitario;
- $\{y_t\}$  è invertibile se  $d > -\frac{1}{2}$  e tutte le radici di  $\psi(z)$  giacciono fuori dal cerchio unitario;
- $\lim_{\omega \rightarrow 0} \omega^{2d} s(\omega)$  esiste ed è finito;
- $\lim_{k \rightarrow +\infty} k^{1-2d} \rho_k$  esiste ed è finito.

### 3.2.3 Le distribuzioni asintotiche

Le distribuzioni asintotiche degli stimatori di media, varianza e autocovarianza di un processo ARFIMA sono state ricavate da Hosking (1996). Tali distribuzioni si rivelano utili per fare inferenza su queste grandezze: si utilizzerà ad esempio la distribuzione asintotica delle autocovarianze per la stima del coefficiente  $d$  secondo il metodo della minima distanza, che esamineremo nel paragrafo 4.3.5.

Ci limiteremo a presentare i risultati ottenuti omettendo le dimostrazioni, spesso assai complesse, per le quali si può vedere, oltre all'articolo già citato, Hosking (1984b e 1994).

Sia  $\{y_t\}$  un generico processo tale che  $\lim_{k \rightarrow +\infty} \gamma_k = Ak^{-d}$ , con  $A$  costante positiva e  $d \in ]0,1[$ ; un processo ARFIMA, sulla base della (3.2.15), soddisfa questa condizione.

Per quanto riguarda la media campionaria, si dimostra che:

- $\bar{y}$  è corretto per  $\mu$ ;
- $\text{var}(\bar{y}) = \lim_{n \rightarrow +\infty} \frac{2An^{-d}}{(1-d)(2-d)}$ ;
- $n^{d/2}(\bar{y} - \mu) \xrightarrow{d} N\left(0, \frac{2A}{(1-d)(2-d)}\right)$ , per  $n \rightarrow +\infty$ .

Siano ora  $\hat{\gamma}_k = \frac{1}{n} \sum_{t=1}^{n-k} (y_t - \bar{y})(y_{t+k} - \bar{y})$  e  $\tilde{\gamma}_k = \frac{1}{n} \sum_{t=1}^{n-k} (y_t - \mu)(y_{t+k} - \mu)$ :

- $\hat{\gamma}_k$  ha distorsione asintotica pari a  $-\frac{2An^{-d}}{(1-d)(2-d)}$ ;
- La covarianza asintotica di  $\hat{\gamma}_k$  è, per  $0 < d < 1/2$  (si veda l'articolo originale per gli altri casi):

$$2A^2 n^{-2d} \int_0^1 \int_0^1 g(x_1, x_2) g(x_2, x_1) dx_1 dx_2, \quad (3.2.17)$$

$$\text{dove } g(x, y) = |x - y|^{-d} - \frac{[x^{1-d} + (1-x)^{1-d} + y^{1-d} + (1-y)^{1-d}]}{1-d} + \frac{2}{(1-d)(2-d)}.$$

Sia ora  $\hat{\rho}_k = \frac{\hat{\gamma}_k}{\hat{\gamma}_0}$ . Si dimostra che:

- $\hat{\rho}_k$  ha distorsione asintotica pari a  $-\frac{2(1-\rho_k)}{(1-d)(2-d)} \frac{A}{\gamma_0} n^{-d}$ ;
- La covarianza asintotica di  $\hat{\rho}_k$  è, per  $0 < d < 1/2$  (di nuovo si rimanda all'articolo originale per gli altri casi):

$$2\left(\frac{A}{\gamma_0}\right)^2 (1-\rho_k)(1-\rho_l) n^{-2d} \int_0^1 \int_0^1 g(x_1, x_2) g(x_2, x_1) dx_1 dx_2. \quad (3.2.18)$$



Uno dei risultati più “spiacevoli” che Hosking ottiene è che le distribuzioni asintotiche di  $\hat{\gamma}_k$  e  $\hat{\rho}_k$ , che per  $d \in \left[\frac{1}{2}, 1\right[$  risultano normali, non godono di tale importante caratteristica nell’intervallo che più ci interessa, cioè  $[0, \frac{1}{2}]$ , ma hanno una distribuzione più complessa (cumulata di Rosenblatt modificata).

### 3.3 L'individuazione dell'integrazione frazionaria

#### 3.3.1 Il test $R/S$

Il primo test per memoria lunga è stato proposto da Hurst (1951). Si introduce la statistica

*rescaled range* definita come  $\frac{R_T}{\hat{\sigma}_T}$ , dove:

$$R_T = \max_{0 \leq k \leq T} \left\{ \sum_{j=1}^k (y_j - k\bar{y}) \right\} - \min_{0 \leq k \leq T} \left\{ \sum_{j=1}^k (y_j - k\bar{y}) \right\}. \quad (3.3.1)$$

Hurst (1951) dimostra che vale:

$$\text{plim}_{T \rightarrow +\infty} \left( T^{-H} \frac{R_T}{\hat{\sigma}_T} \right) = k, \quad (3.3.2)$$

approssima la (3.3.2):

$$\ln \left[ E \left( \frac{R_T}{\hat{\sigma}_T} \right) \right] \cong k + H \ln T, \quad (3.3.3)$$

e stima  $H$  come

$$\hat{H} = \frac{\ln \left( \frac{R_T}{\hat{\sigma}_T} \right)}{\ln T}. \quad (3.3.4)$$

Poiché un processo stazionario deve avere  $H = 1/2$ , un valore di  $\hat{H}$  significativamente superiore è considerato indice di memoria lunga. Uno dei principali problemi di questo approccio è il fatto che è sensibile anche alla dipendenza di breve termine, per cui indica come processi a memoria lunga anche processi che hanno solo dipendenze di breve termine, ad esempio un AR (1) (si veda Lo, 1991 – pagina 1288).

Il modo per ovviare a questo problema è modificare la statistica test in modo da farla deviare solo per i processi a memoria lunga. Lo (1991) propone di utilizzare:

$$Q_T = \frac{R_T}{s^2(l)}, \quad (3.3.5)$$

dove  $s^2(l)$  è lo stimatore della varianza di Newey e West (1987) presentato nella (2.3.14) e  $l < T$ . Siccome il denominatore include anche le prime  $q$  autocovarianze, l'effetto “dipendenza a breve termine” sarà catturato proprio da queste, e sarà il discriminante per gli effetti di memoria lunga che ci interessa isolare.

Abbiamo già detto che il parametro  $H$  caratterizza, nel moto Browniano frazionario di Hurst, l'intensità della memoria lunga. Geweke e Porter-Hudak (1983 – pagina 224) dimostrano che, nel “passaggio” tra la modellazione di Hurst, in tempo continuo, e il *fractional white noise*, definito in tempo discreto, vale la seguente relazione:

$$d = H - \frac{1}{2}. \quad (3.3.6)$$

Questa relazione “trasforma” il test  $R/S$  (con o senza le modifiche di Lo) in un metodo di stima per  $d$ :

$$\hat{d} = \hat{H} - \frac{1}{2} = \frac{\ln \hat{Q}_T}{\ln T} - \frac{1}{2}. \quad (3.3.7)$$

### 3.3.2 Il test KPSS

Uno dei principali problemi nella ricerca di radici unitarie condotta attraverso i test ADF (Dickey e Fuller, 1979) e PP (Phillips e Perron, 1988), per la definizione dei quali si rimanda al paragrafo 2.3.5, è la scarsa potenza dei test suddetti, cioè la loro scarsa attitudine a rifiutare l'ipotesi nulla quando essa è falsa. Se ciò è vero quando l'alternativa è  $I(0)$ , diventa ancora più difficile fare distinzioni nel caso in cui ci si trovi di fronte a un'alternativa  $I(d)$  (si vedano, per un'indagine su questo argomento, Sowell, 1990, e Hassler e Wolters, 1994).

Un modo per ovviare alla bassa potenza dei test ADF e PP è accoppiarli a un nuovo test, stavolta di stazionarietà: se uno dei due non rifiuta  $H_0$  e l'altro la rifiuta dovremmo essere abbastanza sicuri del risultato.

Kwiatkowski, Phillips, Schmidt e Shin (1992) modificano e generalizzano una statistica ricavata da Nabeya e Tanaka (1988). Sia  $\{y_t\}$  la serie da testare, risultante come somma di un *random walk* e un *white noise*:

$$\begin{aligned} y_t &= \xi t + r_t + \varepsilon_t, \\ r_t &= r_{t-1} + u_t. \end{aligned} \quad (3.3.8)$$

Si ha stazionarietà quando il termine *random walk* è ininfluenza, cioè formalmente:

$$H_0 : \sigma_u^2 = 0. \quad (3.3.9)$$

Il test è di tipo LM (si veda il paragrafo 4.2.2) e la statistica che si utilizza è:

$$LM = \sum_{t=1}^T \frac{S_t^2}{\hat{\sigma}_e^2}, \quad (3.3.10)$$

dove  $S_t$  è la somma dei residui, fino a  $t$ , di una regressione di  $y_t$  sulla (3.3.8) privata del termine *random walk*:

$$S_t = \sum_{i=1}^t e_i, \quad (3.3.11)$$

e  $\hat{\sigma}_e^2$  è la varianza stimata dei residui di questa regressione.

Per la stima di  $\hat{\sigma}_e^2$  Kwiatkowski, Phillips, Schmidt e Shin (1992) propongono l'utilizzo dello stimatore di Newey e West (1987) che abbiamo introdotto nella (2.3.14) e che qui riportiamo per comodità:

$$s^2(l) = \frac{\sum_{t=1}^T e_t^2}{T} + 2 \frac{\sum_{s=1}^l \left(1 - \frac{s}{l+1}\right) \sum_{t=s+1}^T e_t e_{t-s}}{T}. \quad (3.3.12)$$

Definiamo ora:

$$\sigma^2 = \lim_{T \rightarrow +\infty} T^{-1} E(S_T^2). \quad (3.3.13)$$

Se gli errori non hanno distribuzione normale, la (3.3.12) è una stima di  $\sigma^2$  invece che di  $\sigma_e^2$ . Sia ora:

$$\eta = T^{-2} \sum_{t=1}^T S_t^2, \quad (3.3.14)$$

e sia  $\eta_\mu$  definito come nella (3.3.14), ma nel caso più semplice in cui si escluda la possibilità di trend deterministici nella serie. Si può dimostrare (si veda Kwiatkowski, Phillips, Schmidt e Shin, 1992 – pagina 165) che  $\eta_\mu$  converge a un ponte Browniano moltiplicato per  $\sigma^2$ , ovvero:

$$\eta_\mu \xrightarrow{d} \sigma^2 \int_0^1 V(r)^2 dr, \quad (3.3.15)$$

dove  $V(r) = B(r) - rB(1)$ . Ciò significa che dividere  $\eta_\mu$  per una stima consistente di  $\sigma^2$  dà una distribuzione asintoticamente priva di disturbi.

Sia ora:

$$\hat{\eta}_\mu = \frac{\eta_\mu}{s^2(l)} = \frac{1}{T^2} \frac{\sum_{t=1}^T S_t^2}{s^2(l)}; \quad (3.3.16)$$

è immediato osservare che, sulla base della (3.3.15):

$$\hat{\eta}_\mu \xrightarrow{d} \int_0^1 V(r)^2 dr. \quad (3.3.17)$$

Si può dimostrare senza troppe difficoltà (si veda l'articolo già citato – pagina 167) che la (3.3.16) vale anche nel caso in cui si abbia nella serie trend deterministico, anche se in questo caso la distribuzione asintotica è diversa è leggermente più complessa.

Senza scendere in ulteriori dettagli riguardanti la consistenza, che risulta  $O(T/l)$ , e la potenza del test, per i quali si rimanda all'articolo citato (pagine 167 e seguenti), introduciamo ora i risultati di Lee e Schmidt (1996), che applicano il test KPSS ai processi  $I(d)$ .

Lee e Schmidt (1996 – pagina 290) dimostrano la consistenza di  $\eta_\mu$  e del suo “gemello” per serie con trend deterministici  $\eta_\tau$ , in particolare che sono  $O(T/l)^{2d}$ , e che quindi, per  $d > 0$  e  $T \rightarrow +\infty$ , convergono in probabilità a infinito.

Per quanto riguarda la potenza del test, si ricavano i seguenti risultati, nell'intervallo  $0 < d < 1/2$ :

- Poiché il test è consistente, la potenza cresce al crescere di  $T$ ;
- La potenza è maggiore quando  $d$  è più grande, cioè quando l'ipotesi alternativa è più distante dalla nulla;
- La potenza di  $\hat{\eta}_\mu$  e quella di  $\hat{\eta}_\tau$  non sono dissimili, il che significa che inserire la possibilità che vi sia un trend deterministico non è “costoso” in termini di potenza;
- La potenza decresce al crescere di  $l$ , indipendentemente dalla dimensione di  $T$ , ciò si poteva intuire dal fatto che i test sono  $O(T/l)^{2d}$ .

### 3.3.3 Il test Durbin-Watson adattato

L'ultima metodologia di test per l'integrazione frazionaria è proposta da Tsay (1998), che considera la possibilità di utilizzare il test di Durbin e Watson (1950).

Il test DW è utilizzato per testare la correlazione tra i residui:

$$DW = \frac{\sum_{t=2}^T (e_t - e_{t-1})^2}{\sum_{t=1}^T e_t^2} = 2(1 - \hat{\rho}_T) - \frac{e_1^2 + e_T^2}{\sum_{t=1}^T e_t^2}; \quad (3.3.18)$$

se il campione è sufficientemente grande il secondo addendo nella (3.3.18) è trascurabile:

$$DW \cong 2(1 - \hat{\rho}_T), \quad (3.3.19)$$

per cui, in assenza di autocorrelazione, si ha un valore di  $DW$  prossimo a due. Questa statistica ha una distribuzione irregolare, ma si è dimostrato che il valore critico di  $DW$  per un certo livello di significatività è compreso tra due valori  $DW_l$  e  $DW_u$ , che dipendono da  $T$  e dal numero di parametri, e che sono stati tabulati da Savin e White (1977).

La statistica  $DW$  è stata modificata da Nabeya e Tanaka (1990) per essere utilizzata come test per radice unitaria:

$$\begin{cases} MDW = \frac{S_1}{S_2} \frac{1}{T \cdot DW} \\ S_1 = \frac{1}{T} \sum_{t=2}^T (e_t - e_{t-1})^2 \\ S_2 = S_1 + \frac{2}{T} \sum_{s=1}^l \left(1 - \frac{s}{l+1}\right) \sum_{t=2}^T (e_t - e_{t-1})(e_{t-s} - e_{t-s-1}), \end{cases} \quad (3.3.20)$$

dove  $l$  indica il numero di ritardi a cui si tronca la serie.

Tsay (1998) studia, su dati simulati, il comportamento della statistica  $MDW$  su un ipotesi alternativa  $I(d)$  invece che  $I(1)$ . Vengono utilizzati tre modelli: il primo con  $d \in \left] \frac{1}{2}, \frac{3}{2} \right[$ , il

secondo con  $d = \frac{1}{2}$ , il terzo con  $0 < d < \frac{1}{2}$ , e si ricava che, per  $T - l \rightarrow +\infty$ :

- $MDW = O\left(\left(\frac{T}{l}\right)^{2(d-1)}\right)$  per  $d \in \left] \frac{1}{2}, \frac{3}{2} \right[$ ,
- $MDW = O\left(\frac{l}{T} \frac{\ln T}{\ln l}\right)$  per  $d = \frac{1}{2}$ ,
- $MDW = O\left(\frac{l}{T}\right)$  per  $0 < d < \frac{1}{2}$ .

La distribuzione asintotica di  $MDW$  è ricavata, sulla base dei risultati di Sowell (1990), per

$$d \in \left] \frac{1}{2}, \frac{3}{2} \right[ \text{ e } T - l \rightarrow +\infty:$$

$$\begin{aligned} \left(\frac{l}{T}\right)^{2d} MDW \xrightarrow{d} & \int_0^1 B_d^2(s) ds - \left[ \int_0^1 B_d(s) ds \right]^2 + \\ & -12 \left[ \int_0^1 s B_d(s) ds - \frac{1}{2} \int_0^1 B_d(s) ds \right]^2, \end{aligned} \quad (3.3.21)$$

dove  $B_d$  indica un moto Browniano frazionario.

La performance della statistica  $MDW$  è molto promettente. Tsay (1998) la compara, su dati simulati, con la  $R/S$  modificata di Lo (1991). Si osserva, com'era prevedibile, che la potenza della statistica  $MDW$  aumenta all'aumentare del campione e diminuisce all'aumentare di  $l$ . In genere, la potenza di  $MDW$  è superiore a quella di  $R/S$ , specie nel caso in cui  $d \in \left] \frac{1}{2}, \frac{3}{2} \right[$ ; si deve comunque osservare che in campioni piccoli si riscontrano gli stessi problemi di distorsione.



### 3.4 Modelli teorici

A questo punto abbiamo definito una nuova classe di modelli che si presta bene a modellare processi a memoria lunga e ne abbiamo ricavato le proprietà principali. Restano da analizzare due distinte problematiche:

- Quali e quanti sono i benefici apportati dall'uso di un modello ARFIMA al posto di un più semplice ARMA? Ovvero, esiste nella realtà evidenza di comportamento  $I(d)$ ?
- Da cosa può essere generata una modellazione complessa come quella ARFIMA? Esistono cioè modelli teorici che possano spiegare l'insorgenza di comportamento  $I(d)$ ?

Ci occupiamo in questa sezione di rispondere alla seconda domanda proponendo due distinti modelli che forniscono una spiegazione teorica sull'occorrenza dell'integrazione frazionaria.

#### 3.4.1 *L'aggregazione di modelli dinamici*

Abbiamo già detto che i modelli a integrazione frazionaria si sono sviluppati al di fuori delle scienze economiche, e sono entrati in questo campo solo negli anni '80. Il lavoro che ha fatto nascere interesse per questo tipo di modellistica, e che soprattutto ne ha messo in risalto l'importanza ai fini dell'analisi delle serie storiche economiche, è di Granger (1980): si mostra che fenomeni di integrazione frazionaria possono nascere dall'aggregazione di variabili economiche. Come abbiamo già rilevato nell'introduzione, visto che buona parte delle variabili che trattano i macroeconomisti risultano da aggregazioni, si comprende l'importanza che può assumere la modellazione ARFIMA.

Seguiamo i passi di Granger (1980) e supponiamo di avere  $N$  serie indipendenti, ciascuna governata da un processo AR (1). È dimostrabile (Granger e Morris, 1976) che la serie che risulta dalla loro aggregazione è ARMA ( $N, N - 1$ ). Poiché la maggior parte delle variabili macroeconomiche risulta dall'aggregazione di "microvariabili" (si pensi ad esempio al

reddito, o al consumo), questo risultato suggerirebbe che le variabili macroeconomiche siano governate da processi con un enorme numero di parametri. È importante ribadire che i risultati che saranno ricavati valgono solo per quanto riguarda l'aggregazione di processi AR (1), ma non MA (1):  $\bar{x}_t = \sum_{j=1}^N x_{j,t}$ , con  $x_{j,t} = \beta_j \varepsilon_{j,t-1} + \varepsilon_{j,t}$  è ancora un MA (1).

Definiamo la questione in maniera più formale. Siano le  $N$  serie indipendenti  $x_{j,t} = \alpha_j x_{j,t-1} + \varepsilon_{j,t}$  con  $j = 1, 2, \dots, N$  e sia  $\bar{x}_t = \sum_{j=1}^N x_{j,t}$ . Lo spettro di  $x_{j,t}$  è, utilizzando la (2.4.7):

$$s_j(\omega) = \frac{1}{|1 - \alpha_j z|^2} \cdot \frac{\text{var}(\varepsilon_{j,t})}{2\pi} \quad \text{con } z = e^{-i\omega}, \quad (3.4.1)$$

Poiché le componenti sono indipendenti, lo spettro di  $\bar{x}$  può essere ottenuto sommando gli  $s_j$ . Se assumiamo che i coefficienti  $\alpha_j$  e le varianze di  $\varepsilon_{j,t}$  siano variabili aleatorie, con funzioni di ripartizione  $F(\alpha)$  e  $G(\varepsilon_t)$ , tra di loro indipendenti, possiamo scrivere la seguente approssimazione:

$$\bar{s}(\omega) \cong \frac{N}{2\pi} E[\text{var}(\varepsilon_{j,t})] \cdot \int \frac{1}{|1 - \alpha z|^2} dF(\alpha). \quad (3.4.2)$$

Se  $\alpha$  è discreta e può assumere  $m$  possibili valori,  $\bar{s}(\omega)$  è lo spettro di un processo ARMA  $(m, m - 1)$ , ma se  $\alpha$  può assumere qualsiasi valore, allora a  $\bar{s}(\omega)$  non corrisponde alcuna rappresentazione ARMA finita.

A questo punto sono necessarie delle ipotesi sulla distribuzione di  $\alpha$ . Utilizzeremo, per avere la massima generalità possibile, la distribuzione Beta che abbiamo introdotto in (2.1.6), limitandoci così alla regione  $\alpha \in [0, 1]$ . Tale limitazione è giustificata dal fatto che al di fuori di tale intervallo si ottengono processi antipersistenti oppure esplosivi, poco interessanti per la nostra analisi. Si avrà:

$$dF(\alpha) = \frac{2}{B(p, q)} \alpha^{2p-1} (1-\alpha^2)^{q-1} d\alpha. \quad (3.4.3)$$

Osservando che:

$$\frac{1}{|1-\alpha z|^2} = \frac{1}{(1-\alpha)^2} \left[ \frac{1+\alpha z}{1-\alpha z} + \frac{1+\alpha \bar{z}}{1-\alpha \bar{z}} \right], \quad (3.4.4)$$

e che:

$$\frac{1+\alpha z}{1-\alpha z} = 1 + 2 \sum_{k=1}^{+\infty} (\alpha z)^k. \quad (3.4.5)$$

Utilizzando la (3.4.4), la (3.4.5) e la (3.4.2) possiamo ricavare il coefficiente di  $z^k$  in  $\bar{s}(\omega)$ :

$$\frac{2}{B(p, q)} \int_0^1 \alpha^{2p+k-1} (1-\alpha^2)^{q-2} d\alpha, \quad (3.4.6)$$

che per la (2.4.2) risulta essere  $\bar{\gamma}_k$ , cioè la k-esima autocovarianza di  $\bar{x}_t$ . La (3.4.6) può essere

scritta anche come:

$$\begin{aligned} \bar{\gamma}_k &= \frac{B(p+k/2, q-1)}{B(p, q)} = \\ &= \frac{\Gamma(q-1)}{B(p, q)} \frac{\Gamma(p+k/2)}{\Gamma(p+k/2+q-1)}, \end{aligned} \quad (3.4.7)$$

che utilizzando l'approssimazione di Stirling diventa, per  $k \rightarrow +\infty$ :

$$\bar{\gamma}_k \cong \frac{\Gamma(q-1)}{B(p, q)} (p+k/2)^{1-q} = A k^{1-q}, \quad (3.4.8)$$

con A termine costante, funzione di  $p$  e  $q$ . Confrontando la (3.4.8) con la (3.2.15) si osserva immediatamente che:

$$\bar{x}_t \sim I(1-q/2). \quad (3.4.9)$$

Poiché il fattore di integrazione frazionaria è  $1-q/2$ , è importante notare che il processo ha varianza finita solo se  $q > 1$ .

A questo punto si può rimuovere l'assunto che gli  $\varepsilon_{j,t}$  siano *white noise*. I processi  $\{x_{j,t}\}$  saranno generati da  $x_{j,t} = \alpha_j x_{j,t-1} + y_{j,t}$ , e lo spettro di  $\{y_{j,t}\}$  dipenderà da un certo vettore di parametri  $\zeta_j$ . Si supponga inoltre che gli  $\{y_{j,t}\}$  siano indipendenti e che gli  $\alpha$  e gli  $\zeta$  provengano da distribuzioni indipendenti. La (3.4.2) diventa:

$$\bar{s}(\omega) \cong NE_{\zeta} \left[ s_y(\omega, \zeta) \right] \int \frac{1}{|1 - \alpha z|^2} dF(\alpha). \quad (3.4.10)$$

Poiché  $\bar{y}_t = \sum_{j=1}^N y_{j,t} \sim I(d_y)$ , si conclude immediatamente che:

$$\bar{x}_t \sim I(d_y + 1 - q). \quad (3.4.11)$$

Supponiamo ora che le  $x_{j,t}$  non siano più indipendenti, ma generate dal semplice modello

$$x_{j,t} = \alpha_j x_{j,t-1} + \beta_j w_t. \text{ Si ha allora che } \bar{x}_t = \left( \sum_{j=1}^N \frac{\beta_j}{1 - \alpha_j L} \right) w_t, \text{ che può essere riportata nel}$$

continuo a:

$$\bar{x}_t \cong NE(\beta) \left( \int \frac{1}{1 - \alpha L} dF(\alpha) \right) w_t. \quad (3.4.12)$$

Utilizzando la (3.4.3), la (3.4.4) e la (3.4.5) sulla (3.4.10) otteniamo:

$$\bar{x}_t \cong NE(\beta) \left( \sum_{k=0}^{+\infty} L^k \frac{B(p + k/2, q)}{B(p, q)} \right) w_t, \quad (3.4.13)$$

per cui il coefficiente di  $L^k$ , per  $k \rightarrow +\infty$ , è di ordine  $k^{-q}$ . Se supponiamo che  $w_t \sim I(d_w)$ , otteniamo dalla (3.2.7) che

$$\bar{x}_t \sim I(1 - q + d_w). \quad (3.4.14)$$

Passando a un modello con variabili dipendenti più generale, espresso nella forma

$$x_{j,t} = \alpha_j x_{j,t-1} + y_{j,t} + \beta_j w_t + \varepsilon_{j,t}, \text{ si può dimostrare (si veda Granger, 1980 – pagina 235) che,}$$

se  $\bar{y}_t \sim I(d_y)$  e  $w_t \sim I(d_w)$ , allora  $\bar{x}_t \sim I(d_x)$ , dove  $d_x$  è il maggiore tra  $1 - q/2$ , che proviene

dalla (3.4.9),  $1 - q/2 + d_y$ , dalla (3.4.11), e  $1 - q + d_w$ , dalla (3.4.14).

### 3.4.2 Il modello Error-Duration

Un altro tipo di analisi interessante per spiegare l'occorrenza dei modelli a memoria lunga in generale è quella proposta da Parke (1999).

Sia  $\varepsilon_t$ , con  $t = 1, 2, \dots, T$ , una serie di shock indipendenti e identicamente distribuiti con media zero e varianza finita e costante  $\sigma^2$ . Supponiamo che ogni shock  $\varepsilon_s$  abbia una durata stocastica  $n_s$ , cioè che persista tra  $s$  e  $s + n_s$ , e chiamiamo  $g_{s,t}$  la funzione indicatore dell'evento "persistenza dello shock  $\varepsilon_s$ ". Si avrà in termini più formali  $g_{s,t} = 1$  per  $t \in [s, s + n_s]$  e  $g_{s,t} = 0$  altrimenti.

La realizzazione al tempo  $t$  di un processo generato a partire da questi shock sarà:

$$y_t = \sum_{s=-\infty}^t g_{s,t} \varepsilon_s. \quad (3.4.15)$$

Abbiamo già detto che la quantità  $n_s$  è stocastica. Sia  $p_k$  la probabilità che  $\varepsilon_s$  sopravviva fino a  $s + k$ : è naturale assumere  $p_0 = 1$  e  $p_i \geq p_{i+1}$ . Avremo allora:

$$E(g_{t-i,t}) = p_i, \quad (3.4.16)$$

e di conseguenza:

$$\text{var}(g_{t-i,t}) = p_i(1 - p_i). \quad (3.4.17)$$

Le autocovarianze di un processo di questo tipo saranno date, seguendo la formulazione generale, da:

$$\gamma_k = E \left[ \sum_{i=0}^{+\infty} g_{t-i,t} \varepsilon_{t-i} \right] \left[ \sum_{j=k}^{+\infty} g_{t-j,t-k} \varepsilon_{t-j} \right]. \quad (3.4.18)$$

Gli shock sono indipendenti, per cui i prodotti incrociati in cui compaiono  $\varepsilon_{t-i}$  e  $\varepsilon_{t-j}$  con  $i \neq j$  si annullano. I termini lungo la diagonale risultano essere del tipo:

$$E(g_{t-i,t}^2 \varepsilon_{t-i}^2) = E(g_{t-i,t}^2) E(\varepsilon_{t-i}^2);$$

osserviamo che  $E(\varepsilon_{t-i}^2) = \sigma^2$  e che  $E(g_{t-i,t}^2) = \text{var}(g_{t-i,t}) + E(g_{t-i,t})^2$ , richiamiamo la (3.4.16) e la (3.4.17) e otteniamo

$$\gamma_k = \sigma^2 \sum_{j=k}^{+\infty} p_j, \quad (3.4.19)$$

da cui si ricava direttamente che:

$$p_k = \frac{\gamma_k - \gamma_{k+1}}{\sigma^2}. \quad (3.4.20)$$

A questo punto richiamiamo la definizione di memoria lunga secondo McLeod e Hipel (1978) e sostituiamo nella (3.1.1) la (3.4.19): otteniamo che il processo  $\{y_t\}$  ha memoria lunga se la

quantità  $\sum_{k=1}^{+\infty} k p_k$  non è finita. Se vogliamo avere un processo a memoria lunga, dobbiamo

intuitivamente fare in modo che gli shock “durino abbastanza” e non vengano dimenticati presto. Questo significa che le  $p$  non devono decadere troppo velocemente; un decadimento iperbolico potrebbe essere adeguato. Supponiamo che le probabilità di sopravvivenza  $p$  siano tali che  $p_{k-1} = k^{2d-1}$ : si ha che il processo ha memoria lunga per  $d > 0$ , cioè esattamente quanto accade per i processi ARFIMA.

### 3.5 Risultati empirici

Siamo ora pronti per rispondere alla prima delle due domande che ci siamo posti nell'introduzione alla sezione precedente, individuando l'evidenza di comportamento  $I(d)$  in alcune importanti categorie di serie storiche. Non citeremo le ricerche condotte nel campo dei rendimenti di attivi finanziari misurati ad alta frequenza, che saranno presentate nel capitolo 5.

Come abbiamo già avuto modo di rimarcare, lo studio dei fenomeni di memoria lunga è nato all'interno delle scienze naturali, in particolare dell'idrologia: non stupisce quindi che le prime applicazioni apparse in letteratura riguardino questo campo. In particolare, la serie in assoluto più nota agli studiosi di memoria lunga è quella sul livello minimo raggiunto del Nilo nei pressi del Cairo, registrato annualmente tra gli anni 622 e 1281. La serie (si veda la figura 4.1) appare, nel suo complesso, stazionaria; se però si esaminano sottoperiodi sembrano essere presenti tendenze o cicli locali. Sulla base di questi dati, il lavoro di Hurst (1951) esaminava il problema di come costruire un bacino di riserva di dimensioni ottimali per un certo intervallo di tempo. La statistica  $R/S$  da egli introdotta si comportava però in maniera contraddittoria rispetto a quanto previsto dai risultati sui tradizionali processi stocastici; Hurst (1951) individuò inoltre lo stesso tipo di comportamento in altre serie storiche idrologiche, geofisiche e meteorologiche.

Mandelbrot e Van Ness (1968) proposero un nuovo processo stocastico definito in tempo continuo atto a spiegare il comportamento anomalo delle serie analizzate da Hurst (1951). Tale processo prese il nome di *fractional Brownian motion*.

Come abbiamo già rilevato nell'introduzione, la modellistica a memoria lunga cominciò a destare interesse negli economisti solo negli anni '80. Presentiamo ora una breve rassegna delle applicazioni economiche finora apparse in letteratura.

### 3.5.1 L'integrazione frazionaria nel PNL

Una delle più importanti applicazioni macroeconomiche dei modelli a integrazione frazionaria è quella proposta da Sowell (1992b), già citata nell'introduzione.

Sowell studia l'annosa questione se il PNL sia generato da un modello a radice unitaria con drift:

$$x_t = \mu + x_{t-1} + u_t, \quad (3.5.1)$$

oppure da un modello a trend deterministico:

$$x_t = \mu t + w_t, \quad (3.5.2)$$

dove  $u_t$  e  $w_t$  sono generici processi stazionari, che possono essere espressi, attraverso la rappresentazione di Wold, come MA ( $\infty$ ):

$$u_t = \sum_{j=0}^{+\infty} a_j \varepsilon_{t-j} \text{ e } w_t = \sum_{j=0}^{+\infty} b_j \varepsilon_{t-j}, \quad (3.5.3)$$

con  $\varepsilon_{t-j}$ , come al solito, *white noise*.

Calcoliamo ora le differenze prime di entrambi i modelli. Per il modello (3.5.1) si ha:

$$\Delta x_t = \mu + \sum_{j=0}^{+\infty} a_j \varepsilon_{t-j}, \quad (3.5.4)$$

che risulta evidentemente I(0), mentre per il (3.5.2):

$$\Delta x_t = \mu + (1-L) \sum_{j=0}^{+\infty} b_j \varepsilon_{t-j}, \quad (3.5.5)$$

che è I(-1).

Un modo interessante per vedere quale dei due modelli si adatti meglio ai dati empirici è quello di modellare la serie con un ARFIMA ( $p, d, q$ ) e vedere come si comporta  $d$ . Sowell stima sedici modelli ARMA ( $p, q$ ) e sedici ARFIMA ( $p, d, q$ ), combinando tra loro diverse modellazioni con i ritardi AR e MA variabili tra zero e tre.



Il risultato è che la modellazione più precisa è l'ARFIMA (3,  $d$ , 2), che viene scelta anche come più parsimoniosa dal criterio di Akaike<sup>5</sup> (1973).

A questo punto si impone rispettivamente  $d = 0$  e  $d = -1$  al modello di cui sopra: entrambi i vincoli risultano accettabili al 5% di significatività, ma non al 10%. Ciò indica, oltre all'importanza di una modellazione ARFIMA, che i due modelli teorici sono indistinguibili ai dati.

### 3.5.2 *L'integrazione frazionaria nell'inflazione*

Procediamo ora con un'altra tematica "popolare" (è in questo campo che, per inciso, è nata la modellistica ARCH), cioè l'inflazione. La ricerca di Baillie, Chung e Tieslau (1996) muove dall'osservazione che le autocorrelazioni delle inflazioni calcolate su diversi paesi presentano una bassa velocità di decadimento. D'altra parte, differenziando le serie, le autocorrelazioni risultano sistematicamente negative, segno che il filtro è "troppo forte": di qui l'intuizione che l'inflazione possa seguire un modello ARFIMA.

Per avere indicazioni più chiare si applicano alle serie il test di radice unitaria PP (Phillips e Perron, 1988) e quello di stazionarietà KPSS (Kwiatkowski, Phillips, Schmidt e Shin, 1992). A parte la serie relativa al Giappone, che appare stazionaria, entrambe le ipotesi vengono rifiutate, segno che ci si trova in qualche modo tra la classe  $I(0)$  e quella  $I(1)$ .

Si passa così alla stima di un modello ARFIMA (0,  $d$ , 1) – GARCH (1, 1) per ogni paese. I valori di  $d$  che si ottengono oscillano tra 0.3 e 0.5 (a parte il Giappone, che consistentemente coi risultati del test KPSS presenta un  $d$  non significativo). I modelli presentano anche forte persistenza nella varianza condizionata: la somma dei parametri  $\alpha$  e  $\beta$  si avvicina al cerchio unitario. Ciò può indicare, e lo vedremo meglio in seguito, fenomeni di integrazione frazionaria anche all'interno della varianza condizionata.

---

<sup>5</sup> Si veda, per una trattazione sommaria dei criteri informativi, il paragrafo A.1.1 in appendice.

### 3.5.3 L'integrazione frazionaria nei tassi di interesse

Crato e Rothman (1994) applicano l'approccio di Sowell (1992b) a un insieme più ricco di serie, che comprende anche i rendimenti dei titoli di stato. Mentre sulle altre grandezze si ottengono risultati simili a quelli già osservati da Sowell, la serie dei rendimenti dei titoli di stato, se differenziata, produce dei residui per i quali  $\hat{d} = -0.19$ ; con questo valore sia l'ipotesi di trend deterministico che quella di *random walk* sono rifiutate. Possiamo quindi aspettarci che il processo generatore dei dati sia effettivamente a integrazione frazionaria, con un valore di  $\hat{d} = 0.81$ ; ciò indicherebbe, in accordo con l'esperienza comune, che i rendimenti dei titoli di stato non sono stazionari, ma sono comunque *mean reverting*.

Un'altra ricerca è condotta da Tsay (2000), che analizza il comportamento del tasso di interesse reale degli Stati Uniti. Vengono applicati in primo luogo i test PP e KPSS, che rifiutano entrambi l'ipotesi nulla; ciò è assunto come "indizio" della presenza di integrazione frazionaria. La stima, condotta col metodo della massima verosimiglianza condizionata, rivela infatti la significatività di  $d$ , che assume un valore stimato di 0.666.

## 3.6 I processi di Gegenbauer

### 3.6.1 Il processo di Gegenbauer

Un'interessante generalizzazione dei processi ARFIMA si ha utilizzando, al posto del semplice operatore  $(1-L)^d$ , una formulazione più complessa. Seguiamo i passi di Gray, Zhang e Woodward (1989), e richiamiamo il polinomio di Gegenbauer, che risulta ortogonale nel senso che abbiamo introdotto nel paragrafo 2.1.3:

$$(1 - 2ax + x^2)^{-b}, \quad (3.6.1)$$

dati  $b \neq 0$ ,  $|x| < 1$  e  $|a| \leq 1$ .

Si può dimostrare (Rainville, 1960) che la (3.6.1) può essere riscritta come:

$$\begin{cases} (1 - 2ax + x^2)^{-b} = \sum_{n=0}^{+\infty} C_{n,b}(a)x^n \\ C_{n,b}(a) = \sum_{k=0}^{[n/2]} \frac{(-1)^k \cdot \frac{\Gamma(b+n-k)}{\Gamma(b)} \cdot (2a)^{n-2k}}{k!(n-2k)!} \end{cases} \quad (3.6.2)$$

A partire dalla (3.6.1), il processo di Gegenbauer può essere definito come (assumiamo per semplicità  $\mu = 0$ ):

$$y_t = (1 - 2\xi L + L^2)^{-d} \varepsilon_t, \quad (3.6.3)$$

dove  $\varepsilon_t$ , come di consueto, è un *white noise* a media nulla e varianza unitaria.

Utilizzando la (3.6.2) possiamo espandere la (3.6.3):

$$\begin{cases} y_t = (1 - 2\xi L + L^2)^{-d} \varepsilon_t = \left[ \sum_{n=0}^{+\infty} C_{n,d}(\xi) L^n \right] \varepsilon_t \\ C_{n,d}(\xi) = \sum_{k=0}^{[n/2]} \frac{(-1)^k \cdot \frac{\Gamma(d+n-k)}{\Gamma(d)} \cdot (2\xi)^{n-2k}}{k!(n-2k)!} \end{cases} \quad (3.6.4)$$

Sotto le condizioni di invertibilità che ricaveremo in seguito possiamo scrivere infine:

$$(1 - 2\xi L - L^2)^d y_t = \varepsilon_t, \quad (3.6.5)$$

cioè la (3.6.3) espressa in maniera analoga al *fractional white noise* di Hosking (1981) che abbiamo introdotto nel paragrafo 3.2.1. Si osserva immediatamente che, nel caso in cui  $\xi = 1$ , la (3.6.5) dà un *fractional white noise* di ordine  $2d$ : il processo di Gegenbauer è dunque una sua generalizzazione.

### 3.6.2 Le proprietà salienti

Ci limiteremo, nella dimostrazione delle proprietà del processo di Gegenbauer, al caso  $|\xi| < 1$ .

Il caso in cui  $\xi = \pm 1$ , oltre ad essere meno interessante ai fini di questa ricerca, presenta comunque dimostrazioni molto più semplici, che in buona parte derivano direttamente da quelle Hosking (1981) per il *fractional white noise*.

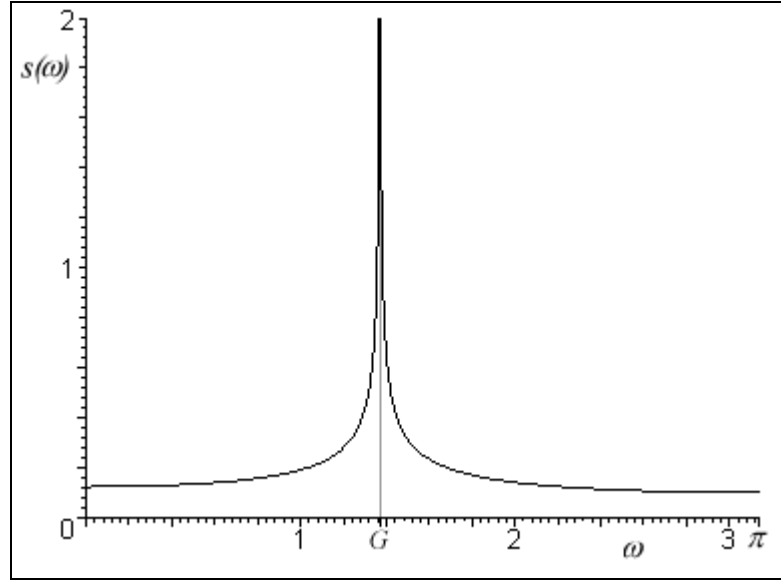
Gray, Zhang e Woodward (1989) dimostrano in primo luogo (si veda l'articolo citato a pagina 236 per la dimostrazione), che il processo di Gegenbauer è stazionario per  $d < 1/2$  se  $|\xi| < 1$  o

$d < 1/4$  se  $\xi = \pm 1$ , e invertibile per  $d > -\frac{1}{2}$  se  $|\xi| < 1$  o  $d > -\frac{1}{4}$  se  $\xi = \pm 1$ .

Lo spettro del processo di Gegenbauer è:

$$\begin{aligned} s(\omega) &= |1 - 2\xi e^{i\omega} + e^{2i\omega}|^{-2d} = \\ &= [4(\cos \omega - \xi)^2]^{-d}. \end{aligned} \quad (3.6.6)$$

Sia ora  $|\xi| < 1$ : dalla (3.6.6) si vede immediatamente che, per frequenze che tendono a  $G = \arccos \xi$ , lo spettro diverge (figura 3.2); ciò significa (confronta la 3.1.3) che il processo ha memoria lunga; la frequenza  $G$  prende il nome di frequenza di Gegenbauer.



**Figura 3.2**

*Lo spettro di un processo di Gegenbauer con  $d = 0.25$  e  $\xi = 0.2$*

Per ricavare la funzione di autocorrelazione dobbiamo fare uso del seguente lemma (per la dimostrazione si veda Gray, Zhang e Woodward, 1989 – pagina 239).

Sia  $R(k) = \int_0^\pi s(\omega) \cos(k\omega) d\omega$ ; sia dato  $\omega_0 \in ]0, \pi[$  e sia lo spettro esprimibile come

$s(\omega) = b(\omega) |\omega - \omega_0|^{-\delta}$ , con  $\delta \in ]0, \frac{1}{2}[$ . Relativamente a  $b(\omega)$ , si ipotizza che sia non

negativo, a variazioni limitate in  $]0, \omega_0 - \varepsilon[ \cup ]\omega_0 + \varepsilon, \pi[$ , con  $\varepsilon$  positivo e sufficientemente piccolo, e che vari lentamente a  $\omega_0$ . Si ha allora che, per  $k \rightarrow +\infty$ :

$$R(k) \cong k^{\delta-1} \sin\left(\frac{\pi\delta}{2} - k\omega_0\right) \left[ b(k^{-1} + \omega_0) + b(-k^{-1} - \omega_0) \right]. \quad (3.6.7)$$

Torniamo ora allo spettro del processo di Gegenbauer, e scriviamo la (3.6.6) come:

$$s(\omega) = \left[ 2 \sin^2\left(\frac{\omega - G}{2}\right) \sin^2\left(\frac{\omega + G}{2}\right) \right]^{-d} = |\omega - G|^{-2d} b(\omega),$$

dove:

$$b(\omega) = 2^{-d} \left[ \sin^2 \left( \frac{\omega + G}{2} \right) \right]^{-d} \left[ \sin \left( \frac{\omega - G}{2} \right) (\omega - G)^{-1} \right]^{-2d}. \quad (3.6.8)$$

Osserviamo ora che, utilizzando la (2.4.8), si ha:

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \frac{2R(k)}{\gamma_0}. \quad (3.6.9)$$

Poiché si dimostra (si veda, di nuovo, Gray, Zhang e Woodward, 1989) che la (3.6.8) gode delle proprietà richieste dal lemma (3.6.7), si ha allora:

$$\rho_k \cong \frac{2k^{2d-1}}{\gamma_0} \sin(\pi d - kG) [b(k^{-1} + G) + b(-k^{-1} - G)]; \quad (3.6.10)$$

poiché si osserva immediatamente che, per  $k \rightarrow +\infty$ , il termine  $b(k^{-1} + G) + b(-k^{-1} - G)$  tende a una costante finita  $A$ , si conclude che:

$$\rho_k \cong \frac{2A}{\gamma_0} k^{2d-1} \sin(\pi d - kG)^6, \quad (3.6.11)$$

cioè che le autocorrelazioni decadono iperbolicamente ma con andamento sinusoidale.

L'ultimo risultato che presentiamo è ricavato da Chung (1996a) e riguarda l'andamento non asintotico delle autocorrelazioni. Si dimostra che vale:

$$\begin{aligned} \gamma_j &= \frac{\sigma^2}{2\sqrt{\pi}} \Gamma(1-2d) (2 \sin G)^{1/2-2d} \times \\ &\times \left[ P_{j-1/2}^{2d-1/2}(\xi) + (-1)^j P_{j-1/2}^{2d-1/2}(-\xi) \right], \end{aligned} \quad (3.6.12)$$

dove  $P_l^m$  è il polinomio di Legendre che abbiamo definito nella (2.1.13).

### 3.6.3 I processi GARMA

Una volta descritto il processo di Gegenbauer, è assolutamente triviale generalizzare il processo ARFIMA  $(p, d, q)$  della (2.2.16) alla forma GARMA  $(p, \xi, d, q)$ :

---

<sup>6</sup> Chung (1996a) osserva però che la (3.6.11) non è corretta per  $G = \pi/2$  e propone di rimpiazzare il termine  $\sin(\pi d - kG)$  con  $\cos(kG)$ .

$$\varphi(L)(1 - 2\xi L + L^2)^d (y_t - \mu) = \psi(L)\varepsilon_t. \quad (3.6.13)$$

Gray, Zhang e Woodward (1989) dimostrano (pagina 243) che per il processo GARMA  $(p, \xi, d, q)$  valgono sostanzialmente le proprietà del processo di Gegenbauer:

- Stazionarietà per  $d < 1/2$  se  $|\xi| < 1$  o  $d < 1/4$  se  $\xi = \pm 1$ ;
- Invertibilità per  $d > -\frac{1}{2}$  se  $|\xi| < 1$  o  $d > -\frac{1}{4}$  se  $\xi = \pm 1$ ;
- Memoria lunga;
- $\rho_k \cong Kk^{2d-1} \sin(\pi d - kG)$  per  $|\xi| < 1$  e  $d < \frac{1}{2}$ , con  $K$  costante finita.

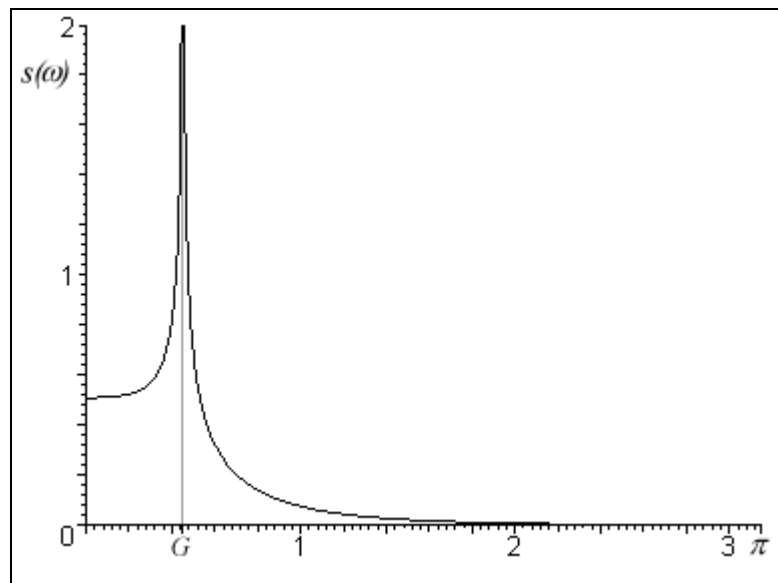
Lo spettro di un processo GARMA è dato ovviamente da:

$$s(\omega) = \frac{\sigma^2}{2\pi} \frac{\left| \psi(e^{-i\omega}) \right|^2}{\left| \varphi(e^{-i\omega}) \right|^2} \left[ 4(\cos \omega - \xi)^2 \right]^{-d}, \quad (3.6.14)$$

per cui, nel caso di un GARMA  $(1, \xi, d, 1)$  si avrebbe:

$$s(\omega) = \frac{\sigma^2}{2\pi} \frac{1 + 2\psi \cos \omega + \psi^2}{1 - 2\varphi \cos \omega + \varphi^2} \left[ 4(\cos \omega - \xi)^2 \right]^{-d}, \quad (3.6.15)$$

disegnato nella figura 3.3.



**Figura 3.3**

*Lo spettro di un processo GARMA  $(1, 0.9, 0.25, 1)$  con  $\varphi = \psi = 1/2$*

Le ricerche su dati reali che utilizzano i modelli GARMA sono assai poche. Interessante ai fini della nostra analisi è il lavoro di Chung (1996b), che esamina l'inflazione USA utilizzando dati trimestrali tra il 1875 e il 1983 e annuali tra il 1749 e il 1983.

I dati trimestrali sono modellati con un GARMA (0, 1), per il quale si ottiene  $\hat{d} = 0.114$  e  $\hat{\xi} = 0.985$ , a cui corrisponde  $G = 0.0552\pi$ ; quelli annuali con un GARMA (2, 2), con  $\hat{d} = 0.211$ ,  $\hat{\xi} = 0.776$  e  $G = 0.8177\pi$ ; è interessante osservare che a entrambe le frequenze di Gegenbauer corrisponde una periodicità di circa nove anni.

Nel caso dei dati trimestrali si osserva che la frequenza  $G$  è molto bassa; lo spettro che ne risulta ha pertanto un picco molto vicino a zero. Osserviamo che una forma di questo tipo richiama molto da vicino i risultati empirici di Granger (1966) che abbiamo presentato nel paragrafo 1.2.2. Un processo GARMA quindi modella la figura 1.3 in maniera molto più precisa rispetto a un ARFIMA, poiché riesce a descrivere anche il decadimento della densità spettrale alle frequenze più vicine allo zero. In generale possiamo dire che, nella modellazione di serie ad alta frequenza, ci aspettiamo di avere una frequenza di Gegenbauer bassa, a cui corrisponde un valore di  $\xi$  prossimo a uno; ci si discosta poco, quindi, dai processi ARFIMA, ma si ha comunque un *plus* che riguarda sia le frequenze inferiori a  $G$ , sia l'andamento sinusoidale delle autocorrelazioni.



# **Capitolo 4**

## **Problemi di stima econometrica**

Dopo aver richiamato brevemente i principi generali della stima di massima verosimiglianza, si introdurrà il concetto di stima di quasi-massima verosimiglianza e si esamineranno sommariamente i test basati sulla verosimiglianza. A questo punto saremo pronti per scendere nel dettaglio ed esaminare le problematiche specifiche della stima di processi ARFIMA e GARMA, con particolare attenzione per i metodi di massima verosimiglianza. Il capitolo si conclude con un esempio di stima su dati reali del modello ARFIMA.

## 4.1 La stima di massima verosimiglianza

La funzione di verosimiglianza (e conseguentemente la stima di massima verosimiglianza) è uno degli strumenti più importanti e più dibattuti di tutta la statistica inferenziale. Si scrivono interi testi di inferenza statistica basata sulla verosimiglianza<sup>1</sup>; in questo contesto riteniamo importante, più che scendere nel dettaglio, presentarne la filosofia ispiratrice.

La funzione di distribuzione congiunta di un campione casuale, dato un certo vettore di parametri  $\theta$ , può essere utilizzata in due modi diversi. Prima dell'estrazione del campione, l'espressione  $f(x_1, \dots, x_n | \theta)$ , assumendo  $\theta$  noto, indica la probabilità<sup>2</sup> che il campione  $x_1, \dots, x_n$  venga estratto.

Se, al contrario, il campione è già stato osservato mentre il vettore di parametri  $\theta$  è ignoto, la funzione  $f(\theta | x_1, \dots, x_n) = \ell(\theta)$  dipende solo dai parametri ed è detta funzione di verosimiglianza (*likelihood function*). Fissato  $\theta$ , essa è proporzionale alla probabilità di ottenere il campione casuale le cui realizzazioni numeriche si sono effettivamente osservate. Tuttavia, anche se per ogni valore di  $\theta$  la funzione di verosimiglianza è una probabilità, essa non è una distribuzione di probabilità<sup>3</sup>.

Se osserviamo che  $\ell(\theta_1) > \ell(\theta_2)$ , significa che il vettore di parametri  $\theta_1$  è, sulla base dei dati osservati, più “plausibile” di  $\theta_2$ . Questo principio sta alla base della stima di massima verosimiglianza: si assume come stima di  $\theta$  quel valore  $\hat{\theta}$  per cui la verosimiglianza è massima – ovvero il valore più plausibile date le osservazioni disponibili.

Supponiamo di avere una popolazione il cui processo generatore dei dati dipenda da un certo vettore di parametri  $\theta$ . Vogliamo, a partire da un certo insieme di osservazioni, stimare questi

---

<sup>1</sup> Si veda, ad esempio, l'ottimo testo di Azzalini (1992).

<sup>2</sup> In realtà ciò è vero solo per variabili aleatorie discrete; nel caso continuo, essa è proporzionale alla probabilità di estrarre un campione in un intervallo infinitesimo  $n$ -dimensionale attorno al vettore campionario.

<sup>3</sup> Una discussione formale di questo risultato, apparentemente controintuitivo, è in Casella e Berger (1990). Si noti che i tentativi di normalizzare la funzione di verosimiglianza per ricondurla a una distribuzione di probabilità – la cosiddetta “inferenza fiduciale” di Fisher – non hanno prodotto risultati apprezzabili.

parametri. I dati disponibili possono essere pensati come un campione estratto da una popolazione formata da tutti gli insiemi di dati che il DGP è suscettibile di generare. La stima di massima verosimiglianza di  $\theta$  è quel vettore  $\hat{\theta}$  che massimizza la probabilità congiunta di aver estratto il campione di cui disponiamo.

#### 4.1.1 La funzione di verosimiglianza per serie storiche

Supponiamo di avere un modello ARMA ( $p, q$ ):

$$y_t = \mu + \varphi_1 y_{t-1} + \cdots + \varphi_p y_{t-p} + \psi_1 \varepsilon_{t-1} + \cdots + \psi_q \varepsilon_{t-q} + \varepsilon_t.$$

Il vettore di parametri da stimare sarà  $\theta = (\mu, \varphi_1, \dots, \varphi_p, \psi_1, \dots, \psi_q)$ .

La funzione di densità di probabilità congiunta della serie di  $T$  osservazioni  $\{y_t\}$  sarà  $f(y_T, y_{T-1}, \dots, y_1 | \theta)$ , dove le osservazioni sono variabili casuali dipendenti dal vettore  $\theta$ . La funzione di verosimiglianza sarà quindi:

$$\ell(\theta) = f(\theta | y_T, y_{T-1}, \dots, y_1). \quad (4.1.1)$$

Per capire come esplicitare la forma funzionale della (4.1.1) nel contesto delle serie storiche, consideriamo una successione di variabili aleatorie  $\{\varepsilon_t\}$  indipendenti e identicamente distribuite. In questo caso si ha semplicemente:

$$f(\varepsilon_T, \varepsilon_{T-1}, \dots, \varepsilon_1 | \theta) = \prod_{t=1}^T f(\varepsilon_t | \theta).$$

In generale però non si avranno osservazioni indipendenti e identicamente distribuite. Supponendo però di avere due sole osservazioni, sappiamo che vale:

$$f(y_2, y_1) = f(y_1) f(y_2 | y_1).$$

Se le osservazioni sono tre, si ha quindi:

$$f(y_3, y_2, y_1) = f(y_1) f(y_3, y_2 | y_1);$$

poiché d'altra parte

$$f(y_3, y_2 | y_1) = f(y_2 | y_1) f(y_3 | y_2, y_1),$$

si conclude che

$$f(y_3, y_2, y_1) = f(y_1) f(y_2 | y_1) f(y_3 | y_2, y_1).$$

Applicando ricursivamente questa metodologia (che prende il nome di *Prediction Error Decomposition*), si può pervenire alla formula generale:

$$f(y_T, y_{T-1}, \dots, y_1) = f(y_1) \prod_{t=2}^T f(y_t | y_{t-1}, \dots, y_1). \quad (4.1.2)$$

Si noti che, utilizzando questo tipo di approccio, abbiamo bisogno di un valore iniziale a cui condizionare gli altri; per questa ragione questa metodologia è nota come verosimiglianza condizionata, che in particolari contesti porta a una *conditional sum of squares* (CSS).

Tornando alla funzione di verosimiglianza della (4.1.1), possiamo scriverla, utilizzando la (4.1.2), come:

$$\begin{aligned} \ell(\boldsymbol{\theta}) &= f(\boldsymbol{\theta} | y_T, y_{T-1}, \dots, y_1) = \\ &= f(\boldsymbol{\theta} | y_1) \prod_{t=2}^T f(\boldsymbol{\theta} | y_t, y_{t-1}, \dots, y_1). \end{aligned} \quad (4.1.3)$$

Resta ora da vedere come trattare le  $f$ . Se conosciamo (o possiamo ipotizzare) la distribuzione dei termini di disturbo, e consideriamo il caso più semplice di normalità, si ha in primo luogo:

$$f(\varepsilon_t) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\varepsilon_t^2}{2\sigma^2}\right).$$

Per procedere oltre abbiamo bisogno di introdurre<sup>4</sup> il concetto di trasformazione di variabili casuali. Supponiamo di avere a disposizione  $T$  osservazioni della serie  $\{y_t\}$  e di sapere che  $y_t = g_t(x_1, \dots, x_T)$ ; supponiamo altresì di conoscere la distribuzione congiunta  $f_x(x_1, \dots, x_T)$  e di voler ricavare quella di  $y_t$ . Si può dimostrare che la densità di probabilità congiunta di  $\{y_t\}$  è data da:

$$\begin{aligned}
f_y(y_1, \dots, y_T) &= \\
&= \sum_{t=1}^T |\det J_t| f_x(g_{1t}^{-1}(y_1, \dots, y_T), \dots, g_{Tt}^{-1}(y_1, \dots, y_T)).
\end{aligned} \tag{4.1.4}$$

La matrice  $J_t$  che prende il nome di matrice Jacobiana, è così definita:

$$J_t = \begin{bmatrix} \frac{\partial g_{1t}^{-1}}{\partial y_1} & \frac{\partial g_{1t}^{-1}}{\partial y_2} & \dots & \frac{\partial g_{1t}^{-1}}{\partial y_T} \\ \frac{\partial g_{2t}^{-1}}{\partial y_1} & \frac{\partial g_{2t}^{-1}}{\partial y_2} & \dots & \frac{\partial g_{2t}^{-1}}{\partial y_T} \\ \dots & \dots & \dots & \dots \\ \frac{\partial g_{Tt}^{-1}}{\partial y_1} & \frac{\partial g_{Tt}^{-1}}{\partial y_2} & \dots & \frac{\partial g_{Tt}^{-1}}{\partial y_T} \end{bmatrix},$$

e il suo determinante, che compare nella (4.1.4), prende il nome di Jacobiano della trasformazione da  $\mathbf{y}$  a  $\mathbf{x}$ .

A questo punto è abbastanza agevole procedere: se utilizziamo un generico modello lineare  $y_t = \mathbf{x}_t' \boldsymbol{\beta} + \varepsilon_t$ , con disturbi indipendenti e identicamente distribuiti, la matrice Jacobiana è l'identità, per cui lo Jacobiano della trasformazione da  $y_t$  a  $\varepsilon_t$  vale uno, e si ha semplicemente:

$$f(y_t | \mathbf{x}_t, \boldsymbol{\beta}) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(y_t - \mathbf{x}_t' \boldsymbol{\beta})^2}{2\sigma^2}\right), \tag{4.1.5}$$

da cui

$$\ell(\boldsymbol{\beta}) = \prod_{t=1}^T \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(y_t - \mathbf{x}_t' \boldsymbol{\beta})^2}{2\sigma^2}\right),$$

e di conseguenza

$$\ln \ell(\boldsymbol{\beta}) = -\frac{T}{2} \ln 2\pi - \frac{T}{2} \ln \sigma^2 - \sum_{t=1}^T \frac{(y_t - \mathbf{x}_t' \boldsymbol{\beta})^2}{2\sigma^2}. \tag{4.1.6}$$

Si può dimostrare inoltre (si rimanda a Davidson e McKinnon, 1993 – pagina 489) che, se nel vettore  $\mathbf{x}_t$  compaiono dei ritardi di  $y_t$ , la matrice Jacobiana è triangolare bassa con “uni” sulla

---

<sup>4</sup> Piuttosto che tirare dritto senza porsi la questione, si è ritenuto opportuno cercare di dare al lettore almeno un'idea di massima sui problemi che pone la trasformazione; la materia è comunque più complessa, per cui si rimanda, per una trattazione più completa, a Casella e Berger (1990).

diagonale, per cui il suo determinante è anche in questo caso uno. Questo risultato ci consente di trascurare, almeno per quanto riguarda i modelli ARMA, il Jacobiano della trasformazione. È immediatamente evidente, dalla (4.1.6), perché sia comodo utilizzare il logaritmo naturale della funzione di verosimiglianza: si evitano infatti sia il calcolo della produttoria che quello di  $T$  esponenziali, mentre rimangono solo da calcolare due logaritmi naturali e una sommatoria.

In generale, quindi, per avere una stima di massima verosimiglianza di  $\theta$ , si dovranno compiere due passi:

- Trovare  $\ell(\theta)$  per il modello che vogliamo utilizzare;
- Massimizzare<sup>5</sup>  $\ell(\theta)$  o, equivalentemente, il suo logaritmo naturale, che prende il nome di log-verosimiglianza (*log-likelihood*).

Un esempio servirà a chiarificare la situazione<sup>6</sup>. Supponiamo di voler trovare la funzione di verosimiglianza di un processo MA (1):

$$y_t = \mu + \psi \varepsilon_{t-1} + \varepsilon_t, \text{ con } \varepsilon_t \sim N(0, \sigma^2).$$

Conoscendo  $\varepsilon_{t-1}$ , sappiamo che  $(y_t | \varepsilon_{t-1}) \sim N(\mu + \psi \varepsilon_{t-1}, \sigma^2)$ , per cui la (4.1.5) diventa:

$$f(y_t | \varepsilon_{t-1}) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y_t - \mu - \psi \varepsilon_{t-1})^2}{2\sigma^2}\right).$$

Partendo da  $\varepsilon_0$ , possiamo calcolare la verosimiglianza di ogni singola osservazione per via ricursiva, partendo dalla prima. Dato  $\varepsilon_0 = 0$ , possiamo scrivere:

$$\ln \ell(\mu, \psi, \sigma^2) = -\frac{T}{2} \ln 2\pi - \frac{T}{2} \ln \sigma^2 - \sum_{t=1}^T \frac{(y_t - \mu - \psi \varepsilon_{t-1})^2}{2\sigma^2}.$$

---

<sup>5</sup> A parte alcuni casi particolarmente semplici nei quali si può ricavare una soluzione analitica, la massimizzazione della funzione di verosimiglianza deve essere affrontata utilizzando metodi di ottimizzazione numerici; tali metodi saranno esaminati nella sezione A.2 in appendice.

<sup>6</sup> L'esempio è tratto da Hamilton (1994 – pagina 127).

Ci sono delle situazioni in cui condizionare ai valori pre-campionari non è opportuno, o può essere fonte di problemi. In questo caso ricavare la funzione di verosimiglianza, che prende il nome di funzione di verosimiglianza esatta o non condizionata, è molto più complesso, ed è opportuno utilizzare una notazione matriciale.

Continuiamo con l'utilizzo esemplificativo di un processo MA (1), e poniamo

$\mathbf{y} = (y_1, y_2, \dots, y_T)'$ ,  $\mathbf{m} = (\mu, \mu, \dots, \mu)'$  e:

$$\mathbf{V} = E[(\mathbf{y} - \mathbf{m})(\mathbf{y} - \mathbf{m})'] = \sigma^2 \begin{bmatrix} (1 + \psi^2) & \psi & 0 & \dots & 0 \\ \psi & (1 + \psi^2) & \psi & \dots & 0 \\ 0 & \psi & (1 + \psi^2) & \dots & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 0 & 0 & 0 & \dots & (1 + \psi^2) \end{bmatrix}. \quad (4.1.7)$$

La funzione di verosimiglianza è allora:

$$\ell(\mathbf{y}) = (2\pi)^{-T/2} |\mathbf{V}|^{-1/2} \exp\left[-\frac{1}{2}(\mathbf{y} - \mathbf{m})' \mathbf{V}^{-1}(\mathbf{y} - \mathbf{m})\right]. \quad (4.1.8)$$

$\mathbf{V}$  è definita positiva e simmetrica: possiamo calcolarne la fattorizzazione triangolare

$\mathbf{V} = \mathbf{A}\mathbf{D}\mathbf{A}'$ :

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 & 0 \\ \frac{\psi}{1 + \psi^2} & 1 & 0 & \dots & 0 & 0 \\ 0 & \frac{\psi(1 + \psi^2)}{1 + \psi^2 + \psi^4} & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & \frac{\psi[1 + \psi^2 + \psi^4 + \dots + \psi^{2(n-2)}]}{1 + \psi^2 + \psi^4 + \dots + \psi^{2(n-1)}} & 1 \end{bmatrix}, \quad (4.1.9)$$

$$\mathbf{D} = \begin{bmatrix} 1 + \psi^2 & 0 & 0 & \dots & 0 \\ 0 & \frac{1 + \psi^2 + \psi^4}{1 + \psi^2} & 0 & \dots & 0 \\ 0 & 0 & \frac{1 + \psi^2 + \psi^4 + \psi^6}{1 + \psi^2 + \psi^4} & \dots & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 0 & 0 & 0 & \dots & \frac{1 + \psi^2 + \psi^4 + \dots + \psi^{2n}}{1 + \psi^2 + \psi^4 + \dots + \psi^{2(n-1)}} \end{bmatrix}. \quad (4.1.10)$$

La (4.1.8) diventa:

$$\begin{aligned}\ell(\boldsymbol{\theta}) &= (2\pi)^{-T/2} |\mathbf{A}\mathbf{D}\mathbf{A}'|^{-1/2} \times \\ &\times \exp\left[-\frac{1}{2}(\mathbf{y} - \mathbf{m})' [\mathbf{A}']^{-1} \mathbf{D}^{-1} \mathbf{A}^{-1} (\mathbf{y} - \mathbf{m})\right].\end{aligned}\quad (4.1.11)$$

La matrice  $\mathbf{A}$  è triangolare bassa e ha solo “uni” sulla diagonale, per cui  $|\mathbf{A}| = 1$ , e di conseguenza:

$$|\mathbf{A}\mathbf{D}\mathbf{A}'| = |\mathbf{D}|. \quad (4.1.12)$$

Sia ora:

$$\tilde{\mathbf{y}} = \mathbf{A}^{-1}(\mathbf{y} - \mathbf{m}); \quad (4.1.13)$$

con la (4.1.12) e la (4.1.13), la (4.1.11) diventa:

$$\ell(\boldsymbol{\theta}) = (2\pi)^{-T/2} |\mathbf{D}|^{-1/2} \exp\left[-\frac{1}{2} \tilde{\mathbf{y}}' \mathbf{D}^{-1} \tilde{\mathbf{y}}\right]. \quad (4.1.14)$$

Notiamo ora che  $\mathbf{D}$  è diagonale, per cui il determinante è dato dal prodotto degli elementi sulla diagonale:

$$|\mathbf{D}| = \prod_{t=1}^T d_{tt}; \quad (4.1.15)$$

mentre l'inversa di  $\mathbf{D}$  è data dai reciproci degli elementi sulla diagonale:

$$\tilde{\mathbf{y}}' \mathbf{D}^{-1} \tilde{\mathbf{y}} = \sum_{t=1}^T \frac{\tilde{y}_t^2}{d_{tt}}. \quad (4.1.16)$$

Sostituiamo la (4.1.15) e la (4.1.16) nella (4.1.11) e otteniamo:

$$\ell(\boldsymbol{\theta}) = (2\pi)^{-T/2} \left[ \prod_{t=1}^T d_{tt} \right]^{-\frac{1}{2}} \exp\left[-\frac{1}{2} \sum_{t=1}^T \frac{\tilde{y}_t^2}{d_{tt}}\right], \quad (4.1.17)$$

da cui:

$$\ln \ell(\boldsymbol{\theta}) = -\frac{T}{2} \ln(2\pi) - \frac{1}{2} \sum_{t=1}^T \ln(d_{tt}) - \frac{1}{2} \sum_{t=1}^T \frac{\tilde{y}_t^2}{d_{tt}}. \quad (4.1.18)$$

È interessante osservare (si veda per la dimostrazione Hamilton, 1994 – pagina 130), che la (4.1.18) vale anche nel caso di un generico MA ( $q$ ).



### 4.1.2 Le proprietà asintotiche

La fortuna degli stimatori di massima verosimiglianza è da ascrivere in larga parte alla quantità di proprietà asintotiche di cui godono, che compensano il rischio di stime distorte in campioni di limitata numerosità.

Seguendo Greene<sup>7</sup> (2000), se valgono le seguenti condizioni, dette di regolarità:

- Le prime tre derivate di  $\ln \ell(\boldsymbol{\theta})$  rispetto a  $\boldsymbol{\theta}$  sono finite per ogni  $\boldsymbol{\theta}$ ;
- Si possono calcolare i valori attesi delle prime due derivate di  $\ln \ell(\boldsymbol{\theta})$  rispetto a  $\boldsymbol{\theta}$ ;
- Il determinante della matrice cubica delle derivate terze di  $\ln \ell(\boldsymbol{\theta})$  rispetto a  $\boldsymbol{\theta}$  è finito;

si dimostra che lo stimatore di massima verosimiglianza:

- È consistente  $\left( \text{p} \lim_{T \rightarrow +\infty} \hat{\boldsymbol{\theta}} = \boldsymbol{\theta} \right)$ ;
- È asintoticamente normale  $\left( \sqrt{T}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \xrightarrow{d} N(0, \mathbf{I}^{-1}(\boldsymbol{\theta})) \right)$ ;
- È asintoticamente efficiente;
- È invariante, cioè dato  $\gamma = f(\boldsymbol{\theta})$ , si ha  $\hat{\gamma} = f(\hat{\boldsymbol{\theta}})$ .

La matrice  $\mathbf{I}(\boldsymbol{\theta})$  che compare nella distribuzione asintotica prende il nome di matrice d'informazione (*information matrix*) ed è definita come segue<sup>8</sup>:

$$\mathbf{I}(\boldsymbol{\theta}) = E \left[ - \frac{\partial^2 \ln \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right], \quad (4.1.19)$$

In generale sarà difficile poter calcolare esplicitamente il valore atteso nella (4.1.19). Per stimare  $\mathbf{I}(\boldsymbol{\theta})$  possiamo allora procedere in due modi. Il primo è quello di inserire  $\hat{\boldsymbol{\theta}}$  nella

(4.1.19) e trascurare il valore atteso; si avrà allora semplicemente  $\hat{\mathbf{I}}_{2D} = - \frac{\partial^2 \ln \ell(\hat{\boldsymbol{\theta}})}{\partial \hat{\boldsymbol{\theta}} \partial \hat{\boldsymbol{\theta}'}}$ .

<sup>7</sup> La presentazione di Greene (2000) è volutamente semplificata; per una definizione più rigorosa delle condizioni di regolarità si rimanda a Davidson e McKinnon (1993).

<sup>8</sup> La matrice di derivate seconde di una funzione rispetto ai suoi argomenti prende il nome di matrice Hessiana.

Se calcolare le derivate seconde analiticamente è complesso e numericamente richiede troppo tempo, un'altra possibilità, proposta da Berndt, Hall, Hall e Hausman (1974) è la seguente:

calcoliamo il gradiente della verosimiglianza di ogni singola osservazione  $\hat{\mathbf{g}}_t = \frac{\partial \ln \ell(\hat{\boldsymbol{\theta}}, y_t)}{\partial \hat{\boldsymbol{\theta}}}$  e

raccogliamo i risultati in una matrice  $\hat{\mathbf{G}} = [\hat{\mathbf{g}}_1, \hat{\mathbf{g}}_2, \dots, \hat{\mathbf{g}}_T]'$ .

Osservando<sup>9</sup> che l'opposto del valore atteso dell'Hessiana è la matrice di varianze e covarianze del gradiente, cioè che  $\text{var}(\mathbf{g}_t) = \mathbf{I}_t(\boldsymbol{\theta})$ , e che  $E(\mathbf{g}_t) = 0$ , otteniamo la formula di stima della matrice d'informazione:

$$\hat{\mathbf{I}}_{OP} = \sum_{t=1}^T \hat{\mathbf{g}}_t \hat{\mathbf{g}}_t' = \hat{\mathbf{G}}' \hat{\mathbf{G}}. \quad (4.1.20)$$

### 4.1.3 La quasi-massima verosimiglianza

Recuperando il discorso del paragrafo precedente sulla matrice d'informazione, possiamo dire che in generale i valori che si ottengono per  $\hat{\mathbf{I}}_{2D}$  e  $\hat{\mathbf{I}}_{OP}$  non dovrebbero essere troppo dissimili. In caso lo siano, ciò può indicare che il modello non è correttamente specificato, cioè che stiamo lavorando con una funzione di verosimiglianza "sbagliata".

L'errata specificazione di un modello può senz'altro indicare degli errori nella fase di definizione dello stesso, ma è di frequente provocata dalla mancata verifica delle assunzioni sulla forma della distribuzione dei termini di disturbo. Nel caso dei mercati azionari, o dei cambi, ad esempio, si hanno generalmente distribuzioni di tipo leptocurtico<sup>10</sup>: cercare di stimare dati di questo tipo con un modello a disturbi normalmente distribuiti potrebbe essere fonte di problemi.

<sup>9</sup> Si veda per la dimostrazione Greene (2000 – pagina 128).

<sup>10</sup> La curtosi ( $K$ ) è il momento quarto di una distribuzione; per una distribuzione normale la curtosi vale tre, per cui una distribuzione leptocurtica ha  $K > 3$ . Ciò significa che le code della distribuzione sono più spesse di quelle della normale. Quella della distribuzione dei rendimenti dei mercati finanziari è una questione complessa e tuttora dibattuta; per una trattazione introduttiva si può vedere Nelson (1991).

Un modo per procedere potrebbe essere quello di ridefinire la (4.1.6) utilizzando una diversa distribuzione per i disturbi (ad esempio sono frequenti, sempre nel caso dei mercati finanziari, le stime che utilizzano disturbi t-distribuiti, i quali sembrano modellare meglio i residui osservati); tuttavia si tratta di una soluzione poco utile nel caso in cui la distribuzione dei residui osservati non “somigli” alla forma della funzione di densità di nessuna distribuzione conosciuta. Sarebbe pertanto utile disporre di strumenti che consentano di procedere alla stima senza alcuna ipotesi sulla forma della distribuzione dei termini di disturbo.

Fortunatamente la stima di massima verosimiglianza può superare questo problema in maniera abbastanza agevole: si parla, in questo caso, di stima di quasi-massima verosimiglianza, concetto introdotto da White (1982).

Supponiamo di avere una serie di dati  $\{y_t\}$  generati a partire da una certa distribuzione  $G$ , della quale non siamo in grado di descrivere a priori la forma. Sia invece  $F(\boldsymbol{\theta})$  una distribuzione nota e “utilizzabile”. Definiamo la funzione di quasi log-verosimiglianza come:

$$\ln \tilde{\ell}(\boldsymbol{\theta}) = \frac{\sum_{t=1}^T \ln f(\boldsymbol{\theta} | y_t, y_{t-1}, \dots, y_1)}{T}, \quad (4.1.21)$$

che differisce dalla (4.1.6) esclusivamente nel fatto che viene calcolata sulla base di una funzione di distribuzione “sbagliata”: la stima di quasi-massima verosimiglianza  $\tilde{\boldsymbol{\theta}}$  sarà ovviamente la soluzione di  $\max_{\boldsymbol{\theta}} \ln \tilde{\ell}(\boldsymbol{\theta})$ .

Supponiamo<sup>11</sup> che la distribuzione “vera” di un certo carattere sia  $f$ . Per trovare la stima dei parametri dovremo allora trovare un vettore  $\hat{\boldsymbol{\theta}}$  tale che  $f_{\hat{\boldsymbol{\theta}}} = f$ . Ovviamente ciò è possibile solo se  $f$  appartiene alla famiglia  $F = \{f_{\boldsymbol{\theta}}, \boldsymbol{\theta}\}$ ; in caso contrario ci troviamo in una situazione in cui il modello è specificato male. In questo caso dovremo trovare un vettore  $\boldsymbol{\theta}^*$  tale che

---

<sup>11</sup> Si è seguito Gouriéroux e Monfort (1997).

$f_{\theta^*}$  sia il più vicino possibile a  $f$ ; la distanza tra  $f_{\theta^*}$  e  $f$  prende il nome di discrepanza di Kullback e può essere utilizzata, tra le altre cose, come criterio informativo:

$$KLIC = E \left[ \ln \frac{f_{\theta^*}(y)}{f(y)} \right].$$

White (1982) dimostra (si veda l'articolo citato, pagina 4) che:

$$\text{p} \lim_{T \rightarrow +\infty} \tilde{\boldsymbol{\theta}} = \boldsymbol{\theta}^*. \quad (4.1.22)$$

Questa proprietà ci consente di utilizzare con una certa sicurezza la stima di massima verosimiglianza anche quando non siamo sicuri di aver specificato bene il modello. Infatti, in caso di errata specificazione, si avrà una stima di quasi-massima verosimiglianza che convergerà, per  $T \rightarrow +\infty$ , al vettore  $\boldsymbol{\theta}^*$ . Tale vettore, pur non corrispondendo necessariamente alla soluzione “vera”<sup>12</sup>, minimizzerà comunque, per come è definito, l'ignoranza sulla struttura del modello.

La seconda importante proprietà asintotica degli stimatori di quasi-massima verosimiglianza è la normalità. Siano:

$$A(\boldsymbol{\theta}) = E \left[ - \frac{\partial^2 \ln \tilde{\ell}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right], \quad (4.1.23)$$

$$B(\boldsymbol{\theta}) = E[\mathbf{g}'_t(\boldsymbol{\theta}) \mathbf{g}_t(\boldsymbol{\theta})], \quad (4.1.24)$$

$$C(\boldsymbol{\theta}) = A(\boldsymbol{\theta})^{-1} B(\boldsymbol{\theta}) A(\boldsymbol{\theta})^{-1}; \quad (4.1.25)$$

White (1982) dimostra (si veda l'articolo citato a pagina 6) che:

$$\sqrt{T}(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}^*) \xrightarrow{d} N(0, C(\boldsymbol{\theta}^*)). \quad (4.1.26)$$

Se vogliamo stimare la matrice di varianze e covarianze di uno stimatore di quasi-massima verosimiglianza, possiamo procedere osservando la stretta somiglianza tra  $A(\boldsymbol{\theta})$  e  $B(\boldsymbol{\theta})$  e,

---

<sup>12</sup> Si noti che, in alcuni importanti casi, si ha che  $\boldsymbol{\theta}$  e  $\boldsymbol{\theta}^*$  coincidono anche sotto una funzione di verosimiglianza non corretta. Per ulteriori approfondimenti, si rimanda a Gouriéroux, Monfort e Trognon (1984).

rispettivamente, la (4.1.19) e la (4.1.20). Utilizzando quest'ultime due espressioni per la stima di  $A(\boldsymbol{\theta})$  e  $B(\boldsymbol{\theta})$  otteniamo:

$$C(\tilde{\boldsymbol{\theta}}) = [\hat{\mathbf{I}}_{2D} \hat{\mathbf{I}}_{OP}^{-1} \hat{\mathbf{I}}_{2D}]^{-1} = \hat{\mathbf{I}}_{2D}^{-1} \hat{\mathbf{I}}_{OP} \hat{\mathbf{I}}_{2D}^{-1}. \quad (4.1.27)$$

Relativamente all'efficienza degli stimatori QML, è facile rendersi conto che essa non sussiste. Pur senza darne una prova formale, si osserva che, utilizzando uno stimatore QML, si trascurava parte dell'informazione disponibile, cioè la forma dei termini di disturbo; in questo senso la matrice d'informazione ha valori più elevati rispetto a quelli di una semplice stima di massima verosimiglianza, e lo stimatore QML è meno efficiente.

## 4.2 I test basati sulla verosimiglianza

Abbiamo visto che la funzione di verosimiglianza misura “quanto bene” un modello si adatta ai dati empirici; questo suggerisce che la si può utilizzare a scopi inferenziali sui coefficienti. I test classici basati sulla verosimiglianza sono tre: il test LR, il test LM e il test di Wald. Nel nostro lavoro utilizzeremo esclusivamente i primi due, per cui, per la definizione del terzo<sup>13</sup>, si rimanda a Greene (2000).

### 4.2.1 Il test LR

Quando dobbiamo fare valutazioni sulla specificazione di un modello si fa quasi sempre ricorso all'imposizione di vincoli sui coefficienti. La logica del test è un confronto tra valori della funzione di verosimiglianza massimizzata liberamente e quella massimizzata imponendo i vincoli. Qualora i vincoli fossero propri del processo generatore dei dati, la loro imposizione in fase di stima non dovrebbe far perdere molta capacità esplicativa, e i valori delle funzioni di verosimiglianza con o senza l'imposizione dei vincoli dovrebbero essere simili.

Nel caso delle stime di massima verosimiglianza si può quindi procedere stimando prima il modello senza restrizioni (ottenendo la stima  $\hat{\theta}$ ) e poi quello vincolato (ottenendo  $\bar{\theta}$ ) e calcolando il rapporto fra verosimiglianze (*likelihood ratio*):

$$LR = \frac{\ell(\bar{\theta})}{\ell(\hat{\theta})}. \quad (4.2.1)$$

Poiché la verosimiglianza del modello vincolato è sicuramente inferiore a quella del modello non vincolato,  $LR$  sarà sicuramente inferiore a uno. Un valore molto basso significa che imponendo i vincoli si perde molta capacità esplicativa, per cui è poco conveniente accettarli.

---

<sup>13</sup> Il test di Wald si rivela utile in particolare quando si ha a disposizione solo la verosimiglianza del modello non vincolato.

Relativamente alla distribuzione di questa statistica, si dimostra (si veda Greene, 2000 – pagina 390) che:

$$-2 \ln LR = 2 \left[ \ln \ell(\hat{\boldsymbol{\theta}}) - \ln \ell(\bar{\boldsymbol{\theta}}) \right] \xrightarrow{d} \chi_m^2, \quad (4.2.2)$$

dove  $m$  indica il numero di vincoli che si impongono.

#### 4.2.2 Il test LM

Uno svantaggio dei test LR è che ci obbligano a calcolare sia la verosimiglianza con le restrizioni che quella ordinaria. Il test *Lagrange multiplier*, o LM, utilizza solo la verosimiglianza ristretta. Tralasciando le dimostrazioni, per le quali si può vedere Greene (2000 – pagina 156), si ottiene che:

$$LM = \left[ \frac{\partial \ln \ell(\bar{\boldsymbol{\theta}})}{\partial \bar{\boldsymbol{\theta}}} \right]' \cdot [\mathbf{I}(\bar{\boldsymbol{\theta}})]^{-1} \cdot \left[ \frac{\partial \ln \ell(\bar{\boldsymbol{\theta}})}{\partial \bar{\boldsymbol{\theta}}} \right] \xrightarrow{d} \chi_m^2. \quad (4.2.3)$$

### 4.3 La stima dei modelli a integrazione frazionaria

Quando ci troviamo a dover stimare un modello ARFIMA possiamo seguire due strade: o ci concentriamo sulla stima di  $d$ , e poi sugli altri parametri, oppure cerchiamo di stimare tutto il modello in un unico passaggio. Ognuna di queste tecniche ha pregi e difetti, che analizzeremo caso per caso; è utile però, prima di cominciare, ribadire un concetto che abbiamo già richiamato più volte nel corso di questo lavoro, cioè il *trade-off* tra precisione e semplicità. Relativamente ai metodi di stima questo concetto diventa di enorme importanza, perché è proprio in questo frangente che la semplicità diventa preziosa nel consentirci di ottenere rapidamente le stime.

Vale la pena tenere bene a mente questo concetto, perché non ci sarà possibile identificare “il metodo di stima migliore”. Potremo identificare quelli che si rivelano più precisi, e quelli che sono computazionalmente meno impegnativi, ma ogni situazione in cui ci si trova a dover affrontare la stima di un processo ARFIMA avrà diversi obiettivi e diversi presupposti; per questo sarà compito del ricercatore o dell’analista identificare la procedura che meglio si adatta al caso specifico che si deve affrontare.

#### 4.3.1 Massima verosimiglianza approssimata

Un primo procedimento che si può scegliere di seguire è il seguente. Richiamiamo la generica rappresentazione di un processo ARFIMA  $(p, d, q)$  data nella (3.2.16):

$$\varphi(L)(1-L)^d y_t = \psi(L)\varepsilon_t, \quad (4.3.1)$$

ed espandiamo in serie di McLaurin il termine  $(1-L)^d$ ; per avere una formulazione chiusa dovremo però arrestarci a un certo valore  $M$ , facendo attenzione a che sia sufficientemente grande per non “tagliare fuori” delle importanti componenti di lungo periodo. Richiamando la (3.2.1), la (4.3.1) sarà così approssimabile dalla:



$$\varphi(L) \sum_{k=0}^M \binom{d}{k} (-1)^k L^k y_t = \psi(L) \varepsilon_t. \quad (4.3.2)$$

Hosking (1984a) propone di generare la serie:

$$w_t = (1 - L)^d y_t,$$

attraverso la quale la (4.3.1) diventa:

$$\varphi(L) w_t = \psi(L) \varepsilon_t, \quad (4.3.3)$$

cioè un normale ARMA con  $p + q$  parametri da stimare. Si noti inoltre che, seguendo questa strada, le approssimazioni che facciamo sono due: la prima deriva dall'aver troncato a  $M$  l'espansione di  $(1 - L)^d$ , la seconda è di aver ignorato il Jacobiano della trasformazione da ARFIMA a ARMA (che non vale esattamente uno). Si può dimostrare (si veda Hosking, 1984a – pagina 1902), in ogni caso, che questo termine tende asintoticamente a uno.

Li e McLeod (1986) studiano le proprietà di questo approccio, ricavandone la matrice d'informazione:

$$\mathbf{I} = \begin{bmatrix} \mathbf{I}_{p,q} & \mathbf{J} \\ \mathbf{J}' & \frac{\pi^2}{6} \end{bmatrix}, \quad (4.3.4)$$

dove  $\mathbf{I}_{p,q}$  è la matrice d'informazione del processo ARMA  $(p, q)$  sottostante, calcolata secondo i metodi ordinari, e  $\mathbf{J} = [\gamma_u(0), \gamma_u(1), \dots, \gamma_u(p-1), \gamma_v(0), \gamma_v(1), \dots, \gamma_v(q-1)]$ , con:

$$\begin{cases} \gamma_u(l) = \sum_{i=0}^{+\infty} \frac{\varphi'_i}{l+i+1} \\ \gamma_v(l) = \sum_{i=0}^{+\infty} \frac{\psi'_i}{l+i+1}; \end{cases}$$

e la distribuzione asintoticamente normale:

$$\sqrt{T}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \xrightarrow{d} N(0, \mathbf{I}^{-1});$$

si dimostra infine che lo stimatore è  $\sqrt{T}$  – consistente.

È importante osservare che, nella (4.3.1), abbiamo implicitamente considerato che la media non condizionata del processo sia zero; quando la media figura tra i parametri da stimare, Li e McLeod (1986) osservano che la velocità di convergenza si riduce a  $d - \frac{1}{2}$ . Si osserva inoltre che la presenza di una media ignota induce distorsione verso il basso nella stima di  $d$ .

Questo comportamento è evidenziato anche da Chung e Baillie (1993), che osservano, attraverso uno studio su dati simulati, che la presenza di  $\mu$  nel vettore  $\theta$  induce distorsione anche negli altri parametri.

Un altro punto debole di questa metodologia risiede nella scelta dei valori a cui condizionare. Per stimare un modello ARMA, servono un numero di valori pre-campionari pari al maggiore tra  $p$  e  $q$ . Normalmente i valori pre-campionari sono estratti dalla serie di cui si dispone semplicemente “mettendoli da parte” ed escludendoli dalla stima; perdere qualche osservazione in genere non modifica significativamente l’ampiezza del campione. Abbiamo già detto che nel caso di un modello ARFIMA non possiamo troncare la serie dei coefficienti binomiali troppo presto; ciò vale anche per i valori iniziali: ci serviranno quindi  $M$  valori pre-campionari. Considerando che in genere  $M$  è dell’ordine delle migliaia e che raramente una serie giornaliera può superare le diecimila osservazioni, se vogliamo applicare questo approccio ci troviamo costretti a utilizzare serie che possono essere più lunghe del doppio rispetto alla dimensione campionaria effettiva che utilizziamo per la stima; ciò è senz’altro poco desiderabile.

Il modo per ovviare a questo problema senza abbandonare l’idea del condizionamento è quello di condizionare a valori pari alla media non condizionata del processo. A dire il vero questo approccio è assai diffuso, e non ci si è preoccupati di osservare gli effetti che ciò può avere sulle stime che si ottengono. Riteniamo che l’argomento valga la pena di essere approfondito, perché condizionare a mille valori tutti uguali è ben diverso dal farlo su un solo

valore<sup>14</sup>. Pur in assenza di ulteriori risultati, crediamo comunque che sia più prudente troncare la serie e privarla dei valori iniziali.

### 4.3.2 Massima verosimiglianza esatta

La funzione di verosimiglianza esatta di un generico processo ARFIMA  $(p, d, q)$  è stata ricavata da Sowell (1992a): seguiremo la sua impostazione.

Sia  $\mathbf{y} = (y_1, y_2, \dots, y_T)$  una serie generata da un modello ARFIMA  $(p, d, q)$ , stazionario e distribuito normalmente con media zero e matrice di varianze e covarianze  $\mathbf{V}$ . La funzione di verosimiglianza è, richiamando la (4.1.8):

$$\ell(\boldsymbol{\theta}) = (2\pi)^{-T/2} |\mathbf{V}|^{-1/2} \exp\left[-\frac{1}{2} \mathbf{y}' \mathbf{V}^{-1} \mathbf{y}\right]; \quad (4.3.5)$$

per massimizzare la (4.3.5) dobbiamo esprimere  $\mathbf{V}$  in funzione di  $\boldsymbol{\theta}$ .

Cominciamo col notare che, poiché  $\mathbf{y}$  è stazionario, la matrice  $\mathbf{V}$  è una forma di Toeplitz, cioè:

$$\mathbf{V}_{ij} = \gamma_{i-j}. \quad (4.3.6)$$

Esprimiamo le autocovarianze in funzione dello spettro, utilizzando la (2.4.8):

$$\gamma_k = \frac{1}{2\pi} \int_0^{2\pi} s(\omega) e^{i\omega k} d\omega. \quad (4.3.7)$$

Nella (3.2.10) abbiamo ricavato lo spettro per un processo ARFIMA  $(0, d, 0)$ . Per ottenere quello di un ARFIMA  $(p, d, q)$  dobbiamo procedere per gradi. In primo luogo liberiamo la serie dall'integrazione frazionaria e calcoliamone lo spettro.

Sia  $u_t = (1-L)^d y_t$ , che, per come è definito  $\{y_t\}$ , risulta ARMA  $(p, q)$ : la rappresentazione di

Wold sarà  $u_t = \frac{\psi(L)}{\varphi(L)} \varepsilon_t$ . Poiché  $\{y_t\}$  è stazionario, le radici di  $\varphi(z)$  giacciono al di fuori del

cerchio unitario. Possiamo pertanto scrivere:

---

<sup>14</sup> Come si vedrà meglio in seguito, la questione del condizionamento è stata affrontata da Teyssière (1996) nel contesto dei modelli a memoria lunga per la varianza condizionata.

$$\varphi(z) = \prod_{j=1}^p (1 - \varphi_j z). \quad (4.3.8)$$

Lo spettro di un generico processo ARMA  $(p, q)$  risulta dalla (2.4.7):

$$s(\omega) = \frac{\sigma^2}{2\pi} \frac{|\psi(z)|^2}{|\varphi(z)|^2}, \quad (4.3.9)$$

dove, come di consueto,  $z = e^{-i\omega}$ .

Sostituendo la (4.3.8) nella (4.3.9) otteniamo:

$$s_u(\omega) = \frac{\sigma^2}{2\pi} |\psi(z)|^2 \prod_{j=1}^p (1 - \varphi_j z)^{-1} (1 - \varphi_j z^{-1})^{-1}. \quad (4.3.10)$$

Se  $\varphi(z)$  ha solo radici reali possiamo scomporre la (4.3.10):

$$s_u(\omega) = \frac{\sigma^2}{2\pi} |\psi(z)|^2 \sum_{j=1}^p z^p \zeta_j \left[ \frac{\varphi_j^{2p}}{1 - \varphi_j z} - \frac{1}{1 - \varphi_j^{-1} z} \right], \quad (4.3.11)$$

dato:

$$\zeta_j = \left[ \varphi_j \prod_{i=1}^p (1 - \varphi_i \varphi_j) \prod_{m=1, m \neq j}^p (\varphi_j - \varphi_m) \right]^{-1}. \quad (4.3.12)$$

Poiché  $\psi(z)$  è di ordine finito, possiamo scriverlo come polinomio in  $z$ :

$$\psi(z) = \sum_{l=-q}^{+q} \xi_l z^l, \quad (4.3.13)$$

dato:

$$\xi_l = \sum_{s=\max[0, l]}^{\min[q, q-l]} \psi_s \psi_{s-l}. \quad (4.3.14)$$

D'altra parte lo spettro di  $\{y_t\}$  risulterà dal prodotto dello spettro di  $\{u_t\}$  per quello di un ARFIMA  $(0, d, 0)$ , nella forma della (3.2.9):

$$s(\omega) = (1 - z)^{-d} (1 - z^{-1})^{-d} s_u(\omega). \quad (4.3.15)$$

Sostituendo per prima cosa la (4.3.13) nella (4.3.11) si ottiene la:

$$s_u(\omega) = \frac{\sigma^2}{2\pi} \sum_{l=-q}^{+q} \xi_l z^l \sum_{j=1}^p \zeta_j z^p \left[ \frac{\varphi_j^{2p}}{1 - \varphi_j z} - \frac{1}{1 - \varphi_j^{-1} z} \right], \quad (4.3.16)$$

che inserita nella (4.3.15) restituisce:

$$s(\omega) = \frac{\sigma^2}{2\pi} (1-z)^{-d} (1-z^{-1})^{-d} \sum_{l=-q}^{+q} \sum_{j=1}^p \xi_l \zeta_j z^{p+l} \left[ \frac{\varphi_j^{2p}}{1 - \varphi_j z} - \frac{1}{1 - \varphi_j^{-1} z} \right]. \quad (4.3.17)$$

A questo punto, per trovare le autocovarianze, possiamo inserire la (4.3.17) nella (4.3.7).

Definiamo, per sintetizzare:

$$C(d, h, r) = \frac{1}{2\pi} \int_0^{2\pi} \left[ \frac{r^{2p}}{1 - rz} - \frac{1}{1 - r^{-1}z} \right] (1-z)^{-d} (1-z^{-1})^{-d} z^h d\omega; \quad (4.3.18)$$

che inserito nella (4.3.17) restituisce:

$$\gamma_k = \sigma^2 \sum_{l=-q}^{+q} \sum_{j=1}^p \xi_l \zeta_j C(d, p+l-k, \varphi_j). \quad (4.3.19)$$

D'altra parte si dimostra che (Sowell, 1992a – pagina 185):

$$C(d, h, r) = \frac{\Gamma(1-2d)\Gamma(d+h)}{\Gamma(1-d+h)\Gamma(1-d)\Gamma(d)} \times \\ \times \left[ r^{2p} F_r(d+h, 1, 1-d+h) + F_r(d-h, 1, 1-d-h) - 1 \right], \quad (4.3.20)$$

il che ci consente di calcolare le autocovarianze senza dover calcolare l'integrale nella (4.3.18), utilizzando la (2.1.9) per la funzione ipergeometrica e dovendo computare solo la funzione  $\Gamma$ .

Un altro vantaggio “computazionale” è che i valori di  $C(d, h, r)$  possono essere ottenuti per via ricursiva. Infatti l'unico parametro variabile per le diverse autocovarianze è  $h$ ; per la (2.1.3) applicata alla funzione ipergeometrica si ha:

$$F_c(a, 1, b) = \frac{b-1}{c(a-1)} [F_c(a-1, 1, b-1) - 1]; \quad (4.3.21)$$

e a partire dalla (4.3.21) possiamo calcolare i valori  $C(d, h, r)$  utilizzando solo  $p$  computazioni della funzione ipergeometrica.

A questo punto, possiamo tracciare un sommario dei passi che servono per calcolare la verosimiglianza di un certo vettore di parametri:

- Fattorizziamo i polinomi AR e MA, calcolando la (4.3.8) e la (4.3.13);
- Utilizziamo i coefficienti AR così trovati per calcolare gli  $\zeta_j$  dalla (4.3.12);
- Calcoliamo i valori di  $C(d, h, r)$ ;
- Calcoliamo la matrice  $\mathbf{V}$ ;
- La invertiamo e ne calcoliamo il determinante;
- Inseriamo infine questi ultimi valori nella (4.3.5), dove finalmente possiamo vedere il valore della verosimiglianza.

Questo stimatore è asintoticamente normale poiché soddisfa le condizioni di normalità asintotica ricavate da Dahlhaus (1989)<sup>15</sup> e valide per tutti gli stimatori di massima verosimiglianza di processi a memoria lunga.

Per valutarne le proprietà in campioni limitati, Sowell confronta, su dati simulati, le performance del suo stimatore, di quello di Fox e Taqqu (1986) e di quello di Geweke e Porter-Hudak (1983), ottenendo risultati assai soddisfacenti in termini di distorsione osservata e MSE<sup>16</sup>. Una grossa limitazione è però rappresentata dal fatto che si ipotizza che il processo sia stazionario, limitandoci di fatto ai casi in cui  $d < 1/2$ .

---

<sup>15</sup> Tali condizioni sono assai complesse e richiedono, per essere analizzate, strumenti di teoria della probabilità avanzati. Non si è ritenuto opportuno appesantire ulteriormente questo lavoro trattandole dettagliatamente, tanto più che sono richiamate solo in questo contesto; si rimanda pertanto il lettore interessato ad una trattazione più approfondita all'articolo citato.

<sup>16</sup> Il Mean Squared Error (MSE) è un indicatore utilizzato per la selezione degli stimatori che pesa sia l'efficienza che l'eventuale distorsione. Per una trattazione più completa si rimanda a Greene (2000 – pagina 104).

### 4.3.3 La verosimiglianza di Fox e Taqqu

Fox e Taqqu (1986) considerano la possibilità di utilizzare l'approccio approssimato di Whittle<sup>17</sup> (1951), cioè massimizzare:

$$\ell_w(\boldsymbol{\theta}) = \frac{1}{\boldsymbol{\sigma}} \exp\left(-\frac{\mathbf{z}'\mathbf{V}\mathbf{z}}{2T\boldsymbol{\sigma}^2}\right), \quad (4.3.22)$$

dove  $\mathbf{z}$  è il vettore degli scarti dalla media campionaria delle osservazioni  $y_t$  e  $\mathbf{V}$  è la matrice di varianze e covarianze delle osservazioni.

Poiché i parametri compaiono solo in  $\mathbf{V}$ , massimizzare la (4.3.22) equivale a minimizzare l'esponente o, ugualmente, a massimizzare:

$$\sigma_T^2(\boldsymbol{\theta}) = \frac{\mathbf{z}'\mathbf{V}\mathbf{z}}{T}. \quad (4.3.23)$$

Si dimostra (si veda Beran, 1994, pagina 116), che la (4.3.22) equivale a:

$$\ell_w(\boldsymbol{\theta}) = \frac{1}{2\pi} \int_{-\pi}^{+\pi} \ln s(\omega; \boldsymbol{\theta}) d\omega + \frac{1}{2\pi} \int_{-\pi}^{+\pi} \frac{\hat{s}(\omega)}{s(\omega; \boldsymbol{\theta})} d\omega,$$

dove  $\hat{s}(\omega)$  è il periodogramma della serie.

Ai fini computazionali è particolarmente fastidiosa e impegnativa la presenza degli integrali in  $\omega$  nella formula precedente: una versione discreta analoga è data semplicemente da:

$$\ell_w(\boldsymbol{\theta}) = \frac{2}{T} \sum_{j=1}^m \ln s(\omega_j; \boldsymbol{\theta}) + \frac{2}{T} \sum_{j=1}^m \frac{\hat{s}(\omega_j)}{s(\omega_j; \boldsymbol{\theta})}, \quad (4.3.24)$$

dove  $m = T - 1 / 2$  è il numero di punti del periodogramma e  $\omega_j = 2\pi j / T$ .

---

<sup>17</sup> L'idea di fondo è quella di scrivere la funzione di verosimiglianza non nel dominio del tempo, ma nel dominio della frequenza. Considerato che una derivazione più formale avrebbe richiesto l'introduzione di ulteriori concetti matematici approfonditi, come la trasformazione di Fourier, si è preferito presentare la materia in termini più intuitivi, proponendo comunque le formule che si utilizzano, all'atto pratico, per l'utilizzo di questo metodo. Per una trattazione più approfondita si rimanda a Gouriéroux e Monfort (1997 – pagina 314).

La (4.3.24) è il “banco di lavoro” del metodo di Fox e Taqqu (1986): come si è già accennato, la stima  $\hat{\boldsymbol{\theta}}$  è il vettore che rende massima la (4.3.24). Fox e Taqqu (1986 – pagina 526) dimostrano, per  $0 < d < 1/2$ , la consistenza e la normalità asintotica di  $\hat{\boldsymbol{\theta}}$ :

$$\sqrt{T}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \xrightarrow{d} N(0, 4\pi \mathbf{W}^{-1}), \quad (4.3.25)$$

dove la matrice  $\mathbf{W}$ , quadrata e di dimensioni pari al numero di parametri da stimare, è così composta:

$$w_{ij} = \int_{-\pi}^{+\pi} s(\omega; \boldsymbol{\theta}) \frac{\partial^2 s^{-1}(\omega, \boldsymbol{\theta})}{\partial \theta_i \partial \theta_j} d\omega. \quad (4.3.26)$$

A causa della necessità di calcolare lo spettro teorico ad ogni “tentativo”, il metodo in questione si presenta piuttosto intensivo a livello computazionale, specialmente con ordini AR e MA superiori al primo, per i quali è richiesta la fattorizzazione dei polinomi in  $L$ . Il tempo che la procedura di ottimizzazione impiega per raggiungere la convergenza è sensibilmente superiore a quello impiegato dal metodo della massima verosimiglianza condizionata.

Ciononostante, in alcune situazioni, il metodo che abbiamo appena presentato si rivela molto utile. Cheung e Diebold (1996) osservano che lo stimatore di massima verosimiglianza esatta di Sowell perde di efficienza, in campioni limitati, quando la media della popolazione è ignota e figura tra i parametri da stimare. La (4.3.24), invece, è invariante rispetto alla media della popolazione, per cui può rappresentare in casi di questo tipo una valida alternativa.

Chiudiamo questo paragrafo menzionando un'altra procedura di massima verosimiglianza che lavora nel dominio della frequenza. L'idea di McCoy e Walden (1996) per gli ARFIMA  $(0, d, 0)$ , estesa da Jensen (1999) agli ARFIMA  $(p, d, q)$ , è quella di applicare allo spettro della serie la *discrete wavelet transform* e utilizzare i coefficienti della funzione madre e le loro varianze come approssimazione della funzione di verosimiglianza. Si tratta senz'altro di metodologie promettenti, anche se molto complesse, ma la loro trattazione richiederebbe l'uso di nozioni matematiche estremamente approfondite ed esula dagli scopi di questo lavoro.



#### 4.3.4 Il metodo GPH

Geweke e Porter-Hudak (1983) sono stati i primi statistici a proporre una metodologia di stima originale per i processi  $I(d)$ . Questo metodo, e gli altri due che presenteremo di seguito, concentrano l'attenzione sulla stima di  $d$ , per utilizzare poi il valore così ottenuto per la stima degli altri parametri. Questo approccio, come abbiamo già accennato nell'introduzione a questa sezione, prende il nome di stima in due passi.

Supponiamo che la serie  $\{y_t\}$ , se differenziata  $d$  volte, produca dei residui  $I(0)$ :  $(1-L)^d y_t = u_t$ , dove  $u_t$  ha una generica formulazione ARMA. Come abbiamo già visto nella (4.3.15) possiamo scrivere:

$$\begin{aligned} s(\omega) &= \frac{\sigma^2}{2\pi} \left(4 \sin^2 \frac{1}{2} \omega\right)^{-d} s_u(\omega) \Rightarrow \\ \Rightarrow \ln s(\omega) &= \ln \frac{\sigma^2 s_u(0)}{2\pi} - d \ln \left(4 \sin^2 \frac{1}{2} \omega\right) + \ln \frac{s_u(\omega)}{s_u(0)}. \end{aligned} \quad (4.3.27)$$

Siano  $\omega_j = \frac{2\pi j}{T}$  per  $j = 0, 1, \dots, T-1$ , e sia  $\hat{s}(\omega_j)$  il valore del periodogramma al punto  $\omega_j$ ,

calcolato secondo la (2.4.11). Riscriviamo la (4.3.27) come:

$$\begin{aligned} \ln \hat{s}(\omega_j) &= \ln \frac{\sigma^2 s_u(0)}{2\pi} - d \ln \left(4 \sin^2 \frac{1}{2} \omega_j\right) + \\ &+ \ln \frac{s_u(\omega_j)}{s_u(0)} + \ln \frac{\hat{s}(\omega_j)}{s_u(\omega_j)}, \end{aligned} \quad (4.3.28)$$

Dalla somiglianza della (4.3.28) con un modello di regressione lineare nasce l'intuizione che ha generato la metodologia GPH: regredire il logaritmo naturale del periodogramma sul logaritmo dello spettro di un processo ARFIMA  $(0, d, 0)$  (il secondo addendo) e su una costante, e utilizzare il coefficiente di tale regressione come  $\hat{d}$ .

Geweke e Porter-Hudak (1983 – pagina 226) osservano, utilizzando uno studio di simulazione, la consistenza di  $\hat{d}$  solo per  $d < 0$ , Robinson (1990) la ricava invece per  $0 < d < 1/2$ .

La stima GPH ha dalla sua parte il grosso pregio della semplicità: non serve calcolare la funzione di verosimiglianza, non servono metodi di ottimizzazione particolari: ci serve solo il periodogramma e una semplice regressione lineare. Questa semplicità è però pagata dalla scarsa precisione della stima che si ottiene. Il confronto operato da Sowell (1992a) mette in evidenza come la stima di massima verosimiglianza sia dalle 2 alle 4 volte più precisa, sia in termini di distorsione che di MSE. Inoltre, Agiakloglou, Newbold e Wohar (1992) osservano che, in campioni limitati, la stima GPH dà risultati ancora meno precisi quando  $\{u_t\}$  è autocorrelato; cioè in pratica in tutti i casi in cui i coefficienti AR sono significativi.

Si sono fatti vari tentativi per “ripescare” il metodo GPH, che è senza ombra di dubbio attraente nella sua semplicità. Robinson (1994) propone di utilizzare una versione smussata del periodogramma (si confronti il paragrafo 2.4.6):

$$S(\omega) = \int_0^{\omega} s(g) dg, \quad (4.3.29)$$

e trova una stima consistente della (4.3.29):

$$\hat{S}(\omega) = \frac{2\pi}{T} \sum_{j=1}^{\omega T/2\pi} \left| \frac{1}{\sqrt{2T\pi}} \sum_{t=1}^T y_t e^{i\omega_j t} \right|^2, \quad (4.3.30)$$

che poi utilizza nella (4.3.28). Lobato e Robinson (1996) dimostrano la normalità asintotica di questo stimatore solo per  $d < 1/4$ .

Né questa né le molte altre modifiche del metodo GPH che sono state proposte (si vedano, tra gli altri, Giriatis, Robinson e Samarov, 1997, oppure Robinson e Henry, 1999, che considerano anche la possibilità che vi sia eteroschedasticità nei residui) danno risultati ragguardevoli. Citando Baillie (1996): «*Despite the amount of theoretical work in attempting*

*to devise robust semiparametric estimators of the long-memory parameter, there is substantial evidence documenting their poor performance in terms of bias and mean squared error [...]. Overall the consensus of evidence is somewhat negative about semiparametric estimation, with adjustments to the periodogram at low frequencies appearing unlikely to radically improve their small sample performance.»*

#### 4.3.5 MDE

Un'ultima possibilità di stima del parametro  $d$ , conosciuta come *minimum distance estimator*, è considerata da Tieslau, Schmidt e Baillie (1996).

Le autocorrelazioni di un processo ARFIMA  $(0, d, 0)$  sono, richiamando la (3.2.14):

$$\rho_j = \frac{\Gamma(1-d)\Gamma(j+d)}{\Gamma(d)\Gamma(j-d+1)} = \prod_{i=1}^j \frac{d+i-1}{i-d}, \quad (4.3.31)$$

mentre le autocorrelazioni osservate sulla serie  $\{y_t\}$ :

$$\hat{\rho}_j = \frac{\sum_{t=1}^{T-j} (y_t - \bar{y})(y_{t-j} - \bar{y})}{\sum_{t=1}^T (y_t - \bar{y})^2}. \quad (4.3.32)$$

La stima di  $d$  è il valore  $\hat{d}$  che minimizza la funzione:

$$S(d) = [\hat{\boldsymbol{\rho}} - \boldsymbol{\rho}(d)]' \mathbf{V}^{-1} [\hat{\boldsymbol{\rho}} - \boldsymbol{\rho}(d)], \quad (4.3.33)$$

dove  $\mathbf{V}$  è la matrice di varianze e covarianze di  $\hat{\boldsymbol{\rho}}$ , definita in (3.2.18).

La MDE è formulata in maniera semplice ed è computazionalmente poco impegnativa, tuttavia soffre di un pesante limite: poiché le autocorrelazioni dei processi ARFIMA  $(p, d, q)$

non sono  $\sqrt{T}$ -consistenti per  $d \geq \frac{1}{4}$  (si veda il paragrafo 3.2.3), non lo sarà neanche lo

stimatore, che convergerà invece a velocità  $T^{\frac{1}{2}-d}$ .

#### 4.3.6 Una tabella riepilogativa

Riteniamo utile a questo punto confrontare direttamente e sinteticamente i metodi che abbiamo presentato:

<b>Metodo</b>	<b>Pro</b>	<b>Contro</b>
<i>Massima verosimiglianza condizionata</i>	- Formulazione molto semplice.	- Rischio di troncare l'espansione di $(1 - L)^d$ troppo presto e di perdere informazioni sulla memoria lunga; - Perdita di osservazioni per il condizionamento; - Distorsione se $\mu$ è tra i parametri da stimare.
<i>Massima verosimiglianza esatta</i>	- Grande precisione.	- Valida solo se il processo è stazionario; - Distorsione se $\mu$ è tra i parametri da stimare; - Lentezza e complessità computazionale.
<i>Massima verosimiglianza di Fox e Taqqu</i>	- Utile quando $\mu$ figura tra i parametri da stimare.	- Lentezza computazionale.
<i>GPH</i>	- Semplicità applicativa; - Grande rapidità computazionale.	- Scarsa precisione, specie se figurano radici AR vicine al cerchio unitario.
<i>MDE</i>	- Semplicità applicativa.	- Convergenza più lenta per $d > 1/4$ .

**Tabella 4.1**

*Confronto tra le metodologie di stima ARFIMA*

## 4.4 La stima dei processi GARMA

Dopo avere introdotto la classe di processi GARMA, Gray, Zhang e Woodward (1989) hanno proposto, per la stima di tali processi, di ricavare i valori di  $d$  e  $\xi$  tramite *grid search*, di filtrare la serie attraverso il polinomio di Gegenbauer così ottenuto, e di stimare, per ciascuna casella della griglia, i parametri ARMA attraverso la massima verosimiglianza condizionata.

Il problema di questo tipo di approccio è che, nonostante i parametri sottoposti a *grid search* siano solo due, è comunque molto dispendioso in termini di tempi di calcolo e di memoria occupata.

Chung (1996b) invece massimizza direttamente la funzione di verosimiglianza, applicando il filtro  $(1 - 2\xi L + L^2)^d$  contestualmente agli altri. I risultati che questo approccio dà su dati simulati sono estremamente promettenti: la distorsione è in genere molto bassa. Problemi si presentano invece quando  $\xi$  si avvicina a uno. In questo caso la “vicinanza” col modello ARFIMA fa sentire il suo peso: la frequenza di Gegenbauer si abbassa e la periodicità, di conseguenza, sale. Ciò significa che servono più osservazioni per coprire un ciclo e che, quindi, si ha bisogno di campioni più grandi.

Chung (1996b) ricava altresì le distribuzioni asintotiche degli stimatori CSS. Si dimostra (Chung, 1996b – pagina 118) in primo luogo che, data la media campionaria  $\bar{x}$  del processo e dato  $\xi \neq 1$ , valgono:

$$\begin{aligned}\sqrt{T}(\bar{x} - \mu) &\xrightarrow{d} N(0, 2\pi s(0)), \\ \sqrt{T}(\hat{\mu} - \mu) &\xrightarrow{d} N(0, 2\pi s(0)).\end{aligned}\tag{4.4.1}$$

Questo risultato è molto importante, perché ci consente di evitare tutti i problemi che pone stimare  $\mu$  contestualmente agli altri parametri nelle stime condotte coi metodi di massima verosimiglianza. Infatti se  $\hat{\mu}$  e  $\bar{x}$  convergono alla stessa distribuzione possiamo utilizzare quest’ultimo come stima di  $\mu$ , inserirlo nella (4.1.6) e “fare finta” di essere in un caso con

media nota, saltando così a piè pari tutti i problemi che pone la stima di  $\mu$ , messi in evidenza per gli ARFIMA da Chung e Baillie (1993) e Cheung e Diebold (1994).

L'altro risultato importante riguarda la consistenza di  $\hat{\xi}$ , che converge più velocemente degli altri parametri:

$$T(\hat{\xi} - \xi) \xrightarrow{d} \frac{\sin G}{d} \frac{\int_0^1 B_1 dB_2 - \int_0^1 B_2 dB_1}{\int_0^1 B_1^2(r) dr + \int_0^1 B_2^2(r) dr}, \quad (4.4.2)$$

dove  $B_1$  e  $B_2$  sono due moti Browniani indipendenti; la (4.4.2) è tabulata da Chung (1996a).

La diversa velocità di convergenza di  $\xi$  rispetto agli altri parametri può provocare problemi se si utilizzano metodi di ottimizzazione basati sul gradiente; Chung (1996a) propone pertanto di utilizzare un algoritmo di ricerca lineare per l'ottimizzazione di  $\xi$  e un generico metodo basato sul gradiente per gli altri parametri.

Per quanto riguarda gli altri parametri, si dimostra invece (Chung, 1996b – pagina 120) che, dato  $\delta = [d \quad \phi' \quad \psi']'$ :

$$\sqrt{T}(\hat{\delta} - \delta) \xrightarrow{d} N(\mathbf{0}, \mathbf{I}_\delta^{-1}), \quad (4.4.3)$$

dove la matrice d'informazione è così definita<sup>18</sup>:

$$\mathbf{I}_\delta = \begin{bmatrix} I_d & \mathbf{I}_{d,\phi\psi} \\ \mathbf{I}'_{d,\phi\psi} & \mathbf{I}_{\phi\psi} \end{bmatrix}, \quad (4.4.4)$$

dove:

$$\mathbf{I}_{\phi\psi} = \begin{bmatrix} E(\mathbf{u}_t \mathbf{u}_t') & E(\mathbf{u}_t \mathbf{v}_t') \\ E(\mathbf{v}_t \mathbf{u}_t') & E(\mathbf{v}_t \mathbf{v}_t') \end{bmatrix}, \quad (4.4.5)$$

$$\mathbf{I}'_{d,\phi\psi} = [\gamma_{d\phi}(1), \dots, \gamma_{d\phi}(p), \gamma_{d\psi}(1), \dots, \gamma_{d\psi}(q)], \quad (4.4.6)$$

<sup>18</sup> Si noti che, ovviamente, la (4.4.4) è una generalizzazione della (4.3.3) di Li e McLeod (1986).

$$I_d = 2 \left( \frac{\pi^2}{3} - \pi G + G^2 \right); \quad (4.4.7)$$

e dove i vettori  $\mathbf{u}_t$  e  $\mathbf{v}_t$  che compaiono nella (4.4.5) sono dati da:

$$\begin{aligned} \mathbf{u}_t &= (U_t, U_{t-1}, \dots, U_{t-p+1})', \\ \mathbf{v}_t &= (V_t, V_{t-1}, \dots, V_{t-q+1})', \end{aligned}$$

coi processi  $\{U_t\}$  e  $\{V_t\}$  generati da:

$$\begin{aligned} \phi(L)U_t &= \varepsilon_t, \\ \psi(L)V_t &= \varepsilon_t, \\ \varepsilon_t &\sim N(0,1). \end{aligned}$$

Per quanto riguarda gli elementi di  $\mathbf{I}'_{d,\phi\psi}$ , essi valgono:

$$\begin{aligned} \gamma_{d\phi}(i) &= 2 \sum_{k=0}^{+\infty} \frac{\cos[(k+i)G]}{k+i} \phi_k, \\ \gamma_{d\psi}(j) &= 2 \sum_{k=0}^{+\infty} \frac{\cos[(k+j)G]}{k+j} \psi_k. \end{aligned}$$

La possibilità di utilizzare la funzione di verosimiglianza esatta è considerata sempre da Chung (1996b). Si osserva, ripercorrendo i passi di Sowell (1992b), che le autocovarianze di un processo GARMA  $(p, \xi, d, q)$  possono essere così espresse:

$$\left\{ \begin{aligned} \gamma_j &= \sum_{l=-q}^{+q} \eta_l \sum_{k=1}^p \zeta_k C(j, p-l, p+l, \phi_k) \\ C(j, p-l, p+l, \phi_k) &= \phi_k^{-2p} \sum_{m=0}^{+\infty} \phi_k^{-m} \tilde{\gamma}_{j-p-l-m} + \sum_{m=0}^{+\infty} \phi_k^{-n} \tilde{\gamma}_{j-p-l+n} \\ \zeta_k &= \left[ \phi_k^{-1} \prod_{i=1}^p (1 - \phi_i^{-1} \phi_k^{-1}) \prod_{m=1}^p (\phi_k^{-1} - \phi_m^{-1}) \right]^{-1} \\ \eta_l &= \sum_{k=0}^{q-|l|} \psi_k \psi_{k+|l|}, \end{aligned} \right. \quad (4.4.8)$$

dove  $\tilde{\gamma}_j$  è l'autocovarianza del processo di Gegenbauer che abbiamo introdotto nella (3.6.12); manca purtroppo una formula ricorsiva come quella della (4.3.21), e ciò rende questa procedura di stima estremamente lenta.

## 4.5 Un esempio pratico

Presentiamo in questa sezione un'esemplificazione di come si può analizzare una serie di dati reali attraverso un modello di classe ARFIMA; tale esemplificazione potrà essere utile sia per valutare meglio il funzionamento delle varie metodologie di stima fin qui proposte che per comprendere come si può procedere quando si lavora con serie a memoria lunga.

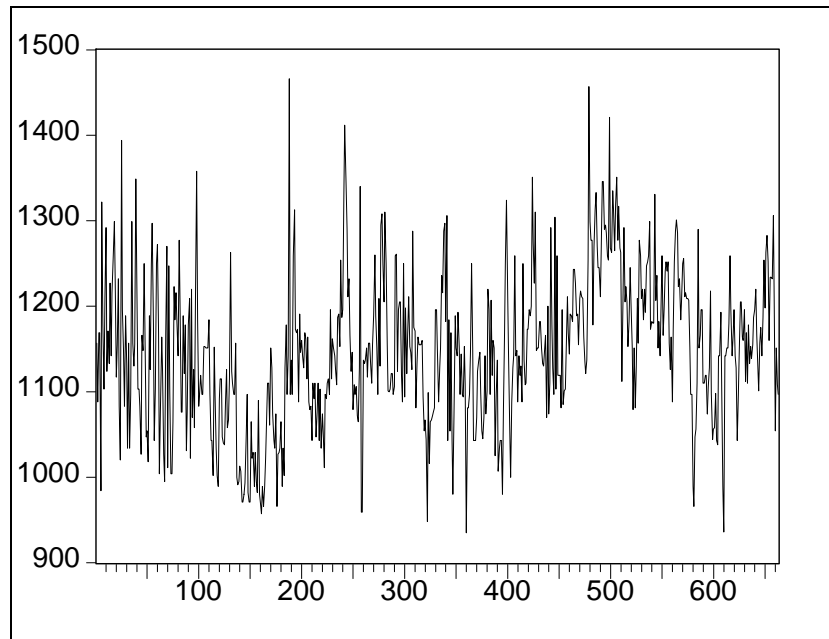
Proporremo un caso particolarmente esemplificativo per la modellistica a memoria lunga, ovvero la “celebre” serie sulle piene del Nilo. Abbiamo avuto modo di notare che la modellistica a memoria lunga ha avuto origine nel campo dell'idrologia; non sorprende quindi che una delle serie più utilizzate e conosciute dagli “addetti ai lavori” sia quella sul livello minimo del fiume Nilo, rilevate nei pressi del Cairo negli anni tra il 622 e il 1281 (663 osservazioni)<sup>19</sup>.

L'andamento della serie è riportato in figura 4.1; nelle figure 4.2 e 4.3 presentiamo, rispettivamente, le prime 500 autocorrelazioni e il periodogramma.

---

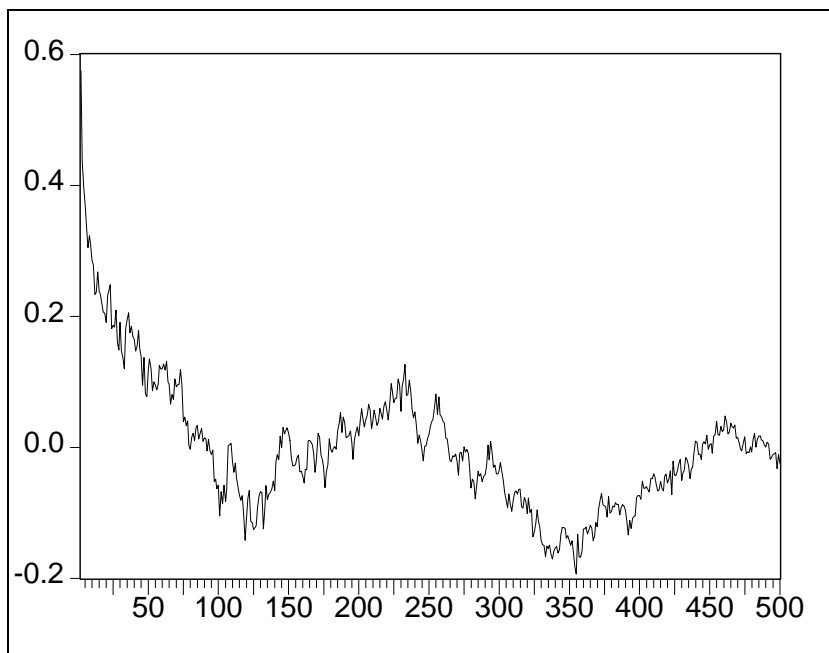
<sup>19</sup> I dati, presentati nel testo di Beran (1994), sono disponibili all'URL <http://lib.stat.cmu.edu/S/beran>.





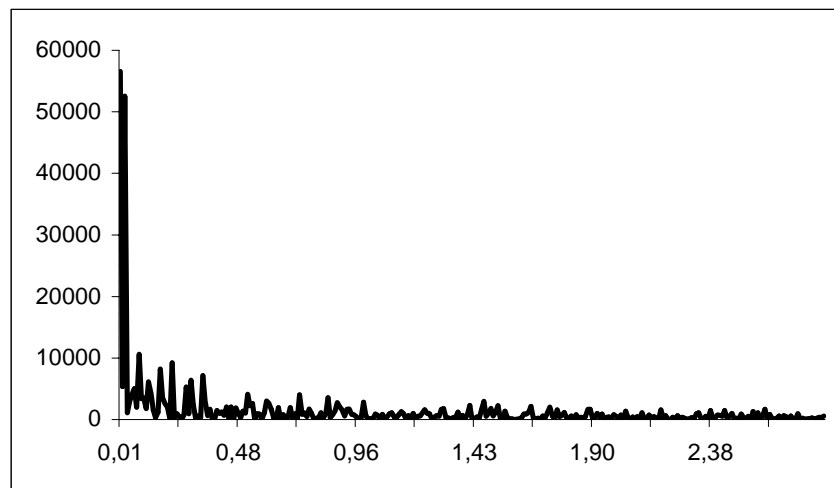
**Figura 4.1**

*Andamento del livello minimo del fiume Nilo*



**Figura 4.2**

*Autocorrelazioni, fino al cinquecentesimo ritardo, della serie della figura 4.1*



**Figura 4.3**

*Periodogramma della serie riportata in figura 4.1*

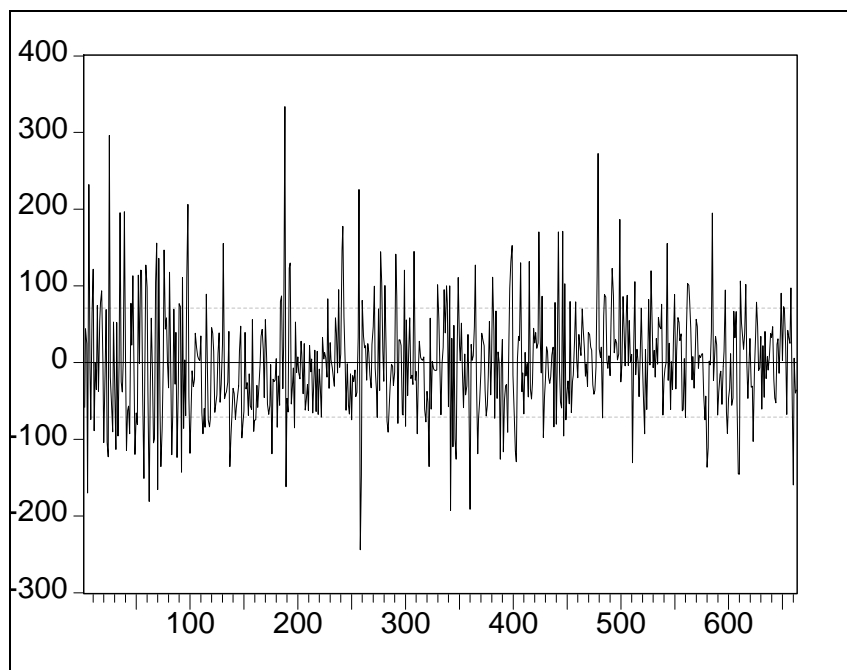
Dall'osservazione della bassa velocità di decadimento del correlogramma e del picco alle basse frequenze nel periodogramma possiamo ipotizzare la presenza di memoria lunga all'interno della serie. Abbiamo quindi provato a filtrare i dati con un modello ARMA (1, 1), ottenendo i risultati che presentiamo nella tabella 4.2.

<i>Variabile</i>	<i>Coefficiente</i>	<i>Standard error</i>	<i>Statistica t</i>	<i>P-value</i>
$\mu$	1147.985	10.67342	107.5555	0.0000
<i>AR (1)</i>	0.869552	0.029296	29.68123	0.0000
<i>MA (1)</i>	-0.495958	0.051522	-9.626194	0.0000
<i>Log-verosimiglianza</i>			-3759.249	
<i>AIC</i>			11.36631	
<i>SC</i>			11.38668	

**Tabella 4.2**

*Stima ARMA (1, 1) sui dati della figura 4.1*

I residui della regressione, presentati nella figura 4.4, sembrano presentare ancora marcata autocorrelazione; ciò è confermato dagli esiti del test Q di Ljung e Box (1979) (si veda il paragrafo A.1.2 in appendice) sui residui, che riportiamo nella tabella 4.3.



**Figura 4.4**

*Residui della regressione della tabella 4.2*

<i>Ritardi</i>	<i>AC</i>	<i>PAC</i>	<i>Statistica Q</i>	<i>P-value</i>
1	0.059	0.059	2.2891	/
2	-0.093	-0.097	8.1082	/
3	-0.045	-0.033	9.4439	0.002
4	-0.003	-0.008	9.4509	0.009
5	-0.013	-0.020	9.5574	0.023
6	-0.035	-0.036	10.385	0.034
7	0.046	0.047	11.790	0.038
8	0.047	0.034	13.281	0.039
9	0.022	0.023	13.608	0.059
10	0.042	0.051	14.786	0.063
11	-0.034	-0.033	15.565	0.077
12	-0.008	0.006	15.614	0.111
13	0.075	0.079	19.463	0.053
14	0.020	0.009	19.724	0.073
15	0.028	0.040	20.267	0.089
20	0.055	0.055	23.346	0.178
25	0.005	0.001	34.452	0.059
30	-0.010	-0.040	44.894	0.023
35	0.074	0.066	56.866	0.006
40	-0.014	0.001	58.441	0.018
45	-0.054	-0.041	65.810	0.014
50	0.065	0.037	76.869	0.005

**Tabella 4.3**

*Test Q per i residui della regressione della tabella 4.2*

Si può osservare come il risultato del filtraggio, soddisfacente per i ritardi tra il decimo e il ventesimo, peggiori decisamente quando si considerano ritardi più elevati, segno della mancata cattura della dipendenza di lungo periodo manifestata dai dati.

Siamo quindi passati alla stima di un modello ARFIMA, provando sia la specificazione  $(1, d, 1)$  che un semplice *fractional white noise*. Abbiamo in primo luogo eseguito la stima con il metodo della massima verosimiglianza condizionata, presentato nel paragrafo 4.3.1. I risultati sono riportati, rispettivamente, nelle tabelle 4.3 e 4.4<sup>20</sup>. Possiamo osservare come i coefficienti AR e MA risultino non significativi, inducendoci così a preferire la specificazione  $(0, d, 0)$ , selezionata anche dal criterio informativo di Schwarz (1978), più sensibile alla parsimonia.

<i>Variabile</i>	<i>Coefficiente</i>	<i>Standard error</i>	<i>Statistica t</i>	<i>P-value</i>
$\mu$	1148.125	/	/	/
$d$	0.372836	0.074981	4.972419	0.0000
$AR(1)$	-0.430947	1.071501	-0.402190	0.6877
$MA(1)$	0.484159	1.014963	0.477022	0.6335
<i>Log-verosimiglianza</i>			-3686.069	
<i>AIC</i>			11.11936	
<i>SC</i>			11.14876	

**Tabella 4.3**

*Stima ARMA (1, d, 1) sui dati della figura 4.1*

<i>Variabile</i>	<i>Coefficiente</i>	<i>Standard error</i>	<i>Statistica t</i>	<i>P-value</i>
$\mu$	1148.125	/	/	/
$d$	0.398681	0.055636	7.165910	0.0000
<i>Log-verosimiglianza</i>			-3686.314	
<i>AIC</i>			11.12010	
<i>SC</i>			11.12990	

**Tabella 4.4**

*Stima ARFIMA (0, d, 0) sui dati della figura 4.1*

<sup>20</sup> A causa dei problemi di distorsione che abbiamo rilevato, la costante è stata esclusa dalla procedura di stima e calcolata come semplice media aritmetica delle osservazioni; pertanto non sono riportati gli standard error e le statistiche t. Le stime sono state ottenute mediante un pacchetto ARFIMA sviluppato dall'autore di questo lavoro.

<i>Ritardi</i>	<i>AC</i>	<i>PAC</i>	<i>Statistica Q</i>	<i>P-value</i>
1	0.023	0.023	0.3542	0.552
2	-0.051	-0.051	2.0582	0.357
3	-0.006	-0.003	2.0811	0.556
4	0.015	0.013	2.2312	0.693
5	-0.007	-0.008	2.2624	0.812
6	-0.041	-0.040	3.4118	0.756
7	0.034	0.035	4.1841	0.758
8	0.027	0.022	4.6886	0.790
9	0.000	0.002	4.6886	0.861
10	0.027	0.031	5.1741	0.879
11	-0.051	-0.054	6.9116	0.806
12	-0.027	-0.024	7.4151	0.829
13	0.056	0.056	9.5364	0.731
14	-0.006	-0.012	9.5623	0.793
15	0.009	0.014	9.6118	0.843
20	0.039	0.037	12.552	0.896
25	-0.014	-0.021	20.682	0.710
30	-0.030	-0.048	30.288	0.451
35	0.061	0.065	40.385	0.244
40	-0.019	-0.005	41.620	0.400
45	-0.071	-0.055	48.439	0.336
50	0.046	0.040	58.472	0.192
75	0.000	-0.029	75.938	0.448
100	-0.099	-0.079	95.667	0.604

**Tabella 4.5**

*Test Q per i residui della regressione della tabella 4.4*

Il test di autocorrelazione sui residui, riportato nella tabella 4.5, dà stavolta risultati molto più soddisfacenti, specie per quanto riguarda i ritardi più elevati.

Passiamo ora a confrontare su questa serie di osservazioni i vari metodi di stima che abbiamo presentato nella sezione 4.3. Nella tabella 4.6 riportiamo le stime ricavate per il coefficienti di integrazione frazionaria  $d$ , il tempo in secondi impiegato dalla procedura<sup>21</sup> e il numero di iterazioni necessarie per la convergenza per quattro diversi metodi di stima: massima verosimiglianza condizionata (CML – paragrafo 4.3.1), massima verosimiglianza nel dominio della frequenza (FML – paragrafo 4.3.3), massima verosimiglianza esatta (EML – paragrafo 4.3.2) e metodo semiparametrico<sup>22</sup> di Geweke e Porter-Hudak (1983) (GPH – paragrafo 4.3.4).

<sup>21</sup> Su un processore Intel Celeron a 700MHz; il valore iniziale per  $d$  è stato posto uguale a 0.3.

<sup>22</sup> Non trattandosi di un metodo euristico, il numero di iterazioni ovviamente non compare.

	<i>CML</i>	<i>FML</i>	<i>EML</i>	<i>GPH</i>
<i>d</i>	0.398681	0.400402	0.392643	0.498082
<i>Tempo</i>	1.32	9.45	3.24	0.11
<i>Iterazioni</i>	3	3	4	/

**Tabella 4.6**

*Confronto tra i metodi di stima*

Possiamo osservare che le proprietà sintetizzate nella tabella 4.1 restano sostanzialmente verificate: il metodo GPH è il più rapido in assoluto, ma fornisce un risultato piuttosto difforme dagli altri; i metodi basati sulla funzione di verosimiglianza forniscono risultati piuttosto simili tra di loro, tra di essi tuttavia la massima verosimiglianza condizionata risulta più rapida, particolarmente in confronto alla massima verosimiglianza di Fox e Taqqu (1986).

# Capitolo 5

## Il modello FIGARCH

Dopo aver richiamato le caratteristiche fondamentali della modellistica ARCH, passeremo in rassegna le ricerche empiriche che hanno messo in evidenza i fenomeni di integrazione frazionaria all'interno della varianza condizionata e introdurremo i due modelli a memoria lunga finora proposti, cioè il LM-ARCH di Ding e Granger (1996) e il FIGARCH di Baillie, Bollerslev e Mikkelsen (1996), di cui si esamineranno dettagliatamente sia le caratteristiche che il comportamento all'atto pratico.

## 5.1 Introduzione alla modellistica ARCH<sup>1</sup>

La maggior parte delle ricerche macroeconomiche sono focalizzate sull'equazione della media della variabile che si analizza. Se ad esempio studiamo l'andamento del tasso di disoccupazione, ci interessa vedere da quali altre variabili esso è influenzato, e quale livello di disoccupazione ci possiamo ragionevolmente attendere in futuro. Poca attenzione è in genere dedicata ai momenti successivi della distribuzione della variabile, a cominciare dalla varianza.

Le ricerche che hanno tentato di isolare delle regolarità nei rendimenti dei mercati finanziari hanno però dato scarsi risultati: tali rendimenti appaiono difatti (specialmente quelli dei mercati dei cambi) decisamente poco autocorrelati. D'altra parte, per i gestori di portafoglio e gli analisti finanziari, non è rilevante solo riuscire a prevedere i rendimenti futuri degli attivi, ma anche la loro volatilità, di modo da poter misurare e coprire adeguatamente il rischio.

Definire il concetto di volatilità è complesso e non può portare a risultati univoci. In prima approssimazione si può dire che la volatilità misura la dispersione dei rendimenti che un certo attivo è suscettibile di generare. Volendo essere più formali, e attenendoci a un contesto tipicamente statistico, osserviamo che, se trattiamo il rendimento come una variabile aleatoria con una certa distribuzione nota, possiamo identificare, dato un livello di confidenza, un intervallo all'interno del quale si troverà, al dato livello di confidenza, il rendimento futuro. Se si modellano i rendimenti tramite una distribuzione con varianza variabile, ad esempio la normale, l'ampiezza dell'intervallo è determinata da questa varianza, per cui possiamo interpretarla come la volatilità del mercato e utilizzarla per calcolare il rischio di un certo portafoglio<sup>2</sup>.

Vale la pena ora approfondire la questione della distribuzione dei rendimenti. L'idea di considerare i rendimenti come normalmente distribuiti è intrigante, ma purtroppo poco

---

<sup>1</sup> In questa sezione si è seguita l'impostazione di Gallo e Pacini (2002).

<sup>2</sup> La questione sarà approfondita nel paragrafo 5.1.8.



realistica. Gli studi empirici che sono stati condotti (si vedano ad esempio Mandelbrot, 1963, e Fama, 1965) hanno messo in evidenza alcuni fatti:

- I rendimenti tendono ad essere leptocurtici, cioè ad avere code della distribuzione più spesse rispetto alla normale;
- La volatilità dei mercati si presenta “a grappoli”, cioè si ha che, citando Mandelbrot (1963): «...*large changes tend to be followed by large changes, of either sign, and small changes tend to be followed by small changes...*»;
- L’informazione che si accumula nelle giornate di chiusura dei mercati ha l’effetto di aumentare la volatilità, per cui quest’ultima tende ad essere più elevata nei giorni di riapertura (il che è noto anche come effetto *week-end* – si veda ad esempio Gallo e Pacini, 2002);
- Le variazioni negative dei prezzi tendono ad aumentare la volatilità, mentre quelle positive tendono a diminuirla (questo effetto, noto come effetto *leverage*, è stato notato da Black, 1976).

L’addensamento della volatilità è particolarmente importante: numerose ricerche (si può vedere ad esempio Taylor, 1986) hanno infatti mostrato che, mentre i rendimenti appaiono poco o addirittura non autocorrelati, i loro quadrati lo sono marcatamente e inequivocabilmente.

Uno studio particolarmente importante in questa direzione è quello condotto da Ding, Granger e Engle (1993) che studiano l’andamento dei valori assoluti di 17055 rendimenti giornalieri dell’indice Standard & Poor 500 elevati a una serie di esponenti  $k$ . I risultati più rilevanti che si ricavano<sup>3</sup> sono i seguenti:

- I rendimenti in sé sono chiaramente non autocorrelati;
- L’autocorrelazione è massima per  $k = 0.75$ ;

- L'autocorrelazione per  $k = 1$  è superiore a quella per  $k = 2$ , ma quest'ultima è comunque assai significativa, e resta positiva fino al ritardo 2598.

### 5.1.1 Il modello ARCH

La sigla ARCH sta per *AutoRegressive Conditional Heteroskedasticity*, ed è stata introdotta nella letteratura econometrica da Engle (1982), riscuotendo subito un enorme successo e fruttando all'autore il premio Nobel per l'economia nel 2003.

Dato un generico modello ARMA a media nulla della forma  $\varphi(L)r_t = \vartheta(L)\varepsilon_t$ , la persistenza nei rendimenti al quadrato che abbiamo menzionato poc' anzi può essere formalmente indicata ipotizzando che i residui al quadrato seguano un processo AR ( $m$ ):

$$\varepsilon_t^2 = \omega + \alpha_1 \varepsilon_{t-1}^2 + \alpha_2 \varepsilon_{t-2}^2 + \cdots + \alpha_m \varepsilon_{t-m}^2 + w_t, \quad (5.1.1)$$

dove:

$$w_t \sim N(0, \lambda^2). \quad (5.1.2)$$

La varianza condizionata del processo  $\{r_t\}$  sarà data da:

$$h_t = E(\varepsilon_t^2 | I_{t-1}) = \omega + \alpha_1 \varepsilon_{t-1}^2 + \alpha_2 \varepsilon_{t-2}^2 + \cdots + \alpha_m \varepsilon_{t-m}^2, \quad (5.1.3)$$

dove  $I_{t-1}$  rappresenta l'insieme informativo disponibile in  $t - 1$ .

Possiamo scrivere la (5.1.3) utilizzando l'operatore di ritardo che abbiamo introdotto nel paragrafo 2.3.2:

$$h_t = \omega + \alpha(L)\varepsilon_t^2. \quad (5.1.4)$$

Il nostro processo ARMA per la media diventa quindi:

$$\begin{cases} \varphi(L)r_t = \vartheta(L)\varepsilon_t \\ \varepsilon_t = u_t \sqrt{h_t}, \end{cases} \quad (5.1.5)$$

---

<sup>3</sup> Altri aspetti di questa ricerca riguardanti la velocità di decadimento delle autocorrelazioni saranno esaminate nel paragrafo 5.2.1.

dove  $\{u_t\}$  è un *white noise* a media nulla e varianza unitaria. Un modello di questo tipo, che considera  $m$  ritardi per la varianza condizionata, prende il nome di ARCH ( $m$ ).

L'utilizzo di un modello ARCH ci consente quindi di spiegare, e isolare, la persistenza nella volatilità dei mercati finanziari, riconducendo alla normalità i disturbi del processo generatore dei dati. Si può inoltre dimostrare<sup>4</sup> che in un modello ARCH la distribuzione degli  $\varepsilon_t$  è leptocurtica. Attraverso l'introduzione di un modello ARCH e l'utilizzo, nella (5.1.4), di una dummy per cogliere gli effetti *week-end* riusciamo quindi a spiegare ben tre dei quattro risultati empirici che abbiamo presentato nel paragrafo precedente.

Consideriamo ora il processo  $w_t = \varepsilon_t^2 - h_t$ : esso rappresenta le discrepanze che si osservano tra la varianza condizionata prevista e i residui al quadrato, ovvero, in altri termini, le “correzioni” alla varianza condizionata. Se inseriamo la (5.1.3) e la (5.1.5) nella (5.1.1) otteniamo:

$$w_t = h_t(u_t^2 - 1), \quad (5.1.6)$$

da cui:

$$\lambda^2 = E(w_t^2) = E(h_t^2)E(u_t^2 - 1)^2. \quad (5.1.7)$$

I valori prodotti dal processo  $\{h_t\}$  devono essere maggiori di zero. Per assicurarci di ciò imponiamo le condizioni:

$$\omega > 0 \wedge \alpha_j \geq 0 \quad \forall j = 1, 2, \dots, m. \quad (5.1.8)$$

Ovviamente un modello ARCH, per essere realistico, dovrà produrre un valore della varianza condizionata che non sia esplosivo. Sia:

$$\mathbf{e}'_t = (\varepsilon_t^2, \varepsilon_{t-1}^2, \dots, \varepsilon_{t-m}^2); \quad (5.1.9)$$

si dimostra (si veda Engle, 1982 – pagina 1004) che:

$$E(\mathbf{e}_t | I_{t-1}) = \mathbf{b} + \mathbf{A}\mathbf{e}_{t-1}, \quad (5.1.10)$$

dove:

$$\mathbf{b}' = (\omega, 0, \dots, 0) \quad (5.1.11)$$

e

$$\mathbf{A} = \begin{bmatrix} \omega & \alpha_1 & \cdots & \alpha_m & 0 \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix}. \quad (5.1.12)$$

La (5.1.10) può essere riscritta, facendo *backsubstitution*:

$$\mathbf{E}(\mathbf{e}_t \mid I_{t-k}) = (\mathbf{I} + \mathbf{A} + \mathbf{A}^2 + \cdots + \mathbf{A}^{k-1})\mathbf{b} + \mathbf{A}^k \mathbf{e}_{t-k}. \quad (5.1.13)$$

Poiché i valori iniziali di  $\mathbf{e}_t$  hanno varianza finita, il limite esiste se e solo se gli autovalori di  $\mathbf{A}$  giacciono tutti all'interno del cerchio unitario; in questo caso si ha:

$$\lim_{k \rightarrow +\infty} \mathbf{E}(\mathbf{e}_t \mid I_{t-k}) = (\mathbf{I} - \mathbf{A})^{-1} \mathbf{b}. \quad (5.1.14)$$

La varianza non condizionata dei residui è il valore atteso del primo elemento di  $\mathbf{e}$  calcolato a  $t = 0$ :

$$\lim_{k \rightarrow +\infty} \mathbf{E}(\mathcal{E}_t^2 \mid I_{t-k}) = \frac{\omega}{1 - \sum_{j=1}^m \alpha_j}. \quad (5.1.15)$$

Poiché imporre che gli autovalori di  $\mathbf{A}$  giacciono tutti all'interno del cerchio unitario equivale a richiedere che le radici dell'equazione caratteristica di  $\mathbf{A}$

$$1 - \alpha_1 z - \alpha_2 z^2 - \cdots - \alpha_m z^m = 0 \quad (5.1.16)$$

stiano all'esterno del cerchio unitario, se i coefficienti sono tutti non negativi la condizione per avere varianza finita è semplicemente:

$$\sum_{j=1}^m \alpha_j < 1. \quad (5.1.17)$$

---

<sup>4</sup> Si rimanda, per la dimostrazione, a Bollerslev, Engle e Nelson (1994).

### 5.1.2 Il modello GARCH

La più importante estensione del modello ARCH è dovuta a Bollerslev (1986). In numerosi studi empirici si era osservato che era frequente dover considerare un numero di ritardi elevato per cogliere appieno la dinamica della varianza condizionata; Bollerslev (1986) propone pertanto di generalizzare la (5.1.4) inserendo tra i regressori di  $h_t$  il proprio passato, utilizzando un polinomio  $[1 - \beta(L)]$  (ovviamente con radici esterne al cerchio unitario):

$$\begin{aligned} h_t &= \omega + \frac{\alpha(L)}{1 - \beta(L)} \varepsilon_t^2 \Rightarrow \\ \Rightarrow [1 - \beta(L)]h_t &= [1 - \beta(1)]\omega + \alpha(L)\varepsilon_t^2, \end{aligned} \quad (5.1.18)$$

che espandendo diventa:

$$\begin{aligned} h_t &= \bar{\omega} + \beta_1 h_{t-1} + \beta_2 h_{t-2} + \dots + \beta_r h_{t-r} + \\ &+ \alpha_1 \varepsilon_{t-1}^2 + \alpha_2 \varepsilon_{t-2}^2 + \dots + \alpha_m \varepsilon_{t-m}^2, \end{aligned} \quad (5.1.19)$$

con:

$$\bar{\omega} = \omega(1 - \beta_1 - \beta_2 - \dots - \beta_r). \quad (5.1.20)$$

Una modello di questo tipo prende il nome di GARCH ( $r, m$ ), dove la G sta per generalizzato. Utilizzando questo tipo di formulazione le dipendenze dai ritardi più lontani vengono colte dai parametri  $\beta$ , e spesso è sufficiente utilizzare  $r = 1$  per avere risultati soddisfacenti. Si potrebbe supporre che le condizioni per la non negatività di  $\{h_t\}$  siano, replicando la (5.1.8):

$$\begin{cases} \omega > 0 \\ \alpha_j \geq 0 \forall j = 1, 2, \dots, m \\ \beta_j \geq 0 \forall j = 1, 2, \dots, r. \end{cases} \quad (5.1.21)$$

In realtà Nelson e Cao (1992) dimostrano che queste condizioni sono sufficienti ma non necessarie; vincoli meno stringenti possono essere ricavati caso per caso. Per un GARCH (2, 1), ad esempio:

$$\begin{cases} \omega > 0 \\ \alpha_1 \geq 0 \\ 0 \leq \beta_1 < 1 \\ \alpha_1 \beta_1 + \alpha_2 \geq 0. \end{cases} \quad (5.1.22)$$

Si potrebbe essere tentati dall'ipotizzare che i termini  $\beta$  rappresentino la parte AR del GARCH, mentre gli  $\alpha$  quella MA; in realtà ciò non è del tutto esatto. Riscriviamo la (5.1.19) aggiungendo  $\varepsilon_t^2$  a entrambi i lati dell'uguaglianza:

$$\begin{aligned} h_t + \varepsilon_t^2 &= \bar{\omega} + \beta_1 h_{t-1} + \beta_2 h_{t-2} + \cdots + \beta_r h_{t-r} + \\ &+ \alpha_1 \varepsilon_{t-1}^2 + \alpha_2 \varepsilon_{t-2}^2 + \cdots + \alpha_m \varepsilon_{t-m}^2 + \varepsilon_t^2 = \\ &= \bar{\omega} - \beta_1 (\varepsilon_{t-1}^2 - h_{t-1}) - \beta_2 (\varepsilon_{t-2}^2 - h_{t-2}) - \cdots + \\ &- \beta_r (\varepsilon_{t-r}^2 - h_{t-r}) + \beta_1 \varepsilon_{t-1}^2 + \beta_2 \varepsilon_{t-2}^2 + \cdots + \beta_r \varepsilon_{t-r}^2 + \\ &+ \alpha_1 \varepsilon_{t-1}^2 + \alpha_2 \varepsilon_{t-2}^2 + \cdots + \alpha_m \varepsilon_{t-m}^2 + \varepsilon_t^2, \end{aligned} \quad (5.1.23)$$

da cui, utilizzando  $w_t = \varepsilon_t^2 - h_t$ , che come abbiamo detto è normalmente distribuito con media nulla e varianza  $\lambda^2$ :

$$\begin{aligned} \varepsilon_t^2 &= \bar{\omega} + (\alpha_1 + \beta_1) \varepsilon_{t-1}^2 + (\alpha_2 + \beta_2) \varepsilon_{t-2}^2 + \cdots + \\ &+ (\alpha_p + \beta_p) \varepsilon_{t-p}^2 + w_t - \beta_1 w_{t-1} - \beta_2 w_{t-2} - \cdots - \beta_r w_{t-r}, \end{aligned} \quad (5.1.24)$$

dove  $p = \max[m, r]$ . La (5.1.24) può essere scritta, utilizzando l'operatore  $L$ , come:

$$\begin{aligned} [1 - \alpha(L) - \beta(L)] \varepsilon_t^2 &= \bar{\omega} + [1 - \beta(L)] w_t \Rightarrow \\ \Rightarrow \psi(L) \varepsilon_t^2 &= \bar{\omega} + [1 - \beta(L)] w_t. \end{aligned} \quad (5.1.25)$$

Si conclude quindi che i residui al quadrato seguono un processo ARMA  $(p, r)$ , e sono stazionari in covarianza se:

$$\sum_{i=1}^p \alpha_i + \beta_i < 1. \quad (5.1.26)$$

Se la (5.1.26) è verificata si ha anche, per la varianza non condizionata:

$$\sigma^2 = E(\varepsilon_t^2) = \frac{\bar{\omega}}{1 - (\alpha_1 + \beta_1) - (\alpha_2 + \beta_2) - \cdots - (\alpha_p + \beta_p)}. \quad (5.1.27)$$

Bollerslev (1986) dimostra che, per un GARCH  $(p, q)$  la funzione di autocorrelazione dei residui al quadrato vale:

$$\rho_k = \sum_{i=1}^{\max[p,q]} (\alpha_i + \beta_i) \rho_{k-i}, \quad (5.1.28)$$

per  $k \geq q+1$ . Per applicare la (5.1.28) ci servono le prime  $q$  autocorrelazioni; nel semplice caso di un GARCH  $(1, 1)$  Bollerslev (1988) dimostra che:

$$\rho_1 = \alpha + \frac{\alpha^2 \beta}{1 - 2\alpha\beta - \beta^2}. \quad (5.1.29)$$

Dalla (5.1.28) e dalla (5.1.29) si ricava che la funzione di autocorrelazione per i disturbi al quadrato del nostro GARCH  $(1, 1)$  vale:

$$\rho_k = \left[ \alpha + \frac{\alpha^2 \beta}{1 - 2\alpha\beta - \beta^2} \right] (\alpha + \beta)^{k-1}, \quad (5.1.30)$$

da cui è evidente il decadimento esponenziale e, di conseguenza, la memoria corta.

Menzioniamo infine una diversa possibilità, considerata da Taylor (1986) e Schwert (1990), cioè quella di utilizzare, al posto della varianza condizionata e i residui al quadrato, la deviazione standard e i residui assoluti:

$$\begin{cases} s_t = \omega + \sum_{i=1}^p \alpha_i |\varepsilon_{t-i}| + \sum_{j=1}^q \beta_j s_{t-j} \\ \varepsilon_t = u_t s_t, \quad u_t \sim N(0,1). \end{cases} \quad (5.1.31)$$

L'idea di questa formulazione nasce dall'osservazione che la persistenza è più forte per i rendimenti assoluti piuttosto che per i rendimenti al quadrato; la loro diffusione è comunque assai limitata.

### 5.1.3 Il modello IGARCH

La famiglia di modelli IGARCH è stata introdotta da Engle e Bollerslev (1986) col proposito di “allungare” la memoria dei GARCH, imponendo semplicemente che la somma dei

coefficienti dei regressori valga 1. È immediato osservare che, in questo caso, la (5.1.27) ha denominatore zero, per cui la varianza non condizionata è infinita, e il processo non è stazionario in covarianza. Inoltre, la presenza di una costante positiva indica, analogamente a ciò che accade per i *random walk*, la presenza di un drift. È però dimostrato da Nelson (1990) e Bougerol e Picard (1992) che i processi IGARCH sono strettamente stazionari ed ergodici anche se non stazionari in covarianza<sup>5</sup>.

Si potrebbe arguire che, essendovi una radice unitaria all'interno della (5.1.24), il processo IGARCH abbia indubbiamente memoria lunga; questo però non è esatto. Ding e Granger (1996) dimostrano infatti che, per un IGARCH (1, 1), le autocorrelazioni dei rendimenti al quadrato valgono:

$$\rho_k = \frac{1}{3}(1 + 2\alpha)(1 + 2\alpha^2)^{-k/2}, \quad (5.1.32)$$

il che implica anche in questo caso un decadimento esponenziale. Ciò può essere visto anche scrivendo un processo IGARCH (1, 1) in forma ARCH ( $\infty$ ):

$$\begin{aligned} h_t &= \omega + \alpha \varepsilon_{t-1}^2 + (1 - \alpha)h_{t-1} = \\ &= \frac{\omega}{\alpha} + \alpha \sum_{k=1}^{+\infty} (1 - \alpha)^{k-1} \varepsilon_{t-k}^2; \end{aligned} \quad (5.1.33)$$

nella (5.1.33), infatti, si osserva che i coefficienti degli  $\varepsilon_{t-k}^2$  decadono esponenzialmente. Ciò può sembrare piuttosto singolare. In realtà bisogna fare chiarezza su cosa si intende per persistenza<sup>6</sup>. Seguiamo Ding e Granger (1996) e consideriamo un semplice IARCH (1):

$$\begin{cases} h_t = \varepsilon_{t-1}^2 \\ \varepsilon_t = u_t \sqrt{h_t}, \end{cases} \quad (5.1.34)$$

con  $\{u_t\}$ , al solito, white noise a media nulla e varianza unitaria.

---

<sup>5</sup> La prova di questo risultato, che appare senz'altro contrario alla comune intuizione, è particolarmente complessa, per cui si rimanda all'articolo citato. Il fatto saliente è che, nel caso in cui la varianza del processo non sia finita, la covarianza non è definita e quindi la stazionarietà stretta non implica la stazionarietà in covarianza.



Uno shock al tempo  $t$  potrà provenire solo da  $u_t$  e ovviamente lascerà intatto  $h_t$ , che dipende solo dalle informazioni passate. Poiché  $E(\varepsilon_{t+k}^2 | I_t) = E(h_{t+k} | I_t) = \varepsilon_t^2 = u_t^2 h_t$ , lo shock proveniente da  $u_t$  influenzerà permanentemente le aspettative sull'andamento futuro del processo. Ma il “vero”  $h_{t+k}$  è dato da:

$$h_{t+k} = \varepsilon_{t+k-1}^2 = u_{t+k-1}^2 \cdot u_{t+k-2}^2 \cdot \dots \cdot u_{t+1}^2 \cdot \varepsilon_t^2, \quad (5.1.35)$$

per cui l'impatto reale di un shock in  $t$  è:

$$\frac{\partial h_{t+k}}{\partial \varepsilon_t^2} = u_{t+k-1}^2 \cdot u_{t+k-2}^2 \cdot \dots \cdot u_{t+1}^2. \quad (5.1.36)$$

Poiché gli  $u_t$  sono distribuiti normalmente con media zero e varianza unitaria, la maggior parte di essi saranno molto vicini a zero (per essere precisi, circa il 68% sarà in valore assoluto minore di uno), per cui a maggior ragione risulteranno “piccoli” i loro quadrati. Ciò significa che il prodotto di tali quadrati convergerà quasi sicuramente a zero per  $k \rightarrow +\infty$ .

Mettersi d'accordo su cosa si intende per persistenza è quindi fondamentale: se con ciò intendiamo l'influenza del presente sulle aspettative future, senz'altro la classe IGARCH ha persistenza indefinita; se invece intendiamo piuttosto l'influenza reale che hanno sul presente gli shock passati, il modello non ha persistenza diversa da un GARCH ordinario. Vale la pena osservare che questo è molto diverso da quello che avviene nei processi per la media: nel caso di un *random walk* infatti previsioni sul futuro e influenza reale del passato coincidono.

#### 5.1.4 I modelli asimmetrici

Abbiamo visto, nell'introduzione a questa sezione, che i cambiamenti negativi nei prezzi azionari tendono ad aumentare la volatilità, e quelli positivi a ridurla. I modelli GARCH

---

<sup>6</sup> L'argomento è particolarmente complesso e verrà illustrato con un semplice esempio. Un punto di vista alternativo che sembra risolvere questo apparente paradosso è dovuto a Davidson (2002) e sarà illustrato nella sezione 5.3.

trattano gli shock positivi e negativi allo stesso modo, per cui non riescono a spiegare questo risultato empirico.

Una prima formulazione ARCH che tiene conto di questa asimmetria è introdotta da Zakoian (1991): a seconda che lo shock sia positivo o negativo, si modifica la (5.1.31) utilizzando un coefficiente diverso:

$$s_t = \omega + \sum_{i=1}^p \alpha_i^+ \varepsilon_{t-i}^+ + \sum_{i=1}^p \alpha_i^- \varepsilon_{t-i}^-, \quad (5.1.37)$$

dove  $\varepsilon_t^+ = \varepsilon_t$  se  $\varepsilon_t > 0$  o 0 altrimenti e  $\varepsilon_t^- = \varepsilon_t - \varepsilon_t^+$ . Questo modello prende il nome di Threshold – ARCH (TARCH).

Più raffinato è il modello di Nelson (1991), che contempla anche una diversa forza di questo effetto a seconda della dimensione dello shock. Si modifica il modello GARCH della (5.1.18) considerando, invece che i livelli, i logaritmi naturali della varianza condizionata. Il modello che ne risulta prende il nome di Exponential GARCH (EGARCH):

$$\begin{aligned} \ln h_t &= \bar{\omega} + \frac{1 + \alpha(L)}{1 - \beta(L)} \left[ \theta \frac{\varepsilon_{t-1}}{\sqrt{h_{t-1}}} + \gamma \left( \left| \frac{\varepsilon_{t-1}}{\sqrt{h_{t-1}}} \right| - \sqrt{\frac{2}{\pi}} \right) \right] = \\ &= \bar{\omega} + \frac{1 + \alpha(L)}{1 - \beta(L)} g \left( \frac{\varepsilon_{t-1}}{\sqrt{h_{t-1}}} \right), \end{aligned} \quad (5.1.38)$$

dove  $\sqrt{\frac{2}{\pi}}$  è il valore atteso del valore assoluto di disturbi normali con media nulla e varianza unitaria.

### 5.1.5 La stima dei modelli ARCH

La stima dei modelli della famiglia GARCH non presenta particolari problemi, se si utilizza il condizionamento. Difatti basta sostituire, nella (4.1.6),  $h_t$  a  $\sigma^2$ , e ovviamente inserire i parametri GARCH nel vettore  $\theta$ , per poi procedere a massimizzare la funzione di

verosimiglianza attraverso i metodi di ottimizzazione descritti nella sezione A.2 in appendice.

Precisamente, la funzione da massimizzare sarà:

$$\ln \ell(\boldsymbol{\theta}) = -\frac{T}{2} \ln 2\pi - \sum_{t=1}^T \frac{\ln h_t}{2} - \sum_{t=1}^T \frac{\varepsilon_t^2}{2h_t}. \quad (5.1.39)$$

Nonostante quanto abbiamo detto a proposito della distribuzione dei termini di disturbo nei modelli ARCH, osserviamo che risulta in ogni caso molto difficile, all'atto pratico, riuscire ad estrarre residui che possano essere considerati normalmente distribuiti.

Bollerslev (1987) propone ad esempio, per superare questo problema, di utilizzare disturbi t-distribuiti, Nelson (1991) utilizza invece un altro tipo di distribuzione nota come GED (*Generalized Error Distribution*).

In alternativa, se non si ritiene di poter specificare una forma parametrica per la distribuzione dei termini di disturbo, si può fare ricorso al principio di quasi-massima verosimiglianza che abbiamo esaminato nel paragrafo 4.1.3. La consistenza e la normalità asintotica degli stimatori QML sono provate formalmente<sup>7</sup>, per tutti i tipi di processi con varianza condizionata variabile, da Bollerslev e Wooldridge (1992).

### 5.1.6 I test ARCH<sup>8</sup>

Testare la presenza di effetti ARCH in una serie è relativamente semplice. La metodologia più comune è quella introdotta da Engle (1982), che propone di regredire i residui al quadrato dell'equazione della media su un certo numero di propri ritardi; i coefficienti di tale regressione dovranno, in caso di presenza di eteroschedasticità condizionata, essere significativamente positivi.

---

<sup>7</sup> Anche in questo caso si è preferito omettere la complessa dimostrazione, per la quale si rimanda all'articolo citato. Si osservi comunque che questo risultato era stato congetturato e utilizzato sin dall'introduzione dei modelli ARCH.

<sup>8</sup> In questo paragrafo si è seguita la presentazione di Gallo e Pacini (2002).

I metodi per condurre il test sono due: il primo è di considerare semplicemente la statistica  $F$  generata dal vincolare tutti i coefficienti a zero; il secondo è quello di utilizzare un test LM. In questo caso, Engle (1982 – pagina 1000) dimostra che la statistica test può essere approssimata da:

$$T \cdot R^2 \xrightarrow{d} \chi_m^2, \quad (5.1.40)$$

dove  $m$  indica il numero di ritardi che si considerano nell'autoregressione dei residui al quadrato.

Alcune metodologie di test per l'individuazione delle asimmetrie negli shock sono proposte da Engle e Ng (1993). Una prima idea può essere quella di utilizzare la regressione:

$$\frac{\hat{\varepsilon}_t^2}{h_t} = \alpha + \beta S_{t-1}^-, \quad (5.1.41)$$

dove  $S_t^-$  è una dummy che assume valore uno se  $\varepsilon_t > 0$ . Sotto l'ipotesi nulla di simmetria il coefficiente  $\beta$  deve valere zero: il modo più semplice per verificarlo è utilizzare la statistica:

$$\frac{\hat{\beta}}{\hat{\sigma}_{\hat{\beta}}} \sim t_{T-2}; \quad (5.1.42)$$

questo approccio prende il nome di *sign bias test*.

L'idea che sta alla base del *negative size bias test* è di considerare la possibilità vi sia non solo un effetto dovuto agli shock negativi, ma che l'entità di tale effetto dipenda dalla dimensione dello shock:

$$\frac{\hat{\varepsilon}_t^2}{h_t} = \alpha + \gamma S_{t-1}^- \hat{\varepsilon}_{t-1}; \quad (5.1.43)$$

per testare la presenza di effetti asimmetrici facciamo di nuovo riferimento alla (5.1.42).

Ovviamente la (5.1.43) può essere modificata per tenere conto di effetti da shock positivi, semplicemente utilizzando una dummy  $S_t^+$ ; si parla in questo caso di *positive size bias test*.

I test che abbiamo presentato possono essere combinati, dando così vita a un test congiunto.

La regressione che si utilizza è:

$$\frac{\hat{\varepsilon}_t^2}{h_t} = \alpha + \beta S_{t-1}^- + \gamma S_{t-1}^- \hat{\varepsilon}_{t-1} + \delta S_{t-1}^+ \hat{\varepsilon}_{t-1}, \quad (5.1.44)$$

e il test si effettua semplicemente vincolando a zero i coefficienti  $\beta$ ,  $\gamma$  e  $\delta$  della (5.1.44): in questo caso si utilizza un test F con distribuzione  $F_{3,T-4}$ .

### 5.1.7 Le previsioni sulla volatilità

Le ragioni dell'enorme successo della modellistica ARCH risiedono anche nel suo grande rilievo pratico: i modelli ARCH non sono solo “giocattoli” per econometrici, ma vengono quotidianamente utilizzati da analisti finanziari e gestori di portafogli. Questo perché un modello ARCH, una volta ottenute le stime dei parametri, può essere utilizzato per fare previsioni sulla varianza condizionata. La questione è relativamente semplice. Rileviamo in primo luogo che, poiché  $w_t$  ha media nulla, si ha (si confronti la (5.1.3)):

$$E(\varepsilon_{t+k}^2 | I_t) = h_{t+k}; \quad (5.1.45)$$

se consideriamo un semplice GARCH (1, 1), la previsione in  $t$  sulla varianza dopo un periodo è:

$$E(\varepsilon_{t+1}^2 | I_t) = E(h_{t+1} | I_t) = \bar{\omega} + \alpha \varepsilon_t^2 + \beta h_t, \quad (5.1.46)$$

per cui si ha anche, per  $s$  periodi in avanti, utilizzando la (5.1.45):

$$\begin{aligned} E(\varepsilon_{t+s}^2 | I_t) &= E(h_{t+s} | I_t) = \\ &= \bar{\omega} + \alpha E(\varepsilon_{t+s-1}^2 | I_t) + \beta E(h_{t+s-1} | I_t) = \\ &= \bar{\omega} + (\alpha + \beta) E(h_{t+s-1} | I_t). \end{aligned} \quad (5.1.47)$$

Effettuando  $s$  *backsubstitution* fino a  $t$  otteniamo infine:

$$E(\varepsilon_{t+s}^2 | I_t) = \bar{\omega} \sum_{j=0}^{s-2} (\alpha + \beta)^j + (\alpha + \beta)^{s-1} (\alpha \varepsilon_t^2 + \beta h_t). \quad (5.1.48)$$

Se  $\alpha + \beta < 1$ , per  $s \rightarrow +\infty$  il secondo addendo tende a zero, mentre la sommatoria nel primo addendo converge a  $\frac{1}{1-\alpha-\beta}$ . Si ha quindi:

$$\lim_{s \rightarrow +\infty} E(\varepsilon_{t+s}^2 | I_t) = \frac{\bar{\omega}}{1-\alpha-\beta}, \quad (5.1.49)$$

che per la (5.1.27) equivale alla varianza non condizionata del processo; se ne conclude che le previsioni convergono esponenzialmente a un livello di equilibrio pari alla varianza non condizionata.

Consideriamo ora il caso IGARCH, in cui  $\alpha + \beta = 1$ . In questo caso, se la costante è diversa da zero, il primo addendo della (5.1.48) non converge, per cui si ha una previsione esplosiva. Se invece la costante vale zero, rimane solo il secondo addendo che vale, indipendentemente da  $s$ ,  $\alpha \varepsilon_t^2 + \beta h_t$ . Le previsioni pertanto non decadono verso un determinato livello di equilibrio, ma rimangono costanti.

Torniamo ora alla (5.1.48) e osserviamo che l'effetto del disturbo attuale<sup>9</sup>  $\varepsilon_t^2$  (che ovviamente non ha effetti su  $h_t$ ) dopo  $s$  periodi è determinato da  $\alpha(\alpha + \beta)^{s-1}$ , che per  $\alpha + \beta < 1$  decade esponenzialmente; se ne conclude che in un modello GARCH i risultati passati vengono “dimenticati” con velocità esponenziale. Al contrario, nel caso IGARCH, il coefficiente di  $\varepsilon_t^2$  resta costante ad  $\alpha$ , per cui l'effetto sulle previsioni future dei risultati attuali persiste indefinitamente, come avevamo già notato nel paragrafo 5.1.3.

---

<sup>9</sup> Se al posto di  $\varepsilon_t^2$  utilizziamo  $w_t$ , cioè la correzione rispetto al valore atteso dei residui al quadrato (in altre parole, lo shock in  $t$ ), otteniamo la cosiddetta *impulse response function*, introdotta da Gallant, Rossi e Tauchen (1993).

### 5.1.8 Il calcolo del Value at Risk<sup>10</sup>

Un altro campo d'impiego dei modelli ARCH riguarda il calcolo del *Value at Risk*. Il *Value at Risk*, o, più brevemente, *VaR*, è un indicatore utilizzato per la gestione e la copertura del rischio di portafoglio.

Il rischio di un certo attivo finanziario è dato in generale da tre componenti:

- Rischio di mercato;
- Rischio liquidità;
- Rischio controparte.

La prima di queste componenti riguarda la possibilità che il valore dell'attivo che si considera risulti diverso, per i naturali movimenti del mercato, da quello atteso; la seconda tiene conto del fatto che possa essere difficile liquidare immediatamente la posizione che si è assunta; la terza infine riguarda la possibilità che l'emittente dei titoli che si detengono risulti insolvente.

Il *Value at Risk* è, come abbiamo detto, un indicatore che tiene conto delle prime due componenti di cui sopra<sup>11</sup>. In sostanza l'idea è di calcolare la massima perdita attesa, a un certo livello di probabilità (in genere 95%), e di moltiplicarla per il numero di giorni che si ritengono necessari per liquidare la posizione.

Supponiamo che i rendimenti siano distribuiti normalmente con media zero e varianza  $\sigma^2$ . In questo caso il punto critico che lascia sulla coda sinistra il 5% della densità di probabilità è  $-1.645\sigma$ , e tale valore rappresenta altresì, in termini percentuali (e ovviamente in valore assoluto), la perdita massima giornaliera di cui sopra.

---

<sup>10</sup> Il VaR è un argomento sconfinato e non pretendiamo di darne una trattazione sistematica; ci limiteremo agli aspetti più interessanti che si relazionano all'utilizzo dei modelli GARCH. Per una trattazione più completa si può vedere Cherubini e Della Lunga (2001) o Metelli (1995).

<sup>11</sup> Il rischio controparte, in mercati operati da *market makers*, risulta senz'altro meno rilevante degli altri due, almeno per quanto riguarda i mercati azionari e dei derivati. Discorso diverso meriterebbero i portafogli che contengono obbligazioni "a rischio", per le quali sono disponibili vari indicatori di rischio stimati da apposite agenzie.

Per fare un semplice esempio, se supponiamo che servano tre giorni per liquidare la posizione e che la varianza dei rendimenti sia 0.0001 si avrebbe:

$$VaR_3 = P_t \cdot 1.645 \sqrt{3 \cdot 0.0001} = 0.02849 P_t,$$

il che significa che il valore “a rischio” del portafoglio è il 2.849% del valore presente. La formula di cui sopra può essere riscritta, utilizzando una notazione generica e considerando un tempo di liquidazione pari a  $F$  giorni:

$$VaR_F = P_t \cdot 1.645 \sigma \sqrt{F}. \quad (5.1.50)$$

Il problema principale nel calcolo del  $VaR$  è la stima della volatilità. Una prima idea può essere quella di stimare la varianza dei rendimenti utilizzando la varianza storica. In generale il  $VaR$  dovrà essere calcolato su un portafoglio composto da  $N$  titoli; in questo caso tale varianza storica è data da:

$$\begin{cases} r_p = w_1 r_1 + \dots + w_N r_N \\ \text{var}(r_p) = \sum_{i=1}^N w_i^2 \text{var}(r_i) + 2 \sum_{i=1}^N \sum_{j=1, j \neq i}^N w_i w_j \text{cov}(r_i, r_j), \end{cases} \quad (5.1.51)$$

per cui nel calcolo del  $VaR$  ci serviranno non solo le varianze dei singoli titoli, ma anche  $N(N-1)/2$  covarianze.

Il primo modo per ovviare a questo problema è stato proposto dalla banca J.P. Morgan, che fornisce giornalmente una matrice di varianze e covarianze per buona parte dei titoli disponibili sul mercato. Tale approccio è noto come RiskMetrics™ (si veda AA. VV., 1993); le varianze così stimate non sono calcolate tramite un modello GARCH, ma tramite il più semplice *exponential smoothing*<sup>12</sup>:

$$\hat{\sigma}_t^2 = \lambda \hat{\sigma}_{t-1}^2 + (1 - \lambda)(r_t - \bar{r})^2. \quad (5.1.52)$$

---

<sup>12</sup> Si tratta sostanzialmente di un modello IGARCH senza costante in cui il parametro  $\lambda$ , invece che essere stimato sulla base dei dati storici, è assegnato dagli analisti (in genere si aggira intorno a 0.94) sulla base di valutazioni soggettive.



Un'altra possibilità che ha incontrato un discreto successo è stata proposta da Barone-Adesi e Giannopoulos (1996). L'idea è che si possano ottenere stime più precise del  $VaR$  utilizzando, al posto della varianza non condizionata, la varianza condizionata stimata tramite un modello GARCH. Nel caso in cui il portafoglio sia formato da un solo titolo, la questione è relativamente semplice: si stima il modello e lo si utilizza per prevedere la varianza condizionata. La (5.1.50) diventa così:

$$VaR_F = P_t \cdot 1.645 \sqrt{\sum_{f=0}^{F-1} h_{t+f}} . \quad (5.1.53)$$

Nel caso in cui il portafoglio sia composto da più attivi, si procede in questo modo: si costruisce una pseudo-serie formata dai rendimenti passati che si sarebbero ottenuti ipotizzando che il portafoglio avesse sempre avuto la composizione attuale, dopodiché si stima il modello GARCH su questa serie e, di nuovo, si calcola la varianza condizionata prevista.

## 5.2 La memoria lunga nei rendimenti ad alta frequenza

### 5.2.1 Le ricerche empiriche

Abbiamo già rilevato, nell'introduzione, come quello dei rendimenti dei mercati azionari o dei tassi di cambio è uno dei settori più promettenti per l'utilizzo efficace dell'integrazione frazionaria. Se il primo studio applicato ai rendimenti giornalieri del mercato azionario statunitense, condotto da Lo (1991) attraverso la statistica  $R/S$  da lui modificata, non ha evidenziato fenomeni di integrazione frazionaria, altri risultati ben più incoraggianti sono arrivati dagli studi sui rendimenti assoluti e quadrati.

Abbiamo già citato la ricerca di Ding, Granger e Engle (1993), che considerano i rendimenti assoluti giornalieri dell'indice Standard & Poor 500 elevati a una serie di esponenti  $k$  e ne osservano le autocorrelazioni. Relativamente alla presenza di memoria lunga nei rendimenti assoluti o quadrati del mercato azionario statunitense, si osserva che le autocorrelazioni restano positive fino ad oltre 2500 ritardi, cioè ben oltre a quanto si sarebbe dovuto osservare in caso di decadimento esponenziale delle stesse.

Ding, Granger e Engle (1993) procedono ricercando per la funzione di autocorrelazione una forma analitica in grado di spiegare questa marcata persistenza. Una modellazione ARMA tradizionale, in cui l'autocorrelazione è  $\rho_k = \alpha\beta^k$ , decresce troppo velocemente all'inizio della serie e troppo lentamente alla fine, per cui l'idea potrebbe essere quella di utilizzare un modello a integrazione frazionaria. Riscriviamo la (3.2.14) in forma ricorsiva, come avevamo già fatto per la (3.2.2), utilizzando la (2.1.3):

$$\rho_k = \rho_{k-1} \frac{k+d-1}{k-d}, \quad (5.2.1)$$

e richiamiamo la (3.2.15) per  $k$  sufficientemente grandi. Utilizzando questa formulazione, però, si ottiene il risultato opposto: decadimento troppo lento all'inizio e troppo veloce alla

fine della serie. L'idea di Ding, Granger e Engle (1993) è di combinare queste due possibilità utilizzando il modello:

$$\rho_k = \frac{\alpha \rho_{k-1}^{\beta_1} \beta_2^k}{k^{\beta_3}}; \quad (5.2.2)$$

linearizzando coi logaritmi e stimando i parametri attraverso il metodo dei minimi quadrati si ottiene:

$$\hat{\rho}_k = \frac{0.893 \hat{\rho}_{k-1}^{0.784} \cdot 0.999955^k}{k^{0.057}}.$$

Tuttavia, Ding, Granger e Engle (1993) ritengono sufficiente a spiegare tale persistenza l'utilizzo di un modello di tipo GARCH e procedono proponendo una generalizzazione, che prende il nome di Asymmetric – Power ARCH (A-PARCH):

$$\begin{cases} s_t^\delta = \omega + \sum_{i=1}^p \alpha_i (|\varepsilon_{t-i}| - \gamma_i \varepsilon_{t-i})^\delta + \sum_{j=1}^q \beta_j s_{t-j}^\delta \\ \varepsilon_t = u_t s_t, \quad u_t \sim N(0,1). \end{cases} \quad (5.2.3)$$

È immediato osservare che, a partire dalla (5.2.3) si possono ricavare:

- Il modello ARCH “originale” per  $\delta = 2$ ,  $\beta_j = 0 \forall j$  e  $\gamma_i = 0 \forall i$ ;
- Il modello GARCH della (5.1.19), per  $\delta = 2$  e  $\gamma_i = 0 \forall i$ ;
- Il GARCH della (5.1.31), per  $\delta = 1$  e  $\gamma_i = 0 \forall i$ ;
- Il TARCH della (5.1.37), per  $\delta = 1$ ,  $\beta_j = 0 \forall j$ ,  $\alpha_i^+ = \alpha_i(1 - \gamma_i)$  e  $\alpha_i^- = \alpha_i(1 + \gamma_i)$ ;

nonché altre tre formulazioni che non abbiamo qui esaminato.

Granger e Ding (1996a e 1996b) utilizzano la stessa serie per stimare, attraverso il metodo GPH, il coefficiente di integrazione frazionaria per il modello:

$$\begin{cases} h_t = [1 - (1 - L)^d] \frac{|r_t|}{E(|u_t|)} \\ r_t = u_t h_t, \quad u_t \sim N(0,1), \end{cases} \quad (5.2.4)$$

ottenendo  $d = 0.474$ .

Crato e de Lima (1994) rilevano invece la presenza di integrazione frazionaria nei residui al quadrato di cinque diversi indici del mercato azionario statunitense, due settimanali e tre giornalieri. Dopo aver filtrato le serie con un AR (5), si sono calcolati i valori di  $d$  attraverso i metodi GPH,  $R/S$  e  $R/S$  modificata di Lo (1991); ricavando indicazioni uniformi e forti nella direzione dell'integrazione frazionaria, specie per gli indici settimanali.

Sul “fronte” dei tassi di cambio una ricerca interessante è quella di Baillie e Bollerslev (1994). Si analizzano i logaritmi del tasso di cambio del dollaro con altre sette valute e si trova il vettore di cointegrazione<sup>13</sup>  $\hat{\alpha}$ ; dalle autocorrelazioni del termine di *error correction*  $\hat{\alpha}'y$  si individuano immediatamente “indizi” della presenza di integrazione frazionaria. Difatti la stima di un semplice *fractional white noise*, condotta col metodo CSS, dà il risultato, decisamente inequivocabile,  $\hat{d} = 0.89$ , con  $\hat{\sigma}_{\hat{d}} = 0.02$ . Si conclude pertanto che il termine di *error correction* non è  $I(0)$ , ma  $I(d)$ ; si può parlare quindi di cointegrazione frazionaria.

Andersen e Bollerslev (1997a) invece lavorano sui rendimenti assoluti, computati ogni cinque minuti per un anno, del tasso di cambio DM/\$. La stima di  $d$ , per il quale si ottiene un valore di 0.359, è condotta tramite il metodo GPH con le correzioni introdotte da Robinson (1994).

Si osserva che dopo l'applicazione alla serie del filtro  $(1 - L)^d$  lo spettro, stimato a partire dal periodogramma, risulta piatto per frequenze minori di  $\frac{2\pi}{288}$ , indicando così che il filtro

“ripulisce” la serie dalle relazioni di lungo periodo.

---

<sup>13</sup> Dati  $N$  processi  $I(1)$   $\{y_{it}\}$ , essi si dicono cointegrati se esiste una matrice  $\alpha$ , di dimensioni  $N \times T$ , diversa da zero tale che il processo che risulta dalla combinazione lineare  $\alpha'y_t$  è stazionario in covarianza. Tale serie prende il nome di termine di *error correction*. Per una trattazione più approfondita si rimanda a Hamilton (1994 – capitolo 21).

### 5.2.2 Il modello LM-ARCH

Ding e Granger (1996) proseguono l'analisi del correlogramma condotta da Ding, Granger e Engle (1993), cercando una modellazione più precisa e confrontando non tanto varie forme per la funzione di autocorrelazione, ma modelli ARCH in grado di generare una funzione di autocorrelazione di forma compatibile con quella del correlogramma.

L'idea è che, se un modello GARCH cattura le dipendenze di breve periodo, e un modello IGARCH quelle di lungo, si possano ottenere dei risultati interessanti aggregandoli:

$$\begin{cases} h_t = wh_{1,t} + (1-w)h_{2,t} \\ h_{1,t} = \alpha_1 \varepsilon_{t-1}^2 + (1-\alpha_1)h_{1,t-1} \\ h_{2,t} = \sigma^2(1-\alpha_2-\beta_2) + \alpha_2 \varepsilon_{t-1}^2 + \beta_2 h_{2,t-1} \\ \varepsilon_t = u_t \sqrt{h_t}, \quad u_t \sim N(0,1), \end{cases} \quad (5.2.5)$$

dove  $w$  rappresenta il peso che viene dato al modello IGARCH.

Espandendo in rappresentazione ARCH ( $\infty$ ) le due componenti della varianza condizionata otteniamo:

$$\begin{cases} h_{1,t} = \alpha_1 \sum_{k=1}^{+\infty} (1-\alpha_1)^{k-1} \varepsilon_{t-k}^2 \\ h_{2,t} = \sigma^2 \frac{1-\alpha_2-\beta_2}{1-\beta_2} + \alpha_2 \sum_{k=1}^{+\infty} \beta_2^{k-1} \varepsilon_{t-k}^2, \end{cases} \quad (5.2.6)$$

da cui:

$$\begin{aligned} h_t = & \sigma^2(1-w) \frac{1-\alpha_2-\beta_2}{1-\beta_2} + \\ & + \sum_{k=1}^{+\infty} \left[ w\alpha_1(1-\alpha_1)^{k-1} + (1-w)\alpha_2\beta_2^{k-1} \right] \varepsilon_{t-k}^2. \end{aligned} \quad (5.2.7)$$

Ritornando dalla rappresentazione della (5.2.7) a quella GARCH otteniamo:

$$\begin{aligned} h_t = & \sigma^2(1-w)\alpha_1(1-\alpha_2-\beta_2) + \\ & + [w\alpha_1 + (1-w)\alpha_2] \varepsilon_{t-1}^2 + \\ & - [w\alpha_1\beta_2 + (1-w)(1-\alpha_1)\alpha_2] \varepsilon_{t-2}^2 + \\ & + (1-\alpha_1 + \beta_2)h_{t-1} - (1-\alpha_1)\beta_2 h_{t-2}, \end{aligned} \quad (5.2.8)$$

cioè un GARCH (2, 2); in generale si dimostra che un modello di evoluzione della varianza condizionata generato da  $N$  modelli GARCH (1, 1) o IGARCH (1, 1) produce un GARCH ( $N, N$ ):

$$\begin{cases} h_t = \sum_{i=1}^N w_i h_{i,t}, & \sum_{i=1}^N w_i = 1 \\ h_{i,t} = \sigma^2(1 - \alpha_i - \beta_i) + \alpha_i \varepsilon_{t-1}^2 + \beta_i h_{i,t-1}, & i = 1, 2, \dots, N \\ \varepsilon_t = u_t \sqrt{h_t}, & u_t \sim N(0, 1). \end{cases} \quad (5.2.9)$$

È evidente a questo punto che ci siamo ritrovati nella situazione già descritta nel paragrafo 3.4.1. Ricalcando i passi che abbiamo già affrontato possiamo ipotizzare che  $\beta$  provenga da una distribuzione Beta generica, e che  $\alpha = (1 - \beta)\alpha^*$ , dove  $\alpha^*$  ha una distribuzione qualunque, nell'intervallo  $[0, 1]$ , con media  $\mu$  positiva. Riscriviamo la (5.2.9) utilizzando l'operatore di ritardo  $L$  e otteniamo:

$$\begin{cases} h_{i,t} = \sigma^2 \frac{1 - \alpha_i - \beta_i}{1 - \beta_i} + \frac{\alpha_i}{1 - \beta_i L} \varepsilon_{t-1}^2 \\ h_t = \sum_{i=1}^N w_i \left[ \sigma^2 \frac{1 - \alpha_i - \beta_i}{1 - \beta_i} + \frac{\alpha_i}{1 - \beta_i L} \varepsilon_{t-1}^2 \right]. \end{cases} \quad (5.2.10)$$

Ripercorrendo i passi che già conosciamo possiamo scrivere:

$$\begin{aligned} h_t &= \sigma^2 \int_0^1 \frac{1 - \alpha - \beta}{1 - \beta} dF(\alpha, \beta) + \int_0^1 \frac{\alpha}{1 - \beta L} \varepsilon_{t-1}^2 dF(\alpha, \beta) = \\ &= \sigma^2 \int_0^1 (1 - \alpha^*) dF(\alpha^*) + \sum_{k=1}^{+\infty} \varepsilon_{t-k}^2 \int_0^1 \alpha^* (1 - \beta) \beta^{k-1} dF(\alpha^*, \beta) = \\ &= \sigma^2 (1 - \mu) + \sum_{k=1}^{+\infty} \varepsilon_{t-k}^2 \int_0^1 \alpha^* dF(\alpha^*) \int_0^1 (1 - \beta) \beta^{k-1} dF(\beta) \Rightarrow \\ h_t &= \sigma^2 (1 - \mu) + \mu \sum_{k=1}^{+\infty} \frac{B(p + k - 1, q + 1)}{B(p, q)} \varepsilon_{t-k}^2. \end{aligned} \quad (5.2.11)$$

Il  $k$ -esimo coefficiente ARCH è:

$$a_k = \frac{B(p + k - 1, q + 1)}{B(p, q)} = \frac{q \Gamma(p + q) \Gamma(p + k - 1)}{\Gamma(q) \Gamma(p + q + k)} \equiv$$

$$\cong \frac{q\Gamma(p+q)}{\Gamma(q)} k^{-1-q}; \quad (5.2.12)$$

ciò implica un decadimento iperbolico degli shock nella varianza condizionata. Se poi confrontiamo la (5.2.12) i coefficienti della rappresentazione AR ( $\infty$ ) di un processo ARFIMA che abbiamo presentato nella (3.2.7) otteniamo che il processo che abbiamo creato è a memoria lunga; lo indichiamo con il nome di Long Memory ARCH di ordine  $q$ . Si osservi che, poiché  $\sum_{k=1}^{+\infty} a_k = 1$ , il modello LM-ARCH è in sostanza un IGARCH con infiniti termini e coi pesi più forti sui ritardi lontani.

Nel caso in cui  $\mu = 1$ , poiché  $\alpha_i = 1 - \beta_i$ , il processo che risulta è un'aggregazione di  $N$  modelli IGARCH, e la (5.2.11) diventa  $h_t = \sum_{k=1}^{+\infty} a_k \varepsilon_{t-k}^2$ .

Se invece ipotizziamo che  $p + q = 1$  la (5.2.11) diventa:

$$\begin{aligned} h_t &= \sigma^2(1-\mu) + \mu \sum_{k=1}^{+\infty} \frac{q\Gamma(k-q)}{\Gamma(1-q)\Gamma(k+1)} \varepsilon_{t-k}^2 = \\ &= \sigma^2(1-\mu) + \mu \left[ 1 - (1-L)^q \right] \varepsilon_t^2 \Rightarrow \\ (1-L)^q \varepsilon_t^2 &= \left( \frac{1}{\mu} - 1 \right) (\sigma^2 - h_t) + h_t (u_t^2 - 1). \end{aligned} \quad (5.2.13)$$

Se  $q < 1/2$ , il processo è stazionario e ha valore atteso pari alla varianza non condizionata:

$E(\varepsilon_t^2) = E(h_t) = \sigma^2$ ; in questo caso la parte destra della (5.2.13) è un *white noise*, per cui i residui al quadrato sono  $I(q)$ . Infine, per  $\mu = 1$ , la (5.2.13) diventa:

$$(1-L)^q \varepsilon_t^2 = h_t (u_t^2 - 1). \quad (5.2.14)$$

### 5.2.3 Il modello FIGARCH

Come abbiamo appena visto, il modello LM-ARCH è ricavato a partire dalle ipotesi sulla sua insorgenza, cioè l'aggregazione di diversi processi GARCH; Baillie, Bollerslev e Mikkelsen (1996) muovono invece a partire da un'intuizione in un certo senso più astratta.

L'idea di fondo, estremamente semplice, è quella di applicare alla formulazione GARCH della (5.1.25) il *fractional difference operator* che abbiamo definito nella (3.2.1):

$$\psi(L)(1-L)^d \varepsilon_t^2 = \omega + [1 - \beta(L)]w_t, \quad (5.2.15)$$

da cui, con semplici passaggi algebrici, si perviene alla rappresentazione ARCH ( $\infty$ ):

$$\begin{aligned} h_t &= \omega + \left[ 1 - \frac{\psi(L)(1-L)^d}{1 - \beta(L)} \right] \varepsilon_t^2 = \\ &= \omega + \lambda(L) \varepsilon_t^2. \end{aligned} \quad (5.2.16)$$

$(1-L)^d$ , per  $z = 1$ , vale zero, per cui nel polinomio  $\lambda(z)$  vi è una radice unitaria.

Analogamente a ciò abbiamo visto verificarsi per la classe IGARCH, la presenza di un valore  $\omega > 0$  indica la presenza di un drift, e implica quindi che il processo non sia stazionario. La prova della stazionarietà stretta fornita nell'articolo è però errata: non sempre i coefficienti della (5.2.16) sono dominati in valore assoluto da quelli della rappresentazione ARCH ( $\infty$ ) di un processo IGARCH. Una dimostrazione più formale della stazionarietà forte, che utilizza un teorema ricavato da Zaffaroni (2000) è fornita in Caporin (2002): la condizione che deve essere soddisfatta è che  $E(\ln w_t) < 0$ .

Per quanto riguarda la non negatività della varianza condizionata, valgono le stesse considerazioni che abbiamo fatto per i modelli GARCH, e le condizioni meno stringenti devono essere ricavate caso per caso. Per il FIGARCH (1,  $d$ , 1), Bollerslev e Mikkelsen

(1996) osservano che i coefficienti del polinomio  $\lambda(L) = \sum_{j=1}^{+\infty} \lambda_j L^j$  della (5.2.16), nel caso

FIGARCH (1,  $d$ , 1), valgono:



$$\begin{cases} \lambda_1 = d - \beta + \psi \\ \lambda_k = \beta\lambda_{k-1} + \left( \frac{k-1-d}{k} - \psi \right) \delta_{k-1} \\ \delta_k = d \frac{\Gamma(k-d)}{\Gamma(1-d)\Gamma(k+1)}. \end{cases} \quad (5.2.17)$$

Per assicurare la non negatività della varianza condizionata è sufficiente che tutti i coefficienti  $\lambda$  siano non negativi. Poiché i  $\delta$  sono tutti negativi (a parte  $\delta_0$ , che comunque non viene utilizzato), la questione è relativamente semplice; Bollerslev e Mikkelsen (1996) propongono di imporre:

$$\begin{cases} \beta - d \leq \psi \leq \frac{2-d}{3} \\ d \left[ \psi - \frac{1-d}{2} \right] \leq \beta(\psi - \beta + d), \end{cases} \quad (5.2.18)$$

mentre Chung (1999) suggerisce:

$$0 \leq \psi \leq \beta \leq d < 1. \quad (5.2.19)$$

Osserviamo che né la (5.2.18) né la (5.2.19) rappresentano condizioni necessarie, ma solo sufficienti e che, quel che è più importante, sono assolutamente alternative; ad esempio i coefficienti  $\psi = 0.6$ ,  $\beta = 0.7$  e  $d = 0.8$  soddisfano la (5.2.19) ma non la (5.2.18).

Relativamente ai coefficienti, vale la pena rimarcare l'osservazione di Chung (1999) relativamente al significato del coefficiente  $\psi$ . Nel FIGARCH di Baillie, Bollerslev e Mikkelsen (1996) non viene imposto alcun vincolo su tale coefficiente, cosicché non è necessariamente  $\psi(L) = 1 - \alpha(L) - \beta(L)$  come nel GARCH; in questa situazione è evidente che un FIGARCH  $(r, 0, m)$  non è equivalente a un GARCH  $(r, m)$ , e il coefficiente  $\psi$  assume un significato autonomo ancora da chiarire.

Vediamo ora quali sono le caratteristiche di memoria dei processi FIGARCH analizzando la persistenza all'interno delle previsioni delle informazioni presenti, sulla falsariga di quanto abbiamo fatto nel paragrafo 5.1.7 per i GARCH e IGARCH.

La questione è relativamente semplice se richiamiamo la possibilità di rappresentare un FIGARCH  $(p, d, q)$  come ARCH  $(\infty)$ :

$$h_t = \omega + \lambda(L)\varepsilon_t^2, \quad (5.2.20)$$

poiché è immediato osservare che, per una previsione di  $s$  periodi in avanti, l'effetto di  $\varepsilon_t^2$  sulla previsione è governato da  $\lambda_s$ . Nel caso di un semplice FIGARCH  $(0, d, 0)$  si ha:

$$h_t = \omega + (1 - L)^d \varepsilon_t^2, \quad (5.2.21)$$

per cui i coefficienti  $\lambda$  coincidono con i  $\varphi$  della (3.2.6)

$$\varphi_s = \frac{\Gamma(s - d)}{\Gamma(s + 1)\Gamma(-d)},$$

che per l'approssimazione di Stirling, come abbiamo già osservato, decadono iperbolicamente al crescere di  $s$ :

$$\lim_{s \rightarrow +\infty} \frac{\Gamma(s - d)}{\Gamma(s + 1)\Gamma(-d)} = \frac{s^{-d-1}}{\Gamma(-d)}. \quad (5.2.22)$$

Per quanto riguarda le specificazioni più complesse, si può procedere in maniera simile: volendo ad esempio fare il caso del FIGARCH  $(1, d, 1)$ , e richiamando i coefficienti  $\lambda$  ricavati da Bollerslev e Mikkelsen, che avevamo già introdotto nella (5.2.17):

$$\lambda_k = \beta\lambda_{k-1} + \left( \frac{k-1-d}{k} - \psi \right) \frac{\Gamma(k-d-1)}{\Gamma(1-d)\Gamma(k)}, \quad (5.2.23)$$

osserviamo che il primo addendo, risolto per *backsubstitution*, decresce esponenzialmente al crescere di  $k$ , mentre per il secondo vale:

$$\lim_{k \rightarrow +\infty} \left( \frac{k-1-d}{k} - \psi \right) \frac{\Gamma(k-d-1)}{\Gamma(1-d)\Gamma(k)} = \frac{1-\psi}{\Gamma(1-d)} k^{-d-1}; \quad (5.2.24)$$

si ha pertanto:

$$\lim_{k \rightarrow +\infty} \lambda_k = A\beta^k + \frac{1-\psi}{\Gamma(1-d)} k^{-d-1}. \quad (5.2.25)$$

che per  $d > 0$  implica un decadimento asintoticamente iperbolico.

Richiamando il risultato che si era ottenuto per i modelli GARCH e IGARCH, cioè rispettivamente il decadimento esponenziale e la persistenza indefinita delle informazioni presenti, osserviamo che il FIGARCH rappresenta un'importante “via di mezzo”.

Parlando di persistenza degli shock, osserviamo che un modello teorico per spiegare l'insorgenza di memoria lunga nella varianza condizionata dei rendimenti delle attività finanziarie è individuato da Parke (1999) a partire dal suo modello *Error-Duration*, che abbiamo esaminato nel paragrafo 3.4.2.

Consideriamo un mercato semplificato, in cui il prezzo è dato da una media delle aspettative degli operatori di quel mercato e varia al variare di tali aspettative. È ragionevole supporre che la volatilità del mercato diminuita dagli operatori bene informati e aumentata dai “dilettanti”: difatti questi ultimi vedono raramente verificate le loro aspettative e di conseguenza sono costretti a cambiare posizione più spesso, incrementando così la volatilità.

Sia  $g_{s,t}$  la funzione indicatore del mantenimento, al tempo  $t$ , di una posizione assunta al tempo  $s$  e sia  $\varepsilon_s$  una misura della qualità media dell'informazione disponibile sul mercato al tempo  $s$ . Poiché la volatilità complessiva del mercato sarà proporzionale alla qualità media

dell'informazione disponibile, otterremo l'usuale espressione (3.4.15):  $\ln(h_t) = \sum_{s=-\infty}^t g_{s,t} \varepsilon_s$ .

Notiamo infine che la stretta relazione che passa tra il modello LM-ARCH e il FIGARCH: la (5.2.14) corrisponde infatti alla (5.2.15) di un FIGARCH  $(0, d, 0)$ , in cui  $\omega = 0$ ,  $\psi(L) = 1$  e  $\beta(L) = 0$ .

Alcuni rilievi critici sulla parametrizzazione FIGARCH della (5.2.15) sono mossi da Chung (1999). Se riscriviamo la (5.2.15) utilizzando la (5.1.27) otteniamo:

$$\psi(L)(1-L)^d \varepsilon_t^2 = \psi(L)(1-L)^d \sigma^2 + [1 - \beta(L)]w_t, \quad (5.2.26)$$

per cui la costante si riduce a essere:

$$\omega = \sigma^2 \psi(1) \sum_{j=0}^{+\infty} \varphi_j, \quad (5.2.27)$$

dove i  $\varphi_j$  sono i coefficienti dell'espansione in serie di McLaurin del polinomio  $(1-L)^d$  presentati nella (3.2.6).

Si ha in primo luogo:

$$\sum_{j=0}^{k-1} \varphi_j = -\frac{k}{d} \varphi_k = \frac{\Gamma(k-d)}{\Gamma(k)\Gamma(1-d)}, \quad (5.2.28)$$

che per l'approssimazione di Stirling, introdotta con la (2.1.4), diventa, per  $k$  sufficientemente grande:

$$\sum_{j=0}^{k-1} \varphi_j \cong \frac{k^{-d}}{\Gamma(1-d)}, \quad (5.2.29)$$

da cui, immediatamente:

$$\sum_{j=0}^{+\infty} \varphi_j = 0. \quad (5.2.30)$$

Si conclude pertanto che la (5.2.27) è identicamente uguale a zero, indipendentemente dal valore di  $\sigma^2$ <sup>14</sup>. Questo risultato è problematico, poiché nelle stime finora condotte si sono osservate costanti significativamente maggiori di zero. Senz'altro ciò può essere dovuto al fatto che necessariamente la (5.2.30) deve essere troncata a fini computazionali, ma lascia comunque aperto il problema di quale significato dare a una costante che risulti significativa.

---

<sup>14</sup> Si è utilizzata la notazione  $\sigma^2$  unicamente per parallelismo con la (5.1.27) per il caso GARCH. Il significato di  $\sigma^2$  nei modelli FIGARCH (come del resto negli IGARCH) è ancora poco chiaro e necessita ulteriori approfondimenti.

Chung (1999) propone quindi per il FIGARCH una forma leggermente differente dalla (5.2.15):

$$\psi(L)(1-L)^d(\varepsilon_t^2 - \sigma^2) = [1 - \beta(L)]w_t, \quad (5.2.31)$$

dove, come abbiamo già osservato nella (5.2.27):

$$\sigma^2 = \frac{\omega}{\psi(L)}(1-L)^{-d}. \quad (5.2.32)$$

Chiaramente, la (5.2.31) è molto più simile della (5.2.15) al modello ARFIMA della (3.2.16).

Volendo esplicitare  $h_t$ , l'analogia della (5.2.16) diventa:

$$h_t = \varepsilon_t^2 - \frac{\psi(L)}{1 - \beta(L)}(1-L)^d(\varepsilon_t^2 - \sigma^2). \quad (5.2.33)$$

#### 5.2.4 La stima del modello FIGARCH

Per stimare un modello FIGARCH il modo senz'altro più semplice è quello di utilizzare la massima verosimiglianza approssimata che abbiamo descritto nel paragrafo 4.3.1. In questa circostanza la log-verosimiglianza è data ancora dalla (5.1.39); ovviamente però tra i parametri da stimare comparirà anche  $d$ .

Un problema che merita di essere affrontato è quello dei valori pre-campionari. In genere nelle stime ARCH si suole condizionare a valori pari alla varianza non condizionata del processo. Ma come abbiamo già osservato, per il FIGARCH come per l'IGARCH, la varianza non condizionata non esiste. Possiamo pertanto procedere in due modi: o condizioniamo a valori "eliminati" dalla serie originale, oppure, seguendo la proposta di Baillie, Bollerslev e Mikkelsen (1996), utilizziamo, al posto della varianza non condizionata, la media dei residui al quadrato.

Per quanto riguarda la teoria asintotica, Lee e Hansen (1994) studiano il caso dell'IGARCH (1, 1), e ricavano la  $\sqrt{T}$ -consistenza dello stimatore di quasi-massima verosimiglianza.

Baillie, Bollerslev e Mikkelsen (1996) ricavano che anche lo stimatore di quasi-massima verosimiglianza del FIGARCH (1,  $d$ , 1) è consistente:

$$\sqrt{T}(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}) \xrightarrow{d} N(0, A(\boldsymbol{\theta})^{-1} B(\boldsymbol{\theta}) A(\boldsymbol{\theta})^{-1}), \quad (5.2.34)$$

dove  $A(\boldsymbol{\theta})$  e  $B(\boldsymbol{\theta})$  sono, rispettivamente, l'Hessiana e il prodotto dei gradienti, calcolati in  $\boldsymbol{\theta}$ , che abbiamo definito nella (4.1.23) e nella (4.1.24). Manca ancora, purtroppo, una dimostrazione formale della consistenza per un generico IGARCH ( $r$ ,  $m$ ), per cui non si può ricavare un risultato analogo alla (5.2.34) per il caso FIGARCH ( $r$ ,  $d$ ,  $m$ ).

Il comportamento di questo stimatore è messo alla prova su dati simulati. Baillie, Bollerslev e Mikkelsen (1996) fissano  $M$  a mille e studiano otto differenti specificazioni FIGARCH. In tutti i modelli la distorsione per  $\tilde{d}$  è molto bassa, e raddoppiando la dimensione del campione si riduce all'incirca di  $\sqrt{2}$ , chiaro segnale della  $\sqrt{T}$ -consistenza. Si osserva inoltre che le stime più precise di  $d$  si hanno quando il processo generatore dei dati è un GARCH ( $d = 0$ ) o un IGARCH ( $d = 1$ ); ciò indica che lo stimatore è in grado di distinguere tra i modelli e può quindi essere utilizzato anche senza un'analisi preventiva sulla presenza di integrazione frazionaria.

Stimando invece i dati simulati con un semplice GARCH (1, 1) si ottiene un altro risultato importante: la somma dei coefficienti si avvicina al cerchio unitario; ciò indica che, se si utilizzano esclusivamente dei modelli GARCH, l'integrazione frazionaria può essere scambiata per la presenza di una radice unitaria.

Per quanto riguarda invece la parametrizzazione alternativa di Chung (1999), presentata nella (5.2.30), lo studio su dati simulati ha evidenziato un comportamento leggermente migliore, in termini di efficienza, rispetto al modello "originale" di Baillie, Bollerslev e Mikkelsen (1996), almeno per  $d < 1/2$ . Rimangono però diverse questioni irrisolte. In primo luogo si osserva che gli *standard error* asintotici per  $\sigma^2$  non seguono la formulazione generale della (4.1.19), ma

sono distribuiti diversamente. Quello che però è più preoccupante è che, per valori di  $d$  maggiori di 0.2<sup>15</sup>, interviene una forte distorsione su  $\sigma^2$ ; ciò è in netto contrasto con quanto avviene per il coefficiente  $\omega$  di Baillie, Bollerslev e Mikkelsen (1996).

Vale infine la pena osservare che il metodo utilizzato per la stima è leggermente diverso. Chung (1999) condiziona a valori pre-campionari pari a zero (e ciò è sensato, visto che si modella  $\varepsilon_t^2 - \sigma^2$  piuttosto che  $\varepsilon_t^2$ ) e utilizza, per l'espansione del *fractional difference operator*, non un numero fisso di ritardi, ma tutti quelli disponibili. Chiariamo meglio questo concetto con un esempio: se ci troviamo in  $t = 41$  utilizzeremo 41 ritardi, se invece siamo in  $t = 3276$  ne utilizzeremo non 1000, come suggerivano Baillie, Bollerslev e Mikkelsen (1996), ma esattamente 3276.

Un'ultima questione relativa al gradiente della funzione di verosimiglianza: nei lavori di Baillie, Bollerslev e Mikkelsen (1996) e Bollerslev e Mikkelsen (1996) si sono utilizzate le derivate numeriche; Chung (1999) ha però trovato le espressioni analitiche per le derivate parziali, per quanto riguarda il modello FIGARCH modificato da lui proposto. La derivata della funzione di verosimiglianza di un generico modello ARCH è:

$$\frac{\partial \ln \ell}{\partial \boldsymbol{\theta}} = -\sum_{t=1}^T \frac{\varepsilon_t}{h_t} \frac{\partial \varepsilon_t}{\partial \boldsymbol{\theta}} + \frac{1}{2} \sum_{t=1}^T \frac{\varepsilon_t^2 - h_t}{h_t} \frac{\partial h_t}{\partial \boldsymbol{\theta}}. \quad (5.2.35)$$

Data la serie di derivate dei coefficienti della (3.2.6):

$$\frac{\partial \varphi_k}{\partial d} = -\varphi_k \sum_{j=1}^k \frac{1}{k - j - d}, \quad (5.2.36)$$

valgono, considerando la sola costante  $\mu$  nell'equazione della media, le:

---

<sup>15</sup> Si può sospettare che ciò sia dovuto al fatto che la varianza campionaria di un processo ARFIMA è  $\sqrt{T}$  – consistente solo per  $d < 1/4$ , come abbiamo osservato nel paragrafo 3.2.3.

$$\begin{aligned}
\frac{\partial \varepsilon_t}{\partial \mu} &= -1, \\
\frac{\partial h_t}{\partial \mu} &= \left[ 1 - \frac{\psi(L)(1-L)^d}{1-\beta(L)} \right] \cdot 2\varepsilon_t \frac{\partial \varepsilon_t}{\partial \mu}, \\
\frac{\partial h_t}{\partial \sigma^2} &= \frac{\psi(1)}{1-\beta(1)} \sum_{k=0}^{t-1} \varphi_k, \\
\frac{\partial h_t}{\partial \beta_j} &= \frac{h_{t-j} - \varepsilon_{t-j}^2}{1-\beta(L)}, \quad j=1,2,\dots,p, \\
\frac{\partial h_t}{\partial \psi_j} &= \frac{1}{1-\beta(L)} \sum_{k=0}^{t-1} \varphi_k (\varepsilon_{t-k-j}^2 - \sigma^2), \quad j=1,2,\dots,q, \\
\frac{\partial h_t}{\partial d} &= -\frac{\psi(L)}{1-\beta(L)} \sum_{k=1}^{t-1} \frac{\partial \varphi_k}{\partial d} (\varepsilon_{t-k-j}^2 - \sigma^2).
\end{aligned} \tag{5.2.37}$$

A partire dalle (5.2.37) è piuttosto facile<sup>16</sup> ricavare le derivate analitiche per quanto riguarda la parametrizzazione “originale” di Baillie, Bollerslev e Mikkelsen (1996):

$$\begin{aligned}
\frac{\partial \varepsilon_t}{\partial \mu} &= -1, \\
\frac{\partial h_t}{\partial \mu} &= \left[ 1 - \frac{\psi(L)(1-L)^d}{1-\beta(L)} \right] \cdot 2\varepsilon_t \frac{\partial \varepsilon_t}{\partial \mu}, \\
\frac{\partial h_t}{\partial \omega} &= \frac{1}{1-\beta(1)}, \\
\frac{\partial h_t}{\partial \beta_j} &= \frac{h_{t-j} - \varepsilon_{t-j}^2}{1-\beta(L)}, \quad j=1,2,\dots,p, \\
\frac{\partial h_t}{\partial \psi_j} &= \frac{1}{1-\beta(L)} \sum_{k=0}^M \varphi_k \varepsilon_{t-k-j}^2, \quad j=1,2,\dots,q, \\
\frac{\partial h_t}{\partial d} &= -\frac{\psi(L)}{1-\beta(L)} \sum_{k=1}^M \frac{\partial \varphi_k}{\partial d} \varepsilon_{t-k-j}^2.
\end{aligned} \tag{5.2.38}$$

Per quanto riguarda infine la presenza di un’eventuale dummy  $D_t$  nella (5.2.15), vale semplicemente:

$$\frac{\partial h_t}{\partial \delta} = \frac{D_t}{1-\beta(L)}. \tag{5.2.39}$$



### 5.2.5 Le applicazioni ai tassi di cambio

La prima applicazione a dati reali del modello FIGARCH è dovuta a Baillie, Bollerslev e Mikkelsen (1996), che stimano i rendimenti giornalieri percentuali del tasso di cambio DM/\$, dal 1979 al 1992, attraverso quattro diverse specificazioni: un GARCH (1, 1), un IGARCH (1,1) e due diversi FIGARCH. I risultati sono presentati nella tabella 5.1:

<i>FIGARCH</i> ( <i>p, d, q</i> )	(1, 0, 1)	(1, 1, 0)	(1, <i>d</i> , 1)	(1, <i>d</i> , 0)
$\omega$	0.016 (0.004)	0.01 (0.003)	0.018 (0.006)	0.017 (0.005)
$\beta_1$	0.867 (0.015)	0.872 (0.015)	0.675 (0.104)	0.762 (0.061)
$\psi_1$	0.98 (0.01)	—	0.094 (0.075)	—
$d$	—	—	0.652 (0.16)	0.823 (0.077)
$Q_{20}$	34.17	36.69	35.52	35.68

**Tabella 5.1**

*Stima QML di vari modelli FIGARCH per i rendimenti giornalieri del tasso di cambio DM/\$;  
tra parentesi i valori degli standard error  
(da Baillie, Bollerslev e Mikkelsen, 1996)*

Osserviamo in primo luogo come i risultati GARCH e IGARCH siano molto simili; in una situazione “tradizionale” avremmo concluso che la serie della varianza condizionata è I(1). In realtà, procedendo ed esaminando i risultati della stima FIGARCH, si osserva che i risultati sono sensibilmente più precisi. Ciò conferma quanto avevamo già osservato, nel paragrafo precedente, a proposito della possibile confusione dell’integrazione frazionaria con l’integrazione di ordine uno. Si osserva inoltre che, nel caso del FIGARCH (1, *d*, 1), il coefficiente GARCH del FIGARCH non è significativo, segno che le dipendenze sono catturate dall’integrazione frazionaria; difatti, eliminandolo (ultima colonna) si ha un incremento di  $\tilde{d}$  e una riduzione della sua deviazione standard. La modellazione migliore sembra dunque il FIGARCH (1, *d*, 0). È interessante infine osservare come la stima  $\tilde{d} = 0.823$  sia in accordo coi risultati ottenuti da Harvey (1993), che stimando la volatilità del

<sup>16</sup> Tali derivate sono state ricavate dall’autore di questo lavoro; un’ulteriore estensione nella quale viene presentata la matrice Hessiana analitica del modello è introdotta da Lombardi e Gallo (2002).

medesimo tasso di cambio (sul periodo più ristretto 1981-1985) con un semplice *fractional white noise* otteneva  $\hat{d} = 0.868$ .

Beine, Laurent e Lecourt (1999) stimano una serie di tassi di cambio utilizzando diverse modellazioni FIGARCH con residui t-distribuiti e con una dummy per effetti week-end. Entrambi questi effetti appaiono significativi, anche se bisogna notare che utilizzare la t di Student, pur migliorando il *fit* complessivo del modello, non modifica sostanzialmente le stime dei parametri; ciò conferma i risultati di Baillie, Bollerslev e Mikkelsen (1996) relativamente alla performance della QMLE per il FIGARCH. A differenza di quanto osservato da Baillie, Bollerslev e Mikkelsen (1996) il modello migliore sembra essere il FIGARCH (1,  $d$ , 1); ciò può essere dovuto sia all'utilizzo delle dummy che alla scelta dei valori iniziali per la routine di massimizzazione. Si osserva infatti che il profilo in  $d$  della funzione di verosimiglianza ha un massimo locale per  $\tilde{d} = 0.87$ , cioè un valore molto vicino a quello trovato da Baillie, Bollerslev e Mikkelsen (1996).

Il valore di  $d$  stimato da Beine, Laurent e Lecourt (1999) risulta essere, per tutti i tassi considerati, intorno a 0.43, cioè sensibilmente al di sotto del valore ottenuto da Baillie, Bollerslev e Mikkelsen (1996). Ciò indica che l'esclusione del ritardo GARCH aumenta la dipendenza di lungo periodo, avvicinando il modello alla classe IGARCH; i risultati di Beine, Laurent e Lecourt (1999) indicano invece che la dipendenza di lungo periodo è meno forte, e che il modello è più vicino a un GARCH tradizionale.

Beine e Laurent (1999) considerano la possibilità che vi sia memoria lunga non solo nella varianza condizionata, ma anche nei livelli del tasso di cambio: si introduce un modello a doppia memoria lunga ARFIMA-FIGARCH (per il quale si può vedere anche Teyssi re, 1997), che viene applicato ai tassi di cambio di FF, DM,   e   rispetto al \$, e si conduce uno studio sulla distribuzione dei residui. L'uso della t di Student non riesce a liberare i residui da un eccesso di skewness, anche a causa della presenza di numerosi valori "anomali". Tali

anomalie possono essere causate da interventi delle banche centrali o da accordi internazionali di svalutazione o rivalutazione; una prima idea potrebbe essere di inserire nell'equazione della varianza condizionata delle dummy che catturino questi effetti.

Questo studio è condotto da Beine, Benassy-Quere e Lecourt (1999), che utilizzano, al posto di semplici dummy, gli importi degli interventi ufficiali, osservando che essi incrementano piuttosto che diminuire la volatilità. Sono presentati altresì una varietà di risultati che distinguono i tipi di intervento (segreto, palese, concordato); non ci occuperemo tuttavia di queste problematiche, che esulano dai limiti naturali di questo lavoro.

Un metodo più raffinato per spiegare la presenza di valori anomali (Beine e Laurent, 1999) è quello di utilizzare distribuzioni per i residui che contemplino la possibilità di “salti”: si osserva che la distribuzione che sembra più adeguata è una mista normale-bernoulliana con probabilità di salto variabili nel tempo.

### 5.2.6 Il modello FIEGARCH e i mercati azionari

Il primo studio che ha applicato la famiglia di modelli che stiamo trattando allo studio dei mercati azionari è dovuto a Bollerslev e Mikkelsen (1996), che introducono il modello FIEGARCH semplicemente applicando il *fractional difference operator* alla (5.1.38):

$$\ln h_t = \omega + \frac{1 + \psi(L)}{\phi(L)} (1 - L)^{-d} g \left( \frac{\varepsilon_{t-1}}{\sqrt{h_{t-1}}} \right). \quad (5.2.40)$$

La (5.2.40) viene applicata alla stima dei rendimenti giornalieri dell'indice Standard & Poor 500 tra il 1953 e il 1990, filtrati da un AR (3) con dummy  $D_t$  (coefficiente  $\delta$ ) per catturare gli effetti week-end. I risultati sono presentati nella tabella 5.2; le ultime righe indicano il valore della statistica  $Q$  di Ljung e Box (1979), calcolata per 10 e 100 ritardi, prima sui residui standardizzati, poi sui residui standardizzati assoluti.

	<i>AR</i> (3)	<i>AR</i> (3) <i>EGARCH</i> (2, 1)	<i>AR</i> (3) <i>IEGARCH</i> (2, 1)	<i>AR</i> (3) <i>FIEGARCH</i> (2, <i>d</i> , 1)
$\mu$	0.00004 (0.0009)	0.00003 (0.00007)	0.00003 (0.00008)	0.00003 (0.00007)
$\varphi_1$	0.12 (0.032)	0.186 (0.011)	0.185 (0.012)	0.184 (0.011)
$\varphi_2$	-0.068 (0.028)	-0.056 (0.012)	-0.053 (0.01)	-0.057 (0.012)
$\varphi_3$	0.0002 (0.024)	0.019 (0.011)	0.020 (0.009)	0.021 (0.011)
$\omega$	-9.696 (0.032)	-10.117 (0.312)	-10.997 (0.413)	-10.273 (0.414)
$\delta$	0.372 (0.199)	0.212 (0.057)	0.214 (0.057)	0.217 (0.058)
$\theta$	—	-0.103 (0.021)	-0.098 (0.02)	-0.118 (0.023)
$\gamma$	—	0.203 (0.033)	0.197 (0.033)	0.231 (0.03)
$\beta_1$	—	-0.948 (0.036)	-0.974 (0.016)	-0.717 (0.142)
$\psi_1$	—	0.997 (0.002)	—	0.774 (0.124)
$\psi_2$	—	0.838 (0.092)	0.894 (0.053)	—
$D$	—	—	—	0.633 (0.063)
<i>AIC</i>	-64351	-67434	-67426	-67460
<i>SC</i>	-64308	-67355	-67354	-67381
$Q_{10}$	15.8	9.4	9.9	10
$Q_{100}$	176.4	99.6	101.1	100.5
$Q_{[10]}$	3877	18.7	18.5	15.1
$Q_{[100]}$	12622	133.2	132.7	122.7

**Tabella 5.2**

*Stima QML di vari modelli FIEGARCH per i rendimenti giornalieri dell'indice S&P 500;  
tra parentesi i valori degli standard error  
(da Bollerslev e Mikkelsen, 1996)*

Come si può osservare, il modello FIEGARCH è quello che ha la performance migliore, sia in termini dei criteri informativi, sia per quanto concerne il test  $Q$ , specie se si considerano i residui assoluti.

### 5.2.7 Il prezzaggio delle opzioni

Bollerslev e Mikkelsen (1996) procedono applicando il modello FIEGARCH al prezzaggio<sup>17</sup> di opzioni *call* simulate secondo i modelli di Black e Scholes (1973), Hull e White (1987) e secondo il metodo del valore attuale scontato. Si genera la distribuzione attesa, a  $T + 70$  e  $T + 260$  giorni, del valore dell'indice Standard & Poor 500 ( $P_T = 330.2$ ) e si considerano sia opzioni *at the money* ( $K = P_T$ ) che *out of the money* ( $K = 1.25P_T$ ). I risultati sono presentati nella tabella 5.3:

	<i>AR</i> (3)	<i>AR</i> (3) <i>EGARCH</i> (2, 1)	<i>AR</i> (3) <i>IEGARCH</i> (2, 1)	<i>AR</i> (3) <i>FIEGARCH</i> (2, 1)
$C_{BS}(70, P_T)$	13.6	16.2	17.5	16.2
$C_{HW}(70, P_T)$	12.6	14.0	15.1	14.1
$C_{VA}(70, P_T)$	15.6	17.5	19.1	17.5
$C_{BS}(260, P_T)$	33.2	37.5	41.8	38.6
$C_{HW}(260, P_T)$	31.4	33.2	36.8	34.2
$C_{VA}(260, P_T)$	39.8	40.9	46.1	41.6
$C_{BS}(260, 1.25P_T)$	4.9	8.5	12.4	9.5
$C_{HW}(260, 1.25P_T)$	3.8	5.3	8.2	6.0
$C_{VA}(260, 1.25P_T)$	5.8	6.4	10.3	7.3

**Tabella 5.3**

*Prezzi di opzioni call simulati sulla base dei modelli della tabella 5.2  
(da Bollerslev e Mikkelsen, 1996)*

Si può osservare come il modello FIEGARCH dia risultati in linea con quelli dell'EGARCH per  $t = 70$ , mentre tenda a discostarsi, conformemente alle sue qualità di memoria lunga, all'aumentare dell'orizzonte temporale. Ciò lo rende senz'altro molto versatile e preferibile all'IEGARCH, che dà invece, specie per le opzioni *out of the money*, valori eccessivi.

<sup>17</sup> Quello del prezzaggio delle opzioni è un argomento estremamente complesso e richiede, per essere trattato in maniera approfondita, l'utilizzo di nozioni matematiche avanzate. Si rimanda, per una trattazione completa, a Geronazzo (1998) o Neftci (1996).

### 5.2.8 Le estensioni più recenti

Il modello FIGARCH è piuttosto recente, e la maggior parte delle estensioni teoriche sono ancora “in cantiere”.

Al di là del già citato FIEGARCH di Bollerslev e Mikkelsen (1996), un'altra possibilità per introdurre nel FIGARCH l'effetto *leverage*, stavolta attraverso l'uso di una dummy, è considerata da Hsieh, Chung e Lin (1999)

Tse (1998) introduce invece il modello FIAPARCH, applicando il *fractional difference operator* alla (5.2.3):

$$\begin{cases} s_t^\delta = \bar{\omega} + \left[ 1 - \frac{1-\psi L}{1-\beta L} (1-L)^d \right] (|\varepsilon_t| - \gamma \varepsilon_t)^\delta \\ \varepsilon_t = u_t s_t, \quad N(0,1). \end{cases} \quad (5.2.42)$$

Un'interessante estensione, nella quale il *fractional difference operator* è rimpiazzato dalla produttoria di più polinomi di Gegenbauer, è stata proposta da Guégan (2001). Un modello di questo tipo permette di avere, oltre alla memoria lunga, più componenti cicliche, di modo che la funzione di autocorrelazione dei disturbi al quadrato presenti un decadimento iperbolico-sinusoidale.

Citiamo infine uno studio su FIGARCH multivariato e cointegrazione frazionaria: Brunetti e Gilbert (1999) applicano al FIGARCH i risultati di Hosoya (1996) sulla stima di quasi-massima verosimiglianza di modelli bivariati a memoria lunga, studiando due serie di prezzi del greggio e individuandone la cointegrazione.

### 5.3 Il modello HYGARCH

Il paradosso concernente i rapporti tra modelli GARCH a integrazione frazionaria e modelli GARCH integrati è stato esaminato in un recente lavoro di Davidson (2002). Si mostra in particolare che memoria e ampiezza sono due caratteristiche distinte: la memoria indica quanto tempo impiegano gli shock ad essere riassorbiti, ed è misurata dalla velocità di decadimento a zero dei coefficienti della rappresentazione ARCH ( $\infty$ ); l'ampiezza indica invece la possibile dimensione degli shock alla varianza condizionata, e determinando l'esistenza dei momenti superiori e, di conseguenza, la stazionarietà in covarianza. In particolare, la condizione

$$S = \sum_{i=1}^{\infty} \lambda_i < 1, \quad (5.3.1)$$

dove  $\lambda_i$  sono i coefficienti della rappresentazione ARCH ( $\infty$ ), risulta essere necessaria e sufficiente per la stazionarietà in covarianza<sup>18</sup>. La proprietà di memoria è quindi del tutto estranea alle questioni relative alla stazionarietà in covarianza: ciò spiega perché un modello a memoria corta (IGARCH) e uno a memoria lunga (FIGARCH) risultano entrambi non stazionari.

Abbiamo rilevato in precedenza come il modello FIGARCH sia nato come generalizzazione dell'IGARCH. Quanto detto poc'anzi relativamente alla stazionarietà in covarianza sembra confermare questa idea. Tuttavia apparrebbe paradossale che un modello a memoria lunga finisse per trasformarsi in un modello a memoria corta per  $d = 1$ . In realtà il fatto è che, contrariamente a quanto avviene per i processi per la media, la memoria dei modelli FIGARCH cresce all'avvicinarsi di  $d$  a zero.

Si propone quindi una nuova cornice all'interno della quale inscrivere sia il modello FIGARCH che i tradizionali GARCH e IGARCH.

L'idea è quella di utilizzare un ulteriore parametro  $\alpha \geq 0$  e definire i coefficienti della rappresentazione ARCH ( $\infty$ ) come:

$$\lambda(L) = 1 - \frac{\psi(L)}{1 - \beta(L)} \{1 + \alpha[(1 - L)^d - 1]\}. \quad (5.3.2)$$

Il modello risultante, che per  $\alpha = 1$  coincide col FIGARCH, per  $\alpha = 0$  col GARCH e per  $\alpha = 1$  e  $d = 1$  con l'IGARCH, prende il nome di HYGARCH, ovvero Hyperbolic GARCH.

Il modello è messo alla prova su dati relativi a tassi di cambio. Se per le valute “tradizionali” si ricavano risultati sostanzialmente non dissimili da quelli che si sarebbero ottenuti vincolando  $\alpha$  a 1 e stimando quindi un semplice modello FIGARCH, il modello HYGARCH si dimostra utile nell'analisi dei tassi di cambio dei paesi coinvolti nella crisi asiatica del 1997. Come prima ipotesi, si potrebbe ritenere che una serie storica formata da dati pre e post crisi sia caratterizzata da un cambiamento di regime. Se applichiamo ai dati il modello HYGARCH si osserva però che le stime dei coefficienti risultano stabili, manifestando un livello di ampiezza assai superiore a quello previsto dal tradizionale FIGARCH.

Una delle caratteristiche interessanti del modello è che riesce a riprodurre la variazione di regime in maniera endogena. L'esempio riportato nell'articolo citato è particolarmente illustrativo: si simula una serie di dati utilizzando i coefficienti stimati sui dati del Won Coreano, si inserisce la “crisi”, ovvero, cinque osservazioni anomale (e di segno uguale) consecutive, e quello che ne risulta è una serie dall'andamento del tutto simile a quello dei dati reali. Il comportamento della serie successivo alla crisi è pertanto determinato unicamente da fattori endogeni, senza dover necessariamente ipotizzare un cambiamento di regime esogeno.

---

<sup>18</sup> Si noti difatti che i modelli IGARCH e FIGARCH non la soddisfano.



# Capitolo 6

## Un'applicazione empirica

In questo capitolo sarà presentata un'applicazione empirica del modello FIGARCH che abbiamo esaminato nel capitolo precedente. Dopo aver presentato i dati e averli analizzati tramite gli strumenti GARCH tradizionali, passeremo alla stima del modello FIGARCH, che verrà condotta sia col metodo proposto da Baillie, Bollerslev e Mikkelsen (1996), sia con quello di Chung (1999). Si utilizzeranno quindi i risultati di stima ottenuti per prevedere la volatilità e calcolare il *Value at Risk* di un ipotetico portafoglio.

## 6.1 Un confronto su dati reali

### 6.1.1 *Analisi preliminare dei dati*<sup>1</sup>

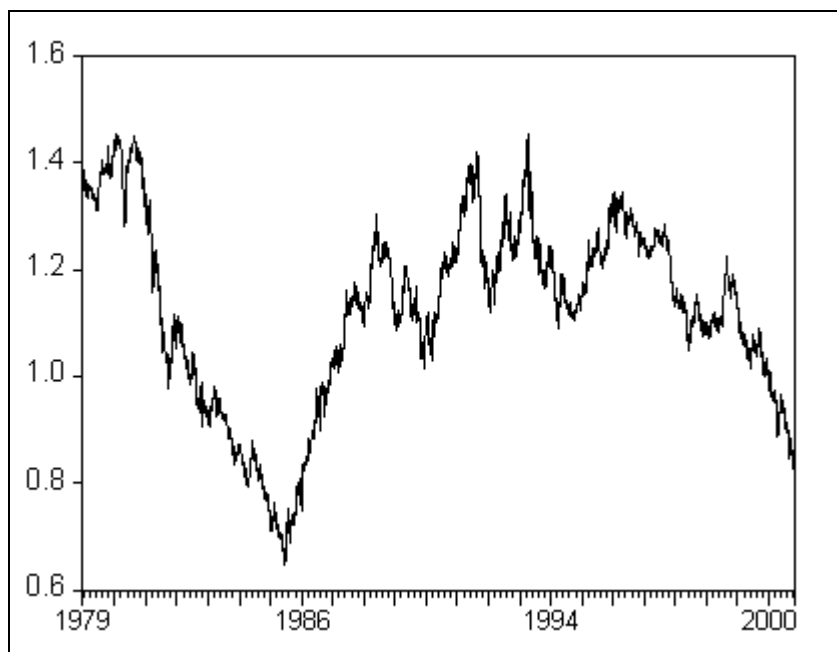
Il tipo di dati che utilizzeremo è il tasso spot medio giornaliero \$/€ rilevato dalla Federal Reserve sul mercato di New York, tra il 28 dicembre 1978 e l'8 novembre 2000, per un totale di 5705 osservazioni. Per quanto riguarda le osservazioni anteriori al 1 gennaio 1999, data di nascita della moneta unica europea, si sono utilizzati i dati relativi all'ECU (che alla nascita dell'euro è stato convertito con quest'ultimo alla pari). A dire il vero questa scelta è discutibile, poiché, com'è noto, la sterlina inglese faceva parte dell'ECU (con un peso dell'11%), mentre non fa parte dell'euro. Tuttavia riteniamo che l'alternativa a questa impostazione, cioè di usare il marco tedesco per i dati anteriori al 1 gennaio 1999, sia senz'altro meno adeguata. Difatti l'utilizzare, per il periodo anteriore alla nascita della moneta unica, un'unica valuta “forte”, avrebbe potuto determinare comportamento anomalo della serie. L'ECU invece, pur avendo una composizione leggermente diversa rispetto all'euro, rappresenta pur sempre un paniere in cui compaiono anche divise “deboli”.

La pratica comune del nostro paese è sempre stata quella di considerare il tasso di cambio come quantità di moneta nazionale necessaria per acquistare un'unità di moneta estera; notiamo tuttavia che dall'introduzione dell'euro si è diffusa la convenzione contraria, pertanto ci atterremo a questa nuova impostazione. Il tempo è misurato in settimane di cinque giorni; per quanto riguarda i dati mancanti per cause di festività, si sono riempiti tali “buchi” coi valori della seduta precedente<sup>2</sup>.

---

<sup>1</sup> Le stime presentate in questo paragrafo sono state condotte con l'applicativo Econometric Views (Eviews).

<sup>2</sup> In realtà anche questa assunzione è tutt'altro che pacifica: volendo essere più precisi, tali dati mancanti dovrebbero essere stimati; per uno studio su tale possibilità si rimanda a Maravall e Pena (1992).

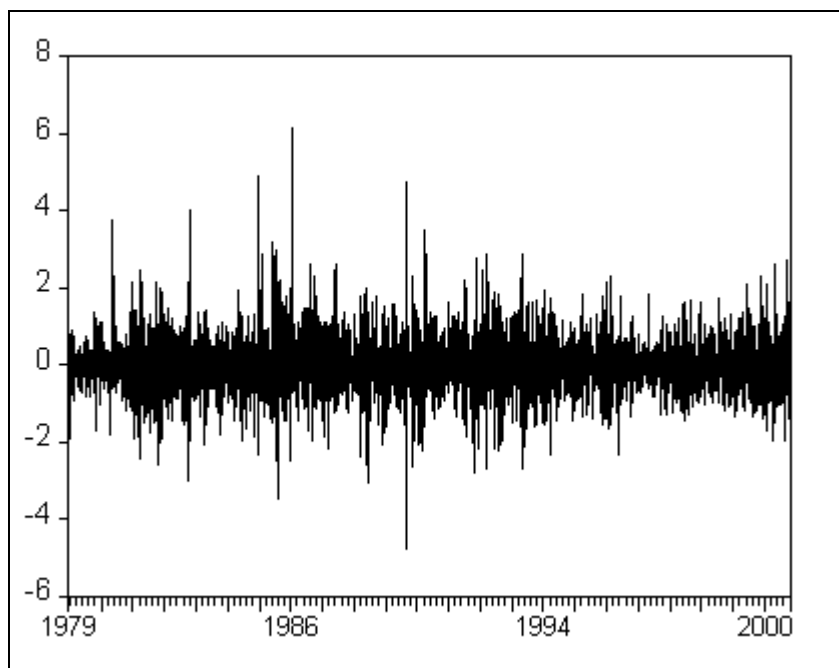


**Figura 6.1**

*Tasso di cambio spot giornaliero \$/€*

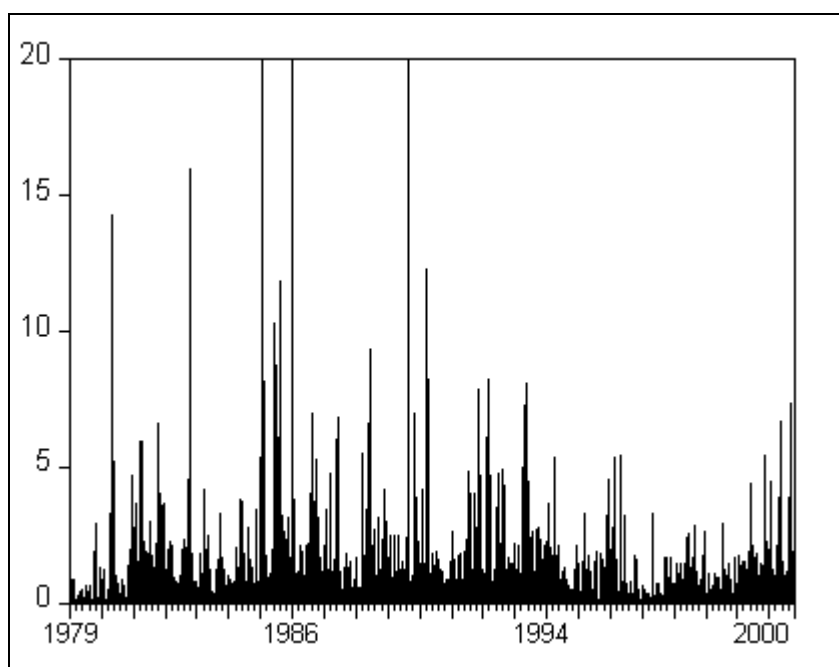
Dalla figura 6.1 possiamo riconoscere gli eventi principali che hanno caratterizzato l'andamento del tasso di cambio delle valute europee contro il dollaro: il rafforzamento della divisa statunitense nella prima metà degli anni '80, il successivo indebolimento a seguito degli accordi del Plaza (22 settembre 1985), nonché la fase di debolezza attraversata dalla moneta europea dopo la sua introduzione.

Nella figure 6.2 e 6.3 vediamo rispettivamente l'andamento dei rendimenti e dei rendimenti al quadrato del tasso di cambio, calcolati come differenza logaritmica del tasso di cambio stesso e riportati su base percentuale.



**Figura 6.2**

*Rendimenti giornalieri percentuali del tasso di cambio \$/€*



**Figura 6.3**

*Rendimenti giornalieri percentuali al quadrato del tasso di cambio \$/€*

Una delle assunzioni di base circa la volatilità dei mercati finanziari è che essa sia generata dall'arrivo continuo di nuove informazioni, che obbligano gli operatori a formulare nuove previsioni, ripensare le proprie strategie e, in ultima analisi, a modificare la propria posizione. In questo contesto è ragionevole supporre (si veda ad esempio Baillie e Bollerslev, 1989) che la volatilità debba essere mediamente più alta nei giorni di riapertura del mercato dopo una chiusura, poiché vi sono più informazioni “accumulate” che devono essere scontate. Il primo passo che abbiamo compiuto nella nostra analisi dei dati è stato quindi quello di generare una variabile *dummy*  $\{D_t\}$ , che assume valore 1 nei lunedì e nei giorni successivi a un giorno di chiusura e 0 negli altri casi, al fine di catturare questi effetti.

Si è passati quindi ad analizzare i dati relativi ai rendimenti attraverso un modello ARMA (1, 1), al fine di identificare eventuali fenomeni di autocorrelazione. I risultati di tale stima sono presentati nella tabella 6.1.

<i>Variabile</i>	<i>Coefficiente</i>	<i>Standard error</i>	<i>Statistica t</i>	<i>P-value</i>
$\mu$	- 0.008358	0.009995	- 0.836220	0.4031
$D$	0.000063	0.020737	0.003044	0.9976
$AR(1)$	0.085824	0.416846	0.205890	0.8369
$MA(1)$	- 0.117265	0.415508	- 0.282221	0.7778
<i>Statistica F</i>			1.892119	
<i>P-value</i>			0.128592	
<i>Log-verosimiglianza</i>			- 5822.575	
<i>AIC</i>			2.042979	
<i>SC</i>			2.047641	

**Tabella 6.1**

*Stima ARMA (1, 1) sui dati della figura 6.2*

È immediato osservare l'assenza di alcun tipo di dinamica nei livelli della serie: nessuno dei coefficienti è significativo, neppure la costante. Forti sono invece gli indizi di effetti ARCH (visibili già dai “grappoli” di volatilità presenti nelle figure 6.2 e 6.3): nella tabella 6.2 presentiamo i risultati del test Q di Ljung e Box (1979) (si veda il paragrafo A.1.2 in appendice) sui residui al quadrato.

<i>Ritardi</i>	<i>AC</i>	<i>PAC</i>	<i>Statistica Q</i>	<i>P-value</i>
1	0.162	0.162	149.48	/
2	0.045	0.020	161.19	/
3	0.105	0.098	224.58	0.000
4	0.078	0.047	259.01	0.000
5	0.051	0.028	273.82	0.000
6	0.096	0.075	326.63	0.000
7	0.082	0.046	364.71	0.000
8	0.050	0.020	378.9	0.000
9	0.033	0.003	385.27	0.000
10	0.044	0.018	396.27	0.000
11	0.096	0.073	449.23	0.000
12	0.022	- 0.020	451.97	0.000
13	0.039	0.020	460.44	0.000
14	0.065	0.034	484.90	0.000
15	0.045	0.016	496.38	0.000
16	0.029	0.006	501.27	0.000
17	0.030	- 0.001	506.41	0.000
18	0.043	0.020	517.15	0.000
19	0.041	0.018	526.77	0.000
20	0.046	0.022	538.80	0.000

**Tabella 6.2**

*Test Q per i residui al quadrato della regressione della tabella 6.1*

Il test ARCH, difatti, dà risultati piuttosto espliciti:

<i>Variabile</i>	<i>Coefficiente</i>	<i>Standard error</i>	<i>Statistica t</i>	<i>P-value</i>
<i>C</i>	0.378071	0.016364	23.10378	0.0000
<i>Res<sup>2</sup> (-1)</i>	0.161846	0.013070	12.38348	0.0000
<i>Statistica F</i>			153.3506	
<i>P- value</i>			0.0000	
<i>TR<sup>2</sup></i>			149.3861	
<i>P-value</i>			0.0000	

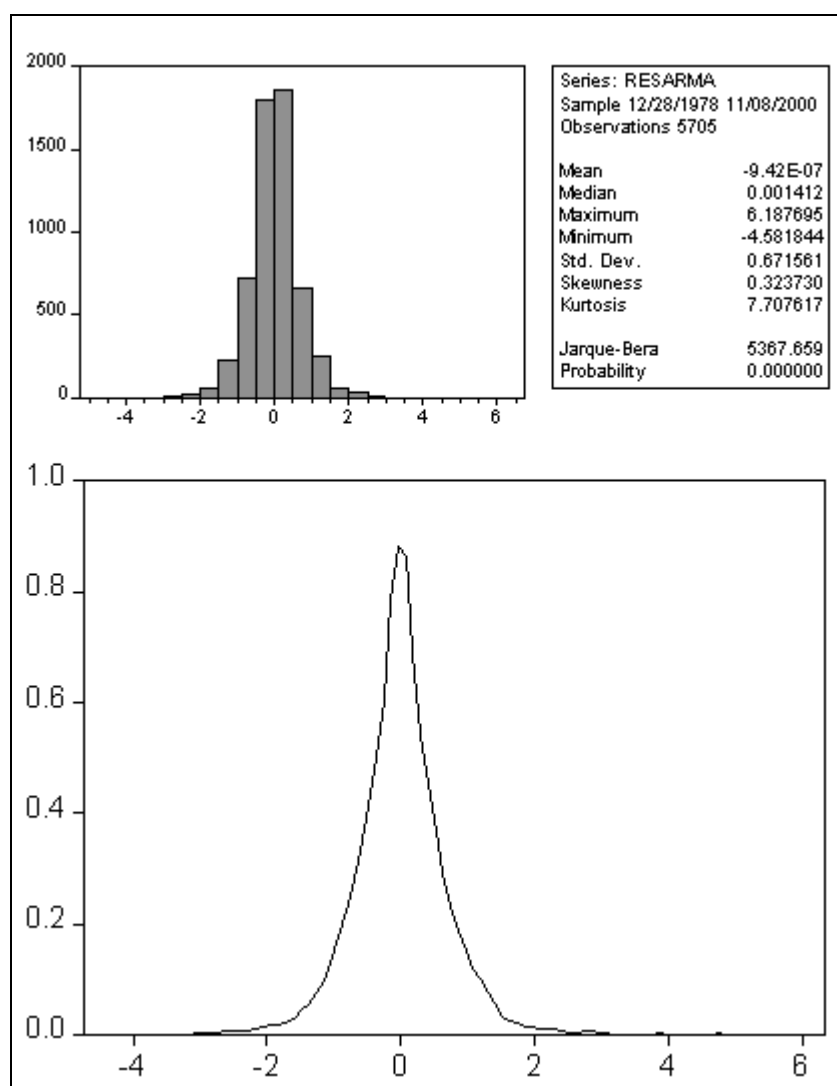
**Tabella 6.3**

*Test ARCH-LM sui residui della tabella 6.1*

Decidiamo pertanto di procedere escludendo dalla stima qualsiasi tipo di effetto sui livelli, a parte la costante, e di concentrarci sulla dinamica della varianza condizionata.

Osserviamo prima di procedere che i residui della tabella 6.1 presentano la caratteristica di non normalità che ci attendevamo: in particolare si rileva un leggero eccesso di skewness (0.323702) e un marcato eccesso di curtosi (7.706266), tanto che la statistica di *JB* di Jarque e Bera (1982) (si veda il paragrafo A.1.2 in appendice) vale addirittura 5363.678 (si veda la figura 6.4); in questo contesto è opportuno procedere utilizzando la stima di quasi-massima

verosimiglianza che abbiamo introdotto, per quanto riguarda i modelli GARCH, nel paragrafo 5.1.5.



**Figura 6.4**

*Statistiche di normalità e stima kernel uniforme della densità dei residui della tabella 6.1, ampiezza di banda 0.1438*

I risultati della stima di un semplice GARCH (1, 1) sono presentati nella tabella 6.4:

<i>Variabile</i>	<i>Coefficiente</i>	<i>Standard error</i>	<i>Statistica z</i>	<i>P-value</i>
$\mu$	- 0.002834	0.008027	- 0.353017	0.7241
$\omega$	- 0.002946	0.003950	- 0.745826	0.4558
$D$	0.056162	0.017710	3.171125	0.0015
$\alpha$	0.078328	0.012060	6.495110	0.0000
$\beta$	0.899728	0.014246	63.15864	0.0000
<i>Log-verosimiglianza</i>			- 5442.729	
<i>AIC</i>			1.909809	
<i>SC</i>			1.915636	

**Tabella 6.4**

*Stima QML GARCH (1, 1) sui dati della figura 6.2*

Come ci si aspettava, la *dummy* ha coefficiente positivo ed è significativa; otteniamo tuttavia una costante negativa (che implicherebbe una varianza non condizionata minore di zero, situazione evidentemente assurda). Ciò è senz'altro dovuto alla grossa forza dell'effetto lunedì; decidiamo pertanto di utilizzare la *dummy* al contrario, cioè come fattore correttivo (stavolta in negativo) per gli altri giorni della settimana. I risultati sono presentati nella tabella 6.5:

<i>Variabile</i>	<i>Coefficiente</i>	<i>Standard error</i>	<i>Statistica z</i>	<i>P-value</i>
$\mu$	- 0.002866	0.008027	- 0.356991	0.7211
$\omega$	0.053240	0.014424	3.690968	0.0002
$1-D$	- 0.056217	0.017714	- 3.173698	0.0015
$\alpha$	0.078403	0.012066	6.497942	0.0000
$\beta$	0.899692	0.014250	63.13835	0.0000
<i>Log-verosimiglianza</i>			- 5442.73	
<i>AIC</i>			1.909809	
<i>SC</i>			1.915636	

**Tabella 6.5**

*Stima QML GARCH (1, 1) sui dati della figura 6.2, dummy “al contrario”*

Come si può osservare dalla tabella 6.6, l'applicazione del filtro GARCH dà buoni risultati anche sul correlogramma dei residui, che appare adeguatamente “ripulito” dalle autocorrelazioni.



<i>Ritardi</i>	<i>AC</i>	<i>PAC</i>	<i>Statistica Q</i>	<i>P-value</i>
1	-0.005	-0.005	0.1647	0.685
2	-0.003	-0.003	0.2174	0.897
3	0.025	0.025	3.7624	0.288
4	0.020	0.020	6.0869	0.193
5	0.017	0.018	7.8318	0.166
6	-0.013	-0.014	8.8396	0.183
7	0.004	0.003	8.9253	0.258
8	0.002	0.001	8.9486	0.347
9	0.027	0.027	13.072	0.159
10	0.021	0.022	15.658	0.110
11	0.015	0.016	16.956	0.109
12	-0.008	-0.010	17.335	0.137
13	-0.001	-0.004	17.346	0.184
14	0.009	0.006	17.792	0.216
15	0.021	0.021	20.356	0.159
16	0.011	0.011	21.005	0.178
17	-0.013	-0.012	21.936	0.187
18	-0.016	-0.019	23.493	0.172
19	0.007	0.004	23.775	0.205
20	0.021	0.019	26.182	0.160

**Tabella 6.6**

*Test Q per i residui del modello della tabella 6.5*

Tuttavia, se effettuiamo un nuovo test ARCH, ci rendiamo conto che vi sono ancora effetti non spiegati:

<i>Variabile</i>	<i>Coefficiente</i>	<i>Standard error</i>	<i>Statistica t</i>	<i>P-value</i>
<i>C</i>	0.950405	0.037059	25.64589	0.0000
<i>Stdres<sup>2</sup> (-1)</i>	0.050344	0.023671	2.126805	0.0335
<i>Statistica F</i>			14.48846	
<i>P-value</i>			0.000142	
<i>TR<sup>2</sup></i>			14.45681	
<i>P-value</i>			0.000143	

**Tabella 6.7**

*Test ARCH-LM sui residui standardizzati della tabella 6.5*

Provando a inserire ulteriori termini ARCH e GARCH, si osserva che né il terzo termine GARCH né il secondo ARCH sono significativi; la specificazione migliore, indicata anche dai criteri di Akaike (1973) e di Schwarz (1978) sembra quindi essere un GARCH (2, 1). Con questa parametrizzazione le dinamiche della varianza condizionata vengono colte appieno, e il test ARCH non rifiuta l'ipotesi nulla ( $P\text{-value} = 0.575$ ) I risultati della stima sono presentati nella tabella 6.8:

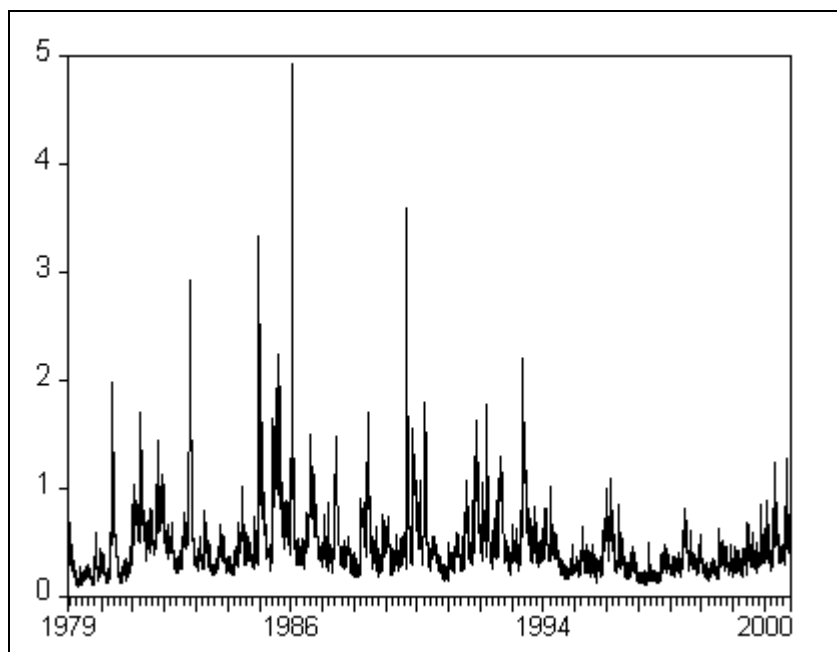
<i>Variabile</i>	<i>Coefficiente</i>	<i>Standard error</i>	<i>Statistica z</i>	<i>P-value</i>
$\mu$	- 0.003920	0.007926	- 0.494630	0.6209
$\omega$	0.071953	0.015949	4.511444	0.0000
$1-D$	- 0.074833	0.019965	- 3.748154	0.0002
$\alpha$	0.108756	0.016727	6.501898	0.0000
$\beta_1$	0.397078	0.140740	2.821370	0.0048
$\beta_2$	0.462078	0.135284	3.415599	0.0006
<i>Log-verosimiglianza</i>			- 5435.074	
<i>AIC</i>			1.907475	
<i>SC</i>			1.914468	

**Tabella 6.8**

*Stima QML GARCH (2, 1) sui dati della figura 6.2*

Osserviamo che la somma dei coefficienti ( $\alpha + \beta_1 + \beta_2 = 0.967912$ ) è molto vicina a uno; se proviamo a imporre il vincolo  $\alpha + \beta_1 + \beta_2 = 1$  (cioè che il processo generatore dei dati sia di fatto un IGARCH) otteniamo  $F = 9.5265$ , a cui è associato un *P-value* di 0.002035. Siamo ancora nella regione di rifiuto, ma si capisce bene che, su serie più corte o con stime meno precise, si poteva facilmente essere tratti in inganno.

Nella figura 6.5 si è tracciato l'andamento della varianza condizionata secondo il modello della tabella 6.8.



**Figura 6.5**

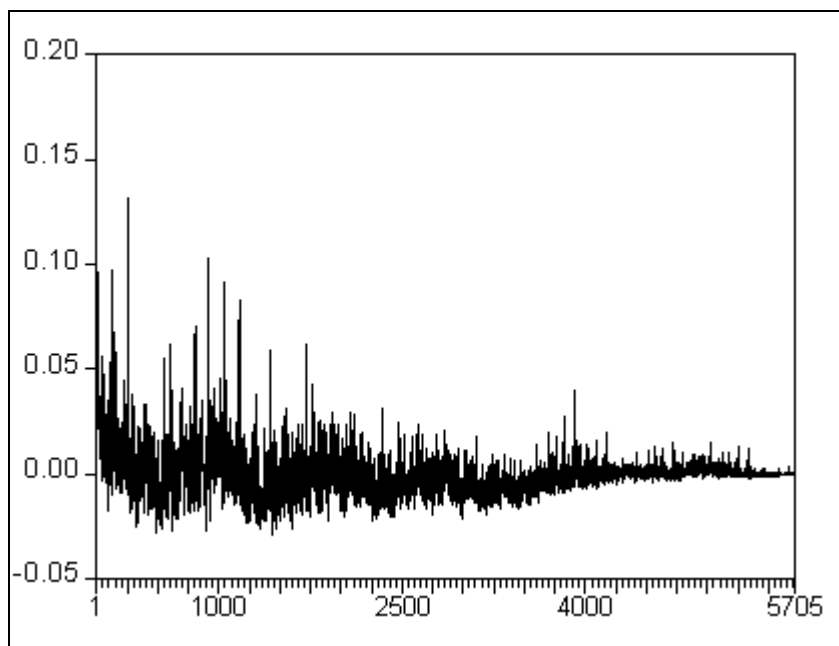
*Varianza condizionata secondo il modello della tabella 6.8*

### *6.1.2 Il modello FIGARCH originale*

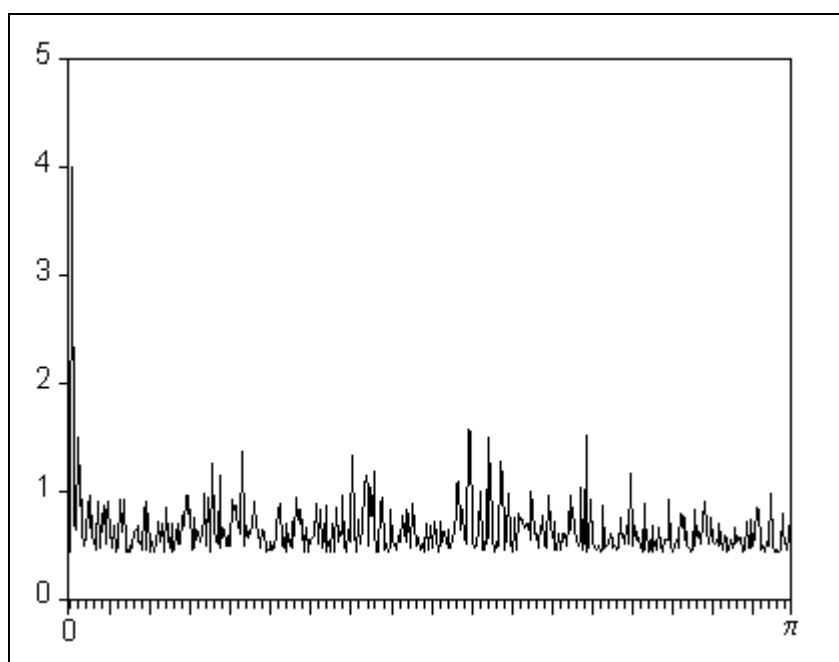
A questo punto, dopo aver identificato il modello GARCH “ottimale”, siamo pronti per introdurre l’integrazione frazionaria. L’utilità di una modellazione a memoria lunga è evidente se si esaminano le autocorrelazioni osservate e il periodogramma dei residui al quadrato<sup>3</sup>, che presenta un marcato picco a frequenze molto basse.

---

<sup>3</sup> Le autocorrelazioni e il periodogramma sono stati ottenuti mediante apposite routine C++ scritte dall’autore di questo lavoro. Il codice sorgente è presentato nel paragrafo B.2.7 in appendice.



**Figura 6.6**  
*Autocorrelazioni empiriche*



**Figura 6.7**  
*Periodogramma dei rendimenti al quadrato*

In questo paragrafo ci occuperemo pertanto della stima di alcuni modelli FIGARCH “classici”, nella parametrizzazione originale di Baillie, Bollerslev e Mikkelsen (1996) che

abbiamo presentato nella (5.2.15). Per tutte le stime si è utilizzata la metodologia proposta da Baillie, Bollerslev e Mikkelsen (1996), cioè mille ritardi per l'espansione del *fractional difference operator* e condizionamento a valori pre-campionari pari alla media dei residui al quadrato.

Per la selezione del modello si è utilizzato il pacchetto per Ox (si veda Doornik, 1999) G@RCH 1.1<sup>4</sup>, sviluppato da Laurent e Peters (2000); una volta identificata la parametrizzazione ottimale si è proceduto allo sviluppo di un'apposita routine di stima in C++, il cui codice è presentato in appendice nella sezione B.2.

Il primo modello che stimeremo sarà un FIGARCH (2,  $d$ , 1). I risultati della stima sono presentati nella tabella 6.9:

<i>Variabile</i>	<i>Coefficiente</i>	<i>Standard error</i>	<i>Statistica t</i>	<i>P-value</i>
$\mu$	- 0.003998	0.00802	- 0.4987	0.6180
$\omega$	0.073967	0.01851	3.995	0.0001
$1-D$	- 0.076961	0.02061	- 3.733	0.0002
$\psi$	0.232262	0.10791	2.152	0.0314
$\beta_1$	0.570320	0.13667	4.173	0.0000
$\beta_2$	0.063359	0.0500	1.267	0.2052
$d$	0.448156	0.0999	4.486	0.0000
<i>Log-verosimiglianza</i>			- 5430.243	
<i>AIC</i>			1.906130	
<i>SC</i>			1.914289	

**Tabella 6.9**

*Stima QML FIGARCH (2,  $d$ , 1) sui dati della figura 6.2*

La log-verosimiglianza aumenta marcatamente rispetto al GARCH (2, 1) della tabella 6.8, ed entrambi i criteri informativi concordano nell'accettare la nuova parametrizzazione.

Il coefficiente di integrazione frazionaria è decisamente significativo (e anche, ovviamente, statisticamente diverso da uno); osserviamo inoltre che il secondo ritardo GARCH diventa ridondante e perde di significatività.

<sup>4</sup> Per ulteriori dettagli su questo pacchetto si rimanda al paragrafo B.1.3 in appendice.

Un secondo tentativo è allora quello di eliminare il secondo coefficiente GARCH e stimare quindi un FIGARCH (1,  $d$ , 1):

<i>Variabile</i>	<i>Coefficiente</i>	<i>Standard error</i>	<i>Statistica t</i>	<i>P-value</i>
$\mu$	-0.003880	0.00801	-0.4843	0.6282
$\omega$	0.066654	0.01666	4.002	0.0001
$1-D$	-0.072209	0.01928	-3.744	0.0002
$\psi$	0.345517	0.05029	6.870	0.0000
$\beta$	0.708376	0.06462	10.96	0.0000
$d$	0.469507	0.10255	4.578	0.0000
<i>Log-verosimiglianza</i>			-5431.738	
<i>AIC</i>			1.906303	
<i>SC</i>			1.913296	

**Tabella 6.10**

*Stima QML FIGARCH (1,  $d$ , 1) sui dati della figura 6.2*

Osserviamo un leggero aumento del coefficiente di integrazione frazionaria (che di fatto raccoglie quello che si è perso eliminando il secondo ritardo GARCH) e una leggera diminuzione della log-verosimiglianza. In questo caso i criteri informativi danno risultati discordi: il criterio di Akaike (1973) preferisce il FIGARCH (2,  $d$ , 1), mentre quello di Schwarz (1978), più sensibile alla parsimonia, indica il FIGARCH (1,  $d$ , 1). Considerato che poter utilizzare la parametrizzazione FIGARCH (1,  $d$ , 1) è preferibile anche per la disponibilità dei vincoli sui coefficienti (5.2.18) e (5.2.19), decidiamo di seguire la strada più semplice.

Un ultimo tentativo è quello di eliminare il coefficiente ARCH, come suggerito da Baillie, Bollerslev e Mikkelsen (1996) e stimare quindi un FIGARCH (1,  $d$ , 0):

<i>Variabile</i>	<i>Coefficiente</i>	<i>Standard error</i>	<i>Statistica t</i>	<i>P-value</i>
$\mu$	-0.003091	0.00821	-0.3766	0.7065
$\omega$	0.130986	0.02451	5.344	0.0000
$1-D$	-0.085138	0.02518	-3.381	0.0007
$\beta$	0.153238	0.04441	3.450	0.0006
$d$	0.248689	0.03508	7.089	0.0000
<i>Log-verosimiglianza</i>			-5464.831	
<i>AIC</i>			1.917557	
<i>SC</i>			1.923384	

**Tabella 6.11**

*Stima QML FIGARCH (1,  $d$ , 0) sui dati della figura 6.2*

Sicuramente con questa specificazione non si fanno dei passi avanti: la perdita di verosimiglianza è notevole; insoliti sono gli effetti sul coefficiente di integrazione frazionaria (ci si sarebbe aspettati un aumento, piuttosto che una diminuzione), e insolito è altresì l'effetto sul coefficiente della dummy, che abbandona l'intorno di  $-0.75$  in cui si era assestato in tutte le altre stime per scendere a  $-0.85121$ .

Volendo essere più scrupolosi, possiamo condurre un semplice test LR: utilizzando la (4.2.2) si ha:

$$2(-5431.738 + 5464.831) = 66.186 \sim \chi_1^2, \quad (6.1.1)$$

e ci troviamo ovviamente nella regione di rifiuto dell'ipotesi nulla; dobbiamo quindi concordare con Beine, Laurent e Lecourt (1999) sull'utilità del coefficiente ARCH.

Un'ultima questione da analizzare riguarda la simmetria degli shock: se si riscontra asimmetria potrebbe essere necessario ricorrere a una modellazione FIEGARCH; fortunatamente, però, non è questo il caso:

<i>Test</i>	<i>Statistica t</i>	<i>P-value</i>
<i>Sign bias test</i>	1.28383	0.19920
<i>Negative size bias test</i>	0.81337	0.41600
<i>Positive size bias test</i>	1.61579	0.10614
<i>Statistica F</i>		
<i>Test congiunto</i>	3.30015	0.34762

**Tabella 6.12**

*Test per effetti asimmetrici sul modello della tabella 6.10*

Concludiamo pertanto che la modellazione più adeguata è il FIGARCH (1,  $d$ , 1).

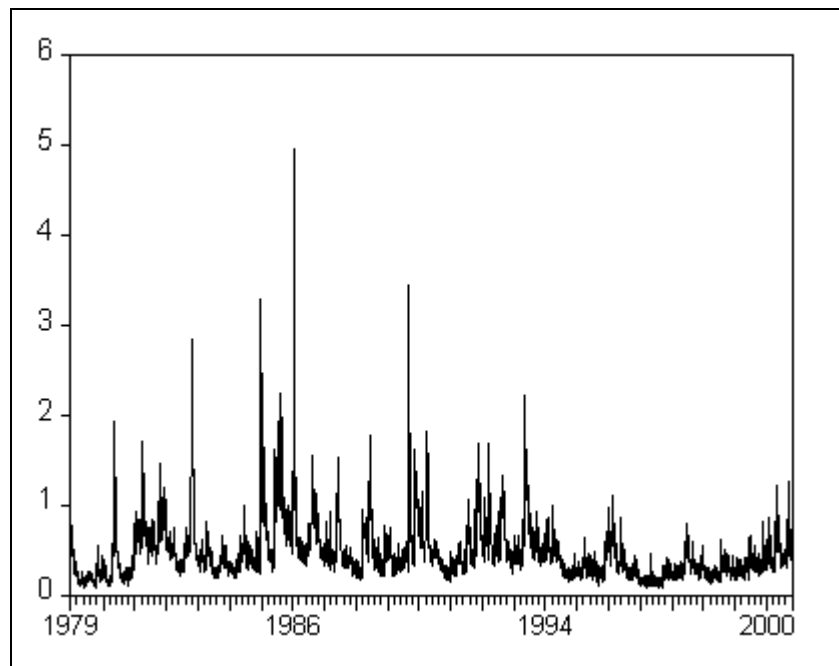
Nella tabella 6.13 presentiamo i risultati della stima di questo modello con la routine da noi sviluppata.

<i>Variabile</i>	<i>Coefficiente</i>	<i>Standard error</i>	<i>Statistica t</i>	<i>P-value</i>
$\mu$	- 0.003858	0.008012	- 0.481608	0.6301
$\omega$	0.066255	0.016606	3.989708	0.0001
$1-D$	- 0.071778	0.019254	- 3.727907	0.0002
$\psi$	0.345168	0.050601	6.821366	0.0000
$\beta$	0.709362	0.064374	11.019356	0.0000
$d$	0.470897	0.102841	4.578884	0.0000
<i>Log-verosimiglianza</i>			- 5431.787692	
<i>AIC</i>			1.904220	
<i>SC</i>			1.913316	

**Tabella 6.13**

*Stima QML FIGARCH (1, d, 1) sui dati della figura 6.2, codice C++*

Nella figura 6.8 abbiamo tracciato l'andamento della varianza condizionata secondo il modello della tabella 6.13:



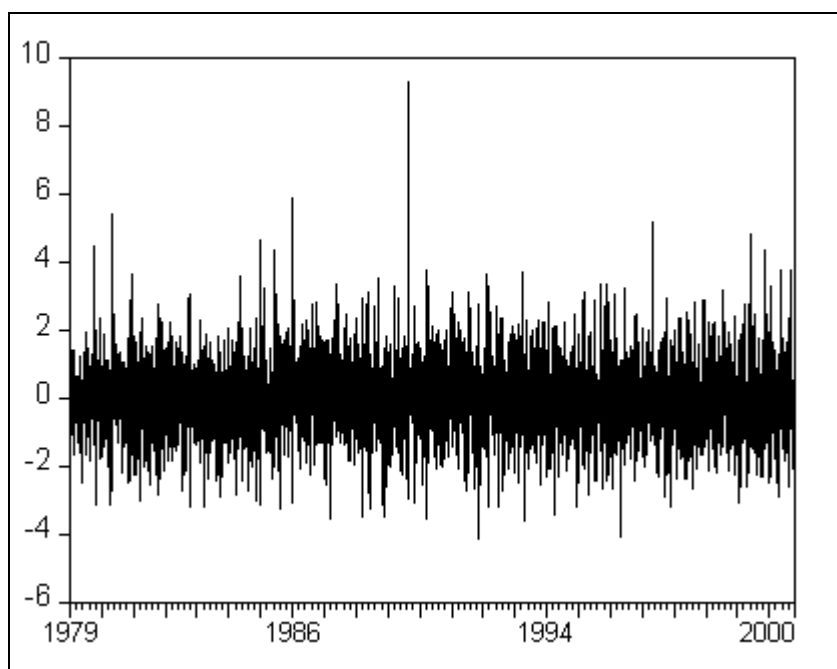
**Figura 6.8**

*Varianza condizionata secondo il modello della tabella 6.13*

Nella figura 6.9 vediamo invece il grafico dei residui standardizzati, che ha in linea di massima un andamento soddisfacente (si noti in particolare la scomparsa del *clustering* della volatilità); se poi eliminiamo i valori più anomali, cioè i rendimenti superiori al 4% o inferiori

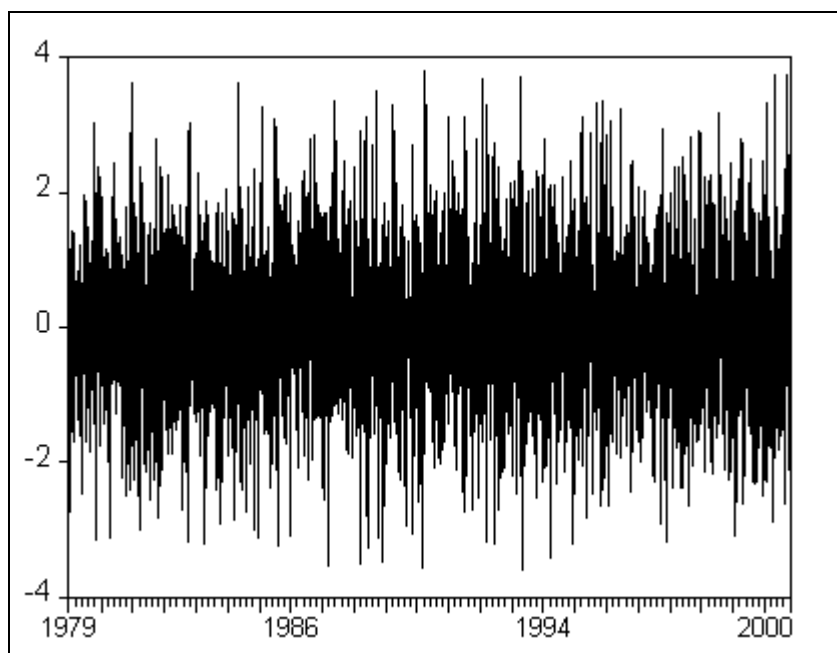


al  $-4\%$ , otteniamo la figura 6.10, le cui statistiche relative alla normalità sono visualizzate nella figura 6.11.



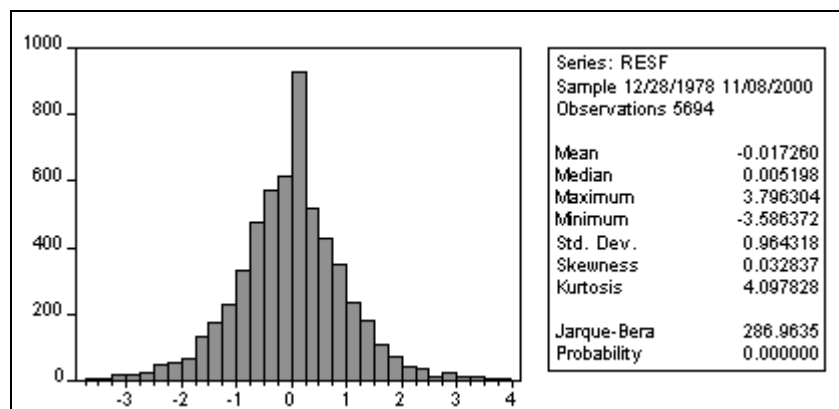
**Figura 6.9**

*Residui standardizzati del modello della tabella 6.13*



**Figura 6.10**

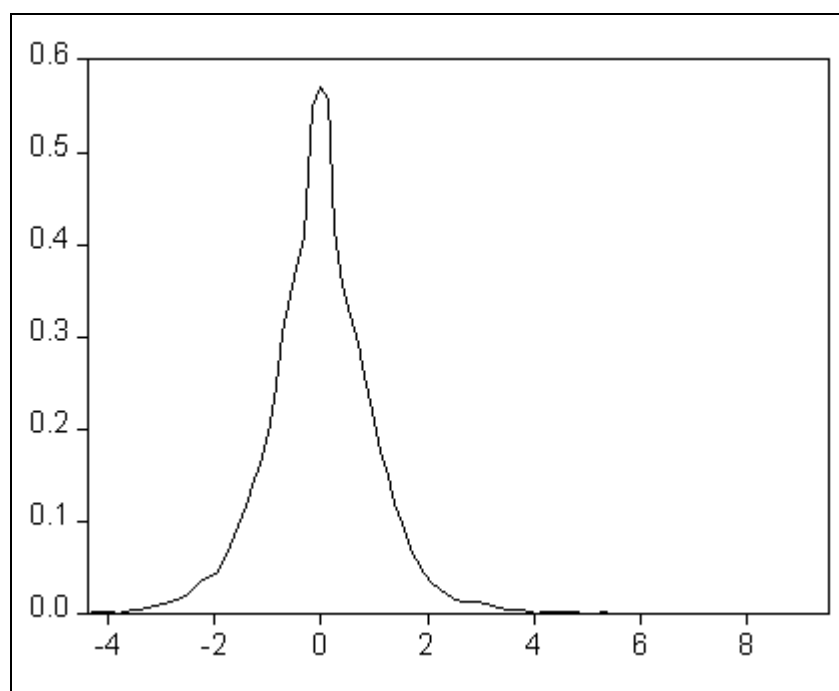
*Residui della figura 6.9 con l'esclusione dei valori esterni all'intervallo  $[-4, 4]$*



**Figura 6.11**

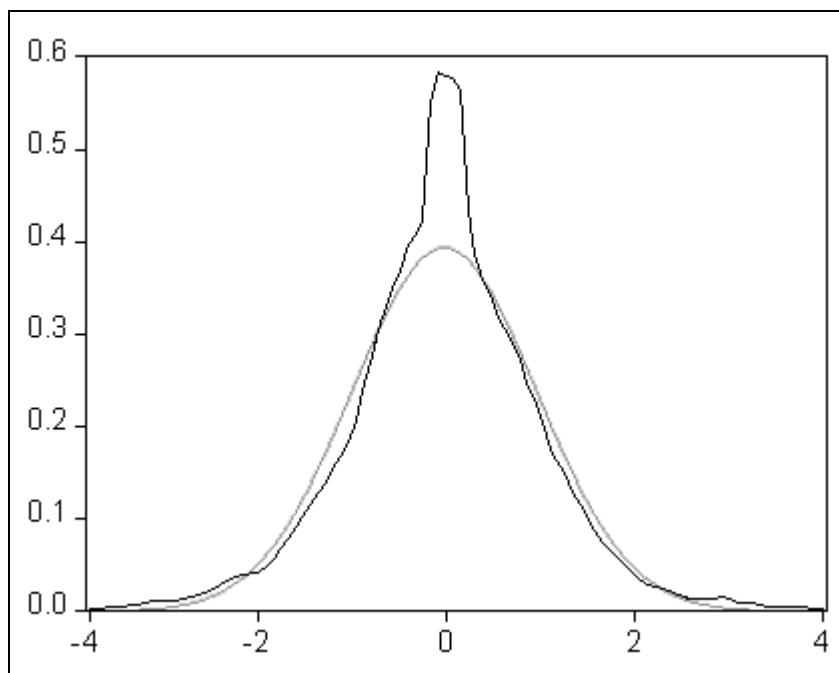
*Statistiche di normalità per i residui della figura 6.10*

Nelle figure 6.12 e 6.13 si sono tracciate le stime kernel della densità dei residui della figura 6.9 e 6.10, mentre la figura 6.14 riporta il grafico quantile-quantile dei residui della figura 6.10.



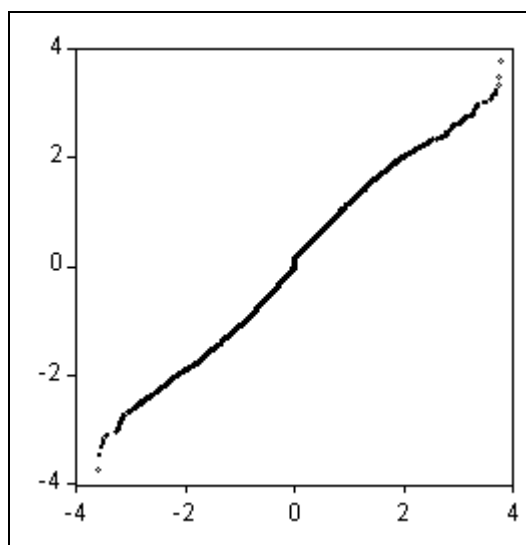
**Figura 6.12**

*Stima kernel uniforme della densità dei residui della figura 6.9, ampiezza di banda 0.2318*



**Figura 6.13**

*Stima kernel uniforme della densità dei residui della figura 6.10, ampiezza di banda 0.2317;  
in grigio la densità di una distribuzione  $N(0, 1)$*



**Figura 6.14**

*Grafico dei quantili della distribuzione dei residui  
della figura 6.10 contro i quantili della normale*

Le informazioni che possiamo ricavare da questi grafici sono assai preziose. Si osserva in primo luogo che, anche eliminando i residui più anomali, non si riesce ad avvicinarsi alla

normalità, perché permane comunque un'eccessiva curtosi, dovuta alla presenza di troppi valori sulle code. Ciò è indicato chiaramente anche dalla figura 6.14: ci si discosta dalla normalità proprio sulle code dello *scatter* quantile-quantile. Ciò indica chiaramente, come avevamo ipotizzato in partenza, che la distribuzione dei termini di disturbo non è normale, ed è quindi stato saggio procedere utilizzando la quasi-massima verosimiglianza.

### 6.1.3 La parametrizzazione alternativa<sup>5</sup>

È interessante confrontare i risultati che abbiamo ottenuto finora con quelli che si sarebbero ottenuti utilizzando la (4.2.31) di Chung (1999). Seguendo quanto suggerito da Chung (1999), si è condizionato a valori pari a zero e si è utilizzato, per l'espansione del *fractional difference operator*, un numero di ritardi variabile pari a  $t$ .

I risultati della stima sono presentati nella tabella 6.14:

<i>Variabile</i>	<i>Coefficiente</i>	<i>Standard error</i>	<i>Statistica t</i>	<i>P-value</i>
$\mu$	- 0.003512	0.00842	- 0.417	0.6766
$\sigma^2$	0.344732	0.09206	3.744	0.0002
$1-D$	- 0.048916	0.02161	- 2.264	0.0236
$\psi$	0.376623	0.04020	9.369	0.0000
$\beta$	0.654078	0.05424	12.06	0.0000
$d$	0.376623	0.05526	6.647	0.0000
<i>Log-verosimiglianza</i>			- 5440.344	
<i>AIC</i>			1.909321	
<i>SC</i>			1.916314	

**Tabella 6.14**

*Stima QML FIGARCH (1, d, 1) nella forma della (4.2.31) sui dati della figura 6.2*

Osserviamo che, al di là della costante, che però in questo caso assume un significato del tutto diverso, non vi sono, sui parametri GARCH tradizionali, grosse differenze con il modello FIGARCH originale stimato nella tabella 6.10. Si rileva in particolare una diminuzione (in valore assoluto) del coefficiente della *dummy* e una leggera diminuzione del coefficiente GARCH.

Più marcate sono invece le discrepanze per quanto riguarda il coefficiente di integrazione frazionaria  $d$ , per il quale si osserva una sensibile diminuzione. Vi è da dire che gli studi su dati simulati presentati da Chung (1999) avevano mostrato sì la presenza di distorsione negativa sulle stime di  $d$ , ma di entità dell'ordine di  $10^{-3}$ , anche su campioni molto più piccoli del nostro; senza dubbio la questione ha bisogno di ulteriori approfondimenti<sup>6</sup>.

#### 6.1.4 Previsioni sulla volatilità

A questo punto, dopo aver stimato una serie di modelli GARCH e FIGARCH e aver individuato le due modellazioni più adeguate, cioè il GARCH (2, 1) e il FIGARCH (1,  $d$ , 1), possiamo occuparci di confrontare le previsioni sulla volatilità che si ricavano applicando ciascuno di questi due modelli. Le previsioni<sup>7</sup> sono condotte su un orizzonte temporale che si conclude al 31 gennaio 2001, per un totale di 60 giorni lavorativi, che diventano 58 se si escludono le festività (comunque inserite nel campione). Considerato che l'effetto della dummy è praticamente identico nei due modelli, si è deciso, al fine di rendere il confronto più immediatamente intelligibile, di non considerarlo all'interno delle previsioni. Si sono pertanto ristimati<sup>8</sup> i modelli delle tabelle 6.8 e 6.13 escludendo la dummy:

<i>Variabile</i>	<i>Coefficiente</i>	<i>Standard error</i>	<i>Statistica t</i>	<i>P-value</i>
$\mu$	-0.004339	0.007937	-0.546669	0.5846
$\omega$	0.012245	0.002853	4.292393	0.0000
$\alpha$	0.103275	0.017531	5.891063	0.0000
$\beta_1$	0.393000	0.148367	2.648831	0.0081
$\beta_2$	0.479605	0.137081	3.498701	0.0005
<i>Log-verosimiglianza</i>			-5452.314	
<i>AIC</i>			1.913239	
<i>SC</i>			1.919066	

**Tabella 6.15**

*Stima QML GARCH (2, 1) sui dati della figura 6.2*

<sup>5</sup> Le stime di questo paragrafo sono state condotte col pacchetto G@RCH 1.1 di Laurent e Peters (2000).

<sup>6</sup> Potrebbe senz'altro essere interessante osservare, su dati simulati, cosa succede se si stima col modello di Chung (1999) un processo generato dal modello di Baillie, Bollerslev e Mikkelsen (1996).

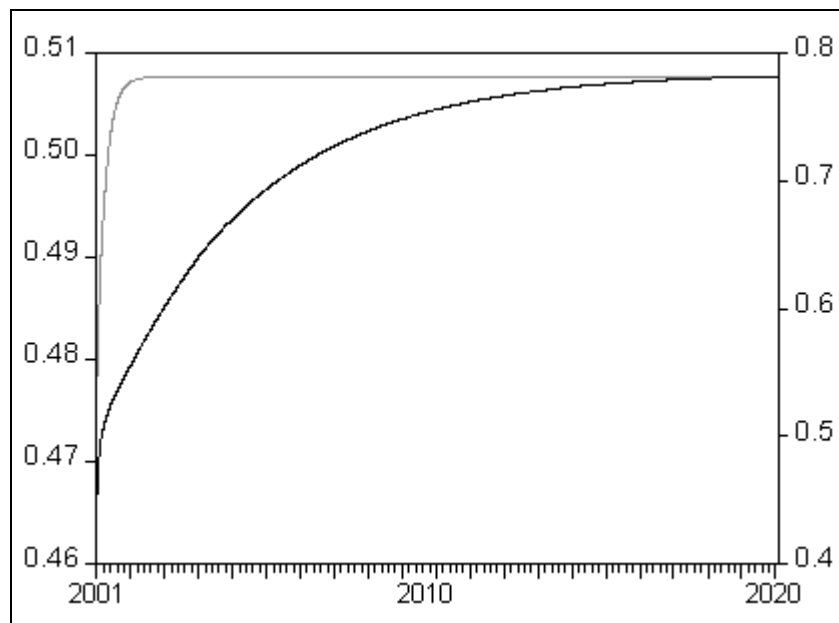
<sup>7</sup> Per quanto riguarda il GARCH (2, 1) si è utilizzato di nuovo Eviews, mentre per il FIGARCH (1,  $d$ , 1) si è sviluppata un'apposita routine C++ presentata nel paragrafo B.2.7 in appendice.

<i>Variabile</i>	<i>Coefficiente</i>	<i>Standard error</i>	<i>Statistica t</i>	<i>P-value</i>
$\mu$	- 0.003642	0.008083	- 0.450611	0.6523
$\omega$	0.011209	0.004081	2.746427	0.0060
$\psi$	0.335278	0.047998	6.986695	0.0000
$\beta$	0.713007	0.063637	11.204248	0.0000
$d$	0.479121	0.097046	4.937070	0.0000
<i>Log-verosimiglianza</i>			- 5447.172	
<i>AIC</i>			1.909613	
<i>SC</i>			1.918710	

**Tabella 6.16**

*Stima QML FIGARCH (1, d, 1) sui dati della figura 6.2*

Sappiamo già cosa attenderci, perché abbiamo già visto nella (5.1.48) e nella (5.2.25) che le previsioni, per i modelli GARCH e FIGARCH, decadono rispettivamente esponenzialmente e iperbolicamente; vale la pena comunque focalizzare meglio la situazione osservando la figura 6.15.

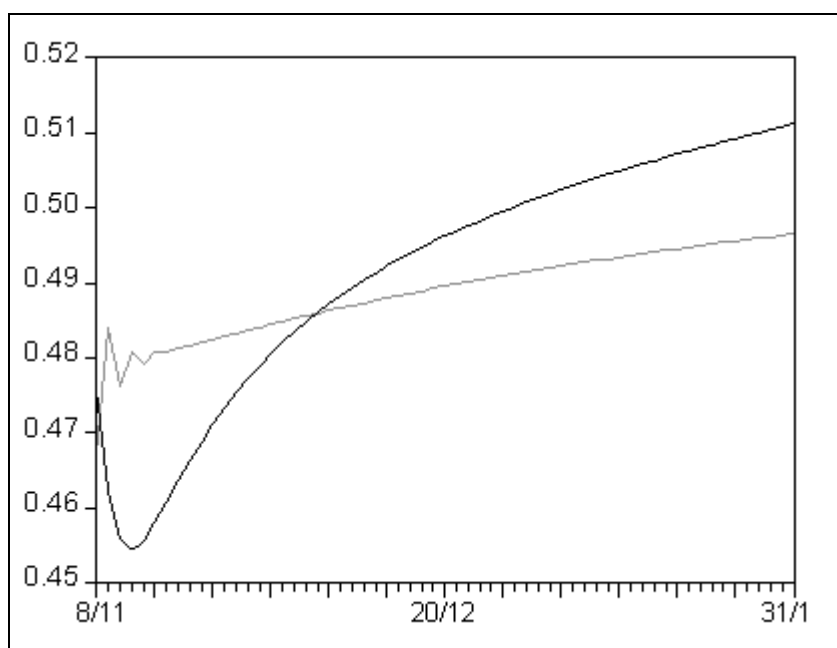


**Figura 6.15**

*Volatilità attesa secondo il modello GARCH (2, 1) della tabella 6.15 (linea grigia – scala sinistra) e FIGARCH (1, d, 1) della tabella 6.16 (linea nera – scala destra)*

<sup>8</sup> La stima del modello GARCH (2, 1) senza dummy è stata condotta utilizzando Eviews, mentre per il modello FIGARCH (1, d, 1) si è utilizzato il codice C++ da noi sviluppato, presentato nella sezione B.2 in appendice.

Osserviamo in primo luogo come i due modelli convergano a due valori differenti: il GARCH (2, 1) a 0.50767 e il FIGARCH (1,  $d$ , 1) a 0.78356. Ciò è dovuto alla controversa interpretazione della costante nei modelli FIGARCH, questione che è stata affrontata nel paragrafo 5.2.3 e che, come abbiamo rilevato, necessita di ulteriori approfondimenti. Per quanto riguarda la velocità di convergenza, è immediatamente evidente che, mentre il GARCH tradizionale converge abbastanza rapidamente (circa 500 osservazioni), il FIGARCH impiega molto più tempo a raggiungere il livello di equilibrio. Da questo confronto emerge immediatamente la maggior adeguatezza dei modelli a integrazione frazionaria per quanto riguarda gli orizzonti temporali di medio e lungo periodo. È però interessante osservare come l'utilizzo del modello FIGARCH (1,  $d$ , 1) abbia effetti rilevanti anche sulla dinamica di breve periodo.



**Figura 6.16**

*Volatilità attesa secondo il modello GARCH (2, 1) della tabella 6.15 (linea grigia)  
e FIGARCH (1,  $d$ , 1) della tabella 6.16 (linea nera)*

L'orizzonte temporale che abbiamo considerato nella figura 6.16 non è particolarmente lungo, ma la stima GARCH tradizionale dà, già per una stima a trenta giorni, un valore assai

prossimo alla varianza non condizionata. Utilizzando invece il modello FIGARCH (1,  $d$ , 1) osserviamo, nei primi tre giorni, una diminuzione della volatilità, che torna poi a crescere verso i livelli indicati dal modello GARCH (2, 1). L'utilizzo dell'integrazione frazionaria, quindi, oltre a modellare in maniera più realistica l'andamento sul lungo periodo, dà anche, pesando anche le osservazioni più lontane nel tempo, una differente dinamica di breve periodo.

Volendo ad esempio utilizzare queste stime di breve periodo per computare il *Value at Risk* di un portafoglio che, per semplicità<sup>9</sup>, ipotizziamo composto da 100\$ (tale portafoglio, all'8 novembre 2000, data finale del nostro campione, sarebbe valso 116.836079€), si otterrebbe:

<i>Metodo</i>	<i>VaR a 1 giorno</i>	<i>VaR a 3 giorni</i>	<i>VaR a 10 giorni</i>
<i>Varianza storica</i>	1.228776€	2.236490€	4.083254€
<i>GARCH (2, 1)</i>	1.313380€	2.296143€	4.208076€
<i>FIGARCH (1, <math>d</math>, 1)</i>	1.343779€	2.296051€	4.140958€

**Tabella 6.17**

*Calcolo del VaR di un portafoglio composto da 100\$ (116.836069€ all'8/11/2000)  
secondo i modelli delle tabelle 6.15 e 6.16*

Osserviamo in primo luogo che la varianza storica (0.451363) non coincide con la varianza non condizionata a cui convergono le previsioni secondo il modello GARCH (2, 1), cioè 0.50767; ci troviamo perciò in una situazione in cui la varianza attuale è superiore a quella storica ma inferiore a quella indicata dai modelli ARCH. In questo caso le stime del *VaR* condotte col metodo di Barone-Adesi e Giannopoulos (1996) saranno comunque più prudentziali rispetto a quelle che si ottengono con la varianza storica.

Osserviamo cosa accade invece se utilizziamo il FIGARCH (1,  $d$ , 1) della tabella 6.16: Come si può vedere agevolmente dalla figura 6.16, nei primi trenta giorni, le stime della volatilità offerte dal modello FIGARCH (1,  $d$ , 1) sono sistematicamente inferiori rispetto a quelle del

<sup>9</sup> È evidente che l'analisi del *VaR* di un portafoglio così composto è assai poco interessante, se non altro perché è difficile pensare che per liquidarlo ci possano volere tre o addirittura dieci giorni. L'esempio ha comunque funzioni esclusivamente illustrative, per cui ci si perdonerà la semplificazione.



GARCH (2, 1). Pertanto, le stime del  $VaR$  del modello FIGARCH (1,  $d$ , 1) sono comprese tra quelle della varianza storica e quelle del GARCH (2, 1); in questo caso l'utilizzo dell'integrazione frazionaria ci consente di avere risultati intermedi tra il metodo della varianza storica e quello GARCH (2, 1).

## 6.2 Conclusioni

Dopo aver introdotto gli strumenti di analisi appropriati, abbiamo presentato i principali risultati teorici relativi ai processi a integrazione frazionaria, le ricerche empiriche che ne hanno individuato la presenza all'interno di alcune serie economiche e i modelli teorici che spiegano l'insorgenza dei fenomeni di integrazione frazionaria sulla base dell'aggregazione di più processi autoregressivi. L'importanza dei modelli a integrazione frazionaria è così palese: è noto infatti che la maggior parte delle grandezze macroeconomiche sono determinate dall'aggregazione di più microvariabili. Si è poi introdotto il processo di Gegenbauer, una generalizzazione del *fractional white noise*, che contempla la possibilità che il decadimento delle autocorrelazioni segua un andamento iperbolico-sinusoidale, e se ne sono esaminate le proprietà principali.

Dopo aver richiamato i principi generali della stima di massima e quasi-massima verosimiglianza, si sono presentate le metodologie di stima più comuni per i processi a integrazione frazionaria; ciò ha completato il quadro di riferimento relativo ai processi  $I(d)$ . Si è quindi presentata un'applicazione empirica di questo tipo di modellistica, utilizzando la serie storica del livello minimo del fiume Nilo.

Si sono quindi introdotti i modelli ARCH (Engle, 1982), GARCH (Bollerslev, 1986) e IGARCH (Engle e Bollerslev, 1986), nonché le altre varianti principali; si sono illustrate le metodologie di stima e di test per effetti ARCH e si è infine mostrato come utilizzare tali modelli per fare previsioni sulla volatilità dei rendimenti dei mercati finanziari e calcolare il *Value at Risk* di un certo portafoglio.

Ci siamo trovati a questo punto in possesso di tutti gli strumenti necessari ad affrontare il tema centrale di questo lavoro: si sono presentate le ricerche che hanno mostrato evidenza di memoria lunga all'interno della volatilità dei rendimenti dei mercati finanziari e si è quindi introdotto il modello *Fractionally Integrated GARCH* (FIGARCH) di Baillie, Bollerslev e

Mikkelsen (1996), illustrandone con dettaglio gli aspetti teorici e le metodologie di stima. Si è rilevato in particolare come il modello FIGARCH, nel quale le previsioni decadono con andamento iperbolico, si ponga come un'interessante "via di mezzo" tra il GARCH (decadimento esponenziale) e l'IGARCH (persistenza indefinita). Si sono infine presentate le prime ricerche empiriche che hanno applicato il modello FIGARCH a dati reali, rilevandone l'importanza ai fini esplicativi, nonché alcune estensioni teoriche del modello, con particolare attenzione alla parametrizzazione alternativa proposta da Chung (1999).

Si è inoltre mostrato come condurre, all'atto pratico, la stima di un modello FIGARCH, utilizzando dati reali sul tasso di cambio \$/€ per il periodo che va dal 28 dicembre 1978 all'8 novembre 2000; da tale analisi si è ricavata l'evidente presenza, all'interno dei dati considerati, di integrazione frazionaria. Abbiamo poi utilizzato il modello FIGARCH così stimato per fare previsioni sulla volatilità futura di questo tasso di cambio, confrontandole con quelle che si sarebbero ottenute utilizzando un tradizionale GARCH (si vedano le figure 6.15 e 6.16); si sono infine utilizzate tali previsioni per calcolare il *Value at Risk* di un ipotetico portafoglio composto da 100\$, mettendo in risalto la differenza tra il risultato ottenuto con la stima GARCH tradizionale e quello ottenuto con la stima FIGARCH.

### *6.2.1 Le previsioni nei mercati finanziari<sup>10</sup>*

Abbiamo già rilevato nell'introduzione come l'econometria produca e studi modelli atti a descrivere, nella maniera più precisa possibile, il comportamento, nel tempo o nello spazio, delle variabili economiche. Si potrebbe dire che, in un certo senso, l'econometrico produce gli attrezzi che l'economista (o comunque l'analista dei fenomeni economici) utilizza per

---

<sup>10</sup> Non vogliamo assolutamente, con queste considerazioni, dare l'impressione di voler suggerire il metodo di previsione "migliore", tanto più che riteniamo che non vi possa essere una soluzione univoca della questione. Ci interessa qui solo mettere in rilievo punti di forza e di debolezza di ciascuno dei metodi che esamineremo, con particolare attenzione al metodo econometrico. Per una trattazione più completa e più approfondita di questo spinoso argomento si può vedere Gabbi (1999), a cui ci si è ispirati nella stesura di questo paragrafo.

svolgere il proprio lavoro. La fortuna di un modello è determinata, pertanto, non solo da chi lo crea (e dai suoi colleghi), ma anche e soprattutto da chi si trova a doverlo utilizzare.

Abbiamo osservato altresì come gli “utenti” d’elezione per gli studi relativi alle serie storiche, e in particolare alle loro capacità previsive, siano le varie istituzioni finanziarie. Il successo della modellistica ARCH (Engle, 1982) e il vasto numero di estensioni proposte deriva proprio da questo: si tratta infatti di un modello, relativamente semplice, per l’analisi e la previsione della volatilità dei mercati finanziari, cioè una tematica estremamente rilevante per gli “utenti” di cui sopra.

Vale la pena rilevare che gli strumenti econometrici non sono però gli unici utilizzati dalle istituzioni finanziarie per formulare previsioni. Si può dire, in termini estremamente approssimativi, che esistono altri due tipi di approccio: l’analisi fondamentale e l’analisi tecnica<sup>11</sup>.

L’analista fondamentale stima il valore di un certo attivo da una prospettiva “aziendale”: nel caso di un’azione, ad esempio, valuta i possibili sviluppi del *business* dell’azienda, i flussi di cassa futuri che la gestione potrà essere in grado di generare, i dividendi attesi. Ovviamente la stima di un “fondamentalista” è centrata sul valore, piuttosto che sui rendimenti.

Tuttavia è semplice spostare il fuoco dell’analisi su questi ultimi. Nel breve periodo possono prevalere movimenti speculativi che spostano il prezzo dell’azione dal suo “valore fondamentale”: l’analista fondamentale potrà così selezionare i titoli il cui prezzo di mercato è inferiore al loro valore fondamentale. Poiché il prezzo di tali titoli dovrà prima o poi convergere al valore fondamentale, acquistandoli si potranno realizzare, nel medio e nel lungo periodo, rendimenti significativamente positivi.

---

<sup>11</sup> Queste due tematiche sono sconfinite, e nella presentazione che seguirà si cercherà semplicemente di dare al lettore un’idea generale sui loro presupposti e sul loro funzionamento. Per una trattazione più approfondita degli strumenti più utilizzati per stimare il valore fondamentale di un’azienda si rimanda a Brealey e Myers (1993) o Cattaneo (1999), per un’introduzione all’analisi tecnica si può vedere Edwards e Magee (1992).

I metodi per giungere alla definizione di un “valore fondamentale” sono però potenzialmente infiniti, per cui ogni analista potrà fornire previsioni diverse; l’attendibilità delle stesse dipenderà quindi in modo decisivo dalle capacità e dall’esperienza di chi le formula.

L’analista tecnico muove invece da prospettive totalmente diverse, cioè dall’idea che l’andamento dei prezzi delle azioni segua determinate regolarità empiriche, che sono visibili dall’analisi del grafico (tanto che alcuni indicano gli analisti tecnici col nome di “graficisti”) e che pertanto prendono il nome di figure. Ad esempio, esistono i canali di prezzo, che indicano l’andamento di medio periodo del prezzo di un titolo, le bandiere, che indicano l’esaurimento della tendenza in atto, o le formazioni testa e spalle, che preludono un’inversione di tendenza. Se si riesce a identificare una figura prima che essa sia completata, si può, sulla base delle regolarità menzionate in precedenza, prevedere l’andamento della serie. È chiaro quindi come l’analisi tecnica si rivolga prevalentemente alle previsioni di breve periodo, e quindi allo speculatore più che al “cassettista”.

L’analisi tecnica non muove da assunzioni economiche o risultati statistici particolari<sup>12</sup>, ma si basa quasi totalmente sull’abilità dell’analista nel riconoscere le figure; perciò, allo stato attuale delle cose, è più che altro un’arte<sup>13</sup>: un detto comune tra gli analisti tecnici è «il problema non è il grafico, è il graficista». Nondimeno, non si può escludere che gli analisti tecnici abbiano realmente individuato delle regolarità empiriche la cui base statistico-probabilistica è, allo stato attuale delle cose, oscura. Del resto, anche i primi medici, pur muovendo da assunzioni teoriche totalmente errate circa il funzionamento del nostro organismo, avevano individuato delle regolarità empiriche che consentivano loro di curare certe patologie ben prima dell’avvento della biochimica, della farmacia e del metodo scientifico.

---

<sup>12</sup> Si noti però che alcuni studi recenti hanno cercato di identificare queste relazioni, si veda ad esempio Lo (2000).

Abbiamo osservato, nella sezione 5.1 come l'analisi di tipo econometrico non fornisca strumenti particolarmente utili per la previsione dei rendimenti futuri; tuttavia abbiamo visto anche che i modelli ARCH (Engle, 1982) possano rappresentare in modo soddisfacente l'andamento della varianza condizionata, e possono quindi essere usati per fare previsioni sulla volatilità.

Vale la pena osservare che questo risultato è comunque molto importante, all'atto pratico: pochi gestori di portafoglio destinano quote rilevanti dei propri fondi alla gestione "speculativa" (e quindi, alla previsione dei rendimenti), mentre quasi tutti coprono adeguatamente il rischio<sup>14</sup> dei propri attivi. Per coprire il rischio è necessario riuscire a quantificarlo, ed è pertanto fondamentale disporre di previsioni sulla volatilità.

Un grosso limite dei modelli ARCH tradizionali, nell'utilizzo a fini previsivi, è la loro miopia, cioè la loro scarsa attitudine a "vedere lontano".

Abbiamo difatti rilevato come le previsioni, nei modelli ARCH, convergano a velocità esponenziale verso la varianza non condizionata del processo. Convergenza a velocità esponenziale significa che, già su un orizzonte temporale di un mese, si avrà una previsione di poco distante dal livello di convergenza. Se per alcune applicazioni, ad esempio il calcolo del *VaR*, un orizzonte di un mese è più che sufficiente (è infatti molto difficile immaginare un portafoglio che possa richiedere un mese per essere liquidato), per altre non lo è affatto: si pensi ad esempio al prezzaggio delle opzioni secondo i metodi di Black e Scholes (1973) o Hull e White (1987).

---

<sup>13</sup> Chi ha un po' di domestichezza con l'argomento sa quanto sia facile riconoscere le "figure" a posteriori, e quanto sia difficile e aleatorio farlo a priori, cioè in ultima analisi quando ce n'è bisogno.

<sup>14</sup> Il rischio, nei mercati finanziari, rappresenta la possibilità che il valore presente di un certo attivo sia eroso a causa di movimenti sfavorevoli del mercato; la volatilità di un attivo, usualmente utilizzata come indicatore di rischio, è qui intesa come misura della dispersione dei rendimenti futuri attorno al loro valore atteso.

### 6.2.2 *L'utilizzo dei modelli a memoria lunga*

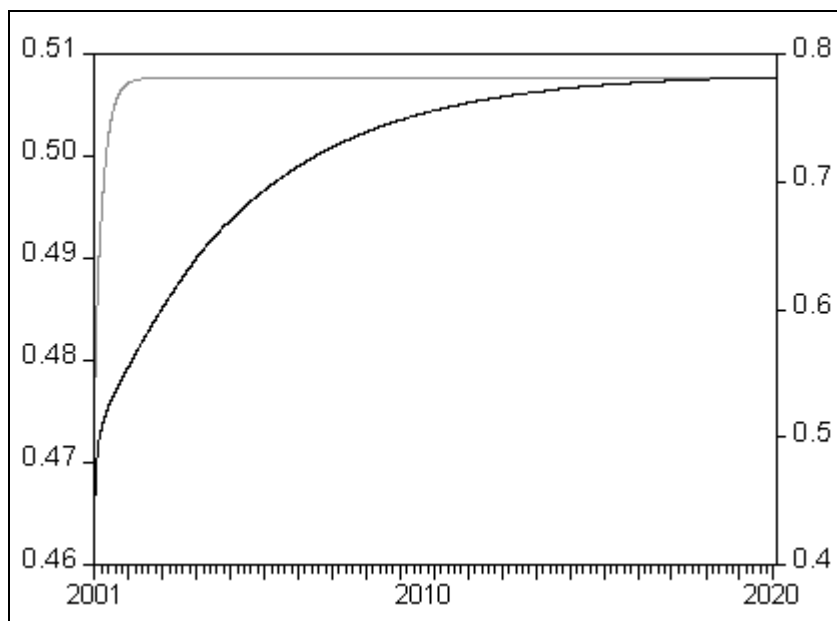
È qui che si rivela l'importanza di una modellistica a memoria lunga per la varianza condizionata. Il “prezzo” dell'utilizzo di un modello FIGARCH è rappresentato dal fatto che la sua stima è indubbiamente e sensibilmente più lenta rispetto a un modello GARCH: la maggiore precisione del FIGARCH costa quindi molto in termini di tempo di esecuzione per la procedura di stima<sup>15</sup>. Se un accademico non ha in genere vincoli temporali troppo stringenti, un gestore di portafoglio ha bisogno di aggiornare in tempo reale le sue previsioni sulla volatilità.

Tuttavia questo prezzo non paga solo la maggiore precisione, ma anche la maggior “lungimiranza” delle previsioni: si è visto infatti che le previsioni sulla varianza condizionata fornite da un modello FIGARCH convergono verso il livello di equilibrio<sup>16</sup> con velocità iperbolica. Ciò ci consente di avere previsioni che risentono dell'influenza del presente per molto tempo. Per apprezzare meglio questo fatto basta osservare la figura 6.15, che qui riproduciamo per comodità.

---

<sup>15</sup> Su un processore Intel Pentium III a 500 Mhz la stima di un modello GARCH per una serie di cinquemila osservazioni richiede pochi secondi, mentre un modello FIGARCH impiega dai due ai cinque minuti.

<sup>16</sup> La teoria è purtroppo ancora carente su questo punto: le previsioni convergono ad un valore strutturalmente diverso dalla varianza non condizionata. D'altra parte, alcuni risultati sembrano suggerire che, nei modelli FIGARCH, la varianza non condizionata sia infinita. Per una trattazione dell'argomento si rimanda al paragrafo 5.2.3.



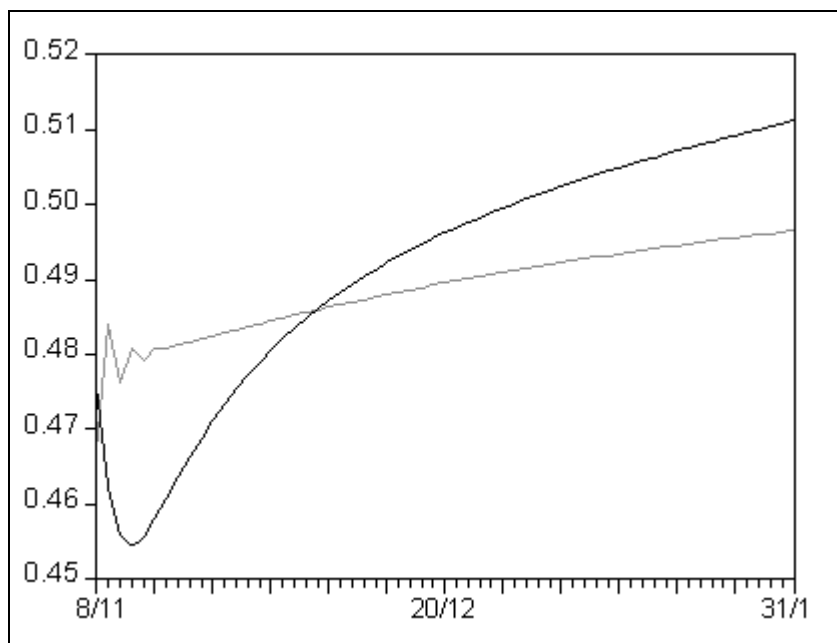
**Figura 6.15**

*Volatilità attesa secondo il modello GARCH (2, 1) della tabella 6.15 (linea grigia – scala sinistra) e FIGARCH (1, d, 1) della tabella 6.16 (linea nera – scala destra)*

Le previsioni del GARCH tradizionale (linea grigia) raggiungono il livello di equilibrio dopo pochi mesi, mentre quelle del FIGARCH impiegano ben venti anni a raggiungere la convergenza. Ciò si rivela assai prezioso in tutte le situazioni in cui si richiedono stime sulla volatilità a medio e lungo termine; ad esempio, come abbiamo detto poc'anzi, per il prezzaggio delle opzioni.

A ciò va aggiunta la maggior precisione del modello FIGARCH e la sua maggior sensibilità all'andamento passato della serie: ciò fa in modo che questo modello fornisca previsioni sensibilmente diverse da quelle dei GARCH tradizionali anche per il breve periodo, come si può notare dalla figura 6.16, riportata qui sotto.





**Figura 6.16**

*Volatilità attesa secondo il modello GARCH (2, 1) della tabella 6.15 (linea grigia) e FIGARCH (1, d, 1) della tabella 6.16 (linea nera)*

Riteniamo pertanto che i modelli FIGARCH non si ridurranno ad essere soltanto “giocattoli” per econometrici, ma potranno avere successo e diffusione specialmente presso quegli operatori professionali che più hanno interesse alla stima della volatilità: cambisti, gestori di portafogli e analisti finanziari. Questa è la ragione per cui abbiamo ritenuto interessante studiarli in maniera approfondita.

### 6.2.3 Gli sviluppi futuri

Quello dell’integrazione frazionaria, è un filone di ricerca relativamente giovane, in particolar modo per quanto riguarda le applicazioni di tipo finanziario. Le problematiche teoriche che si sono sollevate nel corso dell’esposizione non appaiono essere di semplice soluzione, specie quelle relative alla teoria asintotica, e richiederanno ulteriori sforzi.

Le applicazioni più promettenti della modellistica a memoria lunga in finanza sono senza dubbio quelle relative ai dati ad altissima frequenza<sup>17</sup>, ovvero le singole transazioni che avvengono sui mercati registrate in tempo reale.

Fino a pochi anni fa la gestione di una così imponente mole di dati<sup>18</sup> era impensabile e si ricorreva pertanto a dati rilevati giornalmente. Con lo sviluppo di risorse hardware sempre più potenti è oggi possibile sia immagazzinare che analizzare statisticamente tali insiemi di dati. I primi studi in questa direzione (Andersen e Bollerslev, 1997a) hanno evidenziato come diverse misure di volatilità rilevate a livello intragiornaliero, ad esempio la differenza tra massimo e minimo (si veda anche Lombardi, 2002), presentino caratteristiche di memoria lunga. D'altra parte, in un contesto intragiornaliero, il più ampio orizzonte previsivo fornito da questi modelli assume un rilievo ancora maggiore e li rende preferibili rispetto alle tradizionali controparti a memoria corta.

La sfida, in questa direzione, è quella di sviluppare nuovi modelli che tengano conto sia delle caratteristiche empiriche dei dati (memoria lunga, eventuali ciclicità, effetti dovuti ai meccanismi di contrattazione specifici di ogni mercato), che del fatto che le osservazioni ad altissima frequenza non sono equispaziate. Anche in questo senso, si tratta di partire dall'osservazione dei dati e sviluppare una modellistica che fornisca allo stesso tempo un'adeguata rappresentazione degli stessi e una buona capacità previsiva.

---

<sup>17</sup> Si veda, per una semplice introduzione all'argomento, Cecconi e Lombardi (2001).

<sup>18</sup> Una tipica giornata di contrattazioni, su un mercato non eccessivamente ampio come quello italiano, supera abbondantemente le 1500 osservazioni; un database di un anno consterebbe quindi di circa 600000 rilevazioni.

# **Appendice A**

## **Altri strumenti utili**

## A.1 Alcune diagnostiche utili

### A.1.1 I criteri informativi

Abbiamo già rilevato come, quando ci si trovi a modellare una serie, ci si debba confrontare con il *trade-off* che passa tra la semplicità del modello e la sua capacità esplicativa. I criteri informativi sono statistiche che ci aiutano nella selezione del modello, identificando quello che risulta essere più parsimonioso senza perdere eccessivamente in precisione.

Il criterio di utilizzo più comune è dovuto ad Akaike (1973), che propone di utilizzare il modello per cui è minima la quantità:

$$AIC = -\frac{2\ln \ell}{T} + \frac{2k}{T}, \quad (\text{A.1.1})$$

dove  $T$  è il numero di osservazioni e  $k$  il numero di parametri che si utilizzano; si può osservare immediatamente che un numero di parametri elevato penalizza la (A.1.1).

Un altro criterio, che differisce da quello di Akaike per il maggior peso che viene dato alla parsimonia, è proposto da Schwarz (1978):

$$SC = -\frac{2\ln \ell}{T} + k \frac{\ln T}{T}. \quad (\text{A.1.2})$$

### A.1.2 I test sui residui

Il modello lineare classico ha come ipotesi di base la normalità e l'indipendenza dei residui.

Una volta specificato un modello, è utile verificare se tali ipotesi sono confermate sui residui di regressione.

Per quanto riguarda la normalità Jarque e Bera (1982) propongono<sup>1</sup> l'utilizzo della statistica:

---

<sup>1</sup> In realtà questo test è stato sviluppato da Bowman e Shenton (1975), ma in econometria e nei pacchetti di software econometrici è quasi sempre indicato col nome dei ricercatori che per primi ne proposero l'uso per l'analisi dei residui.

$$JB = \frac{T-r}{6} \left[ sk^2 + \frac{1}{4}(K-3)^2 \right], \quad (\text{A.1.3})$$

dove  $r$  indica il numero di regressori,  $sk$  la skewness dei residui e  $K$  la curtosi<sup>2</sup>. La (A.1.3), che sotto normalità vale zero, ha una distribuzione  $\chi^2_2$ .

Un test per l'autocorrelazione dei residui è invece proposto da Ljung e Box (1979) che introducono la statistica:

$$Q_k = T(T+2) \sum_{j=1}^k \frac{\hat{\rho}_j}{T-j}. \quad (\text{A.1.4})$$

Nel caso in cui non vi sia autocorrelazione tra i residui fino al ritardo  $k$ , la (A.1.4) deve valere zero: questa sarà l'ipotesi nulla. La (A.1.4) si distribuisce come un  $\chi^2_{k-r}$ , dove  $r$  è il numero di regressori che si sono utilizzati nel modello che ha generato i residui.

---

<sup>2</sup> La skewness e la curtosi sono rispettivamente il momento centrale terzo e quarto di una distribuzione. Per la normale si ha  $sk = 0$  e  $K = 3$ .

## A.2 Metodi di ottimizzazione<sup>3</sup>

Il secondo passo della procedura di stima di massima verosimiglianza di cui abbiamo parlato nel paragrafo 4.1.1, cioè la massimizzazione della funzione di verosimiglianza, è un tema estremamente delicato. In generale, infatti, le derivate prime della funzione di verosimiglianza (o di log-verosimiglianza) rispetto ai parametri non saranno lineari in detti parametri, precludendo così la possibilità di una soluzione semplice e immediata.

In questa circostanza l'idea che potrebbe venirci in mente è di “provare” diversi  $\hat{\theta}$ , utilizzando ad ogni tentativo un diverso incremento delle loro componenti, calcolare la verosimiglianza per ognuno di essi e scegliere infine il vettore “migliore”. Se i parametri da stimare sono due, e si “provano” dieci diversi valori per ogni parametro, si può immaginare di dover riempire una griglia di  $10^2$  caselle con i valori della verosimiglianza per ogni scelta dei parametri. Di qui il nome di questo metodo: *grid search*.

Purtroppo, all'aumentare del numero di parametri, il numero di caselle da riempire cresce a dismisura: se essi fossero  $n$ , e i valori possibili di ogni parametro  $k$ , ci troveremmo alle prese con un ipercubo composto da  $k^n$  caselle da riempire.

Questo ci suggerisce che abbiamo bisogno di fare ricorso a metodi di ottimizzazione numerici, più adeguati e più parsimoniosi.

### A.2.1 *Steepest ascent*

Supponiamo di poter disporre di un vettore di parametri  $\theta_0$  e di cercare di migliorare la stima a partire da esso, spostandoci di una distanza quadrata prefissata a  $k$ . Ci troveremo allora di fronte a un problema di ottimizzazione vincolata:

---

<sup>3</sup> Per quanto riguarda i primi tre paragrafi si è seguito Hamilton (1994), per i successivi Fletcher (2000). Dobbiamo comunque ringraziare Giovanni Della Lunga di Datamat S.p.A., che ci ha introdotto allo studio di queste problematiche.

$$\begin{cases} \max_{\boldsymbol{\theta}_1} \ell(\boldsymbol{\theta}_1) \\ (\boldsymbol{\theta}_1 - \boldsymbol{\theta}_0)'(\boldsymbol{\theta}_1 - \boldsymbol{\theta}_0) = k \end{cases}$$

di cui possiamo formare il lagrangiano:

$$J(\boldsymbol{\theta}_1) = \ell(\boldsymbol{\theta}_1) + \lambda [k - (\boldsymbol{\theta}_1 - \boldsymbol{\theta}_0)'(\boldsymbol{\theta}_1 - \boldsymbol{\theta}_0)], \quad (\text{A.2.1})$$

e ricavare la soluzione, che si ha per:

$$\frac{\partial J(\boldsymbol{\theta}_1)}{\partial \boldsymbol{\theta}_1} = \frac{\partial \ell(\boldsymbol{\theta}_1)}{\partial \boldsymbol{\theta}_1} - 2\lambda(\boldsymbol{\theta}_1 - \boldsymbol{\theta}_0) = \mathbf{0}; \quad (\text{A.2.2})$$

se indichiamo con  $\mathbf{g}(\boldsymbol{\theta}) = \frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$ , la (A.2.2) si può riscrivere come:

$$\boldsymbol{\theta}_1 = \boldsymbol{\theta}_0 + \frac{1}{2\lambda} \mathbf{g}(\boldsymbol{\theta}_1), \quad (\text{A.2.3})$$

il che equivale a dire che ci dobbiamo spostare da  $\boldsymbol{\theta}_0$ , in direzione  $\mathbf{g}(\boldsymbol{\theta}_1)$ , di un passo pari a

$$\frac{1}{2\lambda}.$$

Si può osservare subito come questa procedura sia attraente al di là del caso specifico (in genere non sarà importante fissare  $k$  a priori). Se invece di utilizzare il moltiplicatore di Lagrange utilizziamo un generico scalare  $s$  scelto a priori, sufficientemente piccolo per evitare passi troppo grandi, la (A.2.3) diventa:

$$\boldsymbol{\theta}_1 = \boldsymbol{\theta}_0 + s\mathbf{g}(\boldsymbol{\theta}_1), \quad (\text{A.2.4})$$

e può essere iterata finché la distanza tra  $\boldsymbol{\theta}_0$  e  $\boldsymbol{\theta}_1$  non si fa sufficientemente piccola. Uno dei principali limiti di questo tipo di approccio è che, utilizzando solo il gradiente, la convergenza può essere molto lenta.

### A.2.2 Newton-Raphson

L'idea che sta alla base del metodo Newton-Raphson è di considerare, invece dell'approssimazione lineare del metodo *steepest ascent*, un'approssimazione in serie di Taylor arrestata al secondo termine, attorno a  $\boldsymbol{\theta}_0$ , della funzione di verosimiglianza e di scegliere il  $\boldsymbol{\theta}_1$  che la massimizza. Per questa ragione questo metodo prende il nome di *curve-fit*. Sia l'Hessiana:

$$\mathbf{H}(\boldsymbol{\theta}_0) = -\frac{\partial^2 \ell(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}_0 \partial \boldsymbol{\theta}_0'}; \quad (\text{A.2.5})$$

l'approssimazione di Taylor sarà:

$$\ell(\boldsymbol{\theta}) = \ell(\boldsymbol{\theta}_0) + \mathbf{g}'(\boldsymbol{\theta}_0)(\boldsymbol{\theta} - \boldsymbol{\theta}_0) - \frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\theta}_0)' \mathbf{H}(\boldsymbol{\theta}_0)(\boldsymbol{\theta} - \boldsymbol{\theta}_0). \quad (\text{A.2.6})$$

Per massimizzare la (A.2.6) dobbiamo porre a zero la sua derivata rispetto a  $\boldsymbol{\theta}$ :

$$\frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \mathbf{g}(\boldsymbol{\theta}_0) - \mathbf{H}(\boldsymbol{\theta}_0)(\boldsymbol{\theta} - \boldsymbol{\theta}_0) = \mathbf{0}, \quad (\text{A.2.7})$$

da cui, inserendo un controllo della dimensione del passo:

$$\boldsymbol{\theta}_1 = \boldsymbol{\theta}_0 + s \mathbf{H}^{-1}(\boldsymbol{\theta}_0) \mathbf{g}(\boldsymbol{\theta}_0), \quad (\text{A.2.8})$$

che, di nuovo, dovrà essere iterata finché non si è raggiunto un qualche criterio di convergenza.

### A.2.3 DFP-BFGS

Nel metodo Newton-Raphson l'Hessiana dev'essere calcolata e invertita ad ogni iterazione, e ciò impegna senz'altro, specie se si utilizza un metodo numerico di derivazione, molta potenza di calcolo.

Davidon (1959) e Fletcher e Powell (1963) osservano che l'Hessiana è formata dalle derivate parziali dei gradienti, per cui ci dice come varia  $\mathbf{g}(\boldsymbol{\theta})$  al variare di  $\boldsymbol{\theta}$ ; delle informazioni sul



comportamento dell'Hessiana le possiamo perciò ricavare anche dal confronto tra  $\Delta \mathbf{g}(\boldsymbol{\theta})$  e  $\Delta \boldsymbol{\theta}$ .

Si può pertanto ricavare una formula per aggiornare, ad ogni iterazione, l'inversa dell'Hessiana, evitando così di doverla calcolare ad ogni passaggio.

La formula di aggiornamento per l'inversa dell'Hessiana data da:

$$\mathbf{A}_{m+1} = \mathbf{A}_m - \frac{\mathbf{A}_m \cdot \Delta \mathbf{g}_{m+1} \cdot \Delta \mathbf{g}'_{m+1} \cdot \mathbf{A}_m}{\Delta \mathbf{g}'_{m+1} \cdot \mathbf{A}_m \cdot \Delta \mathbf{g}_{m+1}} - \frac{\Delta \boldsymbol{\theta}_{m+1} \cdot \Delta \boldsymbol{\theta}'_{m+1}}{\Delta \mathbf{g}'_{m+1} \cdot \Delta \boldsymbol{\theta}_{m+1}}, \quad (\text{A.2.9})$$

mentre quella per  $\boldsymbol{\theta}$  diventa:

$$\boldsymbol{\theta}_{m+1} = \boldsymbol{\theta}_m + s \mathbf{A}_m \mathbf{g}(\boldsymbol{\theta}_m).$$

Una correzione alla (A.2.9) che ha dimostrato ottime performance all'atto pratico è stata proposta da Broyden (1970), Fletcher (1970), Goldfarb (1970) e Shanno (1970), che propongono di aggiungere alla (A.2.9) il termine:

$$\begin{aligned} & [\Delta \mathbf{g}'_{m+1} \cdot \mathbf{A}_m \cdot \Delta \mathbf{g}_{m+1}] \cdot \\ & \cdot \left[ \frac{\Delta \boldsymbol{\theta}_{m+1}}{\Delta \mathbf{g}'_{m+1} \cdot \Delta \boldsymbol{\theta}_{m+1}} - \frac{\mathbf{A}_m \cdot \Delta \mathbf{g}_{m+1}}{\Delta \mathbf{g}'_{m+1} \cdot \mathbf{A}_m \cdot \Delta \mathbf{g}_{m+1}} \right] \cdot \\ & \cdot \left[ \frac{\Delta \boldsymbol{\theta}_{m+1}}{\Delta \mathbf{g}'_{m+1} \cdot \Delta \boldsymbol{\theta}_{m+1}} - \frac{\mathbf{A}_m \cdot \Delta \mathbf{g}_{m+1}}{\Delta \mathbf{g}'_{m+1} \cdot \mathbf{A}_m \cdot \Delta \mathbf{g}_{m+1}} \right]', \end{aligned} \quad (\text{A.2.10})$$

generando così un nuovo metodo, variante del DFP, noto come BFGS.

#### A.2.4 BHHH

Un'ulteriore approssimazione per l'Hessiana è proposta da Berndt, Hall, Hall e Hausman (1974). Come abbiamo già visto nella (4.1.20), l'idea è che:

$$\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \cong \sum_{t=1}^T \mathbf{g}'_t(\boldsymbol{\theta}) \mathbf{g}_t(\boldsymbol{\theta}) = \tilde{\mathbf{H}}, \quad (\text{A.2.11})$$

da cui la (A.2.8) diventa:

$$\begin{aligned}\boldsymbol{\theta}_{m+1} &= \boldsymbol{\theta}_m + s \left\{ \sum_{t=1}^T \mathbf{g}'_t(\boldsymbol{\theta}_m) \mathbf{g}_t(\boldsymbol{\theta}_m) \right\}^{-1} \mathbf{g}(\boldsymbol{\theta}_m) = \\ &\boldsymbol{\theta}_m + s \tilde{\mathbf{H}}^{-1} \mathbf{g}(\boldsymbol{\theta}_m).\end{aligned}\tag{A.2.12}$$

### A.2.5 Marquardt

Un altro metodo che evita il calcolo dell'Hessiana è quello proposto da Marquardt (1963) e generalizzato da Goldfeld, Quandt e Trotter (1966). In sostanza si utilizza la procedura BHHH per approssimare l'Hessiana ma, per evitare problemi nei casi in cui il prodotto dei gradienti è quasi singolare, si aggiunge, agli elementi lungo la diagonale, un fattore correttivo variabile a seconda della distanza dal massimo. Se  $\tilde{\mathbf{H}}(\boldsymbol{\theta})$  è l'Hessiana approssimata della (A.2.11), si utilizzerà nella (A.2.8):

$$\overline{\mathbf{H}}(\boldsymbol{\theta}_m) = \tilde{\mathbf{H}}(\boldsymbol{\theta}_m) + \lambda_m \mathbf{I},\tag{A.2.13}$$

da cui:

$$\boldsymbol{\theta}_{m+1} = \boldsymbol{\theta}_m + s \overline{\mathbf{H}}^{-1} \mathbf{g}(\boldsymbol{\theta}_m).\tag{A.2.14}$$

Il fattore  $\lambda_m$  è di importanza cruciale per il metodo di Marquardt. Osservando la (A.2.13), si capisce subito che, se  $\lambda$  è zero, il metodo di Marquardt equivale al BHHH; se invece  $\lambda$  è molto elevato, l'Hessiana è dominante diagonalmente, e ciò rende la (A.2.13) analoga al prodotto del gradiente per una costante  $s$  che avevamo incontrato nel paragrafo 3.2.1. Quando siamo distanti dal massimo, vale la pena di fare passi ampi senza “sottilizzare”, e quindi utilizzare lo *steepest ascent* con un coefficiente alto; quando invece ci si avvicina al massimo, conviene procedere con cautela e utilizzare il BHHH.

Per essere più precisi, i passi che compie l'algoritmo di Marquardt sono i seguenti (si veda Press, Teukolsky, Vetterling e Flannery, 1992):

- Si parte con un valore di  $\lambda$  arbitrario e con un  $\boldsymbol{\theta}_0$  “ragionevole”;

- Si calcola  $\ell(\boldsymbol{\theta}_0)$ ;
- Si calcola  $\boldsymbol{\theta}_1$  a partire dalla (A.2.14);
- Si calcola  $\ell(\boldsymbol{\theta}_1)$  e lo si confronta con  $\ell(\boldsymbol{\theta}_0)$ :
  - Se è maggiore vuol dire che ci siamo mossi di una quantità adeguata, riduciamo  $\lambda$  e accettiamo  $\boldsymbol{\theta}_1$  come “nuova soluzione”;
  - Se è minore, ci siamo spostati troppo, aumentiamo  $\lambda$  e proviamo con un nuovo  $\boldsymbol{\theta}_1$ ;
- Si riparte dall’inizio finché la soluzione non soddisfa il criterio di convergenza.

#### A.2.6 Il controllo del passo

Abbiamo fin qui ipotizzato che il fattore che dà la dimensione del passo  $s$  sia scelto a priori. In realtà questa non è una mossa molto astuta, poiché scegliere un valore alto rischia di farci “saltare” sul massimo, mentre un valore troppo basso può generare un numero enorme di iterazioni. Si deve quindi utilizzare un algoritmo che, per ogni tentativo, sceglie il valore di  $s$  più adeguato. Seguendo Gouriéroux e Monfort (1997) osserviamo che tutti i metodi di ottimizzazione sopra presentati soggiacciono alla forma generale:

$$\boldsymbol{\theta}_{m+1} = \boldsymbol{\theta}_m + \lambda_m \mathbf{Q}(\boldsymbol{\theta}_m) \mathbf{g}(\boldsymbol{\theta}_m), \quad (\text{A.2.15})$$

dove ad esempio, per il metodo *steepest ascent*,  $\mathbf{Q}$  è la matrice identità. Trovare la dimensione ottimale del passo equivarrà a massimizzare  $\ell(\boldsymbol{\theta}_{m+1} | \boldsymbol{\theta}_m)$  rispetto a  $\lambda_k$ ; poiché si tratta della ricerca di un solo parametro, possiamo anche concederci il lusso di procedere in maniera semplice, applicando il metodo *grid search*.

Un’altra possibilità ingegnosa (si veda Press, Teukolsky, Vetterling e Flannery, 1992) è quella di creare una funzione dummy che assuma i valori della funzione di verosimiglianza sulla linea che parte da  $\boldsymbol{\theta}_m$  in direzione  $\mathbf{g}(\boldsymbol{\theta}_m)$ , si procede poi a massimizzare questa funzione (per

esempio con l'interpolazione parabolica) e si calcola la distanza del punto di massimo da  $\theta_m$ : tale distanza sarà l'ampiezza del prossimo passo. Ovviamente, nel caso dello *steepest ascent*, il punto di massimo sarà anche il nuovo “punto di partenza”  $\theta_{m+1}$ .

### A.2.7 Il problema dei valori iniziali

Chiudiamo questa sezione menzionando brevemente un altro problema computazionale che si deve affrontare.

In generale la funzione di verosimiglianza potrà avere più massimi locali; è possibile che gli algoritmi di massimizzazione che abbiamo introdotto siano “ingannati” e finiscano per indicarci un massimo locale invece che quello globale. Ciò è tanto più probabile quanti più massimi locali separano il punto iniziale dal massimo globale. Pertanto il punto di partenza dell'algoritmo di massimizzazione dev'essere scelto con particolare cura, evitando assolutamente di fornire alla macchina valori casuali<sup>4</sup>.

Ad esempio, nel caso della stima di massima verosimiglianza esatta di un processo ARFIMA, Sowell (1992a) propone di utilizzare come valore iniziale per  $d$  una sua stima ottenuta attraverso il metodo GPH, e per gli altri parametri quelli ottenuti da una regressione lineare senza considerare l'integrazione frazionaria.

Esistono a dire il vero metodi di ottimizzazione globali, che esplorano cioè tutto lo spazio dei parametri senza dipendere pertanto dalla scelta dei valori iniziali: tali metodi prendono il nome di algoritmi genetici. Un'idea sommaria<sup>5</sup> sul loro funzionamento si può avere considerando il vettore di parametri come un cromosoma composto da vari geni (ciascun parametro), e ipotizzando di riprodurre, a partire da una popolazione iniziale di cromosomi, il meccanismo di selezione naturale: i cromosomi vengono selezionati per la riproduzione sulla

---

<sup>4</sup> O, quando questa è l'unica strada percorribile, almeno verificando la fondatezza dei risultati “provando” diversi vettori di valori iniziali.

base delle loro caratteristiche di adattamento all'ambiente (e quindi, nel caso dell'ottimizzazione di una funzione, sulla base del valore assunto dalla stessa nel punto in questione), si riproducono accoppiandosi e subiscono mutazioni. Sulla base di questo procedimento si giunge, dopo un certo numero di "popolazioni" (ovvero di iterazioni dell'algoritmo), ad avere una "specie selezionata" e dotata di caratteristiche ottime.

La trattazione di questo argomento esula comunque dallo scopo di questo lavoro, per cui si rimanda a Mitchell (1998).

---

<sup>5</sup> Ci si è ispirati alla presentazione di Cherubini e Della Lunga (2001).

# **Appendice B**

## **Codice sorgente**

## B.1 Una panoramica introduttiva

### *B.1.1 La programmazione e l'econometria*

La statistica ha risentito (esattamente come la fisica, volendo proseguire il parallelismo azzardato nell'introduzione) in maniera estremamente positiva dell'avvento dei calcolatori programmabili e della rapida crescita della loro potenza. Operazioni che un tempo richiedevano ore di calcoli manuali possono venire effettuate in pochi secondi e, quel che è più importante, operazioni un tempo impensabili sono ora possibili (si pensi a dover trattare manualmente una serie di migliaia di osservazioni!).

Un buono statistico però non può limitarsi a sapere usare gli strumenti di calcolo predefiniti e preparati da altri. Un buono statistico è anche, necessariamente, un programmatore: chi propone un nuovo modello non può permettersi di dipendere da altri per quanto riguarda la sua implementazione.

Questa appendice non ha certo la pretesa di trattare in maniera neppure superficiale l'immenso argomento della programmazione; ci limiteremo a presentare sommariamente gli strumenti più utilizzati in ambito econometrico e quelli a cui si è fatto ricorso nello svolgimento di questo lavoro.

### *B.1.2 Gli ambienti di sviluppo principali*

Lasciando da parte gli applicativi veri e propri (anche se, come Econometric Views, sono in parte programmabili), possiamo dire che l'ambiente di sviluppo di elezione per l'analisi delle serie storiche (e, più in generale, per l'econometria) è GAUSS. Si tratta di un linguaggio di programmazione che può essere sia compilato che interpretato, e che ha il grosso vantaggio di trattare le matrici come struttura di dati nativa. GAUSS è quasi uno standard per la

programmazione econometrica e statistica, ed è disponibile un'enorme quantità di codice di pubblico dominio (ad esempio, relativamente al FIGARCH, il codice sviluppato da Laurent, 1999, per il calcolo della serie delle varianze condizionate). Uno dei principali punti deboli di questo ambiente di sviluppo è l'assenza di funzioni di *debug* avanzate, il che rende l'individuazione degli errori estremamente problematica ed aleatoria.

Più vicino a un linguaggio vero e proprio, piuttosto che a un applicativo programmabile, è l'ambiente Ox, sviluppato da Doornik (1999). Si tratta di un linguaggio compilato molto simile al C++, ma in grado di trattare le matrici in maniera nativa. Anche per Ox sono disponibili numerosi pacchetti per l'analisi di serie storiche (ad esempio quello sviluppato da Doornik e Ooms, 1999, per la stima di processi ARFIMA).

### *B.1.3 Il pacchetto G@RCH 1.1*

Quello che noi abbiamo utilizzato per le prime stime FIGARCH presentate nel capitolo 6 è appunto un pacchetto Ox: G@RCH 1.1, sviluppato da Laurent e Peters (2000).

Questo pacchetto lascia all'utilizzatore numerose opzioni per la stima di varie specificazioni ARMA-GARCH, con la possibilità di utilizzare l'integrazione frazionaria sia nell'equazione della media che in quella della varianza. Senza scendere in ulteriori dettagli sulle funzionalità di questo pacchetto, per le quali si rimanda a Laurent e Peters (2000), soffermiamoci su alcuni aspetti importanti relativi alle stime FIGARCH.

Osserviamo in primo luogo che il pacchetto G@RCH 1.1 utilizza, come metodo di ottimizzazione, il BFGS che abbiamo presentato nel paragrafo A.2.3; il gradiente e l'Hessiana sono calcolati per approssimazione numerica. Le stime di alcuni modelli GARCH tradizionali sono state confrontate con quelle ottenute tramite l'applicativo Econometric Views 3.1, riscontrando una performance di stima accettabile. In generale però si deve rilevare che il metodo di ottimizzazione di Marquardt (1963), che abbiamo utilizzato per tutte le stime in



Eviews, ha dato risultati leggermente migliori. Ad esempio, volendo lavorare sul modello della tabella 6.8:

<i>Variabile</i>	<i>Eviews</i>	<i>G@RCH 1.1</i>
$\mu$	- 0.00392	- 0.00381
$\omega$	0.071953	0.071563
$1-D$	- 0.074833	- 0.074261
$\alpha$	0.108756	0.108945
$\beta_1$	0.397078	0.398593
$\beta_2$	0.462078	0.460282
<i>Log-verosimiglianza</i>	- 5435.074	- 5435.162

**Tabella B.1**

*Confronto tra i risultati di Eviews e quelli di G@RCH 1.1*

A causa di questa, seppur lieve, minore efficienza si è preferito, per quanto concerne le stime GARCH “tradizionali”, lavorare con Eviews.

#### *B.1.4 Il linguaggio C++*

Tra i linguaggi di programmazione veri e propri, il C (Kernighan e Ritchie, 1988), con le successive estensioni C++ (Stroustrup, 1994), rappresenta uno standard riconosciuto: la quasi totalità degli applicativi oggi in commercio sono sviluppati in C++. Il C++ è, tra tutti i linguaggi di programmazione di alto livello, quello di livello più basso, che consente quindi il maggior livello di controllo e di interazione sui processi che avvengono all’interno della macchina; è inoltre un linguaggio portabile e con un’enorme precisione e rapidità di esecuzione.

Abbiamo già detto che il C++ è un linguaggio di livello più basso rispetto agli altri comunemente utilizzati: ciò lo rende indubbiamente più complesso da apprendere e, soprattutto, da utilizzare. Le uniche operazioni che il C++ può eseguire in maniera nativa sono esattamente quelle che il processore può svolgere: operazioni aritmetiche e logiche elementari. In un contesto di questo tipo un’operazione semplice come il prodotto tra matrici,

che in GAUSS o in Ox viene eseguita con un semplice comando, richiede la scrittura di diverse righe di codice per “spiegare” al processore come effettuare tale operazione. Ma questa iniziale complessità è uno dei principali punti di forza del C++, perché implica che ogni cosa debba essere fatta in maniera molto attenta e pedante, riducendo così le possibilità di compiere errori logici.

I moderni compilatori C++ sono indubbiamente superiori agli altri strumenti di programmazione sia per la potenza degli strumenti di *debug*, che per la possibilità di creare programmi sotto forma non solo di file eseguibili, ma di oggetti di diverso tipo. Una delle caratteristiche più interessanti del C++ (o meglio, dei compilatori C++) sotto questo profilo è la possibilità di creare programmi sotto forma di *dll* (*dynamic linked library*), cioè programmi che vengono richiamati, ricevono input e restituiscono l’output ad altre applicazioni. In questo senso si potrebbe dire che il programma è “nascosto” dall’applicativo principale. In questo modo si possono creare programmi interfacciati e gestiti interamente dagli applicativi più diffusi, come Microsoft Word o Microsoft Excel, nei quali l’utente si trova quindi a contatto solo con l’ambiente a lui familiare e non ha necessità di apprendere l’utilizzo di ulteriori strumenti.

Il nostro programma per la stima FIGARCH<sup>1</sup>, che abbiamo qui presentato in versione eseguibile, potrebbe quindi essere facilmente convertito in *dll*, e potrebbe così lavorare ed essere richiamato direttamente dal foglio di lavoro Excel contenente i dati sui quali si desidera condurre la stima. Ci sembra che questa caratteristica sia particolarmente rilevante, visto che buona parte delle istituzioni finanziarie usano Excel come strumento di trattamento dati. Crediamo quindi che l’aver sviluppato una routine di stima in un linguaggio di programmazione più “vicino” alla realtà e agli strumenti di cui usualmente si servono gli

---

<sup>1</sup> A questo proposito non potrò mai ringraziare a sufficienza Giovanni Della Lunga e Marianna Marziliani di Datamat S.p.A, senza la cui pazienza e disponibilità non avrei mai potuto apprendere l’arte della programmazione C++.

operatori professionali possa costituire un valido contributo alla diffusione della modellistica FIGARCH presso questi ultimi.

## B.2 Codice C++

Presentiamo in questa sezione il codice che si è sviluppato per la stima di un modello FIGARCH (1,  $d$ , 1). Per quanto riguarda alcuni calcoli (ad esempio la funzione gamma, il calcolo numerico delle derivate e le routine di ottimizzazione) si è utilizzato, seppur apportando alcune modifiche, il codice proposto da Press, Teukolsky, Vetterling e Flannery (1992). Nonostante siano di utilizzo libero (almeno per i possessori del testo), non ci è sembrato serio pubblicare questi listati, magari presumendo che le nostre piccole modifiche li abbiano resi differenti dagli originali, per cui, per quanto riguarda queste funzioni, ci si limiterà a presentare i prototipi.

Si sono sviluppate due versioni del programma: una con la dummy, l'altra senza. Presenteremo solo la prima versione, poiché ricavare la seconda è assolutamente triviale. Ugualmente procederemo in seguito: per lo sviluppo di questo programma si sono scritte migliaia di righe di codice costituito in buon parte da diverse varianti di funzioni simili; non ci è parso opportuno appesantire questa tesi con del codice privo di funzioni illustrative. Si è deciso pertanto di presentare solo le funzioni “madri” e i prototipi delle loro varianti, illustrandone a parole le differenze. Tutti i listati sono comunque disponibili nel dischetto allegato.

### *B.2.1 La serie della varianza condizionata*

La prima routine che presentiamo calcola, a partire da un vettore di parametri `param` e dalla serie dei residui al quadrato `ressq`, la serie della varianza condizionata.

```
#include <math.h>

void figarchvc(double *ressq,    /*Residui al quadrato*/
               double *h,       /*Varianza condizionata, output*/
               double *param,    /*Vettore dei parametri*/
```

```

        int *dum,          /*Vettore della dummy*/
        double mrq,        /*Media dei residui al quadrato*/
        int nob,           /*Numerosità del campione*/
        int ntr)           /*Troncamento del termine  $(1-L)^d$ */

{
    double *phi,           /*Vettore dei coefficienti phi*/
           *ee,             /*Vettore dei residui corretti*/
           *zeta;           /*Vettore di appoggio*/
    double w, a, b, d, D, den; /*Altre dichiarazioni*/
    int t, k;

    ee=new double [nob+ntr]; /*Allocazione dei vettori*/
    zeta=new double [nob];
    phi=new double [ntr];

    phi[0]=1;              /*Primo valore di phi*/
    w=param[1];             /*Si "dà nome" ai parametri*/
    a=param[2];
    b=param[3];
    d=param[4];
    D=param[5];
    den=refgamma(-d);       /*Gamma(-d), fissato per evitare
                             ntr+nob chiamate*/

    for(k=1;k<ntr;k++)      /*Calcolo dei coefficienti phi*/
        phi[k]=exp(lngamma(k-d)-lngamma(k+1))/den;

    for(t=0;t<nob+ntr;t++)  /*Calcolo dei residui corretti:*/
    {
        if(t<ntr)           /*valori pre-campionari,*/
            ee[t]=mrq;
        else                 /*valori del campione.*/
            ee[t]=ressq[t-ntr];
    }

    for(t=0;t<nob;t++)      /*Calcolo del vettore zeta:*/
    {
        zeta[t]=0;          /*inizializzazione,*/
        for(k=0;k<ntr;k++)  /*filtro  $(1-L)^d$ .*/
            zeta[t]+=phi[k]*ee[t+ntr-k];
    }

    /*Calcolo del vettore della varianza condizionata*/
    h[0]=w+ee[ntr]-zeta[0];
    for(t=1;t<nob;t++)
        h[t]=w+ee[t+ntr]-zeta[t]+b*h[t-1]-b*ee[t+ntr-1]
            +a*zeta[t-1]+D*dum[t];

    delete [] ee;           /*Liberazione dei vettori*/
    delete [] zeta;
    delete [] phi;
}

```

### B.2.2 La funzione di verosimiglianza

Questa funzione calcola l'opposto (questo perché le routine standard di ottimizzazione prevedono la minimizzazione e non la massimizzazione di funzioni) della log-verosimiglianza del campione. Gli argomenti sono un vettore di parametri `param`, la serie dei residui `resid`, la numerosità del campione `nobs` e il termine `ntr` a cui arrestarsi nel troncamento dell'espansione del *fractional difference operator*  $(1-L)^d$ ; il valore della funzione è restituito alla chiamata. Si utilizza un contatore `call` che viene incrementato di uno ogni volta che la funzione viene chiamata.

```
#include <math.h>
#define PI 3.1415926535897932384626433832795

double figarchll(double *y,          /*Vettore delle osservazioni*/
                 double *param,      /*Vettore dei parametri*/
                 int *dum,           /*Vettore della dummy*/
                 int *call,          /*Numero delle chiamate*/
                 int nobs,           /*Numerosità della serie*/
                 int ntr)            /*Troncamento del termine (1-L)^d*/
{
    double *h,                      /*Varianza condizionata*/
            *ressq;                  /*Residui al quadrato*/
    double loglik, lnp, mrq, m; /*Altre dichiarazioni*/
    int t;

    *call+=1;                        /*Incremento del contatore*/

    loglik=0;                        /*Inizializzazioni*/
    mrq=0;
    lnp=log(2*PI);
    m=param[0] ;
    h=new double [nobs];             /*e allocazioni.*/
    ressq=new double [nobs];

    for(t=0;t<nobs;t++)              /*Calcolo dei residui al quadrato*/
    {
        ressq[t]=(y[t]-m)*(y[t]-m);
        mrq+=ressq[t];              /*della loro somma,*/
    }
    mrq/=nobs;                      /*e della media.*/

    /*Calcolo della varianza condizionata*/
    figarchvc(ressq,h,param,nobs,ntr,mrq);

    for(t=0;t<nobs;t++)              /*Calcolo di -Log-verosimiglianza*/
        loglik+=lnp+log(h[t])+ressq[t]/h[t];
    loglik*=0.5;
```

```

delete [] h;           /*Liberazione dei vettori*/
delete [] ressq;
return loglik;         /*Output*/
}

```

La funzione `figarchllvec` è identica a `figarchll`, ma calcola la verosimiglianza di una singola osservazione `t` passata come argomento. Il suo prototipo è:

```

double figarchllvec(double *y,      /*Vettore delle osservazioni*/
double *param,    /*Vettore dei parametri*/
int *dum,         /*Vettore della dummy*/
int t,           /*Osservazione di cui si chiede
la verosimiglianza*/
int nobs,        /*Numerosità della serie*/
int ntr)         /*Troncamento del termine (1-L)d*/

```

### B.2.3 Il gradiente analitico

Le singole derivate del gradiente analitico sono state presentate nella (5.2.36), nella (5.2.38) e nella (5.2.39), che per comodità riportiamo di seguito, applicata al FIGARCH (1,  $d$ , 1):

$$\begin{aligned}
\frac{\partial \varepsilon_t}{\partial \mu} &= -1, \\
\frac{\partial \varphi_k}{\partial d} &= -\varphi_k \sum_{j=1}^k \frac{1}{k-j-d}, \\
\frac{\partial h_t}{\partial \mu} &= \left[ 1 - \frac{(1-\psi L)(1-L)^d}{1-\beta L} \right] \cdot 2\varepsilon_t \frac{\partial \varepsilon_t}{\partial \mu}, \\
\frac{\partial h_t}{\partial \omega} &= \frac{1}{1-\beta}, \\
\frac{\partial h_t}{\partial \beta} &= \frac{h_{t-1} - \varepsilon_{t-1}^2}{1-\beta L}, \\
\frac{\partial h_t}{\partial \psi} &= \frac{1}{1-\beta L} \sum_{k=0}^M \varphi_k \varepsilon_{t-k-1}^2, \\
\frac{\partial h_t}{\partial d} &= -\frac{1-\psi L}{1-\beta L} \sum_{k=1}^M \frac{\partial \varphi_k}{\partial d} \varepsilon_{t-k}^2, \\
\frac{\partial h_t}{\partial \delta} &= \frac{D_t}{1-\beta L}.
\end{aligned} \tag{B.2.1}$$

La funzione `agrad` restituisce il gradiente della funzione di verosimiglianza, calcolato analiticamente, nel punto  $x$ .

```
#include <math.h>

void agrad(double *x,      /*Vettore dei parametri*/
           double *g,      /*Vettore del gradiente - output*/
           double *y,      /*Vettore delle osservazioni*/
           int *dum,       /*Vettore della dummy*/
           int nVar,       /*Numero dei parametri*/
           int nobs,       /*Numerosità della serie*/
           int ntr)        /*Troncamento del termine  $(1-L)^d$ */

{
    double *res,           /*Vettore dei residui*/
           *ressq,         /*Vettore dei residui al quadrato*/
           *zeta,          /*Vettore d'appoggio*/
           *h,             /*Varianza condizionata*/
           *ee,            /*Residui corretti*/
           *eesq,          /*Residui corretti al quadrato*/
           *phi,           /*Coefficienti phi*/
           *dphi,          /*Derivate dphi/dd*/
           *dhda,          /*Derivate dh/da*/
           *dhdb,          /*Derivate dh/db*/
           *dhdd,          /*Derivate dh/dd*/
           *dhdD,          /*Derivate dh/dD*/
           *dhdm;          /*Derivate dh/dm*/

    /*Altre dichiarazioni - dhdw è la derivata (costante per ogni
       osservazione) della varianza condizionata rispetto
       alla costante w*/
    double mrq, dhdw, den, m, w, a, b, d, D;
    int j, k, t;

    ee=new double[nobs+ntr]; /*Allocazione dei vettori*/
    eesq=new double[nobs+ntr];
    ressq=new double[nobs];
    res=new double[nobs];
    zeta=new double[nobs];
    h=new double[nobs];
    phi=new double[ntr];
    dphi=new double[ntr];
    dhda=new double[nobs];
    dhdb=new double[nobs];
    dhdd=new double[nobs];
    dhdD=new double[nobs];
    dhdm=new double[nobs];

    mrq=0;                      /*Inizializzazioni*/
    m=x[0];
    w=x[1];
    a=x[2];
    b=x[3];
    d=x[4];
    D=x[5];

    for(t=0;t<nobs;t++)
    {
```



```

        res[t]=y[t]-m;          /*Calcolo dei residui,*/
        ressq[t]=res[t]*res[t]; /*dei loro quadrati,*/
        mrq+=ressq[t];         /*e della media.*/
    }
    mrq/=nobs;

    for(t=0;t<nobs+ntr;t++)      /*Calcolo dei residui corretti:*/
    {
        if(t<ntr)                /*valori pre-campionari*/
        {
            ee[t]=0;
            eesq[t]=mrq;
        }
        else                      /*e valori del campione.*/
        {
            ee[t]=res[t-ntr];
            eesq[t]=ressq[t-ntr];
        }
    }

    /*Calcolo delle varianze condizionate*/
    figarchvc(ressq,h,x,dum,mrq,nobs,ntr);

    /*Calcolo dei coefficienti phi*/
    phi[0]=1;
    den=refgamma(-d);
    for(k=1;k<ntr;k++)
        phi[k]=exp(lngamma(k-d)-lngamma(k+1))/den;

    /*Calcolo del vettore d'appoggio*/
    for(t=0;t<nobs;t++)
    {
        zeta[t]=0;
        for(k=0;k<ntr;k++)
            zeta[t]+=phi[k]*ee[t+ntr-k];
    }

    /*Calcolo delle derivate secondo la (B.2.1):*/
    /*dphi/dd*/
    for(k=0;k<ntr;k++)
    {
        dphi[k]=0;
        for(j=1;j<=k;j++)
            dphi[k]+=1/(k-j-d);
        dphi[k]*=-phi[k];
    }

    /*dh/dw*/
    dhdw=1/(1-b);

    /*dh/da*/
    dhda[0]=0;
    for(k=0;k<ntr;k++)
        dhda[0]+=phi[k]*eesq[ntr-k-1];
    for(t=1;t<nobs;t++)
    {
        dhda[t]=b*dhda[t-1];
        for(k=0;k<ntr;k++)
            dhda[t]+=phi[k]*eesq[t+ntr-k-1];
    }

```

```

/*dh/db*/
dhdb[0]=-eesq[ntr-1];
for(t=1;t<nobs;t++)
    dhdb[t]=b*dhdb[t-1]+h[t-1]-eesq[t+ntr-1];

/*dh/dd*/
dhdd[0]=0;
for(k=1;k<ntr-1;k++)
    dhdd[0]+=-dphi[k]*eesq[ntr-k]+a*dphi[k]*eesq[ntr-k-1];
for(t=1;t<nobs;t++)
    {
        dhdd[t]=b*dhdd[t-1];
        for(k=1;k<ntr;k++)
            dhdd[t]+=-dphi[k]*eesq[t+ntr-k]
                +a*dphi[k]*eesq[t+ntr-k-1];
    }

/*dh/dD*/
dhdD[0]=dum[0];
for(t=1;t<nobs;t++)
    dhdD[t]=b*dhdD[t-1]+dum[t];

/*dh/dm*/
dhdm[0]=2*zeta[0]-2*res[0];
for(t=1;t<nobs;t++)
    dhdm[t]=b*dhdm[t-1]-2*res[t]+2*b*res[t-1]
        +2*zeta[t]-2*a*zeta[t-1];

for(j=0;j<nVar;j++)          /*Inizializzazione del gradiente*/
    g[j]=0;

for(t=0;t<nobs;t++)          /*Riempimento del gradiente*/
    {
        g[0]-=res[t]/h[t]+0.5*dhdm[t]*(ressq[t]-h[t])/(h[t]*h[t]);
        g[1]-=0.5*dhdw*(ressq[t]-h[t])/(h[t]*h[t]);
        g[2]-=0.5*dhda[t]*(ressq[t]-h[t])/(h[t]*h[t]);
        g[3]-=0.5*dhdb[t]*(ressq[t]-h[t])/(h[t]*h[t]);
        g[4]-=0.5*dhdd[t]*(ressq[t]-h[t])/(h[t]*h[t]);
        g[5]-=0.5*dhdD[t]*(ressq[t]-h[t])/(h[t]*h[t]);
    }

delete [] ee;                /*Liberazione dei vettori utilizzati*/
delete [] eesq;
delete [] res;
delete [] ressq;
delete [] zeta;
delete [] h;
delete [] phi;
delete [] dphi;
delete [] dhda;
delete [] dhdb;
delete [] dhdd;
delete [] dhdD;
delete [] dhdm;
}

```

La funzione `agradmat` è analoga ad `agrad`, ma restituisce i gradienti rispetto alle singole osservazione inseriti in una matrice `G`. Questo è il suo prototipo:

```
void agradmat(double *x,      /*Vettore dei parametri*/
             double **G,    /*Matrice dei gradienti - output*/
             double *y,      /*Vettore delle osservazioni*/
             int *dum,       /*Vettore della dummy*/
             int nVar,       /*Numero dei parametri*/
             int nobs,       /*Numerosità della serie*/
             int ntr)        /*Troncamento del termine (1-L)d*/
```

#### *B.2.4 Il gradiente numerico*

Il calcolo numerico del gradiente della funzione di verosimiglianza è effettuata dalla funzione `ngrad`:

```
void ngrad(double (*func)    /*Funzione da derivare*/
           (double*, /*e suoi argomenti.*/
            double*,
            int*,
            int*,
            int,
            int),
           double *x,        /*Vettore dei parametri*/
           double *g,        /*Vettore del gradiente - output*/
           double *y,        /*Vettore delle osservazioni*/
           int *dum,         /*Vettore della dummy*/
           int *call,        /*Contatore delle chiamate a func*/
           int nVar,         /*Numero di parametri*/
           int nobs,         /*Numerosità della serie*/
           int ntr)          /*Troncamento del termine (1-L)d*/
{
    int i, j;
    int* dummy;

    dummy=new int [nVar]; /*Allocazione del vettore delle dummy*/
    for (i=0;i<nVar;i++) /*e sua inizializzazione.*/
        dummy[i]=0;

    for (j=0;j<nVar;j++)
    {
        dummy[j]=1; /*Per ogni dimensione j si pone a 1 la dummy,*/
        g[j]=dfdrdr(func,x,y,dum,dummy,call,nVar,nobs,ntr);
        /*si calcola la derivata della funzione di verosimiglianza
           rispetto alla variabile indicata dalla dummy,*/
        dummy[j]=0; /*e si azzerla la dummy utilizzata.*/
    }

    delete [] dummy;
}
```

L'analogo della funzione `agradmat` per il gradiente numerico è `ngradvec`, che restituisce il gradiente rispetto a una singola osservazione `t`. Il funzionamento di questa funzione è estremamente semplice, perché basta utilizzare, al posto di `figarchll`, la funzione `figarchllvec` e calcolarne la derivata numerica; ci si limiterà pertanto a presentarne il prototipo:

```
void ngradvec(double (*func) /*Funzione da derivare*/
              (double*, /*e suoi argomenti.*/
               double*,
               int*,
               int,
               int,
               int),
              double *x, /*Vettore dei parametri*/
              double *g, /*Vettore del gradiente - output*/
              double *y, /*Vettore delle osservazioni*/
              int *dum, /*Vettore della dummy*/
              int t, /*Istante di calcolo del gradiente*/
              int nVar, /*Numero di parametri*/
              int nobs, /*Numerosità della serie*/
              int ntr) /*Troncamento del termine (1-L)^d*/
```

Vale la pena osservare, prima di procedere, che l'utilizzo di questa funzione è altamente sconsigliato: difatti l'algoritmo di calcolo numerico della derivata non è ottimizzato per questo tipo di problemi, e impiega, su un processore Intel Pentium III a 500Mhz, circa dieci secondi per ogni osservazione. Pertanto, il calcolo della matrice  $G$ , che è effettuato dalla routine `agradmat` in pochi secondi, richiederebbe, volendo utilizzare questa routine, circa quattordici ore per una serie di cinquemila osservazioni.

### *B.2.5 L'Hessiana*

Per il calcolo numerico dell'Hessiana si è sviluppata la routine `hess`, che fa utilizzo di una versione leggermente modificata, al fine di calcolare le derivate parziali e seconde, della routine di derivazione numerica `dfriidr`. Abbiamo chiamato tale routine `dfpriidr` (si veda il paragrafo B.2.8 per i prototipi).

```

void hess(double (*func) /*Funzione da derivare*/
          (double*, /*e suoi argomenti.*/*
          double*,
          int*,
          int*,
          int,
          int),
          double *x, /*Vettore dei parametri*/
          double **H, /*Matrice Hessiana - output*/
          double *y, /*Vettore delle osservazioni*/
          int *dum, /*Vettore della dummy*/
          int *call, /*Contatore delle chiamate a func*/
          int nVar, /*Numero di parametri*/
          int nobs, /*Numerosità della serie*/
          int ntr) /*Troncamento del termine (1-L)d*/
{
    int *dummy1, *dummy2; /*Dichiarazioni*/
    int i, j;

    dummy1=new int[nVar]; /*Allocazione dei vettori dummy*/
    dummy2=new int[nVar];

    for (i=0;i<nVar;i++)
    {
        dummy1[i]=1; /*Selezione della prima variabile,*/
        for (j=0;j<=i;j++)
        {
            dummy2[j]=1; /*e della seconda.*/
            /*Calcolo della derivata parziale rispetto alle
            variabili indicate da dummy1 e dummy2*/
            H[i][j]=dfpridr(func,x,y,dum,dummy1,
                            dummy2,call,nVar,nobs,ntr);
            dummy2[j]=0; /*Ripristino delle dummy*/
        }
        dummy1[i]=0;
    }

    for (i=0;i<nVar;i++) /*Riempimento del triangolo superiore*/
        for (j=i+1;j<nVar;j++)
            H[i][j]=H[j][i];

    delete [] dummy1;
    delete [] dummy2;
}

```

## B.2.6 Il main

Col termine main si indica, nel linguaggio C++, la parte centrale del programma, a partire dalla quale ha inizio l'esecuzione.

```

#include "Figarch.h"
#define M 1000 /*Troncamento del termine (1-L)d*/
#define T 5705 /*Numero di osservazioni*/
#define F 100 /*Numero di previsioni*/

```

```

#define N 6                /*Numero di parametri*/
#define PTS 300            /*Punti del periodogramma*/
#define TOL 1.0e-6        /*Livello di tolleranza per l'ottimizzazione*/
#define ANALYTICAL 1      /*Switch per le derivate analitiche*/

void main (void)
{
    double theta[N],      /*Parametri*/
           se[N],         /*Standard errors*/
           tstat[N],      /*Statistiche t*/
           pval[N],       /*P-values*/
           rse[N],        /*Standard errors robusti*/
           rtstat[N],     /*Statistiche t robuste*/
           rpval[N],      /*P-values robusti*/
           var[9],        /*Value at risk*/
           ll,            /*Log-verosimiglianza*/
           mrq,           /*Media dei residui al quadrato*/
           elaps, min,    /*Tempo trascorso in secondi e minuti*/
           med, sdev,     /*Media e deviazione standard della serie*/
           aic, sc;       /*Criteri informativi di Akaike e Schwarz*/
    double *r,            /*Osservazioni*/
           *rsq,          /*Residui al quadrato*/
           *acov,         /*Autocovarianze*/
           *acor,         /*Autocorrelazioni*/
           *spec,         /*Periodogramma*/
           *omega,        /*Frequenze del periodogramma*/
           *grad,         /*Gradiente*/
           *fcst,         /*Previsioni sulla varianza condizionata*/
           *gfcst,        /*Previsioni GARCH - input da EViews*/
           *ht;           /*Autocovarianze*/
    double **cov,         /*Matrice di varianze e covarianze robusta*/
           **Iop,         /*Matrice di varianze e covarianze BHHH*/
           **I2d,         /*Matrice Hessiana*/
           **Ipr, **G, **Gt;
    int *dum;             /*Dummy*/
    int i, t,
        iter,            /*Iterazioni della routine di ottimizzazione*/
        call;            /*Chiamate alla funzione di verosimiglianza*/
    FILE *data, *dummy, /*File utilizzati*/
          *fore, *facf, *period, *cvar, *output, *gfore;
    clock_t start, finish;

    start=clock();        /*Partenza dell'orologio*/
    sdev=med=mrq=ll=iter=call=0;

    r=new double[T];      /*Allocazioni*/
    dum=new int[T+F];
    ht=new double[T];
    rsq=new double[T];
    acov=new double[T];
    acor=new double[T];
    fcst=new double[F];
    gfcst=new double[F];
    spec=new double[PTS];
    omega=new double[PTS];
    grad=new double[N];
    cov=new double*[N];
    Iop=new double*[N];
    I2d=new double*[N];
    Ipr=new double*[N];

```

```

Gt=new double*[N];
for(i=0;i<N;i++)
{
    cov[i]=new double[N];
    Iop[i]=new double[N];
    I2d[i]=new double[N];
    Ipr[i]=new double[N];
    Gt[i]=new double[T];
}
G=new double*[T];
for(t=0;t<T;t++)
    G[t]=new double[N];

/*Stampa dell'intestazione sul file di output e a schermo*/
output=fopen("c:\\documenti\\uni\\tesi\\data\\FIGARCHD.out","w");
printf("FIGARCH (1,d,1) estimation routine - Executable V. 1.2D\n");
printf("(C) Marco J. Lombardi, 16/12/2000\n");
fprintf(output,"FIGARCH (1,d,1) estimation routine - Executable V.
1.2D\n");
fprintf(output,"(C) Marco J. Lombardi, 16/12/2000\n");

/*Lettura dei dati*/
data=fopen("c:\\documenti\\uni\\tesi\\data\\reu.txt","rt");
for(t=0;t<T;t++)
    fscanf(data,"%lf",&r[t]);
fclose(data);
dummy=fopen("c:\\documenti\\uni\\tesi\\data\\eudum.txt","rt");
for(t=0;t<T+F;t++)
{
    if(t<T)                /*Osservazioni del campione*/
    {
        fscanf(dummy,"%d",&dum[t]);
        dum[t]=1-dum[t];/*"Inversione" della dummy*/
    }
    else
        dum[t]=dum[t-5];/*Osservazioni future*/
}
fclose(dummy);

for(t=0;t<T;t++)          /*Calcolo delle statistiche descrittive*/
{
    rsq[t]=r[t]*r[t];
    mrq+=rsq[t];
    med+=r[t];
}
mrq/=T;
med/=T;
for(t=0;t<T;t++)
    sdev+=(r[t]-med)*(r[t]-med);
sdev/=T;
sdev=sqrt(sdev);
printf("\nDATASET: $/E daily returns\nSample 28/12/1978 -
08/11/2000\n");
printf("Mean = %f Standard error = %f\nMean of squares = %f\n",
med, sdev, mrq);
fprintf(output,"\nDATASET: $/E daily returns\nSample 28/12/1978 -
8/11/2000\n");
fprintf(output,"Mean = %f Standard error = %f\nMean of squares =
%f\n",med,sdev,mrq);

```

```

/*Calcolo e stampa delle autocorrelazioni*/
acf(rsq,acov,acor,T);
facf=fopen("c:\\documenti\\uni\\tesi\\data\\acf.txt","w");
fprintf(facf,"Autocovariance Autocorelation\n");
for(t=0;t<T;t++)
    fprintf(facf,"%lf      %lf\n",acov[t],acor[t]);
fclose(facf);

/*Calcolo e stampa del periodogramma*/
periodogram(spec,omega,acov,T,PTS);
period=fopen("c:\\documenti\\uni\\tesi\\data\\period.txt","w");
fprintf(period,"Omega      Periodogram\n");
for(t=1;t<PTS;t++)
    fprintf(period,"%lf %lf\n",omega[t],spec[t]);
fclose(period);

/*Valori iniziali per la routine di ottimizzazione*/
theta[0]=0.0;
theta[1]=0.1;
theta[2]=0.3;
theta[3]=0.7;
theta[4]=0.5;
theta[5]=0.0;
printf("\nEstimating on a sample of %d observations",T);
fprintf(output,"\nEstimating on a sample of %d observations",T);
if(ANALYTICAL==1)      /*Stima con derivate analitiche*/
{
    printf(" with analytical derivatives\n");
    fprintf(output," with analytical derivatives\n");
    adfpmmin(figarchll,theta,r,&ll,TOL,dum,
        &call,&iter,N,T,M,output);
}
else                    /*e con derivate numeriche.*/
{
    printf(" with numerical derivatives\n");
    fprintf(output," with numerical derivatives\n");
    ndfpmmin(figarchll,theta,r,&ll,TOL,dum,
        &call,&iter,N,T,M,output);
}

printf("\nComputing standard errors...\n");
hess(figarchll,theta,I2d,/*Calcolo dell'Hessiana*/
    r,dum,&call,N,T,M);
gjin(I2d,N);            /*e sua inversione.*/
for(i=0;i<N;i++)
{
    /*Calcolo degli standard errors*/
    se[i]=sqrt(I2d[i][i]);
    tstat[i]=theta[i]/se[i];
    pval[i]=pstud(tstat[i],T-N);
}
aic=2*(ll/T)+2*(N/T);    /*e dei criteri informativi*/
sc=2*(ll/T)+N*(log(T)/T);

/*Stampa dei risultati*/
printf("\nParameter estimates:\n");
printf("m = %f   se = %f   t = %f   p  = %f\n",
    theta[0],se[0],tstat[0],pval[0]);
printf("w = %f   se = %f   t = %f   p  = %f\n",
    theta[1],se[1],tstat[1],pval[1]);

```



```

printf("a = %f    se = %f    t = %f    p = %f\n",
theta[2],se[2],tstat[2],pval[2]);
printf("b = %f    se = %f    t = %f    p = %f\n",
theta[3],se[3],tstat[3],pval[3]);
printf("d = %f    se = %f    t = %f    p = %f\n",
theta[4],se[4],tstat[4],pval[4]);
printf("D = %f    se = %f    t = %f    p = %f\n",
theta[5],se[5],tstat[5],pval[5]);
printf("Log-likelihood = %f\n",-ll);
printf("AIC = %f    SC = %f\n",aic,sc);
fprintf(output,"\nParameter estimates:\n");
fprintf(output,"m = %f    se = %f    t = %f    p = %f\n",
theta[0],se[0],tstat[0],pval[0]);
fprintf(output,"w = %f    se = %f    t = %f    p = %f\n",
theta[1],se[1],tstat[1],pval[1]);
fprintf(output,"a = %f    se = %f    t = %f    p = %f\n",
theta[2],se[2],tstat[2],pval[2]);
fprintf(output,"b = %f    se = %f    t = %f    p = %f\n",
theta[3],se[3],tstat[3],pval[3]);
fprintf(output,"d = %f    se = %f    t = %f    p = %f\n",
theta[4],se[4],tstat[4],pval[4]);
fprintf(output,"D = %f    se = %f    t = %f    p = %f\n",
theta[5],se[5],tstat[5],pval[5]);
fprintf(output,"Log-likelihood = %f\n",-ll);
fprintf(output,"AIC = %f    SC = %f\n",aic,sc);
fprintf(output,"\nVariance-covariance matrix:\n      m      w
a      b      d      D\n");
fprintf(output,"m %f %f %f %f %f %f\n",
I2d[0][0],I2d[0][1],I2d[0][2],I2d[0][3],I2d[0][4],I2d[0][5]);
fprintf(output,"w %f %f %f %f %f %f\n",
I2d[1][0],I2d[1][1],I2d[1][2],I2d[1][3],I2d[1][4],I2d[1][5]);
fprintf(output,"a %f %f %f %f %f %f\n",
I2d[2][0],I2d[2][1],I2d[2][2],I2d[2][3],I2d[2][4],I2d[2][5]);
fprintf(output,"b %f %f %f %f %f %f\n",
I2d[3][0],I2d[3][1],I2d[3][2],I2d[3][3],I2d[3][4],I2d[3][5]);
fprintf(output,"d %f %f %f %f %f %f\n",
I2d[4][0],I2d[4][1],I2d[4][2],I2d[4][3],I2d[4][4],I2d[4][5]);
fprintf(output,"D %f %f %f %f %f %f\n",
I2d[5][0],I2d[5][1],I2d[5][2],I2d[5][3],I2d[5][4],I2d[5][5]);

printf("\nComputing robust standard errors...\n");
/*Calcolo della matrice dei gradienti G,*/
if(ANALYTICAL==1)
    agradmat(theta,G,r,dum,N,T,M);
else
    for(t=0;t<T;t++)
        ngradvec(figarchllvec,theta,G[t],r,dum,t,N,T,M);
mtransp(G,Gt,T,N); /*sua trasposizione in Gt,*/
mprod(Gt,G,Iop,N,T,T,N); /*prodotto G*Gt in Iop,*/
mprod(I2d,Iop,Ipr,N,N,N,N); /*prodotto I2d*Iop*I2d,*/
mprod(Ipr,I2d,cov,N,N,N,N);

for(i=0;i<N;i++)/*e calcolo degli standard errors robusti.*/
{
    rse[i]=sqrt(cov[i][i]);
    rtstat[i]=theta[i]/rse[i];
    rpval[i]=pstud(rtstat[i],T-N);
}

/*Stampa deli risultati*/

```

```

printf("\n\nRobust parameter estimates:\n");
printf("m = %f  se = %f  t = %f  p = %f\n",
theta[0],rse[0],rtstat[0],rpval[0]);
printf("w = %f  se = %f  t = %f  p = %f\n",
theta[1],rse[1],rtstat[1],rpval[1]);
printf("a = %f  se = %f  t = %f  p = %f\n",
theta[2],rse[2],rtstat[2],rpval[2]);
printf("b = %f  se = %f  t = %f  p = %f\n",
theta[3],rse[3],rtstat[3],rpval[3]);
printf("d = %f  se = %f  t = %f  p = %f\n",
theta[4],rse[4],rtstat[4],rpval[4]);
printf("D = %f  se = %f  t = %f  p = %f\n",
theta[5],rse[5],rtstat[5],rpval[5]);
printf("Log-likelihood = %f\n",-ll);
printf("AIC = %f  SC = %f\n",aic,sc);
fprintf(output,"\n\nRobust parameter estimates:\n");
fprintf(output,"m = %f  se = %f  t = %f  p = %f\n",
theta[0],rse[0],rtstat[0],rpval[0]);
fprintf(output,"w = %f  se = %f  t = %f  p = %f\n",
theta[1],rse[1],rtstat[1],rpval[1]);
fprintf(output,"a = %f  se = %f  t = %f  p = %f\n",
theta[2],rse[2],rtstat[2],rpval[2]);
fprintf(output,"b = %f  se = %f  t = %f  p = %f\n",
theta[3],rse[3],rtstat[3],rpval[3]);
fprintf(output,"d = %f  se = %f  t = %f  p = %f\n",
theta[4],rse[4],rtstat[4],rpval[4]);
fprintf(output,"D = %f  se = %f  t = %f  p = %f\n",
theta[5],rse[5],rtstat[5],rpval[5]);
fprintf(output,"Log-likelihood = %f\n",-ll);
fprintf(output,"AIC = %f  SC = %f\n",aic,sc);
fprintf(output,"\n\nRobust variance-covariance matrix:\n      m      w
a      b      d      D\n");
fprintf(output,"m %f %f %f %f %f %f\n",
cov[0][0],cov[0][1],cov[0][2],cov[0][3],cov[0][4],cov[0][5]);
fprintf(output,"w %f %f %f %f %f %f\n",
cov[1][0],cov[1][1],cov[1][2],cov[1][3],cov[1][4],cov[1][5]);
fprintf(output,"a %f %f %f %f %f %f\n",
cov[2][0],cov[2][1],cov[2][2],cov[2][3],cov[2][4],cov[2][5]);
fprintf(output,"b %f %f %f %f %f %f\n",
cov[3][0],cov[3][1],cov[3][2],cov[3][3],cov[3][4],cov[3][5]);
fprintf(output,"d %f %f %f %f %f %f\n",
cov[4][0],cov[4][1],cov[4][2],cov[4][3],cov[4][4],cov[4][5]);
fprintf(output,"D %f %f %f %f %f %f\n",
cov[5][0],cov[5][1],cov[5][2],cov[5][3],cov[5][4],cov[5][5]);

/*Previsione della varianza condizionata*/
printf("\nForecasting conditional variance over %d periods...\n",
F);
for(t=0;t<T;t++) /*Calcolo della media dei residui al quadrato*/
{
    rsq[t]=(r[t]-theta[0])*(r[t]-theta[0]);
    mrq+=rsq[t];
}
mrq/=T;
/*Calcolo della varianza condizionata*/
figarchvc(rsq,ht,theta,dum,mrq,T,M);
/*Stampa della varianza condizionata*/
cvar=fopen("c:\\documenti\\uni\\tesi\\data\\htD.txt","w");
fprintf(cvar,"Conditional variance\n");
for(t=0;t<T;t++)

```

```

        fprintf(cvar,"%lf\n",ht[t]);
fclose(cvar);
/*Calcolo delle previsioni*/
forecast(rsq,dum,fcst,theta,ht[T-1],mrq,T,M,F);
/*e stampa dell'output.*/
fore=fopen("c:\\documenti\\uni\\tesi\\data\\foreD.txt","w");
fprintf(fore,"Forecasts of conditional variance\n");
for(t=1;t<F;t++)
    fprintf(fore,"%lf\n",fcst[t]);
fclose(fore);

/*Calcolo del VaR*/
printf("\nComputing Value at Risk...\n");
fprintf(output,"\n\nVALUE AT RISK COMPUTATIONS:\n");
/*Lettura dei dati di Eviews per il GARCH (2,1)*/
gfore=fopen("c:\\documenti\\uni\\tesi\\data\\gfore.txt","rt");
for(t=0;t<F;t++)
    fscanf(gfore,"%lf",&gfcst[t]);
fclose(gfore);

/*VaR FIGARCH a 1 giorno*/
var[0]=1.645*sqrt(fcst[0]);
fprintf(output,"FIGARCH (1,d,1) Conditional VaR over 1 day = %f%%\n",
var[0]);
/*VaR GARCH a 1 giorno*/
var[1]=1.645*sqrt(gfcst[0]);
fprintf(output,"GARCH (2,1) Conditional VaR over 1 day =
%f%%\n",var[1]);
/*VaR sigma a 1 giorno*/
var[2]=1.645*sdev;
fprintf(output,"Unconditional VaR over 1 day = %f%%\n",var[2]);
/*VaR FIGARCH a 3 giorni*/
var[3]=0;
for(i=0;i<3;i++)
    var[3]+=fcst[i];
var[3]=sqrt(var[3]);
var[3]*=1.645;
fprintf(output,"\nFIGARCH (1,d,1) Conditional VaR over 3 days =
%f%%\n",var[3]);
/*VaR GARCH a 3 giorni*/
var[4]=0;
for(i=0;i<3;i++)
    var[4]+=gfcst[i];
var[4]=sqrt(var[4]);
var[4]*=1.645;
fprintf(output,"GARCH (2,1) Conditional VaR over 3 days =
%f%%\n",var[4]);
/*VaR sigma a 3 giorni*/
var[5]=1.645*sqrt(3)*sdev;
fprintf(output,"Unconditional VaR over 3 days = %f%%\n",var[5]);
/*VaR FIGARCH a 10 giorni*/
var[6]=0;
for(i=0;i<10;i++)
    var[6]+=fcst[i];
var[6]=sqrt(var[6]);
var[6]*=1.645;
fprintf(output,"\nFIGARCH (1,d,1) Conditional VaR over 10 days =
%f%%\n",var[6]);
/*VaR GARCH a 10 giorni*/
var[7]=0;

```

```

        for(i=0;i<10;i++)
            var[7]+=gfcst[i];
        var[7]=sqrt(var[7]);
        var[7]*=1.645;
        fprintf(output,"GARCH (2,1) Conditional VaR over 10 days =
%f%%\n",var[7]);
        /*VaR sigma a 10 giorni*/
        var[8]=1.645*sqrt(10)*sdev;
        fprintf(output,"Unconditional VaR over 10 days = %f%%\n",var[8]);

        /*Calcolo del tempo trascorso*/
        finish=clock();
        elaps=(double)(finish-start)/CLOCKS_PER_SEC;
        min=elaps/60;
        /*Stampa degli ultimi dati*/
        printf("\nThe Log-likelihood function was called %d times\n",call);
        printf("%.2f seconds (%.2f minutes) elapsed\n",elaps,min);
        fprintf(output,"\nThe Log-likelihood function was called %d
times\n",call);
        fprintf(output,"%.2f seconds (%.2f minutes) elapsed\n",elaps,min);
        fclose(output);
        printf("\nOutput saved as FIGARCHD.out\n");

        /*Liberazione dei vettori*/
        delete [] r;
        delete [] dum;
        delete [] ht;
        delete [] rsq;
        delete [] acov;
        delete [] acor;
        delete [] spec;
        delete [] omega;
        delete [] fcst;
        delete [] gfcst;
        delete [] grad;
        for(i=0;i<N;i++)
        {
            delete [] cov[i];
            delete [] Iop[i];
            delete [] I2d[i];
            delete [] Ipr[i];
            delete [] Gt[i];
        }
        delete [] cov;
        delete [] Iop;
        delete [] I2d;
        delete [] Ipr;
        delete [] Gt;
        for(i=0;i<T;i++)
            delete [] G[i];
        delete [] G;
    }

```

### B.2.7 Le altre funzioni sviluppate

La prima funzione che presentiamo in questo paragrafo calcola, a partire da una serie  $y$  di  $T$  osservazioni, le autocovarianze e le autocorrelazioni empiriche, che vengono restituite come vettori `acov` e `acor`.

```
void acf(double *y,          /*Serie dei dati*/
         double *acov,      /*Autocovarianze, output*/
         double *acor,      /*Autocorrelazioni, output*/
         int T)             /*Numerosità del campione*/
{
    double med=0;           /*Dichiarazioni*/
    int t, j;

    for(t=0;t<T;t++)        /*Calcolo della media della serie*/
        med+=y[t];
    med/=T;

    for(j=0;j<T;j++)        /*Calcolo di acov e acor*/
    {
        acov[j]=0;
        for (t=j;t<T;t++)
            acov[j]+=(y[t]-med)*(y[t-j]-med);
        acov[j]/=T;
        acor[j]=acov[j]/acov[0];
    }
}
```

La funzione `periodogram` calcola, a partire dal vettore delle autocovarianze osservate `acov`, il periodogramma `spectre`, dopo aver diviso l'intervallo  $[0, \pi]$  in `npts` sezioni, le cui ascisse sono riportate in output nel vettore `omega`.

```
#include <math.h>
#define PI 3.1415926535897932384626433832795

void periodogram(double *spectre, /*Periodogramma, output*/
                double *omega,    /*Ascisse, output*/
                double *acov,     /*Autocovarianze*/
                int T,            /*Numero di autocovarianze*/
                int npts)         /*Numero di intervalli*/
{
    double num;                  /*Dichiarazioni*/
    int i, j;
    omega[0]=0;

    for(j=1;j<npts;j++)          /*Calcolo delle ascisse*/
    {
```

```

        num=j;                                /*conversione di j in double*/
        omega[j]=num/(npts-1)*PI; /*e divisione dell'intervallo*/
    }

    for(i=0;i<npts;i++)                        /*Calcolo del periodogramma*/
    {
        spectre[i]=(1/(2*PI))*acov[0];
        for(j=0;j<T;j++)
            spectre[i]+=(1/PI)*acov[j]*cos(omega[i]*j);
    }
}

```

La funzione `forecast` calcola le previsioni della varianza condizionata per i prossimi `nfor` periodi, calcolate secondo i parametri passati col vettore `param`; le previsioni sono restituite nel vettore `forecast`.

```

#include <math.h>

void forecast(double *ressq,                /*Residui al quadrato*/
              int *dum,                    /*Vettore delle dummy*/
              double *forecast,            /*Previsioni - output*/
              double *param,              /*Parametri*/
              double hh,                  /*Ultimo valore della varianza
                                         condizionata*/
              double mrq,                 /*Media dei residui al quadrato*/
              int nobs,                   /*Numero di osservazioni*/
              int ntr,                    /*Troncamento del termine  $(1-L)^d$ */
              int nfor)                   /*Periodi di previsione*/
{
    double *lambda,                      /*Coefficienti lambda (si confronti
                                         la (5.2.17))*/
           *ff,                          /*Serie della varianza
                                         condizionata corretta con
                                         l'aggiunta in coda
                                         delle previsioni*/
           *delta;                       /*Coefficienti delta (si confronti
                                         la (5.2.17))*/

    double w, a, b, d, D;                /*Dichiarazioni*/
    int k, t;

    lambda=new double[ntr];              /*Allocazioni*/
    delta=new double[ntr];
    ff=new double[nobs+nfor];

    w=param[1];                          /*Inizializzazione dei parametri*/
    a=param[2];
    b=param[3];
    d=param[4];
    D=param[5];
    delta[0]=1;
}

```

```

for(k=1;k<ntr;k++)          /*Calcolo dei coefficienti delta*/
    delta[k]=d*exp(lngamma(k-d)-lngamma(k+1)-lngamma(1-d));

for(t=0;t<nobs;t++)          /*Riempimento di ff*/
    ff[t]=ressq[t];

lambda[0]=0;                  /*Calcolo dei coefficienti lambda*/
lambda[1]=d-b+a;
for(k=2;k<ntr;k++)
    lambda[k]=b*lambda[k-1]+((k-1-d)/k)-a)*delta[k-1];

ff[nobs]=hh;                  /*Calcolo dei valori di ff in coda*/
for(t=nobs+1;t<nobs+nfor;t++)
{
    ff[t]=w/(1-b)+D*dum[t];
    for(k=1;k<ntr;k++)
        ff[t]+=lambda[k]*ff[t-k];
}

for(t=0;t<nfor;t++)          /*Passaggio degli ultimi nfor valori
                               di ff al vettore di output*/
    forecast[t]=ff[nobs+t];

delete [] lambda;             /*Liberazione dei vettori*/
delete [] delta;
delete [] ff;
}

```

Le funzioni `mprod` e `mtransp` calcolano rispettivamente il prodotto delle matrici `matrix1` (di dimensioni `nRows1xnCols1`) e `matrix2` (di dimensioni `nRows2xnCols2`) e la trasposta di `matrix` (di dimensioni `nRowsxnCols`); sia il prodotto che la trasposta sono restituiti come output.

```

void mprod(double **matrix1,
           double **matrix2,
           double **output,
           int nRows1,
           int nCols1,
           int nRows2,
           int nCols2)
{
    int i, j, k;
    for (i=0;i<nRows1;i++)
        for (j=0;j<nCols2;j++)
        {
            output[i][j]=0;
            for (k=0;k<nCols1;k++)
                output[i][j]+=matrix1[i][k]*matrix2[k][j];
        }
}

```

```

void mtransp(double **matrix,
             double **output,
             int nRows,
             int nCols)

{
    int i, j;

    for(i=0;i<nCols;i++)
        for(j=0;j<nRows;j++)
            output[i][j]=matrix[j][i];
}

```

La funzione `refgamma` calcola la riflessione della funzione  $\Gamma$  per i reali negativi (si veda la (2.1.2)); `xx` è il punto (minore di zero) in cui calcolare la funzione, il cui valore è restituito alla chiamata.

```

#include <math.h>
#define PI 3.1415926535897932384626433832795

double refgamma(double xx)

{
    double gamma,den;
    gamma=exp( lngamma(1-xx) );
    den=gamma*sin(PI*xx);
    return PI/den;
}

```

### *B.2.8 Prototipi delle funzioni Numerical Recipes*

Come abbiamo già accennato, presentiamo in questo paragrafo i prototipi delle funzioni tratte dal testo di Press, Teukolsky, Vetterling e Flannery (1992).

La funzione `adfpmin` è la routine di minimizzazione di una funzione `func` secondo il metodo BFGS (si veda il paragrafo 4.2.3). La variante `ndfpmin`, con lo stesso prototipo, utilizza le derivate numeriche al posto delle analitiche. Sono state apportate due modifiche: la prima per fare stampare a schermo e su file, ad ogni iterazione, il risultato parziale; la seconda



per fare in modo che la convergenza debole<sup>2</sup> sia ammissibile solo dopo la venticinquesima iterazione.

```
void adfpmmin(double(*func) /*Funzione da minimizzare,*/
              (double*, /*e suoi argomenti.*/
               double*
               int*,
               int*,
               int,
               int),
              double *p, /*Vettore del punto in cui ci si trova,
                           passato come valore iniziale*/
              double *y, /*Vettore delle osservazioni*/
              double *fret, /*Valore della funzione al minimo -
                              output*/
              double gtol, /*Livello di tolleranza per il
                              gradiente*/
              int *dum, /*Vettore delle dummy*/
              int *call, /*Chiamate alla funzione da minimizzare*/
              int *iter, /*Numero di iterazioni effettuate -
                           output*/
              int n, /*Numero dei parametri*/
              int nob, /*Numerosità del campione*/
              int ntr, /*Termine di troncamento per (1-L)d*/
              FILE *output) /*File di output*/
```

La funzione `lnsrch` è utilizzata da `adfpmmin` per effettuare la ricerca lineare dello *stepsize* ottimale.

```
void lnsrch(double (*func) /*Funzione di lavoro,*/
            (double*, /*e suoi argomenti.*/
             double*,
             int*,
             int*,
             int,
             int),
            double *p, /*Direzione per la ricerca*/
            double *x, /*Punto di minimo, output*/
            double *xold, /*Punto di partenza*/
            double *g, /*Gradiente in xold*/
            double *y, /*Vettore delle osservazioni*/
            double *f, /*Valore di func in x*/
            double fold, /*Valore di func in xold*/
            double stpmax, /*Ampiezza massima del passo*/
            int *dum, /*Vettore delle dummy*/
            int *check,
            int *call, /*Numero di chiamate a func*/
            int n, /*Numero di parametri di lavoro*/
            int nob, /*Numerosità del campione*/
            int ntr) /*Termine di troncamento per (1-L)d*/
```

---

<sup>2</sup> Si parla di convergenza forte quando il gradiente vale zero, di convergenza debole quando, pur non verificandosi tale condizione, l'algoritmo di ricerca lineare non riesce a trovare un punto "migliore".

La funzione `dfridr` calcola la derivata numerica della funzione `func` nel punto `x` rispetto alla variabile indicata da `dummy`; la variante `dfpridr` calcola invece la derivata parziale rispetto a `dummy1` e `dummy2`.

```
double dfridr(double (*func) /*Funzione da derivare,*/
              (double*, /*e suoi argomenti.*/
               double*,
               int*,
               int*,
               int,
               int),
              double* x, /*Punto in cui si calcola la derivata*/
              double* y, /*Vettore delle osservazioni*/
              int *dum, /*Vettore delle dummy*/
              int* dummy, /*Dummy per la scelta della variabile*/
              int* call, /*Numero di chiamate a func*/
              int nVar, /*Numero delle variabili*/
              int nobs, /*Numerosità del campione*/
              int ntr) /*Termine di troncamento per (1-L)d*/

double dfpridr(double (*func) /*Funzione da derivare,*/
               (double*, /*e suoi argomenti.*/
                double*,
                int*,
                int*,
                int,
                int),
               double* x, /*Punto in cui si calcola la derivata*/
               double* y, /*Vettore delle osservazioni*/
               int *dum, /*Vettore delle dummy*/
               int* dummy1, /*Dummy per la variabile 1*/
               int* dummy2, /*Dummy per la variabile 2*/
               int* call, /*Numero di chiamate a func*/
               int nVar, /*Numero delle variabili*/
               int nobs, /*Numerosità del campione*/
               int ntr) /*Termine di troncamento per (1-L)d*/
```

La funzione `lngamma` calcola il logaritmo naturale della funzione  $\Gamma$  in `xx`, che viene passato come argomento. Il valore della funzione è restituito in `output` alla chiamata.

```
double lngamma(double xx)
```

La funzione<sup>3</sup> `pstud` calcola il *P-value* (che viene restituito alla chiamata) in `x` di una *t* di Student con `gdl` gradi di libertà:

```
double pstud(double x,int gdl)
```

La funzione `gjinv` inverte la matrice `a`, di dimensioni  $n \times n$ , sovrascrivendola all'originale.

```
void gjinv(double **a, int n)
```

---

<sup>3</sup> Questa funzione è stata sviluppata da Marco Fazi di Datamat S.p.A., che ringraziamo per la gentile collaborazione.

## B.3 La forma dell'output

Presentiamo infine una copia del file di output generato dalla procedura:

```
FIGARCH (1,d,1) estimation routine - Executable V. 1.2D
(C) Marco J. Lombardi, 16/12/2000

DATASET: $/E daily returns
Sample 28/12/1978 - 08/11/2000
Mean = -0.008344 Standard error = 0.671836
Mean of squares = 0.451434

Estimating on a sample of 5705 observations
with analytical derivatives
BFGS iteration #1 : Log-likelihood = -5896.523219
BFGS iteration #2 : Log-likelihood = -5441.685085
BFGS iteration #3 : Log-likelihood = -5523.412701
BFGS iteration #4 : Log-likelihood = -5523.412701
BFGS iteration #5 : Log-likelihood = -5522.559005
BFGS iteration #6 : Log-likelihood = -5501.702193
BFGS iteration #7 : Log-likelihood = -5503.784756
BFGS iteration #8 : Log-likelihood = -5503.784756
BFGS iteration #9 : Log-likelihood = -5503.741299
BFGS iteration #10: Log-likelihood = -5493.713269
BFGS iteration #11: Log-likelihood = -5476.366218
BFGS iteration #12: Log-likelihood = -5464.821487
BFGS iteration #13: Log-likelihood = -5459.153853
BFGS iteration #14: Log-likelihood = -5445.649189
BFGS iteration #15: Log-likelihood = -5440.824242
BFGS iteration #16: Log-likelihood = -5439.895907
BFGS iteration #17: Log-likelihood = -5438.509780
BFGS iteration #18: Log-likelihood = -5436.296883
BFGS iteration #19: Log-likelihood = -5435.508082
BFGS iteration #20: Log-likelihood = -5434.406385
BFGS iteration #21: Log-likelihood = -5433.712884
BFGS iteration #22: Log-likelihood = -5433.160114
BFGS iteration #23: Log-likelihood = -5432.360591
BFGS iteration #24: Log-likelihood = -5431.851984
BFGS iteration #25: Log-likelihood = -5431.835626
BFGS iteration #26: Log-likelihood = -5431.790735
BFGS iteration #27: Log-likelihood = -5431.788446
BFGS iteration #28: Log-likelihood = -5431.787933
BFGS iteration #29: Log-likelihood = -5431.787694
BFGS iteration #30: Log-likelihood = -5431.787692
Strong convergence achieved...

Parameter estimates:
m = -0.003858   se = 0.007638   t = -0.505153   p = 0.613471
w = 0.066255   se = 0.011736   t = 5.645321   p = 0.000000
a = 0.345168   se = 0.033182   t = 10.402284   p = 0.000000
b = 0.709362   se = 0.037854   t = 18.739409   p = 0.000000
d = 0.470897   se = 0.058379   t = 8.066263   p = 0.000000
D = -0.071778   se = 0.013981   t = -5.134000   p = 0.000000
Log-likelihood = -5431.787692
AIC = 1.904220   SC = 1.913316
```

Variance-covariance matrix:

	m	w	a	b	d	D
m	0.000058	-0.000001	-0.000011	-0.000005	0.000003	0.000001
w	-0.000001	0.000138	0.000064	-0.000154	-0.000223	-0.000160
a	-0.000011	0.000064	0.001101	-0.000343	-0.001268	-0.000061
b	-0.000005	-0.000154	-0.000343	0.001433	0.001936	0.000104
d	0.000003	-0.000223	-0.001268	0.001936	0.003408	0.000167
D	0.000001	-0.000160	-0.000061	0.000104	0.000167	0.000195

Robust parameter estimates:

m	= -0.003858	se = 0.008012	t = -0.481608	p = 0.630103
w	= 0.066255	se = 0.016606	t = 3.989708	p = 0.000067
a	= 0.345168	se = 0.050601	t = 6.821366	p = 0.000000
b	= 0.709362	se = 0.064374	t = 11.019356	p = 0.000000
d	= 0.470897	se = 0.102841	t = 4.578884	p = 0.000005
D	= -0.071778	se = 0.019254	t = -3.727907	p = 0.000195
Log-likelihood = -5431.787692				
AIC = 1.904220    SC = 1.913316				

Robust variance-covariance matrix:

	m	w	a	b	d	D
m	0.000064	0.000000	-0.000033	0.000059	0.000075	0.000002
w	0.000000	0.000276	0.000175	-0.000431	-0.000606	-0.000311
a	-0.000033	0.000175	0.002560	-0.001435	-0.003822	-0.000109
b	0.000059	-0.000431	-0.001435	0.004144	0.005980	0.000292
d	0.000075	-0.000606	-0.003822	0.005980	0.010576	0.000384
D	0.000002	-0.000311	-0.000109	0.000292	0.000384	0.000371

VALUE AT RISK COMPUTATIONS:

FIGARCH (1,d,1) Conditional VaR over 1 day = 1.126614%  
GARCH (2,1) Conditional VaR over 1 day = 1.124122%  
Unconditional VaR over 1 day = 1.105171%

FIGARCH (1,d,1) Conditional VaR over 3 days = 2.156788%  
GARCH (2,1) Conditional VaR over 3 days = 1.965269%  
Unconditional VaR over 3 days = 1.914212%

FIGARCH (1,d,1) Conditional VaR over 10 days = 4.075922%  
GARCH (2,1) Conditional VaR over 10 days = 3.601692%  
Unconditional VaR over 10 days = 3.494857%

The Log-likelihood function was called 737 times  
281.93 seconds (4.70 minutes) elapsed

# Bibliografia

- **AA. VV.**, 1993, “RiskMetrics™ (First edition)”, (*J.P. Morgan, New York, NY*).
- **Adenstedt, R.K.**, 1974, “On large sample estimation for the mean of a stationary random sequence”, *Annals of mathematical statistics* 2: 1095-1107.
- **Agiakloglou, C., P. Newbold and M. Wohar**, 1992, “Bias of an estimator of the fractional difference parameter”, *Journal of time series analysis* 14: 235-246.
- **Akaike, H.**, 1973, “Information theory and an extension of the maximum likelihood principle”, *Second international symposium on information theory (Akademiai kiadó, Budapest, H)*.
- **Andersen, T.G.**, 1994, “Stochastic autoregressive volatility: a framework for volatility modelling”, *Mathematical finance* 4: 75-102.
- **Andersen, T.G. and T. Bollerslev**, 1997a, “Heterogeneous information arrivals and returns volatility dynamics: uncovering the long run in high frequency returns”, *Journal of finance* 52: 975-1005.
- **Andersen, T.G. and T. Bollerslev**, 1997b, “DM-Dollar volatility: intraday activity patterns, macroeconomic announcements and longer run dependencies”, *Journal of finance* 53: 219-265.
- **Andersen, T.G. and T. Bollerslev**, 1997c, “Intraday periodicity and volatility persistence in financial markets”, *Journal of empirical finance* 4: 115-158.
- **Azzalini, A.**, 1992, “Inferenza Statistica. Una introduzione basata sul concetto di verosimiglianza”, (*Sprinter-Verlag, Berlin, D*).
- **Baillie, R.T.**, 1996, “Long-memory processes and fractional integration in econometrics”, *Journal of econometrics* 73: 5-59.
- **Baillie, R.T. and T. Bollerslev**, 1989, “The message in daily exchange rates: a conditional variance tale ”, *Journal of business and economic statistics* 7: 297-305.
- **Baillie, R.T. and T. Bollerslev**, 1994, “Cointegration, fractional cointegration and exchange rate dynamics”, *Journal of finance* 49: 737-745.
- **Baillie, R.T., T. Bollerslev and H.O. Mikkelsen**, 1996, “Fractionally integrated generalized autoregressive conditional heteroskedasticity”, *Journal of econometrics* 74: 3-30.

- **Baillie, R.T., C.F. Chung and M.A. Tieslau**, 1996 “Analysing inflation by the ARFIMA-GARCH model”, *Journal of applied econometrics* 11: 23-40.
- **Barone-Adesi, G. and K. Giannopoulos**, 1996, “A simplified approach to the conditional estimation of Value at Risk”, *Futures and options world* October: 68-72.
- **Beine, M. and S. Laurent**, 1999, “Central bank interventions and jumps in double long memory models of daily exchange rates”, *unpublished working paper (CADRE, Lille, F)*.
- **Beine, M., S. Laurent and C. Lecourt**, 1999, “Accounting for conditional leptokurtosis and closing day effects in FIGARCH models of daily exchange rates”, *unpublished working paper (CADRE, Lille, F)*.
- **Beine, M., A. Benassy-Quere and C. Lecourt**, 1999, “The impact of foreign exchange interventions: new evidence from FIGARCH estimation”, *unpublished working paper (CADRE, Lille, F)*.
- **Beran, J.**, 1994, “Statistics for long-memory processes”, (*Chapman & Hall, New York, NY*).
- **Berndt, E.K., B.H. Hall, R.E. Hall and J.A. Hausman**, 1974, “Estimation and inference in nonlinear structural models”, *Annals of economic and social measurement* 3: 653-665.
- **Black, F.**, 1976, “Studies of stock prices volatility changes”, *Proceedings from the American Statistical Association – Business and economic statistics section*: 177-181.
- **Black, F. and M. Scholes**, 1973, “The pricing of options and corporate liabilities”, *Journal of political economy* 81: 637-654.
- **Boes, D.C., R.A. Davis and S.N. Gupta**, 1989, “Parameter estimation in low order fractionally differenced ARMA models”, *Stochastic hydrology and hydraulics* 3: 97-110.
- **Bollerslev, T.**, 1986, “Generalized autoregressive conditional heteroskedasticity”, *Journal of econometrics* 31: 307-327.
- **Bollerslev, T.**, 1987, “A conditional heteroskedastic time series model for speculative prices and rates of return”, *Review of economics and statistics* 69: 542-547.

- **Bollerslev, T.**, 1988, “On the correlation structure for the GARCH process”, *Journal of time series analysis* 9: 121-131.
- **Bollerslev, T., R.F. Engle and D.B. Nelson**, 1994, “ARCH models”, *Handbook of econometrics (Fourth volume)* (Elsevier Science, Amsterdam, NL).
- **Bollerslev, T. and H.O. Mikkelsen**, 1996, “Modelling and pricing long memory in stock market volatility”, *Journal of econometrics* 73: 151-184.
- **Bollerslev, T. and J.M. Wooldridge**, 1992, “Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances”, *Econometric reviews* 11: 143-172.
- **Bougerol, P. and N. Picard**, 1992, “Stationarity of GARCH processes and of some nonnegative time series”, *Journal of econometrics* 52: 115-128.
- **Bowman, K.O. and B.R. Shenton**, 1975, “Omnibus test contours for departures from normality based on  $\sqrt{b_1}$  and  $b_2$ ”, *Biometrika* 62: 243-250.
- **Box, G.E.P. and G.M. Jenkins**, 1976, “Time series analysis: forecasting and control”, (Holden-Day, San Francisco, CA).
- **Brealey, R.A. and S.C. Myers**, 1993, “Principi di finanza aziendale (Seconda edizione)”, (McGraw-Hill Italia, Milano, I).
- **Breidt, F.J., N. Crato and P.J.F. de Lima**, 1997, “On the detection and estimation of long memory in stochastic volatility processes”, *Journal of econometrics* 83: 325-348.
- **Brockwell, P.J. and R.A. Davis**, 1991, “Time series: Theory and methods (Second edition)”, (Springer-Verlag, New York, NY).
- **Broyden, C.G.**, 1970, “The convergence of a class of double rank minimization algorithms”, *Journal of the Institute of mathematics and its applications* 6: 76-90.
- **Brunetti, C. and C.L. Gilbert**, 1999, “Bivariate FIGARCH and fractional cointegration”, *unpublished working paper* (Queen Mary and Westfield College, London, UK).
- **Caparrelli, F.**, 1995, “Il mercato azionario”, (McGraw-Hill Italia, Milano, I).
- **Caporin, M.**, 2002, “FIGARCH models: Stationarity, estimation methods and the identification problem”, *working paper 0202* (GRETA, Venezia, I).
- **Casella, G. and R.L. Berger**, 1990, “Statistical Inference”, (Duxbury Press, Belmont, CA).
- **Cattaneo, C.**, 1999, “Manuale di finanza aziendale”, (Il Mulino, Bologna, I).



- **Cecconi, M. and M.J. Lombardi**, 2001, "Dati finanziari ad alta frequenza: trattamento e applicazioni", *Scienza e Business* 9: 17-24.
- **Cherubini, U. and G. Della Lunga**, 2001, "Il rischio finanziario", (*McGraw-Hill Italia, Milano, I*).
- **Cheung Y.W.**, 1996, "Stock market volatility and fractional integration", *International journal of finance and economics* 1: 263-273.
- **Cheung, Y.W. and F.X. Diebold**, 1994, "On maximum likelihood estimation of the differencing parameter of fractionally integrated noise with unknown mean", *Journal of econometrics* 62: 301-316.
- **Chung, C.F.**, 1996a, "Estimating a generalized long-memory process", *Journal of econometrics* 73: 237-259.
- **Chung, C.F.**, 1996b, "A generalized fractionally integrated ARMA process", *Journal of time series analysis* 17: 111-140.
- **Chung, C.F.**, 1999, "Estimating the Fractionally Integrated GARCH Model", *unpublished working paper (National Taiwan University, Taipei, TW)*.
- **Chung, C.F. and R.T. Baillie**, 1993, "Small sample bias in conditional sum of squares estimators of fractionally integrated ARMA models", *Empirical economics* 18: 791-806.
- **Cochrane, J.H.**, 1988, "How big is the random walk in GNP?", *Journal of political economy* 96: 893-920.
- **Cramér H., and M.R. Leadbetter**, 1967, "Stationary and related stochastic processes", (*John Wiley & Sons, New York, NY*).
- **Crato, N. and P.J.F. de Lima**, 1994, "Long-range dependence in the conditional variance of stock returns", *Economics letters* 45: 281-285.
- **Crato N., and P. Rothman**, 1994, "Fractional integration analysis of long-run behaviour for US macroeconomic time series", *Economics letters* 45: 287-291.
- **Dahlhaus, R.**, 1989, "Efficient parameter estimation for self-similar processes", *Annals of statistics* 17: 1749-1766.
- **Davidon, W.C.**, 1959, "Variable metric method of minimization", *A.E.C. research and development report*.
- **Davidson, J.E.**, 2002, "Moment and memory properties of linear conditional heteroskedasticity models", *unpublished working paper (University of Cardiff, Cardiff, UK)*.

- **Davidson, R., and J.G. McKinnon**, 1993, “Estimation and inference in econometrics”, (*Oxford University Press, Oxford, UK*).
- **Dennis, J.E. Jr. and R.B. Schnabel**, 1977, “Numerical methods for unconstrained optimization and nonlinear equations”, (*Prentice Hall, Englewood Cliffs, NJ*).
- **Dickey, D.A. and W.A. Fuller**, 1979, “Distribution of the estimators for autoregressive time series with a unit root”, *Journal of the American Statistical Association* 74: 427-431.
- **Diebold, F.X.**, 1989, “Random walks versus fractional integration: power comparisons of scalar and joint tests of the variance-time function”, *Advances in econometrics and modelling* (*Kluwer Academic Publishers, Needham, MA*).
- **Diebold, F.X. and G.D. Rudebusch**, 1989, “Long memory and persistence in aggregate output”, *Journal of monetary economics* 24: 189-209.
- **Ding, Z.**, 1994, “Time series analysis of speculative returns”, *Ph.D. dissertation* (*University of California, San Diego, CA*).
- **Ding, Z. and C.W.J. Granger**, 1996, “Modelling volatility persistence of speculative returns: a new approach”, *Journal of econometrics* 73: 185-215.
- **Ding, Z., C.W.J. Granger and R.F. Engle**, 1993, “A long-memory property of stock market returns and a new model”, *Journal of empirical finance* 1: 83-106.
- **Doornik, J.A.**, 1999, “Ox: an object oriented matrix programming language (Third edition)”, (*Timberlake Consultants Press, London, UK*).
- **Doornik, J.A. and M. Ooms**, 1999, “A package for estimating, forecasting and simulating ARFIMA models: ARFIMA package 1.0 for Ox”, *discussion paper* (*Nuffield college, Oxford, UK*).
- **Doornik, J.A., G. Draisma and M. Ooms**, 1998, “Introduction to Ox”, (*Timberlake Consultants Press, London, UK*).
- **Durbin J. and G. Watson**, 1950, “Testing for serial correlation in least squares regression”, *Biometrika* 37: 409-428.
- **Edwards, R.D. and J. Magee**, 1992, “Technical analysis of stock trends (Sixth edition)”, (*John Magee, Boston, MA*).
- **Enders, W.**, 1995, “Applied econometric time series”, (*John Wiley & Sons, New York, NY*).
- **Engle, R.F.**, 1982, “Autoregressive conditional heteroskedasticity with estimates of the variance of U.K. inflation”, *Econometrica* 50: 987-1008.

- **Engle, R.F. and T. Bollerslev**, 1986, “Modelling the persistence of conditional variances”, *Econometric reviews* 5: 1-50.
- **Engle, R.F. and V.K. Ng**, 1993, “Measuring and testing the impact of news on volatility”, *Journal of finance* 48: 1749-1778.
- **Fama, E.F.**, 1965, “The behavior of stock market prices”, *Journal of business* 38: 34-105.
- **Feller, W.**, 1951, “The asymptotic distribution of the range of sums of independent random variables”, *Annals of mathematical statistics* 22: 427-432.
- **Ferrari, G., R. Leoni and G. Marliani**, 1992, “Introduzione al campionamento e all’inferenza statistica”, (*Libreria Alfani, Firenze, I*).
- **Fiorentini, G., G. Calzolari and L. Panattoni**, 1996, “Analytical derivatives and the computation of GARCH estimates”, *Journal of applied econometrics* 11: 399-417.
- **Fletcher, R.**, 1970, “A new approach to variable metric algorithms”, *The computer journal* 13: 317-322.
- **Fletcher, R.**, 2000, “Practical methods of optimization (Second edition)”, (*John Wiley & Sons, New York, NY*).
- **Fletcher, R. and M.J.D. Powell**, 1963, “A rapidly convergent descent method for minimization”, *The computer journal* 6: 163-168.
- **Fox, R. and M.S. Taqqu**, 1986, “Large sample properties of parameter estimates for strongly dependent stationary Gaussian time series”, *Annals of statistics* 14: 517-532.
- **Gabbi, G.**, 1999, “La previsione nei mercati finanziari”, (*Bancaria, Roma, I*).
- **Gallant, A.R., P.E. Rossi and G. Tauchen**, 1993, “Nonlinear dynamic structures”, *Econometrica* 61: 871-907.
- **Gallo, G.M. and B. Pacini**, 2002, “Metodi quantitativi per i mercati finanziari”, (*Carocci, Roma, I*).
- **Geronazzo, L.**, 1998, “Complementi al corso di Modelli matematici per i mercati finanziari”, *unpublished working paper (Università degli studi di Firenze, Firenze, I)*.
- **Geweke, J. and S. Porter-Hudak**, 1983, “The estimation and application of long memory time series models”, *Journal of time series analysis* 4: 221-238.

- **Giriatis, L., P.M. Robinson and A. Samarov**, 1997, “Rate optimal semiparametric estimation of memory parameter of the Gaussian time series with long-range dependence”, *Journal of time series analysis* 18: 49-60.
- **Giusti, E.**, 1988, “Analisi matematica (Seconda edizione)”, (*Bollati-Boringhieri, Torino, I*).
- **Goldfarb, D.**, 1970, “A family of variable metric methods derived by variational means”, *Mathematics of computation* 24: 23-26.
- **Goldfeld, S.M. and R.E. Quandt**, 1972, “Nonlinear methods in econometrics”, (*North-Holland publishing company, Amsterdam, NL*).
- **Goldfeld, S.M., R.E. Quandt and H.F. Trotter**, 1966, “Maximization by quadratic hill climbing”, *Econometrica* 34: 541-551.
- **Gouriéroux, C., A. Monfort and A. Trognon**, 1984, “Pseudo Maximum Likelihood Methods : Theory”, *Econometrica* 52: 680-700.
- **Gouriéroux, C. and A. Monfort**, 1995, “Statistics and econometric models”, (*Cambridge University Press, Cambridge, UK*).
- **Gouriéroux, C. and A. Monfort**, 1997, “Time series and dynamic models”, (*Cambridge University Press, Cambridge, UK*).
- **Gradshteyn, I.S. and I.W. Ryzhik**, 1965, “Table of integrals, series and products”, (*Academic press, New York, NY*).
- **Granger, C.V.J.**, 1966, “The typical spectral shape of an economic variable”, *Econometrica* 16: 150-161.
- **Granger, C.V.J.**, 1980, “Long-memory relationships and the aggregation of dynamic models”, *Journal of econometrics* 14: 227-238.
- **Granger, C.V.J.**, 1981, “Some properties of time series data and their use in econometric model specification”, *Journal of econometrics* 16: 121-130.
- **Granger, C.V.J.**, 1990, “Modelling economic series”, (*Clarendon Press, Oxford, UK*).
- **Granger, C.V.J. and Z. Ding**, 1996a, “Varieties of long-memory models”, *Journal of econometrics* 73: 61-77.
- **Granger, C.V.J. and Z. Ding**, 1996b, “Some properties of absolute return: an alternative measure of risk”, *Annales d'économie et de statistique* 40: 67-91.
- **Granger, C.V.J. and R. Joyeux**, 1980, “An introduction to long-memory time series models and fractional differencing”, *Journal of time series analysis* 1: 15-29.

- **Granger, C.V.J. and O. Morgenstern**, 1963, "Spectral analysis of stock market prices", *Kyklos* 16: 1-27.
- **Granger, C.V.J. and T. Teräsvirta**, 1993, "Modelling nonlinear economic relationships", (*Oxford University Press, Oxford, UK*).
- **Gray, H.L., N.F. Zhang and W.A. Woodward**, 1989, "On generalized fractional processes", *Journal of time series analysis* 10: 233-257.
- **Greene, W.H.**, 2000, "Econometric analysis (Fourth edition)", (*Prentice Hall, Upper Saddle River, NJ*).
- **Guégan, D.**, 2001, "A new model : the k-factor GIGARCH process", *Journal of signal processing* 4: 265-271.
- **Hamilton, J.D.**, 1994, "Time series analysis", (*Princeton University Press, Princeton, NJ*).
- **Harvey, A.C.**, 1993, "Long memory in stochastic volatility", *unpublished working paper* (*London School of Economics, London, UK*).
- **Hassler, U. and J. Wolters**, 1994, "On the power of unit root tests against fractional alternatives", *Economics letters* 14: 1-5.
- **Hauser, M.A., B.M. Potscher, E. Reschenhofer**, 1999, "Measuring persistence in aggregate output: ARMA models, fractionally integrated ARMA models and nonparametric procedures", *Empirical economics* 24: 243-269.
- **Helson, J. and Y. Sarason**, 1967, "Past and future", *Mathematica Scandinavia* 21: 5-16.
- **Hipel, K.W. and A.I. McLeod**, 1978, "Preservation of the rescaled adjusted range", *Water resources research* 14: 508-518.
- **Hosking, J.R.M.**, 1981, "Fractional differencing", *Biometrika* 68: 165-176.
- **Hosking, J.R.M.**, 1984a, "Modelling persistence in hydrological time series using fractional differencing", *Water resources research* 20: 1898-1907.
- **Hosking, J.R.M.**, 1984b, "Asymptotic distributions of the sample mean, autocovariances and autocorrelations of long-memory time series", *Technical summary report 2752* (*Mathematics Research Center, University of Wisconsin, Madison, WI*).
- **Hosking, J.R.M.**, 1994, "Supplement to 'Asymptotic distributions of the sample mean, autocovariances and autocorrelations of long-memory time series'", *Research report RC19831* (*IBM Research Division, Yorktown Heights, NY*).

- **Hosking, J.R.M.**, 1996, "Asymptotic distributions of the sample mean, autocovariances, and autocorrelations of long-memory time series", *Journal of econometrics* 73: 261-284.
- **Hosoya, Y.**, 1996, "The quasi-likelihood approach to statistical inference on multiple time-series with long-range dependence", *Journal of econometrics* 73: 217-236.
- **Hsieh, C.K., C.F. Chung and C.F.J. Lin**, 1999, "Long-memory or regime changes: the Markov switching FIGARCH model", *unpublished working paper*, (National Taiwan University, Taipei, TW).
- **Hull, J.C.**, 1997, "Options, futures and other derivatives", (Prentice Hall, Upper Saddle River, NJ).
- **Hull, J.C. and A. White**, 1987, "The pricing of options on assets with stochastic volatilities", *Journal of finance* 42: 281-300.
- **Hurst, H.E.**, 1951, "Long-term storage capacity of reservoirs", *Transactions of the American Society of Civil Engineers* 116: 770-799.
- **Jarque, C.M. and A.K. Bera**, 1982, "Model specification tests: a simultaneous approach", *Journal of econometrics* 20: 59-82.
- **Jensen, M.J.**, 1999, "An approximate wavelet MLE of short and long-memory parameters", *Studies in nonlinear dynamics and econometrics* 3: 239-253.
- **Kernighan, B.W. and D.M. Ritchie**, 1988, "The C programming language (Second edition)", (Prentice Hall, Upper Saddle River, NJ).
- **Kullback, S. and L.A. Liebler**, 1951, "On information and sufficiency", *Annals of mathematical statistics* 22: 79-86.
- **Kwiatkowski, D., P.C.B. Phillips, P. Schmidt and Y. Shin**, 1992, "Testing the null hypothesis of stationarity against the alternative of an unit root: how sure we are that economic time series have a unit root?", *Journal of econometrics* 54: 159-178.
- **Laurent, S.**, 1999, "Procedure for FIGARCH conditional variances in GAUSS", *working paper* (Université de Liège, Liège, B).
- **Laurent, S. and J.P. Peters**, 2000, "G@RCH 1.1: An Ox package for estimating various ARCH models", *working paper* (Université de Liège, Liège, B).
- **Lee, D. and P. Schmidt**, 1996, "On the power of the KPSS test of stationarity against fractionally-integrated alternatives", *Journal of econometrics* 73: 285-302.

- **Lee, S.W. and B.E. Hansen**, 1994, "Asymptotic theory for the GARCH (1,1) quasi-maximum likelihood estimator", *Econometric theory* 10: 29-52.
- **Li, W.K. and A.I. McLeod**, 1986, "Fractional time series modelling", *Biometrika* 73: 217-221.
- **Lieberman, O., J. Rousseau and D.M. Zucker**, 2000, "Small-sample likelihood-based inference in the ARFIMA model", *Econometric theory* 16: 231-248.
- **Lin, J.Y.**, 1991, "Generalized integrated process and the aggregation of dynamic time series", *Academia economic papers* 19: 207-226.
- **Liu, M.**, 1995, "Modelling long memory in stock market volatility", *unpublished working paper*, (Duke university, Durham, NC).
- **Ljung, G. and G. Box**, 1979, "On a measure of lack of fit in time series models", *Biometrika* 66: 265-270.
- **Lo, A.W.**, 1991, "Long-term memory in stock market prices", *Econometrica* 59: 1279-1313.
- **Lo, A.W.**, 2000, "Finance: a selective survey", *unpublished working paper (M.I.T., Cambridge, MA)*.
- **Lo, A.W., H. Mamaysky and J. Wang**, 2000, "Foundations of technical analysis: computational algorithms, statistical inference, and empirical implementation", *Journal of finance* 55: 1705-1765.
- **Lobato, I. and P.M. Robinson**, 1996, "Averaged periodogram estimation of long memory", *Journal of econometrics* 73: 303-324.
- **Lombardi, M.J.**, 2002, "On the detection of long memory in the intra-daily pattern of stock returns", *Società Italiana di Statistica - Atti della XLI riunione scientifica - Sessioni spontanee*: 657-660, (CLEUP editore, Padova)..
- **Lombardi, M.J. and G.M. Gallo**, 2002, "Analytic Hessian matrices and the computation of FIGARCH estimates", *Statistical methods and applications* 11: 247-264.
- **Mandelbrot, B.B.**, 1963, "The variation of certain speculative prices", *Journal of business* 36: 394-419.
- **Mandelbrot, B.B.**, 1971 "When can price be arbitrated efficiently? A limit to the validity of the random walk and martingale models", *Review of economics and statistics* 53: 225-236.

- **Mandelbrot, B.B.**, 1977, "Fractals: form, chance and dimension", (*Freeman, San Francisco, CA*).
- **Mandelbrot, B.B. and J.W. Van Ness**, 1968, "Fractional Brownian motions, fractional noises and applications", *SIAM Revue* 10: 422-437.
- **Mantegna, R.N. and H.E. Stanley**, 2000, "An introduction to Econophysics", (*Cambridge University Press, Cambridge, UK*).
- **Maravall, A. and D. Pena**, 1992, "Missing observations and additive outliers in time series models", (*European University Institute, Badia Fiesolana, I*).
- **Marquardt, D.W.**, 1963, "An algorithm for least squares estimation of nonlinear parameters", *Journal of the society of industrial and applied mathematics* 11: 431-441.
- **McCoy, E.J. and A.T. Walden**, 1996, "Wavelet analysis and synthesis of stationary long-memory processes", *Journal of computational and graphical statistics* 5: 1-31.
- **McLeod, A.I. and K.W. Hipel**, 1978, "A reassessment of the Hurst phenomenon", *Water resources research* 14: 491-508.
- **Metelli, F.**, 1995, "Il rischio finanziario", (*Il Sole 24 Ore, Milano, I*).
- **Mills, T. C.**, 1990, "Time-series techniques for economists", (*Cambridge University Press, Cambridge, UK*).
- **Mitchell, M.**, 1998, "Introduzione agli algoritmi genetici", (*Apogeo, Milano, I*).
- **Mood, A.M., F.A. Graybill and D.C. Boes**, 1974, "Introduction to the theory of statistics (Third edition)", (*McGraw-Hill, London, UK*).
- **Nabeya, S. and K. Tanaka**, 1988, "Asymptotic theory of a test for the constancy of regression coefficients against the random walk alternative", *Annals of statistics* 16: 218-235.
- **Nabeya, S. and K. Tanaka**, 1990, "Limiting power of unit-root tests in time-series regression", *Journal of econometrics* 46: 247-271.
- **Neftci, S.N.**, 1996, "An introduction to the mathematics of financial derivatives", (*Academic Press, San Diego, CA*).
- **Nelson, D.B.**, 1990, "Stationarity and persistence in the GARCH (1,1) model", *Econometric theory* 6: 318-334.
- **Nelson, D.B.**, 1991, "Conditional heteroskedasticity in asset returns: a new approach", *Econometrica* 59: 347-370.



- **Nelson, D.B. and C.Q. Cao**, 1992, "Inequality constraints in the univariate GARCH model", *Journal of business and economic statistics* 10: 229-235.
- **Newey, W.K. and K.D. West**, 1987, "A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix", *Econometrica* 55: 703-708.
- **Parke, W.R.**, 1999, "What is fractional integration?", *Review of economics and statistics* 81: 632-638.
- **Paruolo, P.**, 1992, "Note sul problema della stima", (*CLUEB, Bologna, I*).
- **Patterson, K.**, 2000, "An introduction to applied econometrics – A time series approach", (*McMillan Press, Houndmills, UK*).
- **Phillips, P.C.B. and P. Perron**, 1988, "Testing for a unit root in time series regression", *Biometrika* 75: 335-346.
- **Porter-Hudak, S.**, 1990, "An application of the seasonally fractionally differenced model to the monetary aggregates", *Journal of the American Statistical Association* 85: 338-344.
- **Press, W.H., S.A. Teukolsky, W.T. Vetterling and B.P. Flannery**, 1992, "Numerical recipes in C: The art of scientific computing (Second edition)", (*Cambridge university press, Cambridge, UK*).
- **Rainville, E.D.**, 1960, "Special functions", (*McMillan Press, Houndmills, UK*).
- **Resnick, O.**, 1987, "Extreme values, regular variation and point processes", (*Springer-Verlag, New York, NY*).
- **Robinson, P.M.**, 1990, "Time series with strong dependence", *Advances in econometrics, Sixth world congress* (*Cambridge University Press, Cambridge, UK*).
- **Robinson, P.M.**, 1994, "Semiparametric analysis of long-memory time series", *Annals of statistics* 22: 515-539.
- **Robinson, P.M.**, 1995, "Log-periodogram regression of time series with long-range dependence", *Annals of statistics* 23: 1048-1072.
- **Robinson, P.M. and M. Henry**, 1999, "Long and short memory conditional heteroskedasticity in estimating the memory parameters of levels", *Econometric theory* 15: 299-336.
- **Rosenblatt, M.**, 1956, "A central limit theorem and a strong mixing condition", *Proceeding of the National Academy of Sciences* 42: 43-47.

- **Savin, E. and K. White**, 1977, "The Durbin-Watson test for serial correlation with extreme sample sizes or many regressors", *Econometrica* 45: 1989-1996.
- **Schwarz, G.**, 1978, "Estimating the dimension of a model", *Annals of statistics* 6: 461-464.
- **Schwert, W.**, 1990, "Stock volatility and the crash of '87", *Review of financial studies* 3: 77-102.
- **Shanno, D.F.**, 1970, "Conditioning of quasi-Newton methods for function minimization", *Mathematics of computation* 24: 647-656.
- **Sowell, F.B.**, 1986, "Fractionally integrated vector time series", *Ph.D. dissertation (Duke university, Durham, NC)*.
- **Sowell, F.B.**, 1990, "The fractional unit root distribution", *Econometrica* 58: 495-505.
- **Sowell, F.B.**, 1992a, "Maximum likelihood estimation of stationary univariate fractionally integrated time series models", *Journal of econometrics* 53: 165-188.
- **Sowell, F.B.**, 1992b, "Modelling long-run behavior with the fractional ARIMA model", *Journal of monetary economics* 29: 277-302.
- **Stroustrup, B.**, 1994, "The C++ language", (*Addison-Wesley, London, UK*).
- **Taylor, S.**, 1986, "Modelling financial time series", (*John Wiley & Sons, New York, NY*).
- **Teyssière, G.**, 1997, "Double long-memory financial time series", *QMW working paper 348 (University of London, London, UK)*.
- **Tieslau, M.A., P. Schmidt and R.T. Baillie**, 1996, "A minimum distance estimator for long-memory processes", *Journal of econometrics* 71: 249-264.
- **Tschernig, R.**, 1992, "Wechselkurse, Unsicherheit und Long Memory", *Ph.D. dissertation (University of Munich, Munich, D)*.
- **Tsay, W.J.**, 1998, "On the power of Durbin-Watson statistic against fractionally integrated processes", *Econometric reviews* 17: 361-386.
- **Tsay, W.J.**, 2000, "Long memory story of the real interest rate", *Economics letters* 67: 325-330.
- **Tse, Y.K.**, 1998, "The conditional heteroskedasticity of the Yen-Dollar exchange rates", *Journal of applied econometrics* 13: 49-56.
- **Weiss, A.A.**, 1986, "Asymptotic theory for ARCH models: estimation and testing", *Econometric theory* 2: 107-131.

- **White, H.**, 1982, “Maximum likelihood estimation of misspecified models”, *Econometrica* 50: 1-25.
- **Whittle, P.**, 1951, “Hypothesis testing in time series analysis”, (*Almquist and Wiksells, Uppsala, S*).
- **Zaffaroni, P.**, 2000, “Stationarity and memory of ARCH ( $\infty$ ) models”, *unpublished working paper (Banca d'Italia, Roma, I)*.
- **Zakoian, J.M.**, 1994, “Threshold heteroskedastic models”, *Journal of economic dynamics and control* 18: 931-955.

Copyright © 2005  
Marco J. Lombardi,  
Giampiero M. Gallo