

POLITECNICO DI MILANO

Facoltà di Ingegneria dei Sistemi
Corso di Studi in Ingegneria Gestionale



Elaborato di Laurea di secondo livello

**Processi long-memory e stime
semiparametriche: aspetti teorici e
applicazioni su dati finanziari
ultra-high frequency**

Relatore: Prof. Rocco Mosconi
Controrelatore: Prof. Carlo Sgarra

Claudio Cecchetto
Matricola 740071

Anno Accademico 2010/2011

**Processi long-memory e stime semiparametriche:
aspetti teorici e applicazione su dati finanziari
ultra-high frequency**

Claudio Cecchetto

Relatore: Prof. Rocco Mosconi

Controrelatore: Prof. Carlo Sgarra

RIASSUNTO. Nella presente tesi si analizzano i processi integrati frazionari ARFIMA(p, d, q) e le tecniche di stima semiparametriche dell'ordine di integrazione d . Tali tecniche, che operano nel dominio delle frequenze, sono in grado di focalizzarsi esclusivamente sul parametro d tralasciando le componenti AR(p) e MA(q), e pertanto forniscono stime più precise rispetto alle tecniche parametriche. Lo stimatore utilizzato viene proposto nell'articolo di P. M. Robinson (1995) *Gaussian semiparametric estimation of long range dependence*, nel quale si derivano le proprietà distributive asintotiche delle stime e si mostra come lo stimatore sia il più preciso tra gli stimatori semiparametrici. Alcuni aspetti critici, quali l'estensione delle stime a processi non-stazionari e la scelta della frequenza λ dello stimatore, sono stati analizzati nel corso della tesi, servendosi dell'ausilio della teoria nel primo caso e di simulazioni nel secondo. La possibilità di sfruttare la trasformata veloce di Fourier congiuntamente ad una programmazione opportuna hanno permesso di abbassare drasticamente i tempi computazionali necessari per eseguire le stime, spesso proibitivi nel caso di stimatori parametrici.

L'applicazione è stata condotta eseguendo inizialmente un'analisi univariata di serie storiche finanziarie ad alta frequenza, campionate ogni 5 secondi per un totale di diversi milioni di dati. Le serie considerate sono relative ai rendimenti e al flusso di ordini di titoli azionari quotati in Borsa Italiana: i risultati ottenuti evidenziano una sostanziale assenza di integrazione frazionaria per quanto riguarda i rendimenti ($d = 0$), e presenza di comportamento 'long memory' per quanto riguarda il flusso di ordini ($d > 0$).

Nella parte finale della tesi si propone un'analisi bivariata, in cui si testa l'eventuale presenza di cointegrazione frazionaria tra la serie dei prezzi e la serie del flusso di ordini cumulati. L'evidenza ottenuta mostra una assenza di cointegrazione frazionaria, in linea con la più generale assenza di cointegrazione suggerita dalla teoria economica. L'allineamento dell'ordine di integrazione delle due serie risulta necessario per l'analisi bivariata: in questo ambito è stato proposto un contributo originale, inerente alla necessità di eliminare la condizione iniziale di una serie prima dell'operazione di filtraggio.

Indice

Elenco delle figure	v
Elenco delle tabelle	vii
Capitolo 1. I processi integrati frazionari	1
1. Analisi nel dominio del tempo	1
2. Analisi nel dominio delle frequenze	4
3. Alcuni esempi	5
Capitolo 2. Stima semiparametrica Gaussiana	9
1. Descrizione del metodo di stima	9
2. Analisi Numerica in Robinson	10
3. Considerazioni teoriche	13
4. Analisi Numerica sul ruolo del parametro m	17
Capitolo 3. Processi non-stazionari	21
1. Elementi teorici sulle finestre	21
2. Serie storiche non-stazionarie e data tapering	27
Capitolo 4. Aspetti computazionali	35
1. Trasformata di Fourier veloce	35
2. Generazione di processi integrati frazionari	40
3. Metodo di Newton	41
4. Codici utilizzati	45
5. Tempi di calcolo ottenuti	48
Capitolo 5. Applicazione: analisi univariata	53
1. Rendimenti e order flow	53
2. Stime ottenute	56
3. Esercizi di simulazione	64
4. Analisi di sottocampioni anomali	67
Capitolo 6. Cointegrazione frazionaria	73
1. Stima del vettore di cointegrazione	73
2. Simulazioni Monte Carlo	75
Capitolo 7. Applicazione: analisi multivariata	79
1. Allineare gli ordini di integrazione	79
2. Analisi di cointegrazione	85
Appendice A. Approssimazione di Whittle	89

Elenco delle figure

1.1 Processo caratterizzato da $d = 0.3$ e relativa funzione di autocorrelazione	6
1.2 Periodogramma del processo caratterizzato da $d = 0.3$	6
1.3 Processo caratterizzato da $d = -0.3$ e relativa funzione di autocorrelazione	7
1.4 Periodogramma del processo caratterizzato da $d = -0.3$	7
2.1 Distribuzioni delle stime \hat{d} al variare di d (valori stazionari) e di T .	13
2.2 Distribuzioni delle stime \hat{d} al variare di d (valori non-stazionari) e di T .	14
2.3 Distribuzione della cumulata di processi con H differente.	15
2.4 Distribuzione della cumulata di un processo con H non-stazionario	15
2.5 Periodogramma medio e toerico e stime al variare di d e m .	19
2.6 Stime al variare di d , m e T .	20
3.1 Rappresentazione del Dirichlet kernel	24
3.2 Rappresentazione della lag windiw di Bartlett	25
3.3 Rappresentazione del Fejer kernel	25
3.4 Rappresentazione della lag window ‘full cosine bell’	26
3.5 Rappresentazione della lag window di Parzen	26
3.6 Funzione dei pesi $W(\theta)$ al variare di θ	27
3.7 Serie storica moltiplicata per le diverse finestre e relativo periodogramma	29
3.8 Distribuzione asintotica delle stime relative a processi stazionari e non-stazionari, variando il taper utilizzato	33
4.1 Trasformate di Fourier e periodogrammi calcolati mediante la notazione in Robinson e Cooley-Tukey	36
4.2 Funzione $R(H)$ al variare di H e di m .	42
4.3 Differenza $\Delta = H_g - H_N$ tra stime mediante metodo a griglia e metodo di Newton, variando H e il numero di passi	43
4.4 Esempi di convergenza del metodo di Newton in 4 simulazioni con H differenti	45
4.5 Tempi impiegati per generare processi integrati frazionari utilizzando i diversi metodi e relativo fit	49
4.6 Differenza tra periodogramma calcolato mediante definizione standard di trasformata di Fourier e mediante FFT	50
4.7 Tempi impiegati per il calcolo della FFT e relativo fit	51

5.1 Serie dei rendimenti r_t e dei prezzi p_t del titolo Enel	55
5.2 Funzione di autocorrelazione ρ_k e periodogramma $I(\lambda)$ dei rendimenti	55
5.3 Serie degli order flow o_t e degli order flow cumulati del titolo Enel	56
5.4 Funzione di autocorrelazione ρ_k e periodogramma $I(\lambda)$ degli order flow	56
5.5 Stime di d della serie r_t relativa al titolo Enel, variando frequenza e sottocampioni considerati	58
5.6 Stime di d della serie o_t relativa al titolo Enel, variando frequenza e sottocampioni considerati	60
5.7 Stime di d della serie r_t relativa al titolo Eni, variando frequenza e sottocampioni considerati	61
5.8 Stime di d della serie r_t relativa al titolo Fiat, variando frequenza e sottocampioni considerati	62
5.9 Stime di d della serie r_t relativa al titolo Snam Rete Gas, variando frequenza e sottocampioni considerati	63
5.10 Stime di d relative ad un white noise, ad un processo MA(5) e ad un processo MA(10)	67
5.11 Stime, rendimenti, periodogramma e funzione di autocorrelazione del sottocampione 1	68
5.12 Stime, rendimenti, periodogramma e funzione di autocorrelazione del sottocampione 24	69
5.13 Stime, rendimenti, periodogramma e funzione di autocorrelazione del sottocampione 2	69
5.14 Stime, rendimenti, periodogramma e funzione di autocorrelazione del sottocampione 5	70
7.1 Differenza tra serie integrata con e senza la condizione iniziale (ordine di integrazione minore, parametri noti)	81
7.2 Differenza tra serie integrata con e senza la condizione iniziale (ordine di integrazione maggiore, parametri noti)	81
7.3 Differenza tra serie integrata con e senza la condizione iniziale (ordine di integrazione minore, parametri incogniti)	83
7.4 Differenza tra serie integrata con e senza la condizione iniziale (ordine di integrazione maggiore, parametri incogniti)	83
7.5 Order flow cumulati e rendimenti cumulati allineati al medesimo ordine di integrazione	86
7.6 Stime \hat{d} delle serie filtrate al medesimo ordine di integrazione	87

Elenco delle tabelle

2.1 Distorsione $\hat{d} - d$ al variare di d e m .	11
2.2 Distorsione $\hat{d} - d$ al variare di d , m e n .	12
2.3 Confronto tra la distorsione $\hat{d} - d$ in Robinson e nelle simulazioni.	12
2.4 Periodogrammi sperimentali e teorici a confronto.	17
3.1 Media e deviazione standard di stime \hat{d} effettuate variando d e il tipo di taper utilizzato	30
3.2 Stime \hat{d} relative a processi stazionari e non-stazionari, variando T , m e il taper utilizzato	32
4.1 Media delle stime \bar{H}_g e \bar{H}_N	44
4.2 Precisione del metodo di Newton	46
4.3 Tempo espresso in secondi per generare serie storiche di lunghezza T indicata utilizzando la rappresentazione MA(∞)	49
4.4 Tempo espresso in secondi per generare serie storiche di lunghezza T indicata utilizzando l'algoritmo di Davies e Harte	49
4.5 Tempo per il calcolo del periodogramma tramite definizione standard di trasformata di Fourier	50
4.6 Tempo espresso in secondi per calcolare il periodogramma, variando il parametro T e m , utilizzando l'algoritmo FFT	51
4.7 Tempo espresso in secondi per il calcolo della trasformata di Fourier mediante algoritmo FFT.	51
4.8 Tempo espresso in secondi per stimare d alla quarta cifra decimale utilizzando la griglia di valori	52
4.9 Tempo espresso in secondi per stimare d utilizzando il metodo di Newton	52
5.1 Stime di d della serie r_t relativa al titolo Enel alla frequenza $\lambda = 0.05$	57
5.2 Stime di d della serie o_t relativa al titolo Enel alla frequenza $\lambda = 0.05$	59
5.3 Stime di d della serie r_t relativa al titolo Eni alla frequenza $\lambda = 0.05$	59
5.4 Stime di d della serie r_t relativa al titolo Fiat alla frequenza $\lambda = 0.05$	64
5.5 Stime di d della serie r_t relativa al titolo Snam Rete Gas alla frequenza $\lambda = 0.05$	64
5.6 Stime di d della serie r_t relativa al titolo Snam Rete Gas alla frequenza $\lambda = 0.05$	64

6.1	Analisi numerica sulla stima del parametro β	75
7.1	Stime di β e degli ordini di integrazione considerando o meno la condizione iniziale (parametri noti)	80
7.2	Stime di β e degli ordini di integrazione considerando o meno la condizione iniziale (parametri incogniti)	82
7.3	Stime di β e degli ordini di integrazione considerando o meno la condizione iniziale (parametro μ noto e Δ incognito)	84
7.4	Stime di β e degli ordini di integrazione considerando o meno la condizione iniziale (parametro μ incognito e Δ noto)	84
7.5	Stime di β e degli ordini di integrazione considerando o meno la condizione iniziale (parametri noti)	85
7.6	Risultati dell'analisi di cointegrazione	87

I processi integrati frazionari

1. Analisi nel dominio del tempo

La famiglia di modelli ARIMA (autoregressive integrated moving-average), largamente utilizzata nell'analisi di serie storiche, prevede l'utilizzo di un valore intero per descrivere l'ordine di integrazione di un processo. L'utilizzo di un valore non intero (frazionario) è invece previsto nei modelli ARFIMA (autoregressive fractional integrated moving-average), dove l'operatore differenza frazionario viene dato dall'espansione binomiale in potenze dell'operatore ritardo. Questi processi sono caratterizzati da 'long memory', nel senso che la dipendenza tra osservazioni molto distanti tra loro nel tempo non è trascurabile, al contrario di quel che accade nella modellazione più tradizionale in cui si presuppone indipendenza tra osservazioni molto distanti nel tempo. Una descrizione formale dei modelli ARFIMA viene data per la prima volta e in maniera indipendente in Hosking [1] e Granger & Joyeux [2], quest'ultimo articolo di stampo più econometrico rispetto al primo.

Nei modelli ARFIMA (p, d, q) il parametro d indica l'ordine di integrazione del processo, può assumere qualunque valore e se compreso nell'intervallo $-\frac{1}{2} \leq d \leq \frac{1}{2}$ rappresenta un processo è stazionario.

Per comprendere le implicazioni dell'utilizzo di d frazionario, iniziamo con il considerare un processo x_t ARIMA $(0,1,0)$, detto comunemente random walk: esso è definito come

$$\Delta x_t = (1 - L)x_t = \varepsilon_t,$$

dove L è l'operatore ritardo ($Lx_t = x_{t-1}$) e ε_t è un processo indipendente e identicamente distribuito. L'integrazione frazionaria prevede l'utilizzo dell'operatore Δ^d , dato dal seguente sviluppo binomiale:

$$\Delta^d = (1 - L)^d = \sum_{k=0}^{\infty} \binom{d}{k} (-L)^k = 1 - dL - \frac{1}{2}d(1-d)L^2 - \frac{1}{6}d(1-d)(2-d)L^3 - \dots$$

Il processo ARFIMA $(0,d,0)$ è definito come

$$(1.1) \quad \Delta^d x_t = \varepsilon_t,$$

dove ε_t è un processo indipendente e identicamente distribuito. Se $d < \frac{1}{2}$, il processo x_t è stazionario e ammette la rappresentazione MA(∞) data da

$$x_t = (1 - L)^{-d} \varepsilon_t = \sum_{j=0}^{\infty} b_j \varepsilon_{t-j},$$

dove

$$(1.2) \quad b_j = \frac{d(1+d) \dots (k-1+d)}{k!} = \frac{(k+d-1)!}{k!(d-1)!}.$$

Spesso in letteratura si utilizza una notazione che si serve della funzione gamma $\Gamma(\cdot)$: tale funzione, il cui dominio è l'insieme dei numeri reali ed immaginari ad eccezione degli interi negativi, estende il concetto di fattoriale ai numeri non interi ed è legata al fattoriale di un numero intero positivo dalla relazione

$$\Gamma(n + 1) = n! .$$

Si può ora riscrivere la ?? come

$$(1.3) \quad b_j = \frac{\Gamma(j + d)}{\Gamma(d)\Gamma(j + 1)},$$

che può essere semplificata utilizzando la formula di Sheppard, secondo la quale per valori di j grandi $\Gamma(j + a)/\Gamma(j + b)$ è ben approssimabile da j^{a-b} , come

$$b_j \simeq A j^{d-1},$$

con j grande e dove A è una opportuna costante (pari evidentemente a $\Gamma(d)^{-1}$). Usando quest'ultima formulazione si consideri il processo y_t nella sua rappresentazione MA(∞) dato da

$$y_t = A \sum_{j=1}^{\infty} j^{d-1} \varepsilon_t - j + \varepsilon_t.$$

Si può vedere come la varianza del processo y_t sia pari a

$$\sigma_y^2 = A^2 \sigma_\varepsilon^2 \left(1 + \sum_{j=1}^{\infty} j^{2(d-1)} \right),$$

e dato che

$$\sum_{j=1}^{\infty} j^{-s}$$

converge con $s > 1$, si capisce come $d < \frac{1}{2}$ sia condizione necessaria per la stazionarietà di y_t . La stessa cosa vale per x_t , dato che le varianze dei due processi differiscono solamente per un valore finito.

Se $d > -\frac{1}{2}$ il processo x_t è invertibile e può essere descritto tramite la rappresentazione AR(∞) data da

$$(1 - L)^d x_t = \sum_{j=0}^{\infty} a_j x_{t-j} = \varepsilon_t,$$

dove

$$(1.4) \quad a_j = \frac{\Gamma(j - d)}{\Gamma(1 - d)\Gamma(j + 1)}.$$

In questo caso l'approssimazione per j grande è data da

$$|a_j| \simeq A j^{-(1+d)},$$

dove A è un'opportuna costante. Si può notare come i coefficienti b_j abbiano tutti lo stesso segno di d , mentre quelli a_j hanno segno opposto (ad eccezione di $a_0 = b_0 = 1$). Da un punto di vista notazionale, il filtro $b(L) = (1 - L)^{-d}$ viene definito un filtro integrante di ordine d (esso alza l'ordine di integrazione della serie a cui è applicato) mentre il filtro $a(L) = (1 - L)^d$ è un filtro integrante di ordine

– d . Nel seguito della tesi con l'espressione ‘applicare un filtro integrante di ordine d ’ alla serie y_t si intenderà generare la serie y_t^* data da

$$y_t^* = (1 - L)^{-d} y_t.$$

Ad esempio se y_t è un white noise, il processo y_t^* risultante sarà un processo ARFIMA $(0, d, 0)$.

Se il parametro d appartiene all'intervallo $-1 < d < \frac{1}{2}$ la funzione di autocorrelazione ρ_k è data da

$$(1.5) \quad \rho_k = \frac{\Gamma(1-d)}{\Gamma(d)} \frac{\Gamma(k+d)}{\Gamma(k+1-d)},$$

dove k indica il lag. La formulazione approssimata è data da

$$(1.6) \quad \rho_k \simeq A k^{2d-1},$$

valida per k grande e $d \neq 0$, dove A è un'opportuna costante. Considerando che in un modello stazionario ARMA i coefficienti di autocorrelazione ρ_k seguono la legge

$$\rho_k \sim A \theta^k,$$

con $|\theta| < 1$ e k grande: questi valori tendono a zero esponenzialmente, molto più velocemente rispetto ai valori dati in (1.6), che tendono a zero iperbolicamente. Proprio questa discesa lenta dei coefficienti verso zero rende non trascurabile la dipendenza che osservazioni anche molto lontane nel tempo hanno tra di loro: da qui nasce l'espressione ‘long memory’, con la quale si descrive questo aspetto dei processi ARFIMA.

Nonostante precedentemente siano stati introdotti degli intervalli per i valori assumibili dal parametro d (per la stazionarietà del processo e per la definizione di funzione di autocorrelazione), si vuole fare presente che le formule (1.3) e (1.4) sono in grado di descrivere i coefficienti b_j e a_j relativamente a processi non stazionari (di qualunque ordine d di integrazione, tranne alcune eccezioni come si vedrà tra poco). Una maniera particolarmente utile per calcolare i suddetti coefficienti prevede l'utilizzo di formule ricorsive, che evitano il dover ricorrere alla funzione $\Gamma(\cdot)$. Data la seguente proprietà della funzione $\Gamma(\cdot)$

$$\Gamma(z+1) = z\Gamma(z),$$

si possono definire ad esempio i coefficienti b_j come

$$b_j = \frac{\Gamma(j+d)}{\Gamma(d)\Gamma(j+1)} = \frac{(j-1+d)\Gamma(j-1+d)}{\Gamma(d)j\Gamma(j)}.$$

Dato che

$$b_{j-1} = \frac{\Gamma(j-1+d)}{\Gamma(d)\Gamma(j)},$$

si nota come una formula ricorsiva può essere ottenuta ponendo

$$(1.7) \quad b_j = -b_{j-1} \frac{1-d-j}{j},$$

con inizializzazione $b_0 = 1$. Tale formula permette il grosso vantaggio di non dover utilizzare direttamente la funzione $\Gamma(\cdot)$ (solitamente non più calcolabile dai tradizionali calcolatori a partire da valori di j di poche centinaia) e di non dover ricorrere a formule approssimate; inoltre permette di calcolare i coefficienti b_j relativi a tutti i processi ARFIMA $(0, d, 0)$ con qualunque ordine di integrazione, anche intero negativo (ciò non è possibile utilizzando la notazione in (1.3)).

2. Analisi nel dominio delle frequenze

I processi frazionari integrati vengono spesso analizzati nel dominio delle frequenze. Consideriamo il processo x_t , con $t = 0, \pm 1, \dots$: se tale processo è stazionario in covarianza, ovvero se $E(x_t) = \mu$ e $\text{Cov}(x_t, x_{t+k}) = \gamma(k)$ non dipendono da t , allora x_t ha una densità spettrale $f(\lambda)$, o semplicemente spettro, definita mediante l'utilizzo della trasformata di Fourier come

$$f(\lambda) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \gamma(k) \exp^{-ik\lambda},$$

dove $\pi \leq \lambda \leq \pi$ è una variabile reale ed indica la frequenza. Tra funzione di covarianza e spettro esiste una corrispondenza biunivoca ($f(\lambda) \leftrightarrow \gamma(k)$): data $f(\lambda)$, $\gamma(k)$ può essere ricavata mediante l'antitrasformata di Fourier come

$$\gamma(k) = \int_{-\pi}^{+\pi} f(\lambda) \exp ik\lambda d\lambda.$$

La densità spettrale $f(\lambda)$ è una funzione non negativa, reale, simmetrica rispetto all'origine e periodica di periodo 2π : inoltre l'area sottesa dalla funzione $f(\lambda)$ tra $-\pi$ e π è pari alla varianza del processo. L'integrale limitato all'intervallo $(\lambda, \lambda + d\lambda)$ fornirà invece il quantitativo di varianza imputabile a quel determinato intervallo; l'analisi spettrale è quindi in grado di misurare l'importanza di determinate frequenze, ad esempio è in grado di separare il contributo dato da frequenze basse piuttosto che da frequenze alte alla varianza complessiva del processo. Spesso ciò che interessa maggiormente nello studio dei processi integrati è infatti la forma dello spettro a basse frequenze, responsabili del comportamento di lungo periodo del processo: come si vedrà nel seguito, si cercherà quindi di modellizzare il comportamento dello spettro nell'intorno di $\lambda = 0$. Si vuole infine precisare che non esiste un approccio preferibile all'altro tra analisi nel dominio del tempo e analisi nel dominio delle frequenze: data la biunivocità tra $f(\lambda)$ e $\gamma(k)$ è come se di fatto i due approcci fornissero lo stesso quantitativo di informazioni.

La definizione di spettro sopra data prevede la conoscenza della funzione di covarianza $\gamma(k)$, dalla quale si ricava $f(\lambda)$ utilizzando la formula della trasformata: tuttavia questa procedura è molto complessa e di solito risulta più conveniente ricorrere alla formulazione

$$f(\lambda) = \frac{\sigma^2}{2\pi} |W(\exp^{k\lambda})|^2,$$

dove $W(\cdot)$ è la funzione di trasferimento e σ^2 è la varianza del processo. Introducendo la variabile complessa $z = \exp^{k\lambda}$, si può riscrivere il risultato precedente nel noto teorema fondamentale dell'analisi spettrale, il quale afferma che se un sistema stabile caratterizzato da funzione di trasferimento $W(z)$ è alimentato da un processo caratterizzato da spettro complesso $\Phi_u(z)$, l'uscita è un processo stazionario con spettro complesso $\Phi_y(z)$ dato da

$$\Phi_y(z) = W(z)W(z^{-1})\Phi_u(z).$$

Ad esempio consideriamo la serie x_t , integrata di ordine $I(1)$, che differenziata una volta genera la serie $y_t \sim I(0)$, ovvero

$$(1 - L)x_t = y_t.$$

Lo spettro $f_x(\lambda)$ di x_t può essere ricavato da quello $f_y(\lambda)$ di y_t mediante il teorema appena descritto: notando che la funzione di trasferimento $W(z)$ è pari a $(1-z)^{-1}$, si può dire che

$$f_x(\lambda) = |(1-z)^{-1}|^2 f_y(\lambda).$$

Tralasciando per un attimo l'esponente negativo, $|1-z|^2$ si risolve nel seguente modo:

$$\begin{aligned} |1-z|^2 &= (1-z)(1-z^{-1}) \\ &= 2 - z - z^{-1} \\ &= 2 - e^{-i\lambda} - e^{i\lambda} \\ &= 2 - 2\cos(\lambda) \\ &\simeq |\lambda|^2, \end{aligned}$$

dove l'approssimazione nell'ultimo passaggio è valida per $\lambda \rightarrow 0$. Riconsiderando l'esponente negativo, tale risultato diventa $|\lambda|^{-2}$, e lo spettro di x_t nell'intorno dell'origine risulta dato da

$$f_x(\lambda) \simeq c|\lambda|^{-2},$$

dove c è una costante positiva, pari a $\lim_{\lambda \rightarrow 0} f_y(\lambda)$. Tale scrittura può essere ora generalizzata ad un processo con generico ordine di integrazione $I(d)$: la funzione di trasferimento sarà data da $W(z) = (1-z)^{-d}$, e ripercorrendo tutti i passaggi precedenti si trova che $|1-z|^{-2d} \simeq |\lambda|^{-2d}$. Il processo integrato frazionario x_t ha quindi spettro nell'intorno dell'origine pari a

$$f_x(\lambda) \simeq c|\lambda|^{-2d},$$

dove c è a costante prima descritta. Il processo x_t è comunemente detto 'long memory' se

$$f(0) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \gamma(k) = \infty,$$

ovvero se $f(\lambda)$ ha un polo in $\lambda = 0$. Se invece il processo ha uno zero in $\lambda = 0$,

$$f(0) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \gamma(k) = 0$$

e il processo x_t viene detto anti-persistente.

3. Alcuni esempi

Mostriamo in figura (1.4) un processo x_t ARFIMA(0, 0.3, 0) e la relativa funzione di autocorrelazione campionaria e teorica. Il processo è ovviamente caratterizzato da lunghezza finita ($t = 1, \dots, T = 2^{20}$), e in figura vengono mostrate solo 1000 di queste osservazioni. La funzione di autocorrelazione è stata invece calcolata sfruttando tutte le osservazioni disponibili. Si possono notare le caratteristiche tipiche dei processi integrati viste nelle sezioni precedenti: ad esempio si nota come il processo x_t sia qualitativamente persistente (le osservazioni cambiano di segno più lentamente rispetto ad un processo random walk), e come la funzione di autocorrelazione decada molto lentamente, in maniera iperbolica. In figura (??) si mostra il periodogramma $I(\lambda)$ del processo x_t (il periodogramma è una stima dello spettro):

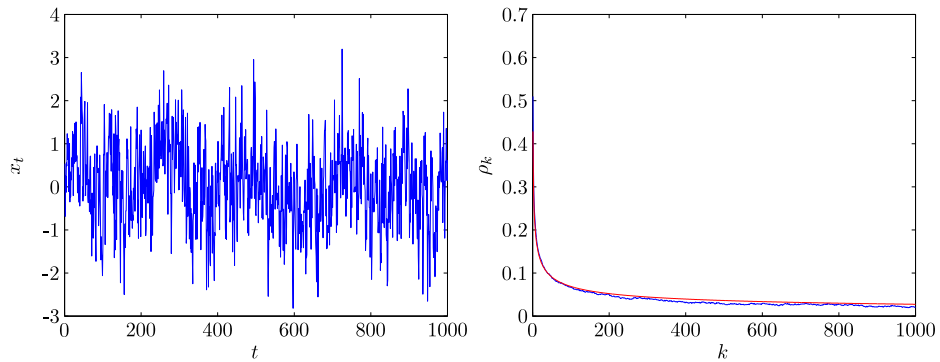


FIGURA 1.1. A sinistra: processo x_t caratterizzato da $d = 0.3$. A destra: funzione di autocorrelazione campionaria (in blu) e teorica (in rosso).

effettivamente gli spettri dei processi ‘long memory’ sono caratterizzati da un polo in $\lambda = 0$.

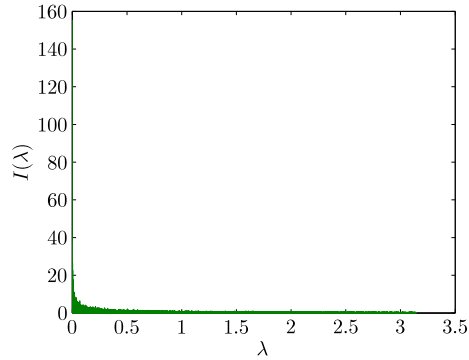


FIGURA 1.2. Periodogramma $I(\lambda)$ del processo x_t con $d = 0.3$.

Ora si genera in maniera analoga al caso precedente un processo x_t anti-persistente ARFIMA(0,-0.3,0) e si mostrano in figura (1.3) 1000 osservazioni di x_t e la funzione di autocorrelazione campionaria e teorica. Si nota come in questo caso il processo x_t sia anti-persistente (oscillazioni più frequenti rispetto ad un white noise). Anche in questo caso il decadimento dei coefficienti di autocorrelazione è più lento rispetto ai processi che ammettono una rappresentazione ARMA (anche se questo aspetto non è così evidente come nel caso ‘long memory’). Infine si mostra il periodogramma $I(\lambda)$ del processo x_t : si può notare come lo spettro di un processo anti-persistente sia caratterizzato da uno zero in $\lambda = 0$.

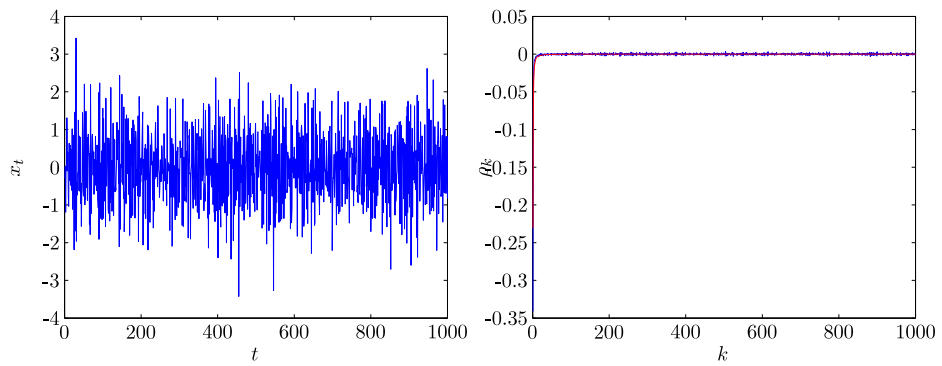


FIGURA 1.3. A sinistra: processo x_t caratterizzato da $d = -0.3$. A destra: funzione di autocorrelazione campionaria (in blu) e teorica (in rosso).

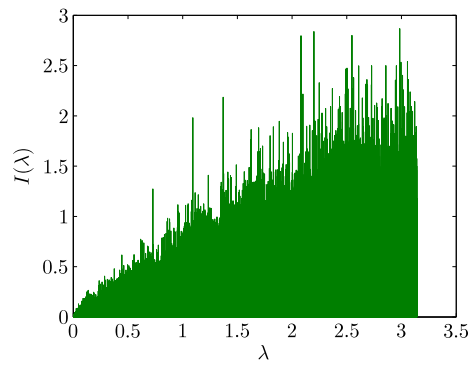


FIGURA 1.4. Periodogramma $I(\lambda)$ del processo x_t con $d = -0.3$.

Stima semiparametrica Gaussiana

1. Descrizione del metodo di stima

La stima semiparametrica Gaussiana è stata proposta da Robinson [3] e si basa sulla forma che assume la densità spettrale di un processo stazionario x_t nelle vicinanze della frequenza $\lambda = 0$. Il parametro stimato viene indicato con H (esponente di Hurst), ed è legato al parametro d dalla relazione $H = d + 0.5$. Nel seguito si cercherà di utilizzare d in conformità con le definizioni date nel capitolo precedente. Il termine semiparametrico si riferisce al fatto che non viene specificato in maniera completa il modello che descrive il comportamento del processo x_t , ma si ci limita ad analizzare tale comportamento solo a basse frequenze. In altre parole, non viene definita la forma funzionale della densità spettrale nell'intero intervallo $(-\pi, \pi]$, ma solo nell'intorno di $\lambda = 0$. Nello specifico, viene assunto che

$$(2.1) \quad f(\lambda) \sim G\lambda^{-d} \quad \text{con} \quad \lambda \rightarrow 0^+,$$

dove $f(\lambda)$ indica la densità spettrale, $G \in (0, \infty)$ è una costante e $d \in (-\frac{1}{2}, \frac{1}{2})$ è l'ordine di integrazione. Quando $d \in (0, \frac{1}{2})$, il processo è detto essere persistente, e $f(\lambda)$ tende ad infinito alla frequenza zero, mentre quando $d \in (-\frac{1}{2}, 0)$ il processo è detto antipersistente e $f(\lambda)$ tende a 0. Infine, quando $d = 0$, il processo è aleatorio e $f(\lambda)$ tende ad una costante finita positiva in $\lambda = 0$.

La stima di d viene fatta basandosi su una forma approssimata di verosimiglianza Gaussiana nel dominio delle frequenze, suggerita dagli studi di Künsch (1986). Approfondimenti sulla funzione di verosimiglianza vengono forniti in Appendice 1. Per prima cosa vengono definiti la trasformata discreta di Fourier $w(\lambda)$ e il periodogramma $I(\lambda)$ di x_t come

$$(2.2) \quad w(\lambda) = (2\pi n)^{-1/2} \sum_{t=1}^n x_t e^{it\lambda}, \quad I(\lambda) = |w(\lambda)|^2.$$

Le frequenze λ sono definite come

$$(2.3) \quad \lambda_j = \frac{2\pi j}{n},$$

con $j = 1, \dots, m$, dove m è un intero minore di $\frac{1}{2}n$. Il parametro m , la cui scelta è arbitraria, definisce la larghezza della banda che si sta considerando, la cui frequenza più alta risulta appunto $2\pi \frac{m}{n}$. La stima di d non viene definita in forma chiusa, ma si ricorre alla funzione

$$(2.4) \quad Q(G, H) = \frac{1}{m} \sum_{j=1}^m \left\{ \log G\lambda_j^{-2d} + \frac{\lambda_j^d}{G} I_j \right\},$$

dove $I_j = I(\lambda_j)$. Le stime \hat{G} e \hat{d} corrispondono a quei particolari valori di G e d che minimizzano tale funzione, ovvero

$$(\hat{G}, \hat{d}) = \arg \min_{-\frac{1}{2} < d < \frac{1}{2}, 0 < G < \infty} Q(G, d).$$

La funzione Q può essere riscritta in modo da dover stimare unicamente \hat{d}

$$\hat{d} = \arg \min_{-\frac{1}{2} < d < \frac{1}{2}} R(d),$$

dove

$$(2.5) \quad R(d) = \log \hat{G}(d) - 2d \frac{1}{m} \sum_{j=1}^m \log \lambda_j, \quad \hat{G}(d) = \frac{1}{m} \sum_{j=1}^m \lambda_j^{2d} I_j.$$

Sempre in Appendice 1 vengono illustrati i passaggi che portano alla formulazione di $R(d)$. Le proprietà distributive asintotiche di questa stima semiparametrica derivano dai lavori precedenti di Fox e Taqu (1986), Dahlhaus (1989) e Giraitis e Surgailis (1990), che hanno analizzato il modello di comportamento della densità spettrale dato dalla (2.1) ed esteso a tutto il dominio delle frequenze $(-\pi, \pi)$, ed hanno trovato proprietà di consistenza con tasso $n^{1/2}$, efficienza e normalità asintotiche. Tale modello, in quanto parametrico, potrebbe risultare inconsistente qualora mal specificato, ovvero nel caso in cui il processo abbia ad alte frequenze un comportamento diverso da quello modellizzato; per tale ragione sono stati proposti vari metodi di stima semiparametrici, come Geweke e Porter-Hudak (1983) e Robinson (1995). Robinson ha dimostrato che la stima semiparametrica sopra descritta gode delle proprietà di consistenza con tasso $m^{1/2}$, efficienza e normalità asintotiche che dominano le altre stime semiparametriche sia per migliore efficienza asintotica sia per assunzioni meno forti. Indicando con d_0 il vero valore del parametro, sotto determinate assunzioni vale infatti che

$$\hat{d} \rightarrow_p d_0 \quad \text{per } n \rightarrow \infty,$$

$$m^{1/2}(\hat{d} - d_0) \rightarrow_d N(0, \frac{1}{4}).$$

2. Analisi Numerica in Robinson

In questa sezione vengono riproposte le simulazioni Monte Carlo presenti nel lavoro di Robinson, dove sono state confrontate le stime \hat{d} di serie integrate frazionarie con le stime \tilde{d} effettuate con una versione semplificata dello stimatore semiparametrico di Geweke e Porte-Hudak (1983). Nelle simulazioni vengono fatti variare la lunghezza n delle serie storiche (che assume i valori $n = 64, 128$ e 256), il parametro m (che assume i valori $m = n/16, n/8$ e $n/4$) e il parametro d (che assume i valori $d = -0.4, -0.2, 0, 0.2$ e 0.4); per ciascuna combinazione vengono effettuate 5000 repliche. I risultati mostrano come \hat{d} abbia una deviazione standard sempre decisamente inferiore rispetto a quella di \tilde{d} , mentre l'interpretazione risulta più difficoltosa per la distorsione: in alcuni casi è \tilde{d} ad essere meno distorto, soprattutto nei casi con $n = 64$ e 128 e $m = n/16$ e $n/8$, mentre per $n = 256$ e $m = n/16$ è \hat{d} a presentare minore distorsione. Tralasciando il confronto tra i due stimatori e soffermandosi solo su \hat{d} , ci sono alcuni aspetti di incertezza riassunti nelle seguenti considerazioni:

- non è chiaro se la distorsione aumenti o diminuisca all'allontanarsi dal valore $d = 0$, come mostrano i valori in tabella (2.1) che riportano i dati presenti nell'articolo;

$N = 256$			
d	$m = n/16$	$m = n/8$	$m = n/4$
-0.4	0.006	-0.044	-0.081
-0.2	-0.016	-0.026	-0.037
0	-0.022	-0.013	-0.005
0.2	-0.020	-0.001	0.009
0.4	-0.034	-0.007	0.016

TABELLA 2.1. Distorsione $\hat{d} - d$ al variare di d e m .

- sempre facendo riferimento alla tabella (2.1) non risulta chiaro se è presente una relazione tra il valore m e l'entità della distorsione;
- infine bisognerebbe capire come mai la distorsione a volte è positiva e diventa negativa al crescere di m (come per $d = -0.4$), altre volte è negativa e diventa positiva sempre al crescere di m (come per $d = 0.4$);

Sembrerebbe quindi fondamentale riuscire a capire il ruolo che ha il parametro m all'interno dell'algoritmo di stima e quale sia il suo valore ottimale, questione che è stata lasciata aperta da Robinson nel suo articolo.

Allo scopo di comprendere i punti sopra elencati e di capire se la stima possa essere estesa anche a processi non stazionari (ovvero con $d > \frac{1}{2}$) vengono presentate delle nuove serie di simulazioni Monte Carlo. Sono state scelte serie storiche di lunghezza $n = 256, 2048$ e 8192 , il parametro m è stato scelto pari a $m = n/4, n/16$ e $n/128$, e d è stato scelto pari a $0, 0.4, 0.8$ e 1.2 . Per ciascuna combinazione sono state effettuate 50000 ripetizioni. In tabella (2.2) vengo riassunti i risultati: nella prima colonna vengono indicati il parametro m in funzione di n e in piccolo la larghezza della banda considerata, mentre all'interno delle celle oltre alla distorsione viene indicato in piccolo il valore m . Leggendo per colonna quindi la banda considerata è costante, e all'aumentare di n (e conseguentemente di m), aumenta la risoluzione con cui si campiona all'interno della banda.

Anche in questo caso permangono gli stessi elementi di incertezza riguardo al ruolo del parametro m visti precedentemente:

- solo per $d = 0$ all'aumentare di m , indipendentemente dalla larghezza di banda considerata, diminuisce la distorsione;
- per $d = 0.4$ la distorsione sembra diminuire, fissata la banda, all'aumentare di n , anche se compare qualche valore anomalo; sembra inoltre diminuire, fissata la lunghezza n della serie, al diminuire della banda, anche in questo caso con qualche risultato anomalo;
- per i valori di d non stazionari lo stimatore sembra funzionare meno bene, in quanto all'aumentare di n la distorsione aumenta; il caso $d = 1.2$ è peggiore rispetto al caso $d = 0.8$.

In corrispondenza di $n = 256$ non sono state eseguite simulazioni per $m = n/128$ a causa dell'insorgere di problemi computazionali, del tutto comprensibili dato che la stima si basa su un campione di soli due elementi. La scelta di $n = 256$ è dovuta principalmente alla volontà di avere un confronto con le simulazioni presenti in

		$m = n/4$	$\pi/2$	$m = n/16$	$\pi/8$	$m = n/128$	$\pi/64$
$d = 0$	$N = 256$	-0.0066	64	-0.0215	16		
	$N = 2048$	-0.0010	512	-0.0033	128	-0.0210	16
	$N = 8192$	-0.0003	2048	-0.0010	512	-0.0068	64
$d = 0.4$	$N = 256$	0.0211	64	-0.0087	16		
	$N = 2048$	0.0194	512	0.0309	128	-0.0095	16
	$N = 8192$	0.0182	2048	0.0028	512	0.0009	64
$d = 0.8$	$N = 256$	-0.0391	64	0.007	16		
	$N = 2048$	-0.0353	512	0.0130	128	0.0108	16
	$N = 8192$	-0.0386	2048	0.0101	512	0.0193	64
$d = 1.2$	$N = 256$	-0.1475	64	-0.1228	16		
	$N = 2048$	-0.1615	512	-0.1394	128	-0.1205	16
	$N = 8192$	-0.1726	2048	-0.1422	512	-0.1242	64

TABELLA 2.2. Distorsione $\hat{d} - d$ al variare di d , m e n .

Robinson (1995), e dal confronto i risultati sembrano coincidere. Si noti soltanto che la stima di H in Robinson (1995) è vincolata ai valori $-\frac{1}{2} < d < \frac{1}{2}$, mentre qua è libera. I valori modificati (sostituendo $d = \frac{1}{2}$ ad ogni $d > \frac{1}{2}$ trovato) sono illustrati in tabella (2.3), e si nota come coincidano con quelli presenti in Robinson.

		$N = 256$	
d		$m = n/4$	$m = n/16$
0	Rob	-0.0050	-0.0220
	Sim	-0.0066	-0.0215
0.4	Rob	0.0160	-0.0340
	Sim	0.0165	-0.0365

TABELLA 2.3. Confronto tra la distorsione $\hat{d} - d$ in Robinson e nelle simulazioni.

Di seguito vengono proposte le distribuzioni di d per $m = T/4$, al variare di T e d . Si può notare come le distribuzioni si comportino in maniera diversa a seconda della stazionarietà o meno del processo:

- nel caso di processi stazionari (figura (2.1)), simulati per $d = -0.35$, $d = 0$ e $d = 0.4$, la distribuzione sembra molto prossima ad una normale (anche se in 4 casi su 6 il P_value suggerisce di rifiutare l'ipotesi di normalità). Si ha comunque in tutti e tre i casi un miglioramento della normalità al crescere della lunghezza T delle serie considerate;
- nel caso di processi non stazionari ($d = 0.8$ e $d = 1.2$), la distribuzione appare molto diversa da una normale, come si può vedere dalla figura (2.2): inoltre al crescere della lunghezza T la normalità non sembra migliorare. Il caso con $d = 0.8$ (non stazionario ma meno di una random walk) non sembra così critico come il caso $d = 1.2$.

Sembra quindi che lo stimatore risulti poco adatto per i processi non stazionari: una motivazione potrebbero essere data, come illustrato più avanti, dalle leggi diverse che governano i due tipi di processo.

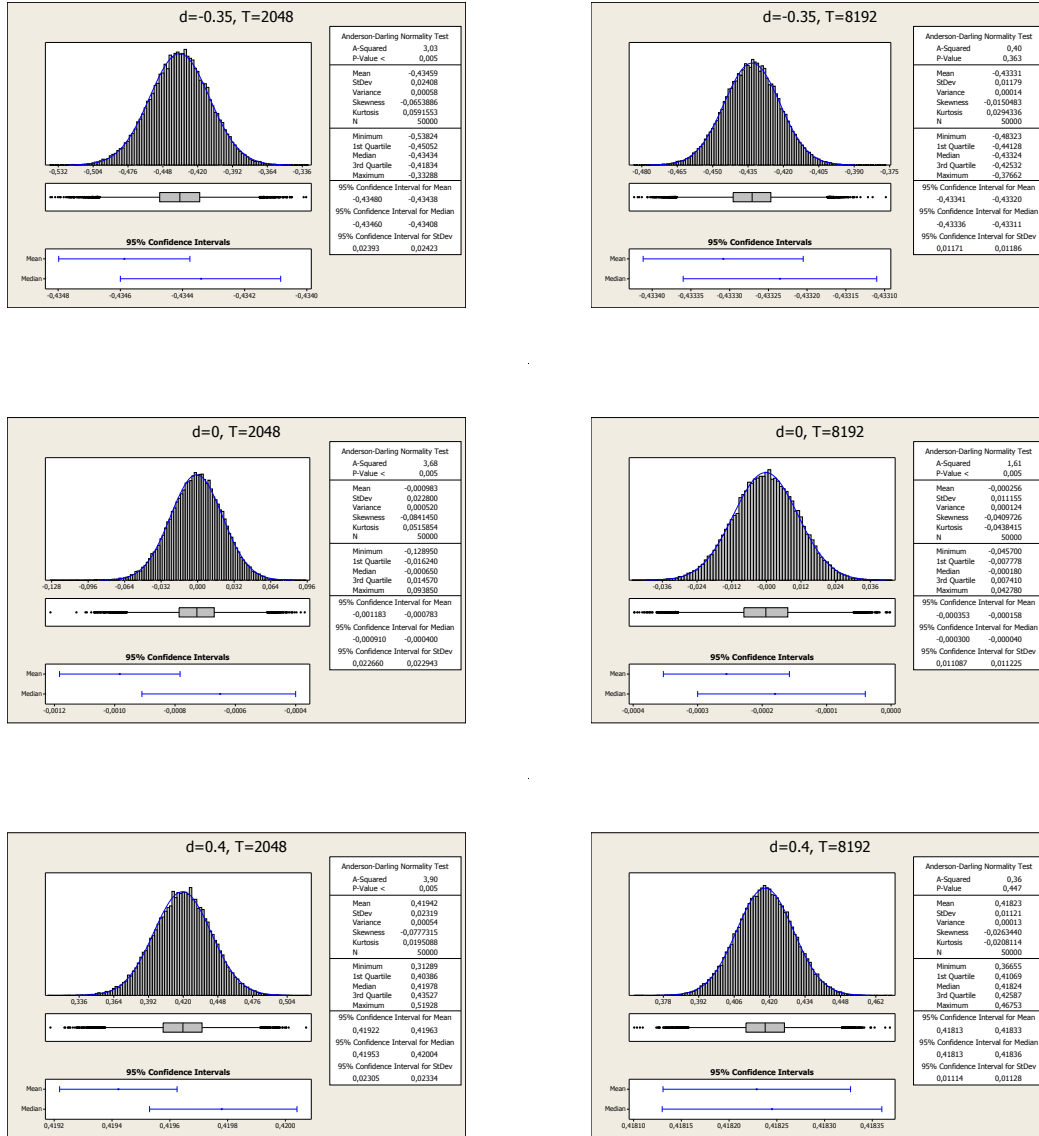


FIGURA 2.1. Distribuzione delle stime \hat{d} di processi caratterizzati da $d = -0.35, d = 0$ e $d = 0.4$ (nell'ordine dall'alto al basso) e $T = 2048$ (colonna di sinistra) e $T = 8192$ (colonna di destra).

3. Considerazioni teoriche

In questa sezione vengono illustrate alcune considerazioni riguardo al mal funzionamento dello stimatore nel caso di processi non stazionari. Prendendo come punto di partenza la letteratura sull'esponente di Hurst, si può affermare che un

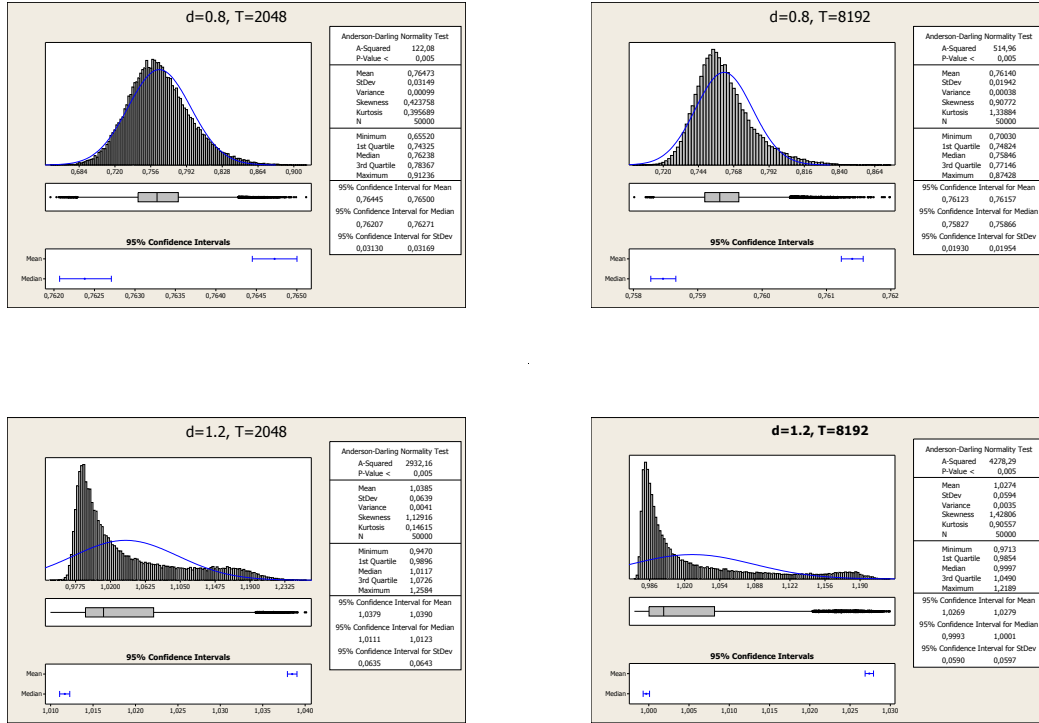


FIGURA 2.2. Distribuzione delle stime \hat{d} di processi caratterizzati da $d = 0.8$ e $d = 1.2$ (nell'ordine dall'alto al basso) e $T = 2048$ (colonna di sinistra) e $T = 8192$ (colonna di destra).

processo caratterizzato dal parametro H , una volta integrato vede la propria deviazione standard crescere secondo la legge σT^H , dove T rappresenta la lunghezza della serie storica. In figura (2.3) viene rappresentata la distribuzione della cumulata di un processo caratterizzato da $\sigma = 1$, $T = 4$ e $H = 0.5$ e 0.9 , dopo 50000 repliche.

Le deviazioni standard attese sono:

- $4^{0.5} = 2$ per quanto riguarda il processo con $H = 0.5$ (deviazione standard ottenuta pari a 2.00138);
- $4^{0.9} = 3.4822$ per quanto riguarda il processo con $H = 0.9$ (deviazione standard ottenuta pari a 3.4988).

In entrambi i casi non si può rifiutare l'ipotesi che il valore ottenuto e quello teorico non siano diversi. Provando a fare la stessa cosa con un processo non stazionario, il risultato cambia: in figura (2.4) viene rappresentata la distribuzione della cumulata di un processo caratterizzato da $\sigma = 1$, $T = 4$ e $H = 1.5$.

Se fosse valida la stessa legge, la deviazione standard sarebbe stata $4^{1.5} = 8$, ma quella ottenuta risulta pari a 5.4646. Entrando nel dettaglio, il processo in questione con $H = 1.5$ è un random walk (in quanto pari ad un processo con $H = 0.5$

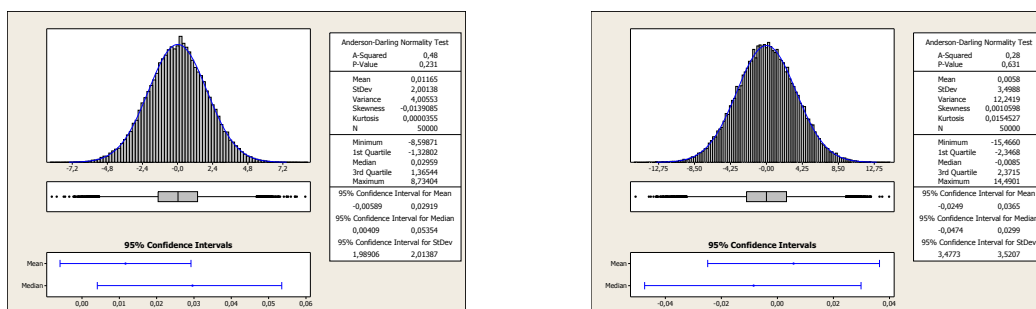


FIGURA 2.3. Distribuzione della cumulata di un processo caratterizzato da $H = 0.5$ (sinistra) e $H = 0.9$ (destra) e deviazione standard delle innovazioni unitaria dopo 4 passi.

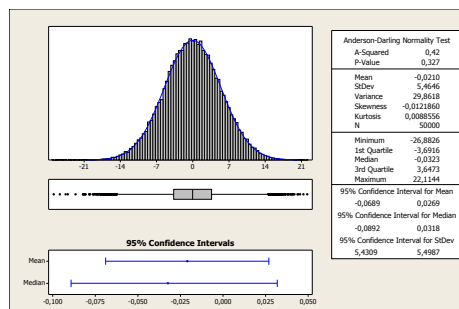


FIGURA 2.4. Distribuzione della cumulata di un processo caratterizzato da $H = 1.5$ e deviazione standard delle innovazioni unitaria dopo 4 passi.

cumulato). L'operazione di cumulare questo processo è quindi pari all'operazione di cumulare due volte un processo white noise.

Consideriamo, per comodità per soli 2 passi, il processo random walk x_t

$$\begin{aligned} x_1 &= \varepsilon_1, \\ x_2 &= x_1 + \varepsilon_2, \\ &\vdots \end{aligned}$$

e la sua cumulata y_t

$$\begin{aligned} y_1 &= x_1, \\ y_2 &= y_1 + x_2, \\ &\vdots \end{aligned}$$

dove

$$\varepsilon_t \sim N(0, 1).$$

Calcoliamo ora la varianza di y_2 :

$$\begin{aligned} & \text{E}[(y_1 + x_2)(y_1 + x_2)] = \\ & = \text{E}[(x_1 + x_2)(x_1 + x_2)] = \\ & = \text{E}[(\varepsilon_1 + \varepsilon_1 + \varepsilon_2)(\varepsilon_1 + \varepsilon_1 + \varepsilon_2)] = \\ & = \text{E}[\varepsilon_1^2 + \varepsilon_1^2 + \varepsilon_1\varepsilon_2 + \varepsilon_1^2 + \varepsilon_1^2 + \varepsilon_1\varepsilon_2 + \varepsilon_1\varepsilon_2 + \varepsilon_1\varepsilon_2 + \varepsilon_2^2] = \\ & = \text{E}[4\varepsilon_1^2 + \varepsilon_2^2] = 5. \end{aligned}$$

Si nota come la prima innovazione compaia $T^2 = 4$ volte, mentre la seconda $(T - 1)^2 = 1$ volte. Generalizzando, si può notare come la varianza condizionale $\text{Var}(y|T)$ sia data da

$$\text{Var}(y|T) = \sigma^2 \sum_{i=0}^T (T - i)^2.$$

Sfruttando il risultato notevole della somma di quadrati, si può scrivere

$$\text{Var}(y|T) = \sigma^2 T(T + 1)(2T + 1)/6.$$

Ritornando al processo prima considerato (caratterizzato da $\sigma = 1$, $T = 4$ e $H = 1.5$), il valore teorico della deviazione standard dovrebbe essere ora pari a

$$(T(T + 1)(2T + 1)/6)^{0.5} = 5.4772,$$

in linea con quello trovato pari a 5.4646.

Provando ora ad ampliare queste considerazioni, si cerca di trovare un possibile collegamento tra il modello nel dominio delle frequenze definito in (2.1) e il dominio del tempo. Indichiamo con F_0 la trasformata discreta di Fourier del processo x_t white noise valutata alla frequenza zero: è stato visto come il modello definisca il comportamento di λ per $\lambda \rightarrow 0^+$, ma, considerando campioni finiti, F_0 esiste finita ed è pari alla cumulata c_T valutata al tempo T del processo x_t , a meno di un fattore moltiplicativo. Si può scrivere

$$F_0 = c_T (2\pi T)^{-1/2}.$$

Le cumulate c_T sono caratterizzate dalla seguente distribuzione condizionale:

$$\text{E}[c_T | c_0] = 0;$$

$$\text{V}[c_T | c_0] = \sigma_\varepsilon^2 T^{2H} = \text{E}[(c_T | c_0)^2].$$

Considerando che il periodogramma I_0 è dato da

$$F_0^2 = \frac{c_T^2}{2\pi T},$$

e sostituendo a c_T^2 il suo valore atteso sopra trovato (si consideri $\sigma_\varepsilon = 1$), si ottiene

$$(2.6) \quad I_0 = \frac{T^{2H}}{2\pi T} = \frac{T^{2H-1}}{2\pi}.$$

Riprendendo ora la formula (2.3), si può dire che

$$T = \frac{2\pi j}{\lambda_j}.$$

Sostituendo T in (2.6) con l'espressione appena trovata, e andando a considerare I_λ , $\lambda \rightarrow 0$ invece che $I(0)$, si ottiene

$$I_\lambda = \frac{(2\pi j)^{2H-1}}{\lambda_j} \frac{1}{2\pi},$$

che, isolando λ_j dagli altri termini, porta all'espressione

$$I_\lambda = G\lambda^{1-2H},$$

analoga alla formula (2.1).

Alcune simulazioni sono state effettuate per trovare conferma di ciò che è stato appena visto. In tabella (2.4) vengono mostrati i valori dei periodogrammi medi alla frequenza zero per vari valori del parametro H : il periodogramma medio è stato ottenuto facendo la media per ogni frequenza di 5000 repliche. I valori trovati vengono indicati con I_0 , mentre quelli teorici calcolati usando la (2.6) con I'_0 : affianco ai periodogrammi, vengono mostrate le deviazioni standard della cumulata in T del processo σ_0 e quelle teoriche σ'_0 . Si nota anche in questo caso come per

H	T	I_0	I'_0	σ_0	σ'_0
0.5	2048	0.16367	0.15915	45.894	45.2548
	8192	0.16115	0.15915	91.082	90.5097
0.9	2048	70.6656	70.9388	953.58	955.426
	8192	207.384	215.046	3266.6	3326.98
1.3	2048	11668.1	31619.0	12254	20171.1
	8192	115022	290565	76943	122294

TABELLA 2.4. Confronto tra periodogrammi sperimentali e teorici in $\lambda = 0$ e deviazioni standard sperimentali e teoriche della cumulata del processo in T .

processi non stazionari ($H = 1.3$) la teoria non sembra funzionare.

4. Analisi Numerica sul ruolo del parametro m

Dopo aver visto le incertezze riguardo alla scelta del parametro m , in questa sezione vengono illustrati i risultati di una nuova serie di simulazioni, questa volta caratterizzate dal fatto che vengono considerati tutti i valori del parametro m all'interno di una data banda. Più precisamente, fissato un valore m_{max} che corrisponde alla banda massima considerata, vengono considerati tutti i valori di m tali che $m_{min} < m < m_{max}$, dove la presenza di m_{min} si rende necessaria per evitare di considerare bande con risoluzione troppo bassa e dalla varianza troppo elevata. In altre parole la prima stima di H viene fatta considerando le frequenze che vanno da λ_1 a $\lambda_{m_{max}}$, la seconda con frequenze che vanno da λ_1 a $\lambda_{m_{max}-1}$ e così via, e l'ultima con frequenze che vanno da λ_1 a $\lambda_{m_{min}}$.

Per fare tutto ciò, le stime \hat{d} non sono state effettuate direttamente sui periodogrammi delle varie serie storiche generate, ma è stata fatta una unica stima

basata sul periodogramma medio di 50000 repliche¹. In figura (2.5) vengono rappresentati sulla sinistra i periodogrammi medi con il corrispettivo periodogramma teorico, mentre sulla destra le stime \hat{d} al variare di m : la lunghezza delle serie storiche è di $T = 8192$, m_{max} è pari a 512 e m_{min} è pari a 21. Possono essere fatte le seguenti considerazioni:

- i periodogrammi medi e teorici sembrano coincidere, a meno di una costante moltiplicativa. Si ricorda che in sede di stima del parametro H quel che incide è la forma del periodogramma, dato dal termine λ^{1-2H} , e non la costante moltiplicativa G ;
- l'incidenza del parametro m sulla distorsione non sembra essere casuale: in particolare nel caso di processi stazionari persistenti (figura (2.5c) e (2.5d)) la distorsione appare sempre positiva, e mentre in una prima fase al diminuire della larghezza di banda la distorsione diminuisce, in una seconda fase alla continua diminuzione della banda si ha un incremento della distorsione;
- nel caso invece di processi antipersistenti (figura (2.5a) e (2.5b)) si ha il fenomeno di distorsioni che cambiano di segno, ovvero per bande sufficientemente larghe si hanno distorsioni negative, e per bande più strette distorsioni positive. Questo fenomeno, a ben guardare, era già stato osservato, riprendendo la tabella (2.2), per $H = 1.3$ (processo antipersistente integrato), dove si riscontravano distorsioni positive per alcune bande e negative per altre.

Si potrebbe ora cercare di capire se l'aumento della distorsione per bande eccessivamente ristrette sia un fenomeno sempre valido o che tende a scomparire all'aumentare della lunghezza n delle osservazioni. Il primo caso sarebbe confermato se, all'aumentare di n , il punto in cui la distorsione è minima nel caso persistente o è nulla nel caso antipersistente cade sempre sulla stessa frequenza; si avrebbe così un valore di larghezza di banda ottimale da scegliere. Il secondo caso invece troverebbe conferma se all'aumentare di n la larghezza di banda ottimale tende a decrescere. In figura (2.6) vengono rappresentati gli stessi casi appena visti con sovrapposti i grafici caratterizzati da $n = 16384$ e $n = 32768$. Osservando i grafici, sembrerebbe essere confermato il secondo caso, e a sostegno di ciò vengono fatte le seguenti considerazioni:

- osservando i casi antipersistenti (figura (2.6a) e (2.6b)), H in entrambi i casi viene stimato correttamente (distorsione nulla) per valori λ^* sempre più piccoli all'aumentare di n . Considerando inoltre bande la cui λ_{max} considerata sia minore di λ^* , è sempre vero che la distorsione diminuisce all'aumentare di n . Non si può dire altrettanto per λ_{max} maggiori di λ^* , anche se, osservando il grafico per $H = 0.35$, la tendenza è che per n ancora maggiori si potrebbe avere un miglioramento della distorsione anche in tale direzione;
- osservando i casi persistenti (figura (2.6c) e (2.6d)), la tendenza sembra essere anche in questa circostanza quella per cui all'aumentare di n la stima \hat{H} migliore si ha per λ^* sempre più piccole; all'aumentare di n

¹ciò è dovuto a motivi computazionali. Supponiamo di avere $n = 8192$, $m_{max} = 512$ e $m_{min} = 21$: per ogni serie storica generata bisognerebbe stimare $512-21=491$ volte H , il che andrebbe poi moltiplicato per 50000. Così facendo invece le stime sono solamente 491, anche se probabilmente i risultati non sono coincidenti.

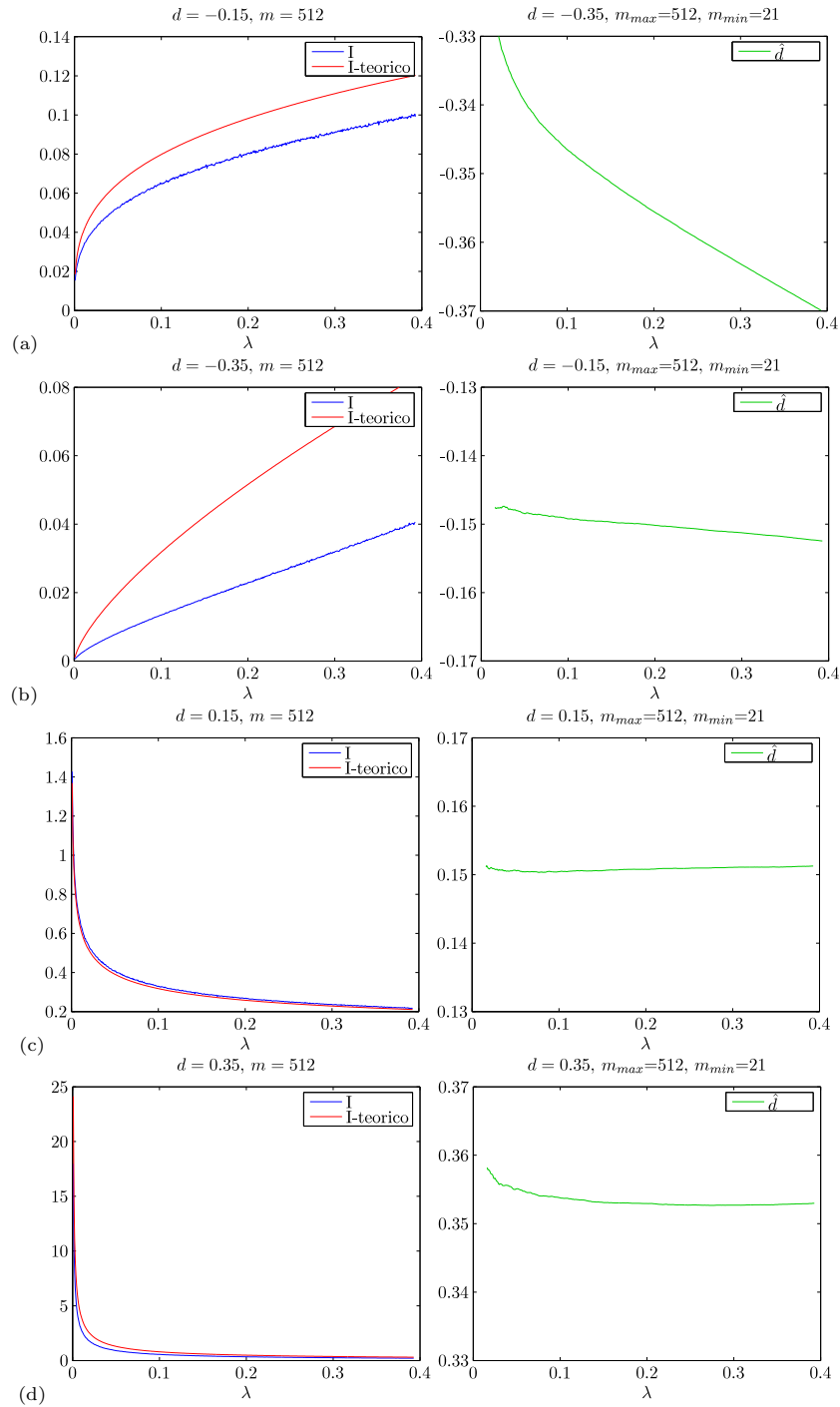


FIGURA 2.5. (a) periodogrammi medio (rosso) e teorico (blu) sulla sinistra per $d = -0.35$, e stime di d al variare di m sulla destra; (b) periodogrammi e stime per $d = -0.15$; (c) periodogrammi e stime per $d = 0.15$; (d) periodogrammi e stime per $d = 0.35$.

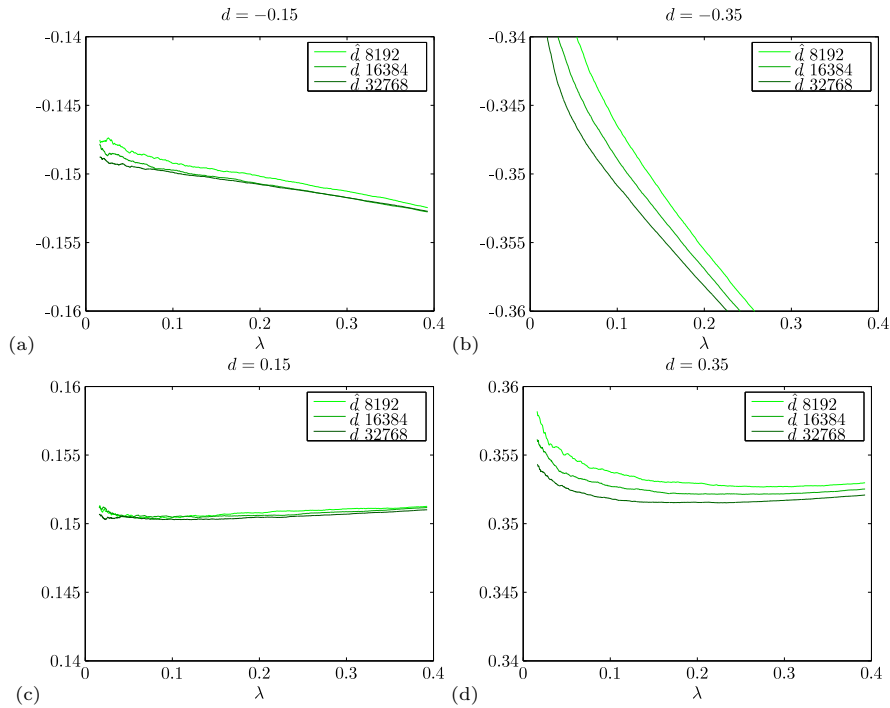


FIGURA 2.6. (a) stime \hat{d} al variare della banda e di n (diverse scale di verde) per $d = -0.35$; (b) stime \hat{d} per $d = -0.15$; (c) stime \hat{d} per $d = 0.15$; (d) stime \hat{d} per $d = 0.35$.

inoltre la distorsione sembra diminuire lungo tutto la banda considerata.

Queste considerazioni appaiono molto più evidenti per $H = 0.85$ rispetto a $H = 0.65$, dove c'è maggiore volatilità nelle stime.

Si può inoltre ora rispondere ad uno dei quesiti posti nelle sezioni precedenti, ovvero se la distorsione aumenta o diminuisce all'allontanarsi di H dal valore centrale 0.5: osservando i grafici, sembrerebbe che, fissata una lunghezza n ed una banda, la distorsione compaia in maniera maggiore per i valori di H più periferici.

Processi non-stazionari

1. Elementi teorici sulle finestre

In questa sezione si descrivono rapidamente alcuni elementi teorici riguardanti l'utilizzo delle finestre in analisi spettrale: il testo preso come riferimento è Priestley [4]. Definiamo il periodogramma I_λ relativo alla generica serie x_t con $t = 1, 2, \dots, T$ a cui è già stata sottratta la media μ come

$$(3.1) \quad I_\lambda = \frac{2}{T} \left| \sum_{t=1}^T x_t e^{i\lambda t} \right|^2,$$

con $\lambda = 2\pi j/T$, $j = 0, 1, \dots, [T/2]$. Tale periodogramma può essere riscritto come

$$I_\lambda = \sum_{s=-(T-1)}^{(T-1)} \hat{R}(s) \cos s\lambda,$$

dove

$$\hat{R}(s) = \frac{1}{T} \sum_{t=1}^{T-|s|} x_t x_{t+|s|},$$

con $s = 0, \pm 1, \dots, \pm(T-1)$.

Tale uguaglianza può essere dimostrata nella seguente maniera: a partire dalla definizione 3.1 segue che

$$\begin{aligned} I_\lambda &= \frac{2}{T} \left[\sum_{t=1}^T x_t e^{i\lambda t} \right] \left[\sum_{r=1}^T x_r e^{-i\lambda r} \right] = \\ &= \frac{2}{T} \left[\sum_{t=1}^T \sum_{r=1}^T x_t x_r \cos(t-r)\lambda \right]. \end{aligned}$$

La doppia sommatoria su t e r può essere portata ad una sommatoria su $s = (t-r)$ definita da $-(T-1)$ a $(T-1)$ e una su t da 1 a $T-|s|$:

$$\begin{aligned} I(\lambda) &= 2 \sum_{s=-(T-1)}^{T-1} \left\{ \frac{1}{T} \sum_{t=1}^{T-|s|} x_t x_{t+|s|} \right\} \cos s\lambda \\ &= 2 \sum_{s=-(T-1)}^{T-1} \hat{R}(s) \cos s\lambda \end{aligned}$$

$$(3.2) \quad = 2 \left\{ \hat{R}(0) + 2 \sum_{s=1}^{T-1} \hat{R}(s) \cos s\lambda \right\}.$$

La funzione di autocovarianza $\hat{R}(s)$ sopra descritta è uno stimatore distorto (ma asintoticamente non distorto) di $R(s)$. Lo stimatore non distorto \hat{R}^* è definito come

$$\hat{R}^* = \frac{1}{T-|r|} \sum_{t=1}^{T-|r|} x_t x_{t+|r|},$$

con $r = 0, \pm 1, \pm 2, \dots, \pm(T-1)$. Tale stimatore risulta tuttavia distorto se si considerano gli effetti di non conoscere la media del processo e aver stimato μ con \bar{x} : si può infatti scrivere (in queste righe x_t non è più un processo a media zero)

$$\begin{aligned} \hat{R}^*(r) &= \frac{1}{T-|r|} \sum_{t=1}^{T-|r|} [\{x_t - \mu\} \{x_{t+|r|} - \mu\}] \\ &= \frac{1}{T-|r|} \sum_{t=1}^{T-|r|} [\{(x_t - \bar{x}) + (\bar{x} - \mu)\} \{(x_{t+|r|} - \bar{x}) + (\bar{x} - \mu)\}] \\ &= \frac{1}{T-|r|} \left\{ \sum_{t=1}^{T-|r|} [\{x_t - \bar{x}\} \{x_{t+|r|} - \bar{x}\}] + (T-|r|)(\bar{x} - \mu)^2 \right\}. \end{aligned}$$

Segue quindi che

$$E[\hat{R}^*(r)] = R(r) - E[(\bar{x} - \mu)^2] \sim R(r) - (2\pi\sigma^2/T)f(0),$$

dato che, per T grande, vale l'approssimazione

$$\text{var}(\bar{x}) = (2\pi\sigma^2/T)f(0),$$

dove $f(0)$ indica lo spettro del processo valutato alla frequenza 0. Si nota quindi come la distorsione di $E[\hat{R}^*(r)]$ sia di ordine $O(1/T)$. Una stima alternativa della funzione di autocovarianza, proposta da alcuni autori come ad esempio Parzen e Schaerf, è data da

$$\hat{R}(r) = \frac{1}{T} \sum_{t=1}^{T-|r|} (x_t - \bar{x})(x_{t+|r|} - \bar{x}).$$

Tale stima è distorta in quanto, trascurando gli effetti della stima di μ , si ha

$$E[\hat{R}(r)] = \left(1 - \frac{|r|}{T}\right) R(r),$$

mentre considerando gli effetti della stima di μ si ha

$$[E[\hat{R}(r)]] = R(r) - \frac{|r|}{T} R(r) - \frac{2\pi\sigma^2}{T^2} f(0).$$

Tuttavia $\hat{R}(r)$ è la stima che solitamente viene usata per i seguenti motivi: innanzitutto se si rappresenta l'entità della distorsione (ignorando la stima della media) come

$$E[\hat{R}(r) - R(r)] = \frac{|r|}{T} R(r)$$

si nota come essa dipenda sia da $|r|$ (e dal suo rapporto con T) che da $R(r)$. La distorsione maggiore si avrebbe all'avvicinarsi di $|r|$ a $T - 1$, ma quando ciò si verifica se T è grande in relazione al tasso di decadimento della funzione di autocovarianza accade che $R(r)$ è molto piccola, e di conseguenza si può pensare che l'entità della distorsione rimanga piccola per tutti i valori di r . In secondo luogo si può dimostrare che $\text{var}(\hat{R}(r)) = O(1/T)$ per tutti i valori di r , mentre $\text{var}(\hat{R}^*(r)) = O(1/(T - |r|))$. All'avvicinarsi di $|r|$ a $(T - 1)$ la varianza dello stimatore non distorto cresce considerevolmente, determinando una coda nella funzione $\hat{R}^*(r)$ abbastanza erratica.

Tutte queste considerazioni aiutano a capire come mai il periodogramma $I(\lambda)$ è uno stimatore inconsistente dello spettro $f(\lambda)$. Ciò potrebbe a prima vista sembrare sorprendente, in quanto come visto $\hat{R}(r)$ è uno stimatore consistente di $R(r)$ caratterizzato da $\text{var}(\hat{R}(r)) = O(1/T)$: tuttavia nella computazione del periodogramma vengono considerate T funzioni di autocovarianza, come si può vedere in 3.2. La varianza di $I(\lambda)$ non diminuisce quindi al crescere di T ma rimane di ordine $O(1)$. Un modo abbastanza intuitivo per migliorare la stima dello spettro è il seguente: immaginando di avere un processo tale che $R(r) \rightarrow 0$ per $|r| \rightarrow \infty$, si può pensare che omettendo dal calcolo le autocovarianze più lontane la stima dello spettro non ne risenta eccessivamente in termini di distorsione ma migliori in termini di varianza (in generale eliminare autocovarianze "a caso" riduce sempre la varianza ma potrebbe causare delle distorsioni). Una stima dello spettro potrebbe quindi essere data da

$$\hat{f}(\lambda) = \frac{1}{2\pi} \sum_{s=-M}^M \hat{R}(s) \cos s\lambda$$

dove il parametro $M < (T - 1)$ è il punto di troncamento e $\hat{f}(\lambda)$ è il periodogramma troncato. Intuitivamente, dato che sono presenti $M + 1$ autocovarianze invece che T , la varianza di $\hat{f}(\lambda)$ sarà $(M/T) \cdot \text{var}I(\lambda)$ ovvero di ordine $O(M/T)$, e considerando che tale stimatore è non distorto in quanto

$$(3.3) \quad E[\hat{f}(\lambda)] = \frac{1}{2\pi} \sum_{s=-M}^M \left(1 - \frac{|s|}{T}\right) R(s) \cos s\lambda \rightarrow h(\lambda) \quad \text{con } M \rightarrow \infty$$

si può concludere che se M e T tendono entrambe ad infinito ma M molto più lentamente (in maniera tale che $(M/T) \rightarrow 0$ con $T \rightarrow \infty$) allora lo stimatore $\hat{f}(\lambda)$ è non distorto e consistente. Si può ora vedere la stima $\hat{h}(\lambda)$ come una forma speciale della forma più generale data da

$$(3.4) \quad \hat{g}(\lambda) = \frac{1}{2\pi} \sum_{s=-(T-1)}^{(T-1)} h(s) \hat{R}(s) \cos s\lambda,$$

dalla quale si può ricavare la 3.3 ponendo

$$(3.5) \quad h(s) = \begin{cases} 1, & \text{se } |s| < M \\ 0, & \text{se } |s| > M. \end{cases}$$

Dato che non c'è alcun motivo per cui la funzione $h(s)$ debba essere discontinua, si può pensare ad una vasta gamma di funzioni che decrescano gradualmente verso zero e che, se tale decrescita è appropriata, portino a stime consistenti di $f(h)$. La

formula 3.4 può essere riscritta (dopo aver notato che altro non è che la trasformata di Fourier del prodotto tra $h(s)$ e $R(s)$, e quest'ultima a sua volta la trasformata di $I(\lambda)$) come convoluzione della trasformata di $h(s)$ e $I(\lambda)$, ovvero come

$$\hat{g}(\lambda) = \int_{-\pi}^{\pi} I(\theta)W(\lambda - \theta) d\theta, \quad \text{dove} \quad W(\theta) = \frac{1}{2\pi} \sum_{s=-(T-1)}^{(T-1)} h(s)e^{-is\theta}.$$

Dunque $\hat{g}(\lambda)$ può essere visto come un integrale pesato del periodogramma $I(\theta)$, dove i pesi sono rappresentati dalla funzione $W(\theta)$. Le sequenze $h(s)$ sono solitamente costruite in maniera tale che i pesi siano maggiormente concentrati per $\theta = 0$, e i vari termini $\hat{g}(\lambda)$ altro non sono che una media pesata di $I(\theta)$ dove i pesi maggiori si hanno nell'intorno di $\theta = \lambda$. In altre parole quello che si sta facendo è semplicemente rendere il periodogramma erratico $I(\lambda)$ un po' più regolare median-done i valori tra frequenze contigue. Dal punto di vista notazionale, la sequenza $h(s)$ è definita in letteratura con il nome di 'lag window', mentre la funzione dei pesi $W(\theta)$ è detta 'spectral window'.

Di seguito vengono elencate alcune finestre tra le più conosciute ed utilizzate, con le corrispettive sequenze $h(s)$ e $W(\theta)$.

1.1. Periodogramma troncato. La funzione $h(s)$ è già stata mostrata in 3.5. La funzione dei pesi $W(\theta)$ è data da

$$W(\theta) = \frac{1}{2\pi} \sum_{s=-M}^M \cos s\theta = \frac{1}{2\pi} \left\{ \frac{\sin[(M + \frac{1}{2})\theta]}{\sin(\theta/2)} \right\} = D_M(\theta),$$

dove $D_M(\theta)$ è noto come 'Dirichlet kernel'. Di seguito (figura 3.1) viene rappresentato il Dirichlet kernel fissando il parametro M pari a 200. La magnitudine del

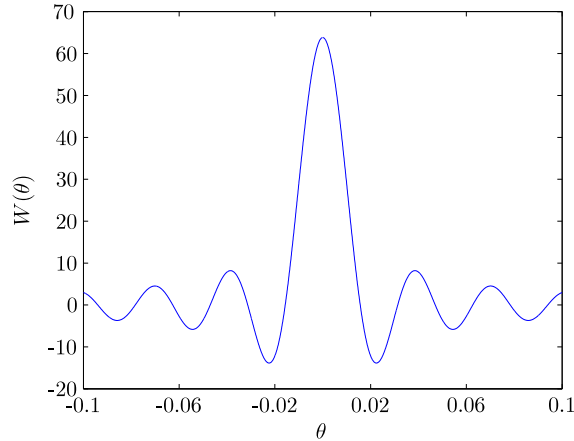


FIGURA 3.1. Dirichlet kernel rappresentato per $M = 200$.

picco più elevato (per $\theta \simeq 0$) è pari a $(2M+1)/2\pi$ (che in questo caso è pari a 63.82) mentre gli zeri si hanno per $\theta = \pm(2\pi)/(2M+1), \pm(4\pi)/(2M+1), \dots$. Un difetto di questo kernel è che in alcuni punti assume valori negativi e dunque non esclude la possibilità di avere valori di $\hat{g}(\lambda)$ negativi, il che ovviamente non è desiderabile in quanto lo spettro è definito non-negativo.

1.2. Finestra di Bartlett. In questo caso la ‘lag window’ $h(s)$ ha la forma

$$h(s) = \begin{cases} 1 - |s|/M, & \text{se } |s| \leq M, \\ 0, & \text{se } |s| > M. \end{cases}$$

mostrata in figura 3.2. La funzione dei pesi $W(\theta)$ è data da

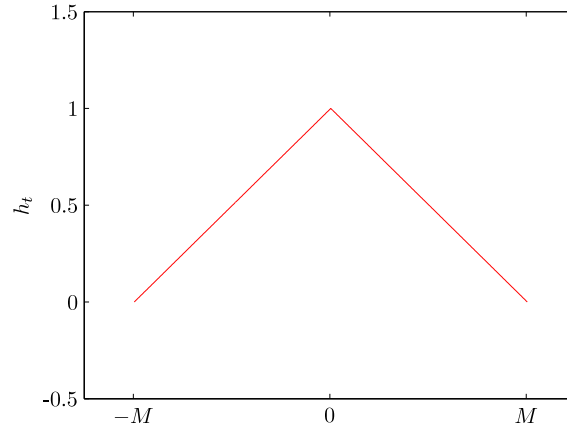


FIGURA 3.2. Rappresentazione della lag window h_t di Bartlett.

$$W(\theta) = \frac{1}{2\pi} \sum_{s=-M}^M \left(1 - \frac{|s|}{M}\right) \cos s\theta = \frac{1}{2\pi M} \left\{ \frac{\sin(M\theta/2)^2}{\sin(\theta/2)} \right\} = F_M(\theta),$$

dove $F_M(\theta)$ è noto come ‘Fejer kernel’. Di seguito (figura 3.3) viene rappresentato il Fejer kernel fissando il parametro M pari a 200. La magnitudine del picco più

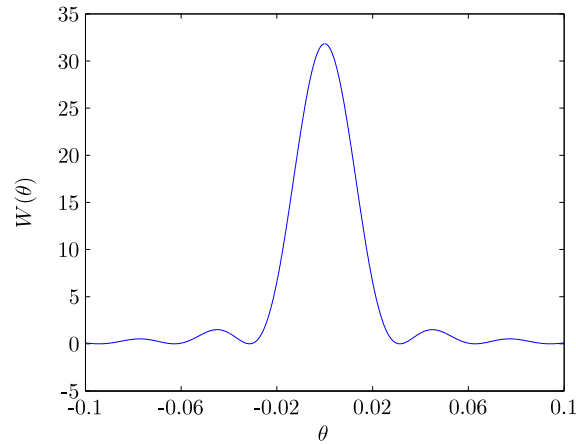


FIGURA 3.3. Fejer kernel rappresentato per $M = 200$.

elevato (per $\theta \simeq 0$) è pari a $(M)/2\pi$ (che in questo caso è pari a 31.83) mentre gli zeri si hanno per $\theta = \pm(2\pi)/M, \pm(4\pi)/M, \dots$. Si può notare come in questo kernel tutti i pesi siano positivi.

1.3. Finestra ‘full cosine bell’. Questa finestra è caratterizzata dalla ‘lag window’

$$h_t = \frac{1}{2} \left(1 + \cos \left[\frac{2\pi t}{2M+1} \right] \right),$$

rappresentata in figura 3.4.

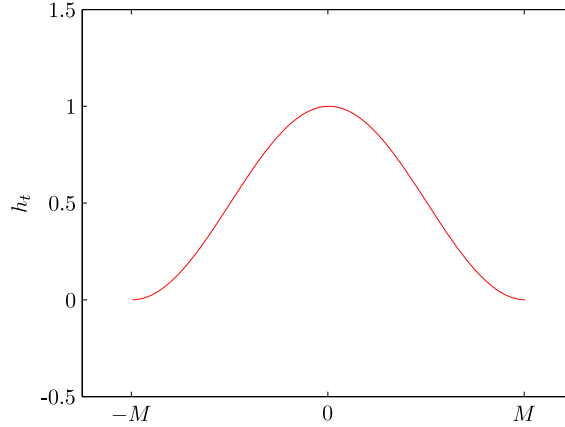


FIGURA 3.4. Rappresentazione della lag window h_t ‘full cosine bell’.

1.4. Finestra di Parzen. Parzen nel 1961 ha proposto la seguente ‘lag window’ h_t :

$$h(s) = \begin{cases} 1 - 6(s/M)^2 + 6(|s|/M)^3, & \text{se } |s| \leq M/2, \\ 2(1 - |s|/M)^3, & \text{se } M/2 \leq |s| \leq M, \\ 0, & \text{se } |s| > M, \end{cases}$$

rappresentata in figura 3.5. La funzione dei pesi $W(\theta)$ è data da

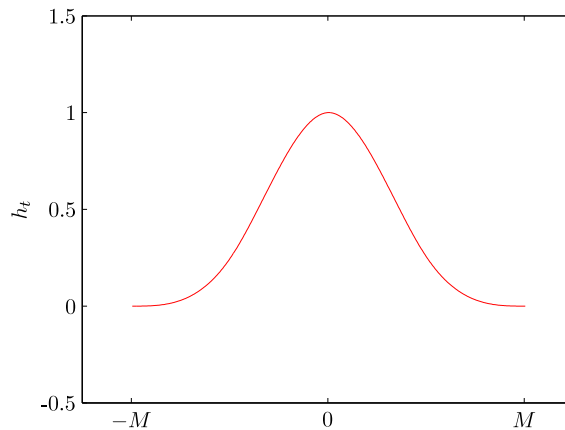


FIGURA 3.5. Rappresentazione della lag window h_t di Parzen.

$$W(\theta) = \frac{3}{8\pi M^3} \left(\frac{\sin(M\theta/4)}{\frac{1}{2}\sin(\theta/2)} \right)^4 \left\{ 1 - \frac{2}{3} \sin^2(\theta/2) \right\},$$

rappresentata in figura 3.6.

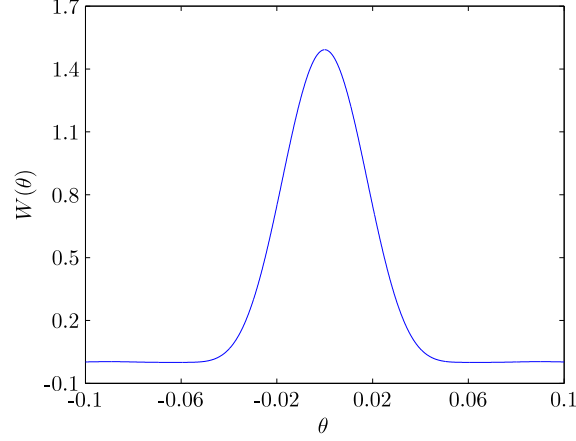


FIGURA 3.6. Funzione dei pesi $W(\theta)$ al variare di θ , per $M = 200$.

2. Serie storiche non-stazionarie e data tapering

In questa sezione si descrive la tecnica ‘data tapering’ che diversi autori hanno proposto come soluzione al problema di dover stimare il parametro d relativo a serie storiche non-stazionarie ($d > 0.5$). L’articolo di riferimento è quello di Velasco [5], dove si studia il problema nello specifico dello stimatore Gaussiano di Robinson.

Innanzitutto Velasco dimostra che per certi valori di d non stazionari è possibile utilizzare la stima tradizionale (in Robinson definita nell’intervallo $-0.5 < d < 0.5$); in particolare per i valori non stazionari $\frac{1}{2} < d < \frac{3}{4}$ valgono le stesse proprietà di consistenza e asintotica normalità che caratterizzano i processi stazionari, ovvero

$$m^{1/2}(\hat{d} - d_0) \rightarrow_d N\left(0, \frac{1}{4}\right),$$

dove \hat{d} è la stima e d_0 il vero valore. In questo caso, definendo con x_t , $t = 1, 2, \dots, T$, il processo non stazionario e con u_t il relativo processo stazionario dato da $u_t = \Delta x_t$, si può dire che lo spettro $f_u(\lambda)$ ha forma

$$f_u(\lambda) \sim G\lambda^{-2(d-1)} \quad \text{per } \lambda \rightarrow 0^+,$$

dove G è una costante positiva.

Per i valori non stazionari $\frac{3}{4} < d < 1$ si ha invece solo consistenza, ovvero

$$\hat{d} \rightarrow_p d.$$

Per $d > 1$ lo stimatore Gaussiano non fornisce più una stima consistente, ed è necessario ricorrere alla tecnica del ‘data tapering’ per una stima corretta di d . Tale tecnica può essere usata anche nella regione $d < 1$ in quanto può ridurre la distorsione della stima: al crescere dei valori di d la convergenza della stima diventa infatti sempre più lenta. Si possono generalizzare le considerazioni sullo spettro

sopra riportate dicendo che il processo stazionario $u_t^{(s)} = \Delta^s x_t$ ha uno spettro con forma

$$f_{u^{(s)}}(\lambda) \sim G\lambda^{-2(d-s)} \quad \text{per } \lambda \rightarrow 0^+,$$

dove $s = \lfloor d + 1/2 \rfloor$. Queste serie non stazionarie presentano in sede di stima di d il problema che il continuo accumulo di informazioni crea dei disturbi nel calcolo della trasformata di Fourier, in particolare dei trend deterministici; questi disturbi possono essere eliminati utilizzando un opportuno ‘data taper’, ovvero una della ‘lag window’ viste nella sezione precedente. Da un punto di vista pratico, la ‘lag window’ h_t si applica moltiplicandola punto a punto direttamente per la serie storica x_t . Il vettore h_t viene in questo caso scelto di lunghezza pari a T (non c’è un punto di troncamento M), simmetrico nell’intorno di $T/2$ con valore massimo pari ad 1. La trasformata di Fourier tapered $w^T(\lambda_j)$ della serie $h_t x_t$ è ora data da

$$w^T(\lambda_j) = \frac{1}{(2\pi \sum_{t=1}^T h_t^2)^{1/2}} \sum_{t=1}^T h_t x_t e^{i\lambda_j t}.$$

La proprietà fondamentale che deve valere affinché si possa fare inferenza su d per serie non stazionarie di ordine s è che

$$(3.6) \quad \sum_{t=1}^T h_t (1 + t + t^2 + \dots + t^s) e^{i\lambda_j t} = 0.$$

Se questa proprietà viene soddisfatta i trend deterministici fino all’ordine s vengono rimossi dal calcolo della trasformata senza bisogno di stimarli direttamente. A seconda del taper utilizzato si rende necessario considerare solo alcune frequenze del periodogramma: indicando con p l’ordine del taper in grado di fare inferenza su processi non stazionari di ordine $s \leq p - 1$, tale taper soddisfa la proprietà 3.6 utilizzando le frequenze λ_{jp} , $0 < j < N$. Nell’articolo di Velasco viene considerata una classe di tapers proposta da Zhurbenko e Kolmogorov in grado di produrre diversi vettori h_t a seconda del parametro p scelto: tali tapers sono molto simili a quelli visti nella sezione precedente, quindi nel seguito considereremo questi ultimi tapers che sono maggiormente conosciuti.

Ora proviamo a classificare i diversi tapers a seconda del loro ordine p : i tapers sono gli stessi visti nella sezione precedente, con l’accorgimento che non viene scelto un punto di troncamento M . Operativamente abbiamo fissato una lunghezza T della serie storica ($T = 1024$), generato i vettori h_t , eseguito la sommatoria descritta in 3.6 per le prime 20 frequenze e verificato per quali di queste tale sommatoria risultasse pari a 0:

- il primo taper (periodogramma troncato) non è propriamente un taper ma è la trasformata di Fourier pura ($h_t = 1$ per ogni t). La sommatoria per $s = 0$ da 0 per tutte le frequenze; per valori di s superiori non converge a 0, e quindi si può dire che la trasformata è un taper di parametro $p = 1$;
- il secondo taper è la finestra triangolare di Bartlett: per $s = 0$ e per $s = 1$ la sommatoria è pari a 0 se si considerano le frequenze $\lambda_2, \lambda_4, \dots$. Per valori di s superiori non c’è convergenza: pertanto questo taper è caratterizzato da parametro $p = 2$;
- il taper ‘full cosine bell’ è un po’ particolare: guardando alla sommatoria sarebbe di ordine $p = 1$, con l’accorgimento di togliere la prima frequenza e le ultime due ($2 \leq j \leq T - 2$) per le quali non si ha convergenza. Per altre caratteristiche può però essere considerato di ordine $p = 3$;

- infine analizziamo il taper proposto da Parzen: in questo caso la sommatoria è pari a 0 per $s = 0, 1, 2, 3$, e questo solo per le frequenze $\lambda_4, \lambda_8, \lambda_{12}, \dots$, mentre per $s = 4$ non si ha convergenza. Il taper può essere classificato di ordine $p = 4$.

In figura 3.7 viene mostrato un processo frazionario con $d = 0.2$, e come questa serie storica si deforma una volta moltiplicata per i vettori h_t . Viene poi mostrato il relativo periodogramma, prendendo solo le frequenze $\lambda_p, \lambda_{2p}, \dots, \lambda_m$.

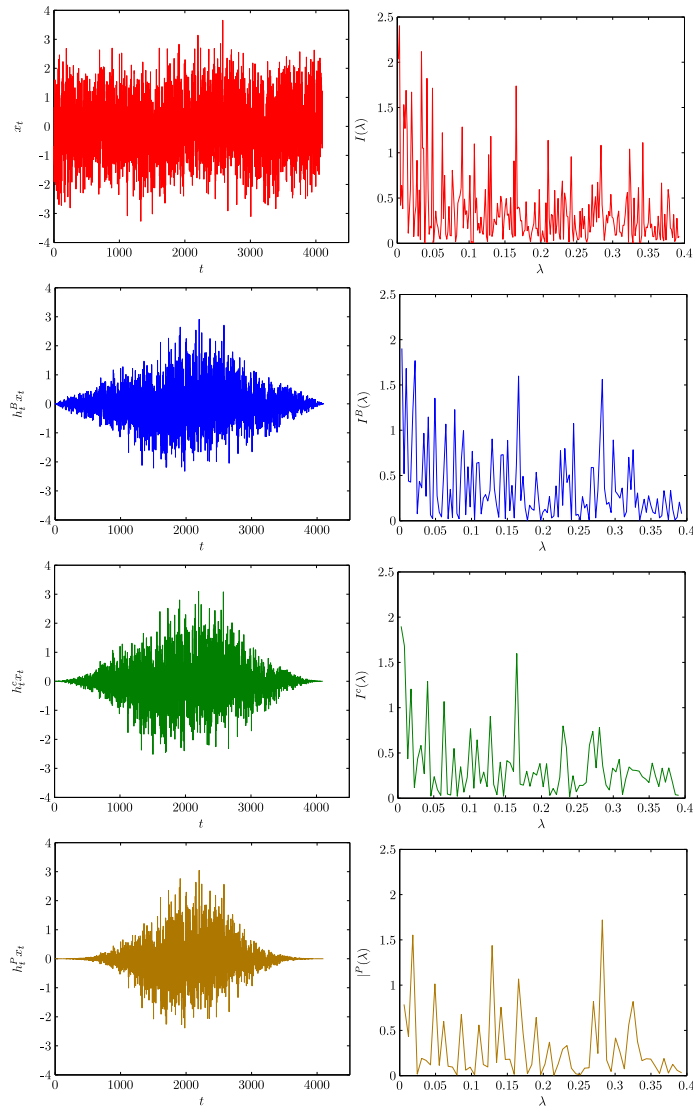


FIGURA 3.7. Sulla sinistra una serie storica caratterizzata da $d = 0.2$ viene moltiplicata per diversi vettori h_t . Sulla destra i corrispettivi periodogrammi (considerando soltanto le frequenze $\lambda_p, \lambda_{2p}, \dots$).

La stima di d , una volta scelto un taper, non si discosta di molto da quella vista per lo stimatore Gaussiano tradizionale: per prima cosa si definisce il nuovo periodogramma tapered $I_p^T(\lambda_j)$ come

$$I_p^T(\lambda_j) = |w_p^T \lambda_j|^2.$$

Il valore \hat{d}_p è come al solito dato da

$$\hat{d}_p = \arg \min R_p(d),$$

dove però questa volta la funzione $R_p(d)$ deve tener conto del fatto che non tutte le frequenze devono essere considerate: la nuova funzione diventa

$$R_p(d) = \log \hat{G}_p(d) - 2d \frac{p}{m} \sum_j^m \log \lambda_j, \quad \text{con} \quad \hat{G}_p(d) = \frac{p}{m} \sum_j^m \lambda_j^{2d} I_p^T(\lambda_j),$$

con $j = p, 2p, \dots, m$.

Come prima serie di simulazioni, si propone un'analisi preliminare per vedere quali taper siano effettivamente adeguati per analizzare le diverse non-stazionarietà. Dopo aver generato un processo stazionario frazionario, si è proceduto a cumulare tre volte tale processo ottenendo quattro serie (una stazionaria e tre non-stazionarie): per ognuna di queste serie sono state effettuate quattro stime di d utilizzando i quattro taper sopra esposti. Il tutto è stato ripetuto 1000 volte, con lo scopo finale di poter calcolare media e deviazione standard di ognuno dei sedici vettori di stime \hat{d} . In tabella 3.1 sono riportati tali valori, calcolati su serie storiche di parametro $T = 2^{12}$ e $m = 2^8$, e $d = 0.2$ per la serie stazionaria.

		$d = 0.2$			
		\hat{d}	\hat{d}^B	\hat{d}^c	\hat{d}^P
mean		0.1987	0.1915	0.1982	0.1996
std dev		0.0332	0.0484	0.0625	0.0688
		$d = 1.2$			
		\hat{d}	\hat{d}^B	\hat{d}^c	\hat{d}^P
mean		1.0656	1.1992	1.2026	1.2032
std dev		0.0666	0.0498	0.0628	0.0685
		$d = 2.2$			
		\hat{d}	\hat{d}^B	\hat{d}^c	\hat{d}^P
mean		1.0123	2.0802	2.2767	2.2157
std dev		0.0253	0.0821	0.0911	0.0689
		$d = 3.2$			
		\hat{d}	\hat{d}^B	\hat{d}^c	\hat{d}^P
mean		1.0034	2.0251	3.0329	3.2463
std dev		0.0214	0.0628	0.0463	0.0729

TABELLA 3.1. Media e deviazione standard di 100 stime \hat{d} effettuate variando d e il tipo di taper utilizzato. In grassetto i valori per i quali il relativo taper dovrebbe fornire stime consistenti.

Da questa analisi preliminare si nota come effettivamente utilizzando taper di ordine $p = 2, 3$ e 4 si possa stimare d di processi caratterizzati da d minore

rispettivamente di 2, 3 e 4, mentre per valori di d al di sopra di questa soglia la stima si appiattisce nell'intorno del valore massimo stimabile.

Per quanto riguarda la distribuzione asintotica di questi taper, Velasco afferma che le stime dei taper risultano consistenti e asintoticamente normali (ovviamente se il taper è opportuno per il tipo di processo), ovvero

$$\hat{d}_p \rightarrow_p d_0,$$

e

$$m^{1/2} (\hat{d} - d_0) \rightarrow_D N \left(0, \frac{1}{4} p \Phi \right),$$

dove

$$\Phi = \lim_{T \rightarrow \infty} \left(\sum_{t=1}^T h_t^2 \right)^{-2} \sum_{k=0, p, 2p, \dots}^{T-p} \left\{ \sum_{t=1}^T h_t^2 \cos(t\lambda_k) \right\}^2.$$

I valori di Φ relativi ai taper analizzati sono: $\Phi = 1.05$ per quanto riguarda la finestra di Bartlett, $\Phi = 1$ per la finestra 'full cosine bell' e $\Phi = 1.00086$ per la finestra di Parzen.

Ora proviamo ad effettuare simulazioni più approfondite per cogliere le distribuzioni asintotiche che caratterizzano questi stimatori e verificarne la loro corrispondenza con la teoria appena vista. Dato che nelle applicazioni finanziarie si ci aspetta tipicamente che le serie storiche abbiano parametro d nell'intorno di 1 (nel caso dei prezzi) o nell'intorno di 0 (nel caso dei rendimenti), abbiamo considerato due gruppi di simulazioni, uno caratterizzato da parametro $d = 0.1$ e $d = 1.1$ e uno da parametro $d = -0.1$ e $d = 0.9$. Per ognuno dei due gruppi sono state condotte le simulazioni nel seguente modo: per 10000 volte è stata generata una serie storica e stimato d utilizzando i quattro differenti taper, poi la serie è stata cumulata e di nuovo è stato stimato d con i differenti taper. Il tutto è stato fatto quattro volte, combinando tra di loro i valori $m = 2^{11}, 2^{12}$ e $T = 2^{14}, 2^{15}$. I risultati sono mostrati in tabella 3.2, dove vengono riportati media e deviazione standard delle stime, la deviazione standard teorica e il P value del test di normalità (di Kolgomorov Smirnov).

In figura 3.8 vengono rappresentate le distribuzioni relative alla prima serie di simulazioni ($m = 2^{11}$ e $T = 2^{14}$). La distribuzione chiaramente non normale è quella relativa alla stima \hat{d} senza taper del processo caratterizzato da $d = 1.1$.

		$d = 0.1$				$d = 1.1$			
		\hat{d}	\hat{d}^B	\hat{d}^c	\hat{d}^P	\hat{d}	\hat{d}^B	\hat{d}^c	\hat{d}^P
$m = 2^{11}, T = 2^{14}$	Mean	0.1021	0.1020	0.1024	0.1017	1.0375	1.0977	1.0974	1.0968
	StdDev	0.0112	0.0163	0.0196	0.0229	0.0325	0.0164	0.0196	0.0229
	σ teo	0.0110	0.0160	0.0191	0.0221	0.0110	0.0160	0.0191	0.0221
	Pvalue	0.9846	0.4867	0.4326	0.0193	0.0000	0.6203	0.2317	0.0290
$m = 2^{11}, T = 2^{15}$	Mean	0.1006	0.1005	0.1015	0.0999	1.0419	1.1005	1.1010	1.0995
	StdDev	0.0111	0.0162	0.0195	0.0227	0.0324	0.0163	0.0195	0.0227
	σ teo	0.0110	0.0160	0.0191	0.0221	0.0110	0.0160	0.0191	0.0221
	Pvalue	0.1060	0.8003	0.5811	0.6926	0.0000	0.9797	0.6235	0.4555
$m = 2^{12}, T = 2^{14}$	Mean	0.1070	0.1069	0.1067	0.1067	1.0200	1.0843	1.0836	1.0835
	StdDev	0.0078	0.0114	0.0135	0.0158	0.0335	0.0115	0.0136	0.0159
	σ teo	0.0078	0.0113	0.0135	0.0156	0.0078	0.0113	0.0135	0.0156
	Pvalue	0.8580	0.2506	0.8658	0.6777	0.0000	0.2180	0.6831	0.5845
$m = 2^{12}, T = 2^{15}$	Mean	0.1023	0.1022	0.1023	0.1020	1.0363	1.0973	1.0969	1.0967
	StdDev	0.0079	0.0114	0.0138	0.0159	0.0323	0.0114	0.0138	0.0159
	σ teo	0.0078	0.0113	0.0135	0.0156	0.0078	0.0113	0.0135	0.0156
	Pvalue	0.7332	0.3630	0.1483	0.5851	0.0000	0.3488	0.2529	0.5464
		$d = -0.1$				$d = 0.9$			
$m = 2^{11}, T = 2^{14}$	Mean	-0.1046	-0.1050	-0.1048	-0.1058	0.9057	0.8903	0.8900	0.8890
	StdDev	0.0112	0.0162	0.0198	0.0229	0.0198	0.0163	0.0198	0.0230
	σ teo	0.0110	0.0160	0.0191	0.0221	0.0110	0.0160	0.0191	0.0221
	Pvalue	0.7914	0.7910	0.2255	0.3150	0.0000	0.9061	0.2540	0.1346
$m = 2^{11}, T = 2^{15}$	Mean	-0.1017	-0.1021	-0.1013	-0.1025	0.9126	0.8975	0.8979	0.8969
	StdDev	0.0111	0.0163	0.0199	0.0230	0.0194	0.0163	0.0199	0.0230
	σ teo	0.0110	0.0160	0.0191	0.0221	0.0110	0.0160	0.0191	0.0221
	Pvalue	0.8065	0.5551	0.4898	0.7412	0.0000	0.4432	0.6598	0.5610
$m = 2^{12}, T = 2^{14}$	Mean	-0.1110	-0.1113	-0.1113	-0.1116	0.8816	0.8663	0.8658	0.8655
	StdDev	0.0079	0.0114	0.0139	0.0159	0.0180	0.0115	0.0139	0.0160
	σ teo	0.0078	0.0113	0.0135	0.0156	0.0078	0.0113	0.0135	0.0156
	Pvalue	0.4065	0.3672	0.3070	0.0548	0.0000	0.2307	0.3029	0.0438
$m = 2^{12}, T = 2^{15}$	Mean	-0.1045	-0.1047	-0.1047	-0.1049	0.9044	0.8902	0.8899	0.8896
	StdDev	0.0079	0.0114	0.0138	0.0158	0.0173	0.0114	0.0138	0.0158
	σ teo	0.0078	0.0113	0.0135	0.0156	0.0078	0.0113	0.0135	0.0156
	Pvalue	0.7566	0.4876	0.9214	0.9070	0.0000	0.5594	0.8657	0.8424

TABELLA 3.2. Stime \hat{d} relative a processi stazionari e non-stazionari, variando T , m e il taper utilizzato.

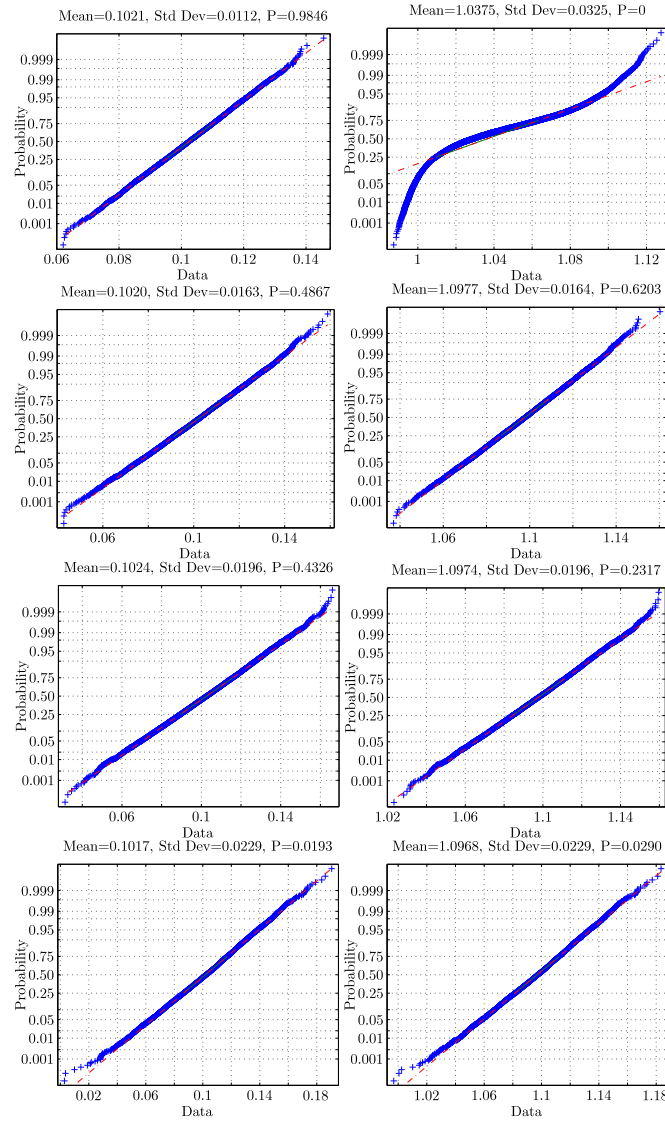


FIGURA 3.8. Distribuzione asintotica delle stime relative a processi stazionari e non-stazionari, variando il taper utilizzato.

Aspetti computazionali

1. Trasformata di Fourier veloce

La trasformata di Fourier veloce (FFT, dall'inglese Fast Fourier Transform), è un'algoritmo proposto per la prima volta da Cooley e Tukey [6] che permette di ridurre notevolmente i tempi di calcolo della trasformata di Fourier. Per prima cosa si definisce la trasformata di Fourier $w(j)$ come

$$(4.1) \quad w(j) = \sum_{t=0}^{T-1} x_t W^{jt}, \quad j = 0, 1, \dots, T-1$$

dove i termini x_t possono essere sia reali che complessi e dove W è la radice T -esima dell'unità,

$$(4.2) \quad W = e^{2\pi i/T}.$$

Con radice T -esima dell'unità si intende un qualunque numero reale o complesso che elevato alla T -esima potenza dia come risultato 1. Si può notare come W^T dia effettivamente 1 come risultato, questo perchè

$$(4.3) \quad W^T = (e^{2\pi i/T})^T = e^{2\pi i} = \cos(2\pi) + i \sin(2\pi) = 1.$$

Si può inoltre osservare come una qualunque potenza intera di una radice T -esima dell'unità dia sempre uno come risultato:

$$(4.4) \quad W^{Tk} = 1^k = 1, \quad k = 0, 1, 2, \dots$$

La formulazione in 4.1 non risulta sostanzialmente diversa da quella vista nel capitolo precedente:

$$(4.5) \quad w(\lambda) = (2\pi T)^{-1/2} \sum_{t=1}^T x_t e^{it\lambda}, \quad \lambda_j = \frac{2\pi j}{T}, \quad j = 1, 2, \dots, T.$$

Il termine $(2\pi T)^{-1/2}$ è infatti una normalizzazione usata per il calcolo successivo dello spettro, mentre il fatto che w dipenda da λ piuttosto che da j è conseguenza di non aver esplicitato λ . Ciò che invece interessa capire è se il fatto che le sommatorie e i vettori di elementi j siano definiti da 0 a $T-1$ in una formula e da 1 a T nell'altra generi due risultati differenti o meno. Nella figura 4.1 viene mostrato white noise y_t di lunghezza $T = 16$ (a) e le trasformate di Fourier corrispondenti calcolate con la formulazione in Robinson (in rosso) e con quella in Cooley-Tukey (in verde)(b). Le due serie appaiono diverse, tranne che per un valore (quello ottenuto ponendo $j=T$ in Robinson e $j=0$ in Cooley-Tukey, in quanto in entrambi i casi l'esponenziale da come risultato 1).

Tuttavia a queste serie diverse corrisponde il medesimo periodogramma $I(j)$, come si vede in figura (c), con l'unica differenza che a $I(T)$ in Robinson corrisponde $I(0)$ in Cooley-Tukey. Di seguito viene riportata una possibile illustrazione del

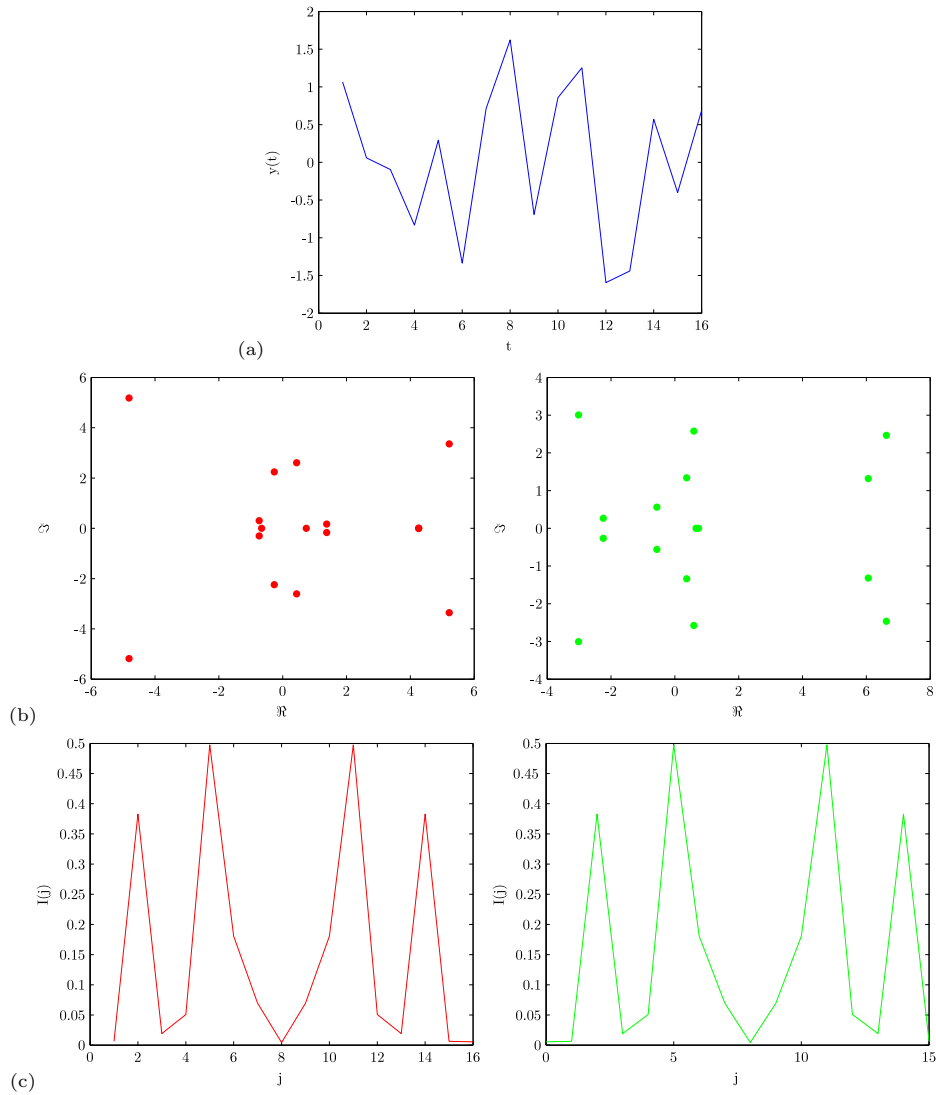


FIGURA 4.1. (a) Processo white noise y_t ; (b) trasformate di Fourier relative al processo y_t utilizzando la notazione in Robinson (sinistra) e quella in Cooley-Tukey (destra); (c) relativi periodogrammi utilizzando la notazione in Robinson (sinistra) e Cooley-Tukey (destra).

perchè si ottengano periodogrammi uguali a partire da scritte diverse. Scriviamo per esteso il periodogramma calcolato alla generica frequenza di indice \tilde{j} secondo la notazione in Robinson e ponendo $W = e^{2\pi i/T}$:

$$\begin{aligned} I(\tilde{j}) &= (x_1 W^{1\tilde{j}} + x_2 W^{2\tilde{j}} + \dots + x_T W^{T\tilde{j}})(x_1 W^{-1\tilde{j}} + x_2 W^{-2\tilde{j}} + \dots + x_T W^{-T\tilde{j}}) = \\ &= x_1^2 + x_1 x_2 W^{1\tilde{j}-2\tilde{j}} + \dots + x_1 x_T W^{1\tilde{j}-T\tilde{j}} + x_2^2 + \dots \end{aligned}$$

Scriviamo ora per esteso il periodogramma calcolato alla generica frequenza di indice \check{j} secondo la notazione in Cooley-Tukey:

$$\begin{aligned} I(\check{j}) &= (x_1 W^{0\check{j}} + x_2 W^{1\check{j}} + \dots + x_T W^{(T-1)\check{j}})(x_1 W^{-0\check{j}} + x_2 W^{-1\check{j}} + \dots + x_T W^{-(T-1)\check{j}}) = \\ &= x_1^2 + x_1 x_2 W^{0\check{j}-1\check{j}} + \dots + x_1 x_T W^{0\check{j}-(T-1)\check{j}} + x_2^2 + \dots \end{aligned}$$

Ora poniamo $I(\tilde{j}) = I(\check{j})$ e vediamo quali condizioni devono valere su \tilde{j} e \check{j} affinché sia soddisfatta l'uguaglianza: i due periodogrammi sono uguali se lo sono i vari termini che lo compongono presi uno ad uno, ovvero

$$\begin{aligned} x_1^2 &= x_1^2 \\ x_1 x_2 W^{1\tilde{j}-2\tilde{j}} &= x_1 x_2 W^{0\check{j}-1\check{j}} \\ x_1 x_T W^{1\tilde{j}-T\tilde{j}} &= x_1 x_T W^{0\check{j}-(T-1)\check{j}} \\ &\dots = \dots \end{aligned}$$

Ognuna di queste equazioni ha come soluzione $(\tilde{j}) = (\check{j})$, si trova quindi conferma di come valutando i periodogrammi alla medesima frequenza si ottengano valori identici nonostante le diverse notazioni.

Passiamo ora all'algoritmo della FFT, utilizzando la notazione in Cooley-Tukey: per prima cosa si suppone che T sia un numero composto, ovvero $T = r_1 r_2$. Si passa poi ad una riscrittura degli indici j e t :

$$\begin{aligned} j &= j_1 r_1 + j_0, \quad j_0 = 0, 1, \dots, r_1 - 1, \quad j_1 = 0, 1, \dots, r_2 - 1, \\ t &= t_1 r_1 + t_0, \quad t_0 = 0, 1, \dots, r_2 - 1, \quad t_1 = 0, 1, \dots, r_1 - 1. \end{aligned}$$

Si può quindi definire $w(j)$ come

$$w(j_1, j_0) = \sum_{t_0} \sum_{t_1} x(t_1, t_0) W^{j t_1 r_2} W^{j t_0}.$$

Dato che

$$W^{j t_1 r_2} = W^{j_1 r_1 t_1 r_2} W^{j_0 k_1 r_2} = W^{T k} W^{j_0 k_1 r_2} = W^{j_0 k_1 r_2},$$

la sommatoria su t_1 dipende solo da j_0 e t_0 e può essere definito il nuovo vettore

$$(4.6) \quad x^{(1)}(j_0, t_0) = \sum_{t_1} x(t_1, t_0) W^{j_0 t_1 r_2}.$$

La trasformata può essere scritta come

$$(4.7) \quad w(j_1, j_0) = \sum_{t_0} x^{(1)}(j_0, t_0) W^{(j_1 r_1 + j_0) t_0}.$$

Dato che ci sono T elementi in x_1 , ognuno dei quali richiede r_1 operazioni, le operazioni totali per il calcolo di x_1 sono $T r_1$; analogamente servono $T r_2$ operazioni per il calcolo di w a partire da x_1 . Il totale di operazioni è quindi $T(r_1 + r_2)$, che generalmente risulta inferiore a T^2 , che è l'ordine di grandezza delle operazioni da effettuare utilizzando l'algoritmo standard. Inoltre se r_1 e r_2 risultassero a loro volta essere numeri composti, possono essere scomposti ulteriormente ripercorrendo i passi dell'algoritmo per avere ulteriori riduzioni nel numero di operazioni. Di seguito viene proposto un esempio, prendendo una serie di lunghezza $T = 8$ i cui elementi sono indicati con x_1, x_2, \dots, x_8 . Utilizzando la formulazione classica della trasformata di Fourier, si ottiene la seguente rappresentazione:

$$\begin{aligned}
w(0) &= x_1 W^0 & +x_2 W^0 & +x_3 W^0 & +x_4 W^0 & +x_5 W^0 & +x_6 W^0 & +x_7 W^0 & +x_8 W^0 \\
w(1) &= x_1 W^0 & +x_2 W^1 & +x_3 W^2 & +x_4 W^3 & +x_5 W^4 & +x_6 W^5 & +x_7 W^6 & +x_8 W^7 \\
w(2) &= x_1 W^0 & +x_2 W^2 & +x_3 W^4 & +x_4 W^6 & +x_5 W^8 & +x_6 W^{10} & +x_7 W^{12} & +x_8 W^{14} \\
w(3) &= x_1 W^0 & +x_2 W^3 & +x_3 W^6 & +x_4 W^9 & +x_5 W^{12} & +x_6 W^{15} & +x_7 W^{18} & +x_8 W^{21} \\
& & & & \vdots & & & & \\
w(7) &= x_1 W^0 & +x_2 W^7 & +x_3 W^{14} & +x_4 W^{21} & +x_5 W^{28} & +x_6 W^{35} & +x_7 W^{42} & +x_8 W^{49}.
\end{aligned}$$

Tale schema si può semplificare data la proprietà espressa in formula (4.4) nel seguente modo:

$$\begin{aligned}
w(0) &= x_1 & +x_2 & +x_3 & +x_4 & +x_5 & +x_6 & +x_7 & +x_8 \\
w(1) &= x_1 & +x_2 W^1 & +x_3 W^2 & +x_4 W^3 & +x_5 W^4 & +x_6 W^5 & +x_7 W^6 & +x_8 W^7 \\
w(2) &= x_1 & +x_2 W^2 & +x_3 W^4 & +x_4 W^6 & +x_5 & +x_6 W^2 & +x_7 W^4 & +x_8 W^6 \\
w(3) &= x_1 & +x_2 W^3 & +x_3 W^6 & +x_4 W^1 & +x_5 W^4 & +x_6 W^7 & +x_7 W^2 & +x_8 W^5 \\
w(4) &= x_1 & +x_2 W^4 & +x_3 & +x_4 W^4 & +x_5 & +x_6 W^4 & +x_7 & +x_8 W^4 \\
w(5) &= x_1 & +x_2 W^5 & +x_3 W^2 & +x_4 W^7 & +x_5 W^4 & +x_6 W^1 & +x_7 W^6 & +x_8 W^3 \\
w(6) &= x_1 & +x_2 W^6 & +x_3 W^4 & +x_4 W^2 & +x_5 & +x_6 W^6 & +x_7 W^4 & +x_8 W^2 \\
w(7) &= x_1 & +x_2 W^7 & +x_3 W^6 & +x_4 W^5 & +x_5 W^4 & +x_6 W^3 & +x_7 W^2 & +x_8 W.
\end{aligned}$$

Si può a questo punto notare come i coefficienti di x_t compaiano seguendo uno schema ben preciso: ad esempio si nota che i coefficienti di x_1 e x_5 o sono uguali (entrambi W^0) o differiscono del fattore W^4 . Una cosa analoga succede per i coefficienti di x_2 e x_6 e così via: si può quindi pensare che basti una combinazione opportuna di queste coppie di elementi sommati tra loro per generare la serie $w(j)$ sopra rappresentata. Elenchiamo queste coppie e notiamo che esse corrispondono agli elementi del vettore $x^{(1)}(j_0, t_0)$ calcolati come in formula (4.6) (in questo esempio $r_1 = 2$ e $r_2 = 4$): Per questo calcolo servono massimo 16 operazioni, dove

$$\begin{array}{ll}
x_1 + x_5 & x^{(1)}(0, 0) \\
x_2 + x_6 & x^{(1)}(0, 1) \\
x_3 + x_7 & x^{(1)}(0, 2) \\
x_4 + x_8 & x^{(1)}(0, 3) \\
x_1 + x_5 W^4 & x^{(1)}(1, 0) \\
x_2 + x_6 W^4 & x^{(1)}(1, 1) \\
x_3 + x_7 W^4 & x^{(1)}(1, 2) \\
x_4 + x_8 W^4 & x^{(1)}(1, 3).
\end{array}$$

per operazione si intende una moltiplicazione seguita da una addizione (massimo perchè, come in questo caso, spesso il fattore moltiplicativo è W^0). La trasformata w_j può ora essere ricavata prendendo questi elementi e moltiplicandoli per un opportuno coefficiente, e questo è ciò che viene sintetizzato in formula (4.7):

$$\begin{aligned}
w(0) &= x^{(1)}(0, 0)W^0 & +x^{(1)}(0, 1)W^0 & +x^{(1)}(0, 2)W^0 & +x^{(1)}(0, 3)W^0 \\
w(1) &= x^{(1)}(1, 0)W^0 & +x^{(1)}(1, 1)W^1 & +x^{(1)}(1, 2)W^2 & +x^{(1)}(1, 3)W^3 \\
w(2) &= x^{(1)}(0, 0)W^0 & +x^{(1)}(0, 1)W^2 & +x^{(1)}(0, 2)W^4 & +x^{(1)}(0, 3)W^6 \\
w(3) &= x^{(1)}(1, 0)W^0 & +x^{(1)}(1, 1)W^3 & +x^{(1)}(1, 2)W^6 & +x^{(1)}(1, 3)W^1 \\
w(4) &= x^{(1)}(0, 0)W^0 & +x^{(1)}(0, 1)W^4 & +x^{(1)}(0, 2)W^0 & +x^{(1)}(0, 3)W^4 \\
w(5) &= x^{(1)}(1, 0)W^0 & +x^{(1)}(1, 1)W^5 & +x^{(1)}(1, 2)W^2 & +x^{(1)}(1, 3)W^7 \\
w(6) &= x^{(1)}(0, 0)W^0 & +x^{(1)}(0, 1)W^6 & +x^{(1)}(0, 2)W^4 & +x^{(1)}(0, 3)W^2 \\
w(7) &= x^{(1)}(1, 0)W^0 & +x^{(1)}(1, 1)W^7 & +x^{(1)}(1, 2)W^6 & +x^{(1)}(1, 3)W^5.
\end{aligned}$$

Il numero di operazioni effettuate è $8 \cdot 4 = 32$, il che porta il totale a 48, minore delle $8^2 = 64$ necessarie nella risoluzione standard. Tuttavia, osservando l'ultimo

set di equazioni, si nota come anche in questo caso ci sia uno schema ripetitivo e quindi possibilità di ulteriori semplificazioni: i coefficienti di $x^{(1)}(0, 0)$ e $x^{(1)}(0, 2)$ o sono uguali o differiscono per il fattore W^4 , stessa cosa per quelli di $x^{(1)}(0, 1)$ e $x^{(1)}(0, 3)$, mentre i coefficienti di $x^{(1)}(1, 0)$ e $x^{(1)}(1, 2)$ differiscono per il fattore W^2 o W^4 , stessa cosa per quelli di $x^{(1)}(1, 1)$ e $x^{(1)}(1, 3)$. Si possono compattare questi elementi generando il seguente vettore: A questo punto la trasformata di Fourier

$$\begin{array}{ll} x^{(1)}(0, 0) + x^{(1)}(0, 2) & x^{(2)}(0) \\ x^{(1)}(0, 1) + x^{(1)}(0, 3) & x^{(2)}(1) \\ x^{(1)}(1, 0) + x^{(1)}(1, 2)W^2 & x^{(2)}(2) \\ x^{(1)}(1, 1) + x^{(1)}(1, 3)W^2 & x^{(2)}(3) \\ x^{(1)}(0, 0) + x^{(1)}(0, 2)W^4 & x^{(2)}(4) \\ x^{(1)}(0, 1) + x^{(1)}(0, 3)W^4 & x^{(2)}(5) \\ x^{(1)}(1, 0) + x^{(1)}(1, 2)W^6 & x^{(2)}(6) \\ x^{(1)}(1, 1) + x^{(1)}(1, 3)W^6 & x^{(2)}(7). \end{array}$$

$w(j)$ è data da Risulta chiaro come ad ogni passaggio il numero di operazione da

$$\begin{array}{l} w(0) = x^{(2)}(0) + x^{(2)}(1) \\ w(1) = x^{(2)}(2) + x^{(2)}(3)W^1 \\ w(2) = x^{(2)}(4) + x^{(2)}(5)W^2 \\ w(3) = x^{(2)}(6) + x^{(2)}(7)W^3 \\ w(4) = x^{(2)}(0) + x^{(2)}(1)W^4 \\ w(5) = x^{(2)}(2) + x^{(2)}(3)W^5 \\ w(6) = x^{(2)}(4) + x^{(2)}(5)W^6 \\ w(7) = x^{(2)}(6) + x^{(2)}(7)W^7. \end{array}$$

effettuare sia pari a $2 \cdot T$, e il numero di passaggi, se T è una potenza di 2, sia pari a $\log_2 T$: il numero di operazioni necessarie per il calcolo di $w(j)$ risulta pertanto $2T \log_2 T$. La scelta $T = 2^m$ risulta pertanto particolarmente vantaggiosa in quanto si ha una grossa riduzione dei tempi computazionali; inoltre, nell'implementazione dell'algoritmo mediante compilatore, permette di utilizzare la vantaggiosa rappresentazione binaria. Gli indici j e t possono essere infatti espressi come:

$$\begin{aligned} j &= j_{m-1}2^{m-1} + \dots + j_12 + j_0, \\ t &= t_{m-1}2^{m-1} + \dots + t_12 + t_0, \end{aligned}$$

dove j_v e t_v assumono valori 0 o 1 e sono uguali al corrispondente valore della cifra di j o t espressi in codifica binaria. I diversi vettori sono ottenuti tramite il seguente procedimento ricorsivo:

$$\begin{aligned} &x^{(l)}(j_0, \dots, j_{l-1}, t_{m-l-1}, \dots, t_0) = \\ &= \sum_{t_{m-1}} x^{(l-1)}(j_0, \dots, j_{l-2}, t_{m-l}, \dots, t_0) W^{(j_{l-1}2^{l-1} + \dots + j_0)t_{m-1}2^{m-l}}, \end{aligned}$$

con $l = 1, 2, \dots, m$. Il vettore finale $x^{(m)}$ contiene tutti gli elementi di $w(j)$ ma in ordine invertito rispetto alla rappresentazione binaria. L'ultimo passaggio per avere la trasformata è pertanto un riordinamento:

$$(4.8) \quad w(j_{m-1}, \dots, j_0) = x^{(m)}(j_0, \dots, j_{m-1}).$$

In seguito a questo primo algoritmo, sono stati proposti negli anni svariati altri algoritmi, allo scopo di ridurre i tempi di calcolo della FFT in situazioni più generiche (ad esempio nel caso in cui T non fosse multiplo di 2) o in situazioni

più particolari (ad esempio nel caso in cui le serie storiche fossero reali o simmetriche). Da un punto di vista pratico durante questa tesi si utilizzerà la funzione `fft` disponibile in Matlab, basata sulla libreria `fftw`. In questa libreria, nota per la sua velocità, sono presenti svariate routines, che di volta in volta vengono scelte in maniera opportuna per minimizzare i tempi. Se ad esempio T dovesse essere una potenza di 2, la scelta ricadrebbe ovviamente sull'algoritmo di Cooley-Tukey ed eventuali varianti.

2. Generazione di processi integrati frazionari

Esistono vari algoritmi per generare processi integrati frazionari. Il metodo più intuitivo è quello che prende spunto dalla definizione stessa di processo integrato frazionario x_t , con $t = 1, \dots, T$:

$$x_t = (1 - L)^{-d} \varepsilon_t,$$

dove L è l'operatore ritardo e ε_t è un rumore bianco. Il processo x_t , tralasciando i dettagli già visti nel Capitolo 1, ammette la rappresentazione $MA(\infty)$ data da

$$x_t = \sum_{j=0}^{\infty} b_j \varepsilon_{t-j},$$

con

$$b_j = \frac{\Gamma(j+d)}{\Gamma(d)\Gamma(j+1)}, \quad j \geq 0.$$

Usando quest'ultima formula, dopo aver generato un rumore bianco e calcolato i termini b_j (sempre nel Capitolo 1 si è vista una procedura computazionalmente conveniente per il loro calcolo), è quindi possibile ottenere un processo integrato frazionario di qualsiasi ordine di integrazione. Il numero di operazioni da eseguire risulta piuttosto elevato: dato l'avvicinarsi molto lento dei termini b_j a 0 è infatti consigliabile non trascurare nessuno di essi (e quindi è consigliabile non troncature il filtro): pertanto alla prima iterazione serve una operazione, alla seconda ne servono due ... e alla generica t -esima iterazione servono t operazioni per il calcolo di x_t . Il numero di operazioni è quindi paragonabile alla somma di interi da 1 a T , che è proporzionale a T^2 . Tuttavia facendo così rimane il problema che i primi elementi di x_t sono ottenuti a partire da pochi elementi ε_t , e si può supporre che ciò implichi che il processo integrato sia inizialmente da considerarsi approssimato piuttosto che esatto: è consigliabile pertanto filtrare una serie ε_t , $t = 0, \dots, K$, $K > T$ e considerare come processo integrato frazionario soltanto gli ultimi T elementi di x_t .

Un altro metodo per generare rumore Gaussiano frazionario esatto è l'algoritmo di Davies e Harte (1987), di cui di seguito se ne illustrano i passi utilizzando la scrittura in Dieker e Mandjes [7]. In questo articolo, come spesso accade nella letteratura matematica e fisica, si utilizza il parametro H , legato a d dalla relazione $H = d + 0.5$. Sempre in questa letteratura spesso si fa riferimento ai processi integrati frazionari stazionari con il termine di 'rumore Gaussiano frazionario' e alle loro cumulate con il termine di 'moto Browniano frazionario' (a tempo discreto).

Come prima cosa viene calcolata la funzione di autocovarianza $\gamma(k)$ del processo x_t , data da

$$(4.9) \quad \gamma(k) = \mathbb{E}x_{t+k}x_t = \frac{1}{2} \left[|k-1|^{2H} - 2|k|^{2H} + |k+1|^{2H} \right],$$

con $k = 0, 1, \dots, T - 1$. Si definisce quindi la serie α_k nel seguente modo:

$$\begin{aligned} \alpha_k &= \gamma(k) && \text{per } k = 0, 1, \dots, T - 1, \\ \alpha_k &= 0 && \text{per } k = T, \\ \alpha_k &= \gamma(2t - K) && \text{per } k = T + 1, T + 2, \dots, 2T - 1. \end{aligned}$$

A questo punto si calcola la trasformata discreta di Fourier $w^{(0)}(k)$ di α_k , usando l'algoritmo della FFT. α_k e conseguentemente $w(k)$ sono costituiti da $2T$ elementi, e il numero di operazioni necessarie per il calcolo di $w(k)$ è di ordine $T \log_2 T$. Si generano ora due processi white noise tra di loro indipendenti $U_k^{(0)}$ e $U_k^{(1)}$ con media nulla e varianza unitaria, con $k = 0, \dots, T$, e si definisce la serie u_k nella seguente maniera:

$$\begin{aligned} u_k &= \sqrt{\frac{w_k}{2T}} U_k^{(0)} && \text{per } k = 0; \\ u_k &= \sqrt{\frac{w_k}{4T}} \left(U_k^{(0)} + iU_k^{(1)} \right) && \text{per } k = 1, \dots, T - 1; \\ u_k &= \sqrt{\frac{w_k}{2T}} U_k^{(0)} && \text{per } k = T; \\ u_k &= \sqrt{\frac{w_k}{4T}} \left(U_{2T-k}^{(0)} + iU_{2T-k}^{(1)} \right) && \text{per } k = T + 1, \dots, 2T - 1. \end{aligned}$$

Si calcola infine la trasformata di Fourier $w^{(1)}(k)$ di u_k , entrambi sempre di dimensione $2T$ per un totale di operazioni di ordine $T \log_2 T$. Il rumore Gaussiano frazionario x_t con $t = 1, \dots, T$ è dato dai primi T elementi del vettore w_k^1 con $k = 0, \dots, T - 1$.

3. Metodo di Newton

Il metodo di Newton è un metodo iterativo per il calcolo degli zeri di una funzione $f(x)$. Definita con \bar{x} la soluzione di $f(x)$, tale metodo afferma che partendo da un generico punto $x(1)$ attraverso le iterazioni definite dall'equazione

$$x(v+1) = x(v) - \frac{f(x_v)}{f'(x_v)},$$

dove $v = 1, 2, \dots, \Upsilon$ sono i passi, si converge rapidamente al punto \bar{x} , a patto di partire da una condizione iniziale sufficientemente vicina alla soluzione. L'interesse verso questi metodi tipici del calcolo numerico è tanto maggiore quanto la funzione risulti difficile, o impossibile, da esplicitare rispetto alla variabile in questione, e questo è il caso della funzione A.6 descritta nel capitolo precedente

$$R(H) = \log \hat{G}(H) - (2H - 1) \frac{1}{m} \sum_{j=1}^m \log \lambda_j, \quad \hat{G}(H) = \frac{1}{m} \sum_{j=1}^m \lambda_j^{2H-1} I_j.$$

Si ricorda che la stima \hat{H} coincide con il valore di H che minimizza tale funzione, che però non è esplicitabile rispetto al parametro H . Ne segue che un possibile approccio al problema è quello di sostituire ad H diversi valori presi da una griglia, e scegliere poi quello ottimale. Se si fosse interessati ad esempio ad avere una stima precisa alla quarta cifra decimale, bisognerebbe variare H da 0 ad 1 con passo 0.0001, il che porta a dover calcolare 10000 volte la funzione sopra indicata. In generale una stima precisa alla n -esima cifra decimale comporta dover calcolare 10^n volte la funzione. Intervalli più piccoli possono essere scelti qualora si abbia già idea di dove possa situarsi il vero valore H_0 , ma l'ordine di grandezza delle volte che si implementa la funzione rimane lo stesso.

Utilizzando il metodo di Newton, si può notevolmente ridurre questo numero. Per prima cosa, si nota come la funzione $f(x)$ di cui bisogna trovare lo zero sia $R'(H)$, in quanto trovando il punto in cui si annulla la derivata prima di $R(H)$ si trova il punto di minimo. Di conseguenza $f'(x)$ è data da $R''(H)$. Partendo da un valore di H generico, ma ragionevole (esempio $H = 0.5$), si può impostare la seguente equazione:

$$(4.10) \quad H(v+1) = H(v) - \frac{R'(H_v)}{R''(H_v)}.$$

Affianco alla rapidità computazionale, possono presentarsi tuttavia alcuni svantaggi che solitamente caratterizzano le tecniche del calcolo numerico. In primo luogo (punto a) si ha generalmente una minor robustezza del metodo, ad esempio nel caso in cui la funzione da minimizzare presenti diversi minimi locali o sia discontinua. Bisognerebbe poi investigare (punto b) se il metodo converga sempre al valore esatto e se questo accada in tutte le regioni (per diversi valori di H). Nello specifico del metodo qui proposto, è necessario inoltre osservare che le derivate (prima e seconda) sono ottenute per via numerica (utilizzando la definizione di rapporto incrementale: $R' = (R(H) + R(H + \epsilon)) / \epsilon$), e a seconda del valore di ϵ utilizzato il metodo potrebbe convergere a valori più esatti ma anche diventare più instabile. Occorre quindi determinare il grado di precisione del metodo e trovare per via sperimentale un set di valori (Υ, ϵ) in grado di essere sufficientemente precisi, stabili e veloci.

Come prima cosa (punto a) mostriamo alcune funzioni $R(H)$ per diversi valori di H e m . In figura 3 vengono mostrate 10 funzioni per diversi valori di m , il tutto fatto con $H = 0.05$ e $H = 0.95$.

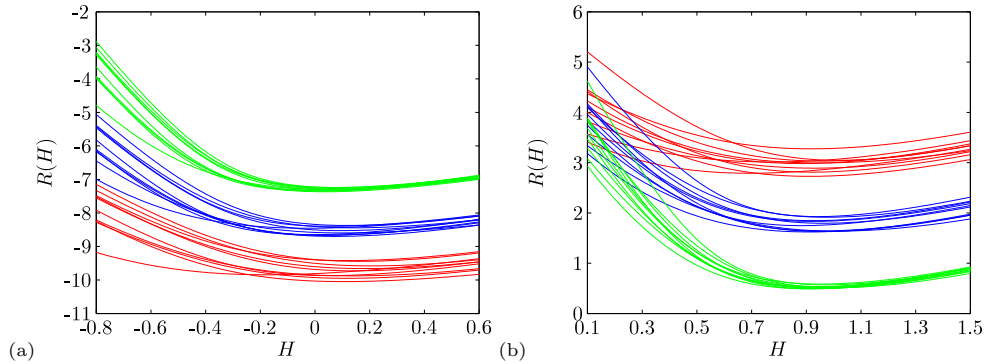


FIGURA 4.2. Funzioni $R(H)$ per diversi valori di m : $m = 2^5$ in rosso, $m = 2^7$ in blu e $m = 2^9$ in verde. Le serie storiche di lunghezza $T = 2^{16}$ sono state simulate con $H = 0.05$ (a) e $H = 0.95$ (b)

Oltre a notare come effettivamente le funzioni siano regolari con un solo punto di minimo, si nota anche come siano più disperse per valori di m più piccoli (causando una maggiore varianza nelle stime).

Passando ora (punto b) ad un confronto tra il risultati ottenuti con un metodo a griglia (con passo 0.0001) e con il metodo di Newton, si determinano le differenze tra le stime ottenute $\Delta = H_g - H_N$ su 1000 simulazioni, caratterizzate da $m = 2^9$, $T = 2^{16}$ e diversi valori di H , e si mostra in figura (3) sotto forma di istogramma la frequenza di tale differenza. I parametri relativi al metodo di Newton sono $\epsilon = 10^{-5}$

e $\Upsilon = 4$ e 5. Considerando che il metodo a griglia tronca il valore della stima di

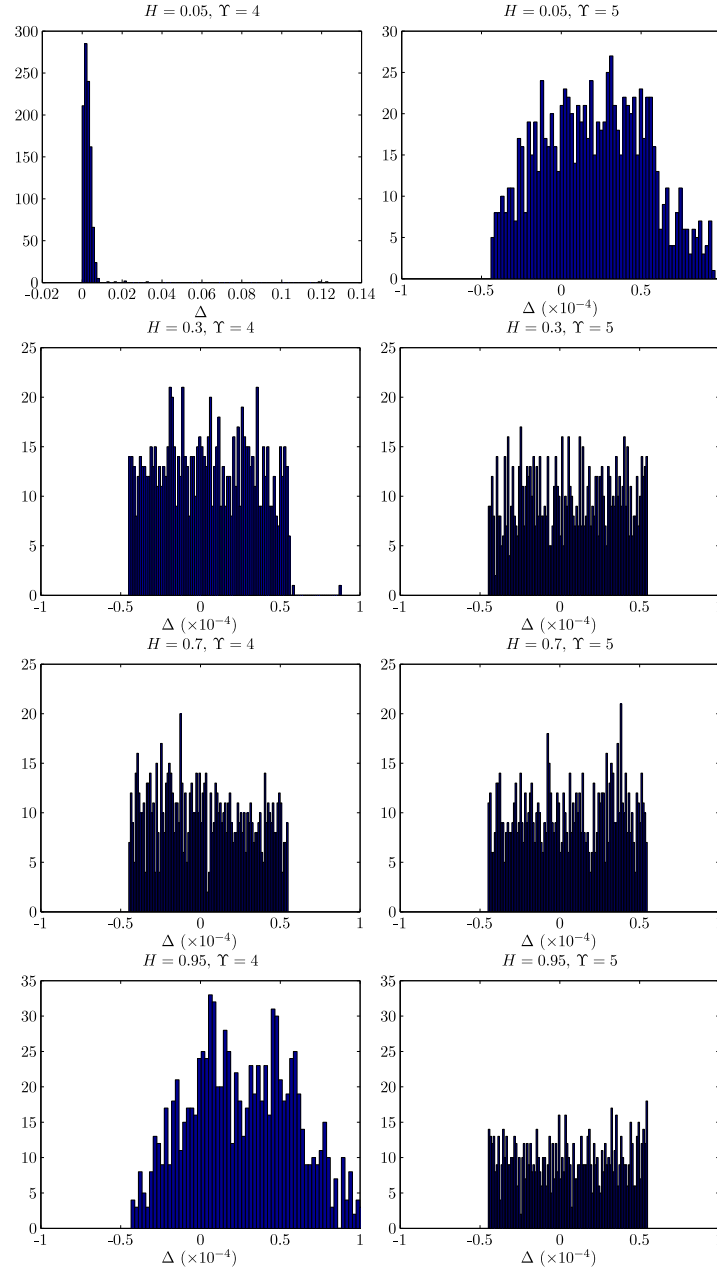


FIGURA 4.3. Rappresentazione mediante istogramma delle differenze tra le stime ottenute $\Delta = H_g - H_N$ su 1000 simulazioni: $m = 2^9$, $T = 2^{16}$, $\epsilon = 10^{-5}$, $\Upsilon = 4$ nella colonna di sinistra, $\Upsilon = 5$ nella colonna di destra, H indicato in figura.

H alla quarta cifra decimale (e quindi l'approssimazione è dell'ordine $5 \cdot 10^{-5}$), si ci aspetta che se il metodo di Newton stima correttamente la quarta cifra decimale

allora la differenza tra le due misure non sia maggiore di $5 \cdot 10^{-5}$, e, data l'uniformità di questa approssimazione, la media del valore assoluto delle differenze sia di circa $2.5 \cdot 10^{-5}$. Osservando le figure si può dunque pensare che per valori di H centrali basti $\Upsilon = 4$ per stimare correttamente la 4 cifra; per valori di H più periferici la convergenza è più lenta (ma non eccessivamente, infatti con $\Upsilon = 5$ si hanno risultati nettamente migliori). In queste 8000 simulazioni c'è stato un caso in cui il metodo è diverso (per $H = 0.05$ e $\Upsilon = 5$, l'osservazione è stata eliminata). In tabella (4.1) vengono rappresentate alcune statistiche riguardanti le simulazioni, in particolare media delle stime \bar{H}_g e \bar{H}_N e media dei valori assoluti delle differenze $\bar{\Delta}$. Si può notare come sembra sempre esserci una bias negativa di H_N rispetto a

	H			
	0.05	0.3	0.7	0.95
$\Upsilon = 4$	$\bar{H}_g = 0.04989$ $\bar{H}_N = 0.04693$ $\bar{\Delta} = 0.00296$	$\bar{H}_g = 0.29948$ $\bar{H}_N = 0.29948$ $\bar{\Delta} = 2.47 \cdot 10^{-5}$	$\bar{H}_g = 0.69888$ $\bar{H}_N = 0.69888$ $\bar{\Delta} = 2.45 \cdot 10^{-5}$	$\bar{H}_g = 0.95083$ $\bar{H}_N = 0.95080$ $\bar{\Delta} = 3.78 \cdot 10^{-5}$
$\Upsilon = 5$	$\bar{H}_g = 0.04912$ $\bar{H}_N = 0.04910$ $\bar{\Delta} = 3.37 \cdot 10^{-5}$	$\bar{H}_g = 0.29908$ $\bar{H}_N = 0.29908$ $\bar{\Delta} = 2.54 \cdot 10^{-5}$	$\bar{H}_g = 0.69866$ $\bar{H}_N = 0.69865$ $\bar{\Delta} = 2.55 \cdot 10^{-5}$	$\bar{H}_g = 0.95206$ $\bar{H}_N = 0.95205$ $\bar{\Delta} = 2.61 \cdot 10^{-5}$

TABELLA 4.1. Media delle stime \bar{H}_g e \bar{H}_N e media del valore assoluto della loro differenza $\bar{\Delta}$; $m = 2^9$, $T = 2^{16}$ e $\epsilon = 10^{-5}$.

H_g (ovvero il metodo di Newton trova valori mediamente un po' più piccoli rispetto alla griglia). Questo può essere dovuto al fatto che il metodo, per come è fatta la funzione, converge sempre da sinistra, nonostante H sia sempre inizializzato a 0.5; alcuni esempi di convergenza del metodo per diverse funzioni $R(H)$ sono mostrati in figura (4.4) (i passi iterati sono 10). Sempre dalla figura 4.4 si vede come il caso più critico si ha per $H = 0.05$, dove spesso il primo passo porta ad un peggioramento (si ci allontana dal minimo in termini di $R(H)$). Se tale peggioramento è marcato, talvolta, come nel caso prima visto, il metodo inizia a divergere risalendo lungo la funzione verso infinito.

Passando infine alla determinazione del livello di precisione raggiungibile (punto c), mostriamo in tabella (4.2) i valori assunti da H_N ad ogni passo delle precedenti simulazioni, al variare di ϵ , e confrontiamo questi valori con H_g determinato con griglia molto fine (10 cifre decimali). Un valore di ϵ troppo grande (come ad esempio 10^{-5}) porta a risultati più grossolani (precisione massima alla quarta cifra decimale), mentre con $\epsilon = 10^{-6}$ e $\epsilon = 10^{-7}$ la precisione giunge alla sesta cifra. Tuttavia valori di ϵ troppo piccoli rendono il metodo instabile, come si può vedere per $\epsilon = 10^{-8}$. La velocità di convergenza non sembra dipendere da ϵ , quanto piuttosto dalla inizializzazione più o meno vicina: per i valori centrali $H = 0.3$ e $H = 0.7$ in cui l'inizializzazione $H = 0.5$ è più vicina bastano 4 passi per raggiungere il grosso delle cifre stimabili con precisione.

Concludendo, nella scelta tra metodo di Newton e metodo a griglia bisogna tener conto che il primo risulta più veloce ma presenta problemi di precisione finita e in alcuni casi di instabilità; ovviamente si può pensare a diverse soluzioni intermedie, ad esempio un metodo ibrido in cui una griglia molto larga identifichi l'inizializzazione più opportuna, successivamente il metodo di Newton converga rapidamente alla 4/5 cifra decimale e un'ulteriore griglia più fine giunga al livello di

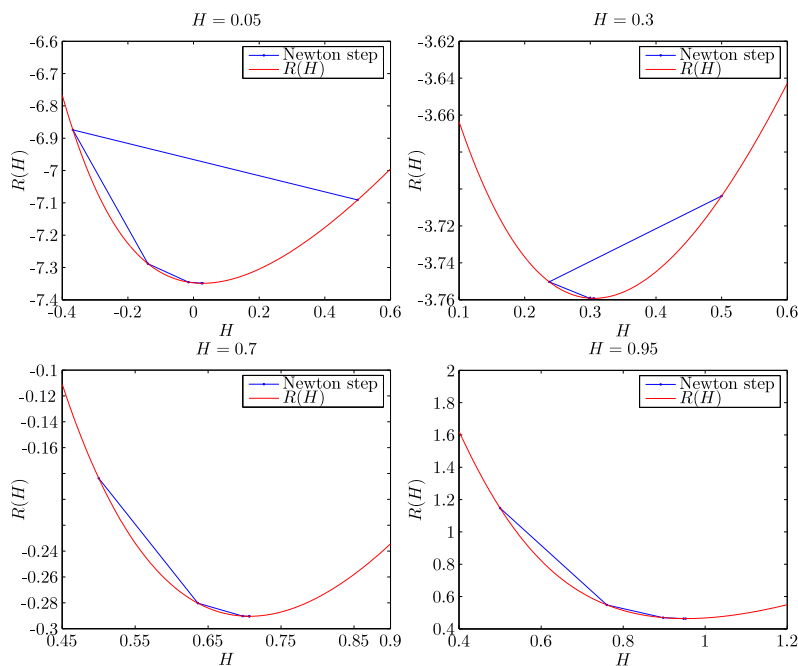


FIGURA 4.4. Esempi di convergenza del metodo di Newton, in 4 simulazioni con H differenti: in rosso la funzione $R(H)$, in blu i diversi passi.

precisione desiderato. Infine, tornando a considerazioni di rapidità computazionale, si nota che la funzione $R(H)$ dipende linearmente da m : si ci aspetta quindi che i tempi computazionali di entrambi i metodi crescano secondo la stessa legge, ovvero proporzionalmente ad m .

4. Codici utilizzati

In questa sezione vengono riportati e rapidamente descritti i codici utilizzati in sede di generazione e stima di processi integrati frazionari. Si inizia mostrando il codice per la generazione di un processo integrato frazionario usando la rappresentazione $MA(\infty)$:

```

1 b1=zeros(1,T);
2 b1(1)=-d;
3 for i=2:T
4 b1(i)=-b2(i-1)*(d-i+1)/i;
5 end
6 e=normrnd(0,1,1,T);
7 f=e(1:T);
8 for i=1:T
9 f(i)=f(i)+e(1:i-1)*b1(i-1:-1:1)';
10 end .

```

		$H = 0.05$	$H = 0.3$	$H = 0.7$	$H = 0.95$
$\epsilon = 10^{-5}$	$v = 1$	0.5000000000	0.5000000000	0.5000000000	0.5000000000
	$v = 2$	-0.3676729687	0.2374086764	0.6358458769	0.7601916383
	$v = 3$	-0.1391739668	0.2983885358	0.6973297697	0.8971818411
	$v = 4$	-0.0151453651	0.3054550357	0.7063956033	0.9472000173
	$v = 5$	0.024535470	0.3055379084	0.7065554620	0.9521943604
	$v = 6$	0.0276294617	0.3055379184	0.7065555081	0.9522373923
	$v = 7$	0.0276458804	0.3055379185	0.7065555081	0.9522373932
	$v = 8$	0.0276458805	0.3055379185	0.7065555081	0.9522373933
	$v = 9$	0.0276458806	0.3055379185	0.7065555081	0.9522373933
	$v = 10$	0.0276458806	0.3055379185	0.7065555081	0.9522373934
	H_g	0.0276508366	0.3055429225	0.7065605068	0.9522423617
	$R_N(H)$	0.0000000002	0	0	-0.0000000001
$R_N(H)$	-7.3487700736	-3.7591852735	-0.2904278986	0.4637849438	
$R_g(H)$	-7.3487700737	-3.7591852736	-0.2904278986	0.4637849437	
$\epsilon = 10^{-6}$	$v = 1$	0.5000000000	0.5000000000	0.5000000000	0.5000000000
	$v = 2$	-0.3673375771	0.2375275500	0.6358605366	0.7601629487
	$v = 3$	-0.1389598107	0.2984185016	0.6973443399	0.8970883005
	$v = 4$	-0.0151153919	0.3054569573	0.7064004857	0.9470520570
	$v = 5$	0.0245193842	0.3055423698	0.7065599945	0.9521891877
	$v = 6$	0.0276412658	0.3055424188	0.7065600074	0.9522417807
	$v = 7$	0.0276503914	0.3055424180	0.7065600085	0.9522418935
	$v = 8$	0.0276503806	0.3055424182	0.7065600082	0.9522418932
	$v = 9$	0.0276503796	0.3055424189	0.7065600076	0.9522418939
	$v = 10$	0.0276503806	0.3055424189	0.7065600084	0.9522418916
	$v = 10$	0.0276503816	0.3055424189	0.7065600078	0.9522418931
	H_g	0.0276508366	0.3055429225	0.7065605068	0.9522423617
$R_N(H)$	0.0000000035	0	-0.0000000022	-0.0000000044	
$R_N(H)$	-7.3487700737	-3.7591852736	-0.2904278986	0.4637849437	
$R_g(H)$	-7.3487700737	-3.7591852736	-0.2904278986	0.4637849437	
$\epsilon = 10^{-7}$	$v = 1$	0.5000000000	0.5000000000	0.5000000000	0.5000000000
	$v = 2$	-0.3528815000	0.2661899460	0.6381330759	0.7570010681
	$v = 3$	-0.1222152439	0.3024327330	0.6976190222	0.9067580068
	$v = 4$	-0.0179266150	0.3056075001	0.7065959465	0.9537075362
	$v = 5$	0.0230981995	0.3055472071	0.7065611277	0.9521898119
	$v = 6$	0.0271613551	0.3055423654	0.7065604981	0.9522303509
	$v = 7$	0.0276153272	0.3055428834	0.7065604569	0.9522462526
	$v = 8$	0.0276624805	0.3055428684	0.7065604603	0.9522417864
	$v = 9$	0.0276508705	0.3055428733	0.7065604651	0.9522423154
	$v = 10$	0.0276508333	0.3055428595	0.7065604503	0.9522423363
	$v = 10$	0.0276508333	0.3055428679	0.7065604599	0.9522423463
	H_g	0.0276508366	0.3055429225	0.7065605068	0.9522423617
$R_N(H)$	0	-0.0000000177	0.0000000111	-0.0000000004	
$R_N(H)$	-7.3487700737	-3.7591852736	-0.2904278986	0.4637849437	
$R_g(H)$	-7.3487700737	-3.7591852736	-0.2904278986	0.4637849437	
$\epsilon = 10^{-8}$	$v = 1$	0.5000000000	0.5000000000	0.5000000000	0.5000000000
	$v = 2$	0.3976542300	0.4766190040	-0.0454531237	0.5894363885
	$v = 3$	0.3679954866	-Inf	0.0139373409	0.7246703973
	$v = 4$	0.4525389066	NaN	-0.1445275832	0.7688099050
	$v = 5$	0.5490233566	NaN	-0.2042315748	Inf
	$v = 6$	0.4411494766	NaN	-0.1256747091	NaN
	$v = 7$	0.5361442766	NaN	0.0872967188	NaN
	$v = 8$	-Inf	NaN	0.1554290144	NaN
	$v = 9$	NaN	NaN	-0.3330670622	NaN
	$v = 10$	NaN	NaN	0.0445836844	NaN
	$v = 10$	NaN	NaN	0.1967716794	NaN
	H_g	0.0276508366	0.3055429225	0.7065605068	0.9522423617
$R_N(H)$	NaN	NaN	-5.7797536534	NaN	
$R_N(H)$	NaN	NaN	0.7947524410	NaN	
$R_g(H)$	-7.3487700737	-3.7591852736	-0.2904278986	0.4637849437	

TABELLA 4.2. Primi 10 passi del metodo numerico al variare di ϵ per diversi valori di H . In grassetto sono indicate le cifre esatte.

Dato che spesso i compilatori calcolano la funzione $\Gamma(\cdot)$ solo entro determinati valori (ad esempio in Matlab fino a circa 160), è stato utilizzato il metodo iterativo (si veda il Capitolo 1) per il calcolo dei coefficienti b_j (righe 3-5), che ovvia a questo problema evitando di dover ricorrere a formule approssimate. Il rumore bianco è indicato con e (riga 6), e il processo integrato frazionario con f (riga 9).

Ora si mostra il codice per la generazione di un processo integrato frazionario utilizzando l'algoritmo di Davies e Harte:

```

1 gamma=0.5*(abs([0:T]-1).^ (2*H)-2*abs([0:T]).^(2*H)+abs([0:T]+1).^ (2*H));
2 gamma(T+1)=0;
3 gamma(T+2:2*T)=gamma(T:-1:2);
4 trasf=real(fft(gamma));
5 R1=normrnd(0,1,1,(T+1));
6 R2=normrnd(0,1,1,(T+1));
7 i=sqrt(-1);
8 w(1)=sqrt(trasf(1).*(1/(2*T)))*R1(1);
9 w(2:T)=sqrt(trasf(2:T).*(1/(4*T))).*(R1(2:T)+R2(2:T).*i);
10 w(T+1)=sqrt(trasf(T+1).*(1/(2*T)))*R1(T+1);
11 w(T+2:2*T)=sqrt(trasf(T+2:2*T).*(1/(4*T))).*(R1(T:-1:2)-R2(T:-1:2).*i);
12 trasf2=real(fft(w));
13 f=trasf2(1:T) .

```

Il codice ripercorre passo a passo il procedimento illustrato nella sezione 2 di questo capitolo. Risulta interessante notare che, a differenza del metodo precedente, in questo caso i rumori bianchi in ingresso sono due (riga 5 e 6): mentre il primo metodo può essere utilizzato per trasformare un processo integrato in un altro processo integrato con ordine di integrazione differente (basta dare in ingresso il processo integrato desiderato al posto del white noise), questo secondo metodo non presenta apparentemente una relazione tra la serie in ingresso e quella in uscita e non può essere utilizzato per la finalità appena esposta.

Passando ora ai codici per le stime, come prima cosa si mostra il codice per il calcolo del periodogramma che sfrutta la definizione standard di trasformata di Fourier:

```

1 j=sqrt(-1);
2 for i=1:m
3 vett=fgn(1:T).*exp(j*(1:T)*lambda(i));
4 tr(i)=sum(vett);
5 tr_conj(i)=conj(tr(i));
6 I(i)=(2*pi*T)^(-.5)*(2*pi*T)^(-.5)*(tr(i)*tr_conj(i));
7 end .

```

Il codice per il calcolo del periodogramma con algoritmo FFT è il seguente:

```

1 tr=fft(fgn);
2 tr=tr(2:m+1);
3 tr_conj=conj(tr);

```

```
4 I=(2*pi*T)^(-.5)*(2*pi*T)^(-.5)*(tr(1:m).*tr_conj(1:m)) .
```

La seconda riga del codice serve a traslare di un passo le trasformate, allo scopo di avere un perodogramma la cui prima posizione sia occupata dalla frequenza di indice $j = 1$ e non $j = 0$.

La stima di H (la stima di d si ottiene in maniera analoga sottraendo 0.5 dal risultato) utilizzando la griglia di valori si ottiene mediante il seguente codice:

```
1 cont=0;
2 for H=0:0.0001:1
3   [f0]=funzione(lambda,m,I,H);
4   cont=cont+1;
5   D(cont,1)=f0;
6   D(cont,2)=H;
7 end
8 A=D(:,1);
9 [a,b]=min(A);
10 H_hat=D(b,2) .
```

Utilizzando il metodo di Newton la stima di H è invece ottenuta tramite il seguente codice:

```
1 H_hat=0.5;
2 epsilon=1.e-6;
3 for i=1:5
4   F=funzione(lambda,m,I_1,H_hat);
5   F_delta=funzione(lambda,m,I_1,H_hat+epsilon);
6   F_2delta=funzione(lambda,m,I_1,H_hat+2*epsilon);
7   D1=(F_delta-F)/epsilon;
8   D2=(F_2delta-2*F_delta+F)/epsilon^2;
9   H_hat=H_hat-D1/D2;
10 end .
```

Infine si mostra il codice rappresentante la funzione $R(H)$:

```
1 G=(1/m)*lambda(1:m).^(2*H-1).*I(1:m);
2 G=sum(G);
3 R= -(2*H-1)*(1/m)*log(lambda(1:m));
4 R=sum(R);
5 f_H=R+log(G) .
```

5. Tempi di calcolo ottenuti

In questa sezione vengono mostrati i tempi di calcolo ottenuti nell'implementare i vari algoritmi descritti nelle sezioni precedenti, e si cerca di capire quali possano essere le leggi che governano tali tempi. Iniziamo con la generazione di serie storiche:

in tabella (4.3) vengono indicati i tempi computazionali per generare processi integrati frazionari utilizzando la rappresentazione $MA(\infty)$, al variare della lunghezza T delle serie (eseguite tre simulazioni per ogni T).

Serie	T								
	2^9	2^{10}	2^{11}	2^{12}	2^{13}	2^{14}	2^{15}	2^{16}	2^{17}
1	0.008	0.0155	0.047	0.141	0.520	1.870	7.223	41.30	248.1
2	0.015	0.0155	0.043	0.150	0.519	1.980	7.730	39.65	253.1
3	0.004	0.0160	0.047	0.145	0.520	1.970	7.425	39.06	250.0

TABELLA 4.3. Tempo espresso in secondi per generare serie storiche di lunghezza T indicata utilizzando la rappresentazione $MA(\infty)$.

In tabella (4.4) vengono mostrati i tempi relativi alla generazione di processi frazionari utilizzando il metodo di Davies e Harte: In figura (4.5) si mostrano i

Serie	T										
	2^{12}	2^{13}	2^{14}	2^{15}	2^{16}	2^{17}	2^{18}	2^{19}	2^{20}	2^{21}	2^{22}
1	0.008	0.023	0.039	0.092	0.203	0.450	0.910	1.730	3.390	6.712	13.98
2	0.008	0.020	0.045	0.078	0.187	0.400	0.876	1.650	3.170	6.430	13.40
3	0.015	0.024	0.047	0.094	0.195	0.435	0.916	1.728	3.160	6.310	13.35

TABELLA 4.4. Tempo espresso in secondi per generare serie storiche di lunghezza T indicata utilizzando l'algoritmo di Davies e Harte.

tempi in funzione di T relativi ai due procedimenti: si nota come essi a parità di lunghezza T siano decisamente a favore del secondo algoritmo; inoltre le leggi teoriche che governano l'andamento dei tempi (T^2 per quanto riguarda la rappresentazione $MA(\infty)$ e $T \log T$ per quanto riguarda l'algoritmo di Davies e Harte) hanno trovato conferma dai dati ottenuti.

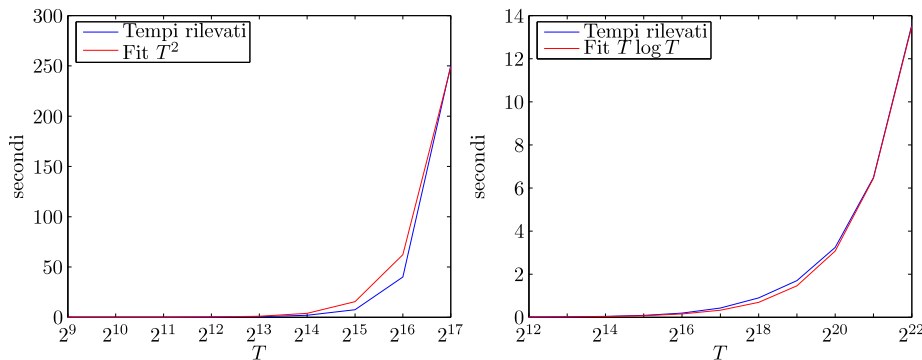


FIGURA 4.5. Tempi impiegati per generare processi integrati frazionari in funzione della lunghezza T e relativo fit utilizzando: la rappresentazione $MA(\infty)$ (sinistra) e l'algoritmo di Davies e Harte (destra).

Passando ora al calcolo del periodogramma, per prima cosa si confrontano due periodogrammi ottenuti con le metodologie viste per avere conferma che essi siano

uguali, ovvero che l'algoritmo utilizzato in Matlab sia esatto e non approssimato. In figura (4.6) viene mostrato, per diversi valori di m e T , l'andamento delle differenza logaritmica $\Delta = \log I_s - \log I_{FFT}$ in funzione della frequenza di indice j , dove I_s indica il periodogramma calcolato mediante trasformata di Fourier da definizione standard e I_{FFT} indica il periodogramma calcolato mediante algoritmo FFT. Tra

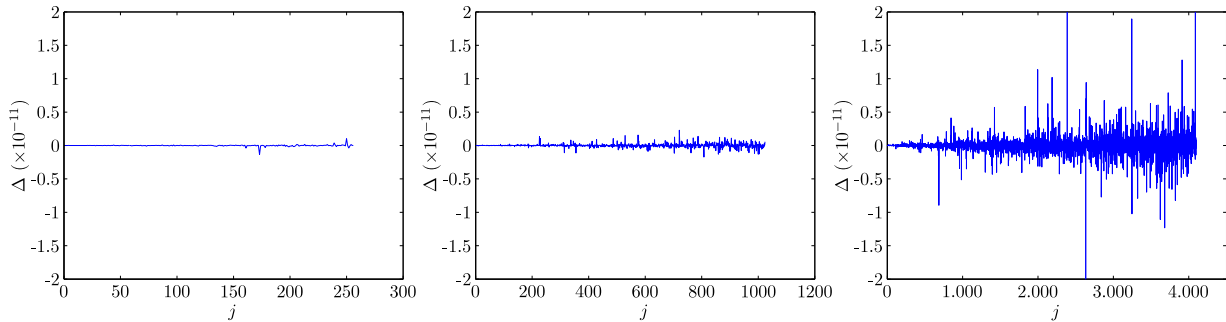


FIGURA 4.6. Differenza logaritmica $\Delta = \log I_s - \log I_{FFT}$ tra periodogramma calcolato mediante definizione standard e mediante FFT per diversi valori di T e m , indicati in figura.

i due metodi c'è effettivamente una lieve differenza, che sembra crescere con T e m : tuttavia tale differenza, probabilmente dovuta al calcolatore, è così lieve che verrà trascurata. In tabella (4.5) vengono mostrati i tempi computazionali relativi al calcolo del periodogramma utilizzando la definizione di trasformata di Fourier. I tempi sono influenzato sia da T che da m , in quanto da un lato tutte le osservazioni x_t con $t = 1, \dots, T$ rientrano nella formula, dall'altro vengono utilizzate solo le frequenze con indice $j = 1, \dots, m$. Come si ci aspettava, i tempi crescono

m	T								
	2^9	2^{10}	2^{11}	2^{12}	2^{13}	2^{14}	2^{15}	2^{16}	2^{17}
2^7	0.043	0.086	0.156	0.296	0.593	1.077	2.199	5.008	10.35
2^8		0.160	0.310	0.608	1.200	2.387	4.509	10.17	20.61
2^9			0.624	1.230	2.420	4.800	9.550	20.39	43.38
2^{10}				2.470	4.920	9.720	19.80	41.44	83.79
2^{11}					9.920	19.70	39.60	83.50	169.5
2^{12}						39.83	84.28	168.9	348.0

TABELLA 4.5. Tempo espresso in secondi per calcolare il periodogramma di serie storiche di lunghezza T , variando il parametro T e m , utilizzando la definizione standard di trasformata di Fourier.

linearmente sia in T , che in m . Crescono però esponenzialmente secondo la legge mT nel caso in cui si decidesse di fissare una banda e analizzare serie frazionarie di lunghezza T crescente. In tabella 4.6 vengono mostrati i tempi relativi al calcolo del periodogramma utilizzando l'algoritmo FFT. L'algoritmo FFT calcola la trasformata a tutte le T frequenze, pertanto m non incide sul tempo. Incide tuttavia, anche se di poco, per il successivo calcolo del periodogramma (si guardi al codice nella sezione precedente). I tempi ottenuti sono decisamente ridotti rispetto

m	T								
	2^{14}	2^{15}	2^{16}	2^{17}	2^{18}	2^{19}	2^{20}	2^{21}	2^{22}
2^{12}	0.004	0.003	0.008	0.015	0.035	0.125	0.132	0.249	0.495
2^{13}		0.008	0.008	0.012	0.035	0.125	0.140	0.250	0.499
2^{14}			0.008	0.008	0.031	0.129	0.140	0.250	0.494
2^{15}				0.015	0.047	0.129	0.152	0.250	0.499
2^{16}					0.047	0.135	0.156	0.260	0.510
2^{17}						0.156	0.168	0.285	0.525
2^{18}							0.180	0.280	0.538
2^{19}								0.312	0.573
2^{20}									0.624

TABELLA 4.6. Tempo espresso in secondi per calcolare il periodogramma di serie storiche di lunghezza T , variando il parametro T e m , utilizzando l'algoritmo FFT.

all'algoritmo precedente; risulta difficile identificare la legge che governa tali numeri in quanto sono molto vicini a zero e soggetti ad alta variabilità, anche se si nota comunque l'incremento dei tempi sia variando T sia in misura minore variando m . Isolando la FFT dal resto dei calcoli si possono usare serie storiche più lunghe ed è più facile cogliere la legge che governa i tempi: in tabella (4.7) si mostrano i tempi computazionali del solo algoritmo FFT. Tali tempi sembrano ben adattarsi al fit

T								
2^{17}	2^{18}	2^{19}	2^{20}	2^{21}	2^{22}	2^{23}	2^{24}	
0.016	0.023	0.054	0.125	0.234	0.473	1.248	2.683	

TABELLA 4.7. Tempo espresso in secondi per il calcolo della trasformata di Fourier mediante algoritmo FFT.

ottenuto usando come regressore $T \log_2 T$, come viene mostrato in figura (4.7).

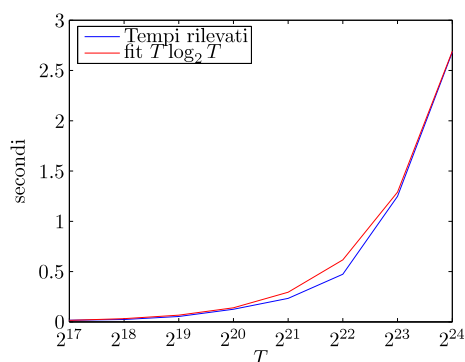


FIGURA 4.7. Tempi impiegati per il calcolo della FFT e relativo fit.

Infine si analizzano i tempi di stima del parametro d . Utilizzando la griglia di valori, si ottengono i tempi indicati in tabella (4.8). I tempi crescono linearmente in m , come da attese, e non sono influenzati dal parametro T . Infine utilizzando il

m	T									
	2^9	2^{10}	2^{11}	2^{12}	2^{13}	2^{14}	2^{15}	2^{16}	2^{17}	2^{18}
2^7	1.193	1.250	1.220	1.217	1.217	1.264	1.217	1.248	1.233	1.217
2^8		1.919	1.950	1.950	1.950	1.919	1.919	1.903	1.918	1.935
2^9			3.340	3.380	3.430	3.338	3.369	3.401	3.417	3.416
2^{10}				6.230	6.310	6.420	6.287	6.350	6.287	6.287
2^{11}					11.94	12.17	12.26	11.98	12.18	12.15
2^{12}						23.37	25.29	24.03	23.45	23.87
2^{13}							46.22	46.95	47.75	46.89
2^{14}								91.90	96.07	95.34

TABELLA 4.8. Tempo espresso in secondi per stimare d alla quarta cifra decimale utilizzando una griglia di valori (da 0 a 1 con passo 0.001).

m	T									
	2^{13}	2^{14}	2^{15}	2^{16}	2^{17}	2^{18}	2^{19}	2^{20}	2^{21}	2^{22}
2^{11}	0.016	0.016	0.015	0.016	0.016	0.016	0.016	0.015	0.016	0.015
2^{12}		0.027	0.030	0.023	0.027	0.031	0.031	0.032	0.031	0.031
2^{13}			0.063	0.060	0.055	0.055	0.062	0.047	0.047	0.062
2^{14}				0.102	0.109	0.109	0.109	0.109	0.109	0.109
2^{15}					0.218	0.210	0.218	0.211	0.219	0.203
2^{16}						0.428	0.437	0.445	0.450	0.453
2^{17}							0.878	0.880	0.905	0.904
2^{18}								1.763	1.778	1.810
2^{19}									3.520	3.570
2^{20}										7.020

TABELLA 4.9. Tempo espresso in secondi per stimare d utilizzando il metodo di Newton: i parametri sono $\Upsilon = 4$ e $\epsilon = 10^{-0.5}$.

metodo di Newton si ottengono i risultati espressi in tabella (4.9). Anche in questo caso la crescita dei tempi è lineare secondo m , mentre non è influenzata da T .

Applicazione: analisi univariata

1. Rendimenti e order flow

In questa sezione e nella prossima si analizzeranno delle serie storiche finanziarie con le metodologie viste nei capitoli precedenti. Si cercherà di stimare il parametro d relativo alle serie dei rendimenti di titoli azionari e alle serie degli order flow (flussi di ordini). I titoli azionari che verranno analizzati sono quotati sul Mercato Telematico Azionario della Borsa Italiana; i dati sono stati presi effettuando campionamenti ad alta frequenza (pochi secondi) delle serie in questione nell'intervallo temporale tra Novembre 2003 e Ottobre 2008 (ad esclusione di qualche periodo in cui i dati non erano disponibili).

I rendimenti sono stati determinati nel seguente modo:

- si è determinato il prezzo p_t^* al generico istante t come media aritmetica tra il miglior prezzo in acquisto (o anche Denaro o BID) e in vendita (o anche Lettera o ASK) nel book di negoziazione: a tale prezzo si ci riferisce con il termine di prezzo al mid-point;
- gli istanti t di campionamento sono equidistanti nel tempo; sono stati scelti i seguenti intervalli temporali Δt : 5 secondi, 20 secondi, 1 minuti, 5 minuti, 20 minuti, 1 ora, 4 ore e 1 giorno. Sono stati omessi dalle serie i prezzi determinati in sede di asta (di apertura o di chiusura);
- i rendimenti r_t sono stati ottenuti come differenza logaritmica dei prezzi: $r_t = \log p_t^* - \log p_{t-1}^*$. La nuova serie dei prezzi p_t è stata ottenuta cumulando quella dei rendimenti e facendone l'esponenziale.

Seguendo questo procedimento si ottengono i seguenti vantaggi:

- il fenomeno del bid-ask bounce è ridotto; tale fenomeno consiste nel fatto che, definendo il prezzo come il valore dell'ultima contrattazione, l'autocorrelazione ad un passo dei rendimenti corrispondenti appare negativa. Intuitivamente ciò può essere spiegato immaginando di avere un book i cui primi livelli siano stabili nel tempo e con una quantità sufficiente affinché non si esauriscano, il prezzo avrà un comportamento oscillatorio tra i due valori in acquisto e vendita (a prescindere da come si susseguano acquisti e vendite, se il prezzo si trova al valore in acquisto può solo rimanere stabile o salire, se si trova al valore in vendita può solo rimanere stabile o scendere). Ovviamente ciò non avviene con il prezzo scelto al mid-point, in quanto a book stabile può corrispondere soltanto un prezzo stabile;
- il fatto non considerare le aste di apertura e chiusura evita di inserire all'interno della serie dei rendimenti possibili elementi atipici, come variazioni di prezzo dovute ad esempio a stacco dei dividendi, stacco dei diritti. . . . Così facendo ovviamente la serie dei prezzi ricostruiti p_t differirà da quella dei prezzi osservati realmente sui mercati.

Per spiegare cosa sia la serie degli order flow, occorre premettere che ogni qualvolta avviene uno scambio sul mercato di Borsa è sempre presente una parte che acquista e una controparte che vende, senza prevalenze di acquisti o vendite: tuttavia se si considerano le compravendite generate da ordini eseguiti a mercato (o al meglio), si può pensare di separare le transazioni originate da acquisti a mercato da quelle originata da vendite a mercato. Attribuendo a queste transazioni rispettivamente il segno più e meno, e andndole a raggruppare all'interno di un dato intervallo di tempo, si ottiene la serie degli order flow (flusso di ordini), indicato con o_t . Con o_t si intende quindi la sommatoria degli ordini a mercato con segno sopra descritto nell'intervallo $(t - \Delta t, t)$. Per quanto riguarda i tempi di campionamento gli intervalli sono gli stessi visti per i rendimenti, e anche in questo caso non sono state considerate le aste.

Emerge subito una sostanziale differenza nelle due serie: prendendo come riferimento la letteratura che riguarda l'analisi di processi integrati in un contesto di aggregazione temporale (ad esempio di veda Hassler [8]), sembrerebbe che la serie dei rendimenti rientri nella categoria dello 'skip sampling' (campionamento ad intervalli predefiniti senza considerare ciò che avviene all'interno d un intervallo), mentre la serie degli order flow nella categoria del 'cumulating' (cumulare ciò che avviene all'interno di un intervallo). Immaginando di conoscere la serie completa sottostante, tra le due catogorie quella che presenta un comportamento (a livello di stime con le tecniche viste) più vicino a quello della serie di parteza è la seconda ('cumulating') mentre la prima presenta delle problematiche che si accentuano al crescere dell'intervallo di campionamento. In ogni caso nello specifico delle serie dei rendimenti va considerato che tipicamente (per i titoli visti) i livelli del book restano costanti anche per periodi abbastanza protratti (decine di secondi) e utilizzando un intervallo di campionamento piccolo (esempio $\Delta t = 5$ secondi) è pensabile che l'informazione persa sia poca, e che la serie sia più simile a quella completa piuttosto che ad un 'skip sampling' (in altre parole, molto spesso $r_t = 0$ e la maggior parte delle volte è perchè non è successo niente nell'intervallo, molto più raramente perchè due movimenti opposti si sono compensati). Considerando che le tecniche viste permettono di analizzare serie di lunghezza molto elevata, si procederà nell'analisi di serie con intervallo di campionamento $\Delta t = 5$ secondi ($T = 5802000$), così da sfruttare tutta l'informativa possibile ed evitare il fenomeno del 'skip samplng'.

Procediamo ora con un'analisi preliminare qualitativa delle serie che andremo ad analizzare con maggiore dettaglio nella prossima sezione. Mostriamo in figura (5.1) un estratto della serie dei rendimenti r_t riferita al titolo Enel, e la serie storica relativa ai prezzi p_t . Si è rappresentata solo una parte della serie dei rendimenti (12000 osservazioni consecutive, pari a 2 giornate di contrattazione) per riuscire a visualizzare meglio il comportamento della serie: si nota come già anticipato la presenza di molte osservazioni pari a 0. Le prime 6000 osservazioni sono antecedenti al cambio tick e le seconde 6000 successive: infatti si nota come i rendimenti possano assumere valori più piccoli nella seconda parte (in quanto il tick è più piccolo).

In figura (5.2) si mostra la funzione di autocorrelazione ρ_k della serie dei rendimenti, dove k indica il lag, e il periodogramma $I(\lambda)$, con $\lambda \in (0, \pi)$, sempre di r_t . Entrambi la funzione di autocorrelazione e il periodogramma presentano le caratteristiche qualitative tipiche che caratterizzano i processi integrati frazionari anti-persistenti, ovvero autocorrelazione negativa che tende iperbolicamente a 0 e

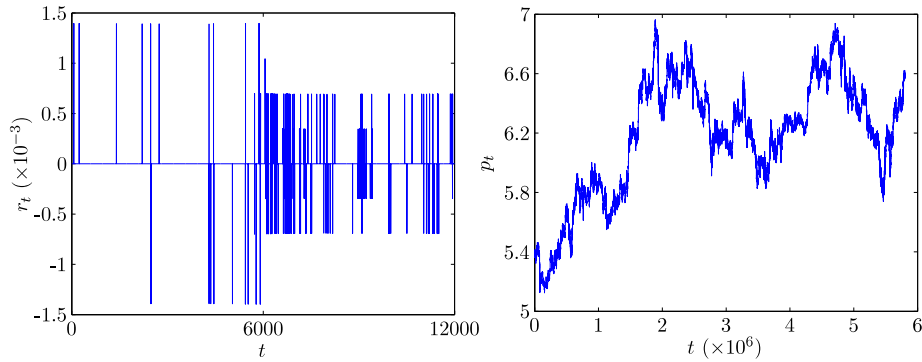


FIGURA 5.1. A sinistra: serie dei rendimenti r_t del titolo Enel (12000 osservazioni); a destra: serie dei prezzi p_t (5802000 osservazioni).

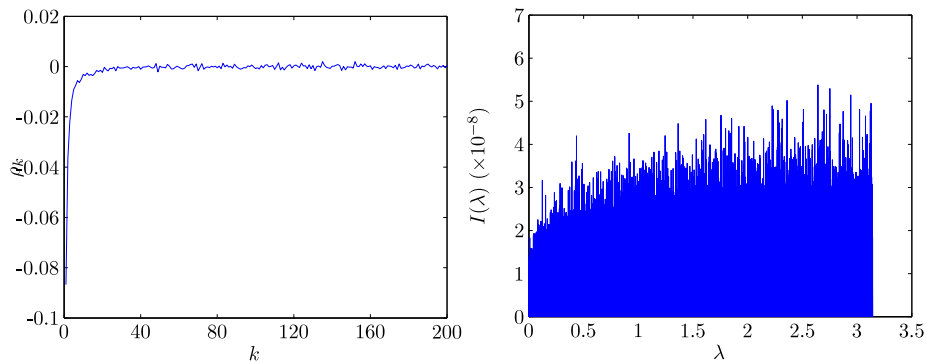


FIGURA 5.2. A sinistra: funzione di autocorrelazione ρ_k dei rendimenti; a destra: periodogramma $I(\lambda)$ dei rendimenti.

spettro con uno zero nell'origine (si veda il capitolo 1). Pertanto nelle stime che si effettueranno si ci aspetta, se diverso da zero, un valore di d negativo.

Passando alla serie degli order flow o_t , si mostra in figura (5.3) un estratto della serie o_t (12000 osservazioni, prese nello stesso intervallo descritto per r_t) e la serie completa cumulata. La scelta di rappresentare la cumulata è unicamente dovuta a fattori grafici (le osservazioni della serie apparirebbero indistinguibili, data la lunghezza della serie): le stime saranno effettuate sulla serie o_t . In questo caso il cambio del tick non produce effetti evidenti. In figura (5.4) si mostra la funzione di autocorrelazione ρ_k della serie degli order flow, dove k indica il lag, e il periodogramma $I(\lambda)$, con $\lambda \in (0, \pi)$, sempre di o_t . In questo caso la funzione di autocorrelazione e il periodogramma presentano qualitativamente le caratteristiche tipiche dei processi integrati frazionari persistenti, ovvero autocorrelazione positiva che tende iperbolicamente a 0 e spettro con un polo nell'origine. Si ci aspetta dunque di trovare, se diverso da zero, un valore di d positivo per la serie degli order flow.

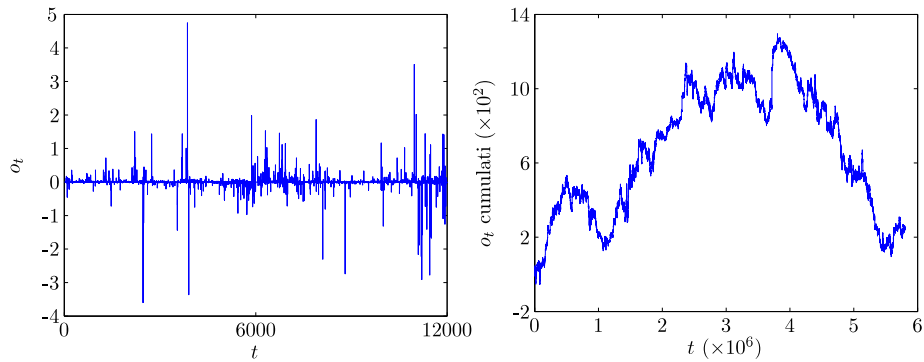


FIGURA 5.3. A sinistra: serie degli order flow o_t del titolo Enel (12000 osservazioni); a destra: serie degli order flow cummulate (5802000 osservazioni).

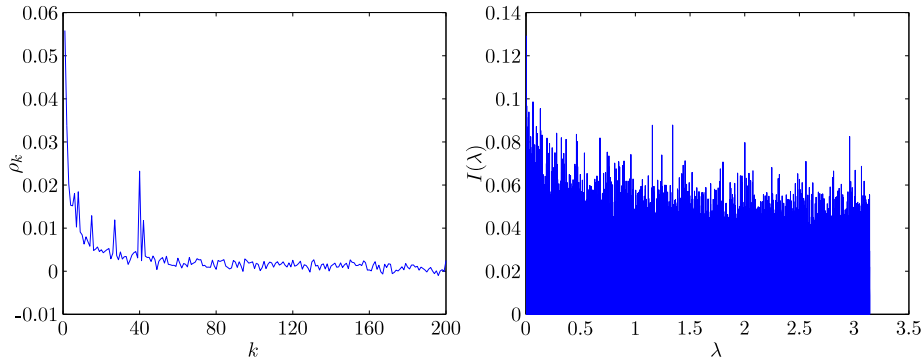


FIGURA 5.4. A sinistra: funzione di autocorrelazione ρ_k della serie o_t ; a destra: periodogramma $I(\lambda)$ di o_t .

2. Stime ottenute

Ora proviamo a condurre una serie di stime considerando:

- diverse tipologie di serie storiche: quelle relative al rendimento del titolo azionario Enel, e quelle relative all'order flow del medesimo titolo. Si cerca di verificare se effettivamente i rendimenti possano essere modellizzati con un processo integrato frazionario con $d < 0$ e se l'order flow da un processo con $d > 0$;
- frequenza di campionamento $\Delta t = 5$ secondi (lunghezza delle serie pari a $T = 5802000$);
- diversi sottocampioni all'interno della medesima serie storica, per eventualmente verificare se le stime si mantengano più o meno inalterate nel tempo: verrà proposta una stima riguardante l'intero campione, una riguardante la parte antecedente al cambio del tick e una riguardante la parte successiva, e ognuna di queste parti sarà poi ulteriormente scomposta in diversi sottocampioni di lunghezza comparabile;

- diverse frequenze all'interno dello stesso campione: in particolare le frequenze, per facilitare il confronto tra sottocampioni con lunghezza diversa, saranno sempre comprese nell'intervallo $0.01 \leq \lambda \leq 0.5$. All'interno di questo intervallo la scelta delle frequenze considerate segue una regola logaritmica, nel senso che saranno considerate tante frequenze 'basse' e poche frequenze 'alte', questo perchè: è più interessante sapere cosa accade a basse frequenze, le alte frequenze essendo più regolari non necessitano di numerose stime e l'onere computazionale per le stime ad alte frequenze è maggiore (m più grande).

Di seguito (figura (5.5)) vengono mostrate su dei grafici le stime ottenute: i dettagli relativi alla tipologia di serie, alla frequenza di campionamento dei dati e al sottoinsieme considerato sono riportati nella didascalia. L'andamento della stima al variare della frequenza è rappresentato da una linea blu. Nel caso di un unico campione oltre alla stima viene rappresentata la banda di confidenza al 95% utilizzando la deviazione standard teorica: tale banda è rappresentata da una linea punteggiata rossa. Nelle figure in cui sono rappresentate le stime sovrapposte di diversi sottocampioni, oltre alle stime stesse sempre in linea blu vengono rappresentate: la media delle stime tramite 'tondini' rossi, la banda di confidenza al 95% utilizzando la deviazione standard teorica sempre tramite una linea punteggiata rossa, la banda di confidenza al 95% utilizzando la deviazione standard campionaria (stimata a partire dai valori di \hat{d} ottenuti nei diversi sottocampioni) tramite una linea tratteggiata rossa.

Analizzando i risultati per le serie dei rendimenti, la prima cosa che si nota è la netta differenza di valori assunti dalle stime al variare delle frequenze considerate: si può supporre che il comportamento del processo ad alte frequenze abbia una certa influenza sulle stime. Si nota inoltre laddove vengano considerati diversi sottocampioni una varianza degli stessi generalmente molto più alta di quella che avrebbero dovuto avere teoricamente: possibili spiegazioni di questo fenomeno possono essere dati particolarmente anomali che influiscono sulle stime, o un cambiamento del valore dei parametri del processo nel corso del tempo.

Teoricamente più si considerano frequenze basse e più si evitano distorsioni nella stima dovute dal comportamento del processo ad alte frequenze: conviene quindi scegliere un valore di λ vicino a 0, varianza permettendo. In queste applicazioni la grossa mole di dati permette di poter considerare frequenze molto basse, come ad esempio $\lambda = 0.05$ (la scelta verrà illustrata in seguito). In tabella 5.1 si elencano le stime ottenute alla frequenza $\lambda = 0.05$, con relativa deviazione standard teorica. Le stime relative alla serie intera, al sottocampione antecedente al cambio

	T	#	\hat{d}	σ_d
Serie intera	5802000	1	-0.0229	0.0023
Serie ante cambio tick	1848000	1	-0.0393	0.00408
Serie post cambio tick	3954000	1	-0.0149	0.00279

TABELLA 5.1. Stime di d della serie r_t relativa al titolo Enel alla frequenza $\lambda = 0.05$, considerando # sottocampioni di lunghezza T .

del tick e a quello successivo, appaiono tutte negative e significativamente diverse da $d = 0$. Tuttavia riprendendo la figura (5.5) si nota come in tutti i casi considerati la stima tende a $d = 0$ per frequenze inferiori a quella considerata, facendo

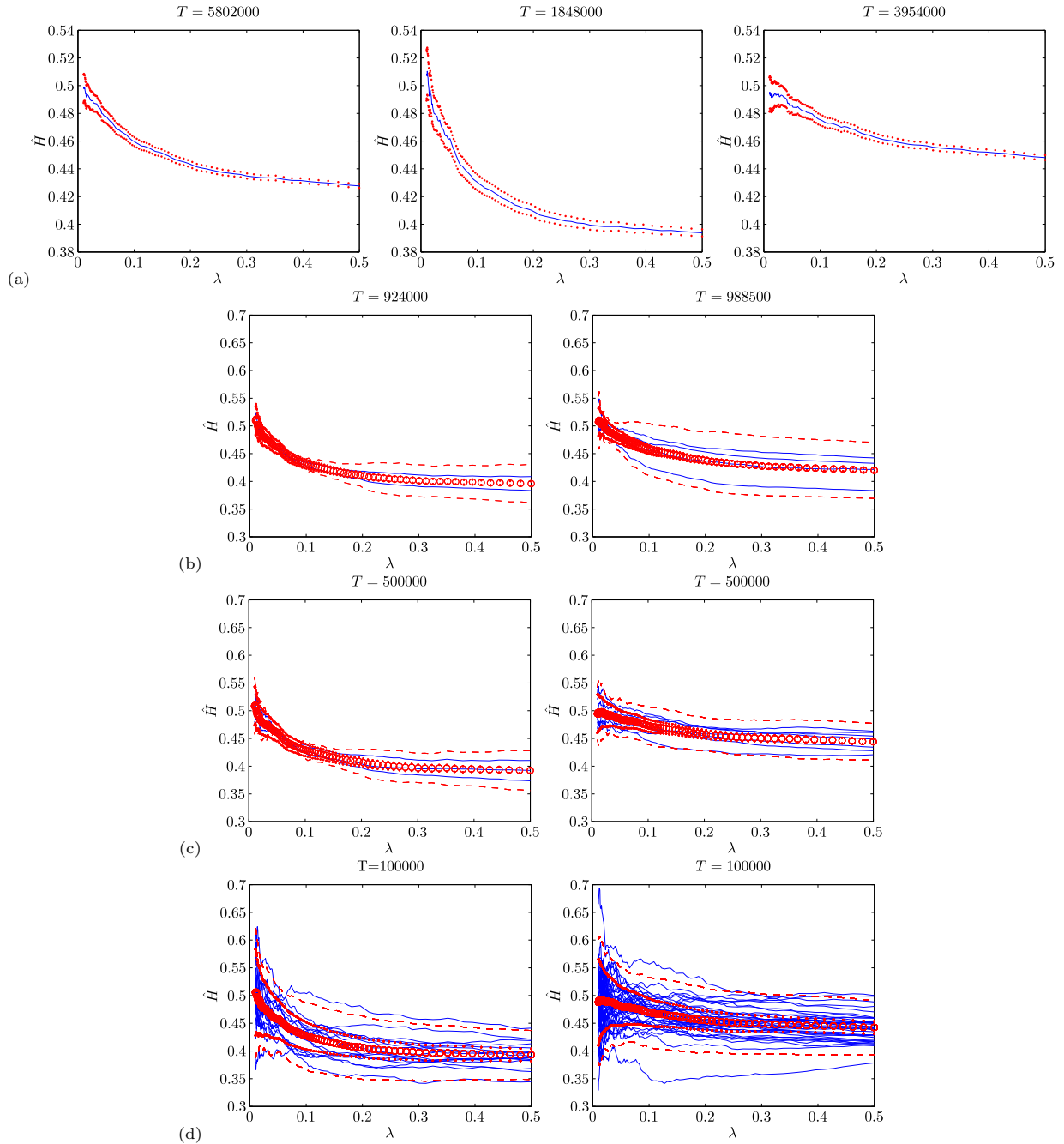


FIGURA 5.5. Stime di H della serie r_t relative al titolo Enel considerando (a) l'intera serie (sinistra), il sottocampione antecedente al cambio tick (centro) e successivo (destra); (b) 2 sottocampioni antecedenti (sinistra) e 4 successivi (destra); (c) 3 sottocampioni antecedenti (sinistra) e 7 successivi (destra); (d) 18 sottocampioni antecedenti (sinistra) e 39 successivi (destra). La dimensione del campione è espressa in figura.

emergere il sospetto in realtà il processo analizzato è caratterizzato da $d = 0$ e che comportamenti ad alte frequenze (ad esempio componenti MA) riescano ad influenzare le stime fino alla frequenza considerata $\lambda = 0.05$. Ciò sarà analizzato con maggior dettaglio nella prossima sezione. Le stime ottenute prima del cambio del tick appaiono diverse rispetto a quelle ottenute dopo: c'è il sospetto che il cambio del tick abbia portato ad una modifica del comportamento della serie r_t . Di seguito verranno analizzati i rendimenti di altri titoli caratterizzati da un diverso impatto ddl cambio del tick, allo scopo di verificare se effettivamente questo fenomeno si riscontra anche in altre situazioni.

Analizzando i risultati ottenuti per la serie o_t (figura (5.6)), le stime sembrano essere più stabili al variare di λ rispetto alle stime dei rendimenti. Anche in questo caso la variabilità sembra essere decisamente maggiore rispetto alla teoria. In tabella (5.2) si riportano le stime ottenute considerando il campione totale, la parte antecedente al cambio tick e quella successiva. Anche in questo caso le stime \hat{d} ap-

	T	\hat{d}	σ_d
Serie intera	5802000	0.0767	0.0023
Serie ante cambio tick	1848000	0.0701	0.0041
Serie post cambio tick	3954000	0.0805	0.0028

TABELLA 5.2. Stime di d della serie o_t relativa al titolo Enel alla frequenza $\lambda = 0.05$.

paiono significativamente diverse da $d = 0$: inoltre rimangono significativamente diverse da $d = 0$ anche per frequenze più basse di $\lambda = 0.05$ (figura (5.6)). Si può concludere che la serie o_t è senza dubbio un processo integrato frazionario, caratterizzato da parametro d indicato in tabella (5.2). Infine si osserva che la stima antecedente al cambio del tick e quella successiva non appaiono molto differenti tra di loro (o comunque non tanto quanto nel caso dei rendimenti): l'impatto del cambio del tick sembra dunque essere trascurabile nel caso degli order flow.

Di seguito (figura (5.7), (5.8) e (5.9)) si mostrano le stime \hat{d} relative alla serie r_t del titolo Eni, Fiat e Snam Rete Gas. In tabella (5.3), (5.4) e (5.5) si riportano i valori relativi alla frequenza $\lambda = 0.05$. Anche in questo caso le stime sono

	T	\hat{d}	σ_d
Serie intera	5802000	-0.0045	0.0023
Serie ante cambio tick	1848000	-0.0045	0.0041
Serie post cambio tick	3954000	-0.0046	0.0028

TABELLA 5.3. Stime di d della serie r_t relativa al titolo Eni alla frequenza $\lambda = 0.05$.

molto vicine a $d = 0$ (in alcuni casi non sono significative) e tendono a convergere a $d = 0$ per valori di $\lambda < 0.05$, confermando i timori già mostrati circa una possibile influenza del comportamento ad alte frequenze sulle stime. L'unico caso che mostra indubbiamente $d \neq 0$ si ha per il titolo Snam Rete Gas, nel campione antecedente al cambio del tick, dove la serie in questione è un processo anti-persistente. Analizzando l'effetto del cambio del tick sulle serie, si nota per i titoli analizzati un generale avvicinamento al valore $d = 0$ dopo il cambio del tick. La tabella

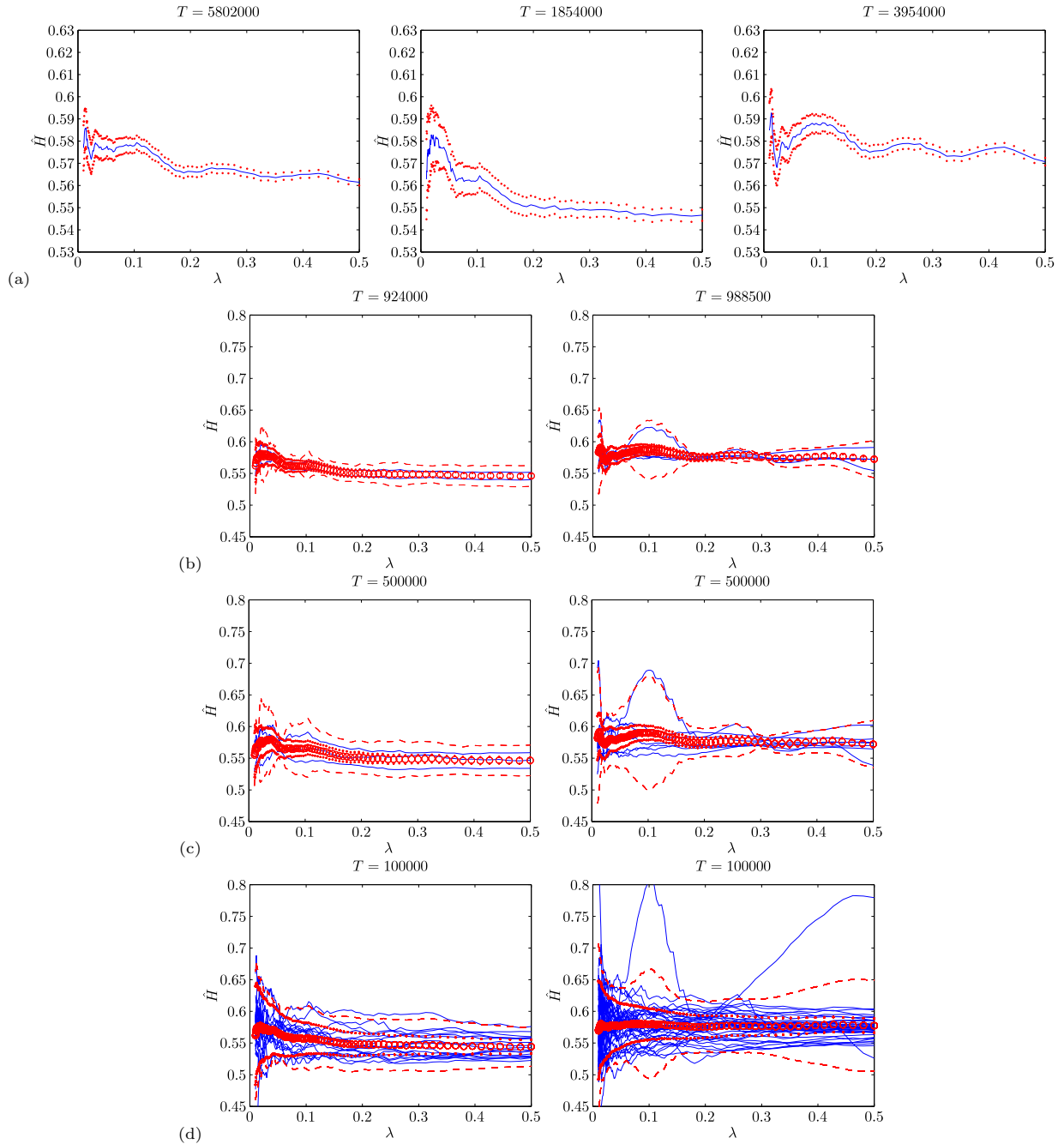


FIGURA 5.6. Stime di H della serie o_t relativa al titolo Enel considerando (a) l'intera serie (sinistra), il sottocampione antecedente al cambio tick (centro) e successivo (destra); (b) 2 sottocampioni antecedenti (sinistra) e 4 successivi (destra); (c) 3 sottocampioni antecedenti (sinistra) e 7 successivi (destra); (d) 18 sottocampioni antecedenti (sinistra) e 39 successivi (destra). La dimensione del campione è espressa in figura.

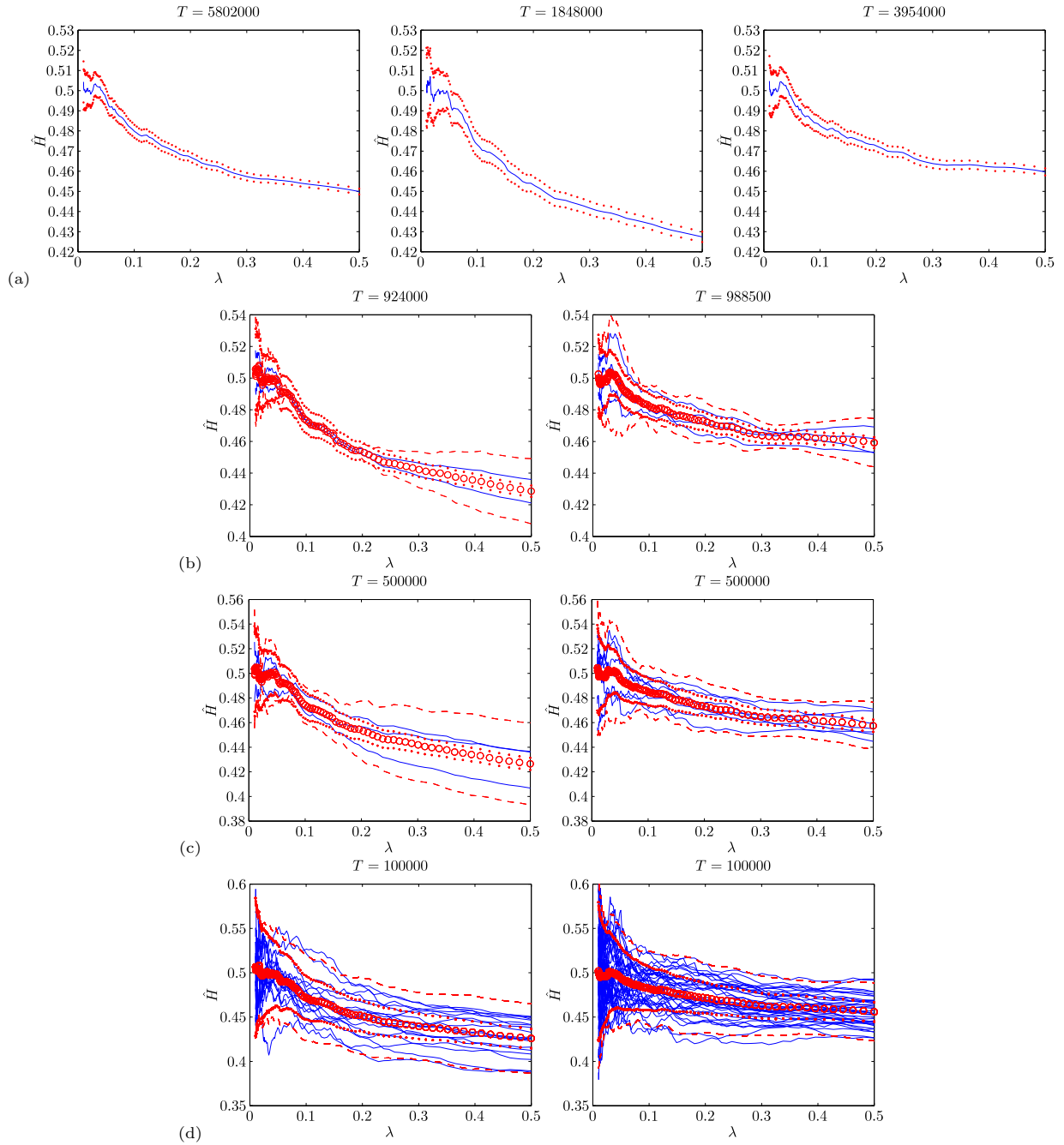


FIGURA 5.7. Stime di H della serie r_t relativa al titolo Eni, considerando (a) l'intera serie (sinistra), il sottocampione antecedente al cambio tick (centro) e successivo (destra); (b) 2 sottocampioni antecedenti (sinistra) e 4 successivi (destra); (c) 3 sottocampioni antecedenti (sinistra) e 7 successivi (destra); (d) 18 sottocampioni antecedenti (sinistra) e 39 successivi (destra). La dimensione del campione è espressa in figura.

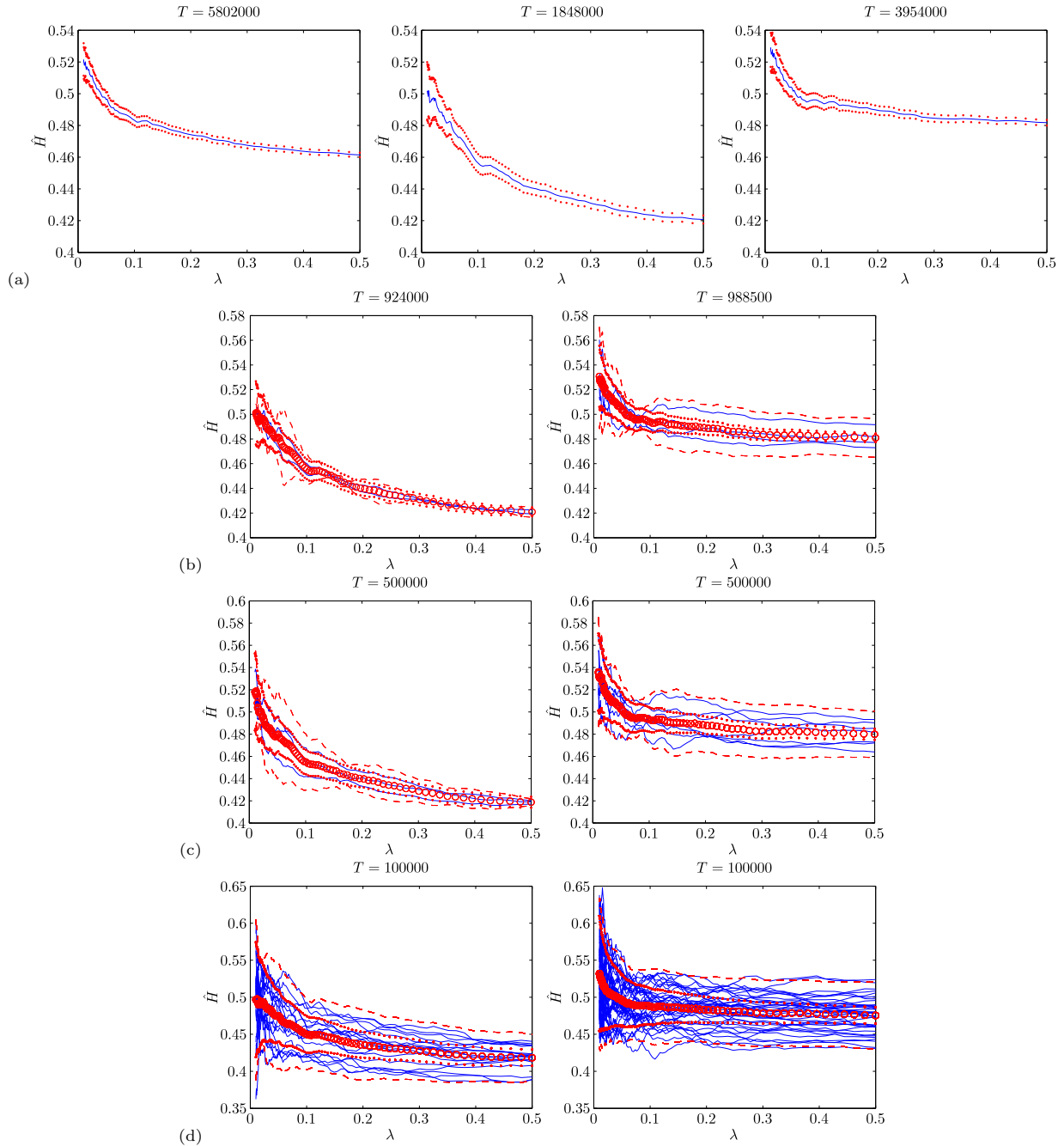


FIGURA 5.8. Stime di H della serie r_t relativa al titolo Fiat, considerando (a) l'intera serie (sinistra), il sottocampione antecedente al cambio tick (centro) e successivo (destra); (b) 2 sottocampioni antecedenti (sinistra) e 4 successivi (destra); (c) 3 sottocampioni antecedenti (sinistra) e 7 successivi (destra); (d) 18 sottocampioni antecedenti (sinistra) e 39 successivi (destra). La dimensione del campione è espressa in figura.

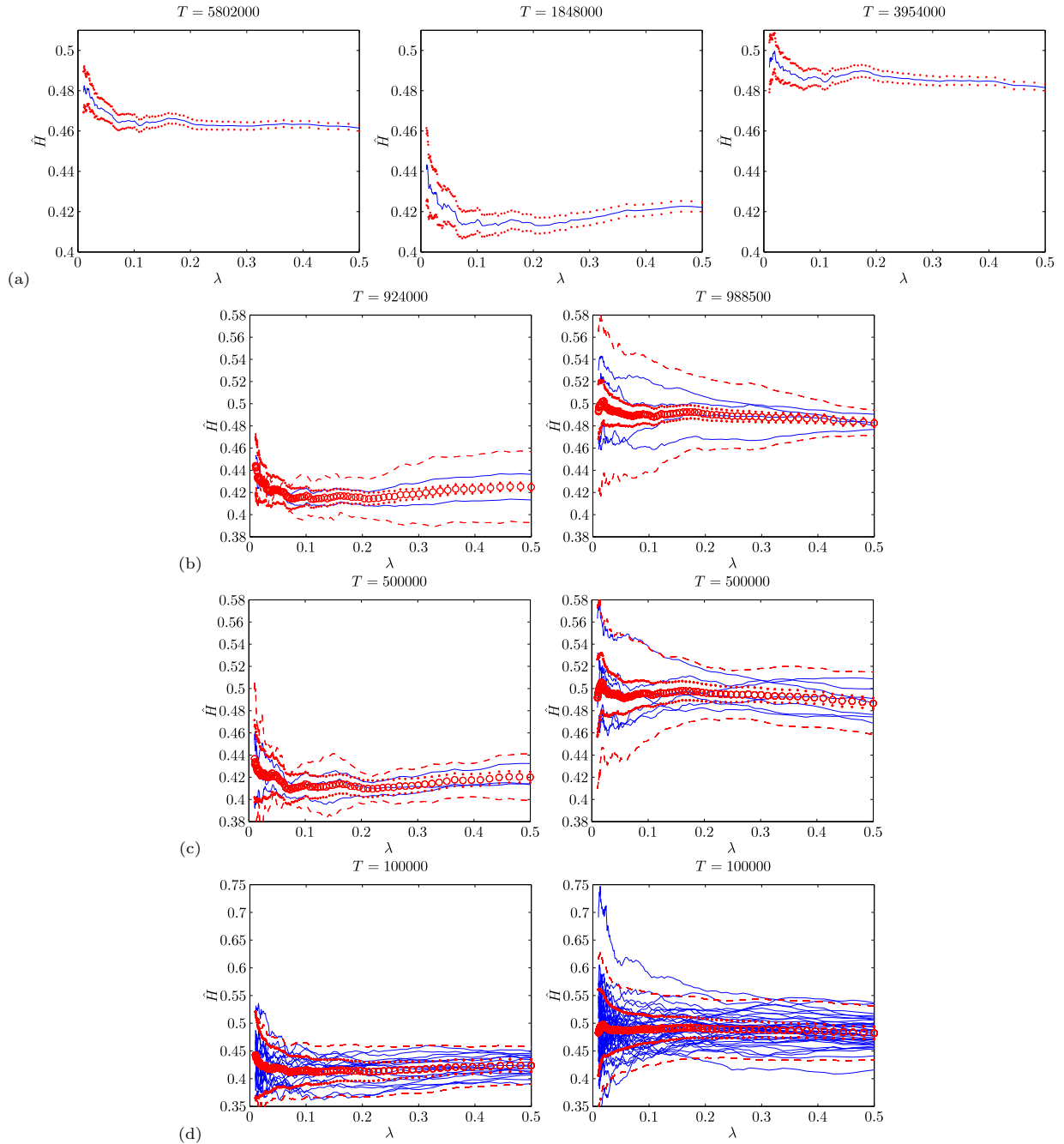


FIGURA 5.9. Stime di H della serie r_t riferita al titolo Snam Rete Gas, considerando (a) l'intera serie (sinistra), il sottocampione antecedente al cambio tick (centro) e successivo (destra); (b) 2 sottocampioni antecedenti (sinistra) e 4 successivi (destra); (c) 3 sottocampioni antecedenti (sinistra) e 7 successivi (destra); (d) 18 sottocampioni antecedenti (sinistra) e 39 successivi (destra). La dimensione del campione è espressa in figura.

	T	\hat{d}	σ_d
Serie intera	5802000	-0.0040	0.0023
Serie ante cambio tick	1848000	-0.0192	0.0041
Serie post cambio tick	3954000	0.0019	0.0028

TABELLA 5.4. Stime di d della serie r_t relativa al titolo Fiat alla frequenza $\lambda = 0.05$.

	T	\hat{d}	σ_d
Serie intera	5802000	-0.0304	0.0023
Serie ante cambio tick	1848000	-0.0781	0.0041
Serie post cambio tick	3954000	-0.0116	0.0028

TABELLA 5.5. Stime di d della serie r_t relativa al titolo Snam Rete Gas alla frequenza $\lambda = 0.05$.

(5.6) vuole mostrare una possibile evidenza del fatto che titoli interessati da un cambiamento del tick più marcato (in proporzione al proprio prezzo) hanno successivamente mostrato delle differenze nelle stime più marcate. In tabella è stato

Titolo	$\bar{p}_t(1)$	Tick(1)	%(1)	$\bar{p}_t(2)$	Tick(2)	%(2)	%(2)/%(1)	Δ_d
Eni	14.880	0.01	0.067%	15.028	0.01	0.067%	1	-0.0001
Enel	5.7753	0.01	0.173%	6.3614	0.005	0.078%	0.45	0.0244
Fiat	4.8366	0.01	0.207%	3.7832	0.0025	0.066%	0.32	0.0211
Snam	3.4659	0.01	0.289%	3.7574	0.0025	0.067%	0.23	0.0665

TABELLA 5.6. Stime di d della serie r_t relativa al titolo Snam Rete Gas alla frequenza $\lambda = 0.05$.

indicato il prezzo medio del titolo antecedente al cambio del tick, il valore del tick e l'incidenza percentuale del tick sul prezzo medio (tick/prezzo medio), indicati con (1); prezzo medio, tick e incidenza successivi al cambio del tick sono indicati con (2). Il tick e l'incidenza nel titolo Eni non sono cambiati, ed effettivamente è il titolo le cui stime \hat{d} sono cambiate meno nelle due parti del campione; per contro, l'incidenza nel titolo Snam Rete Gas dopo il cambio del tick è diventata un quarto rispetto a prima, e il titolo ha mostrato una maggiore differenza nelle stime \hat{d} .

3. Esercizi di simulazione

In questa sezione si andranno a fare delle simulazioni volte a chiarire alcuni elementi di incertezza circa il comportamento delle stime ottenute. In particolare il dubbio principale che si cercherà di chiarire è se sia possibile o meno che il comportamento di un processo ad alta frequenza (ad esempio componenti MA di ordine anche elevato ma comunque a memoria finita) possa condizionare le stime a frequenze basse come quelle considerate ($\lambda = 0.05$). L'importanza di comprendere questo aspetto è dovuta al fatto che se queste componenti hanno una certa influenza all'interno della banda $\lambda \in (0.01, 0.5)$ considerata nelle stime, allora è meglio scegliere frequenze più basse possibili, ottenendo valori di d pressochè pari a zero (assenza di long-memory o anti-persistenza); se invece le componenti ad alta

frequenza non influenzano questa banda, allora una scelta di λ più elevata può portare a stime più precise (minore varianza) e a valori di \hat{d} decisamente minori di $d = 0$ (presenza di anti-persistenza).

3.1. Primo esercizio. Si considerino i dati relativi ai rendimenti r_t del titolo Enel prima del cambio del tick (dati le cui stime \hat{d} sono riportate in figura 5.5a in centro). Di seguito si prova ad eseguire l'esercizio definito dai seguenti passi:

- calcolare la funzione di autocorrelazione di r_t e stimare i primi $q = 5, 10$ coefficienti θ_k del modello

$$r_t = \sum_{k=1}^q \theta_k e_{t-k} + e_t.$$

Come si vedrà, tali coefficienti sono tutti significativi e negativi;

- generare un processo white noise x_t di lunghezza $T = 1848000$ e filtrarlo con i coefficienti stimati, ottenendo un processo MA(5) che chiamiamo y_t^1 e un processo MA(10) y_t^2 ;
- stimare il parametro d dei processi y_t^1 e y_t^2 considerando la larghezza di banda prima utilizzata. A questo punto si può analizzare qualitativamente il comportamento delle stime e vedere se somigliano o meno a quelle trovata per r_t .

Di seguito riportiamo la funzione di autocorrelazione di r_t :

LAG	ACF	LAG	ACF
1	-0,0739 ***	11	-0,0064 ***
2	-0,0389 ***	12	-0,0064 ***
3	-0,0315 ***	13	-0,0031 ***
4	-0,0241 ***	14	-0,0050 ***
5	-0,0158 ***	15	-0,0030 ***
6	-0,0128 ***	16	-0,0030 ***
7	-0,0087 ***	17	-0,0056 ***
8	-0,0095 ***	18	-0,0033 ***
9	-0,0094 ***	19	-0,0033 ***
10	-0,0058 ***	20	-0,0027 ***.

I coefficienti sono tutti negativi e significativi, e la modellizzazione con processo MA(5) e MA(10) è giustificata, anche se incompleta: ordini più elevati non sono stati considerati per i tempi di calcolo troppo lunghi. Di seguito riportiamo le stime dei coefficienti θ_k^1 della modellizzazione MA(5):

VARIABILE	COEFFICIENTE	ERRORE STD	STAT T	P-VALUE
theta_1	-0,0822104	0,000735437	-111,78	<0,00001 ***
theta_2	-0,0457037	0,000737598	-61,963	<0,00001 ***
theta_3	-0,0374837	0,000737849	-50,801	<0,00001 ***
theta_4	-0,0295973	0,000737598	-40,127	<0,00001 ***
theta_5	-0,0218322	0,000735437	-29,686	<0,00001 ***.

Ora riportiamo le stime dei coefficienti θ_k^2 della modellizzazione MA(10):

VARIABILE	COEFFICIENTE	ERRORE STD	STAT T	P-VALUE
theta_1	-0,0827859	0,000735577	-112,55	<0,00001 ***
theta_2	-0,0458470	0,000738033	-62,121	<0,00001 ***
theta_3	-0,0370508	0,000738741	-50,154	<0,00001 ***
theta_4	-0,0284984	0,000739190	-38,553	<0,00001 ***
theta_5	-0,0196378	0,000739390	-26,560	<0,00001 ***
theta_6	-0,0162493	0,000739390	-21,977	<0,00001 ***
theta_7	-0,0121393	0,000739190	-16,422	<0,00001 ***
theta_8	-0,0129956	0,000738741	-17,592	<0,00001 ***
theta_9	-0,0128683	0,000738033	-17,436	<0,00001 ***
theta_10	-0,00981261	0,000735577	-13,340	<0,00001 ***.

Ora si genera il processo white noise x_t e lo si filtra con i 5 coefficienti sopra stimati ottenendo il processo y_t^1 , che presenta i seguenti valori di ACF (fino al sesto lag) e stime dei coefficienti $\hat{\theta}_k^1$ (fino al quinto ordine):

LAG	ACF	VARIABILE	COEFFICIENTE	ERRORE STD	STAT T	P-VALUE
1	-0,0735 ***	theta_1	-0,0817089	0,000735437	-111,10	<0,00001 ***
2	-0,0407 ***	theta_2	-0,0464184	0,000737555	-62,936	<0,00001 ***
3	-0,0338 ***	theta_3	-0,0376532	0,000737825	-51,033	<0,00001 ***
4	-0,0281 ***	theta_4	-0,0302029	0,000737555	-40,950	<0,00001 ***
5	-0,0217 ***	theta_5	-0,0219181	0,000735438	-29,803	<0,00001 ***
6	0,0002					

I primi cinque coefficienti di autocorrelazione sono molto simili ai primi cinque del processo r_t , mentre i coefficienti $\hat{\theta}_k^1$ stimati sono pressochè identici a quelli veri θ_k^1 . Analoga cosa viene fatta prendendo le stime dei 10 coefficienti θ_k^2 e generando il processo y_t^2 (non riportiamo ACF e stime $\hat{\theta}_k^2$).

Mostriamo in figura (5.10) le stime \hat{d}_1 e \hat{d}_2 calcolate sui processi descritti: entrambe le stime forniscono il valore corretto d solo per frequenze molto basse (fino a $\lambda = 0.1$ e $\lambda = 0.05$ circa rispettivamente per i processi y_t^1 e y_t^2). Prendendo una frequenza λ più elevata tra quelle illustrate (che sono comunque frequenze basse se si considera l'intero dominio $(0, \pi)$) si possono avere stime molto distorte, nonostante la numerosità del campione e la bassa varianza delle stime: sembrerebbe che processi a memoria breve possano influenzare la regione delle basse frequenze, ma non quelle degeneranti verso $\lambda = 0$. Confrontando le stime \hat{d}_1 e \hat{d}_2 ottenute in figura con le stime trovate per r_t (figura (5.5)(a) in centro) si nota una certa somiglianza tra \hat{H}_r e \hat{H}_{y_2} . In particolare si nota che queste due stime prese in $\lambda = 0.5$ hanno un valore molto simile (nell'intorno di $d = -0.1$): si potrebbe affermare che questo valore è determinato principalmente dal comportamento a memoria breve del processo. Tuttavia le stime di r_t si allontanano da $d = 0$ molto più velocemente di quanto non lo facciano le stime \hat{d}_2 : la modellizzazione MA(10) è pertanto insufficiente a descrivere il processo (si ricorda l'ACF significativo fino al 20-esimo lag e oltre).

Se però il processo fosse 'long memory' come da simulazione, la stima non dovrebbe tendere verso $H = 0.5$ per $\lambda \rightarrow 0$, ma appiattirsi verso il valore H vero

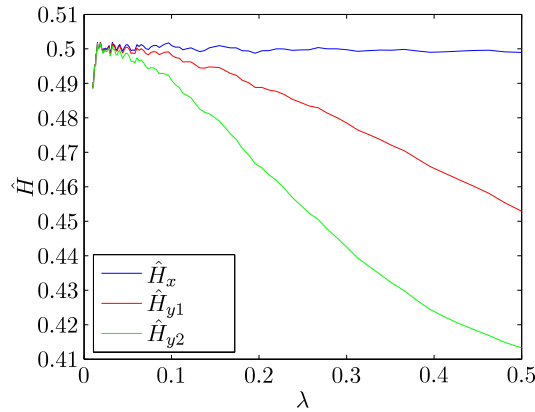


FIGURA 5.10. Stime di H relative al processo white noise x_t (in blu), al processo MA(5) y_t^1 (in rosso) e al processo MA(10) y_t^2 (in verde).

(come si potrà vedere in figura ?? in cui sono simulati processi caratterizzati da $H = 0.4$). Considerare il processo r_t come caratterizzato da $H = 0.5$ presenta la problematica di dover evidentemente considerare un modello MA(q) di ordine molto elevato (di ordine circa 30) per descrivere il processo, il che solitamente non è desiderabile, oltre ad essere oneroso computazionalmente. Una soluzione potrebbe essere la seguente: considerare la stima \hat{H}_r valida per $\lambda = 0.05$. Empiricamente si è visto come un processo MA(10) non giunga a modificare vistosamente questa frequenza; si può supporre inoltre che usare tale stima e filtrare il processo r_t con $(1 - L)^{-(\hat{H}_r - 0.5)}$ porti ad un processo che non è un white noise ma presenti ancora qualche coefficiente di autocorrelazione significativo tra i primi (da modellizzare successivamente tramite MA(q) di pochi ordini). D'altra parte se si considerasse $\hat{H}_r = 0.4$ circa probabilmente si riesce ad avere un white noise (nessun coefficiente significativo) perchè tale valore di \hat{H}_r si adatta bene ai coefficienti dell'ACF per lag vicini (ma magari non per lag lontani, anche se si potrebbe non accorgersene), mentre il valore $\hat{H}_r = 0.5$ sicuramente non si adatta bene all'ACF, necessitando come visto di una modellizzazione MA(q) di ordine molto elevato. Considerando la frequenza sopra riportata si ha la stima $\hat{H}_r = 0.4607$ (ottenuta mediando 6 valori nell'intorno), con deviazione standard $\sigma_H = 0.0041$.

4. Analisi di sottocampioni anomali

Come visto nei grafici mostranti le stime, in particolare nel caso di numerosi (e brevi) sottocampioni, spesso la variabilità di questi è molto elevata. Affianco a sottocampioni che presentano valori di \hat{d} per la frequenza considerata $\lambda = 0.05$ ben al di sotto di $d = 0$, ne seguono altri che presentano valori di \hat{d} significativamente al di sopra di $d = 0$. Di seguito proviamo ad isolare e analizzare questi sottocampioni. Prendiamo le stime mostrate in figura 5.5 (d) a destra (39 sottocampioni con $T = 100000$ presi dopo il cambio tick): il primo sottocampione che analizziamo è quello che presenta le stime di H più elevate (che risulta essere il primo). In figura (5.11) mostriamo nell'ordine: le stime di d (questa volta considerando una banda più ampia $\lambda \in (0, 1.5)$), la cumulata dei rendimenti, il periodogramma, la cumulata dei primi 200 coefficienti di autocorrelazione, tutti relativi al sottocampione r_t

considerato. Si può notare come le stime di d siano inizialmente maggiori di $d = 0$

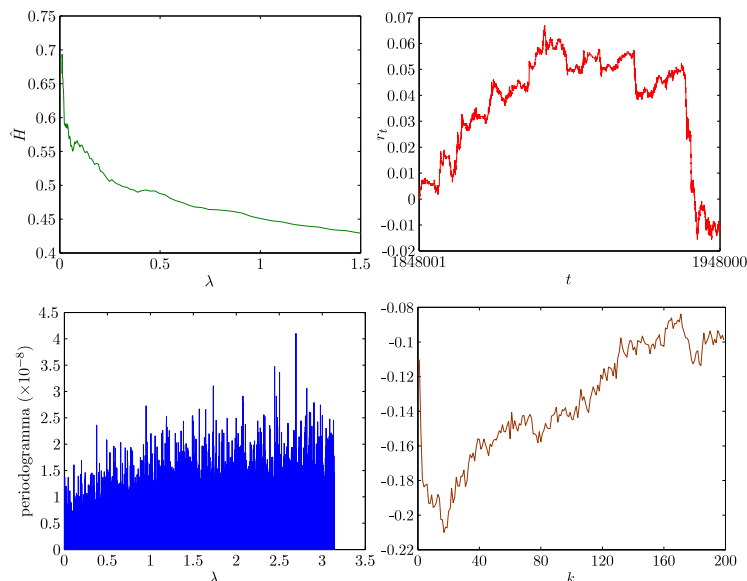


FIGURA 5.11. Sottocampione 1. Nell'ordine (sinistra-destra alto-basso): stima di H al variare di λ , cumulata dei rendimenti r_t , periodogramma di r_t , cumulata dei primi 200 coefficienti dell'ACF di r_t .

per poi diventare minori: i coefficienti d'altra parte sono tendenzialmente negativi all'inizio, per poi (a partire dal 20-esimo coefficiente circa) essere prevalentemente di segno positivo. In figura (5.12) si analizza il 24-esimo sottocampione, caratterizzato da stime di d molto basse. In questo campione le stime sono costantemente e nettamente sotto il valore $d = 0$, per tutte le frequenze: i coefficienti dell'ACF di r_t tendono ad essere più negativi che positivi anche per lag lontani, e il processo r_t sembra essere senza dubbio un processo antipersistente. Ciò che si nota immediatamente è che in questo sottocampione r_t risulta molto meno volatile rispetto al sottocampione precedente: potrebbe esserci una relazione tra volatilità del processo e parametro d , che verrà analizzata in seguito.

Si passa ora ad analizzare i sottocampioni più particolari riguardanti la parte antecedente al cambio tick. Si mostra il secondo sottocampione, caratterizzato da stime di d basse (figura (5.13)) e il quinto sottocampione (\hat{d} alte, figura (5.14)). Si può notare come il secondo sottocampione (con stime basse) sia anche in questo caso il meno volatile; il campioni con stime \hat{d} più elevate appare più volatile.

Per il momento si è parlato di volatilità riferendosi al comportamento qualitativo che si vede in figura; verranno proposte come definizione di volatilità il range assunto dai prezzi e la deviazione standard dei rendimenti. Di seguito si è provato a testare se esiste o meno una relazione tra il range assunto dai prezzi in un intervallo e il corrispondente valore assunto da \hat{d} . Il valore di \hat{d} (uno per ogni sottocampione) è stato ottenuto mediando i 6 valori di \hat{d} assunti nell'intorno della frequenza $\lambda = 0.05$. Sono state presi inizialmente 39 sottocampioni di lunghezza $T = 100000$, ed è stata eseguita una regressione in cui la variabile dipendente è \hat{d} e quella esplicativa è il range. Il range è stato preventivamente trasformato per

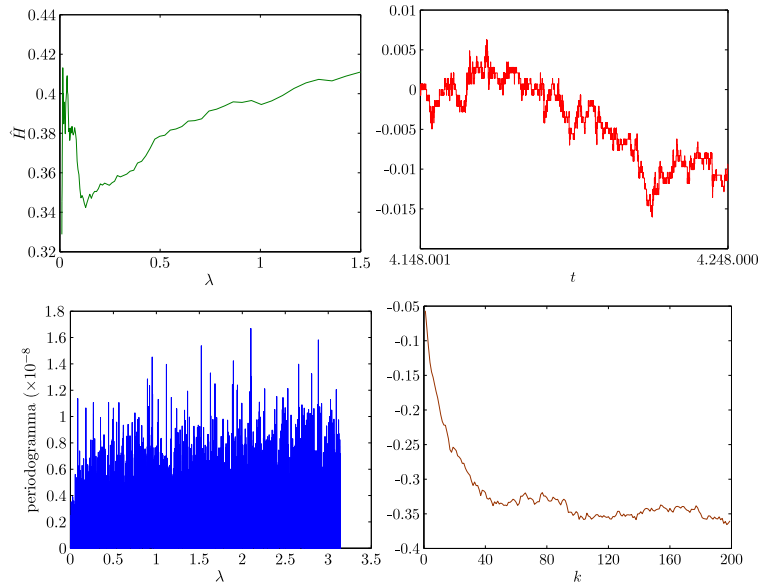


FIGURA 5.12. Sottocampione 24. Nell'ordine (sinistra-destra alto-basso): stima di H al variare di λ , prezzi (cumulata dei rendimenti r_t), periodogramma di r_t , cumulata dei primi 200 coefficienti dell'ACF di r_t .

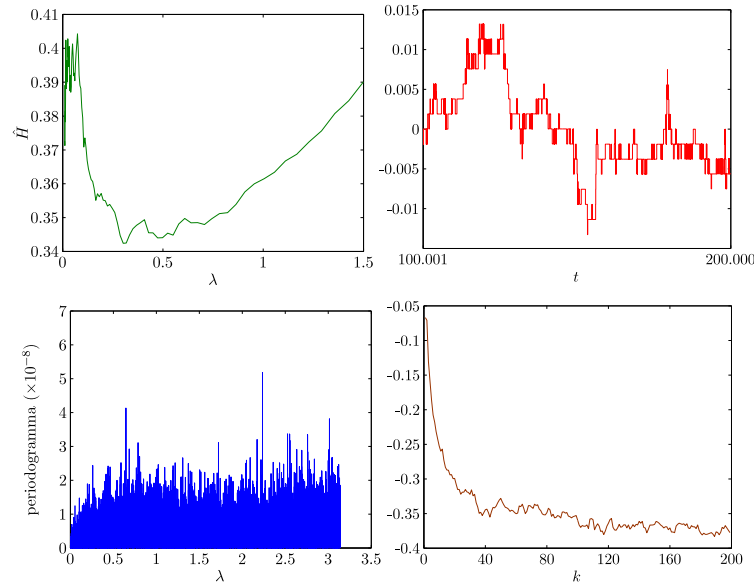


FIGURA 5.13. Sottocampione 2. Nell'ordine (sinistra-destra alto-basso): stima di H al variare di λ , prezzi (cumulata dei rendimenti r_t), periodogramma di r_t , cumulata dei primi 200 coefficienti dell'ACF di r_t .

non violare l'ipotesi di normalità della serie (trasformazione logaritmica). Dopo

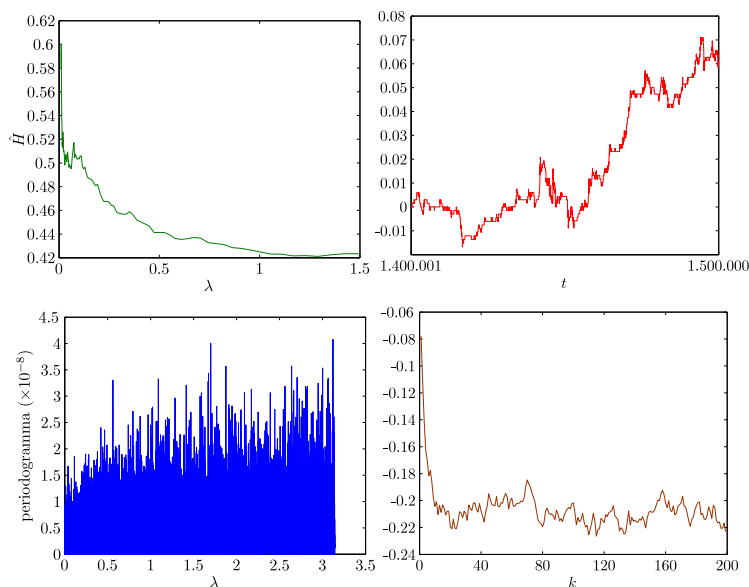


FIGURA 5.14. Sottocampione 5. Nell'ordine (sinistra-destra alto-basso): stima di H al variare di λ , prezzi (cumulata dei rendimenti r_t), periodogramma di r_t , cumulata dei primi 200 coefficienti dell'ACF di r_t .

aver verificato che i residui siano normali e con autocorrelazione nulla, riportiamo il modello che ne è risultato:

The regression equation is
 $H = 0,634 + 0,0500 \log_Range$

Predictor	Coef	SE Coef	T	P
Constant	0,63446	0,03898	16,28	0,000
log_Range	0,04996	0,01255	3,98	0,000

$S = 0,0296361$ $R\text{-Sq} = 30,0\%$ $R\text{-Sq}(\text{adj}) = 28,1\%$.

Riportiamo ora il modello che descrive sempre la stessa relazione ma considerando 79 sottocampioni di lunghezza $T = 50000$:

The regression equation is
 $H = 0,663 + 0,0540 \log_Range$

Predictor	Coef	SE Coef	T	P
Constant	0,66343	0,03243	20,45	0,000
log_Range	0,054020	0,009393	5,75	0,000

$S = 0,0334421$ $R\text{-Sq} = 30,0\%$ $R\text{-Sq}(\text{adj}) = 29,1\%$.

Sembra esserci una relazione positiva tra le due variabili, pressochè identica nei due modelli considerati: tuttavia in questo secondo modello è stato notato che la variabile `log_Range` appare autocorrelata positivamente e significativamente ad un passo. Si potrebbe pensare che questa seppure debole autocorrelazione possa aiutare a prevedere d per il periodo successivo; in altre parole si potrebbe pensare che la stima di d al tempo t sia legata dal (logaritmo del) range al tempo $t - 1$. Tuttavia, dopo aver stimato i parametri del modello, si nota come non ci sia una relazione significativa.

The regression equation is
 $H = 0,535 + 0,0166 \text{ log_Range_1}$

Predictor	Coef	SE Coef	T	P
Constant	0,53466	0,03845	13,90	0,000
log_Range_1	0,01658	0,01113	1,49	0,140

S = 0,0395166 R-Sq = 2,8% R-Sq(adj) = 1,6.%

Nonostante si sia trovata una relazione significativa tra range e \hat{d} , non è però chiaro il nesso causale che lega le due variabili: in certi momenti i prezzi diventano più volatili determinando un maggiore valore di \hat{d} , o ad un certo punto il processo r_t diventa più persistente determinando maggiore volatilità nei prezzi? D'altra parte va considerato che un processo più si allontana dalla condizione di white noise e più assume un range maggiore rispetto al white noise stesso. Il fatto che c'è una relazione tra d e range può essere considerato come abbastanza scontato.

Successivamente lo stesso lavoro è stato fatto considerando la deviazione standard al posto del range. La deviazione standard è stata calcolata sulle serie dei rendimenti. Proponendo un modello simile al precedente, non è stata trovata alcuna relazione significativa per le serie con $T = 100000$, mentre per $T = 50000$ si è al confine tra significatività e non, come mostrato in seguito, e la statistica R^2 indica che la relazione spiega una parte molto limitata della variabilità.

The regression equation is
 $H = 0,435 + 0,317 \text{ Std}$

Predictor	Coef	SE Coef	T	P
Constant	0,43461	0,02351	18,49	0,000
Std	0,3173	0,1683	1,89	0,063

S = 0,0390927 R-Sq = 4,4% R-Sq(adj) = 3,2%

La conclusione che si può trarre è che i sottocampioni che nella sezione precedente apparivano particolarmente turbolenti sono legati ad un valore di d maggiore per il fatto logico che range maggiori sono determinati da processi più persistenti; i rendimenti in questi sottocampioni non hanno una variabilità (la cui proxy usata è la deviazione standard) di regola maggiore rispetto a quella dei sottocampioni meno volatili.

Cointegrazione frazionaria

1. Stima del vettore di cointegrazione

In questa sezione si fa riferimento all'articolo di Marinucci e Robinson [9], in cui si parla di cointegrazione frazionaria. Scopo dell'analisi di cointegrazione è determinare la stima $\hat{\beta}$ nel modello

$$y_t = \beta' x_t + e_t,$$

dove $z_t = (x_t', y_t)'$ è un vettore osservabile per $t = 1, 2, \dots, T$, y_t è uno scalare e $x_t = (x_{1t}, \dots, x_{p-1,t})$. Si può dire che z_t è cointegrata (di ordini $d_1, \dots, d_{p-1}, d_y; d_e$) se x_{it} è $I(d_i)$, $i = 1, \dots, p-1$, y_t è $I(d_y)$ e se esiste un vettore β di dimensione $(p-1) \times 1$ tale che il processo non osservabile $e_t = y_t - \beta' x_t$ sia $I(d_e)$, con $d_e < d_y$. La definizione comprende che ci sia $d_i = d_y > d(e)$ per almeno un i .

La stima di β avviene mediante la tecnica del cross-periodogramma, che presenta alcuni vantaggi descritti più avanti rispetto alla più tradizionale tecnica dei minimi quadrati (OLS). Consideriamo un generico vettore a_t , con $t = 1, \dots, T$, la cui trasformata di Fourier $w_a(\lambda)$ è definita in maniera analoga a quanto visto nei capitoli precedenti

$$w_a(\lambda) = \frac{1}{(2\pi T)^{1/2}} \sum_{t=1}^T a_t e^{it\lambda},$$

dove con $\lambda = 2\pi j/T$, $j = 1, 2, \dots, T/2$ si indicano le frequenze.

Dato un secondo processo b_t , si può definire il cross-periodogramma $I_{ab}(\lambda)$ come

$$(6.1) \quad I_{ab} = w_a(\lambda) w_b'(-\lambda),$$

e si può definire il cross-periodogramma medio $\hat{F}_{ab}(m)$ come

$$(6.2) \quad \hat{F}_{ab}(m) = 2\text{Re} \left\{ \frac{2\pi}{T} \sum_{j=1}^m I_{ab}(\lambda_j) \right\} - \frac{2\pi}{T} 1 \left(m = \frac{n}{2} \right),$$

dove $1 \leq m \leq T/2$ e $1(\cdot)$ è una funzione indicatrice che si attiva con se T è pari e m raggiunge il massimo valore $T/2$. Il caso $m = \lfloor T/2 \rfloor$ è particolarmente interessante in quanto si ha

$$\hat{F}_{ab}(\lfloor T/2 \rfloor) = \frac{1}{T} \sum_{t=1}^T (a_t - \bar{a})(b_t - \bar{b})',$$

ovvero la covarianza campionaria, con \bar{a} e \bar{b} le medie campionarie. Si può comprendere come $\hat{F}_{ab}(m)$ rappresenta quindi il contributo delle frequenze $[1, \lambda_m]$ alla

covarianza campionaria. La stima di β mediante statistica FDLS è data da

$$\hat{\beta}_m = \hat{F}_{xx}(m)^{-1} \hat{F}_{xy}(m).$$

Nel caso si considerasse $m = \lfloor T/2 \rfloor$, si avrebbe la classica stima OLS, ovvero

$$\hat{\beta}_{\lfloor T/2 \rfloor} = \left(\sum_{t=1}^T (x_t - \bar{x})(x_t - \bar{x})' \right)^{-1} \sum_{t=1}^T (x_t - \bar{x})(y_t - \bar{y})'.$$

Anche in questo caso il parametro m viene scelto dall'utente secondo due possibili regole nel caso asintotico in cui $T \rightarrow \infty$, ovvero

$$(6.3) \quad m \sim CT, \quad 0 < C < \frac{1}{2},$$

e

$$(6.4) \quad \frac{1}{m} + \frac{m}{n} \rightarrow 0.$$

Nel primo caso viene considerato un sottoinsieme di frequenze non degeneranti, in maniera tale da considerare sempre la stessa banda al crescere di T : questo tipo di approccio è stato chiamato 'band-spectrum regression'. Nel secondo caso invece si considera un sottoinsieme di frequenze che aumentano al crescere di T ma che degenerano nell'intorno dell'origine, considerando bande via via sempre più strette.

Nonostante nel caso di analisi di cointegrazione di serie $I(1)$ o $I(2)$ ci possa essere mancanza di ortogonalità tra x_t e e_t , le stime di β utilizzando l'approccio OLS risultano consistenti in quanto asintoticamente x_t domina e_t . Nel caso invece in cui si considerano processi stazionari ($d_i < 1/2, \forall i$) questa mancanza di ortogonalità porta ad una perdita di consistenza dei minimi quadrati. Se si utilizza l'approccio FDLS le stime di β risultano invece consistenti, sempre nel caso ci sia correlazione tra x_t e e_t , questo perchè lo spettro del processo x_t domina quello di e_t nell'intorno della frequenza zero (ciò però accade se si considera la (6.4)). Nel caso di processi stazionari $\hat{\beta}_m$ sotto la condizione in (6.4) è quindi superiore alla tecnica OLS; Robinson e Marinucci hanno tuttavia esteso lo studio di $\hat{\beta}$ anche a serie non stazionarie perchè in un contesto frazionario si può non avere conoscenza certa sulla stazionarietà di un processo a priori, e perchè la cointegrazione è un fenomeno che riguarda principalmente le basse frequenze. Le proprietà di $\hat{\beta}_m$ dipendono dai diversi ordini di integrazione d_1, \dots, d_{p-1} e d_e , e dalle differenti regioni dello spazio d_1, \dots, d_{p-1}, d_e . Vengono riportati i diversi tassi di convergenza di $\hat{\beta}_{im}$ (dove con $\hat{\beta}_{im}$ si indica l' i -esimo elemento di $\hat{\beta}_m$). Verrà adottata la scrittura $X_n \sim_d h(n)$ per indicare che $h(n)^{-1} X_n$ converge in distribuzione ad una variabile aleatoria ben definita e non-degenere, il che implica $X_n = O_p(h(n))$. Per $i = 1, \dots, p-1$ si hanno i seguenti casi:

- (1) (Non-stazionarietà minore dell'unità): $d_i > \frac{1}{2}, d_e \geq 0, d_i + d_e < 1$.

$$\hat{\beta}_{i\lfloor T/2 \rfloor} - \beta_i \sim_d T^{1-d_i-d_{min}}, \quad \hat{\beta}_{im} - \beta_i \sim_d T^{d_e-d_i} m^{1-d_{min}-d_e}.$$

- (2) (Caso di confine): $\frac{1}{2} \leq d_i = 1 - d_e < 1$.

$$\hat{\beta}_{i\lfloor T/2 \rfloor} - \beta_i \sim_d T^{2d_e-1} \log T, \quad \hat{\beta}_{im} - \beta_i \sim_d T^{2d_e-1} \log m.$$

- (3) (Caso $I(1)/I(0)$): $d_i = \dots = d_{p-1} = 1, d_e = 0$.

$$\hat{\beta}_{i\lfloor T/2 \rfloor} - \beta_i \sim_d T^{-1}, \quad \hat{\beta}_{im} - \beta_i \sim_d T^{-1}.$$

- (4) (Non-stazionarietà maggiore dell'unità): $d_i = \dots = d_{p-1} > \frac{1}{2}$, $d_e > 0$, $d_i + d_e > 1$.

$$\hat{\beta}_{i|T/2} - \beta_i \sim_d T^{d_e - d_i}, \quad \hat{\beta}_{im} - \beta_i \sim_d T^{d_e - d_i}.$$

2. Simulazioni Monte Carlo

Nel medesimo articolo sono stati riportati i risultati di simulazioni Monte Carlo su dei piccoli campioni per confrontare le performance dei due metodi di stima (OLS e FDLS). Le serie cointegrate y_t e x_t sono rappresentate dal seguente modello:

$$(6.5) \quad \begin{aligned} y_t &= \beta x_t + e_t, \quad \beta = 2, \\ x_t &= \Delta^{-d_x} u_{1t}, \quad e_t = \Delta^{-d_e} u_{2t}, \end{aligned}$$

dove u_{1t} e u_{2t} sono due sequenze di variabili indipendenti normali caratterizzate da media zero, varianza unitaria e correlazione 0.5. Le simulazioni sono state condotte scegliendo diversi ordini di integrazione d_e e d_x in maniera tale da includere tutti i quattro casi elencati nella sezione precedente. Sono state generate serie storiche di lunghezza $T = 64$ e $T = 128$ (5000 repliche), e per ogni simulazione sono stati scelti 3 diversi valori di m , oltre al valore $\lfloor T/2 \rfloor$ corrispondente alla stima OLS.

In tabella 6.1 vengono riportati alcuni dei risultati in termini di bias: nella prima riga i risultati sono quelli presenti nell'articolo, nella seconda quelli ottenuti provando ad eseguire nuove simulazioni.

T	d_e	$d_x = 0.5$				$d_x = 0.8$				$d_x = 1.2$			
		$\hat{\beta}_{m_1}$	$\hat{\beta}_{m_2}$	$\hat{\beta}_{m_3}$	$\hat{\beta}_{\lfloor T/2 \rfloor}$	$\hat{\beta}_{m_1}$	$\hat{\beta}_{m_2}$	$\hat{\beta}_{m_3}$	$\hat{\beta}_{\lfloor T/2 \rfloor}$	$\hat{\beta}_{m_1}$	$\hat{\beta}_{m_2}$	$\hat{\beta}_{m_3}$	$\hat{\beta}_{\lfloor T/2 \rfloor}$
128	0.2	0.53	0.55	0.57	0.66	0.23	0.25	0.26	0.34	0.03	0.03	0.04	0.05
128	0.2	0.20	0.21	0.22	0.35	0.05	0.06	0.06	0.13	0.00	0.00	0.00	0.01

TABELLA 6.1. Analisi numerica sulla stima del parametro β .

I risultati appaiono piuttosto diversi, ma è stato verificato che la stima in $\beta_{\lfloor T/2 \rfloor}$ coincide con quella mediante OLS. Di seguito vengono mostrati e commentati i codici usati per le analisi.

```

1 function[u1,u2,x,e,y,Beta]=Simcoint(Rip,beta,dx,de,T,corr,m,flag)
2
3 for i=1:Rip
4 [u1,u2,x,e,b1,b2]=arfima_corr(dx,de,T,corr);
5
6 y=beta*x+e;
7
8 [Beta(i,:)]=fdls(y,x,m,flag);
9 end

```

Innanzitutto i parametri che il codice riceve in ingresso sono nell'ordine: numero di ripetizioni da effettuare, parametri β , d_x e d_e del modello riportato in 6.5, lunghezza della serie storica T , correlazione tra i rumori bianchi u_{1t} e u_{2t} , valori di m da considerare (m può essere un vettore), ed infine se **flag** assume il valore 'OLS' viene aggiunta una m corrispondente a $T/2$. Il codice restituisce i rumori u_t , le serie

y_t , x_t e e_t , e la matrice Beta ($\text{Rip} \times m$) delle stime ($m+1$ in caso 'OLS' sia attivato). Per fare ciò si utilizzano le funzioni per la creazione dei rumori `arfima_corr` e per la stima `fdls`.

```

1 function [u1,u2,x,e,b1,b2]=arfima_corr(dx,de,T,corr);
2
3 T1=T+500;
4
5 u1=normrnd(0,1,1,T1);
6 b1(1)=1;
7 for i=1:T1-1
8     b1(i+1)=-b1(i)*(-dx-i+1)/i;
9 end
10 for i=1:T1
11     x(i)=u1(1:i)*b1(i:-1:1)';
12 end
13
14 u3=normrnd(0,1,1,T1);
15 u2=corr*u1+sqrt(1-corr^2)*u3;
16
17 b2(1)=1;
18 for i=1:T1-1
19     b2(i+1)=-b2(i)*(-de-i+1)/i;
20 end
21 for i=1:T1
22     e(i)=u2(1:i)*b2(i:-1:1)';
23 end
24
25 x=x(T1-T+1:T1);
26 e=e(T1-T+1:T1);

```

Il codice innanzitutto (riga 3) incrementa la lunghezza delle serie generate (in questo caso di 500, anche se variando questo valore non si hanno avuti significativi cambiamenti nelle stime), definendo il nuovo valore `T1`. In riga 5 si genera il rumore u_{1t} con media nulla e varianza unitaria. Da riga 6 a riga 12 si genera il processo arfima x_t sfruttando l'algoritmo ricorsivo: il vettore `b1` è quello dei coefficienti della rappresentazione $\text{MA}(\infty)$. Nelle righe 14 e 15 si genera u_{2t} , che sarà costituito in parte dal u_{1t} e in parte da un nuovo white noise u_{3t} in modo tale da avere correlazione desiderata e varianza unitaria. Da riga 17 a riga 23 si genera e_t a partire da u_{2t} nella stessa maniera descritta sopra (i coefficienti sono ora chiamati `b2`). Infine (riga 25 e 26) si riportano le serie alla lunghezza T desiderata eliminando i primi (500) dati.

```

1 function[beta]=fdls(y,x,m,flag)
2
3 T=size(x);
4 T=T(2);
5

```

```

6 M=max(m);
7   if strcmp(flag,'OLS')
8     M=T/2;
9   end
10
11 w_x=fft(x);
12 w_x=w_x(2:M+1);
13 w_x_conj=conj(w_x);
14 I_xx=(2*pi*T)^(-.5)*(2*pi*T)^(-.5)*(w_x(1:M).*w_x_conj(1:M));
15
16 w_y=fft(y);
17 w_y=w_y(2:M+1);
18 w_y_conj=conj(w_y);
19 I_yy=(2*pi*T)^(-.5)*(2*pi*T)^(-.5)*(w_y(1:M).*w_y_conj(1:M));
20
21 I_xy=(2*pi*T)^(-.5)*(2*pi*T)^(-.5)*(w_x(1:M).*w_y_conj(1:M));
22
23 cont=size(m);
24 cont=cont(2);
25
26 for i=1:cont
27   F_xx(i)=2*real(2*pi/T*sum(I_xx(1:m(i))));
28   F_xy(i)=2*real(2*pi/T*sum(I_xy(1:m(i))));
29 end
30 if strcmp(flag,'OLS')
31   F_xx(i+1)=2*real(2*pi/T*sum(I_xx(1:T/2)))-2*pi*I_xx(T/2);
32   F_xy(i+1)=2*real(2*pi/T*sum(I_xy(1:T/2)))-2*pi*I_xy(T/2);
33 end
34 for i=1:cont
35   beta(i)=(F_xx(i)^(-1))*F_xy(i);
36 end
37 if strcmp(flag,'OLS')
38   beta(i+1)=(F_xx(i+1)^(-1))*F_xy(i+1);
39 end

```

Infine il codice sopra riportato stima i coefficienti β_m utilizzando i FDLS. Il codice determina T a partire dalle serie date in ingresso (riga 3 e 4): determina inoltre la m più grande M tra quelle date (righe 6-9). Si calcola poi il periodogramma I_{xx} (righe 11-14) sfruttando la trasformata di Fourier veloce (riga 11): stesso discorso per il periodogramma I_{yy} (righe 16-19), anche se in realtà ai fini della stima serve solo la trasformata coniugata complessa espressa in riga 18. Il cross-periodogramma si calcola come in 6.1, moltiplicando punto a punto la trasformata w_x per il coniugato complesso della trasformata w_y (ciò è analogo a considerare $w_y(-\lambda)$). Dopo aver calcolato la dimensione del vettore contenente le diverse m (righe 22-23), si esegue un ciclo che per ogni valore di m determina il relativo cross-periodogramma medio (righe 25-28). Nel caso in cui si desideri la stima OLS, verrà calcolato anche il cross-periodogramma medio valutato in $T/2$ con i dovuti accorgimenti (si veda la 6.2). Infine (righe 33-38) si calcola il vettore di stime $\hat{\beta}_m$.

Applicazione: analisi multivariata

1. Allineare gli ordini di integrazione

Nel seguito si proverà a testare se le serie storiche relative ai prezzi (logaritmici) e agli order flow cumulati siano cointegrate o meno: tali serie presentano però diversi ordini di integrazione (sotto $d = 1$ per quanto riguarda i prezzi e sopra $d = 1$ per quanto riguarda gli order flow cumulati), ed è stato visto come y_t e x_t debbano avere lo stesso ordine di integrazione se si ipotizza un modello $y_t = \beta x_t + e_t$, dove e_t è un processo $I(d_e)$ con $d_e < d_y$. Come prima cosa è necessario quindi allineare i due diversi ordini di integrazione. In questa sezione si vedrà come procedere, in particolare si cercherà di capire se innalzare (o abbassare) l'ordine di integrazione dei prezzi sia equivalente o meno a fare ciò sulla corrispondente serie dei rendimenti cumulati (la differenza tra le due serie è che la prima ha valore $p_0 \neq 0$ in $t = 0$ e la seconda $r_0 = 0$). Per fare ciò sono state effettuate simulazioni utilizzando il modello

$$y_t^* = \beta x_t + e_t,$$

dove il simbolo * sta ad indicare che il relativo processo non è osservabile, e_t (anch'esso non osservabile) è un processo generato con parametro $d_e = 0.6$ (non stazionario), x_t è generato con parametro $d_x = 0.9$ in un primo insieme di simulazioni (Gruppo 1) e con $d_x = 1.1$ nel Gruppo 2, β è posto pari 0.25 e la lunghezza delle serie storiche è $T = 30000$. Il vettore y_t^* così determinato avrà un ordine di integrazione d_{y^*} pari a circa quello di x_t , dato che $I(d_x) > I(d_e)$. La serie y_t (osservabile) è data da

$$y_t = (1 - L)^\Delta y_t^* + \mu,$$

dove Δ è stato posto a -0.2 nel Gruppo 1 e 0.2 nel Gruppo 2 e μ a 10. Per ogni simulazione (100 per ogni gruppo) si procederà a generare le serie come descritto sopra e a stimare alcuni elementi considerando i seguenti modelli:

Modello	Descrizione	Elementi stimati
1	$y_t^* = \beta x_t + e_t, \hat{e}_t = y_t^* - \hat{\beta} x_t,$	$\hat{d}_y, \hat{d}_{y^*}, \hat{d}_x, \hat{d}_e, \hat{\beta}, \hat{d}_{\hat{e}}$
2	$(1 - L)^{-\Delta} y_t = \beta x_t + e'_t \Rightarrow y'_t = \beta x_t + e'_t$	$\hat{d}_{y'}, \hat{\beta}', \hat{d}_{e'}$
3	$(1 - L)^{-\Delta} (y_t - \mu) = \beta x_t + e''_t \Rightarrow y''_t = \beta x_t + e''_t$	$\hat{d}_{y''}, \hat{\beta}'', \hat{d}_{e''}$

Il modello 1 presuppone di avere a disposizione la serie y_t^* non osservabile, e fornisce quindi le migliori stime possibili di β e degli ordini di integrazione. Quando invece si ci pone il problema di ricostruire y_t^* dalla serie osservabile y_t i procedimenti qui analizzati sono due: modificare l'ordine di integrazione della serie y_t applicando il filtro $(1 - L)^{-\Delta}$ direttamente alla serie senza considerare la condizione iniziale data da μ (modello 2) oppure modificare l'ordine di integrazione di y_t applicando il filtro alla serie preventivamente riportata alla condizione iniziale originaria (sottraendo

quindi μ). Si ci aspetta che il modello 3 dia i medesimi risultati del modello 1, in quanto y'' e y^* sono esattamente la stessa cosa se si ipotizza che i parametri Δ e μ siano noti (come verrà fatto inizialmente). In una fase successiva si procederà alla stima di questi due parametri e i modelli 1 e 3 risulteranno diversi. Nel seguito (tabella 7.1) sono schematizzati i risultati ottenuti nel caso di parametri noti.

	Gruppo 1	Gruppo 2
\hat{d}_y	1.106274 (0.0598)	0.901792 (0.0527)
\hat{d}_x	0.904137 (0.0600)	1.105208 (0.0526)
\hat{d}_e	0.596330 (0.0538)	0.599462 (0.0528)
\hat{d}_{y^*}	0.904091 (0.0600)	1.105199 (0.0526)
$\hat{d}_{\hat{e}}$	0.598585 (0.0523)	0.601839 (0.0515)
$\hat{d}_{y'}$	0.904124 (0.0600)	1.105415 (0.0526)
$\hat{d}_{e'}$	0.872412 (0.0726)	1.115102 (0.0708)
$\hat{d}_{y''}$	0.904091 (0.0600)	1.105199 (0.0526)
$\hat{d}_{\hat{e}''}$	0.598585 (0.0523)	0.601839 (0.0515)
	Gruppo 1	Gruppo 2
$\hat{\beta}$	0.250044 (0.000165)	0.250007 (0.000038)
$\hat{\beta}'$	0.250204 (0.006996)	0.245028 (0.052256)
$\hat{\beta}''$	0.250044 (0.000165)	0.250007 (0.000038)

TABELLA 7.1. Stime ottenute ponendo $d_x = 0.9$, $\Delta = -0.2$ nel Gruppo 1 e $d_x = 1.1$, $\Delta = 0.2$ nel Gruppo 2: parametri μ e Δ noti.

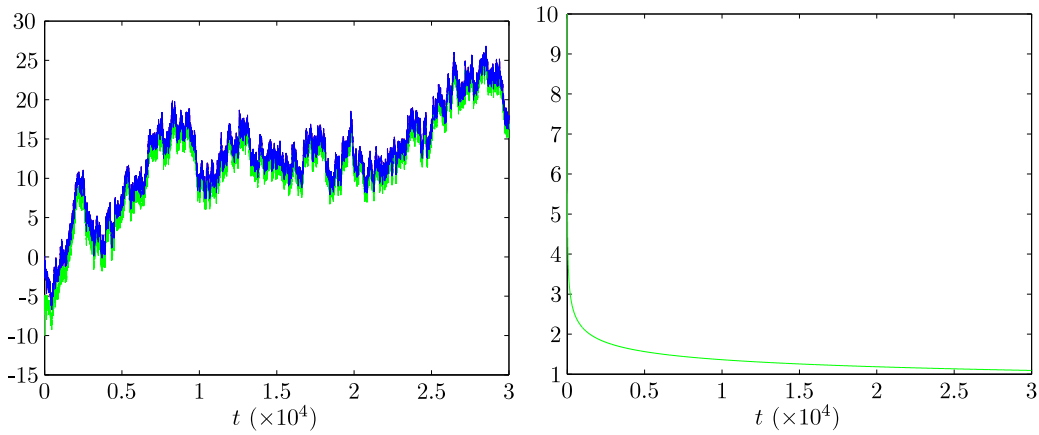
Si può notare come, nonostante le serie y'_t e y''_t presentino il medesimo ordine di integrazione di y_t^* , tuttavia gli ordini di integrazione dei residui associati al modello corrispondente sono nettamente diversi. Nel caso si consideri il modello 2 il residuo \hat{e}'_t ha ordine di integrazione pari a circa quello della variabile dipendente, sinonimo che le serie y'_t e x_t non appaiono cointegrate. Considerando invece il modello 3, le serie y''_t e x_t appaiono cointegrate in quanto l'ordine di integrazione di \hat{e}''_t è molto simile a $d_e = 0.6$. Anche osservando le stime di β si ci accorge come ci sia qualcosa di anomalo nel modello 2: le stime sono infatti caratterizzate da varianza molto maggiore rispetto al modello 3 (che, come atteso, fornisce risultati identici al modello 1). Il caso peggiore si ha nel caso di integrazione verso l'ordine superiore (Gruppo 2).

In figura 7.1 vengono mostrate a sinistra le serie $y_t^* = y'_t$ in blu e y'_t in verde, e a destra la serie $\Delta_t = y_t^* - y'_t$. La simulazione presa in considerazione è l'ultima del Gruppo 1. Si può notare come la differenza tra le serie y'_t e y_t^* non sia una costante, ma segua un andamento che ricorda quello dei termini dati dallo sviluppo di $(1 - L)^d$, con $d > 0$. Quello che avviene è infatti che la condizione iniziale (pari a -10) viene anch'essa filtrata: i primi valori Δ_t sono

$$\Delta_1 = 10.00, \quad \Delta_2 = 8.00, \quad \Delta_3 = 7.20, \quad \Delta_4 = 6.72,$$

mentre i primi coefficienti dello sviluppo di $(1 - L)^d$, con $d = 0.2$, sono

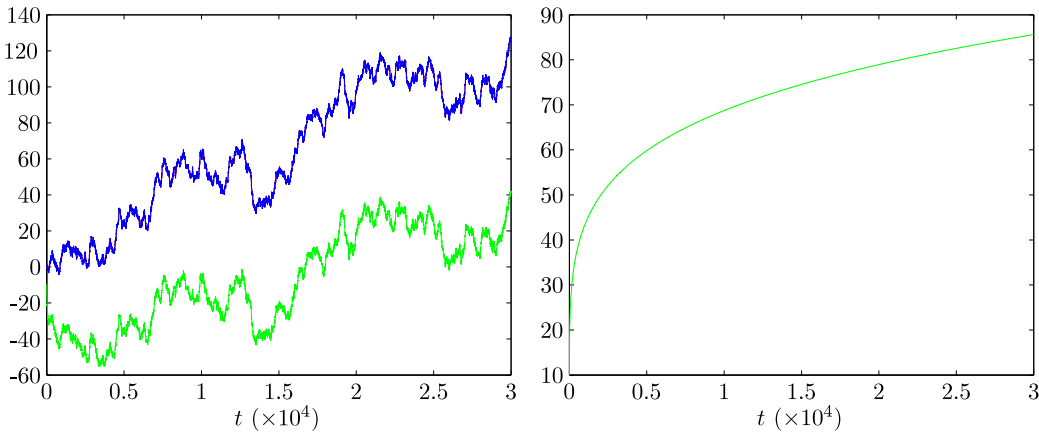
$$b_1 = 1.000, \quad b_2 = -0.200, \quad b_3 = -0.080, \quad b_4 = -0.048.$$

FIGURA 7.1. A sinistra $y_t^* = y_t''$ in blu e y_t' in verde. A destra Δ_t .

Segue che Δ_t può essere ricavata come

$$\Delta_1 = 10b_1, \quad \Delta_2 = 10(b_1 + b_2), \quad \Delta_3 = 10(b_1 + b_2 + b_3), \quad \Delta_4 = 10(b_1 + b_2 + b_3 + b_4).$$

In figura 7.2 mostriamo invece la medesima simulazione ma riferita al Gruppo 2. In questo caso il filtraggio sulla serie senza aggiustamento della condizione iniziale

FIGURA 7.2. A sinistra $y_t^* = y_t''$ in blu e y_t' in verde. A destra Δ_t .

produce distorsioni a prima vista maggiori, anche se il fenomeno che avviene è lo stesso, ovvero quello di un filtraggio della condizione iniziale, questa volta con un filtro $(1 - L)^d$ con $d < 0$. I primi valori Δ_t sono

$$\Delta_1 = 10.00, \quad \Delta_2 = 12.00, \quad \Delta_3 = 13.20, \quad \Delta_4 = 14.08,$$

mentre i primi coefficienti dello sviluppo di $(1 - L)^d$, con $d = -0.2$, sono

$$b_1 = 1.000, \quad b_2 = 0.200, \quad b_3 = 0.120, \quad b_4 = 0.088.$$

Segue che Δ_t può essere ricavata come

$$\Delta_1 = 10b_1, \quad \Delta_2 = 10(b_1 + b_2), \quad \Delta_3 = 10(b_1 + b_2 + b_3), \quad \Delta_4 = 10(b_1 + b_2 + b_3 + b_4).$$

I coefficienti b_j dello sviluppo del filtro $(1 - L)^d$ si avvicinano a zero molto più velocemente con $d > 0$ (quando sono negativi) piuttosto che con $d < 0$ (quando sono positivi): questo comporta che Δ_t assume più velocemente un comportamento simile ad una costante c nel primo caso piuttosto che nel secondo (anche se in entrambi i casi $\Delta_t \rightarrow c$ con $b_j \rightarrow 0$, e questo avviene per $j \rightarrow \infty$). Tanto più Δ_t è simile ad una costante tanto più la stima $\hat{\beta}$ è vicina al valore corretto (ottenuto senza filtrare la costante): per questo motivo l'allineamento 'verso il basso' ha portato a risultati meno dannosi che 'verso l'alto'.

Proponiamo ora le medesime simulazioni questa volta ipotizzando che Δ e μ siano incogniti e debbano essere stimati: $\hat{\Delta}$ sarà ottenuta facendo la differenza tra \hat{d}_y e \hat{d}_x , mentre $\hat{\mu}$ sarà data dal primo valore assunto dal vettore osservabile y_t con segno opposto. In tabella 7.2 mostriamo i risultati ottenuti.

	Gruppo 1	Gruppo 2
\hat{d}_y	1.106274 (0.0598)	0.901792 (0.0527)
\hat{d}_x	0.904137 (0.0600)	1.105208 (0.0526)
\hat{d}_e	0.596330 (0.0538)	0.599462 (0.0528)
\hat{d}_{y^*}	0.904091 (0.0600)	1.105199 (0.0526)
$\hat{d}_{\hat{e}}$	0.598585 (0.0523)	0.601839 (0.0515)
$\hat{d}_{y'}$	0.901924 (0.0603)	1.109019 (0.0528)
$\hat{d}_{e'}$	0.853674 (0.1075)	1.115551 (0.0752)
$\hat{d}_{y''}$	0.901891 (0.0603)	1.108799 (0.0528)
$\hat{d}_{\hat{e}''}$	0.645787 (0.0650)	0.910469 (0.0940)

	Gruppo 1	Gruppo 2
$\hat{\beta}$	0.250044 (0.000165)	0.250007 (0.000038)
$\hat{\beta}'$	0.245870 (0.010441)	0.252920 (0.055724)
$\hat{\beta}''$	0.245715 (0.007031)	0.257886 (0.010641)

TABELLA 7.2. Stime ottenute ponendo $d_x = 0.9$, $\Delta = -0.2$ nel Gruppo 1 e $d_x = 1.1$, $\Delta = 0.2$ nel Gruppo 2: parametri μ e Δ incogniti.

Si può notare come gli ordini di integrazione dei residui $\hat{d}_{e'}$ e $\hat{d}_{\hat{e}''}$ presentino una varianza maggiore rispetto al caso con parametri noti, e oltre a quelli riferiti al modello 2 anche i residui riferiti al modello 3 hanno un ordine di integrazione che inizia ad allontanarsi da quello originario $d_e = 0.6$. La situazione è nettamente peggiore nel Gruppo 2 (in cui si alza l'ordine di integrazione) rispetto al Gruppo 1.

Anche in questo caso mostriamo le diverse serie riferite all'ultima simulazione. In figura 7.3 si mostrano a sinistra le serie y_t^* in rosso e y_t'' in blu (questa volta non coincidono) e y_t' in verde, e a destra la serie $\Delta_{1t} = y_t^* - y_t''$ in blu e $\Delta_{2t} = y_t^* - y_t'$ in verde, riferite al Gruppo 1. Infine in figura 7.4 si mostrano le stesse serie riferite al Gruppo 2.

Volendo scomporre i risultati appena visti nei due casi in cui sia solo uno il parametro incognito (alternativamente prima d e poi μ), si ottiene nel caso di d incognito e μ noto i risultati riportati in tabella (7.3), e nel caso di d noto e μ incognito i risultati illustrati in tabella (7.5).

Si nota come nel caso di allineamento dell'ordine di integrazione di y_t verso il basso (Gruppo 1), il comportamento del residuo \hat{e}'' è molto simile al caso con i

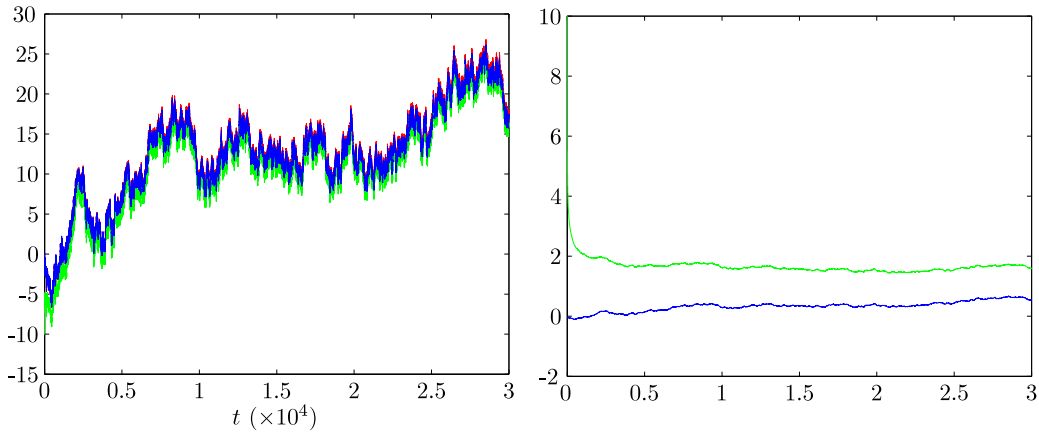


FIGURA 7.3. A sinistra y_t^* in rosso, y_t'' in blu e y_t' in verde. A destra Δ_{1t} in blu e Δ_{2t} in verde.

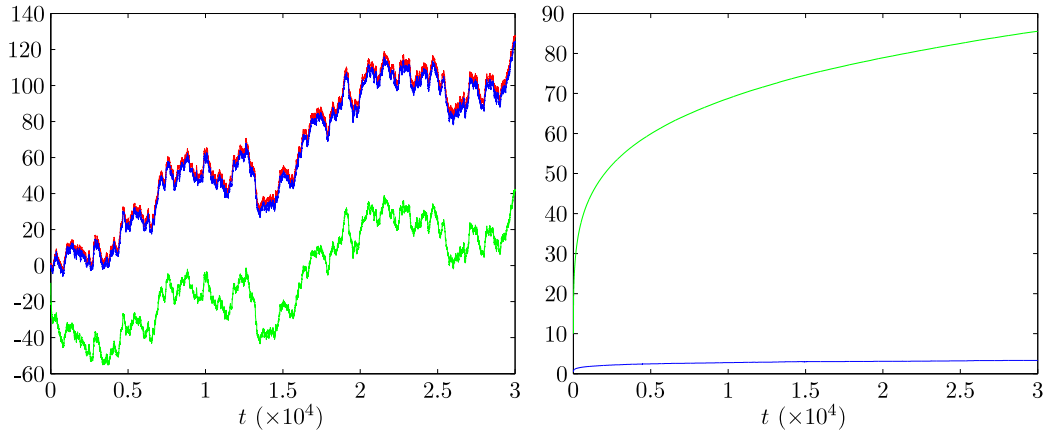


FIGURA 7.4. A sinistra y_t^* in rosso, y_t'' in blu e y_t' in verde. A destra Δ_{1t} in blu e Δ_{2t} in verde.

parametri incogniti per quanto riguarda la tabella (7.3) (d incognito e μ noto) e molto simile al caso con i parametri noti per quanto riguarda la tabella (7.5) (d noto e μ incognito): in altre parole il fatto di aver stimato μ ha degli effetti trascurabili sul risultato finale. Per quanto riguarda invece l'allineamento verso l'alto (Gruppo 2) si ha una situazione diversa, in quanto in entrambe le simulazioni il fatto di aver stimato un parametro ha degli effetti non trascurabili: sembrerebbe che il dover stimare μ porti a risultati peggiori rispetto al dover stimare d .

Infine si riesegono le simulazioni in condizione di parametri noti stressando maggiormente il parametro Δ (viene posto pari a 0.6): scopo della simulazione è verificare se la deviazione standard del parametro β , che è stato visto essere più piccola nel Gruppo 2 rispetto al Gruppo 1 (nella simulazione con parametri noti), mantiene o meno lo stesso comportamento. Anche in questo caso le stime β relative

	Gruppo 1	Gruppo 2
\hat{d}_y	1.106274 (0.0598)	0.901792 (0.0527)
\hat{d}_x	0.904137 (0.0600)	1.105208 (0.0527)
\hat{d}_e	0.596330 (0.0538)	0.599462 (0.0528)
\hat{d}_{y^*}	0.904091 (0.06009)	1.105199 (0.0526)
$\hat{d}_{\hat{e}}$	0.598586 (0.0523)	0.601839 (0.0515)
$\hat{d}_{y'}$	0.901924 (0.0603)	1.109019 (0.0528)
$\hat{d}_{e'}$	0.853680 (0.1078)	1.115551 (0.0752)
$\hat{d}_{y''}$	0.9018908 (0.0603)	1.108796 (0.0528)
$\hat{d}_{e''}$	0.645352 (0.0639)	0.848868 (0.0961)
	Gruppo 1	Gruppo 2
$\hat{\beta}$	0.250044 (0.000165)	0.250007 (0.000038)
$\hat{\beta}'$	0.245870 (0.010441)	0.252920 (0.055724)
$\hat{\beta}''$	0.245701 (0.007049)	0.258132 (0.010689)

TABELLA 7.3. Stime ottenute ponendo $d_x = 0.9$, $\Delta = -0.2$ nel Gruppo 1 e $d_x = 1.1$, $\Delta = 0.2$ nel Gruppo 2: parametro μ noto e Δ incognito.

	Gruppo 1	Gruppo 2
\hat{d}_y	1.106274 (0.0598)	0.901792 (0.0527)
\hat{d}_x	0.904137 (0.0600)	1.105208 (0.0526)
\hat{d}_e	0.596330 (0.0538)	0.599462 (0.0528)
\hat{d}_{y^*}	0.904091 (0.0600)	1.105199 (0.0527)
$\hat{d}_{\hat{e}}$	0.598585 (0.0523)	0.601839 (0.0515)
$\hat{d}_{y'}$	0.904124 (0.0600)	1.105418 (0.0527)
$\hat{d}_{e'}$	0.872412 (0.0726)	1.115102 (0.0708)
$\hat{d}_{y''}$	0.904092 (0.0600)	1.105202 (0.0527)
$\hat{d}_{e''}$	0.603742 (0.0545)	0.892026 (0.1710)
	Gruppo 1	Gruppo 2
$\hat{\beta}$	0.250044 (0.000165)	0.250007 (0.000038)
$\hat{\beta}'$	0.250204 (0.006996)	0.245028 (0.052258)
$\hat{\beta}''$	0.250056 (0.000246)	0.249788 (0.001307)

TABELLA 7.4. Stime ottenute ponendo $d_x = 0.9$, $\Delta = -0.2$ nel Gruppo 1 e $d_x = 1.1$, $\Delta = 0.2$ nel Gruppo 2: parametro μ incognito e Δ noto.

al modello ben specificato presentano una deviazione standard minore nel Gruppo 2 rispetto al Gruppo 1, amplificando il fenomeno riscontrato per valori di Δ minori.

Come ultima cosa resta da verificare se applicare il filtro $(1 - L)^d$ alla serie dei rendimenti cumulati p_t (con condizione iniziale pari a 0) ottenendo la serie $y_t^{(1)}$ o applicare il medesimo filtro alla serie dei rendimenti r_t e cumulare successivamente ottenendo la serie $y_t^{(2)}$ porti al medesimo risultato o meno. In teoria si, in quanto quello che viene fatto è

$$y_t^{(1)} = (1 - L)^d p_t = (1 - L)^d (1 - L)^{-1} r_t$$

	Gruppo 1	Gruppo 2
\hat{d}_y	1.410410 (0.0610)	0.800158 (0.0532)
\hat{d}_x	0.801682 (0.0604)	1.414102 (0.0523)
\hat{d}_e	0.596330 (0.0538)	0.599462 (0.0528)
\hat{d}_{y^*}	0.801615 (0.0604)	1.414099 (0.0523)
$\hat{d}_{\hat{e}}$	0.598046 (0.0523)	0.600972 (0.0512)
$\hat{d}_{y'}$	0.801619 (0.0604)	1.421743 (0.0538)
$\hat{d}_{y''}$	0.801615 (0.0604)	1.414099 (0.0523)
$\hat{d}_{\hat{e}'}$	0.623906 (0.0590)	1.418557 (0.0559)
$\hat{d}_{\hat{e}''}$	0.598046 (0.0523)	0.600972 (0.0512)

	Gruppo 1	Gruppo 2
$\hat{\beta}$	0.250085 (0.000370)	0.250000 (0.000003)
$\hat{\beta}'$	0.250081 (0.001144)	0.257796 (0.574237)
$\hat{\beta}''$	0.250085 (0.000370)	0.250000 (0.000003)

TABELLA 7.5. Stime ottenute ponendo $d_x = 0.8$, $\Delta = -0.6$ nel Gruppo 1 e $d_x = 1.4$, $\Delta = 0.6$ nel Gruppo 2: parametri noti.

nel primo caso e

$$y_t^{(2)} = (1 - L)^{-1}(1 - L)^d r_t$$

nel secondo. Sperimentalmente è stato verificato che $y_t^{(1)} = y_t^{(2)}$: è stato inoltre verificato che al medesimo risultato si può giungere applicando un unico filtro, ovvero facendo

$$y_t^{(3)} = (1 - L)^{d-1} r_t = y_t^{(2)} = y_t^{(1)}$$

2. Analisi di cointegrazione

Ora proviamo a stimare il vettore di cointegrazione $[1, -\beta]$ del modello

$$(7.1) \quad Y_t = \beta X_t + E_t,$$

dove X_t rappresenta l'order flow cumulato e Y_t i rendimenti cumulati. Secondo la teoria più classica di efficienza dei mercati, le due serie non dovrebbero essere cointegrate, in quanto le dinamiche di acquisto e vendita del titolo non dovrebbero interagire sul prezzo del medesimo: in altre parole il book (e di conseguenza il prezzo mid-point) si dovrebbe aggiornare esclusivamente sulla base di nuove informazioni disponibili e non sulla base del flusso di ordini. Da un'analisi di cointegrazione mediante tecniche classiche si è infatti trovato che entrambi i processi considerati sono $I(1)$, così come il residuo dato da una loro combinazione lineare, ovvero si è trovata assenza di cointegrazione. Tuttavia se si considera un contesto di integrazione frazionaria, è già stato visto come i processi in questione non siano esattamente $I(1)$ ($d > 1$ per quanto riguarda gli order flow cumulati e $d < 1$ per quanto riguarda i prezzi), ed è possibile che anche il residuo del modello rappresentato in (7.1) abbia un ordine di integrazione non intero: in particolare le due serie X_t e Y_t si possono considerare cointegrate frazionariamente se l'ordine di integrazione del residuo è inferiore rispetto a quello delle due serie. Inoltre il fatto di trovare un eventuale ordine di integrazione frazionario del residuo, purchè maggiore di 0.5, non

violerebbe quanto affermato dalla teoria classica, in quanto la serie risulta sempre non stazionaria (indicando assenza di cointegrazione tra X_t e Y_t).

Verrà analizzato il titolo Enel, considerando separatamente la parte del campione prima del cambio del tick e quella successiva. Il primo passo da fare è quello di allineare gli ordini di integrazione di X_t e Y_t : dato che le stime \hat{d} variano a seconda della frequenza considerata, segue che anche l'allineamento risulta diverso a seconda della frequenza scelta. In queste applicazioni è stato visto come la frequenza che si ritiene più idonea a modellizzare un comportamento di lungo termine sia $\lambda^{(a)} = 0.05$: tuttavia verrà proposta la stima anche per $\lambda^{(b)} = 0.5$, per coglierne eventualmente le differenze. Indicheremo con $\hat{d}_x^{(a)}$, $\hat{d}_x^{(b)}$, $\hat{d}_y^{(a)}$, $\hat{d}_y^{(b)}$ gli ordini di integrazione stimati riferiti alle serie differenziate (x_t sono gli order flow e y_t sono i rendimenti), e con $\Delta^{(a)} = \hat{d}_x^{(a)} - \hat{d}_y^{(a)}$ e $\Delta^{(b)} = \hat{d}_x^{(b)} - \hat{d}_y^{(b)}$ le differenze di queste stime.

Ad ognuna delle due frequenze considerate corrisponderanno due stime, in quanto l'allineamento è stato fatto in due diversi modi:

- (1) allineando i rendimenti agli order flow ($x_t \sim I(d_x)$ e $y_t \sim I(d_x)$);
- (2) allineando gli order flow ai rendimenti ($x_t \sim I(d_y)$ e $y_t \sim I(d_y)$).

Il numero di stime del parametro β proposte sono quindi in totale 8 (due sottocampioni, due frequenze e due allineamenti). Per ognuna di queste stime verrà riportato, oltre il valore assunto da β , l'ordine di integrazione $I(d_e)$ dei residui.

Prima di presentare i risultati, si mostrano alcune figure riguardo l'allineamento delle serie. In figura (7.5) sono riportate le serie allineate all'ordine di integrazione d_x dell'order flow (sinistra) e d_y del rendimento (destra) e successivamente cumulate: gli ordini di integrazione sono associati a $\lambda^{(a)} = 0.05$.

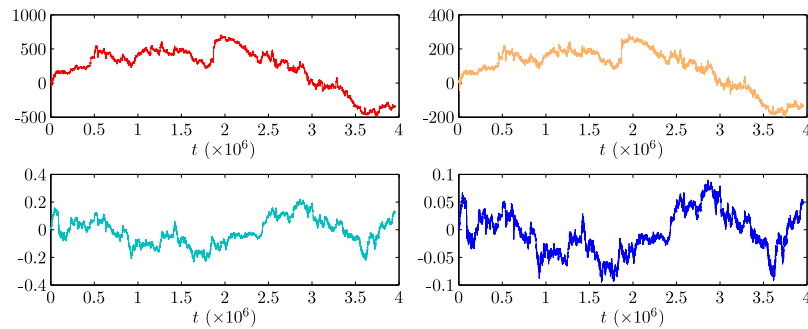


FIGURA 7.5. Order flow (sopra) e rendimenti (sotto) allineati al medesimo ordine di integrazione $d_x = d_y = 0.0805$ (sinistra), e $d_x = d_y = -0.0149$ (destra), con successiva cumulazione della serie.

Si precisa che l'aver filtrato i rendimenti con successiva cumulazione equivale ad aver filtrato la serie dei prezzi depurata dalla condizione iniziale (ovvero inizializzata a 0). Questo espediente permette di evitare di inserire nel filtraggio possibili elementi deterministici, come mostrato nella sezione precedente. In figura 7.6 sono mostrate le stime dell'ordine di integrazione \hat{d} delle serie appena viste: le stime sono relative alle serie con ordine di integrazione pari a d_x (order flow) a sinistra, e pari a d_y (rendimenti) a destra. Si può notare come effettivamente le serie una volta allineate presentino un ordine di integrazione pressochè identico alla frequenza

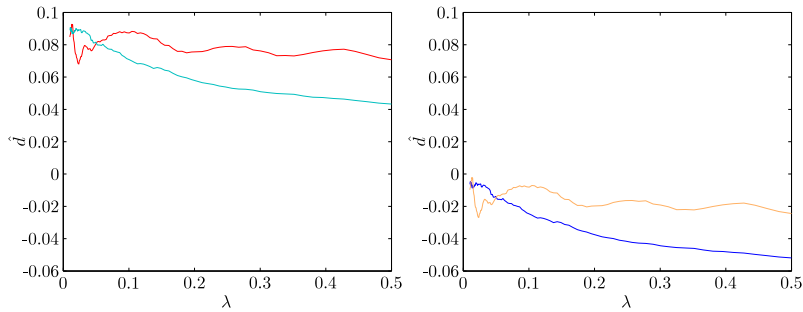


FIGURA 7.6. Stime \hat{d} riferite alle serie mostrate in figura 7.5 differenziate. Le serie una volta filtrate presentano il medesimo ordine di integrazione nell'intorno della frequenza considerata $\lambda = 0.05$.

considerata (ricordiamo pari a $\lambda = 0.05$ in questo caso), e ciò può essere indicativo del buon esito della tecnica di allineamento. Ovviamente se si fossero considerati gli ordini di integrazione corrispondenti alla frequenza $\lambda = 0.5$, le serie ad allineamento avvenuto avrebbero avuto il medesimo ordine di integrazione proprio per $\lambda = 0.5$.

Ora si riportano i dati ottenuti: in tabella (7.6) si riportano la frequenza considerata, la differenza Δ degli ordini di integrazione delle serie X_t e Y_t , l'ordine di integrazione comune delle due serie dopo l'allineamento, la stima $\hat{\beta}$ ottenuta e l'ordine di integrazione stimato \hat{d}_E del residuo del modello, ottenuto come

$$E_t = \hat{\beta}X_t - Y_t.$$

La stima di questa serie è calcolata alla medesima frequenza considerata per la determinazione delle stime di X_t e Y_t . La relazione tra order flow cumulati e

	λ	Δ	$d_X = d_Y$	β	\hat{d}_E
Ante cambio tick	0.05	0.1094	0.9607	0.001063	0.9168
	0.05	0.1094	1.0701	0.001267	1.0323
	0.5	0.1528	0.8938	0.001653	0.8614
	0.5	0.1528	1.0465	0.002218	1.0190
Post cambio tick	0.05	0.0954	0.9851	-0.000054	0.9818
	0.05	0.0954	1.0805	-0.000082	1.0778
	0.5	0.1226	0.9481	-0.000075	0.9731
	0.5	0.1226	1.0707	-0.000126	1.0960

TABELLA 7.6. Risultati dell'analisi di cointegrazione

rendimenti cumulati sembra non essere particolarmente evidente, in particolare nel sottocampione successivo al cambio del tick, dove addirittura compare con il segno negativo. Nella parte antecedente al cambio del tick, i valori di β sono positivi e più significativi: l'ordine di integrazione del residuo appare più basso rispetto all'ordine di integrazione delle serie di partenza, ma si tratta di valori modesti (la differenza tra i due ordini di integrazione è di circa 0.03). Sempre guardando la parte prima del cambio del tick, si osserva una differenza tra i valori di $\hat{\beta}$ sia per quanto riguarda il diverso allineamento fatto, sia per quanto riguarda la differente frequenza considerata. Dato che è stato visto infatti come allineare le serie ad un

ordine di integrazione piuttosto che ad un'altro non dovrebbe portare a stime di β differenti, il fatto di aver trovato valori diversi fa sospettare che i dati abbiano bisogno di ulteriori trattamenti per un'analisi di cointegrazione più accurata. Ad esempio si potrebbe pensare che in un contesto di cointegrazione la modellizzazione ad alte frequenze non vada trascurata: ciò spiegherebbe anche il fatto di aver trovato diversi valori di $\hat{\beta}$ per diverse frequenze a parità di allineamento.

Approssimazione di Whittle

Si consideri il processo stazionario Gaussiano y_t , $t = 1, \dots, T$, con media nulla. La sua funzione di densità di probabilità è rappresentata dalla normale multivariata

$$(A.1) \quad f(y, \theta) = \frac{1}{(2\pi)^{T/2} |\Sigma(\theta)|^{1/2}} \exp \left\{ -\frac{1}{2} y' |\Sigma(\theta)|^{-1} y \right\},$$

dove $\Sigma(\theta)$ indica la matrice di covarianza, θ indica il vettore di parametri e $|\Sigma|$ indica il determinante della matrice. La funzione di verosimiglianza $\mathfrak{L}(y, \theta)$ è data dal logaritmo della formula (A.1), ovvero

$$(A.2) \quad \mathfrak{L}(y, \theta) = -\frac{T}{2} \log(2\pi) - \frac{1}{2} \log |\Sigma(\theta)| - \frac{1}{2} y' |\Sigma(\theta)| y.$$

La stima di massima verosimiglianza si ottiene andando a trovare quel particolare vettore di parametri $\hat{\theta}$ che rende massima il valore di $\mathfrak{L}(y, \theta)$; così facendo si rende massima la probabilità di osservare il particolare processo y_t dato $\hat{\theta}$. Una maniera ben nota per ridurre l'onere computazionale di tale massimizzazione è quella di ricorrere all'approssimazione di Whittle, ovvero andare a sostituire ai termini di $\mathfrak{L}(y, \theta)$ sopra descritti una loro approssimazione basata sul periodogramma del processo $I(\lambda)$, il quale può essere agevolmente calcolato mediante l'utilizzo della fast Fourier transform FFT. Per prima cosa si riscrive la funzione di verosimiglianza $\mathfrak{L}(y, \theta)$ come

$$(A.3) \quad \mathfrak{L}(y, \theta) = -\frac{1}{2T} \log |\Sigma(\theta)| - \frac{1}{2T} y' |\Sigma(\theta)| y.$$

Rispetto alla formula (A.2) è stato eliminato il primo termine (non incide sul risultato in quanto non dipende da θ o y_t) e diviso per T . A questo punto vengono fatte le seguenti approssimazioni: il primo termine della (A.3) è semplificato come

$$\frac{1}{2T} \log |\Sigma(\theta)| \approx \frac{1}{4\pi} \int_{-\pi}^{\pi} \log [2\pi f_{\theta, y}(\lambda)] d\lambda,$$

mentre il secondo come

$$\frac{1}{2T} y' |\Sigma(\theta)| y \approx \frac{1}{4\pi} \int_{-\pi}^{\pi} \frac{I(\lambda)}{f_{\theta, y}} d\lambda,$$

dove il periodogramma $I(\lambda)$ viene calcolato così come descritto nel capitolo 1, formula (2.6). L'unione di questi due termini porta all'approssimazione integrale di Whittle: discretizzando tali termini, si ottiene l'approssimazione discreta di Whittle, definita come

$$(A.4) \quad \mathfrak{L}(y, \theta) = -\frac{1}{2T} \left[\sum_{t=1}^T \log f_{\theta, y}(\lambda_t) + \sum_{t=1}^T \frac{I(\lambda)}{f_{\theta, y}} \right].$$

Si può notare come la funzione $Q(G, d)$ da minimizzare

$$(A.5) \quad Q(G, d) = \frac{1}{m} \sum_{j=1}^m \left\{ \log G \lambda_j^{-2d} + \frac{I_j}{G \lambda_j^{-2d}} \right\},$$

nell'ambito dello stimatore Gaussiano di Robinson altro non è che l'approssimazione discreta di Whittle cambiata di segno (per questo si minimizza), limitata alle prime m frequenze λ_j , $j = 1, \dots, m$ e m elementi del periodogramma I_j , e i cui elementi $f_\theta(\lambda)$ sono stati sostituiti da $G \lambda^{-2d}$. Resta da comprendere come si ricava la nuova e più maneggevole funzione $R(d)$

$$(A.6) \quad R(d) = \log \hat{G}(d) - 2d \frac{1}{m} \sum_{j=1}^m \log \lambda_j, \quad \hat{G}(d) = \frac{1}{m} \sum_{j=1}^m \lambda_j^{2d} I_j.$$

Ipotizzando che d sia noto, esso può essere trattato come un parametro, e pertanto la funzione $Q(G, d)$ diventa $Q(G)$, definita come

$$Q(G) = \frac{1}{m} \sum_{j=1}^m \log G - \frac{2d}{m} \sum_{j=1}^m \log \lambda_j + \frac{1}{m} \sum_{j=1}^m \frac{I_j}{G \lambda_j^{-2d}},$$

il cui minimo può essere ricavato nella maniera usuale come

$$\frac{dQ}{dG} = \frac{1}{m} \sum_{j=1}^m \frac{1}{G} - \frac{1}{m} \sum_{j=1}^m \frac{I_j(\lambda)}{G^2 \lambda_j^{-2d}} = 0.$$

L'unica soluzione di questa equazione è data da

$$\hat{G} = \frac{1}{m} \sum_{j=1}^m \frac{I_j(\lambda)}{\lambda_j^{-2d}},$$

in quanto l'altra soluzione $G = 0$ non appartiene all'intervallo $(0, \infty)$. Tornando a $Q(G, d)$ e sostituendo a G la soluzione \hat{G} , si ottiene

$$Q(d) = \log \hat{G}(d) - 2d \frac{1}{m} \sum_{j=1}^m \log \lambda_j + 1, \quad \hat{G}(d) = \frac{1}{m} \sum_{j=1}^m \lambda_j^{2d} I_j,$$

equivalente a $R(d)$ a meno di una traslazione (che non modifica il punto di minimo).

Bibliografia

- [1] Hosking, J. R. M. (1981). Fractional differencing. *Biometrika* **68**, 165-176.
- [2] Granger, C. W. J. & Joyeux, R. (1980). An introduction to long-memory time series models and fractional differencing. *Journal of Time Series Analysis* **1**, 15-29.
- [3] Robinson, P. M. (1995). Gaussian semiparametric estimation of long range dependence. *The annals of Statistics* **23**, 1630-1661.
- [4] Priestley, M. B. (1981). *Spectral analysis and time series*. Academic Press Inc., London.
- [5] Velasco, C. (1999). Gaussian semiparametric estimation of non stationary time serie. *Journal of Time Series Analysis* **20**, 87-127.
- [6] Cooley, J. W. & Tukey, J. W. (1964). *An algorithm for the machine calculation of complex Fourier series*. IBM Watson Research Center, Bell Telephone Laboratories & Princeton University.
- [7] Dieker, A. B. & Mandjes M. (2003). On spectral simulation of fractional brownian motion. *Probability in the Engineering and Informational Sciences* **17**, 417-434.
- [8] Hassler, U. (2011). Estimation of fractional integration under temporal aggregation. *Journal of Econometrics* **162**, 240-247.
- [9] Marinucci, D. & Robinson P.M. (2001). Semiparametric fractional cointegration analysis *Journal of Econometrics* **105** 225-247.