



UNIVERSITA' DEGLI STUDI DI PADOVA
FACOLTA' DI SCIENZE STATISTICHE

Corso di Laurea: SCIENZE STATISTICHE E GESTIONE DELLE IMPRESE

TESI DI LAUREA

**TRE TEST PER RADICI UNITARIE FRAZIONARIE:
UN ESPERIMENTO MONTE CARLO**

Three tests for fractional unit roots: a Monte Carlo experiment

RELATORE

Prof. LUISA BISAGLIA

LAUREANDO

MICHELE MARIANI

MATRICOLA

599210

ANNO ACCADEMICO
2010/2011

INDICE

INTRODUZIONE	1
CAPITOLO 1: I PROCESSI A MEMORIA LUNGA	
1.1. Origini, definizione e caratteristiche	2
1.2. I modelli ARFIMA (p,d,q): Descrizione e proprietà	5
CAPITOLO 2: I TEST DI RADICI UNITARIE	
2.1. Dall'utilizzo classico all'applicazione nei processi frazionari	10
2.2. I tre Test:	13
2.2.1 Il test efficiente di Wald	13
2.2.2. Il test Im di Robinson	17
2.2.3 Il test LR	19
CAPITOLO 3: UN ESPERIMENTO MONTE CARLO	
3.1. Le simulazioni, i risultati ed il confronto dei tre test	20
CONCLUSIONI	25
APPENDICE	26
RIFERIMENTI BIBLIOGRAFICI	30

INTRODUZIONE

Molte delle serie economico-finanziarie per dati reali sono non stazionarie in media.

Per eliminare tale non stazionarietà esistono diversi modi, tra cui uno dei principali è la differenziazione.

Molto spesso, però, il numero di volte per cui una serie temporale deve essere differenziata per divenire stazionaria viene identificato mediante metodi intuitivi e poco formali e per numerose serie storiche osservate considerare la differenza prima o seconda sembra essere eccessivo.

Granger e Joyeux (1980) e Hosking (1981) hanno così introdotto la classe dei più generali modelli a memoria lunga, tra cui la classe dei modelli ARFIMA o ARIMA frazionari. Per tali modelli il parametro di differenziazione, d , può assumere anche un valore reale. In questo caso, però, gli usuali test per radici unitarie sembrano avere bassa potenza quando viene considerata un'alternativa frazionaria e questo è ancor più vero se il valore di d è prossimo ad 1.

Lo scopo di questo lavoro è quello di confrontare la potenza di tre test per radici unitarie frazionarie in campioni finiti per determinare quale tra i tre ha più capacità discriminatorie tra l'ipotesi nulla di radice unitaria ed un'alternativa frazionaria. Questo studio è dunque strutturato come segue.

Nel primo capitolo sono descritti rapidamente i processi a memoria lunga ed introdotti i processi ARFIMA(p,d,q).

Nel secondo capitolo, dopo una breve esposizione sulle modalità e le motivazioni per cui, nella più recente letteratura, i test per radici unitarie frazionarie hanno assunto importanza, si descrivono i tre test che verranno confrontati: il test efficiente di Wald, il test LM di Robinson e il test LR.

Nel terzo capitolo sono riportati gli esiti di un breve esperimento Monte Carlo che sembrano decretare il test LM come il più potente tra i tre.

CAPITOLO 1: I PROCESSI A MEMORIA LUNGA

1.1 Origini, definizioni e caratteristiche

Hurst (1951), un idrologo inglese, lavorando negli anni '50 ad un progetto di una diga sul Nilo, nel tentativo di prevedere l'afflusso idrico e dunque riuscire a controllare le riserve di acqua, introdusse il concetto di memoria lunga. Solitamente quello che influenza maggiormente l'innalzamento del livello dell'acqua in una diga è provocato dalla quantità di pioggia caduta ed è consuetudine pensare che tale quantità segua un andamento casuale (dunque modellabile attraverso un Random Walk); Hurst volle verificare se era realmente così.

Mise a punto un nuovo strumento statistico denominato "esponente di Hurst (H)" che era in grado di distinguere una serie casuale da una non casuale anche se la serie casuale non era normalmente distribuita, a seconda del valore che l'esponente H assumeva. Hurst aveva ormai capito che ci sono molti fenomeni per cui le osservazioni non sono totalmente indipendenti tra loro, come in un fenomeno casuale, ma piuttosto che esiste, molto spesso, una dipendenza di lungo termine per cui ogni osservazione porta memoria degli eventi passati. Gli effetti di tali eventi continuano a farsi sentire, anche se in modo sempre più smorzato, sulle vicende future, anche se lontane nel tempo, e teoricamente all'infinito.

Negli anni '60, con Mandelbrot e con Wallis, il concetto di memoria lunga si è ampliato e sviluppato maggiormente, e soltanto all'inizio degli anni '80 questo tipo di modellistica ha cominciato a suscitare vero interesse anche tra gli studiosi di econometria.

Dopo questo breve excursus storico è opportuno passare in rassegna i comportamenti e le principali caratteristiche dei processi a memoria lunga. Un processo a memoria lunga viene utilizzato per modellare serie storiche con dipendenza di lungo periodo, cioè serie in cui ogni osservazione dipende in maniera significativa da quelle precedenti.

I processi a memoria lunga hanno caratteristiche ben precise, per questo è possibile riscontrarle nella maggior parte delle serie in cui è presente un effetto memoria. Una prerogativa comune di questo tipo di dati è che le osservazioni assumono valori o al di sopra o al di sotto della media per periodi lunghi. Si alternano infatti andamenti crescenti e decrescenti con frequenza bassa; a valori bassi della serie sembrano seguire valori bassi e viceversa a valori alti invece sembrano seguire valori alti. Questo comportamento, generalmente, non permette di individuare un trend univoco in tutta la serie dei dati, che è perciò stazionaria sia in media che in varianza; risulta anche difficile trovare cicli periodici. Si può notare, per questo tipo di serie, che la varianza della media campionaria converge a zero più lentamente di n^{-1} .

Per i processi a memoria lunga è essenziale studiare il comportamento della funzione di autocorrelazione globale $\rho(\cdot)$, capace di indicarci quanto le osservazioni ad un dato istante sono in relazione e vengono influenzate dalle osservazioni precedenti. Difatti, in questo tipo di procedimenti, le autocorrelazioni decadono molto lentamente a zero all'aumentare della distanza temporale tra le osservazioni a dimostrazione del fatto che c'è forte dipendenza tra queste ultime.

Anche le osservazioni più lontane nel tempo continuano ad influenzare significativamente le più attuali (ovviamente in maniera minore rispetto a quelle più recenti).

A questo punto è necessario dare una definizione più formale di processo a memoria lunga, ovvero:

Definizione

Se X_t è un processo stazionario per il quale esiste un numero reale $0 < \alpha < 1$ ed una costante $C_\rho > 0$ tale che:

$$\lim_{k \rightarrow \infty} \frac{\rho(k)}{C_\rho k^{-\alpha}} = 1$$

o alternativamente

$$\rho(k) \approx C_\rho k^{-\alpha} \quad k \rightarrow \infty$$

Allora X_t è chiamato processo stazionario a memoria lunga o con forte dipendenza o con dipendenza di lungo periodo.

Definizione: vedi Beran (1994)

Per i processi con dipendenza di lungo periodo

$$\sum_{k=-\infty}^{\infty} |\rho(k)| = \infty$$

la somma delle autocorrelazioni non deve essere finita; alcuni autori utilizzano questa proprietà come definizione di processo a memoria lunga perché chiara e di semplice utilizzo.

Per lo studio dei processi a memoria lunga i modelli più comuni sono gli ARFIMA(p,d,q); tuttavia prima di introdurre questa classe di modelli è necessario stabilire il legame esistente tra il parametro α , utilizzato nella precedente definizione di processo a memoria lunga (Beran (1994)), ed il parametro d, parametro di dipendenza che si utilizzerà nello studio dei processi ARFIMA(p,d,q).

La semplice relazione che intercorre tra di essi è data da: $\alpha = 1-2d$.

1.2. I modelli ARFIMA: Descrizione e proprietà

I modelli ARFIMA(p,d,q), AutoRegressive Fractional Integrated Moving Average, sono stati proposti la prima volta da Granger e Joyeux(1980) e Hosking(1981) con lo scopo di studiare quei casi in cui sussiste forte dipendenza tra le osservazioni. Proprio per analizzare queste circostanze, molto frequenti in ambito economico-finanziario, è necessaria una generalizzazione dei processi ARIMA(p,d,q), studiati anni prima da Box e Jenkins (1976), che erano capaci di cogliere solamente la dipendenza di breve periodo.

Nei nuovi modelli ARIMA frazionari si riesce invece ad esplicitare la componente di lungo periodo e la generalizzazione, rispetto agli ARIMA, sta nel fatto che il parametro d, il quale misura l'intensità del legame tra le osservazioni successive, può assumere un qualsiasi valore reale.

Infatti, con opportune generalizzazioni, partendo dall'equazione di un ARIMA(p,d,q) è possibile arrivare a definire un processo ARFIMA(p,d,q).

Formalmente, l'equazione di un ARIMA(p,d,q) assume questa forma:

$$\Phi(B)(1-B)^d X_t = \theta(B)\varepsilon_t$$

Osserviamo che, se d è intero, allora:

$$(1-B)^d = \Delta^d = \sum_{k=0}^d \binom{d}{k} (-1)^k B^k$$

dove il coefficiente binomiale è pari a:

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

e $\Gamma(\cdot)$ è la funzione gamma. Poiché la funzione gamma è definita per tutti i valori reali, la precedente definizione di coefficiente binomiale può essere estesa a $d \in \mathbb{R}$.

Formalmente:

$$(1-B)^d = \sum_{k=0}^{\infty} \binom{d}{k} (-1)^k B^k = \sum_{k=0}^{\infty} \pi_k B^k$$

dove:

$$\pi^k = \frac{\Gamma(k-d)}{\Gamma(k+1)\Gamma(-d)} = \prod_{0 < k \leq j} \frac{j-1-d}{j} \quad j = 0, 1, 2, \dots$$

Arrivati a questo punto, si può dare la seguente definizione di processo ARFIMA(p,d,q):

Definizione

$X = (X_t, t \in Z)$ è un processo ARFIMA(p,d,q) con $d \in (-1/2, 1/2)$ se è stazionario e se soddisfa l'equazione alle differenze:

$$\phi(B)\Delta^d X_t = \theta(B)\varepsilon_t$$

dove $\varepsilon \sim W.N.(0, \sigma^2)$ e $\phi(\cdot)$ e $\theta(\cdot)$ sono polinomi di grado p e q rispettivamente.

La funzione di autocorrelazione $\rho(\cdot)$, al ritardo h, di un processo ARFIMA(p,d,q), con $d \in (-1/2, 1/2)$, gode di questa proprietà:

$$\rho(h)h^{1-2d} \rightarrow c \quad h \rightarrow \infty$$

Questo significa che $\rho(h)$ converge a zero, quando $h \rightarrow \infty$, più lentamente rispetto all'autocorrelazione $\rho(h)$ di un processo ARMA(p,q). Infatti, la funzione di autocorrelazione $\rho(\cdot)$, al ritardo h, di un processo ARMA(p,q), per $h \rightarrow \infty$, converge rapidamente a zero, nel senso che per $r > 1$:

$$r^h \rho(h) \rightarrow 0 \quad h \rightarrow \infty$$

In dettaglio si veda: Brockwell e Davis (1996)

Il lento decadimento delle autocorrelazioni dei processi ARFIMA indica forte dipendenza tra le osservazioni successive, mentre il rapido decadimento a zero delle autocorrelazioni dei processi ARMA indica una dipendenza tra le osservazioni successive valida solamente nel breve periodo.

Per questo motivo, i processi ARFIMA sono utilizzati per lo studio dei processi a memoria lunga, al

contrario, i processi ARMA per quelli a memoria corta.

Di notevole importanza, per lo studio dei processi a memoria lunga, è il caso particolare dei modelli ARIMA frazionari dove p e q , numero di parametri rispettivamente della parte autoregressiva e della parte a media mobile, sono pari a zero.

Il processo in esame $X = (X_t, t \in Z)$, con $p=q=0$, viene denominato in letteratura “White Noise frazionario” ed indicato con l'annotazione $FI(d)$ o $ARFIMA(0,d,0)$.

Tale processo è una soluzione stazionaria con media zero dell'equazione delle differenze:

$$\Delta^d X_t = \varepsilon_t$$

dove $\varepsilon_t \sim W.N. (0, \sigma^2)$ per ogni $t \in Z$ e $\Delta^d = (1-B)^d$ è definito come in precedenza.

L'importanza del $FI(d)$ sta nel fatto che, pur nella sua semplicità, riesce a descrivere al meglio serie di dati che presentano un effetto memoria. Questo risulterà sicuramente più chiaro se si tiene conto della possibilità di scomporre un processo $ARFIMA(p,d,q)$ in due processi più semplici, ovvero:

- un $ARMA(p,q)$ che cattura la dipendenza di breve periodo
- un $ARFIMA(0,d,0)$ che descrive, invece, solamente la dipendenza di lungo periodo, se presente.

Più formalmente, questa scomposizione può essere espressa come segue:

- $\phi(B)Y_t = \theta(B)X_t \quad ARMA(p,q)$
- $\Delta^d X_t = \varepsilon_t \quad ARFIMA(0,d,0)$

Pertanto si può considerare un processo $ARFIMA$ come un processo “a due velocità”.

Per un processo $FI(d)$ vale il seguente Teorema:

Teorema

Se $d \in (-1/2, 1/2)$ allora esiste un'unica soluzione stazionaria X_t di data da:

$$X_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j} = \Delta^{-d} \varepsilon_t$$

dove $\psi_0 = 1$ e

$$\psi_j = \frac{\Gamma(j+d)}{\Gamma(j+1)\Gamma(d)} = \prod_{k=1}^j \frac{k-1+d}{k} \quad j = 0, 1, 2, \dots$$

Inoltre, la funzione di densità spettrale è:

$$f(\lambda) = \frac{\sigma^2}{2\pi} |2 \sin(\lambda/2)|^{-2d} \quad -\pi \leq \lambda \leq \pi$$

la varianza del processo è

$$\gamma(0) = \sigma^2 \frac{\Gamma(1-2d)}{\Gamma^2(1-d)}$$

e la funzione di autocorrelazione è:

$$\rho(k) = \frac{\Gamma(k+d)\Gamma(1-d)}{\Gamma(k-d+1)\Gamma(d)} = \prod_{j=1}^k \frac{j-1+d}{j-d} \quad k = 1, 2, \dots$$

Dimostrazione: Brockwell e Davis (1991)

Quando j tende ad ∞ , è possibile dimostrare che:

$$\psi_j \sim \frac{j^{d-1}}{\Gamma(d)}$$

$$\rho(k) \sim k^{2d-1} \frac{\Gamma(1-d)}{\Gamma(d)}$$

Sapendo che un ARFIMA(p,d,q) gode delle stesse proprietà asintotiche di un FI(d), si possono interpretare i seguenti risultati ottenuti per questo caso specifico, in un'ottica più generale, dal momento che, la differenza tra questi due processi è data solamente dalla dipendenza di breve periodo (processo ARMA), in dettaglio vedi Hosking (1981).

Allora, esaminando quest'ultima espressione di $\rho(k)$, si può cogliere l'andamento delle autocorrelazioni, al variare del parametro d , anche per un generale ARFIMA(p,d,q).

Quando $0 < d < 1/2$ il processo ARFIMA è un processo stazionario con memoria lunga infatti le sue autocorrelazioni decadono iperbolicamente (dunque lentamente) a zero.

Quando $-1/2 < d < 0$ il processo ARFIMA è un processo stazionario con memoria intermedia o dipendenza negativa o antipersistente, infatti le sue autocorrelazioni decadono più velocemente a zero rispetto al caso precedente e sono tutte negative.

Quando $d=0$ il processo ritorna ad essere un processo ARMA(p,q) che è un processo a memoria corta.

Per $-1/2 < d < 1/2$ il processo è stazionario ed invertibile, mentre per $d > 1/2$ il processo è non stazionario ma può essere reso tale attraverso un opportuno numero di differenziazioni.

CAPITOLO 2: TEST DI RADICI UNITARIE

2.1. Dall'utilizzo classico all'applicazione nei processi frazionari

Molte serie economiche e finanziarie sono non stazionarie e la causa principale di questo è, molto spesso, la presenza di un trend.

Nei dati temporali si possono osservare fondamentalmente due tipi di trend a seconda del modello generatore dei dati: trend deterministici, che sono funzioni non aleatorie del tempo, e stocastici, che sono invece funzioni aleatorie del tempo, per ulteriori dettagli si può vedere J.H.Stock e M.W.Watson (2005). Nella realtà questa distinzione non è così netta, ma spesso le due tipologie coesistono in una stessa serie di dati.

A seconda del tipo di trend esiste una diversa procedura per detrendizzare la serie: la regressione temporale con stima del trend per eliminare la componente deterministica e la differenziazione per quella stocastica. La procedura di regressione prevede una stima del trend tramite gli OLS e successivamente una rimozione dalla serie di tale componente, mentre la procedura di differenziazione consiste nel differenziare la serie sino a renderla stazionaria. Una serie differenziata d volte si dice essere integrata di ordine d ed è indicata con $I(d)$ (generalmente di ordine 1 per serie economico-finanziarie).

Una serie resa stazionaria tramite regressione è detta trend-stazionaria e si indica con $I(0)$, una serie resa stazionaria tramite differenziazione è invece chiamata differenza-stazionaria $I(1)$.

Il primo passo, ed anche il principale, da compiere per rimuovere questa non stazionarietà in media in maniera corretta, è riuscire dunque a determinare quale tipologia di trend è presente nei dati per poter conseguentemente applicare il giusto processo di detrendizzazione.

Per fare questo esistono dei test capaci di discriminare tra trend stocastico e stazionario: i test di radici unitaria, chiamati così perché sotto l'ipotesi nulla il polinomio che descrive il processo ha una radice pari all'unità.

I più utilizzati sono il test DF e DF aumentato (ADF) di Dickey D. e Fuller W. (1979), il test PP di Phillips P.B.C e Perron P. (1988) ed i più recenti test ERS di Elliot, Rotthenberg, Stock (1996) e di Ng e Perron (2001).

Tuttavia si è constatato che in molti test l'ipotesi nulla di radice unitaria viene respinta, suggerendo che molte serie temporali non sono ben rappresentate né come differenza-stazionarie I(1) né come trend-stazionarie I(0). Alla luce di questo risultato e basandosi sulla teoria economico-finanziaria si è notato che la classe degli ARIMA(p,d,q) frazionari, dove l'ordine di integrazione d è esteso ad un qualunque numero reale, è molto utile a catturare la proprietà di persistenza di molti processi a memoria lunga che però ovviamente necessitano dell'utilizzo di test di radici unitarie differenti rispetto a quelli di uso classico essendo differente il processo generatore dei dati (DGP).

Infatti per i test di radici unitaria di uso classico il processo generatore di dati è, nel caso più generale, un ARIMA(p,d,q), mentre per i test di radici unitaria applicati ai processi a memoria lunga il DGP che si considera è un ARIMA(p,d,q) frazionario.

Come visto nel precedente capitolo, per rappresentare un processo a memoria lunga si può considerare senza perdere alcuna informazione, anche da un punto di vista interpretativo, il caso più semplice dei processi ARFIMA: il processo frazionario integrato FI(d). Infatti recentemente sono stati sviluppati molti test di radici unitaria per processi frazionari che hanno come processo generatore di dati proprio un FI(d).

Questi test vengono utilizzati per discriminare tra l'ipotesi nulla di radice unitaria e l'alternativa frazionaria ovvero:

$$H_0 : d = 1 \tag{1}$$

contro l'alternativa semplice:

$$H_A : d = d_A$$

dove $d_A < 1$

o composta:

$$H_1 : d < 1$$

ed il DGP è appunto un FI(d) che, rispetto all'annotazione del precedente capitolo, verrà ora indicato con y_t

$$\Delta^d y_t 1\{t > 0\} = \varepsilon_t \quad t = 1, 2, \dots \quad (1)$$

dove gli $\varepsilon_t \sim \text{W.N.}(0, \sigma^2)$ con $\sigma^2 < \infty$ ed $1\{\}$ è la funzione indicatrice.

L'operatore di differenza frazionario:

$$\Delta^d = (1 - L)^d$$

può essere definito in termini dell'operatore ritardo L dall'espressione:

$$\Delta^\alpha := \sum_{i=0}^{\infty} \pi_i(\alpha) L^i$$

per ogni valore di α reale, dove

$$\pi_i(\alpha) = \frac{\Gamma(i - \alpha)}{\Gamma(i + 1)\Gamma(-\alpha)} \quad \alpha \neq 1, 2, \dots$$

e Γ è la funzione Gamma, con

$$\Gamma(0)/\Gamma(0) = 1$$

ed i primi due coefficienti pari a:

$$\pi_0(\alpha) = 1 \quad \pi_1(\alpha) = -\alpha$$

D'ora in avanti non si terrà più conto dell'annotazione (1) per i valori di t non positivi, assumendo

implicitamente che $y_t = \varepsilon_t = 0$ per $t \leq 0$.

In conclusione mentre nei test classici si discrimina tra I(1) ed I(0) in questo caso più generale si discrimina tra FI(1) e l'alternativa composita FI(d) con $d < 1$ (o anche l'alternativa semplice FI(d_A)).

2.2. I TRE TEST:

Introduciamo ora tre test per radici unitarie frazionarie che confronteremo mediante un esperimento Monte Carlo: il test efficiente di Wald, il test LM di Robinson ed il test LR.

Questi tre test sono localmente asintoticamente equivalenti e sono consistenti, vedi I.N.Lobato e Velasco C.(2007).

2.2.1 IL TEST EFFICIENTE DI WALD

Il test efficiente di Wald (Lobato e Velasco, 2007) si è rivelato molto utile per lo studio dei processi frazionari.

Come preannunciato, per semplicità, si consideri come DGP un FI(d) che, per ogni valore di d, è possibile riscrivere come:

$$\Delta y_t = (\Delta - \Delta^d)y_t + \varepsilon_t = (1 - \Delta^{d-1})\Delta y_t + \varepsilon_t \quad (2)$$

dove il termine di errore ε_t è i.i.d. e dove la variabile

$$(1 - \Delta^{d-1})\Delta y_t$$

non contiene Δy_t perchè:

$$(1 - \Delta^{d-1})\Delta y_t = (d-1)\Delta y_{t-1} + \sum_{j=2}^{t-1} \pi_j (d-1)\Delta y_{t-j} \quad (3)$$

In dettaglio si veda: Lobato e Velasco (2007).

L'equazione (2), tuttavia, può anche essere riscritta come:

$$\Delta y_t = \phi (\Delta^{d-1} - 1) \Delta y_t + \varepsilon_t \quad (4)$$

dove $\phi = 0$ sotto l'ipotesi nulla e $\phi = -1$ sotto l'alternativa.

L'equazione (4) suggerisce l'utilizzo del regressore

$$(\Delta^{d-1} - 1) \Delta y_t \quad (5)$$

ma bisogna notare che d non può assumere il valore $d=1$ altrimenti, come evidenziato dall'equazione (3), il regressore si annullerebbe.

Inoltre, un test statistico per la significatività di ϕ non può essere utilizzato per valori di $d > 1$, in quanto il segno dei coefficienti di (5) nell'equazione (4) cambierebbe a seconda del segno di $d-1$ (positivo o negativo). Spinti dalla volontà di superare questi problemi, Lobato e Velasco hanno proposto un nuovo modello di regressione, riscaldato di $1-d$ unità, ovvero:

$$\Delta y_t = \phi z_{t-1}(d) + \mu_t \quad t = 1, \dots, T \quad (6)$$

dove

$$z_{t-1}(d) = \frac{(\Delta^{d-1} - 1)}{1 - d} \Delta y_t \quad (7)$$

Il test applicato a questo modello è il test efficiente di Wald.

Il test efficiente di Wald verifica la significatività dei coefficienti di $z_{t-1}(d)$ con $d > 0.5$ nella (6) attraverso un t-test unilaterale denotato con t_ϕ .

Sotto l'ipotesi nulla, quando Δy_t è i.i.d., $\phi = 0$ e per ogni valore di d , si può considerare il più semplice $\mu_t = \varepsilon_t$ al posto del modello (6). Questo è possibile anche sotto l'ipotesi alternativa quando si considera il parametro di dipendenza, d , in corrispondenza del suo vero valore (non della stima).

Quando d , invece, è diverso dal suo valore reale, il modello ritorna nella sua forma standard, ovvero

quella espressa nell'equazione (6). La maggiore potenza del t-test si ha quando si massimizza la correlazione tra regressione e regressore cioè quando d corrisponde al suo valore reale; negli altri casi il test risulta consistente ma inefficiente (se confrontato con quello in cui d è selezionato in maniera ottimale).

Si può notare inoltre che il caso $d = 1$ produce un'indeterminazione $0/0$ nella (7) risolvibile con la regola di De l'Hopital e così facendo la regressione (6) può essere riscritta come:

$$\Delta y_t = \phi \sum_{j=1}^{t-1} j^{-1} \Delta y_{t-j} + \mu_t \quad (8)$$

Nonostante il t-test basato sull'equazione (8) sia asintoticamente localmente equivalente al t-test basato invece sull'equazione (6), rispetto un'alternativa fissata, si preferisce il test basato sulla (6).

Questo perché nella (8) non esistono valori di ϕ che rendono Δy_t i.i.d, infatti, per nessun valore il regressore

$$\sum_{j=1}^{t-1} j^{-1} \Delta y_{t-j}$$

massimizza la correlazione con Δy_t .

A questo punto introduciamo un Teorema che enuncia la distribuzione del test efficiente di Wald quando il processo, come in questo caso, è generato da un FI(d).

Teorema

Sotto l'assunzione che il DGP è:

$$\Delta^d y_t 1\{t > 0\} = \varepsilon_t$$

dove ε_t è i.i.d ed ha i primi quattro momenti finiti; le proprietà asintotiche del t-test $t\phi$ per testare $\phi=0$ nella (6), dove d^ di z_{t-1} soddisfa:*

$$d^* = d + o_p(T^{-\tau}) \quad \tau > 0 \quad d^* > 0.5$$

per alcuni valori fissati $d > 0.5$, soddisfano i seguenti enunciati:

- Sotto l'ipotesi nulla ($d=1$),

$$t_\phi \rightarrow_d N(0,1)$$

- Sotto un'alternativa fissata ($d < 1$), il test basato su t_ϕ è consistente.
- Sotto un'alternativa locale ($d = 1 - d/\sqrt{T}$, $d > 0$),

$$t_\phi \rightarrow_d N(-\delta h(d), 1)$$

dove

$$h(1) = \sqrt{\sum_{j=1}^{\infty} j^{-2}} = \sqrt{\pi^2 / 6}$$

e

$$h(d) = \frac{\sum_{j=1}^{\infty} j^{-1} \pi_j (d-1)}{\sqrt{\sum_{i=1}^{\infty} \pi_i (d-1)^2}} \quad d \neq 1 \quad d > 0.5$$

Dimostrazione: Appendice di Lobato e Velasco (2007)

Dunque sotto l'ipotesi nulla, per dei valori fissati $d > 0.5$, la distribuzione del test efficiente di Wald è una Normale standard.

E' da notare che il test efficiente Wald ha diverse analogie con il test Dickey-Fuller basato sul modello

$$\Delta y_t = \phi y_{t-1} + \mu_t \quad \phi = \rho - 1$$

dove ρ è il coefficiente di autocorrelazione. Infatti l'ipotesi nulla $\phi = 0$ ($d = 1$) coincide con $\rho = 1$ e

l'alternativa $\phi < 0$ ($d < 1$) con $\rho < 1$. Anche il test Dickey-Fuller si può classificare tra i test di tipo Wald per la relazione diretta che c'è tra il coefficiente della regressione ed il parametro di interesse.

2.2.2. IL TEST LM DI ROBINSON

Robinson nei suoi due articoli del 1991 e 1994 ha proposto un test dei Moltiplicatori di Lagrange (LM) per lo studio delle radici unitarie frazionarie, successivamente sviluppato ed analizzato da Tanaka (1999).

Per capire più specificamente la formulazione della statistica del test LM, si può tener conto che questo test è in realtà identico al test LBI (Locally Best Invariant) utilizzato per testare l'esistenza o meno di radici unitari in processi a media mobile, come dimostrato da Tanaka K.(1996). Considerare ed ottenere dei risultati per il test LBI quindi significa averli anche per il test LM.

Si consideri allora il DGP più semplice per poter applicare il test LBI:

$$z_j = x'_j B + y_j \quad (9)$$

$$(1 - L)^{d+\theta} y_j = \varepsilon_j \quad j = 1, \dots, T$$

dove x_j è una sequenza di variabili non stazionarie fissate opportunamente, y_j è distribuito come una $N(X\beta, \sigma^2 \Omega(\theta))$ dove $\Omega(\theta) = V(y)/\sigma^2$, β è un vettore di valori ignoti, ε_j sono dei processi stazionari (molto spesso per semplicità si considerano White Noise), d è un valore assegnato preventivamente, mentre θ è il parametro sul quale si vuole fare inferenza.

Calcolando la derivata della funzione di log-verosimiglianza (di parametri θ , β e σ^2) per il modello (9), rispetto al parametro θ , otteniamo un'espressione che nella sua forma compatta diventa:

dove ρ_k è l'autocorrelazione dei residui per il k -esimo ordine, per ulteriori dettagli si veda Tanaka

$$S_{T1} = \frac{1}{\sigma^2} \sum_{j=1}^T \left(\sum_{k=1}^{j-1} \frac{1}{k} \widehat{\varepsilon}_{j-k} \right) \widehat{\varepsilon}_j = T \sum_{k=1}^{T-1} \frac{1}{k} \widehat{\rho}_k$$

(1999). Il test LBI vuole testare $H_0: \theta = 0$ contro l'alternativa $H_1: \theta < 0$ (o $\theta > 0$) ed il test rifiuta l'ipotesi nulla per valori piccoli (grandi) di S_{T1} . Normalizzando S_{T1} (utilizzando l'informazione di Fisher) si ottiene la seguente espressione:

$$S'_{T1} = \frac{1}{\sqrt{T}} S_{T1} / \sqrt{\frac{\pi^2}{6}} = \sqrt{T} \sum_{k=1}^{T-1} \frac{1}{k} \widehat{\rho}_k / \sqrt{\frac{\pi^2}{6}} \quad (10)$$

che va poi confrontata con il percentile di una $N(0,1)$, vedi Tanaka (1999)

Considerando la formulazione della statistica del test LM di Robinson:

$$LM = T^{1/2} \left(\frac{\pi^2}{6} \right)^{-1/2} \sum_{j=1}^{T-1} \widehat{\rho}_{\Delta y}(j)$$

si può notare che essa è praticamente identica alla precedente espressione del test LBI.

L'unica differenza con la (10), infatti, è la funzione di autocorrelazione che si prende in considerazione. Mentre per il test LM è $\widehat{\rho}_{\Delta y}(j)$, cioè l'autocorrelazione semplice di ordine j di Δy_t , per il test LBI è $\widehat{\rho}_k$. Questa piccola discordanza è dovuta ai differenti processi generatori dei dati; per il test LBI è stato utilizzato il (9) mentre per il test LM si considera, come è stato fatto anche per il test efficiente di Wald, il DGP (1).

Il test di Robinson è, anch'esso, unilaterale e viene utilizzato per verificare l'esistenza di una radice unitaria contro un'alternativa frazionaria rispetto al parametro d . Anche il test LM, come il test di efficiente di Wald, rifiuta l'ipotesi nulla per valori piccoli della statistica test che è distribuita come una Normale standard.

Questo test, sotto le ipotesi di Gaussianità, è asintoticamente localmente ottimale.

2.2.3. IL TEST LR

Il test likelihood ratio (LR), sotto le assunzioni di Gaussianità, assume la forma:

$$LR = -2T \log \left(\frac{\sum (\Delta^{d_A} y_t)^2}{\sum (\Delta y_t)^2} \right)^{1/2}$$

Bisogna tener conto che sotto l'ipotesi alternativa composta H_1 , d_A viene sostituito con la stima di massima verosimiglianza di d .

In dettaglio si veda: Lobato e Velasco (2008)

Il test LR è anch'esso un test unilaterale ed asintoticamente presenta la stessa distribuzione nulla $N(0,1)$ del test efficiente di Wald e del test LM di Robinson.

CAPITOLO 3: UN ESPERIMENTO MONTE CARLO

3.1. Le simulazioni, i risultati ed il confronto dei tre test

Si è affermato che i tre test sono localmente asintoticamente equivalenti e sono consistenti, ma che cosa succede quando questi vengono utilizzati per lo studio di campioni finiti? Quale tra i tre test presi in considerazione risulta il più potente per discriminare l'ipotesi di radice unitaria contro un'alternativa frazionaria?

Per rispondere a questi quesiti ho condotto un esperimento Monte Carlo. Innanzitutto, è stato considerato come DGP per il test efficiente di Wald, il test LM di Robinson ed il test LR un processo $FI(d)$. Quest'esperimento è stato ripetuto per quattro diverse numerosità campionarie rispettivamente 100, 250, 500, 1000 e le simulazioni, per DGP, sono state 10000 (i comandi del software statistico R utilizzati sono stati riportati nell'Appendice). Nelle quattro tabelle che seguono sono riportate le percentuali di rifiuto dell'ipotesi nulla di radice unitaria dei tre test (per le quattro diverse numerosità campionarie) al variare di d tra 0,5 ed 1 con un incremento di 0,1 unità fino a $d = 0.9$. Tra 0,9 ed 1 invece sono riportati anche i valori $d = 0.95$ e $d = 0.975$ questo perché la probabilità di errore del secondo tipo (accettare l'ipotesi nulla quando questa è in realtà falsa) aumenta man mano che ci si avvicina ad 1, essendo appunto $d = 1$ l'ipotesi nulla. I valori di d tra 0,1 e 0,4 non sono riportati nelle tabelle poiché è stato verificato che i tre test rifiutano l'ipotesi nulla con probabilità 1. E' da notare che nelle tabelle i risultati sono espressi in percentuale per permettere una più facile interpretazione.

In corrispondenza di ogni tabella è presente un grafico esplicativo della funzione potenza empirica dei tre test: la linea verde per il test efficiente di Wald, la rossa per il test LM di Robinson, la blu per il test LR. In ascissa sono riportati i valori di d a partire da 0,1, mentre in ordinata sono riportati i valori ottenuti dalle simulazioni. Questi valori non sono più espressi in termini percentuali poiché la funzione potenza assume solamente valori compresi tra 0 ed 1.

Tabella I. T=100, S=10000

Test/d	0,5	0,6	0,7	0,8	0,9	0,95	0,975	1
Wald	100	98,98	89,65	69,36	28,98	18,29	14,02	11,29
LR	100	96,68	80,29	66,18	17,21	8,26	4,89	6,61
LM	100	98,96	90,24	86,6	24,71	11,97	7,29	4,7

Figura I.

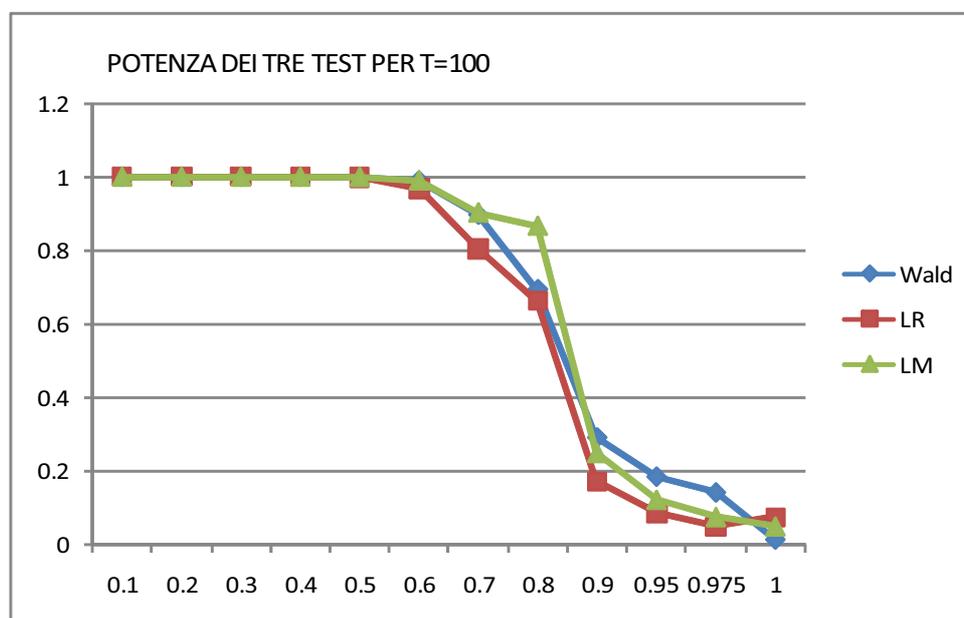


Tabella II. T=250, S=10000

Test/d	0,5	0,6	0,7	0,8	0,9	0,95	0,975	1
Wald	100	99,99	99,04	82,32	36,84	17,82	12,39	9,36
LR	100	99,8	92,28	61,13	21,11	7,7	3,8	4,26
LM	100	100	99,99	97,52	56,67	23,47	12,6	5,44

Figura II.

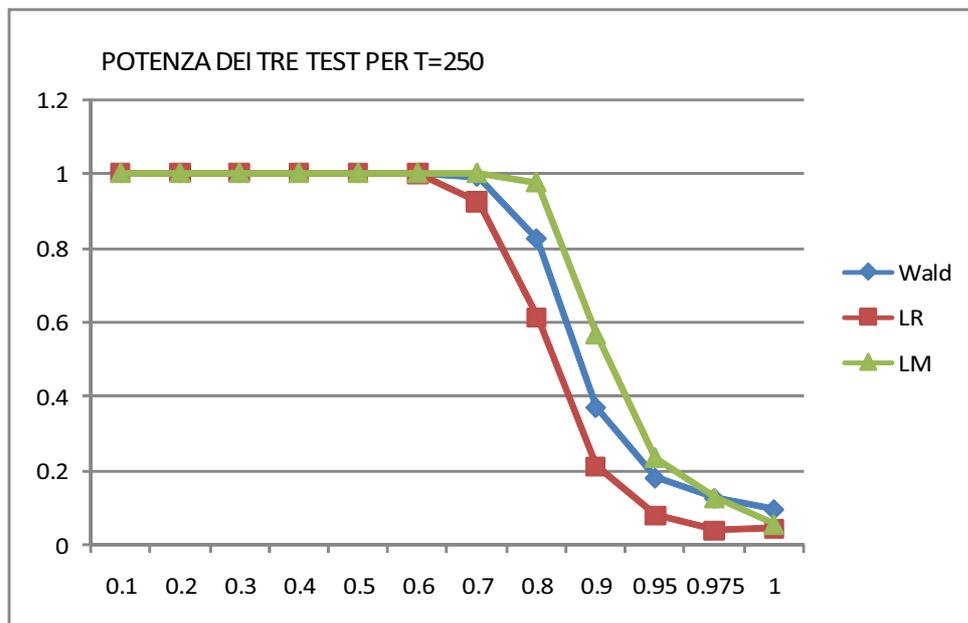


Tabella III. T=500, S=10000

Test/d	0,5	0,6	0,7	0,8	0,9	0,95	0,975	1
Wald	100	100	100	98,33	57,47	19,26	11,56	7,98
LR	100	100	99,29	76,45	26,67	8,26	3,05	3.1
LM	100	100	100	100	85,47	39,78	17,28	5.69

Figura III.

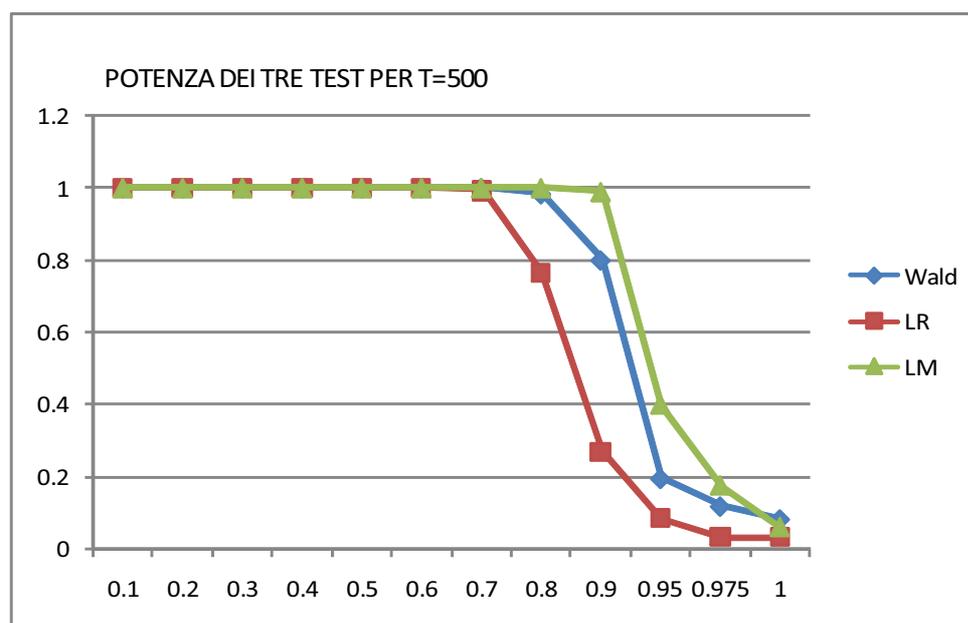
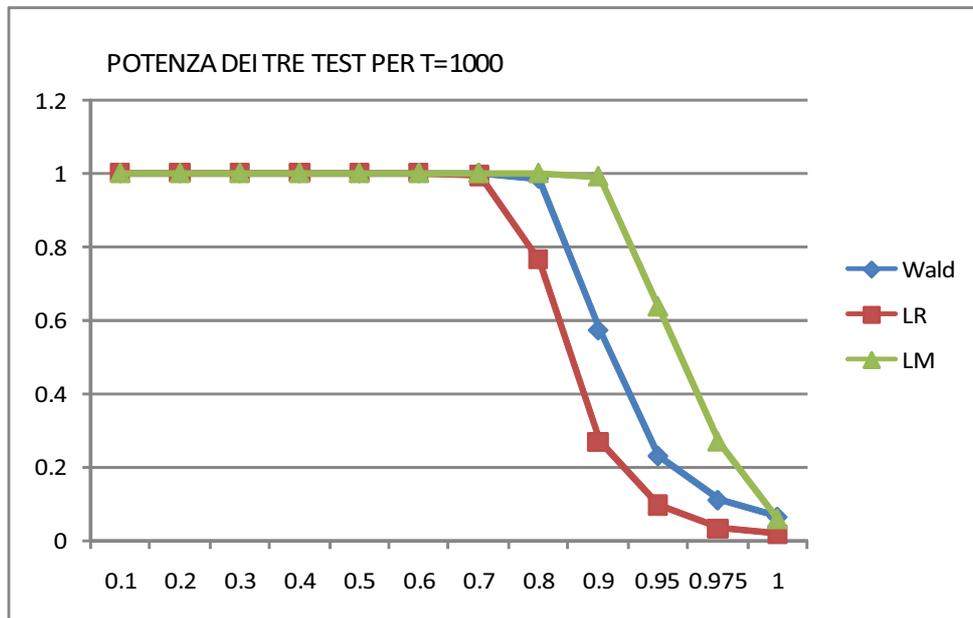


Tabella IV. T=1000, S=10000

Test/d	0,5	0,6	0,7	0,8	0,9	0,95	0,975	1
Wald	100	100	100	98,46	57,25	23,05	11,05	6,47
LR	100	100	99,33	76,4	26,76	9,55	2,94	1,61
LM	100	100	100	100	99,11	63,55	26,75	5,56

Figura IV.



Prima di commentare l'esito dell'esperimento, è bene ricordare che la potenza di un test, $1-\beta$, è la probabilità di rifiutare un'ipotesi nulla quando questa è falsa, ovvero la probabilità di non commettere un errore del secondo tipo, β (la probabilità di accettare un'ipotesi nulla quando in realtà è falsa). La potenza di un test, allora, sarà più alta quanto più piccolo sarà l'errore β del secondo tipo e viceversa, più dettagliatamente Azzalini (2001).

Il livello di significatività considerato in questo studio è pari per tutti e tre i test al 5% ovvero un $\alpha = 0.05$, dove α è la massima probabilità di errore del primo tipo (probabilità di rifiutare l'ipotesi nulla quando in realtà questa è vera).

Un test che rifiuterà l'ipotesi nulla, sbagliando, all'incirca nel 5% dei casi, sarà un test con capacità discriminatorie, tra ipotesi nulla ed alternativa, ottimali.

Alla luce di questo, se si analizzano i risultati, per la *size* ($d=1$), riportati nelle quattro tabelle, il test che si avvicina di più alla percentuale di rifiuto del 5% sembra essere il test LM di Robinson per tutte e quattro le numerosità campionarie prese in considerazione.

Infatti il test efficiente di Wald rifiuta più volte del previsto l'ipotesi nulla, anche se questa distorsione sembra diminuire all'aumentare del campione, il test LR invece per $T=250, 500$ e 1000 sembra rifiutare meno volte del previsto l'ipotesi nulla quando questa è in realtà corretta mentre i risultati per il test LM di Robinson si mantengono sempre intorno al 5%.

Per quanto riguarda i valori di $d < 1$, tutti e tre i test per $d \leq 0.5$ rifiutano H_0 nel 100% dei casi ma già a partire da $d=0.6$, l'alternativa frazionaria, sembra perdere validità contro l'ipotesi di radice unitaria, soprattutto per i campioni con minore numerosità. Anche in questo caso il test LM sembra essere il più potente arrivando a rifiutare H_0 per $d=0.9$ e $T=1000$ più del 99% delle volte (contro circa un 58% del test efficiente di Wald ed un 27% del test LR) perdendo l'ottima capacità di discriminare tra le due ipotesi solo per valori molto prossimi ad uno ($d=0.95$ e 0.975).

Quanto detto fino ad ora si può notare anche dall'analisi grafica, infatti, la linea verde della funzione potenza del test LM, tranne per i valori di d fino a 0.5 compreso, sovrasta sempre le funzioni potenza degli altri due test.

CONCLUSIONI

In questo lavoro il punto di partenza è stata una descrizione generale dei processi a memoria lunga dalle origini ai loro sviluppi. Il passo successivo è stato introdurre i processi ARIMA frazionari ed analizzare le principali motivazioni (una tra tutte è la capacità di questi processi di cogliere la dipendenza di lungo periodo) per cui questi processi hanno assunto notevole importanza, soprattutto nella più recente letteratura.

Al crescente affermarsi degli ARFIMA, di conseguenza, si è ampliato lo studio sui test di radici unitarie e sono diventati oggetto di interesse anche i più generali test per radici unitarie frazionarie, applicabili appunto quando il DGP è un ARIMA frazionario. Questa tesi è stata svolta con lo scopo di confrontare la potenza di tre test per radici unitarie frazionarie (attraverso la conduzione di un esperimento Monte Carlo), per campioni con numerosità finite, nel discriminare un' ipotesi nulla di radice unitaria contro un' alternativa frazionaria.

I tre test esaminati in questo lavoro sono: Il test efficiente di Wald, il test LM di Robinson ed il test LR e le conclusioni a cui siamo giunti dall'analisi dei risultati sono le seguenti.

Il test LM di Robinson sembra essere il test con più potere discriminatorio tra le due ipotesi (rispetto agli altri due test analizzati), sia per campioni con numerosità relativamente bassa che con numerosità alta.

Tuttavia, anche il test LM di Robinson, pur essendo il migliore tra i tre, non riesce del tutto a cogliere la differenza tra ipotesi nulla di radice unitaria ed alternativa frazionaria per valori molto prossimi all'unità. Il processo generatore dei dati, analizzato in questo lavoro è un processo frazionario integrato, poiché, nella sua semplicità, è capace di descrivere molto bene le dipendenze di lungo periodo.

Dunque un possibile sviluppo per lo studio di questi tre test, e non solo, potrebbe essere quello di considerare come DGP un qualsiasi ARFIMA(p,d,q) con un numero maggiore di parametri sia nella parte a media mobile che in quella autoregressiva rispetto al FI(d) esaminato in questo lavoro.

APPENDICE

COMANDI DI R UTILIZZATI PER LE SIMULAZIONI:

```
library(fracdiff) # Introduce la libreria fracdiff per poter utilizzare le
                  # funzioni di simulazione di una serie di dati generati da
                  # un processo ARIMA frazionario

T=100,250,500,1000 # Quattro numerosità campionarie (le utilizzo una alla
                  # volta)

S=10000 # Numero di simulazione da svolgere

val=-1.64 # valore della coda sinistra di una Normale standard per
          #  $\alpha = 0.05$ 

contw=0 # contatore per il test di Wald

contlm=0 # contatore per il test di Robinson

contrlr=0 # contatore per il test LR

for(i in 1:S){

# Devo lavorare separatamente sotto  $H_0$  e sotto  $H_1$ :

# Sotto  $H_1$  si simula un processo frazionario integrato con i valori di d che vengono fatti variare tra
# 0.5 e 0.975; viene utilizzato un d negativo e poi la funzione somma cumulata perchè il comando
# "fracdiff.sim" genera delle serie ARFIMA solo per  $-0.49 < d < 0.49$ . Si calcola inoltre la serie
# differenziata che risulterà utile per i tre test con il comando "diff"

x=fracdiff.sim(T, d= - 0.5(,-0.4,-0.3,-0.2,-0.1,-0.05, -0.025))$series

# simulo un FI(d) con d=- 0.5 ma devo farlo anche per i casi elencati tra parentesi

y=cumsum(x)

# somma cumulata del FI(d) con d=- 0.5 (-0.4,-0.3,-0.2,-0.1,-0.05,- 0.025) che diventa una
# FI(d) con d= 0.5 (0.6, 0.7, 0.8, 0.9, 0.95, 0.975)

deltay=diff(y)

# Sotto  $H_0$  al posto di un FI(d) si può prendere in considerazione un semplice processo Random
# Walk, creato dalla somma cumulata di un insieme di valori casuali generati da una Normale
```

standard. La serie differenziata, ora, coincide con i valori assunti dai vari epsilon.

```
eps <- rnorm(T+100)
```

```
y=ts(cumsum(eps))
```

```
y=y[101:(T+100)]
```

```
deltay=eps[101:(T+100)]
```

#TEST DI WALD:

Dopo aver stimato il valore del parametro d (fitd) tramite la funzione "whittle", si esegue una regressione lineare tra la serie differenziata ed il regressore "z" dell'espressione (7). Da questa regressione si ottiene un t-value che estraggo tramite il comando "summary.lm(lm(...))\$coefficient[\"z\", \"t value\"]" e successivamente si confronta con il valore di una Normale standard per $\alpha = 0.05$ (vedi "val"). Se il t value è minore di "val" significa che l'ipotesi nulla di radice unitaria viene rifiutata e dunque il contatore ("contw") del test di Wald aumenta, altrimenti resta invariato.

```
fitd=whittle(y)$d
```

```
z=(diffseries(y[2:T],fitd)-deltay)/(1-fitd)
```

```
tvalue=summary.lm(lm(deltay~z - 1))$coefficients[\"z\", \"t value\"]
```

```
if(tvalue<val){contw=contw+1} else {contw=contw}
```

#TEST LM

Si calcolano le funzioni di autocorrelazione per la serie y differenziata ("deltay"), si dividono per un vettore di $j-2$ elementi (a partire dal valore 1 con un incremento di 1) e si sommano; poi si moltiplica il tutto per la radice quadrata di T e per π -greco al quadrato diviso 6. In questo modo si ottiene l'espressione (11) del test LM. Il test funziona allo stesso modo di quello efficiente di Wald (questa volta il contatore è chiamato "contlm").

```
rho=acf(deltay,(T-1), plot=F)$acf[2:(T-1)]
```

```
j=seq(1,(T-2),1)
```

```
jrho=rho/j
```

```
somma_jrho=sum(jrho)
```

```
LM=sqrt(T)*(pi^2/6)^(-1/2)*somma_jrho
```

```
if(LM<val){contlm=contlm+1} else {contlm=contlm}
```

#TEST LR

Si differenzia la serie y d volte (tramite "diffseries") dove d è il parametro stimato tramite la funzione "whittle", si eleva al quadrato e si sommano i valori della serie differenziata ed elevata al quadrato ("somma1"). Si sommano i valori elevati al quadrato della serie y differenziata una volta ("deltay") ottenendo "somma2". Il test LR, come nell'espressione (12), è data dal rapporto dei due valori ottenuti "somma1" e "somma2". Il contatore "contlr" calcola la percentuale di rifiuto dell'ipotesi nulla per il test LR proprio come nei due casi precedenti.

```
fitd=whittle(y)$d

somma1=sum((diffseries(y,fitd)^2))

somma2=sum(deltay^2)

LR= 2*T*log(sqrt(somma1/somma2))

if(LR<val){contlr=contlr+1} else {contlr=contlr}

}
print((contw/S)*100)           # stampa la percentuale di rifiuto dell'ipotesi nulla per il
                              # test di Wald
print((contlr/S)*100)        # stampa la percentuale di rifiuto dell'ipotesi nulla per il
                              # test LR
print((contlm/S)*100)       # stampa la percentuale di rifiuto dell'ipotesi nulla per il
                              # test LM
```

FUNZIONE WHITTLE PER LA STIMA DEL PARAMETRO d

```
spettro<-function(t)
{
  tt<-trunc((t-1)/2)
  y<-array(1,tt)
  i<-c(1:t)

  for(j in 1:tt){
    y[j]<-abs(2*sin(pi*j/t))
  }
  return(y)
}

whittle<-function(x) {

result <- list()

d<-real

dhat<-real
```

```

t <- length(x)
FFT<-Mod(fft(x))^2/(2*pi*t)
period <- FFT[2:(t%/2)]
#period <- periodogram(x)
spec <- spetro(t)
q<-function(d) t(period)%*%spec^(2*d)
min<-optimize(q,c(0,1))
dhat<-min$minimum
sigma2 <- q(dhat)/(2*(2*pi)^2)
result$d <- dhat
result$sigma <- sigma2
return(result)
}

```

RIFERIMENTI BIBLIOGRAFICI

- Azzalini.A (2001). *Inferenza statistica: una presentazione basata sul concetto di verosimiglianza*. Milano:Springer
- Beran (1994). *Statistics for Long-Memory Processes*. New York: Chapman&Hall
- Box G.E.P., G.M. Jenkins (1976). *Time series analysis: forecasting and control*.San Francisco: Holden-Day.
- Brockwell P., Davis R. (1991). *Times Series: Theory and Methods*, New York: Springer
- Brocwell P.J., Davis R.A (1996). *Introduction to time series and forecasting*,.New York: Springer
- Dickey D.A, Fuller W.A. (1979). *Distribution of Estimators of Autoregressive Time Series with a Unit Root*. Journal of the American Statistical Association 74, 427-431
- Elliot G., Rothenberg T.J., Stock J.H. (1996). *Efficient Tests for an Autoregressive Unit Root*. Econometrica 64, 813-836
- Granger C., Joyeux J. (1980). *An Introduction to Long Memory Time Series Models and Fractional Differencing*. Journal of Time Series Analysis, 1, 15-30.
- J.H.Stock, M.W.Watson (2005). *Introduzione all'econometria*. Milano: Pearson
- Hosking J.R.M (1981). *Fractional Differencing*. Biometrika, 68, 165-176.
- Hurst H.E (1951). *Long term storage capacity of reservoir*. Transactions of the American Society of Civil Engineers 116, 770-799.
- Lobato I.N., Velasco C. (2007). *Efficient Wald tests for fractional unit roots*. Econometrica 75(2), 575-589
- Lobato I.N., Velasco C. (2008). *Power comparison among tests for fractional unit roots*. Economics Letters 99, 152-154
- Ng S., Perron P. (2001). *Lag Length Selection and the Construction of Unit Root Tets with Good Size and Power*. Econometrica 69, 1519-1554
- Phillips P.C.B., Perron P. (1988). *Testing for Unit Roots in Time Series Regression*. Biometrika 75, 355-346
- Robinson P.M. (1991). *Testing for strong serial correlation and dyanmic conditional heteroskedasticity in multiple regression*. Journal of Econometrics 47,67-84
- Robinson P.M. (1994). *Efficient tests of nonstationary hypotesis*. Journal of the American Statistical Association 89, 1420-1437

- Tanaka K. (1996). *Time series analysis: nonstationary and noninvertible distribution theory*. New York: JOHN WILEY&SONS,INC.
- Tanaka K. (1999). *The nonstationary fractional unit root*. *Econometric Theory* 15,249-264