

## 1. Processi Stocastici Stazionari

In matematica un processo stazionario è un processo stocastico la cui distribuzione di probabilità ha la stessa forma per ogni scala temporale o spaziale. Ne consegue che parametri come media e varianza, se esistono, sono invarianti nel tempo e nello spazio.

Un esempio di processo stazionario è dato dalla *white noise* gaussiana. D'altro canto, il suono del piatto di una batteria non rappresenta un processo stazionario poiché la sua potenza acustica (quindi la sua varianza), a seguito del colpo, decade col passare del tempo, risultando quindi legata alla storia pregressa del processo stesso.

La stazionarietà viene spesso usata nell'analisi delle serie temporali, dove il flusso dei dati è spesso trasformato in modo da divenire stazionario. I processi sono descritti da andamenti o trend stazionari se sono combinazione lineare di processi stazionari, uno o più dei quali esibiscono un andamento ben definito o trend. Trasformare una serie di dati per renderla processo stazionario è un'operazione che prende il nome di *de-trending* (ne è un esempio la DFA – De-trended Fluctuations Analysis).

Entriamo nei dettagli e consideriamo un processo stocastico  $\{X_n, n \in Z\}$ , dove  $Z$  rappresenta l'insieme dei numeri interi relativi, tale che la sua varianza esista e risulti finita, cioè  $Var(X_n) < \infty, \forall n \in Z$ . La funzione di *autocovarianza*  $\gamma_x(\cdot, \cdot)$  di  $\{X_n\}$  è definita dall'espressione seguente

$$\gamma_x(n, m) = Cov(X_n, X_m) = \langle \{ [X_n - \langle X_n \rangle] \cdot [X_m - \langle X_m \rangle] \} \rangle$$

dove con  $\langle X_n \rangle$  abbiamo rappresentato il valore medio della variabile aleatoria  $X_n$ .

La varianza di  $X_n$  che deve essere finita per ipotesi, è quella definita per ogni istante di tempo. In questo modo vogliamo indicare che si costruiscono diverse realizzazioni del processo stocastico, allo stesso istante di tempo, con varianza finita.

Una serie temporale  $\{X_n, n \in Z\}$  forma un processo stocastico stazionario se soddisfa le seguenti proprietà:

- i.  $\langle |X_n|^2 \rangle < \infty, \forall n \in Z$
- ii.  $\langle X_n \rangle = \mu$ , che è il valore medio,  $\forall n \in Z$
- iii.  $\gamma_x(n, m) = \gamma_x(n+t, m+t), \forall n, m, t \in Z$

Definiamo la proprietà iii *stazionarietà debole*.

L'autocovarianza di un processo stazionario la possiamo definire in termini di una funzione di una sola variabile  $h = n - m$ , ossia

$$\gamma_x(h) \equiv \gamma_x(h, 0) = Cov(X_{n+h}, X_n), \forall n, m \in Z \quad (1.1)$$

Questa è la funzione di autocovarianza alla scala  $h$ , ossia alla separazione  $h$ .

La (1.1) segue dal fatto che, se consideriamo un processo stocastico stazionario  $\{X_n, n \in Z\}$ , per la stazionarietà debole, scegliendo  $t = -m$ , risulta

$$\gamma_x(n, m) = \gamma_x(n-m, m-m) = \gamma_x(n-m, 0), \forall n, m \in Z$$

Definiamo ora la funzione di *autocorrelazione* del processo stocastico  $\{X_n, n \in Z\}$  nel modo seguente

$$\rho_x(h) = \frac{\gamma_x(h)}{\gamma_x(0)} = \text{Corr}(X_{n+h}, X_n)$$

Possiamo definire anche una stazionarietà forte (*strict stationarity*) usando la funzione di distribuzione della variabile aleatoria  $X_n$ .

Consideriamo il caso in cui il processo stocastico sia formato da una serie temporale (variabile aleatoria definita ad incrementi temporali)  $\{X_n, n \in T\}$ , dove  $T$  è un insieme temporale. Essa presenta una stazionarietà forte se la distribuzione della sottosequenza  $(X_1, \dots, X_k)$  è uguale alla distribuzione della sottosequenza  $(X_{1+h}, \dots, X_{k+h})$ , per ogni coppia di interi positivi  $k$  e  $h$ . In modo intuitivo questo significa che il grafico ottenuto da una qualsivoglia realizzazione della serie temporale deve mostrare le stesse caratteristiche statistiche su intervalli di eguale lunghezza, ma scelti in modo arbitrario.

Da quanto detto siamo portati a concludere che se  $\{X_n, n \in T\}$  è una serie temporale con stazionarietà forte, allora, scegliendo  $k = 1$ , segue che  $X_n$  ha la stessa distribuzione qualunque sia  $n \in T$ . Se  $\langle |X_n|^2 \rangle < \infty$  abbiamo che il valore medio  $\langle X_n \rangle$  e la varianza  $\text{Var}(X_n)$  della variabile aleatoria  $X_n$  sono entrambi costanti. Inoltre, prendendo  $k = 2$ , risulta che  $X_{n+h}$  e  $X_n$  debbono avere la stessa distribuzione e quindi la stessa covarianza per tutti i valori di  $h \in Z$ . Questo è un risultato notevole poiché ci permette di concludere che un processo con stazionarietà forte, per cui il momento del secondo ordine (varianza) è finito, è un processo stazionario. L'inverso è generalmente falso, ossia un processo stazionario non implica necessariamente che esso sia anche stazionario in senso forte. Tuttavia ciò risulta vero se e solo se la serie temporale in esame è un processo stocastico gaussiano.

I processi stazionari giocano un ruolo cruciale nello studio delle serie temporali. In molti casi dall'osservazione diretta dell'andamento o trend di una serie temporale si deduce un'apparente assenza di stazionarietà. Tuttavia con diverse tecniche che discuteremo in seguito, è possibile operare delle trasformazioni sull'insieme dei dati in modo che il processo ottenuto coi nuovi dati trattati possa ritenersi stazionario, al fine di essere pronto per essere trattato coi modelli statistici propri delle serie temporali stazionarie.

## 2. Autocovarianza e autocorrelazione di un processo stazionario

Analizziamo adesso le proprietà della funzione di autocovarianza di un qualsiasi processo stazionario  $\{X_n, n \in T\}$ . Poiché un processo stazionario ha varianza finita per ipotesi, cioè  $Var(X_n) < \infty$ , otteniamo immediatamente che la funzione di autocovarianza dell'incremento temporale  $h$  è non negativo, ossia  $\gamma(h) \geq 0$ .

Consideriamo ora la seguente disuguaglianza di Cauchy-Schwarz

$$|\gamma(h)| = |Cov(X_{n+h}, X_n)| \leq \sqrt{Var(X_{n+h}) \cdot Var(X_n)}. \quad (2.1)$$

Dalla (2.1) otteniamo immediatamente che  $|\gamma(h)| \leq \gamma(0), \forall h \in Z$ .

Un'ulteriore proprietà della funzione di autocovarianza di un processo stazionario è che

$$\gamma(-h) = Cov(X_{n-h}, X_n) = Cov(X_n, X_{n+h}) = \gamma(h)$$

Esiste anche un teorema che afferma che la funzione di autocovarianza  $\gamma(h)$  è definita positiva per ogni serie temporale  $\{X_n, n \in T\}$  formante un processo stazionario. Infine, per ogni funzione di autocovarianza  $\gamma(h)$  esiste sempre un processo stazionario gaussiano, la cui funzione di autocovarianza è proprio  $\gamma(h)$  stessa.

Dall'osservazione di una particolare realizzazione  $x_n$  di un processo stocastico  $\{X_n, n \in T\}$ , possiamo ottenere la funzione di autocovarianza del campione in esame, definita nel modo seguente

$$\gamma(h) = \frac{1}{n} \sum_{j=1}^{n-h} (x_{j+h} - \langle x \rangle)(x_j - \langle x \rangle) \quad (2.2)$$

per  $0 \leq h \leq n$ , e dove con  $\langle x \rangle$  abbiamo indicato il valor medio degli  $n$  elementi  $x_j$ . Inoltre sappiamo che  $\gamma(h) = \gamma(-h)$  e  $-n < h \leq 0$ .

Osserviamo che nella (2.2) abbiamo diviso per  $n$  anziché per  $n-h$  poiché vogliamo essere sicuri che la matrice  $\Gamma_n = [\gamma(i-j)]$ , calcolata per  $i, j = 1, \dots, n$ , sia positiva definita.

La funzione di autocorrelazione del campione è definita tramite la funzione di autocovarianza

$$\rho(h) = \frac{\gamma(h)}{\gamma(0)}$$

per  $|h| < n$ . La matrice corrispondente  $R_n = [\rho(i-j)]$  risulta anch'essa positiva definita.

Notiamo che sia la funzione di autocorrelazione che la funzione di autocovarianza possono essere calcolate per qualsiasi serie temporale, non necessariamente per quelle che costituiscono la realizzazione di un processo stocastico. Se i dati contengono una componente secolare, allora  $|\rho(h)|$  esibisce un decadimento molto lento al crescere di  $h$ . Se invece i dati hanno una componente periodica, allora la funzione di autocorrelazione  $\rho(h)$  mostrerà la stessa identica periodicità.

Concludiamo affermando che la funzione di autocorrelazione rappresenta uno strumento potente per stabilire l'assenza di stazionarietà di un processo stocastico.

### 3. Componenti secolari e periodiche

In precedenza abbiamo parlato della possibilità di utilizzare la funzione di autocorrelazione  $\rho(h)$  per stabilire se in un processo stocastico qualsiasi ci sia o meno assenza di stazionarietà. Essa, infatti, ci permette di stabilire se all'interno della serie temporale  $\{X_n, n \in T\}$  ci siano o meno componenti secolari e/o periodicità che, con metodi che fra poco presenteremo, possono essere ridotte o addirittura eliminate.

Data una qualsiasi serie temporale  $\{X_n, n \in T\}$ , possiamo immaginare che essa sia divisa nel modo seguente

$$X_n = s_n + p_n + Y_n, n \in T \tag{3.1}$$

dove  $s_n, p_n$  e  $Y_n$  sono tre variabili aleatorie tra loro generalmente indipendenti.

La variabile aleatoria  $s_n$  è detta *componente secolare* se è una funzione dipendente dal tempo in modo non banale e che varia molto lentamente.

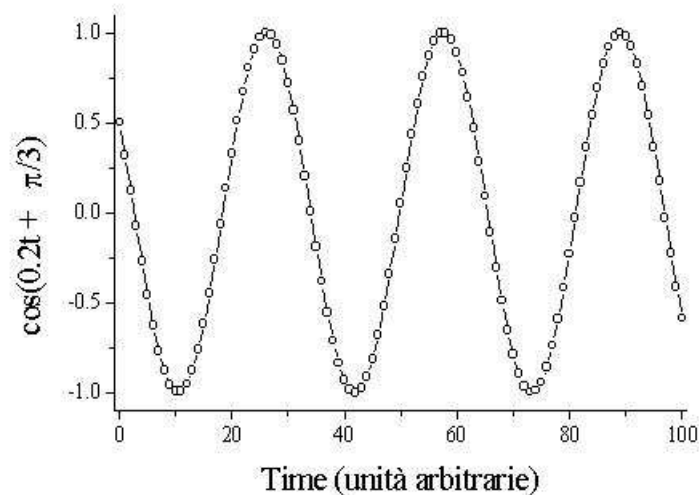
Definiamo invece la variabile aleatoria  $p_n$  *componente periodica* se rappresenta una funzione del tempo periodica, avente un certo periodo apparente  $d$ .

La variabile aleatoria rimanente  $Y_n$  rappresenta un processo stocastico che è almeno debolmente stazionario.

La speranza è che, dopo aver individuato e quindi eliminato le componenti secolari e periodiche, il resto della serie sia un processo stocastico stazionario interessante, in modo da poter usare la teoria che li descrive statisticamente, al fine di trovare un modello probabilistico per  $Y_n$ , studiarne le caratteristiche principali, quindi usare l'intera serie  $X_n$  per operare previsioni o controlli su di essa.

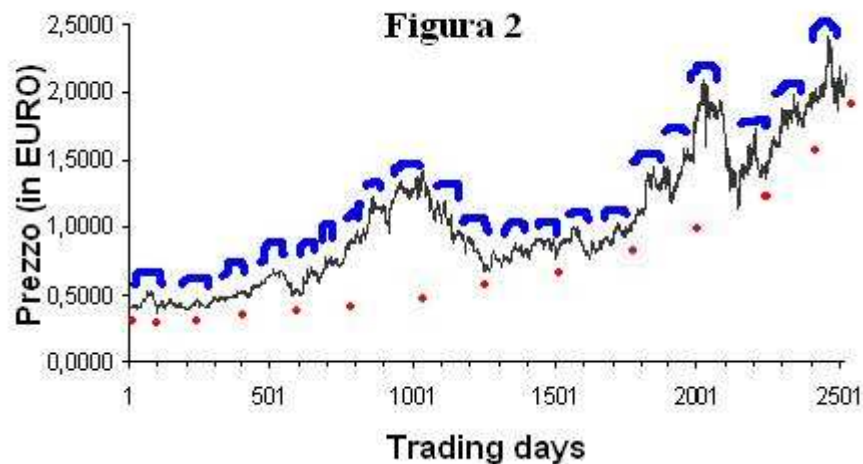
Vediamo alcuni esempi di serie temporali aventi periodicità e componenti secolari.

Figura 1

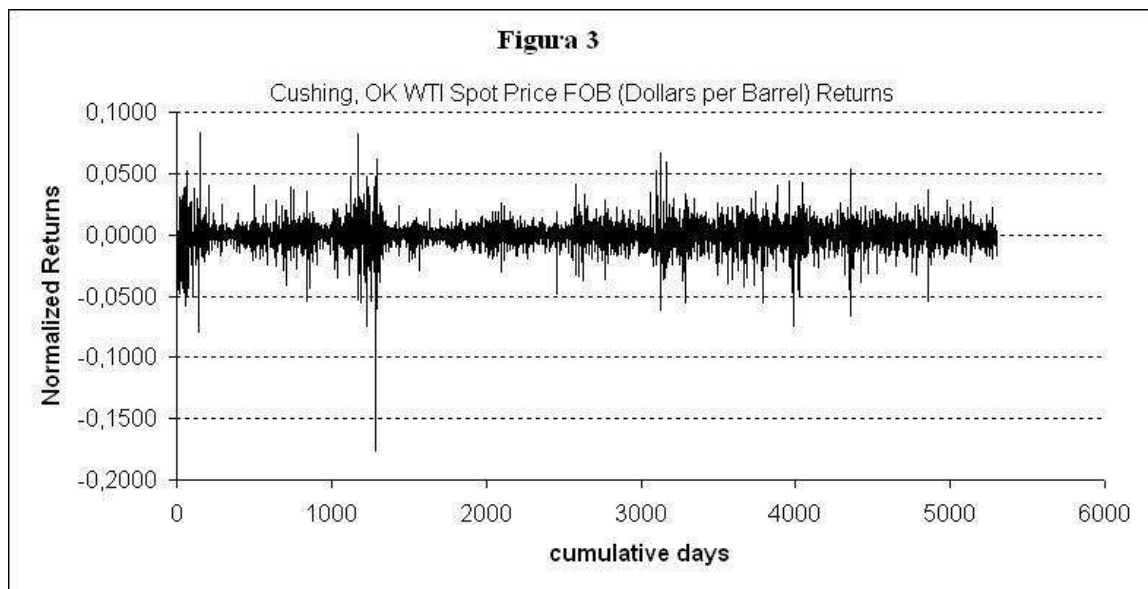


La figura 1 mostra come la serie temporale generata dalla funzione  $f(t) = \cos\left(0.2t + \frac{\pi}{3}\right)$  possiede una componente periodica.

La figura 2 seguente invece mostra una serie temporale finanziaria che presenta sia una componente secolare (la traccia in rosso che sembra essere una crescita esponenziale) che una componente periodica (le piccole concavità segnate in blu, che sembrano ripetersi ad intervalli temporali pressoché costanti).



La figura 2 si riferisce a dieci anni di quotazione giornaliera del titolo azionario Benetton Group. Nella prossima figura (figura 3), per completare il quadro, mostriamo come si presenta graficamente un processo stocastico stazionario depurato dalle componenti secolare e periodica.



La figura 3 mostra la serie temporale formata dai ritorni normalizzati di dieci anni di quotazione giornaliera del prezzo del petrolio in dollari per barile. Parleremo in seguito di questa particolare trasformazione della serie di dati. Quello che ci interessa ora è notare che il processo stocastico ottenuto in figura è stazionario (ma questo lo dimostreremo in seguito) e depurato da qualsiasi componente secolare e/o periodica.

#### 4. Eliminazione delle componenti secolari e periodiche

Come abbiamo detto nel paragrafo precedente, è possibile eliminare o ridurre al minimo le componenti secolari e periodiche operando sulla serie temporale rappresentante il processo stocastico in esame, al fine di ottenere un processo stocastico che sia almeno debolmente stazionario.

Consideriamo ora un processo stocastico  $\{X_n, n \in T\}$  privo di componenti periodiche. La (3.1) diventa

$$X_n = s_n + Y_n$$

e per semplicità possiamo supporre che la variabile aleatoria  $Y_n$  sia stata normalizzata in modo da avere media zero, ossia  $\langle Y_n \rangle = 0$ . Per avere una stima della componente secolare  $s_n$ , possiamo operare in generale un *fit* (approssimazione) con una famiglia di  $n$  funzioni polinomiali di secondo grado, per esempio

$$s_n = a_0 + a_1 n + a_2 n^2$$

scegliendo le costanti  $a_0$ ,  $a_1$  e  $a_2$  in modo da minimizzare la distanza quadratica tra  $s_n$  e la serie temporale osservata  $X_n$ . Questo metodo è detto *stima dei minimi quadrati*.

Un metodo alternativo è rappresentato dallo *smoothing* tramite *medie correnti*. Sia  $q$  un numero intero positivo e consideriamo la media corrente del processo stocastico  $\{X_n, n \in T\}$ , ossia definiamo la nuova variabile casuale

$$W_n = \frac{1}{2q+1} \sum_{j=-q}^q X_{n+j}$$

che rappresenta il valore medio dei  $2q+1$  valori adiacenti a  $X_n$  (un intorno di valori di  $X_n$  di cui  $X_n$  è il centro). Se l'insieme dei dati è composto da  $N$  termini, per  $q+1 \leq n \leq N-q$  abbiamo la seguente espressione

$$W_n = \frac{1}{2q+1} \sum_{j=-q}^q s_{n+j} + \frac{1}{2q+1} \sum_{j=-q}^q Y_{n+j} \approx s_n$$

risultato che si ottiene immediatamente se supponiamo che la componente secolare  $s_n$  è approssimativamente costante nell'intorno di  $X_n$  di raggio  $q$  usato per il calcolo della media (infatti, per definizione, la componente secolare  $s_n$  è una funzione lentamente variabile nel tempo, quindi può essere considerata costante per medio-piccoli intervalli temporali), e se si suppone inoltre che il valor medio della variabile aleatoria  $Y_n$  sia nullo, ossia  $\langle Y_n \rangle = 0$ . Di conseguenza la media corrente ci fornisce un valore  $\mu_n \approx s_n$  che rappresenta la stima per  $q+1 \leq n \leq N-q$

$$\mu_n = \frac{1}{2q+1} \sum_{j=-q}^q X_{n+j}$$

Ovviamente la media corrente non trova applicazione per  $n \leq q$  e per  $n > N - q$ , in quanto l'intorno di  $X_n$ , in questi due casi, non risulterebbe completo. In questi casi bisogna usare una media corrente su mezzo intorno di  $X_n$ . Una stima del termine stocastico dato dalla variabile aleatoria  $Y_n$  si ottiene quindi da  $Y_n = X_n - \mu_n$ . Si può pensare alla serie temporale  $\{\mu_n, n \in T\}$  come ad un processo stocastico ottenuto da  $\{X_n, n \in T\}$  mediante l'applicazione di un operatore lineare, o filtro lineare

$$\mu_n = \sum_{j=-\infty}^{\infty} a_j X_{n+j}$$

i cui pesi sono così distribuiti:

$$\begin{cases} a_j = \frac{1}{2q+1}, -q \leq j \leq q \\ a_j = 0, |j| > q \end{cases}$$

Tale filtro è definito *passa basso* perché la sua applicazione elimina le componenti che oscillano rapidamente dalla serie  $X_n$ , conservando, di conseguenza, le componenti secolari che oscillano molto lentamente nel tempo. Questo è uno dei tanti filtri che si possono usare per eseguire uno smoothing sulla serie temporale di dati. Ovviamente non possiamo scegliere un valore di  $q$  troppo elevato se la componente secolare non dovesse variare lentamente nel tempo, altrimenti possiamo incontrare dei problemi, per cui il metodo risulta privo di significato.

La componente secolare  $s_n$  di una serie temporale  $\{X_n, n \in T\}$  può essere eliminata anche attraverso il metodo della *differenziazione* che adesso presentiamo. Definiamo un operatore il cui risultato sia la differenza del primo ordine tra due eventi successivi della serie temporale

$$\nabla X_n = X_n - X_{n-1} = (1 - B)X_n, \tag{4.1}$$

dove indichiamo con  $B$  l'operatore di traslazione temporale all'indietro, ossia  $BX_n = X_{n-1}$ . Gli operatori di ordine superiore sono definiti in modo ricorsivo nel seguente modo

$$\nabla^j X_n = \nabla(\nabla^{j-1} X_n), \text{ con } \nabla^0 X_n = X_n$$

Se questo operatore viene applicato ad un termine lineare  $s_n = an + b$ , si ottiene una funzione costante  $\nabla s_n = a$ . Allo stesso modo, ogni termine secolare polinomiale di grado  $k$  può essere eliminato, usando  $k$  volte un operatore differenziale del tipo (4.1).

Se la serie temporale  $\{X_n, n \in T\}$  è tale da risultare  $X_n = s_n + Y_n$ , per ogni  $n \in T$ , dove  $s_n$  è la componente secolare rappresentata da un polinomio di grado  $k$

$$s_n = a_0 + a_1 n + a_2 n^2 + \dots + a_{k-1} n^k$$

e  $Y_n$  è un processo stazionario a media nulla, si ha

$$\nabla^k X_n = k! a_{k-1} + \nabla^k Y_n \tag{4.2}$$

La (4.2) rappresenta un processo stocastico stazionario con media  $a_{k-1}$ . Questo risultato mostra come, data una sequenza temporale  $\{X_n, n \in T\}$ , si può applicare l'operatore differenziale (4.1)  $k$  volte, fino a quando non si ottiene una nuova sequenza  $\{\nabla^k X_n, n \in T\}$  che può essere modellizzata come serie temporale stazionaria. In pratica basta applicare l'operatore (4.1) una o due volte per ottenere processo stocastico stazionario, e questo è dovuto al fatto che ogni funzione può essere approssimata con un polinomio di grado basso su un intervallo non troppo esteso. Spesso si nota che l'intensità della serie temporale  $\{\nabla^k X_n, n \in T\}$  ottenuta aumenta con  $n$ . Questo è un fenomeno abbastanza comune e può essere facilmente ridotto o del tutto eliminato se, ad esempio, si opera prima con una funzione logaritmo  $z_n = \log X_n$ , e poi si applica l'operatore differenziale (4.1).

Nella figura 3 del paragrafo 3 mostriamo un esempio di applicazione del metodo della differenziazione di tipo (4.1). La variabile aleatoria rappresentata nel grafico (*ritorni normalizzati*) l'abbiamo ottenuta attraverso la differenziazione del primo ordine dei logaritmi di prezzi del petrolio in due istanti successivi. Il processo stocastico  $\{r_n, n \in T\}$  risulta stazionario (verrà dimostrato in seguito) ed è stato ottenuto nel modo seguente

$$r_n = \log(g_n) - \log(g_{n-1})$$

dove  $g_n$  e  $g_{n-1}$  sono rispettivamente i prezzi del petrolio (in dollari per barile) agli istanti  $n$  ed  $n - 1$ . Successivamente il processo è stato normalizzato in modo che risultasse  $\langle r_n \rangle = 0$ .

Una tecnica alternativa consiste nell'applicare un operatore di differenziazione a cui è associato un "ritardo"  $d$ , definito nel modo seguente

$$\nabla_d X_n = X_n - X_{n-d} = (1 - B^d)X_n \tag{4.3}$$

dove  $B$  è sempre l'operatore di traslazione temporale all'indietro, mentre non bisogna confondere l'operatore appena introdotto  $\nabla_d$  con l'operatore precedente  $\nabla^d$ .

La potenzialità di questo nuovo operatore sta nell'intercettare ed eliminare anche le componenti periodiche di periodo apparente  $d$ . Infatti esso si applica ad una serie temporale  $\{X_n, n \in T\}$ , tale che  $X_n = s_n + p_n + Y_n$ , dove  $p_n$  è una componente periodica di periodo apparente  $d$ .

Applicando l'operatore (4.3) alla serie temporale  $\{X_n, n \in T\}$  otteniamo

$$\nabla_d X_n = s_n - s_{n-d} + Y_n - Y_{n-d}$$

che elimina la componente periodica  $p_n$  dando quindi origine ad un nuovo processo stocastico con componente secolare  $s_n - s_{n-d}$  ed una processo stocastico stazionario  $Y_n - Y_{n-d}$ . Quindi, applicando questa volta l'operatore (4.1), la componente secolare  $s_n - s_{n-d}$  può essere rimossa, come descritto in precedenza.

### 5. Un esempio di processo stazionario: Il moto browniano classico

Il *moto browniano* è un processo stocastico continuo nel tempo che può essere descritto da una *random walk* (passeggiata stocastica) omogenea. Per la sua universalità è a buona ragione considerato il processo stocastico di base. Un modo intuitivo per rappresentare un moto browniano è considerare i semplici “saltelli” stocastici di un punto su di una retta ordinata ed orientata. Supponiamo che il punto parta dalla posizione  $x_0 = 0$  e si muova, verso destra o verso sinistra, di un tratto pari a  $\sqrt{dt}$ , essendo  $dt$  l’intervallo di tempo impiegato da punto per compiere il passo (ritenuto costante). La nuova posizione del punto, dopo  $n$  passi, è legata alla posizione immediatamente precedente tramite la relazione

$$x_n = x_{n-1} \pm \sqrt{dt}$$

dove il segno (+, -) è attribuito in modo del tutto casuale, con uguale probabilità  $\frac{1}{2}$ . Otteniamo una traiettoria continua tramite l’interpolazione lineare operata su ogni sottosequenza  $ndt \leq t \leq (n+1)dt$ , ossia

$$X(t) = x_n + (t - ndt)(x_{n+1} - x_n). \tag{5.1}$$

Si definisce così un processo stocastico  $\{X(t), t \in T\}$ .

Il moto browniano gode delle seguenti proprietà

- i. Se  $t = ndt$  ed  $a > 0$ , gli incrementi  $X(t+a) - X(t)$  sono indipendenti dal “passato” del moto, in altre parole il moto browniano è stazionario, cioè privo di memoria.
- ii. Il valor medio è sempre nullo, cioè  $\langle X(t) \rangle = 0$ .
- iii. La varianza è finita e vale  $Var(X(t)) = \langle X(t)^2 \rangle = t$ .

Per passi molto piccoli, tali che  $dt \ll 1$ , il moto browniano ha incrementi  $X(t+a) - X(t)$  che sono indipendenti tra di loro, a parte una debole persistenza dovuta all’interpolazione (5.1).

Ognuno di questi incrementi rappresenta la somma di variabili binomiali casuali aventi media nulla e varianza finita. Quindi, in accordo col teorema del limite centrale e per  $dt$  infinitesimo, la distribuzione di probabilità limite degli incrementi è gaussiana. Più precisamente, per ogni  $x$  abbiamo

$$\lim_{dt \rightarrow 0} P\{X(t+a) - X(t) \geq x\} = \frac{1}{\sqrt{2\pi a}} \int_x^{+\infty} e^{-\frac{y^2}{2a}} dy. \tag{5.2}$$

La proprietà (5.2), insieme all’indipendenza degli incrementi, caratterizza la statistica dei processi browniani.

Una proprietà importante dei moti browniani è la loro continuità. Essi sono continui pur presentando diverse irregolarità. Infatti, chi studia questi processi, come ad esempio le serie temporali nell’ambito finanziario, ha avuto modo di notare come questi non siano lisci, ma spesso presentano inclinazioni quasi infinite, in modo da renderne difficile la predizione. Una traiettoria browniana e quindi in generale non differenziabile.

Un’altra proprietà notevole dei moti browniani è quella che spesso viene chiamata col nome di *autosimilarità* statistica (conosciuta col termine inglese di *self similarity*). Se effettuiamo una

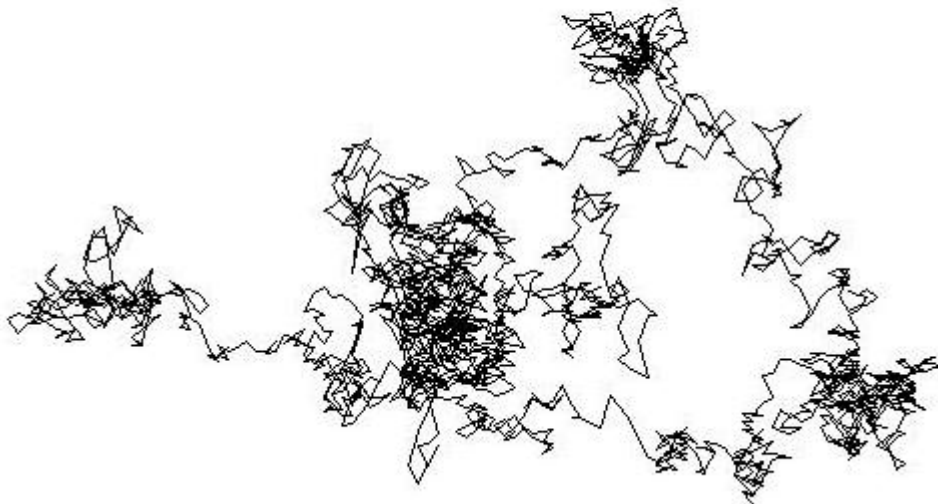
trasformazione della scala temporale (dilatazione o contrazione)  $t \rightarrow \lambda t$  (con  $\lambda > 0$ ), per un moto browniano vale la seguente applicazione

$$X(t) \rightarrow \lambda^{\frac{1}{2}} X(\lambda t) \quad (5.3)$$

che è un'applicazione tra moti browniani. Infatti se consideriamo, per esempio, un moto browniano definito sull'intervallo  $[0,1]$  e lo "dilatiamo" secondo la (5.3) usando  $\lambda = 0.5$ , il risultato è sempre un moto browniano definito però sull'intervallo  $[0,2]$ . L'autosimilarità dei moti browniani indica che questi sono oggetti appartenenti alla geometria *frattale*. Questo significa che le serie statistiche appaiono identiche, dilatando o contraendo la scala temporale di un qualche fattore  $\lambda$  e il "balzo"  $X$  di  $\lambda^{\frac{1}{2}}$ .

### Figura 4

Benetton Normalized Returns - Random Walk



In figura 4 mostriamo la realizzazione di un moto browniano bidimensionale utilizzando come serie temporale di riferimento i ritorni normalizzati di dieci anni di quotazione giornaliera del titolo Benetton Group.

## 6. Misura della stazionarietà di un processo stocastico: l'esponente di Hurst e il Moto Browniano Frazionario

Nel paragrafo precedente abbiamo affermato che un moto browniano altro non è che un random walk le cui ampiezze dei salti sono date da una funzione di distribuzione gaussiana con media zero. Poiché ogni salto è definito temporalmente indipendente da tutto ciò che è avvenuto nel moto in precedenza, abbiamo che

$$\langle X(t) - X(0) \rangle = 0 \quad (6.1)$$

e

$$\sqrt{\langle (X(t) - X(0))^2 \rangle} \propto |t|^{\frac{1}{2}}. \quad (6.2)$$

La (6.1) indica che la media del moto browniano è nulla, mentre la (6.2) afferma che la deviazione standard del moto è proporzionale ad una funzione potenza del tempo con esponente  $\frac{1}{2}$ , ossia aumenta secondo una legge di potenza della variabile temporale  $t$ . Possiamo generalizzare la (6.2) nel modo seguente

$$\sqrt{\langle (X_H(t) - X_H(0))^2 \rangle} \propto |t|^H \quad (6.3)$$

con  $0 < H < 1$  che prende il nome di *esponente di Hurst*.

La (6.3) è la funzione deviazione standard di un moto browniano generalizzato, proporzionale ad una legge di potenza del tempo con esponente  $H$ . Ritroviamo il classico moto browniano per  $H = \frac{1}{2}$ .

Costruiamo ora una nuova serie temporale  $\{\xi_H(t), t \in T\}$  formata dagli incrementi di un moto browniano generalizzato, secondo la (6.3). La nuova variabile aleatoria è

$$\xi_H(t) = X_H(t + \tau) - X_H(t) \quad (6.4)$$

dove  $\tau$  è un certo ritardo temporale. La serie temporale  $\{\xi_H(t), t \in T\}$  prende il nome di *rumore gaussiano frazionario*.

Mendelbrot e Van Ness introdussero il *moto browniano frazionario* come generalizzazione del classico moto browniano, la cui funzione deviazione standard è rappresentata da una legge di potenza del tempo con esponente  $H$ , in accordo con la (6.3). Nel moto browniano frazionario i salti restano comunque distribuiti secondo una distribuzione gaussiana, ma questa volta gli incrementi dati dalla (6.4) non sono indipendenti, ma dipendono dalla storia del moto.

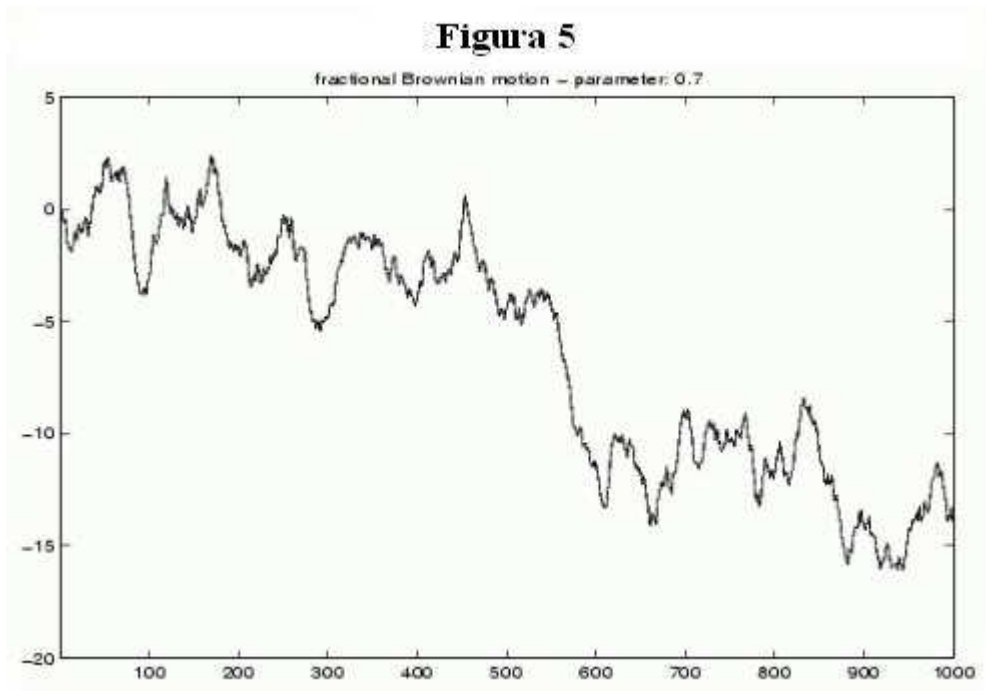
Alla fine del paragrafo 2 abbiamo affermato che la funzione di autocorrelazione di un processo stocastico rappresenta un valido strumento per stabilirne la stazionarietà. Così definiamo formalmente la funzione di autocorrelazione per un moto browniano in questo modo

$$\rho(\tau) \equiv \frac{\langle (\xi_H(t) - \langle \xi_H(t) \rangle)(\xi_H(t - \tau) - \langle \xi_H(t - \tau) \rangle) \rangle}{\sqrt{\langle (\xi_H(t) - \langle \xi_H(t) \rangle)^2 \rangle \langle (\xi_H(t - \tau) - \langle \xi_H(t - \tau) \rangle)^2 \rangle}} \quad (6.5)$$

La (6.5) prende una forma più facile se è riferita ad un moto browniano frazionario. Infatti, per un moto browniano frazionario la funzione di autocorrelazione è definita nel seguente modo

$$\rho(\tau) = \frac{1}{2} (|\tau + 1|^{2H} - 2|\tau|^{2H} + |\tau - 1|^{2H}) \quad (6.6)$$

La (6.6) si annulla in due casi: per  $H = \frac{1}{2}$ , cioè se il moto è browniano classico, e per  $\tau = 0$ , caso banale in cui la correlazione è ovviamente uno.



Nella figura 5 mostriamo un esempio di realizzazione di un moto browniano frazionario monodimensionale con esponente di Hurst  $H = 0.7$ .

Nel paragrafo 2 abbiamo affermato che se la funzione di autocorrelazione di un processo stocastico decade velocemente a zero allora il processo è stazionario, proprio perché gli incrementi rappresentano delle variabili aleatorie indipendenti. Possiamo dunque concludere che un moto browniano classico è un processo stocastico stazionario.

Cosa possiamo dire invece del moto browniano frazionario, per il quale gli incrementi non sono variabili aleatorie indipendenti? Per rispondere a questa domanda è necessario sviluppare la (6.6)

per valori di  $H \neq \frac{1}{2}$  e per ritardi temporali  $\tau$  sempre più ampi.

Attraverso alcuni calcoli (che al momento omettiamo) si può dimostrare che

$$\lim_{\tau \rightarrow \infty} \rho(\tau) \propto \tau^{2H-2} \quad (6.7)$$

ossia che per ritardi temporali  $\tau$  via via più grandi la funzione di autocorrelazione (6.6) tende ad una legge di potenza del tempo con esponente  $2H - 2$ , dove  $H$  è sempre l'esponente di Hurst.

La (6.7) dimostra che la funzione di autocorrelazione di un moto browniano frazionario dipende dal tempo (decade secondo una legge di potenza) in modo non banale. Possiamo quindi concludere che un moto browniano frazionario è un processo stocastico non stazionario.

Così come avviene per il moto browniano classico, anche il moto browniano frazionario presenta caratteristiche di autosimilarità con esponente caratteristico  $H$

$$X_H(\lambda t) = \lambda^H X_H(t).$$

Questo significa che le serie statistiche ci appaiono identiche, dilatando o contraendo l'asse temporale di un fattore  $\lambda$  e lo spostamento  $X_H$  di un fattore  $\lambda^H$ . Quindi il moto browniano frazionario non possiede una scala temporale caratteristica e, quando consideriamo un moto browniano frazionario, ogni salto arbitrario può essere considerato di lunghezza unitaria senza per questo perdita di alcuna generalità.

Possiamo descrivere l'autosimilarità statistica di un moto browniano frazionario monodimensionale ricorrendo alla *dimensione frattale*  $D$  che è legata all'esponente di Hurst secondo la relazione

$$D = 2 - H. \text{ Per un moto browniano classico monodimensionale avremo semplicemente } D = \frac{3}{2}.$$

Come abbiamo avuto modo di dire e vedere precedentemente, in un moto browniano frazionario la funzione di autocorrelazione (6.6) decade lentamente nel tempo (legge di potenza). Il moto browniano frazionario presenta quindi effetti di memoria che possono essere più o meno estesi nel tempo, a seconda del valore dell'esponente di Hurst  $H$ .

Se le correlazioni degli incrementi del moto browniano frazionario sono positive (tendenza a mantenere lo stesso trend), cioè se  $H > \frac{1}{2}$ , il moto si dice essere *persistente*. Al contrario, se

$H < \frac{1}{2}$  le correlazioni sono negative (tendenza ad invertire il trend) e il moto risulta essere *antipersistente*.

### 7. Stima dell'esponente di Hurst: la De-trended Fluctuations Analysis (DFA)

Il metodo DFA è uno di quelli che vengono spesso utilizzati per misurare le correlazioni esistenti negli incrementi di processi stocastici. Esso è uno dei metodi migliori, almeno dal punto di vista operativo e computazionale.

Vediamo in che modo esso è strutturato. Dobbiamo dapprima considerare un processo stocastico (per esempio un moto browniano) formato da un random walk unidimensionale  $\{X(t), t \in T\}$ , che dividiamo in un certo numero di sottosequenze di ampiezza nota  $n$ .

Il prossimo passo è quello di operare un de-trending del processo stocastico  $\{X(t), t \in T\}$  utilizzando una regressione lineare, per ogni sottosequenza. Otteniamo quindi un de-trended walk interpolato come segue

$$X'(t) = a + b(t - t_0)$$

dove con  $t_0$  abbiamo indicato l'istante iniziale riferito ad ogni sottosequenza. Definiamo ora le *fluttuazioni DFA* come

$$F_D(n) = \sqrt{\langle \Delta X^2 \rangle},$$

dove

$$\Delta X = X(t) - X'(t)$$

rappresenta la distanza tra il processo stocastico di partenza  $\{X(t), t \in T\}$  e il de-trended walk  $\{X'(t), t \in T\}$ .

A questo punto non ci resta che definire l'*esponente* della DFA  $\alpha(n)$  come segue

$$\alpha(n) \equiv \frac{d \log F_D(n)}{d \log(n+3)} \tag{7.1}$$

Nella (7.1) la correzione (+3) nell'argomento del logaritmo al denominatore si rende necessaria per valori di  $n$  troppo piccoli. Possiamo operare una stima dell'esponente  $\alpha(n)$  attraverso un fit di questo tipo

$$\alpha(n) \approx \alpha = \frac{\Delta \log F_D(n)}{\Delta \log(n)},$$

dove  $\Delta \log(n) = 2^{\frac{1}{4}}$ , essendo  $2^{\frac{1}{4}}$  il rapporto tra due valori consecutivi di  $n$ .

Se gli incrementi del processo stocastico  $\{X(t), t \in T\}$  sono indipendenti tra di loro, quindi non correlati, otteniamo un valore di  $\alpha = \frac{1}{2}$ , mentre otteniamo  $\alpha \neq \frac{1}{2}$  in caso contrario (presenza di correlazione tra gli incrementi di  $\{X(t), t \in T\}$ ).

Un valore costante di  $\alpha$  denota anche una stabilità di scala, mentre la sua eventuale non costanza indicherebbe la perdita di stabilità di scala, risultando il metodo non applicabile al processo stocastico in esame.

Se il processo stocastico  $\{X(t), t \in T\}$  rappresenta un moto browniano frazionario, allora il coefficiente della DFA (7.1) coincide con l'esponente di Hurst. La DFA rappresenta quindi un metodo alternativo per stimare l'autocorrelazione di un moto browniano frazionario.

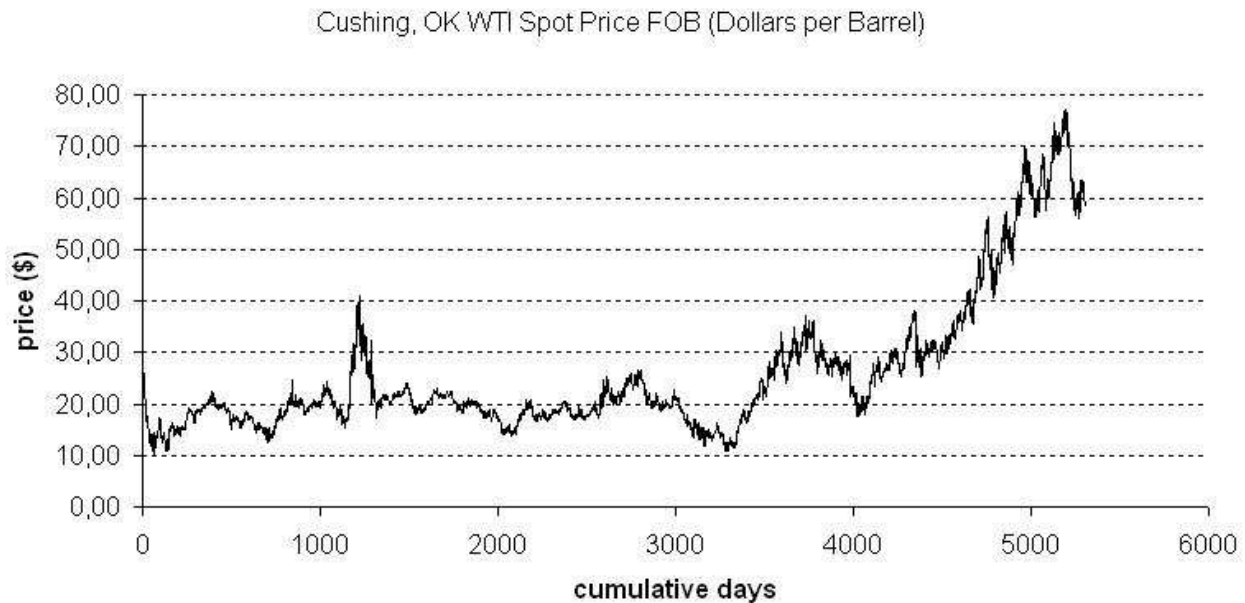
Bisogna però prestare molta attenzione! Se il processo stocastico  $\{X(t), t \in T\}$  non rappresenta un moto browniano frazionario, valori di  $\alpha(n) \neq \frac{1}{2}$  non necessariamente implicano presenze di effetti di memoria tra gli incrementi del processo stocastico in esame. Ne costituiscono esempi i processi stocastici markoviani, noti per essere stazionari e, di conseguenza, privi di memoria. In questi processi il passaggio ad un nuovo stato di sistema dipende unicamente da quello immediatamente precedente e non dal come si è giunti a tale stato. Alcuni processi di Markov presentano una stima dell'esponente DFA  $\alpha(n) \neq \frac{1}{2}$ , pur essendo stazionari e privi di memoria.

Per essere sicuri che il processo che stiamo analizzando sia un moto browniano frazionario bisogna che ci sia coerenza tra il comportamento della funzione di autocorrelazione e la stima dell'esponente di Hurst. Solo in questo caso possiamo affermare con certezza che il processo stocastico con cui abbiamo a che fare è un moto browniano frazionario.

### 8. Costruzione di un moto browniano classico da una serie temporale finanziaria

Dopo tanti spunti teorici cerchiamo di costruire un moto browniano partendo da un processo stocastico particolare, rappresentato da una serie temporale finanziaria. Si tratta di ventuno anni (1986 - 2007) di quotazione giornaliera del prezzo del petrolio in dollari per barile al NYSE (New York Stocks Exchange), ossia la borsa valori americana.

Nella figura che segue rappresentiamo il grafico storico della serie temporale appena presentata.



Facciamo una prima analisi qualitativa semplicemente osservando l'andamento del grafico. Il primo dato allarmante, ma che c'entra poco con la nostra analisi, è la pressoché costante crescita del prezzo del petrolio col passare del tempo. Invece, giusto per tornare ai nostri scopi, notiamo la presenza di una forte componente secolare (rialzo del prezzo nel tempo per le evidenti e risapute varie influenze esterne, quali invasione dell'Iraq, tagli e/o aumenti di produzione da parte dell'OPEC, fattori internazionali destabilizzanti). Non risultano tuttavia evidenti componenti periodiche significative. Possiamo dunque considerare questa serie di dati come un processo stocastico che presenta sicuramente una componente secolare e (bisogna dimostrarlo) una componente almeno debolmente stazionaria

$$X(t) = s(t) + Y(t)$$

dove, coerentemente con quanto affermato nei paragrafi precedenti, indichiamo con  $X(t)$  il processo stocastico di base, con  $s(t)$  la componente secolare, con  $Y(t)$  la parte stazionaria.

Adesso concentriamo i nostri sforzi per eliminare dapprima la componente secolare al fine di ottenere un processo stocastico stazionario che, in fine, dimostreremo essere un moto browniano classico.

Costruiamo, partendo dalla nostra serie temporale iniziale, un nuovo processo stocastico applicando il metodo della differenziazione esposto nel paragrafo 4. Chiameremo questo processo *ritorno del prezzo*, sarà del tipo  $\{r(t), t \in T\}$ , e definito in questo modo

$$r(t) \equiv \log \frac{X(t+1)}{X(t)} \tag{8.1}$$

dove la variabile aleatoria  $X(t)$ , come abbiamo già detto in precedenza, rappresenta il prezzo del petrolio al tempo  $t$ .

La variabile aleatoria (8.1), così come è stata definita, presenta un piccolo problema eliminabile, cioè rappresenta un processo stocastico che, per la serie che stiamo esaminando, è tale da risultare  $\langle r(t) \rangle \neq 0$ . Eliminiamo il problema normalizzando il processo, semplicemente traslandolo in modo da ottenerne uno nuovo che abbia media nulla. Introduciamo un nuovo processo stocastico, chiamandolo *ritorno normalizzato* del prezzo (NR), che indichiamo con  $\{\overline{r}(t), t \in T\}$ , definito nel modo seguente

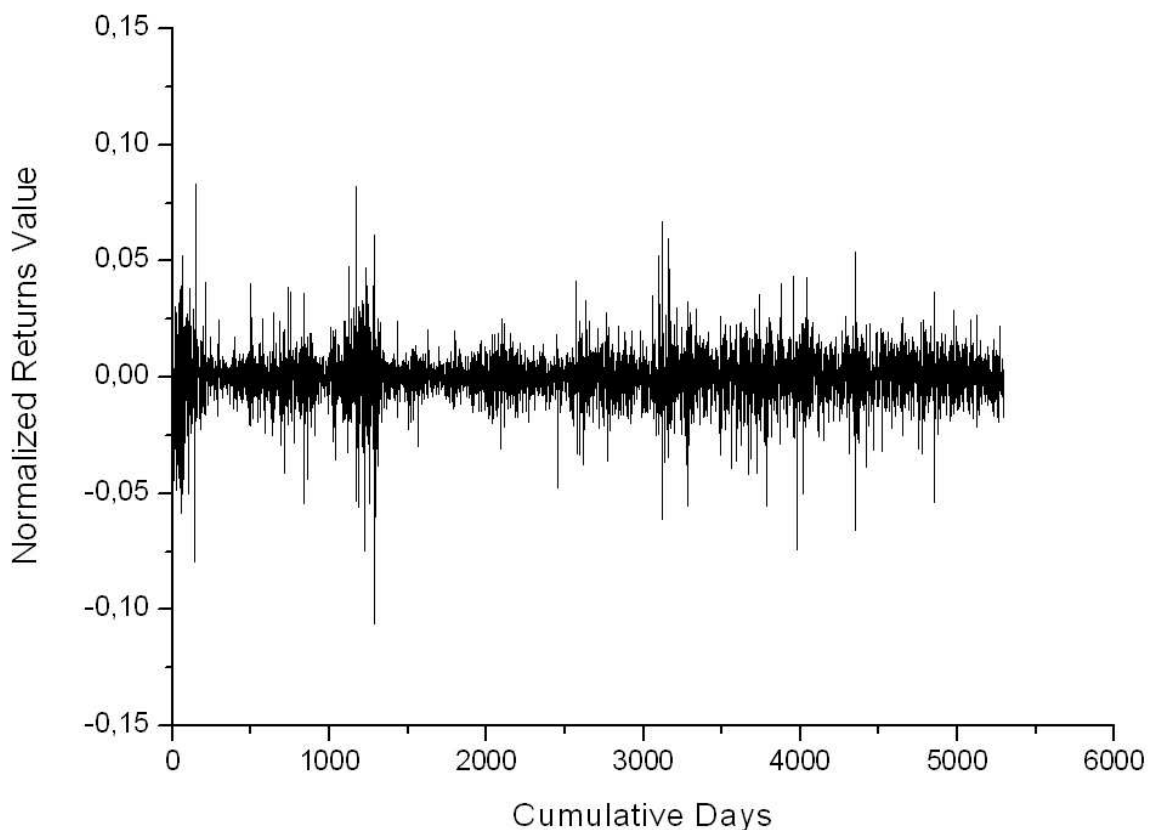
$$\overline{r}(t) \equiv r(t) - \langle r(t) \rangle \tag{8.2}$$

dove con  $\langle r(t) \rangle$  abbiamo indicato il valor medio dei ritorni su tutta la serie temporale  $\{r(t), t \in T\}$ .

La (8.2) rappresenta un processo stocastico di buona definizione, tale che  $\langle \overline{r}(t) \rangle = 0$ .

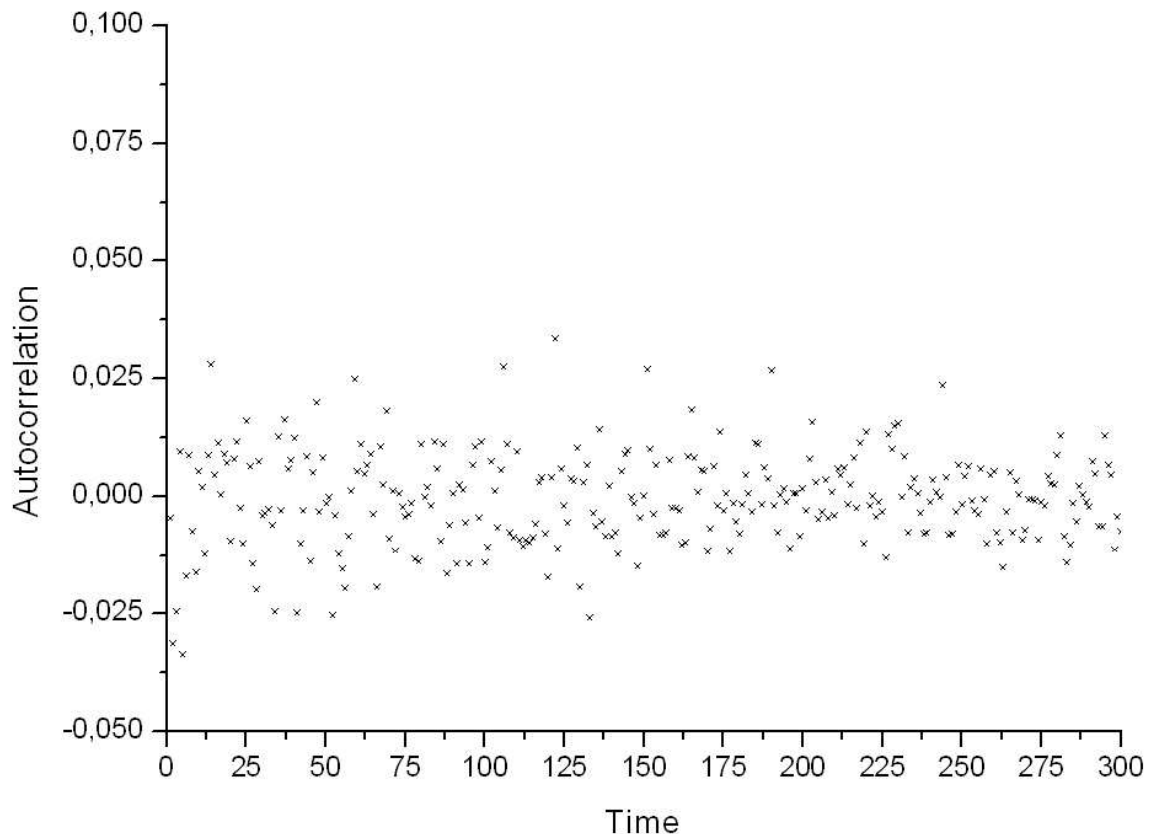
Il passo successivo è di dimostrare che la serie temporale (8.2) è un moto browniano classico, quindi un processo stocastico stazionario, identificabile con il sopra citato  $Y(t)$ .

Nella figura che segue mostriamo il grafico delle fluttuazioni temporali dei ritorni normalizzati del prezzo del petrolio.



Il grafico mostra chiaramente come la media dei ritorni normalizzati sia nulla, mentre sono ben visibili fluttuazioni “estreme” in corrispondenza di bruschi salti del valore del prezzo del petrolio in istanti di tempo immediatamente successivi.

Ci chiediamo se questi incrementi abbiano o meno delle correlazioni temporali, ossia se siano variabili aleatorie indipendenti o se invece dimostrano un legame temporalmente quantificabile con le precedenti. Per far ciò calcoliamo la funzione di autocorrelazione (6.5) riferita al processo stocastico (8.2) (come indicato nei paragrafi precedenti), mostrandone il grafico.



Il grafico non lascia spazi a dubbi. La funzione di autocorrelazione collassa istantaneamente sullo zero e questo dimostra che gli incrementi definiti nella (8.2) sono variabili aleatorie indipendenti e che non presentano correlazioni temporali apprezzabili. Dal risultato ottenuto possiamo concludere certamente che il processo stocastico  $\{r(t), t \in T\}$  è stazionario, ma non abbiamo ancora dimostrato che si tratta di un moto browniano classico.

Per raggiungere il nostro risultato ultimo, cioè di generare un moto browniano classico partendo dall'analisi statistica di una serie di dati finanziari, dobbiamo cercare la coerenza tra l'assenza di memoria sancita dalla funzione di autocorrelazione e la stima dell'esponente di Hurst operata col metodo DFA, introdotto nel paragrafo 7.

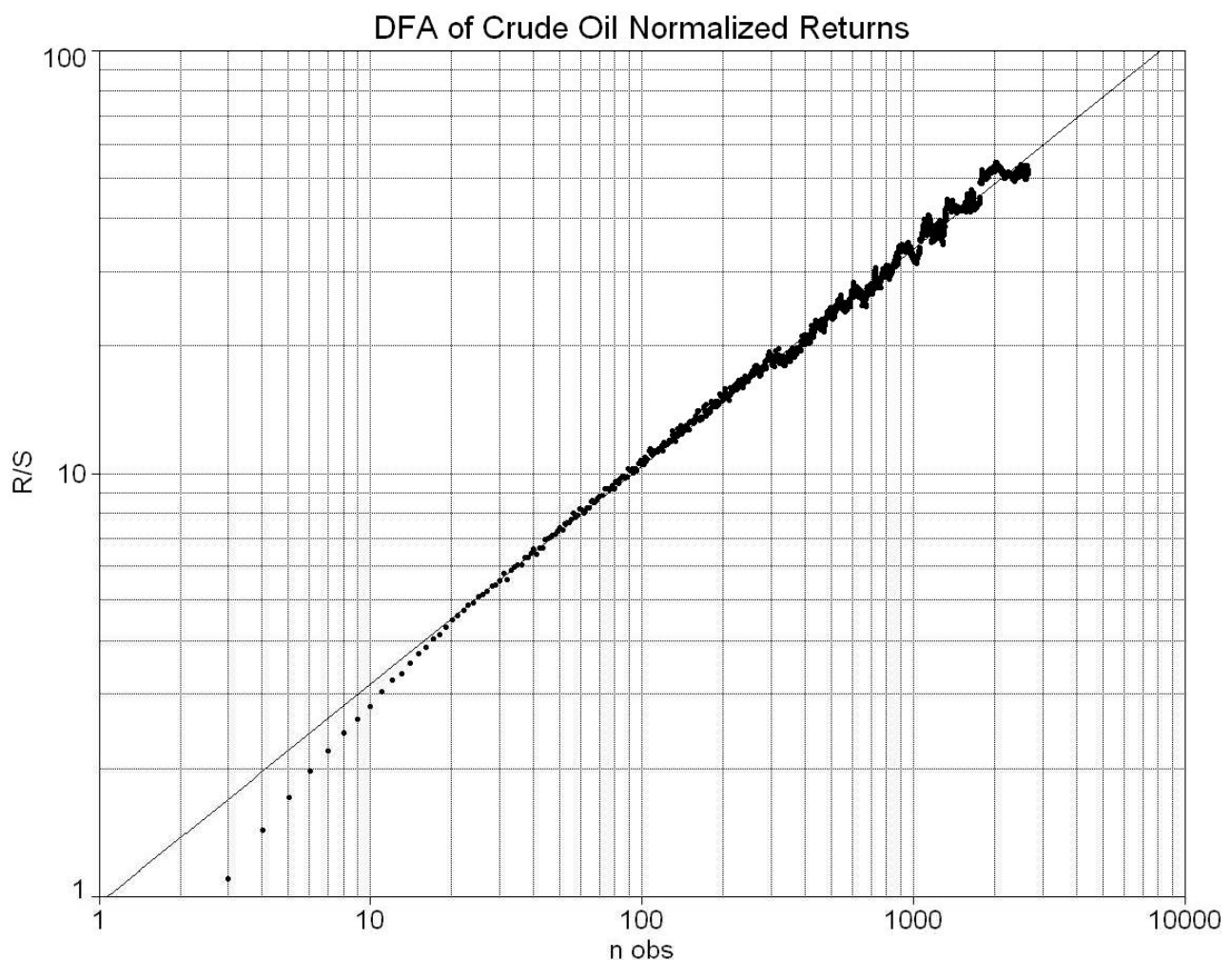
L'analisi DFA svolta sulla nostra serie temporale dei ritorni normalizzati restituisce una stima del valore dell'esponente di Hurst  $H = 0.54 \pm 0.03$ . Dal fit che fra poco vedremo in figura risulta evidente che questo valore è stabile, rendendo quindi il risultato del metodo affidabile, in accordo con quanto detto nel paragrafo 7.

Ricordando che per quanto riguarda un moto browniano classico ci aspettiamo un valore teorico di  $H = 0.5$ , essendo gli incrementi dei ritorni normalizzati variabili aleatorie indipendenti (dimostrato in precedenza dall'analisi della funzione di autocorrelazione), possiamo concludere dicendo che esiste coerenza tra la stima dell'esponente di Hurst operata col metodo DFA ed il risultato ottenuto con la funzione di autocorrelazione.

Possiamo quindi concludere dicendo che la serie temporale  $\{\overline{r}(t), t \in T\}$  dei ritorni normalizzati del prezzo del petrolio, è un processo stocastico stazionario identificabile con un moto browniano classico. Ora possiamo identificare il processo stocastico  $\{\overline{r}(t), t \in T\}$  col processo stocastico stazionario  $\{Y(t), t \in T\}$  che stavamo cercando. Risulta quindi

$$X(t) = s(t) + \overline{r}(t)$$

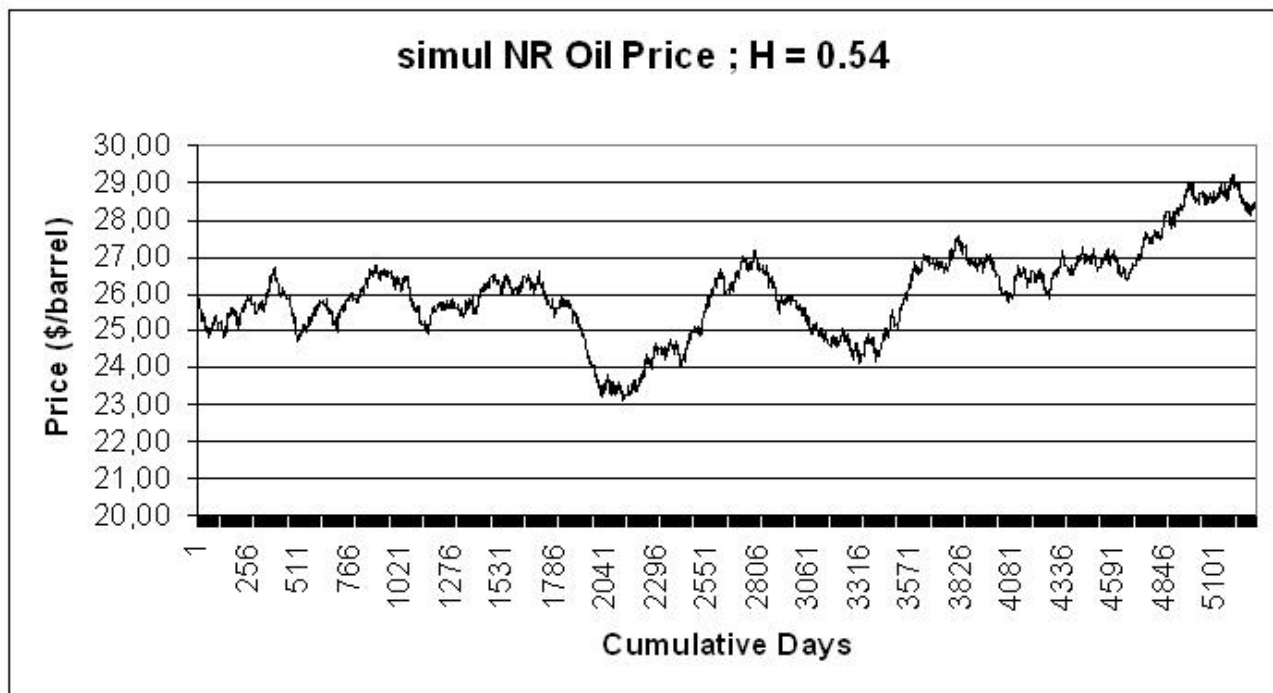
dove  $X(t)$  è il prezzo del petrolio all'istante  $t$ ,  $s(t)$  la sua componente secolare ed  $\overline{r}(t)$  il ritorno del prezzo normalizzato all'istante  $t$ .



Nella figura mostriamo il fit lineare operato col metodo DFA per stimare l'esponente di Hurst per la serie temporale dei ritorni normalizzati del prezzo del petrolio  $\{\overline{r}(t), t \in T\}$ .

Utilizzando la serie temporale dei ritorni normalizzati abbiamo, inoltre, costruito un random walk unidimensionale che simula l'andamento del prezzo del petrolio per lo stesso periodo temporale indicato nella serie temporale  $X(t)$ .

Il moto browniano monodimensionale risultante lo mostriamo nella figura che segue



Dalla figura possiamo notare come il prezzo oscilla veramente poco (si va da un minimo di poco superiore ai 23\$ ad un massimo di poco superiore ai 29\$). Questo risultato lo avremmo in caso di mercato perfettamente efficiente, cioè depurato da qualsiasi influenza esterna. Concludiamo dicendo che dall'analisi dell'andamento del prezzo del petrolio è stato possibile costruire un moto browniano classico, ma che non risulta utile al fine di operare delle previsioni.

### 9. Costruzione di un moto browniano frazionario da una serie temporale finanziaria

Consideriamo la stessa sequenza temporale  $\{X(t), t \in T\}$  utilizzata nel paragrafo 8, ossia la quotazione giornaliera ventennale del prezzo del petrolio per barile, quotato alla NYSE dal 1986 al 2007.

Come abbiamo avuto modo di osservare, essa presenta una chiara componente secolare che elimineremo così come abbiamo fatto nel paragrafo 8. Consideriamo la serie temporale dei ritorni normalizzati del prezzo del petrolio  $\{\overline{r(t)}, t \in T\}$  e la modifichiamo in modo da ottenere un nuovo processo stocastico  $\{\overline{r_{abs}(t)}, t \in T\}$ , che definiamo in questo modo

$$\overline{r_{abs}(t)} \equiv |\overline{r(t)}| - \langle \overline{r(t)} \rangle. \quad (9.1)$$

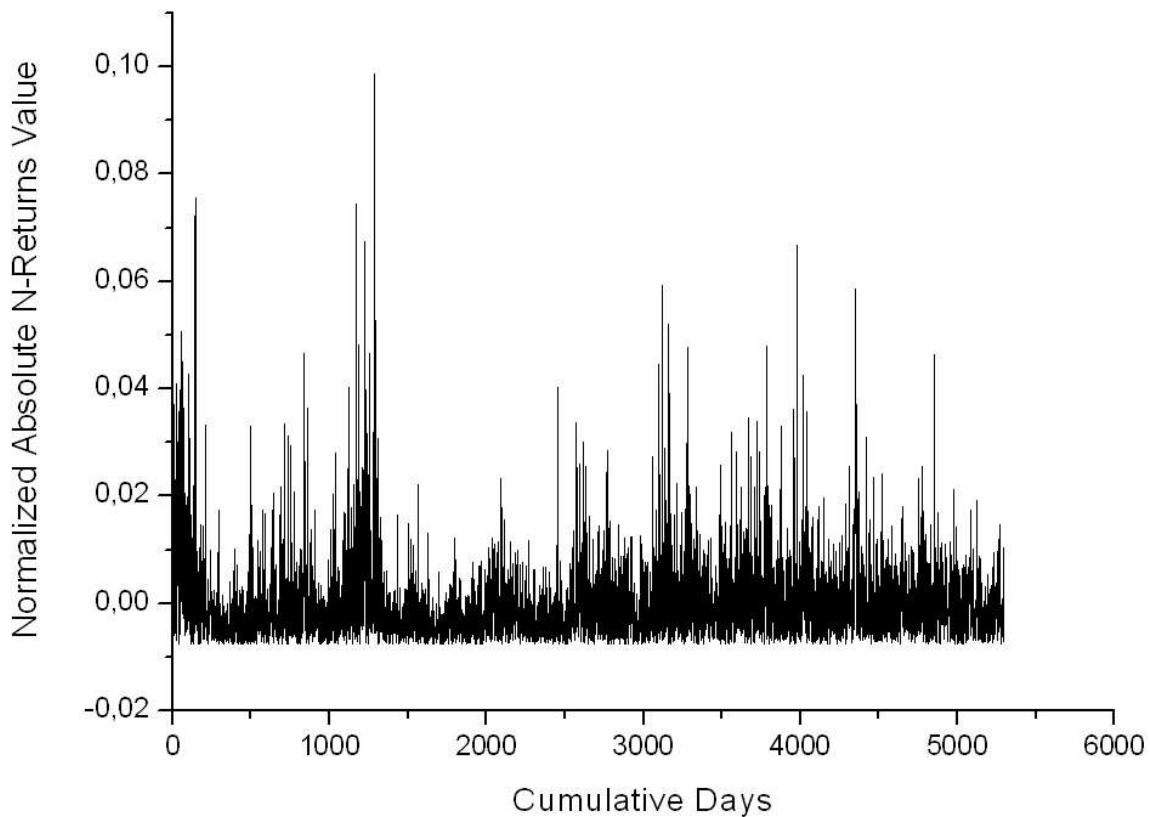
Il processo stocastico (9.1) lo chiamiamo *ritorno assoluto normalizzato* del prezzo (NANR).

Il nostro scopo è ora quello di trovare un processo stocastico non stazionario  $\{F(t), t \in T\}$ , assimilabile ad un moto browniano frazionario, in modo che risulti

$$X(t) = s'(t) + F(t) \quad (9.2)$$

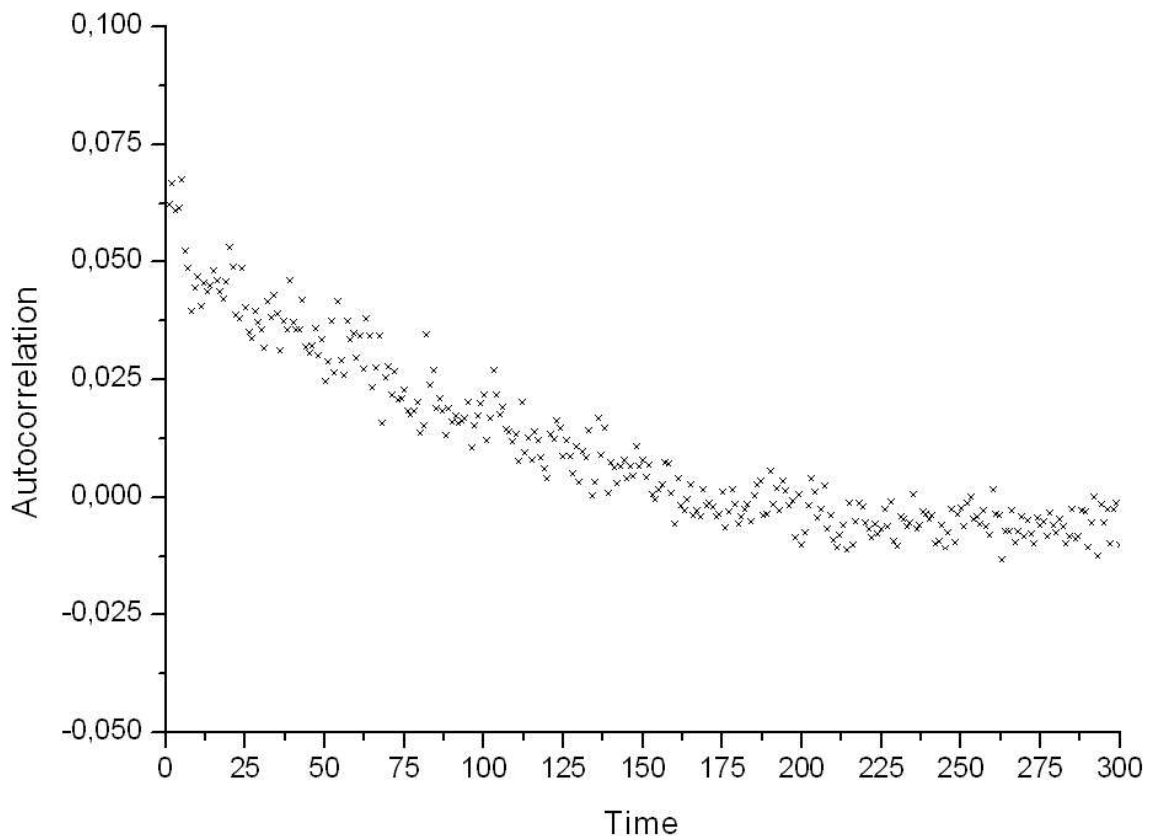
dove  $s'(t)$  è una componente secolare generalmente diversa da quella individuata nel paragrafo 8.

Nella figura seguente mostriamo le fluttuazioni nel tempo della variabile aleatoria dei ritorni assoluti normalizzati del prezzo del petrolio.



Per come è stata costruita la serie temporale  $\{\overline{r_{abs}(t)}, t \in T\}$  possiamo dire che anch'essa ha media nulla. Anche in questo grafico possiamo notare la presenza di eccessive fluttuazioni dovute a sbalzi improvvisi del prezzo del petrolio ad istanti di tempo immediatamente successivi.

In accordo con quanto fatto nel paragrafo precedente, calcoliamo dapprima la funzione di autocorrelazione della serie temporale dei ritorni assoluti normalizzati e mostriamone il grafico nella figura seguente



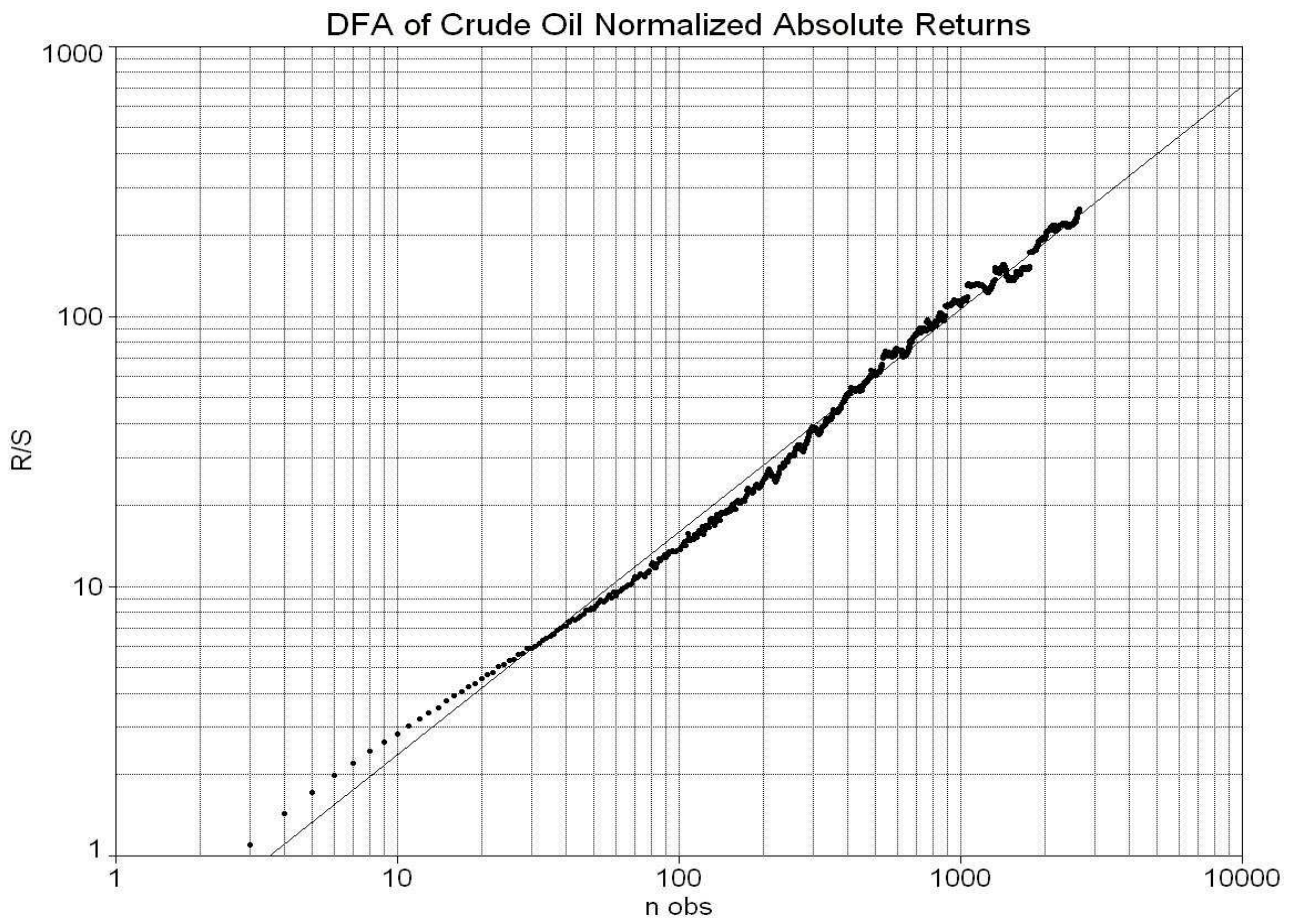
Questa volta notiamo che la funzione di autocorrelazione non decade istantaneamente a zero, così come invece accadeva per la serie temporale dei ritorni normalizzati. Questo comportamento della funzione di autocorrelazione denota che gli incrementi (9.1) non sono indipendenti, ma mostrano correlazioni temporali non banali. Il processo stocastico  $\{\overline{r_{abs}(t)}, t \in T\}$  è quindi non stazionario così che possiamo avere speranza che possa essere assimilabile ad un moto browniano frazionario. Questo però lo dobbiamo dimostrare attraverso la ricerca della coerenza tra comportamento della funzione di autocorrelazione e la stima dell'esponente di Hurst col metodo DFA.

Possiamo intanto escludere che il processo sia markoviano, proprio per come si presenta il grafico della funzione di autocorrelazione. Il mancato decadimento improvviso della funzione di autocorrelazione per la serie temporale  $\{\overline{r_{abs}(t)}, t \in T\}$  non è compatibile con le proprietà dei processi markoviani, per i quali lo stato ad ogni passo successivo dipende esclusivamente dal precedente e non da come esso sia stato raggiunto storicamente.

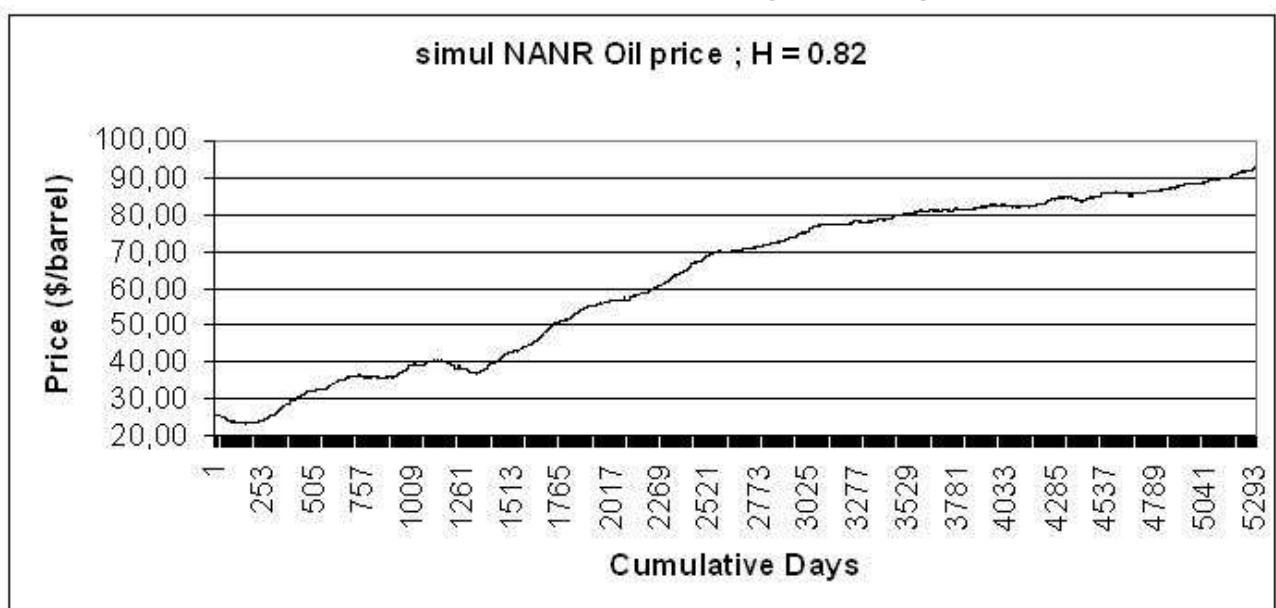
Fortunatamente riusciamo a trovare quella coerenza che stavamo cercando perché l'analisi DFA per la serie temporale  $\{\overline{r_{abs}(t)}, t \in T\}$  mostra un valore per il coefficiente di Hurst stimato come  $H = 0.82 \pm 0.02$ , ben diverso da quello stimato per la serie temporale dei ritorni normalizzati del

paragrafo precedente. La persistenza è evidente, così come lo è la presenza di memoria nel processo in esame.

Nella figura che segue mostriamo proprio il fit operato con l'analisi DFA dell'esponente di Hurst. Anche in questo caso è ben visibile la stabilità del parametro.



Quest'ultima analisi condotta ci permette di concludere che il processo stocastico non stazionario  $\{\overline{r}_{abs}(t), t \in T\}$  rappresenta un moto browniano frazionario, per cui possiamo concludere che la serie temporale  $\{F(t), t \in T\}$  che cercavamo nella (9.2) è proprio  $\{\overline{r}_{abs}(t), t \in T\}$ .



---

Come abbiamo fatto nel paragrafo precedente, utilizzando la serie temporale dei ritorni assoluti normalizzati abbiamo operato una simulazione di come oscillerebbe il prezzo del petrolio nello stesso tempo della serie temporale  $X(t)$ .

Come possiamo vedere nella figura precedente, un moto browniano frazionario risulta essere un buon modello per operare delle previsioni. La forte presenza di effetti di memoria, sancita anche dall'analisi DFA, rende il moto persistente in modo abbastanza robusto e visibile. La nostra simulazione prevede un prezzo finale di poco superiore ai 90\$ al barile contro una quotazione reale, allo stesso periodo, di circa 80\$ al barile.

---

## 10. Bibliografia

- **B. B. Mandelbrot**, *Fractals and Scaling in Finance*. Springer, 1997
- **E. Montroll, M. Shlesinger**, *The Wonderful World of Random Walks*. Nonequilibrium Phenomena II, North-Holland, Amsterdam, 1966
- **K. E. Bassler, G. H. Gunaratne, J. L. McCauley**, *Markov Processes, Hurst Exponent and Nonlinear Diffusion Equations (with application to finance)*, Physics Department, University of Houston, Houston, Tx. 77204
- **V. Plerou, P. Gopikrishnan, B. Rosenow, L. A. N. Amaral, H. E. Stanley**, *Econophysics: financial time series from a statistical point of view*. Center for Polymer Studies and Department of Physics, Boston University, Boston, MA02215, USA
- **V. Plerou, P. Gopikrishnan, L. A. N. Amaral, X. Gabaix, H. E. Stanley**, *Economic fluctuations and anomalous diffusion*. Physical Review E, vol 62, n°3, set 2000