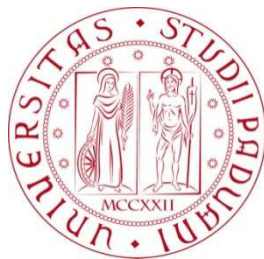


Università degli Studi di Padova
Dipartimento di Scienze Statistiche
Corso di Laurea Triennale in
Statistica e Gestione delle Imprese



RELAZIONE FINALE
UN TEST CUSUM PER MODELLI A MEMORIA LUNGA

Relatore Prof. Luisa Bisaglia
Dipartimento di Scienze Statistiche

Laureando: Valeria Sfragara
Matricola 1012420

Anno Accademico 2013/2014

Sommario:

Introduzione.....	5
1. Processi stazionari a memoria lunga.....	7
1.1 Cosa sono i processi a memoria lunga.....	7
1.2 Il modello ARFIMA.....	8
1.3 Il modello HAR di ordine infinito.....	10
2. Il test CUSUM.....	15
3. L'esperimento di Monte Carlo.....	17
3.1 Risultati.....	19
4. Conclusioni.....	27
Appendice: codice del software R.....	29
Bibliografia.....	35

Introduzione

Molte variabili economiche e finanziarie sono caratterizzate da memoria lunga. È, quindi, stato proposto da alcuni autori, tra cui Corsi (2009) e Hwang e Shin (2013), un modello autoregressivo eterogeneo (HAR) per affrontare questa caratteristica presente in molte serie storiche reali.

La memoria lunga è occasionalmente accompagnata da cambiamenti strutturali ed è stato dimostrato che spesso una parte della memoria lunga può essere causata in modo spurio ignorando dei piccoli cambiamenti di regime.

Alcuni test sono stati sviluppati per verificare l'ipotesi di break strutturali in presenza di memoria lunga. Se i break temporali non sono noti, viene spesso utilizzato un test CUSUM (CUmulative SUM).

In questo lavoro si intende applicare un test CUSUM per verificare l'ipotesi di costanza parametrica in un modello a memoria lunga. La distribuzione limite del test proposto è il massimo valore assoluto di un ponte Browniano standard che non contiene parametri di disturbo.

Tale analisi muove principalmente dall'articolo di Eunju Hwang e Dong Wan Shin "A CUSUM test for a long memory heterogeneous autoregressive model" (2013).

Questo studio è quindi articolato come segue. Nel primo capitolo si introdurrà il concetto di processo a memoria lunga, nello specifico verranno affrontati i modelli eterogenei autoregressivi *HAR* ed i modelli *ARFIMA* (AutoRegressive Fractional Integrated Moving Average). Nel secondo capitolo vengono introdotti i test CUSUM, in particolare verrà trattato il test CUSUM basato sui residui ai minimi quadrati ordinari (OLS). Utilizzando tale test nel terzo capitolo verrà formulato un esperimento Monte Carlo, i cui risultati permetteranno di osservare il comportamento empirico del livello e della potenza del test, costruito a partire da una serie storica a memoria lunga.

1. Processi stazionari a memoria lunga

1.1 Cosa sono i processi a memoria lunga

In questo capitolo vengono descritti alcuni processi stocastici che possono essere utilizzati per modellare dati con dipendenza di lungo periodo.

Le serie temporali generate da processi a memoria lunga presentano delle caratteristiche ben precise. In primo luogo le osservazioni tendono ad assumere valori al di sopra o al di sotto della media, per periodi di tempo relativamente lunghi: valori elevati tendono ad essere seguiti da valori elevati, mentre valori bassi sono tendenzialmente seguiti da valori bassi. Inoltre, nelle serie sono presenti cicli non periodici in successione casuale; risulta, infatti, evidente l'assenza di trend persistenti o di cicli se si guarda alla serie intera. Le serie, poi, appaiono complessivamente stazionarie sia in media che in varianza, ma la varianza della media campionaria converge a zero più lentamente di $1/n$.

È inoltre rilevante notare che la funzione di autocorrelazione (autocovarianza) converge molto lentamente a zero; ciò implica che la dipendenza tra osservazioni successive decresce lentamente all'aumentare della distanza temporale tra le stesse, ovvero decresce con un andamento iperbolico e non esponenziale.

Se si assume che il processo osservato sia stazionario, è possibile definire un processo a memoria lunga o con dipendenza a lungo periodo o dipendente in senso forte, nel seguente modo:

Definizione 1.1.1 *Il processo $X = (X_t, t \in \mathbf{Z})$ viene chiamato processo stazionario con memoria lunga o dipendenza di lungo periodo, se è stazionario e se esiste un numero reale $d \in (-1/2, 1/2)$ e una funzione $L_{\rho(k)}$ che varia lentamente all'infinito, tale che:*

$$\rho(k) \sim L_{\rho(k)} k^{2d-1},$$

quando k tende a ∞ , dove $\rho(k)$ è la funzione di autocorrelazione.

Si osserva che la funzione di autocorrelazione $\rho(k)$ converge a zero così lentamente che $\sum_{k=-\infty}^{\infty} \rho(k)$ diverge.

Esistono diverse classi di modelli per cui sono soddisfatte le condizioni date dalle definizioni precedenti. In quest'elaborato ne verranno trattate solamente due: i modelli *ARFIMA* ed i modelli *HAR* di ordine infinito.

1.2 I modelli ARFIMA

I modelli *ARFIMA*(p, d, q), detti anche *ARIMA*(p, d, q) frazionari, sono stati introdotti indipendentemente da Granger e Joyeux (1980) e da Hosking (1981) come una generalizzazione dei modelli *ARMA*(p, d, q) per descrivere la dinamica di fenomeni che presentano una struttura di dipendenza che persiste nel tempo.

In pratica si verifica spesso che molte serie siano non stazionarie, Box e Jenkins (1976), quindi, hanno proposto un'estensione dei processi *ARMA*(p, q) che studiano questo tipo di serie. Considerano, cioè, processi omogenei non stazionari di grado d , che possono essere resi stazionari attraverso opportune trasformazioni. Se X_t è la d -esima differenza di Y_t , $X_t = (1 - B)^d Y_t$, con $d \in \mathbf{N}$, diremo che Y_t è processo integrato di ordine d e lo denoteremo come *ARIMA*(p, d, q), d viene chiamato parametro di integrazione o di differenziazione. L'equazione corrispondente è perciò:

$$\Phi(B)(1 - B)^d X_t = \theta(B)\epsilon_t \quad (1.2.1)$$

dove $\epsilon_t \sim w.n.(0, \sigma^2)$ e $\Phi(\cdot)$ e $\theta(\cdot)$ sono polinomi di grado p e q rispettivamente¹.

Si noti che i processi *ARMA*(p, q) possono essere pensati come *ARIMA*($p, 0, q$). I modelli *ARFIMA*(p, d, q) generalizzano i modelli *ARMA*(p, q) in quanto il parametro d può assumere un qualsiasi valore reale. In tal caso è possibile

¹ $\Phi(\cdot)$ e $\theta(\cdot)$ sono due polinomi di grado p e q rispettivamente, tali che:

$$\begin{aligned} \Phi(B) &= 1 - \phi_1 B - \dots - \phi_p B^p, \\ \theta(B) &= 1 + \theta_1 B + \dots + \theta_q B^q \end{aligned}$$

e B , l'operatore ritardo, è tale che $B^j X_t = X_{t-j}, j = 0, 1, \dots$

generalizzare l'equazione (1.2.1) nel senso seguente. Dapprima si osservi che se d è intero, allora:

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

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 valori reali, la precedente definizione del coefficiente binomiale può essere estesa a $d \in \mathbf{R}$, come:

$$(1 - B)^d = \sum_{k=0}^{\infty} \binom{d}{k} (-1)^k B^k = \sum_{k=0}^{\infty} \pi_k B^k, \quad (1.2.2)$$

dove:

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

A questo punto si può definire un processo $ARFIMA(p, d, q)$ come segue:

Definizione 1.2.1 $Y = (Y_t, t \in \mathbf{Z})$ è un processo $ARFIMA(p, d, q)$ con $d \in (-1/2, 1/2)$, se è stazionario² e se soddisfa l'equazione alle differenze:

$$\Phi(B)\Delta^d Y_t = \theta(B)\epsilon_t \quad (1.2.3)$$

dove $\epsilon_t \sim w.n.(0, \sigma^2)$ dove $\Phi(\cdot)$ e $\theta(\cdot)$ sono polinomi di grado p e q rispettivamente.

Se $p = q = 0$, il processo $Y = (Y_t, t \in \mathbf{Z})$ è spesso chiamato rumore integrato frazionario e viene denotato come $ARFIMA(0, d, 0)$.

Per tale processo vale il seguente teorema:

Teorema 1.2.1 Se $d \in (-1/2, 1/2)$ allora esiste un'unica soluzione stazionaria Y_t di (1.2.4) data da:

$$Y_t = \sum_{j=0}^{\infty} \psi_j \epsilon_{t-j} = \Delta^d \epsilon_t$$

dove $\psi_0 = 1$ e

²Il processo è stazionario se le radici di $\Phi(B) = 0$ cadono fuori del circolo unitario.

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

La varianza del processo è:

$$\gamma_0 = \sigma^2 \frac{\Gamma(1-2d)}{\Gamma^2(1-d)}$$

e la funzione di autocorrelazione è:

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

Quando $d \in (0, 1/2)$ il processo $ARFIMA(p, d, q)$ è un processo stazionario con memoria lunga. Quando $d = 0$ si ottiene nuovamente l'usuale modello $ARMA(p, q)$. Quando $d \in (-1/2, 0)$ il processo è stazionario con memoria intermedia o dipendenza negativa o antipersistente. Il parametro d , quindi, misura l'intensità della dipendenza. In questo elaborato si farà riferimento solo al caso in cui $d \in (0, 1/2)$, cioè al caso di memoria lunga in senso stretto.

1.3 I modelli HAR di ordine infinito

Il modello eterogeneo autoregressivo (HAR) è stato proposto da Corsi (2004) come un'alternativa per modellare e prevedere la volatilità realizzata, e si ispira al modello $HARCH$ di Müller et al. (1997) e Dacorogna et al. (1998).

Il modello $HAR(3)$ proposto da Corsi (2009) è $Y_t = \beta_0 + \sum_{j=1}^p \beta_j Y_{t,h_j} + a_t$, con $p = 3$ e dove a_t è una successione di errori di regressione,

$$Y_{t,h_j} = \frac{1}{h_j} (Y_{t-1} + Y_{t-2} + \dots + Y_{t-h_j}), \quad j = 1, 2, \dots$$

e $h_1 = 1, h_2 = 5, h_3 = 22$. È possibile notare che Y_{t,h_j} con $j = 2, 3$ sono, rispettivamente, medie mobili settimanali e mensili. Il modello $HAR(3)$ è, tuttavia, un modello a memoria corta $AR(22)$ e non è un vero e proprio modello a memoria lunga. Perciò, per ottenere un processo a memoria lunga, Hwang e Shin (2013) hanno proposto un modello eterogeneo autoregressivo di ordine infinito, $HAR(\infty)$.

Un'estensione del modello $HAR(3)$ proposto da Corsi(2009) è il modello $HAR(\infty)$ proposto da Hwang e Shin (2013) e definito come:

$$Y_t = \beta_0 + \beta_1 Y_{t,h_1} + \beta_2 Y_{t,h_2} + \dots + \epsilon_t, \quad (1.3.1)$$

dove $\{\beta_j: j = 0, 1, 2, \dots\}$ è una successione di numeri reali che tende a zero, $\{h_j: j = 1, 2, \dots\}$ è una successione data di numeri interi positivi che cresce verso infinito e $\{\epsilon_j\}$ è una successione di variabili casuali i.i.d. con media zero e varianza $E[\epsilon_t^2] = \sigma^2$.

Si può notare che Y_t al punto (1.3.1) è un processo $AR(\infty)$:

$$\begin{aligned} Y_t &= \beta_0 + \sum_{j=1}^{\infty} \frac{\beta_j}{h_j} (Y_{t-1} + Y_{t-2} + \dots + Y_{t-h_j}) + \epsilon_t \\ &= \beta_0 + \sum_{j=1}^{\infty} \alpha_j \left(\sum_{k=h_{j-1}+1}^{h_j} Y_{t-k} \right) + \epsilon_t = \beta_0 + \sum_{i=1}^{\infty} \phi_i Y_{t-i} + \epsilon_t \end{aligned}$$

dove $\alpha_j = \sum_{k=j}^{\infty} \beta_k / h_k$ con $j = 1, 2, \dots$; $\phi_1 = \alpha_1$, $\phi_{h_j+r} = \alpha_{j+1}$ con $r = 1, 2, \dots, h_{j+1} - h_j$ e $h_0 = 0$.

Affinché Y_t sia un processo stazionario, sono necessarie le seguenti condizioni:

(A1.3.1) (a) Il polinomio

$$A(z) = 1 - \sum_{i=1}^{\infty} \phi_i z^i$$

si allontana da zero per $|z| \leq 1$,

(b) $\sum_{j=1}^{\infty} |\beta_j| < \infty$,

(c) $\{\epsilon_j\}$ è una successione di variabili casuali i.i.d con media zero e varianza σ^2 .

Teorema 1.3.1 *Assumendo che la condizione (A1.3.1) sia rispettata, Y_t è stazionario ed ha una rappresentazione a media mobile infinita unilaterale*

$$Y_t = \sum_{k=0}^{\infty} \xi_k \epsilon_{t-k},$$

dove i pesi ξ_i sono calcolati ricorsivamente come $\xi_0 = 1$, $\xi_k = \sum_{l=0}^{k-1} \xi_l \phi_{k-l}$ con $k = 1, 2, \dots$. Inoltre i valori ξ_k sono assolutamente sommabili.

Per definire la proprietà di memoria lunga della serie Y_t , è necessaria la seguente condizione:

(A1.3.2) Per una generica costante c ,

$$\beta_j \sim c\lambda^j \text{ per } |\lambda| < 1 \text{ e } h_j \sim cv^j \text{ per } v > 1.$$

Teorema 1.3.2 Sotto le assunzioni (A1.3.1) e (A1.3.2), con $\rho = \frac{(\log v - \log |\lambda|)}{\log v} > 1$ e una generica costante c , si ha

(i) $|\xi_k| \sim ck^{-\rho}$ e

(ii) $\gamma_k = \text{Cov}(Y_t, Y_{t+k}) = \sigma^2 \sum_{i=0}^{\infty} \xi_i \xi_{i+k} \sim ck^{-\rho}$.

Teorema 1.3.3 Si assuma la condizione (A1.3.1) e sia $h_j \sim cv^j$ con $v > 1$ e $|\xi_k| \sim ck^{-\rho}$ con $\rho > 1$. Allora $\beta_j \sim c\lambda^j$ con $\lambda = v^{1-\rho} \in (0, 1)$, dove ξ_k è definito alla proposizione 1.3.1.

Attraverso gli ultimi due teoremi citati, viene mostrato che, affinché un modello HAR(∞) sia a memoria lunga, una condizione necessaria e sufficiente è che i coefficienti nel modello decrescano esponenzialmente.

Si consideri, ora, il modello (1.3.1), dove gli h_j sono dati e i β_j non sono noti.

Per stimare il modello, si adatta un modello HAR di ordine finito come segue:

$$Y_t = \beta_0 + \sum_{j=1}^{\infty} \beta_j Y_{t,h_j} + a_t \quad (1.3.2)$$

con p che dipende da n e dove $a_t(p) = \epsilon_t + \sum_{j=p+1}^{\infty} \beta_j Y_{t,h_j} \equiv \epsilon_t + \eta_t(p)$.

Supponendo che i dati siano definiti come segue: $\{Y_{h_{p+1}}, \dots, Y_0, Y_1, Y_2, \dots, Y_n\}$, una forma matriciale dell'espressione (1.3.2) è data da:

$$Y = \mathbf{X}(p)\beta(p) + a(p),$$

dove $Y = (Y_1, \dots, Y_n)'$, $\mathbf{X}(p) = (X_1(p), \dots, X_n(p))'$ ³, $\beta(p) = (\beta_0, \beta_1, \beta_2, \dots, \beta_p)'$ e $a(p) = (a_1(p), \dots, a_n(p))'$. La stima a minimi quadrati ordinari (OLSE) di $\beta(p)$ è data da

$$\hat{\beta}(p) = (\mathbf{X}(p)'\mathbf{X}(p))^{-1}\mathbf{X}(p)'Y = \hat{\mathbf{R}}(p)^{-1}\hat{\mathbf{r}}(p),$$

dove $\hat{\mathbf{R}}(p) = \frac{1}{n} \sum_{t=1}^n X_t(p) X_t(p)'$ e $\hat{\mathbf{r}}(p) = \frac{1}{n} \sum_{t=1}^n X_t(p) Y_t$.

³ Dove gli $X_t(p)$ sono definiti come $X_t(p) = (1, Y_{t,h_1}, Y_{t,h_2}, \dots, Y_{t,h_p})'$.

Sia $\mathbf{R}(p) = E[X_t(p)X_t(p)']$ e sia $\mathbf{r}(p) = E[X_t(p)Y_t]$; allora, sotto l'assunzione (A1.3.1), che implica la stazionarietà di Y_t , segue che $\sup_{p \geq 1} \|\mathbf{R}(p)\| < \infty$ e $\sup_{p \geq 1} \|\mathbf{R}(p)^{-1}\| < \infty$.

Oltre alle assunzioni (A1.3.1) e (A1.3.2), per provare che la stima OLS è consistente e normale asintoticamente, è necessaria un'altra condizione sul momento di ϵ_t :

$$(A1.3.3) \ E|\epsilon_t|^{q+\delta} < \infty \text{ per } q > 2 \text{ e } 0 < \delta < 1.$$

Attraverso i seguenti teoremi si mostreranno i vincoli ai momenti della matrice di autocovarianza e la sua inversa (Teorema 1.3.4 e 1.3.5), e la consistenza e la normalità asintotica della stima OLS (Teorema 1.3.6 e 1.3.7).

Teorema 1.3.4 *Si assumano le condizioni (A1.3.1)-(A1.3.3). Se $\rho > q(q + \delta)/(2\delta)$, allora per qualche $C > 0$ che non dipende da n ,*

$$E\|\widehat{\mathbf{R}}(p) - \mathbf{R}(p)\|^q \leq C \left(\frac{p^2}{n}\right)^{q/2}$$

e

$$E\|\widehat{\mathbf{r}}(p) - \mathbf{r}(p)\|^q \leq C \left(\frac{p^2}{n}\right)^{q/2}.$$

Teorema 1.3.5 *Si assumano le condizioni (A1.3.1)-(A1.3.3). Se $\rho > q(q + \delta)/(2\delta)$ e $p^{2+\varepsilon} = O(n)$ per $\varepsilon > 0$, allora per qualche $C > 0$ che non dipende da n ,*

$$E\|\widehat{\mathbf{R}}(p)^{-1}\|^q \leq C$$

e

$$E\|\widehat{\mathbf{R}}(p)^{-1} - \mathbf{R}(p)^{-1}\|^{\frac{q}{2}} \leq C \left(\frac{p^2}{n}\right)^{\frac{q}{4}}.$$

Inoltre, se $p^3/n \rightarrow 0$, allora

$$p^{1/2}\|\widehat{\mathbf{R}}(p)^{-1} - \mathbf{R}(p)^{-1}\| \xrightarrow{p} 0$$

con $n \rightarrow \infty$.

Teorema 1.3.6 *Si assumano le condizioni (A1.3.1)-(A1.3.3). Se $\rho > q(q + \delta)/(2\delta)$ e $p^{2+\varepsilon} = O(n)$ per $\varepsilon > 0$,*

$$\|\widehat{\beta}(p) - \beta(p)\| \xrightarrow{p} 0$$

con $n \rightarrow \infty$, e quindi $\widehat{\beta}(p)$ è uno stimatore consistente di β .

Teorema 1.3.7 *Si assumano le condizioni (A1.3.1)-(A1.3.3). Si assuma, inoltre, che $E|\epsilon_t|^{2q+2\delta} < \infty$, $\rho > q(q + \delta)/(2\delta)$ e $p = o(n^{1/3})$. Sia $\ell(p)$ un vettore $(p + 1) \times 1$ tale che $\|\ell(p)\| = O(p)$ e*

$$\lim_{n \rightarrow \infty} \ell(p)' \mathbf{R}(p)^{-1} \ell(p) =: \ell \text{ esista}$$

Allora

$$\sqrt{n} \left[\ell(p)' \left(\hat{\beta}(p) - \beta(p) \right) \right] \xrightarrow{d} n(0, \sigma^2 \ell)$$

con $n \rightarrow \infty$.

I risultati ottenuti sopra convalidano l'inferenza statistica sulla stima dei parametri nei modelli $HAR(p)$.

2. Il test CUSUM

Una delle questioni più importanti della letteratura econometrica è verificare l'ipotesi di cambiamenti strutturali in un modello di regressione.

Sono stati proposti dei test nel caso in cui i punti di break siano in un momento noto. Sfortunatamente nella maggior parte delle situazioni pratiche questo non è conosciuto, si utilizza, quindi, spesso un test CUSUM per studiare i cambiamenti strutturali. Sono stati sviluppati alcuni test per verificare l'ipotesi di costanza parametrica in presenza di memoria lunga. Brown, Durbin ed Evans (1975) proposero un test CUSUM basato sui residui ricorsivi per verificare l'ipotesi di costanza nel tempo dei coefficienti di un modello di regressione lineare che tende in distribuzione ad un processo di Wiener standard. Ploberger e Krämer (1992) invece, hanno studiato un test CUSUM basato sui residui ai minimi quadrati ordinari (OLS), derivando il limite della statistica CUSUM a un ponte Browniano standard. In questo elaborato si prenderà in considerazione soltanto l'ultimo test citato.

Il modello di regressione considerato è il seguente:

$$y_t = x_t' \beta_t + u_t \quad (t = 1, \dots, T),$$

dove al tempo t , y è l'osservazione della variabile dipendente, $x_t = [1, x_{t2}, \dots, x_{tk}]' = [1, \tilde{x}_t']'$ è un vettore $K \times 1$ delle osservazioni delle variabili indipendenti, con il primo elemento pari ad uno, u sono i disturbi e β è il vettore $K \times 1$ dei coefficienti di regressione. Si assume che gli errori siano indipendenti e identicamente distribuiti (non necessariamente come una normale), con media zero e varianza uguale ad una costante, σ^2 , per ogni t . Inoltre, si assume che il vettore dei parametri possa variare nel tempo. L'ipotesi nulla, H_0 , è che i coefficienti del modello rimangano costanti nel tempo, quindi $\beta_t = \beta$ con $t = 1, \dots, T$.

La distribuzione nulla del test è derivata sotto le seguenti assunzioni:

Assunzione (A2.1) I regressori x_t ed i disturbi u_t sono definiti su uno spazio di probabilità comune, tale che

$$\lim_{T \rightarrow \infty} \sup \frac{1}{T} \sum_{t=1}^T \|x_t\|^{2+\delta} < \infty$$

con $\delta > 0$ (e dove $\|\cdot\|$ indica la norma Euclidea).

Assunzione (A2.2)

$$\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T x_t x_t' = R$$

con R matrice ($K \times K$) non singolare e non stocastica.

Assunzione (A2.3) Gli errori u_t sono stazionari ed ergodici, con

$$E(u_t | \mathfrak{U}_t) = 0, \quad E(u_t^2 | \mathfrak{U}_t) = \sigma^2,$$

dove \mathfrak{U}_t è il campo di esistenza di σ generato da $\{y_{t-s}, x_{t-s+1}, u_{t-s} | s \geq 1\}$.

Il test CUSUM è basato sulla somma cumulata dei residui OLS

$$\hat{u}_t^{(T)} = y_t - x_t' \hat{\beta}^{(T)}, \quad (2.1)$$

dove $\hat{\beta}^{(T)} = (\sum_{t=1}^T x_t x_t')^{-1} \sum_{t=1}^T x_t y_t$ è la stima ai minimi quadrati ordinari dei coefficienti. I residui OLS spesso cambiano quando un'altra osservazione viene aggiunta al campione. Inoltre anche sotto H_0 essi sono correlati ed eteroschedastici.

La statistica test è definita da

$$\sup_{0 \leq z \leq 1} |B^{(T)}(z)| \quad (2.2)$$

dove $B^{(T)}(z)$ è la somma cumulata dei residui OLS, citata sopra, ed è definita come:

$$B^{(T)}(z) = \frac{1}{\hat{\sigma} \sqrt{T}} \sum_{t=1}^{[Tz]} \hat{u}_t^{(T)}$$

con $0 \leq z \leq 1$ e dove $\hat{\sigma} = \left(\frac{1}{T} \sum_{t=1}^T \hat{u}_t^2\right)^{1/2}$.

Grandi valori di questo test portano al rifiuto dell'ipotesi nulla, H_0 , ovvero della costanza parametrica. Il seguente teorema indica che valori critici asintoticamente possono essere ottenuti da un ponte Browniano standard.

Teorema 2.1 Sotto H_0 , e date le Assunzioni (A1)-(A3),

$$B^{(T)}(z) \xrightarrow{d} B(z),$$

con $T \rightarrow \infty$ dove $B(z)$ è un ponte Browniano standard.

Alcuni valori critici sono 1.22 ($\alpha = 10\%$), 1.36 ($\alpha = 5\%$) e 1.63 ($\alpha = 1\%$). Nell'esperimento Monte Carlo, descritto nel capitolo successivo, verrà utilizzato solamente il valore critico al 5%.

3. L'esperimento di Monte Carlo

Scopo del seguente esperimento è quello di verificare l'ipotesi di costanza dei parametri di un modello a memoria lunga. Per farlo verrà utilizzato il test CUSUM basato sui minimi quadrati ordinari (OLS) di Ploberger e Krämer (1992), il quale converge in distribuzione ad un ponte Browniano standard. Tutti i procedimenti attuati in questa tesi, sono stati eseguiti a partire dall'articolo di Hwang e Shin (2013).

In questo capitolo verrà mostrato come sono state ottenute "size" e potenza del test ad un livello del 5%. In primo luogo verrà presentato come sono stati generati i dati, poi la stima del modello di regressione considerato ed infine la creazione del test CUSUM. Successivamente, saranno presentati i risultati ottenuti, con relative Tabelle, ed il commento di questi.

Per realizzare l'esperimento di Monte Carlo, abbiamo generato i dati a memoria lunga e creato, quindi, la serie storica, attraverso due metodi differenti.

Nel primo caso, seguendo l'articolo di riferimento Hwang e Shin (2013), i dati sono stati generati approssimando un $HAR(\infty)$ a un modello $HAR(8)$, secondo l'equazione:

$$Y_t = \beta_0 + \sum_{j=1}^p \beta_j Y_{t,h_j} + e_t,$$

con $p = 8$, $e_t \sim N(0, 1)$ i.i.d. ed i seguenti parametri:

$$h_1 = 1, h_2 = 5, h_3 = 22, h_4 = 66, h_5 = 132, h_j = 132\nu^{j-5}, j = 6, 7, 8; \nu = 2, 3;$$

$$\beta_0 = 1, \beta_j = c_\beta \lambda^{j-1} / \sum_{j=1}^p \lambda^{j-1}, j = 1, \dots, p; \lambda = 0.4, 0.8; c_\beta = 0.3, 0.6.$$

I primi cinque h_j corrispondono a un giorno, una settimana, un mese, un trimestre e a un semestre. Il processo $HAR(8)$ viene inizializzato a $t = -10000$.

Nel secondo caso, invece, i dati sono stati generati a partire da un modello $ARFIMA(0, d, 0)$, cioè:

$$(1 - B)^d Y_t = e_t,$$

con $e_t \sim N(0, 1)$ i.i.d.

Poiché siamo interessati solo a processi che presentano dipendenza di lungo periodo in senso stretto, abbiamo considerato $d = (0.1, 0.3, 0.45)$. Tale modello viene inizializzato a $t = -2n$.

Le serie simulate hanno numerosità campionaria $n = 500, 2000$, ed il numero di replicazioni indipendenti è $s = 1000$. Dal modello $HAR(p)$ è possibile notare che $\sum_{j=1}^p \beta_j = c_\beta$, questo implica che la media della serie è $\mu = \beta_0 / (1 - c_\beta)$. La stessa media è stata imposta nella generazione dei vari modelli $ARFIMA$.

Il programma per creare le serie storiche e calcolare, poi, il test CUSUM, è stato scritto con il software statistico R. I comandi di R utilizzati per effettuare le simulazioni ed ottenere i risultati che seguiranno (trascritti nelle varie tabelle), sono riportati nell'appendice.

Si procede, poi, alla costruzione del test CUSUM. Per verificare l'ipotesi di nessun cambiamento strutturale, è stato adattato, sia alle serie simulate da un modello HAR sia a quelle simulate da un modello $ARFIMA$, un modello HAR di ordine finito, $HAR(p)$:

$$Y_t = \beta_0 + \sum_{j=1}^p \beta_j Y_{t,h_j} + a_t \quad (3.1)$$

dove $p = 1, \dots, 5$ e a_t è l'errore di regressione.

Per verificare l'ipotesi di costanza nel tempo dei parametri incogniti $\beta_0, \beta_1, \dots, \beta_p$, l'equazione (3.1) viene riscritta come

$$Y_t = \zeta_0 + \sum_{j=1}^p \beta_j (Y_{t,h_j} - \mu_Y) + a_t, \quad (3.2)$$

dove $\zeta_0 = (\beta_0 + \sum_{j=1}^p \beta_j \mu_Y)$ e $\mu_Y = E(Y_t)$. Si può notare che $E(Y_{t,h_j}) = \mu_Y$.

L'equazione (3.2) si può esprimere in forma matriciale come

$$Y = \bar{X}(p) \bar{\beta}(p) + a,$$

dove $Y = (Y_1, \dots, Y_n)'$, $\bar{X}(p) = (\bar{X}_1(p), \dots, \bar{X}_n(p))'$ con $\bar{X}_t(p) = (1, Y_{t,h_1} - \mu_Y, Y_{t,h_2} - \mu_Y, \dots, Y_{t,h_p} - \mu_Y)'$, $\bar{\beta}(p) = (\zeta_0, \beta_1, \beta_2, \dots, \beta_p)'$ e $a = (a_1, \dots, a_n)'$. Una stima OLS modificata di $\bar{\beta}(p)$ è data da

$$\bar{\beta}(p)^{mols} \equiv \left(\hat{\zeta}_0(p), \hat{\beta}_1(p), \dots, \hat{\beta}_p(p) \right)' = \left(\tilde{X}(p)' \tilde{X}(p) \right)^{-1} \tilde{X}(p)' Y$$

dove $\tilde{X}(p) = (\tilde{X}_1(p), \dots, \tilde{X}_n(p))'$ con $\tilde{X}_t(p) = (1, Y_{t,h_1} - \hat{\mu}_Y, Y_{t,h_2} - \hat{\mu}_Y, \dots, Y_{t,h_p} - \hat{\mu}_Y)'$ e $\hat{\mu}_Y = \frac{1}{n} \sum_{i=1}^n Y_i$. Viene definito ora uno stimatore, $\hat{\beta}(p)$, che è leggermente diverso del solito stimatore ai minimi quadrati ordinari nel gestire il termine costante. Infatti, uno stimatore di $\beta(p) \equiv (\beta_0, \beta_1, \dots, \beta_p)'$ si ottiene come $\hat{\beta}(p) \equiv (\hat{\beta}_0(p), \hat{\beta}_1(p), \dots, \hat{\beta}_p(p))'$, dove $\hat{\beta}_0(p) = \hat{\zeta}_0(p) - \sum_{j=1}^p \hat{\beta}_j(p) \hat{\mu}_Y$.

Adesso si procede con il calcolo dei residui OLS e la definizione della statistica CUSUM:

$$\hat{a}_t(p) = Y_t - \hat{\beta}_0(p) - \sum_{j=1}^p \hat{\beta}_j(p) Y_{t,h_j}$$

sono le stime dei residui OLS e le somme cumulate di questi sono definite da:

$$B^{(n)}(z) = \frac{1}{\hat{\sigma}_a \sqrt{n}} \sum_{t=1}^{[nz]} \hat{a}_t(p)$$

dove $0 \leq z \leq 1$ e $\hat{\sigma}_a = \left(\frac{1}{n} \sum_{t=1}^n \hat{a}_t(p)^2 \right)^{1/2}$.

La statistica CUSUM è $B_n = \sup_{0 \leq z \leq 1} |B^{(n)}(z)|$ e si rifiuterà l'ipotesi nulla, H_0 , di costanza dei parametri, per valori grandi di questo test. Quindi la statistica test B_n è stata confrontata con il valore critico di un ponte Browniano standard.

La potenza del test CUSUM è stata calcolata sotto tre scenari differenti. Nel primo caso i dati vengono generati da un modello in cui β_0 viene modificato in $1.2\beta_0$ dopo $t > n/2$, nel secondo β_0 viene modificato a $t > n/4$ e null'ultimo viene modificato a $t > \frac{3}{4}n$.

3.1 Risultati

Tutte le Tabelle, contenenti i risultati relativi al livello e alla potenza empirici del test effettuato sui diversi modelli, sono riportate alla fine di questo paragrafo.

Dall'esame del blocco "size" di tali Tabelle, risulta evidente che, in generale, aumentando p , ovvero aumentando la dipendenza di lungo periodo, nel modello di regressione che adattiamo ai dati di memoria lunga, il test si stabilizza. Mentre, quando p è piccolo, quindi non è presente una forte dipendenza di lungo periodo, il modello adattato $HAR(p)$ non spiega completamente la memoria lunga, questo perché la memoria lunga non spiegata viene erroneamente identificata come un break.

Dall'analisi, invece, della potenza nelle Tabelle, è possibile notare che questa aumenta all'aumentare di n , ciò sta ad indicare la consistenza del test.

Dall'analisi della Tabella 3.1, si può vedere, analizzando il blocco relativo al livello empirico, che già da $p \geq 2$, ad eccezione di $\lambda = 0.8$ e $c_\beta = 0.6$ dove in $p = 2$ sono presenti valori leggermente più elevati rispetto agli altri parametri, il test assume valori ragionevolmente bassi.

Come nella Tabella precedente, dall'esame della Tabella 3.2, si assumono valori ragionevolmente bassi già da $p \geq 2$, ma, per $p = 2, 3$ con $\lambda = 0.8$ e $c_\beta = 0.6$, il test è di grandi dimensioni. Ciò corrisponde ad una memoria lunga più forte rispetto agli altri parametri, che, comunque, si stabilizza col tendere di p a 5.

Dallo studio della Tabella 3.3, si evince che, analizzando il blocco "size", si ottengono valori sufficientemente bassi con $p \geq 3$. Ciò è indipendentemente da quale d si adotti.

Infine, dall'analisi della Tabella 3.4, si può vedere che, sempre analizzando il blocco "size", per $p = 4, 5$, nei casi in cui $d = 0.3, 0.45$, si ottengono valori sufficientemente bassi del test. Mentre per $d = 0.1$, dove la memoria lunga è meno forte rispetto agli altri due casi, si ottengono valori ragionevoli già a partire da $p = 3$.

Mettendo a confronto i due metodi utilizzati per generare le serie storiche, si può notare che i risultati ottenuti con il modello $ARFIMA(0, 0.1, 1)$ sono quelli che si avvicinano di più al modello $HAR(8)$. Inoltre, analizzando il test "size", è evidente che, nei modelli $ARFIMA$, all'aumentare di d , ovvero generando

modelli con dipendenza di lungo periodo sempre più forte, il test impieghi più tempo per stabilizzarsi.

Infine, dall'analisi delle Tabelle, è possibile assodare che, cambiando l'andamento della serie, sia prima che dopo $n/2$, la potenza del test è in entrambi i casi inferiore. Ciò è dovuto al fatto che, inserendo un break a metà della serie, i due andamenti che si creano hanno la stessa lunghezza, quindi è più facile notare il cambiamento; invece, inserendolo prima o dopo la metà, uno dei due andamenti ha una maggiore lunghezza rispetto all'altro, quindi il break è più difficile da rilevare.

Per quanto riguarda le serie simulate con gli *HAR*, i risultati ottenuti sono in linea con quelli prodotti dall'articolo di riferimento.

(Tabella 3.1): livello e potenza empirici del test, $n = 500$, PGD $HAR(8)$.

A		$.4$				$.8$					
N		2		3		2		3			
c_β		$.3$	$.6$	$.3$	$.6$	$.3$	$.6$	$.3$	$.6$		
<i>size</i>	p	1	7	19	9	20	9	20	8	20	
		2	4	7	5	6	5	10	6	11	
		3	2	2	2	2	3	4	2	4	
		4	1	1	2	1	2	1	1	1	
		5	1	1	1	2	1	1	2	1	
<i>power</i>	$n/2$	p	1	59	73	57	72	62	84	65	81
			2	47	56	45	55	54	70	54	71
			3	22	17	21	19	24	23	23	25
			4	3	2	3	2	3	2	2	3
			5	2	2	2	2	2	2	2	1
<i>power</i>	$n/4$	p	1	39	59	39	58	47	74	45	73
			2	27	32	25	32	32	56	34	58
			3	8	7	9	6	10	9	10	10
			4	1	1	0,3	1	1	0,2	1	1
			5	1	1	1	1	1	1	1	0,2
<i>power</i>	$(3/4)n$	p	1	35	53	37	56	41	60	38	58
			2	25	32	24	29	29	42	28	41
			3	8	7	9	7	10	10	11	9
			4	1	1	1	1	2	1	2	1
			5	1	2	2	1	2	2	2	2

(Tabella 3.2): livello e potenza empirici del test, $n = 2000$, PGD HAR(8).

λ		$.4$				$.8$					
ν		2		3		2		3			
c_β		$.3$	$.6$	$.3$	$.6$	$.3$	$.6$	$.3$	$.6$		
<i>size</i>	p	1	7	24	9	21	12	36	15	39	
		2	6	9	4	9	9	24	8	25	
		3	4	5	4	5	6	13	5	14	
		4	3	2	3	2	4	5	3	4	
		5	2	1	2	2	3	2	1	3	
<i>power</i>	$n/2$	p	1	99	100	99	100	100	100	100	
			2	98	99	98	99	100	100	99	100
			3	96	98	97	97	98	100	98	100
			4	78	74	78	72	77	61	80	59
			5	19	17	19	16	14	5	15	6
<i>power</i>	$n/4$	p	1	94	99	94	98	97	100	97	100
			2	89	94	90	92	94	99	95	100
			3	79	83	77	83	89	97	89	97
			4	35	28	32	24	34	21	33	19
			5	3	2	3	2	3	1	2	1
<i>power</i>	$(3/4)n$	p	1	93	98	93	97	96	99	96	100
			2	89	93	87	93	94	97	93	98
			3	79	78	79	81	83	91	86	92
			4	47	36	47	37	45	35	42	34
			5	8	6	8	7	7	3	7	3

(Tabella 3.3): livello e potenza empirici del test, $n = 500$, PGD ARFIMA con $d = 0.1, 0.3, 0.45$.

			ARFIMA						
			$d=0.1$		$d=0.3$		$d=0.45$		
c_β			.3	.6	.3	.6	.3	.6	
<i>size</i>	p	1	20	57	61	17	74	73	
		2	10	9	26	24	28	28	
		3	4	5	5	6	4	3	
		4	1	1	2	2	2	1	
		5	2	1	2	1	2	1	
<i>power</i>	$n/2$	p	1	68	96	71	78	87	87
			2	51	91	44	52	55	56
			3	22	40	7	13	6	7
			4	3	3	3	2	3	3
			5	3	2	2	3	3	3
<i>power</i>	$n/4$	p	1	53	87	67	74	85	89
			2	36	74	34	43	44	44
			3	9	16	5	6	4	4
			4	1	2	2	2	2	3
			5	2	2	2	1	3	2
<i>power</i>	$(3/4)n$	p	1	52	88	68	77	88	87
			2	34	73	36	41	32	33
			3	8	8	4	3	1	1
			4	1	1	1	1	1	1
			5	1	0,2	1	1	1	1

(Tabella): livello e potenza empirici del test, $n = 2000$, PGD ARFIMA con $d = 0.1, 0.3, 0.45$.

		ARFIMA							
		$d=0.1$		$d=0.3$		$d=0.45$			
c_β		.3	.6	.3	.6	.3	.6		
<i>size</i>	p	1	30	30	87	88	95	97	
		2	19	16	59	59	69	68	
		3	9	9	26	25	30	26	
		4	4	4	8	7	6	6	
		5	2	2	3	3	2	2	
<i>power</i>	$n/2$	p	1	98	100	92	96	99	99
			2	96	100	73	82	87	88
			3	91	100	46	64	53	53
			4	63	70	16	23	10	11
			5	16	7	4	6	4	5
<i>power</i>	$n/4$	p	1	92	100	92	96	99	99
			2	86	100	70	78	86	85
			3	75	99	38	53	43	47
			4	9	13	9	11	7	12
			5	7	5	3	4	2	2
<i>power</i>	$(3/4)n$	p	1	94	100	93	95	99	98
			2	86	100	69	77	84	84
			3	73	99	38	48	31	31
			4	7	8	8	8	7	8
			5	3	1	2	2	1	1

4. Conclusioni

In questo lavoro è stato applicato un test CUSUM per verificare l'ipotesi di costanza parametrica in un modello a memoria lunga. È stato, quindi, introdotto il concetto di processo a memoria lunga, nello specifico i modelli eterogenei autoregressivi *HAR* ed i modelli *ARFIMA*, ed il concetto di test CUSUM, in particolare il test CUSUM basato sui residui ai minimi quadrati ordinari (OLS). Utilizzando tale test è stato, poi, formulato un esperimento Monte Carlo, i cui risultati hanno permesso di osservare il comportamento empirico del livello e della potenza del test, costruito a partire da una serie storica a memoria lunga.

Dall'esame degli esiti del livello empirico, risulta evidente che, in generale, aumentando la dipendenza di lungo periodo nel modello di regressione che adattiamo ai dati di memoria lunga, il test si stabilizza. Mentre, quando questa dipendenza non è presente, il modello adattato *HAR(p)* non spiega completamente la memoria lunga, questo perché viene erroneamente identificata come un break.

Analizzando, invece, la potenza empirica dei risultati ottenuti, è possibile notare che questa aumenta all'aumentare di n , ciò sta ad indicare la consistenza del test.

I risultati che sono stati ottenuti generando la serie con un modello HAR, sono in linea con quelli prodotti dall'articolo di riferimento Hwang e Shin (2013).

Appendice: Codice del software R

Si riportano di seguito i principali comandi R utilizzati per le simulazioni dell'esperimento Monte Carlo:

```
n=500, n=2000
s=10000
lambda=0.4, 0.8
c=0.3, 0.6
ni=2, 3
beta0=1
vc=1.3582          #valore critico
cont<-0

#comandi per creare la size del test
#in test si mettono i valori del test
test<-vector(length=s)
somma_l = (1 + lambda + lambda^2 + lambda^3 + lambda^4 + lambda^5 +
lambda^6 + lambda^7)
beta  =  c(c/somma_l,  c*lambda/somma_l,  (c*lambda^2)/somma_l,
(c*lambda^3)/somma_l,  (c*lambda^4)/somma_l,  (c*lambda^5)/somma_l,
(c*lambda^6)/somma_l, (c*lambda^7)/somma_l)
y=NULL

for (i in 1:s){

#generazione dei dati attraverso un modello HAR(8)
t=10000
e = rnorm(t)
y[1] = beta0 + e[1]
for (j in 2:5) {
y[j] = beta0 + beta[1]*y[j-1] + e[j]}
for (j in 6:22) {
  y[j] = beta0 + beta[1]*y[j-1] + beta[2]* sum(y[(j-1):(j-5)])/5 + e[j]}
```

```

for (j in 23:66) {
  y[j] = beta0 + beta[1]*y[j-1] + beta[2]* sum(y[(j-1):(j-5)])/5 + beta[3]*
sum(y[(j-1):(j-22)])/22 + e[j]}
for (j in 67:132) {
  y[j] = beta0 + beta[1]*y[j-1] + beta[2]* sum(y[(j-1):(j-5)])/5 + beta[3]*
sum(y[(j-1):(j-22)])/22 + beta[4]* sum(y[(j-1):(j-66)])/66 + e[j]}
for (j in 133:(132*ni)) {
  y[j] = beta0 + beta[1]*y[j-1] + beta[2]* sum(y[(j-1):(j-5)])/5 + beta[3]*
sum(y[(j-1):(j-22)])/22 + beta[4]* sum(y[(j-1):(j-66)])/66 + beta[5]*
sum(y[(j-1):(j-132)])/132+ e[j]}
for (j in (132*ni+1):(132*ni^2)) {
  y[j] = beta0 + beta[1]*y[j-1] + beta[2]*sum(y[(j-1):(j-5)])/5 +
beta[3]*sum(y[(j-1):(j-22)])/22 + beta[4]*sum(y[(j-1):(j-66)])/66 +
beta[5]*sum(y[(j-1):(j-132)])/132 + beta[6]*sum(y[(j-1):(j-
132*ni)])/ (132*ni + e[j])}
for (j in (132*ni^2+1):(132*ni^3)) {
  y[j] = beta0 + beta[1]*y[j-1] + beta[2]* sum(y[(j-1):(j-5)])/5 + beta[3]*
sum(y[(j-1):(j-22)])/22 + beta[4]* sum(y[(j-1):(j-66)])/66 + beta[5]*
sum(y[(j-1):(j-132)])/132 + beta[6]* sum(y[(j-1):(j-(132*ni))])/ (132*ni) +
beta[7]* sum(y[(j-1):(j-(132*ni^2))])/ (132*ni^2) + e[j]}
for (j in (132*ni^3+1):t) {
  y[j] = beta0 + beta[1]*y[j-1] + beta[2]* sum(y[(j-1):(j-5)])/5 + beta[3]*
sum(y[(j-1):(j-22)])/22 + beta[4]* sum(y[(j-1):(j-66)])/66 + beta[5]*
sum(y[(j-1):(j-132)])/132 + beta[6]* sum(y[(j-1):(j-(132*ni))])/ (132*ni) +
beta[7]* sum(y[(j-1):(j-(132*ni^2))])/ (132*ni^2) + beta[8]* sum(y[(j-1):(j-
(132*ni^3))])/ (132*ni^3) + e[j]}

#generazione dei dati attraverso un modello ARFIMA(0, d, 0)
d=0.1, 0.3, 0.45
t=2*n
library(fracdiff)
y <- ts(fracdiff.sim(t, ar=NULL, ma=NULL, d=..., mu=(beta0/(1-c)))$series)

```

```

#vengono considerati soltanto gli ultimi n valori della serie creata
Y=y[(t-n+1):t]
mu.hat<-mean(Y)
#creazione degli  $Y_{t,h_j}$  nel caso massimo in cui p=5
Y.h1<-y[(t-n):(t-1)]
Y2<-vector(length=n)
Y3<-vector(length=n)
Y4<-vector(length=n)
Y5<-vector(length=n)

for(j in (t-n+1):t){
Y2[j]<-sum(y[(j-1):(j-5)])/5
Y3[j]<-sum(y[(j-1):(j-22)])/22
Y4[j]<-sum(y[(j-1):(j-66)])/66
Y5[j]<-sum(y[(j-1):(j-132)])/132}

Y.h2<-c(Y2[(t-n+1):t])
Y.h3<-c(Y3[(t-n+1):t])
Y.h4<-c(Y4[(t-n+1):t])
Y.h5<-c(Y5[(t-n+1):t])

x1<-(Y.h1-mu.hat)
x2<-(Y.h2-mu.hat)
x3<-(Y.h3-mu.hat)
x4<-(Y.h4-mu.hat)
x5<-(Y.h5-mu.hat)

#stima del modello per ricavare i parametri  $\beta$ 
model<-lm(Y~x1+x2+x3+x4+x5)

b.mols=as.vector(summary(model)$coef[,1]) #estrazione dal modello di  $\beta_0$ 
b0.hat<-b.mols[1]-(sum(b.mols[2:6])*mu.hat)

```

```

#vengono calcolati I residui
res<-Y-b0.hat-b.mols[2]*Y.h1-b.mols[3]*Y.h2-b.mols[4]*Y.h3-b.mols[5]*Y.h4-
b.mols[6]*Y.h5

sigma.hat<-sqrt(sum(res^2)/n)

#creazione del vettore contenente la somme cumulata dei residui per ogni z
Bz = NULL
for(k in 1:100){
z<-k*0.01
Bz[k]<-sum(res[1:(n*z)])/(sigma.hat*sqrt(n))
}
test[i]<-max(abs(Bz))
#viene verificato che si rifiuta l'ipotesi nulla per valori grandi del test
if (test[i] >vc) cont<- cont + 1
}
#viene calcolata la size del test
pot<- cont/s*100

#Comandi per creare la potenza del test
#cambia rispetto al procedimento precedente nel generare i dati

#generazione dei dati attraverso un modello HAR(8)con break a n/2
t=10000
e = rnorm(t)
y[1] = beta0 + e[1]
for (j in 2:5) {
y[j] = beta0 + beta[1]*y[j-1] + e[j]}
for (j in 6:22) {
y[j] = beta0 + beta[1]*y[j-1] + beta[2]* sum(y[(j-1):(j-5)])/5 + e[j]}
for (j in 23:66) {
y[j] = beta0 + beta[1]*y[j-1] + beta[2]* sum(y[(j-1):(j-5)])/5 + beta[3]*
sum(y[(j-1):(j-22)])/22 + e[j]}

```



```

for (j in 67:132) {
  y[j] = beta0 + beta[1]*y[j-1] + beta[2]* sum(y[(j-1):(j-5)])/5 + beta[3]*
sum(y[(j-1):(j-22)])/22 + beta[4]* sum(y[(j-1):(j-66)])/66 + e[j]}
for (j in 133:(132*ni)) {
  y[j] = beta0 + beta[1]*y[j-1] + beta[2]* sum(y[(j-1):(j-5)])/5 + beta[3]*
sum(y[(j-1):(j-22)])/22 + beta[4]* sum(y[(j-1):(j-66)])/66 + beta[5]*
sum(y[(j-1):(j-132)])/132+ e[j]}
for (j in (132*ni+1):(132*ni^2)) {
  y[j] = beta0 + beta[1]*y[j-1] + beta[2]*sum(y[(j-1):(j-5)])/5 +
beta[3]*sum(y[(j-1):(j-22)])/22 + beta[4]*sum(y[(j-1):(j-66)])/66 +
beta[5]*sum(y[(j-1):(j-132)])/132 + beta[6]*sum(y[(j-1):(j-
132*ni)])/ (132*ni + e[j])}
for (j in (132*ni^2+1):(132*ni^3)) {
  y[j] = beta0 + beta[1]*y[j-1] + beta[2]* sum(y[(j-1):(j-5)])/5 + beta[3]*
sum(y[(j-1):(j-22)])/22 + beta[4]* sum(y[(j-1):(j-66)])/66 + beta[5]*
sum(y[(j-1):(j-132)])/132 + beta[6]* sum(y[(j-1):(j-(132*ni))])/ (132*ni) +
beta[7]* sum(y[(j-1):(j-(132*ni^2))])/ (132*ni^2) + e[j]}
for (j in (132*ni^3+1):(t-n/2)) {
  y[j] = beta0 + beta[1]*y[j-1] + beta[2]* sum(y[(j-1):(j-5)])/5 + beta[3]*
sum(y[(j-1):(j-22)])/22 + beta[4]* sum(y[(j-1):(j-66)])/66 + beta[5]*
sum(y[(j-1):(j-132)])/132 + beta[6]* sum(y[(j-1):(j-(132*ni))])/ (132*ni) +
beta[7]* sum(y[(j-1):(j-(132*ni^2))])/ (132*ni^2) + beta[8]* sum(y[(j-1):(j-
(132*ni^3))])/ (132*ni^3) + e[j]}
for (j in (t-(n/2)+1):t) {
  y[j] = 1.2*beta0 + beta[1]*y[j-1] + beta[2]* sum(y[(j-1):(j-5)])/5 +
beta[3]* sum(y[(j-1):(j-22)])/22 + beta[4]* sum(y[(j-1):(j-66)])/66 +
beta[5]* sum(y[(j-1):(j-132)])/132 + beta[6]* sum(y[(j-1):(j-
(132*ni)])/ (132*ni) + beta[7]* sum(y[(j-1):(j-(132*ni^2))])/ (132*ni^2) +
beta[8]* sum(y[(j-1):(j-(132*ni^3))])/ (132*ni^3) + e[j]}

```

#generazione dei dati attraverso un modello ARFIMA(0, d, 0) con break a n/2
d=0.1, 0.3, 0.45

```
t=2*n
library(fracdiff)
y1 <- ts(fracdiff.sim((t-n/2), ar=NULL, ma=NULL, d, mu=(beta0/(1-
c))))$series)
y2 <- ts(fracdiff.sim((n/2), ar=NULL, ma=NULL, d, mu=(1.2*beta0/(1-
c))))$series)
y=ts(c(y1,y2))
```

Bibliografia:

- BISAGLIA L., (1998), "Processi a memoria lunga: problemi di stima, identificazione e previsione", Tesi di dottorato. Dip.to di Scienze Statistiche. Università degli studi di Padova, Febbraio 1998.
- DI FONZO T., LISI F., (2005), "Serie storiche economiche", Carrocci, Roma.
- HWANG E., SHIN D.W., (2013), "A CUSUM test for a long memory heterogeneous autoregressive model", Economics Letters.
- HWANG E., SHIN D.W., (2013), "Infinite-order, long- memory heterogeneous autoregressive models", Computational Statistics and Data Analysis.
- MCALEER M., MEDEIROS M.C., (2008), "A multiple-regime smooth transition heterogeneous autoregressive model for long memory and asymmetries", Journal of Econometrics 147, 104-119.
- MCCABE B.P.M., HARRISON M.J., (1980), "Testing the constancy of regression relationships over time using least squares residuals", Journal of the Royal Statistical Society C, 29, 142-148.
- PLOBERGER W., KRÄMER W., (1992), "The CUSUM test whit OLS residuals", Econometrica 60, 465-515.
- SIBBERTSEN P., (2004), "Long memory versus structural breaks: an overview", Statistical Paper 45, 465-515.