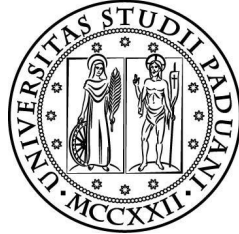


DIPARTIMENTO DI FISICA E ASTRONOMIA “GALILEO GALILEI”

CORSO DI LAUREA MAGISTRALE IN FISICA



UNIVERSITÀ DEGLI STUDI DI PADOVA

TESI DI LAUREA MAGISTRALE

Analisi e modellizzazione di serie temporali finanziarie

Relatore: prof. Attilio Stella,

Corelatore Dott. Michele Caraglio

Controrelatore: prof. Baiesi

Laureando: Giambattista Canilli

Anno accademico 2014 2015

Sommario

Introduzione.....	5
Capitolo 1.....	8
1.1 I MERCATI FINANZIARI	8
1.2 ECONOFISICA: CENNI STORICI	9
1.2.1 Random walk	9
1.2.2 Black and Scholes e il moto Browniano Geometrico	10
1.2.3 Mandelbrot e le distribuzioni non gaussiane.....	10
1.2.4 Engle e il modello ARCH.....	11
1.3 PROPRIETA' STATISTICHE DELLE SERIE FINANZIARIE	12
1.3.1 Stazionarietà	13
1.3.2 Ergodicità	13
1.3.3 Distribuzione dei ritorni: Fat tails	14
1.4 APPROCCIO DI ENSEMBLE.....	14
1.4.1 Analisi del momento secondo dei ritorni logaritmici	16
1.5 ANALISI DELLE CORRELAZIONI LINEARI	18
1.5.1 Evoluzione dell'autocorrelazione.....	19
Capitolo 2 LO SCALING	23
2.1 PROCESSI STOCASTICI	23
2.2 EVIDENZE DI SCALING	25
2.3 SCALING	27
2.3.1 Estensione del modello alla sessione pomeridiana	31
Capitolo 3 IL TEMPO IN FINANZA.....	33
3.1 Tempo e scaling.....	33
3.2 Oltre l'intraday: ritorni notturni e giornalieri ed unificazione	35
Capitolo 4 TRADING.....	41
4.1 FUNZIONE QUANTILE	41
4.2 STRATEGIE DI TRADING	43
4.3 STRATEGIE BASATE SUI QUANTILI.....	45
4.3.1 Strategia barriere Q Semplici (SQS).....	45
4.3.2 Strategia barriere Q a Bande divise (SQD).....	47
4.3.3 Strategia barriere Q a Bande di ST.DEV (SQV)	50
4.3.4 Strategia a Barre di quantile disaccoppiate	51

4.3.5	Strategia con take profit condizionato	52
4.4	STRATEGIE TRADING VELOCI	55
4.4.1	Strategia Trading attivo con densità di correlazione.....	55
	Conclusioni	58
Appendice A	Il test di Kolmogorov-Smirnov.....	60
	Bibliografia.....	61

Introduzione

Questo lavoro tratta l'analisi e la modellizzazione delle serie temporali finanziarie. Si inserisce in quel campo di ricerca che ha visto molti fisici contribuire nel settore della finanza, iniziato da Bachelier nei primi anni del secolo scorso [1], per poi proseguire con Mandelbrot [2] fino ad arrivare alle moderne teorie dello scaling di Mantegna e Stanley [3]. In questa tesi si vuole utilizzare un approccio basato appunto sullo scaling per analizzare in maniera semplice dati presi a differenti scale di tempo e modellizzare le dinamiche del mercato.

Ci si è mossi a partire da considerazioni di proprietà statistiche robuste [4], che possono essere riconosciute analizzando le serie storiche finanziarie di differenti mercati e relative a differenti periodi. Su questi temi la meccanica statistica può apportare un importante contributo, cercando di trovare modelli capaci di riprodurre quanto più correttamente possibile i comportamenti empirici osservati.

Quando ci si deve confrontare con lunghe serie temporali, una importante questione da risolvere è il modo con cui trattare le serie di dati. Se è valida un'ipotesi di stazionarietà e di ergodicità per il processo stocastico che si postula abbia generato la serie, allora è possibile attraverso delle medie temporali risalire ai valori di aspettazione delle quantità prese in esame. Tuttavia questo approccio, quando ci si confronta con una unica realizzazione del processo può risultare problematico. In alcuni casi i dati presentano palesi comportamenti non stazionari [5]. Si è seguito, in questo lavoro, l'approccio proposto da *Bassler et al* [6]. Facendo riferimento a quantità come il momento secondo dei ritorni, o l'autocorrelazione di volatilità, è emerso un comportamento non stazionario nella scala temporale inferiore al giorno (*intraday*), ma è stata ravvisata anche una forte periodicità sulla scala temporale superiore al giorno (*interday*). In questo senso, la periodicità su scale interday rafforza l'idea che possa valere una stazionarietà a scale superiori al giorno, mentre la non stazionarietà nell'*intraday* indica la necessità di ipotizzare un unico processo stocastico che si sviluppi per l'intera durata della finestra *intraday*. Alla luce di quanto detto, è risultato naturale considerare che il processo stocastico responsabile della generazione dei ritorni abbia durata pari ad una giornata borsistica. In questo modo è possibile applicare alla serie di dati un approccio di ensemble, in cui il singolo processo dura da apertura a chiusura dei mercati. In questo modo, per ottenere i valori di aspettazione delle quantità cercate si utilizzano medie di ensemble, in cui ciascuna sottoserie relativa ad una stessa giornata di mercato funge da realizzazione indipendente del medesimo processo. I dati sono stati organizzati in maniera da suddividere l'unica realizzazione iniziale del processo in tante sottoserie che avessero durata pari ad una giornata borsistica ovvero la periodicità delle quantità analizzate.

Una volta definito l'ensemble come strumento appropriato per trattare correttamente le lunghe serie di dati, si è passati ad analizzare il ruolo che le correlazioni lineari svolgono nel mercato reale. Sono state analizzate le correlazioni lineari dei ritorni e le correlazioni del valore assoluto dei ritorni. Le prime sono responsabili di possibilità di

arbitraggio nel mercato, le seconde di fenomeni di persistenza di volatilità. Inoltre sono state studiate queste due quantità per tempi diversi nell'arco di 30 anni, corrispondente alla durata dei dati a nostra disposizione, dal 1985 al 2013, ed è risultato evidente una marcata evoluzione in questo senso: la correlazione diminuisce col procedere del tempo. Dato che se il mercato ha una minore correlazione questo fatto significa una minore possibilità di arbitraggio e dunque una maggiore efficienza, allora l'evoluzione riscontrata delle correlazioni significa una progressiva evoluzione del mercato verso una maggiore efficienza.

Il passo successivo è stato quello di descrivere il sistema a differenti scale temporali, con lo scopo di unificarle attraverso lo scaling, come proposto da Kadanoff [7], sia per permettere una più accurata analisi delle proprietà, sia per mettere in luce lo stretto legame che un sistema basato sulla finanza può avere con la meccanica statistica.

Un'analisi di scaling significa trovare delle relazioni fra le variabili che descrivono il sistema a differenti scale, in maniera da unificare dati presi in condizioni differenti sotto un'unica descrizione globale. Nel contesto preso in esame la variabile considerata è il ritorno, che rappresenta la variazione del prezzo dell'asset calcolato in un ben definito intervallo di tempo. Si scopre analizzando le distribuzioni empiriche dei ritorni rispetto a diversi intervalli di tempo, che queste soddisfano una simmetria di scaling. Riscalando opportunamente la funzione di distribuzione e riscalando gli argomenti, tutte le distribuzioni dei ritorni relativi a diversi intervalli di tempo collassano approssimativamente in una medesima distribuzione.

La simmetria dello scaling riveste una particolare importanza in questo lavoro:, potremmo dire che essa ne costituisca la vera ossatura. Lo scaling infatti è stato sfruttato sotto diversi punti di vista:

Innanzitutto ci si collega alla linea già seguita in [8] per la costruzione di un modello dinamico basato sullo scaling che renda conto delle evidenze empiriche osservate. E, sulla scia dei risultati ottenuti col modello, è stata utilizzata la teoria dello scaling per definire delle strategie operative di trading che sfruttino le correlazioni lineari dei mercati finanziari. Infine, la simmetria dello scaling è servita per proporre una definizione consistente del tempo in finanza, in questo modo definito operativamente. Nelle analisi fatte finora il tempo o segue l'evolversi delle nostre azioni (tempo fisico), o può essere valutato in base al volume di affari del mercato (tempo di volume). In questo lavoro, invece, si cerca di definire il tempo sulla base dell'effetto che produce nell'evoluzione delle distribuzioni dei rendimenti. In questo modo l'idea di tempo trae la propria origine dall'elemento cardine dei mercati: l'evoluzione dei ritorni secondo le leggi dello scaling.

Il lavoro è così strutturato: si apre con una introduzione sui mercati finanziari, per spiegarne brevemente il comportamento ed il lessico di base, con una contestualizzazione storica dell'Econofisica, branca della scienza di relativa giovane età. Si accenna ai lavori di Bachelier [1], di Black and Scholes [17], di Mandelbrot [2] e di Engle.

A seguire si definiscono le caratteristiche principali del mercato, i cosiddetti fatti stilizzati, l'approccio di ensemble che si è voluto seguire e l'analisi delle correlazioni, nodo molto importante per quelle che saranno le strategie di trading.

Il capitolo secondo è dedicato alla tematica dello scaling. In particolare dopo aver spiegato quali sono le evidenze che hanno portato ad utilizzare lo scaling come linea guida del lavoro, seguendo le idee proposte in [5], [14], si procede alla formulazione di un modello basato sullo scaling capace di spiegare le evidenze empiriche descritte in precedenza.

Il capitolo terzo è dedicato al concetto di Tempo in finanza: in questo caso si utilizza lo scaling come linea guida per una definizione operativa di tempo. Si ricava la relazione che lega il fattore di scala delle distribuzioni di probabilità dei ritorni aggregati agli intervalli di tempo ai quali sono state generate, e si utilizza questo fatto per la definizione del tempo stesso in finanza.

Il quarto ed ultimo capitolo è dedicato alle strategie di trading. Utilizzando la formulazione teorica costruita nel capitolo secondo, sono state definite le funzioni quantile, funzioni che hanno lo scopo di tracciare per ogni istante della finestra temporale dell'intraday, (quella che va da apertura mercati a chiusura mercati) la regione nella quale è possibile trovare il valore dell'asset con una probabilità prestabilita. Le funzioni quantile vengono utilizzate nella formulazione delle strategie di trading. Dalle relazioni che legano la funzione quantile al prezzo dell'indice si traggono una serie di regole operative che vanno a formare una vera e propria strategia applicabile ai mercati. Si mettono a confronto varie strategie di trading per proporre diverse soluzioni che trovino migliore adattamento al contesto nel quale si opera e alle attitudini personali nei confronti del rischio.

In appendice troviamo la spiegazione del test di Kolmogorov-Smirnov per valutare la compatibilità fra distribuzioni cumulative, utilizzata nel corso di questo lavoro per verificare che i fit delle distribuzioni empiriche dei ritorni aggregati con le distribuzioni teoriche fossero consistenti.

Capitolo 1

1.1 I MERCATI FINANZIARI

I mercati Finanziari rappresentano senza dubbio un sistema complesso, a partire dalla loro struttura: milioni di investitori operano in diverse borse nelle quali si scambiano bonds, azioni, commodities, etc. Le decisioni su quali diverse strategie adottare influenzano le variazioni di prezzo e le variazioni di prezzo influenzano a loro volta la scelta delle strategie. Il mercato si ipotizza essere efficiente, vista l'impossibilità (quantomeno a livello teorico) dell' *arbitraggio*. Per arbitraggio si intende la possibilità di guadagno con assunzione di rischio nullo.

Si può semplificare la complessità del comportamento umano considerando la possibilità che un trader possa intraprendere solo tre semplici azioni: comprare, vendere o non partecipare al mercato. Se consideriamo casuale la decisione di un trader di intraprendere una di queste azioni, ognuna da assumersi con egual probabilità, abbiamo un modello semplificato di mercato che storicamente è stato anche il primo ad essere formulato, ad opera di L. Bachelier [1]. Egli era studente di Poincarè, e già ad inizio 1900 propose un modello di random walk a variabili indipendenti identicamente distribuite, quale processo alla base delle fluttuazioni degli assets. Tuttavia questo modello proposto da Bachelier non era sufficientemente sofisticato per essere di una qualche utilità pratica, in quanto portava a risultati non ottenibili in pratica, con prezzi degli asset negativi. Successivamente Black and Scholes [17] definirono un modello chiamato *Moto Browniano Geometrico*. Questo modello prevedeva l'utilizzo di processi con ritorni logaritmici casuali che seguissero la distribuzione Gaussiana. Non veniva tuttavia tenuta in considerazione l'esistenza di eventi estremi, che modificavano le code delle distribuzioni, invalidando la presunta gaussianità delle distribuzioni di ritorni. Storicamente si deve a Mandelbrot [2], l'introduzione per primo di distribuzioni di probabilità che non fossero gaussiane ma che tenessero conto delle "fat tail", riscontrate dall'analisi dei dati empirici.

Solamente negli ultimi venti anni ha preso piede la cosiddetta Econofisica [9], ovvero quella disciplina che applica metodi e tecniche sviluppate nell'ambito della matematica e della fisica a discipline economiche. Verso la fine degli anni novanta infatti molti fisici statistici approdarono allo studio dei complessi problemi posti dall'economia, soprattutto nell'ambito della finanza. Non ritenendo i metodi tradizionali applicati dagli economisti soddisfacenti, essi cercarono in primo luogo di analizzare e replicare i dati osservati, per poi andare a spiegare i processi che generano i fenomeni economico finanziari più generali.

1.2 ECONOFISICA: CENNI STORICI

Durante il secolo passato molti scienziati, fisici, matematici ed economisti hanno affrontato il problema della modellizzazione [10] dei mercati finanziari e più in generale della finanza. Di seguito, in estrema sintesi, si riporta una breve panoramica di questa storia.

1.2.1 **Random walk**

Il primo tentativo concreto di analizzare quantitativamente i mercati finanziari si ebbe con la modellizzazione basata sul concetto di *random walk*. Nel 1900 Bachelier sviluppò un modello di mercato [1] basato su un processo stocastico discreto a infiniti steps. Egli analizzò serie storiche finanziarie di varie decine di anni.

Si identifica con t un particolare istante di tempo all'interno della finestra temporale presa in esame, con $S(t)$ si identifica il prezzo di un dato asset al tempo t , e con

$$r_{\tau}^*(t) = S(t + \tau) - S(t) \quad 1.1$$

il ritorno semplice calcolato rispetto l'istante di tempo t ed al suo incremento $(t + \tau)$, dove τ rappresenta il *time lag*.

Considerando la distribuzione empirica dei ritorni, confrontando sottoserie temporali della serie principale, Bachelier osservò i seguenti fatti stilizzati:

1. Le distribuzioni empiriche non sono dipendenti dalla particolare finestra temporale presa in esame
2. I ritorni calcolati ad intervalli τ seguono distribuzioni Gaussiane a media zero e varianza $\sigma^2\tau$
3. i ritorni sono variabili casuali indipendenti .

Da queste proprietà egli ipotizzò che la distribuzione dei rendimenti sia del tipo:

$$p_{\tau}(r) = \frac{1}{\sqrt{2\pi\sigma^2\tau}} \exp\left(\frac{-r^2}{2\sigma^2\tau}\right) \quad 1.2$$

Dal momento che ipotizzò l'indipendenza statistica dei singoli ritorni la distribuzione di n ritorni successivi assume la forma:

$$p_{\tau}(r_1, r_2, \dots, r_n) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2\tau}} \exp\left(\frac{-r_i^2}{2\sigma^2\tau}\right) \quad 1.3$$

Il maggior difetto riscontrato da Bachelier fu che era aperta la possibilità a valori negativi dell'asset $S(t)$.

1.2.2 Black and Scholes e il moto Browniano Geometrico

Questo problema fu risolto da Black and Scholes [17] che introdussero i ritorni logaritmici,

$$r_\tau(t) = \ln S(t + \tau) - \ln S(t) \quad 1.4$$

Sviluppando un modello noto come moto Browniano Geometrico [12]. Le proprietà del modello di *Bachelier* (1,2,3) vengono mantenute, in particolare la proprietà di indipendenza dei ritorni, proprietà espressa da una autocorrelazione nulla, è alla base della gaussianità della distribuzione dei ritorni.

Il teorema del limite centrale, afferma [11] che qualora i ritorni siano statisticamente indipendenti la distribuzione della loro somma converge ad una normale. Il processo considerato, essendo i ritorni statisticamente indipendenti, non presenta correlazioni lineari, dunque non ha memoria. Tuttavia, anche in presenza di correlazione lineare strettamente nulla, la dipendenza dei ritorni, e quindi la non applicabilità del teorema, deriva dal fatto che i ritorni sono correlati non linearmente, ad esempio in modulo. Confrontando tuttavia le PDF ottenute con il moto Browniano Geometrico con quelle estratte dai dati empirici, si può notare come esse siano strutturalmente differenti.

Rispetto infatti ad una distribuzione gaussiana, che decade esponenzialmente, la distribuzione empirica dei ritorni presenta un andamento nelle code che si ipotizza essere a potenza.

1.2.3 Mandelbrot e le distribuzioni non gaussiane

Storicamente il primo che introdusse l'ipotesi di non gaussianità dei rendimenti nell'analisi di mercato fu Mandelbrot nel 1963 [2]. Egli studiò i prezzi del cotone lungo un arco temporale di un secolo, e modellizzò i dati secondo un modello a ritorni indipendenti che seguissero una distribuzione di Lèvy. Le ipotesi che fondano il suddetto processo (Lèvy flights) sono le seguenti:

- 1) *Ritorni considerati variabili casuali indipendenti*
- 2) *PDF dei ritorni stazionaria*
- 3) *CF (funzione caratteristica) della forma:*

$$p(r) = e^{-a|r|^\mu} \quad a > 0; 0 < \mu < 2 \quad 1.5$$

Con a, μ parametri della distribuzione.

Anche se la formulazione analitica delle distribuzioni di Lèvy è complessa, l'andamento asintotico [9] è del tipo:

$$L_\mu(r) \approx \frac{\mu A^\mu}{r^{1+\mu}} \quad 1.6$$

Questo particolare comportamento nelle code della distribuzione prende il nome di “*Pareto Tails*”. Se una variabile casuale si distribuisce seguendo una distribuzione a “*code grosse*”, allora eventi 5-sigma o 7-sigma, possibili ma improbabili qualora la distribuzione fosse normale, diventano in questo caso più frequenti.

Tuttavia utilizzando questa particolare distribuzione rimangono aperti problemi relativi al fatto che la varianza diverga e le correlazioni differenti dai dati empirici. Al primo problema pose soluzione utilizzando una Truncated Lèvy Distribution. Essa assume la forma (k parametro):

$$p(r) = \begin{cases} 0 & \text{se } r < -k \\ L_\mu(r) & \text{se } -k < r < +k \\ 0 & \text{se } r > +k \end{cases} \quad 1.7$$

Rimane tuttavia aperto il problema delle correlazioni. Solo nel 1968 Mandelbrot propose un modello che potesse spiegare fenomeni di clustering di volatilità e non indipendenza degli incrementi

1.2.4 Engle e il modello ARCH

Engle introdusse l'Auto regressive Linear Heteroskedasticity protocol (ARCH) nel quale i ritorni sono del tipo $r_k = \varepsilon_k \sigma_k$ con ε_k variabili casuali i.i.d. a media zero e varianza unitaria, e σ_k processi casuali legati alla grandezza delle fluttuazioni che determinano la varianza. In questo modo la volatilità aumenta in periodi di crescita o decrescita e diminuisce in periodi di movimenti laterali, comportamento anche conosciuto come heteroschedasticità. In particolare i modelli basati su questo protocollo presentano una volatilità che dipende attraverso una serie di parametri dagli N valori dei ritorni pregressi. Da come sono modulati i parametri deriva anche la forma della PDF.

$$\sigma_k^2 = a_0 + a_1 r_{k-1}^2 + \dots + a_N r_{k-N}^2 \quad 1.8$$

Essendo tutti gli a_i positivi questo assicura il fatto che la varianza sia positiva. Dunque, come richiesto dalla heteroschedasticità, i vari ritorni precedenti il ritorno r_k ne influenzano la varianza. Una evoluzione dei processi ARCH sono i processi GARCH, nei quali il processo di memoria oltre che tenere in considerazione gli N passati ritorni contempla anche le N passate varianze, secondo la:

$$\sigma_k^2 = a_0 + a_1 r_{k-1}^2 + \dots + a_N r_{k-N}^2 + \beta_1 \sigma_{k-1}^2 + \dots + \beta_M \sigma_{k-M}^2 \quad 1.9$$

Tuttavia questi processi, altamente utilizzati in finanza, non presentano marcati comportamenti di scaling.

1.3 PROPRIETA' STATISTICHE DELLE SERIE FINANZIARIE

Vengono definite e discusse in questo paragrafo alcune proprietà statistiche delle serie finanziarie [4].

In questo lavoro è stata utilizzata la serie storica dei ritorni dell'indice S&P 500 . Esso raccoglie al suo interno le 500 maggiori compagnie USA. I dati a disposizione, sono campionati a frequenza di 1 minuto, coprono l'arco di tempo che va dal 1985 al 2013, dalle 8:40 alle 15:00 di ogni giorno borsistico, con 380 minuti di contrattazione ogni giorno per una serie totale di durata $\sim 3 * 10^6$ minuti.

In generale il mercato è composto da numerosi asset, i quali, in base a dinamiche di domanda ed offerta, variano di prezzo. Il prezzo di un asset è definito in base all'ultimo prezzo utile al quale sia avvenuta una contrattazione terminata con successo, ovvero l'ultimo prezzo al quale l'asset sia stato comprato o venduto.

Si procede ora a fissare brevemente la notazione. Come già detto viene definito con $S(t)$ il prezzo dell'asset al tempo t . Risulta utile definire il ritorno logaritmico del prezzo del sottostante calcolato ad un tempo t e relativamente ad un intervallo $n\theta = \tau$. La notazione è tale per cui θ rappresenta una unità di tempo ed n un numero intero, dato che il processo è discreto e non continuo. E' possibile calcolare ritorni solo per multipli del tempo minimo ($\theta = 1$ minuto, con i dati a disposizione) Risulta quindi:

$$r_{n\theta}(t) = \ln (S(t)) - \ln (S(t - n\theta)) \quad 1.10$$

Appare quindi chiaro che se si prendono tre tempi in successione (equamente intervallati), t_0, t_1, t_2 , e tali che

$$t_1 - t_0 = \theta; \quad t_2 - t_1 = \theta$$

$$\begin{aligned} r_{2\theta}(t_2) &= r_{\theta}(t_2) + r_{\theta}(t_1) & 1.11 \\ &= \ln (S(t_2)) - \ln (S(t_2 - \theta)) + \ln(S(t_1)) \\ &\quad - \ln(S(t_1 - \theta)) = \ln(S(t_2)) - \ln(S(t_2 - 2\theta)) \\ &= r_{2\theta}(t_2) = r_{\tau}(t_2) \end{aligned}$$

Ne segue che si possono aggregare i singoli ritorni, effettuando una somma su n ritorni di durata θ , ottenendo così il ritorno relativo all'intervallo di tempo $n\theta = \tau$ su cui si effettua la sommatoria:

$$\begin{aligned}
r_\tau(t) = r_{n\theta}(t) &= r_\theta(t) + r_\theta(t - \theta) + \dots + r_\theta(t - (n - 1)\theta) \\
&= \sum_{i=1}^{n-1} r_\theta(t - i\theta)
\end{aligned}
\tag{1.12}$$

Per semplicità da ora in poi si chiamerà $r_\tau(t)$ il ritorno aggregato al tempo t e relativo ad un intervallo $\tau = n\theta$.

1.3.1 Stazionarietà

Affrontando l'analisi di lunghe serie di dati in successione temporale, al fine di avere assicurata la validità di una analisi statistica, deve valere l'ipotesi di stazionarietà. L'ipotesi di stazionarietà prevede che dati appartenenti a differenti periodi possono essere analizzati insieme in quanto le proprietà statistiche non dipendono dal tempo.

Ovviamente vi possono essere vari fattori che influenzano la validità di questa ipotesi, e possiamo dire che per quanto riguarda la stazionarietà delle serie che sono state analizzate sono emerse evidenze che non permettono di utilizzare l'ipotesi di stazionarietà in senso stretto, perché, anticipando ora ciò che verrà trattato più ampiamente nel seguito, la serie temporale ad alta frequenza presa in esame manifesta proprietà periodiche legate alla scala temporale del giorno. Questo fatto non preclude l'analisi dei dati, ma suggerisce piuttosto una estensione del concetto di stazionarietà.

1.3.2 Ergodicità

L'ergodicità assicura che le medie temporali empiriche di un processo stocastico convergono alle quantità statistiche che si suppone rappresentino. Se si ipotizza che valga l'ergodicità, e che il processo sia stazionario, vale la:

$$\langle f(r_\tau) \rangle = \frac{1}{N} \sum_{t=1}^N f(r_\tau(t))
\tag{1.13}$$

Dove con il simbolo $\langle \ \rangle$ si indica una media statistica nella variabile aleatoria $f(r_\tau(t))$, ed N indica il numero di unità di tempo di cui è composta la serie, per cui al secondo membro compare una media temporale lungo la intera realizzazione del processo.

Al pari della stazionarietà, è difficile dimostrare sia l'assenza che la presenza dell'ergodicità. Tuttavia è comune che sistemi che manifestano proprietà di dipendenza a lungo raggio non godano dell'ergodicità.

1.3.3 Distribuzione dei ritorni: Fat tails

Ipotizzando per i ritorni una distribuzione come in 1.2, ci si chiede se questa ipotesi sia comprovata dalle distribuzioni empiriche. Un fatto evidente andando ad analizzare la distribuzione empirica dei ritorni è che questa si discosta sostanzialmente da una distribuzione gaussiana. Il motivo per cui se ne discosta è che eventi estremi risultano assai più frequenti nella distribuzione empirica dei ritorni rispetto a quanto non siano in una normale che abbia come parametri, media e varianza, quelli estratti dall'insieme di dati da cui la distribuzione empirica è stata generata. Questo fenomeno, detto anche “*f. delle code larghe*”, (*fat tails*) rende necessaria l'adozione di una *pdf* che modella il processo sostanzialmente diversa da una gaussiana. Quanto appena detto risulta evidente andando a considerare il grafico seguente.

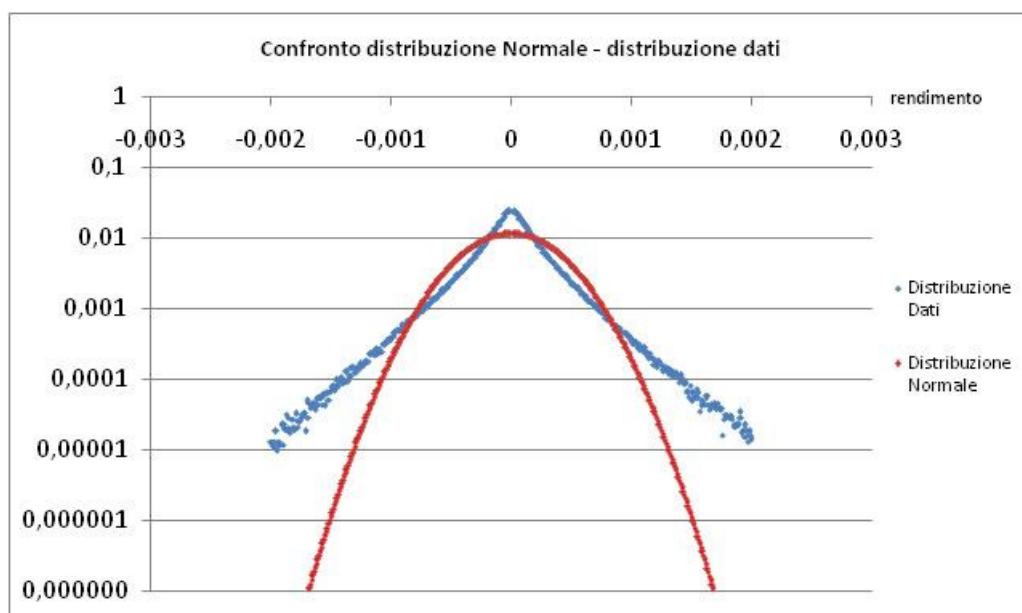


Figura 1.1 S&P500 : Confronto distribuzione empirica dei ritorni a frequenza $\tau=1$ minuto con distribuzione gaussiana con media e varianza derivate dalla distribuzione empirica

1.4 APPROCCIO DI ENSEMBLE

Come anticipato, un problema che si è dovuto affrontare preliminarmente è stato quello di analizzare i dati in maniera da chiarire come trattarli per calcolare le medie necessarie ad ottenere i valori di aspettazione delle quantità prese in esame.

In finanza, come è noto, si ha a disposizione una unica realizzazione di una lunga serie temporale. Per applicare concetti statistici, si pensa alla serie come alla realizzazione di un ipotetico processo aleatorio, e sulla serie si vanno ad effettuare quindi le medie temporali. Nel fare l'assunzione che la media temporale converga a proprietà medie del processo, bisogna assumere che vi sia ergodicità e stazionarietà. Quindi per prima

cosa si è proceduto come in [5] per verificare l'effettiva stazionarietà del processo. È stato analizzato il momento secondo dei ritorni, che è legato alla volatilità, seguendo le idee di *Bassler et al.* [6] e *Baldovin et al.* [7]. Il risultato è stato trovare una violazione palese della stazionarietà, come si può vedere in figura 1.2. È presente, tuttavia, un robusto comportamento periodico su intervalli temporali ben definiti e tutti della stessa lunghezza, pari ad un giorno.

Va notato che il tipico pattern ad U del momento secondo pur denunciando la non stazionarietà, dimostra che c'è una certa dose di periodicità. La periodicità rafforza l'idea che possa valere una stazionarietà oltre le scale di tempo superiori al giorno, ma contraddice la stazionarietà a scale di tempo inferiori al giorno.

Per risolvere questo problema, è possibile una estensione del concetto di stazionarietà che la renda valida solo in certe scale di tempo [7].

Proprio in questo senso è indirizzato il lavoro di *Bassler et al.* [6], il quale analizzando le grandezze di cui sopra (il secondo momento dei ritorni e la autocorrelazione di volatilità), ha dimostrato una caratteristica intrinseca di periodicità nella finestra temporale dell'intraday. Per intraday si intende l'arco di tempo che va dall'orario di apertura della giornata borsistica all'orario di chiusura della stessa, dalle 8:40 alle 15:00.

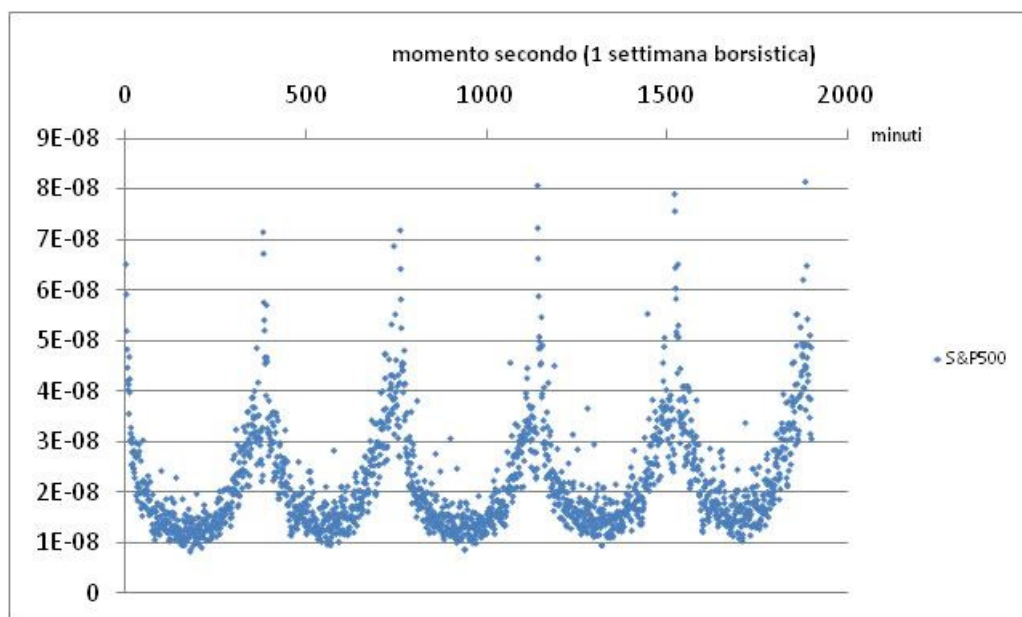


Figura 1.2 S&P500 momento secondo dei ritorni di intervallo $\tau = 1$ minuto relativi ad una settimana borsistica (5gg di contrattazioni)

La periodicità su scale interday rafforza infatti l'idea che possa valere una stazionarietà a scale superiori al giorno, mentre la non stazionarietà nell'intraday indica la necessità di ipotizzare un unico processo stocastico che si sviluppi per l'intera durata della finestra intraday. Alla luce di quanto detto, e vista la periodicità in figura 1.2, è risultato naturale considerare che il processo stocastico responsabile della

generazione dei ritorni abbia durata pari ad una giornata borsistica. In questo modo è possibile applicare alla serie di dati un approccio di ensemble, in cui il singolo processo dura da apertura a chiusura dei mercati. È grazie a questo carattere di forte periodicità che è possibile pensare ai mercati finanziari nell'ottica dell'ensemble.

Per guardare a scale di tempo nell'intraday, dove si realizza un processo non stazionario, si utilizza quindi un'ottica di ensemble. Invece per guardare a scale di tempo superiori all'intraday si utilizzano medie temporali e si assume ergodicità e stazionarietà.

Basandosi su queste evidenze, sono stati organizzati i dati in maniera da suddividere l'unica realizzazione iniziale del processo in sottoserie di durata pari ad una giornata borsistica. È stato poi utilizzato un approccio di ensemble per ottenere i valori di aspettazione delle quantità cercate come medie di ensemble.

La media di ensemble della quantità $f(r_\tau(t))$ è

$$\langle f(r_\tau(t)) \rangle_e = \frac{1}{M} \sum_{l=1}^M f(r_\tau(t))^l \quad 1.14$$

Dove $\langle \rangle_e$ indica la media di ensemble, t corre lungo la finestra temporale dell'intraday e M rappresenta il numero di giorni di cui è composta la serie di dati, cioè il numero di elementi del nostro ensemble. Da notare che rispetto alla 1.13 non scompare la dipendenza da t , in conseguenza della non stazionarietà riscontrata nell'intraday, ma a differenza della 1.13 ora t corre lungo la finestra intraday.

Per ottenere le evidenze empiriche in figura 1.2, si rende necessaria una analisi del momento secondo dei ritorni:

1.4.1 Analisi del momento secondo dei ritorni logaritmici

Il momento secondo è un ottimo indicatore del comportamento periodico. risulta massimo ad inizio e fine giornata borsistica, come conseguenza del maggior numero di trade effettuati in questi due momenti, e presenta un minimo a metà giornata.

$$m_2^e(t, \tau) = \frac{1}{M} \sum_{l=1}^M (r_\tau^l(t))^2 \quad 1.15$$

Dove r rappresenta il ritorno logaritmico definito dalla 1.10, τ l'intervallo di tempo (1 minuto) su cui è calcolato il ritorno, t rappresenta l'istante di tempo all'interno della finestra temporale dell'intraday, M il numero di giorni da cui è composto l'ensemble.

Nel caso in esame i τ sono alla frequenza di 1 minuto e t varia in maniera discreta da 1 a 380 minuti.

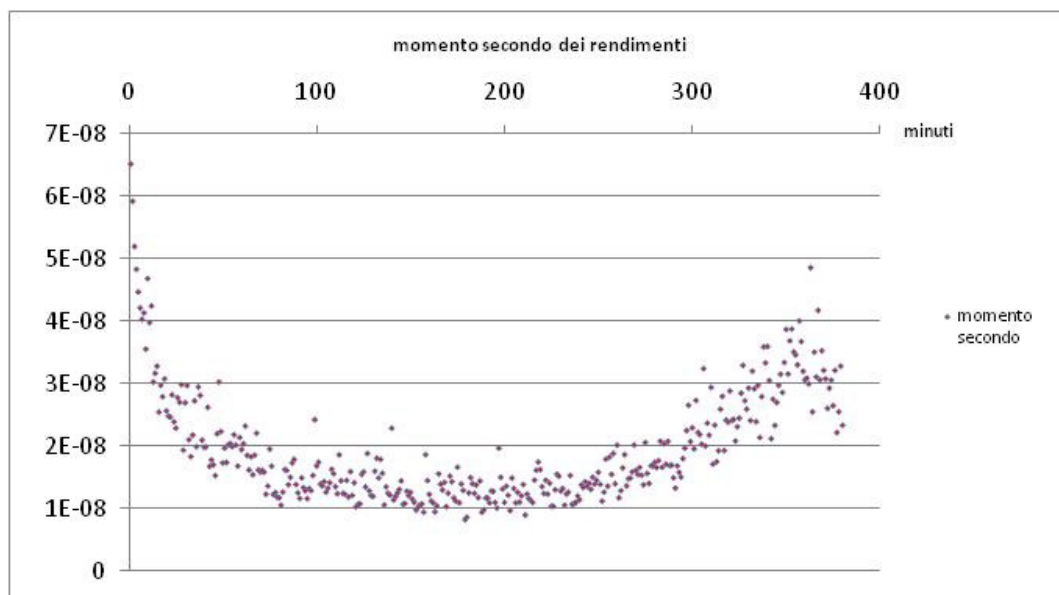


Figura 1.3 S&P500 : Momento secondo dei ritorni di intervallo $\tau = 1$ minuto relativo alla finestra temporale intraday

Si nota la tipica forma ad U, che decade approssimativamente da inizio giornata per le successive tre ore ed aumenta nelle successive tre e mezza.

1.5 ANALISI DELLE CORRELAZIONI LINEARI

Si è proceduto, seguendo l'analisi dei cosiddetti *Stylized facts* [4]; [9], nell'analisi dei dati volta a valutare l'esistenza o meno delle correlazioni lineari e se vi sia una evoluzione temporale nelle suddette correlazioni. Analizzare le correlazioni dei ritorni logaritmici fornisce importanti informazioni sul mercato. L'esistenza di correlazioni lineari non nulle rende possibile la riuscita di strategie di trading che, se invece applicate in un mercato assolutamente efficiente, con correlazioni lineari nulle e senza quindi possibilità di arbitraggio, non produrrebbero risultati soddisfacenti. Ricordando che si sta lavorando in un contesto di ensemble, su scala intraday, si definisce la correlazione lineare dei ritorni secondo la:

$$c_{lin}^e(1, t) = \frac{\frac{1}{M} \sum_{l=1}^M r_1^l(1) r_1^l(t)}{\sqrt{m_2(1,1) m_2(t, 1)}} \quad 1.16$$

Dove M rappresenta il numero di giorni della serie, $m_2^e(t, \tau)$ il momento secondo definito dalla 1.14 con τ l'intervallo di tempo su cui è calcolato il ritorno posto ora pari ad un minuto, e t l'istante di tempo all'interno della finestra temporale dell'intraday.

La correlazione del valore assoluto dei ritorni è definita secondo la

$$c_{ass}^e(1, t) = \frac{\frac{1}{M} \sum_{l=1}^M |r_1^l(1)| |r_1^l(t)|}{\sqrt{m_2(1,1) m_2(t, 1)}} \quad 1.17$$

Il grafico che segue confronta queste due quantità:

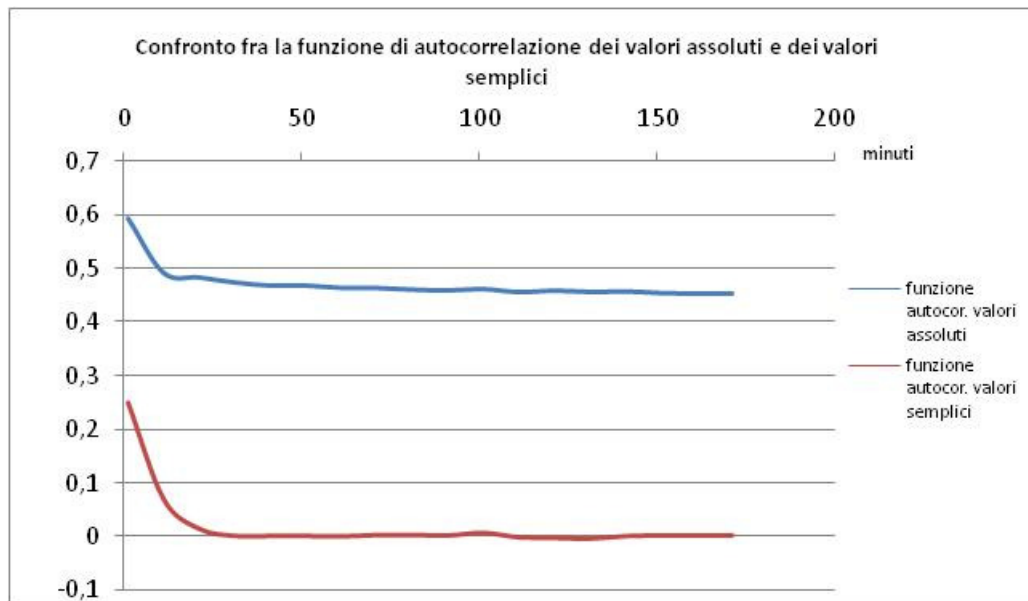


Figura 1.4 Confronto fra la funzione di autocorrelazione e la funzione di autocorrelazione a valori assoluti dei ritorni

Risulta chiaro che la correlazione semplice definita dalla 1.16 è molto superiore alla correlazione lineare semplice.

Questo fatto può essere spiegato associando alla correlazione lineare semplice dei rendimenti un effetto di memoria di trend, ovvero in presenza di correlazione lineare positiva se ad un dato istante di tempo il mercato ha un ritorno positivo, anche negli istanti successivi tenderà ad avere ritorni positivi. E' presente quindi un fenomeno di persistenza. E' possibile invece associare alla correlazione lineare dei valori assoluti dei rendimenti un effetto di memoria di volatilità, ovvero in presenza di correlazione lineare in valore assoluto dei rendimenti positiva, se ad un dato istante di tempo il mercato ha un forte aumento di volatilità, anche negli istanti successivi tenderà ad avere una volatilità elevata. E' presente quindi anche in questo caso un fenomeno di persistenza.

Risulta quindi chiara la presenza di rilevanti fenomeni di persistenza della volatilità e in minore quantità, di persistenza di memoria di trend. Questo argomento verrà ripreso nel capitolo 2 relativamente alla formulazione di un modello di mercato consistente con lo scaling.

Si vuole ora vedere se vi sia una evoluzione di queste quantità nell'arco di tempo che va dall'anno 1985 all'anno 2013.

1.5.1 Evoluzione dell'autocorrelazione

Risulta utile definire come in [9] la funzione di autocorrelazione lineare calcolata su medie di ensemble di ritorni di M giorni, come:

$$\begin{aligned} \langle r_\tau(t)r_\tau(t+k\tau) \rangle_e & \\ &= \frac{1}{M} \sum_{l=1}^M \frac{1}{\text{Var}(r_\tau^l)} \frac{1}{n-k+1} \sum_{i=0}^{n-k} r_\tau^l(t+i\tau)r_\tau^l(t+(i+k)\tau) \end{aligned} \quad 1.18$$

Dove M rappresenta il numero di giorni pari al numero di elementi dell'ensemble, τ l'intervallo di tempo su cui è calcolato il ritorno, t l'istante di tempo all'interno della finestra temporale dell'intraday, n il numero di diversi ritorni nell'intraday.

Il grafico seguente rappresenta la funzione di autocorrelazione calcolata con ritorni su intervalli τ di 1 minuto.

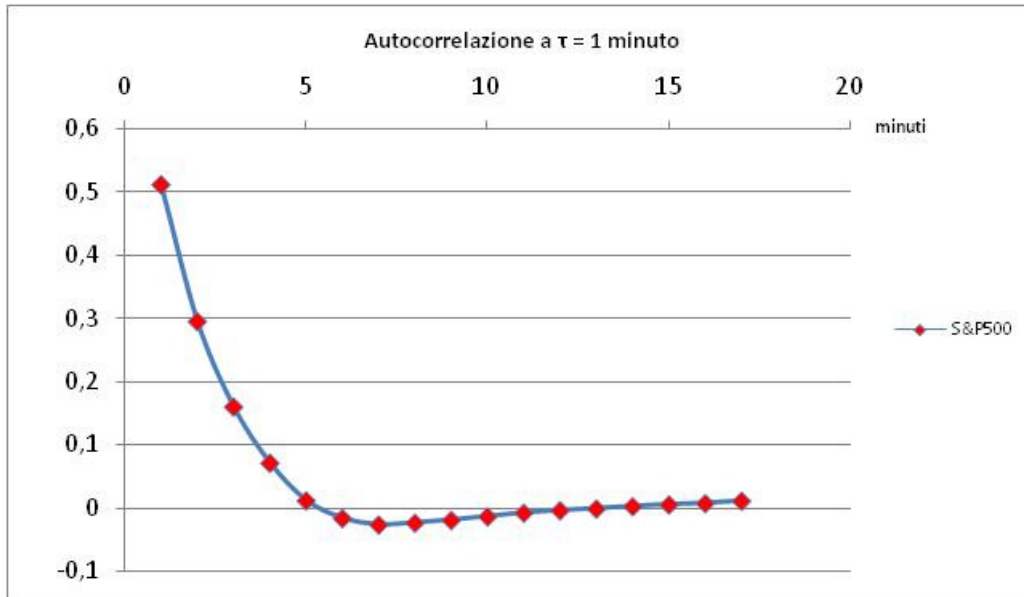


Figura 1.5 Funzione di autocorrelazione calcolata per ritorni di intervallo 1 minuto

Come si può vedere si ha autocorrelazione positiva massima fra il ritorno al tempo t e quello al tempo $t+1$ minuto, per poi passare a valori negativi dal tempo $T+7$ minuti e nulli dal tempo 12 minuti in avanti.

Dopo aver esaminato la funzione di autocorrelazione calcolata sulla base dell'insieme completo dei dati disponibili, si è proceduto ad analizzare la stessa funzione di autocorrelazione per singoli intervalli di tempo, compresi fra il 1985 ed il 2013, cercando di valutare se vi fosse una qualche evoluzione. I grafici 1.6 e 1.7 rappresentano l'evoluzione della funzione di autocorrelazione nei tre periodi analizzati: [1985-1994]; [1995-2004]; [2005-2013].

Come è possibile notare, vi è una sempre minore durata del tempo medio di autocorrelazione, dunque la conclusione che se ne trae è che il mercato si sia drasticamente evoluto fra il 1985 ed il 2013. Nel periodo compreso fra il 1985 ed il 1994 la durata di persistenza media di un trend è stato di 3 minuti, per poi scendere nella decade successiva a 1,6 minuti e ad un minuto nel periodo 2005-2013.

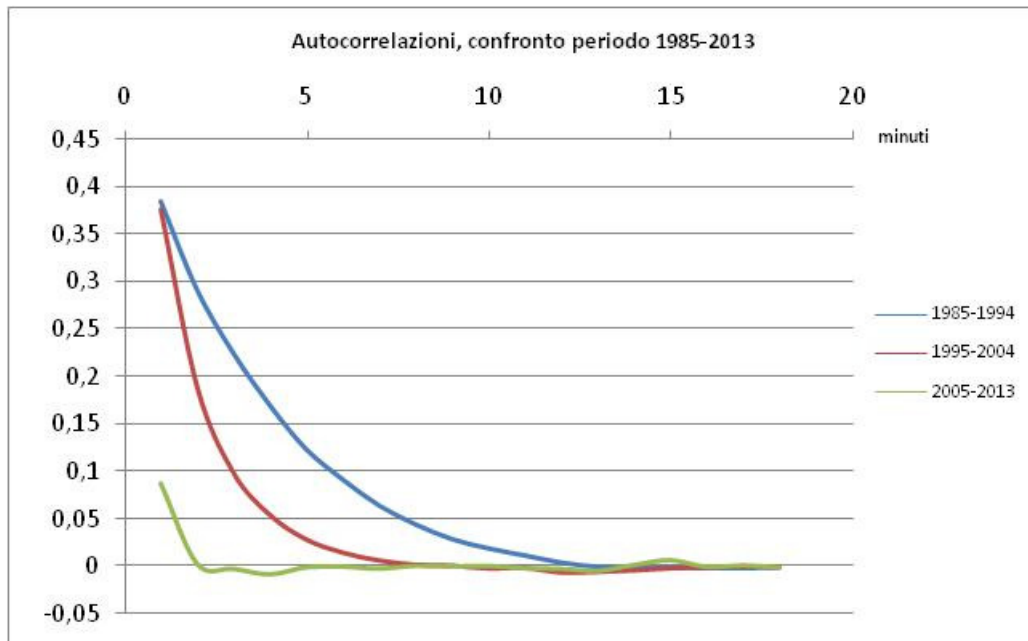


Figura 1.6 Confronto fra la funzione di autocorrelazione calcolata con serie di dati di durata pari a 10 anni dal 1985 al 2013

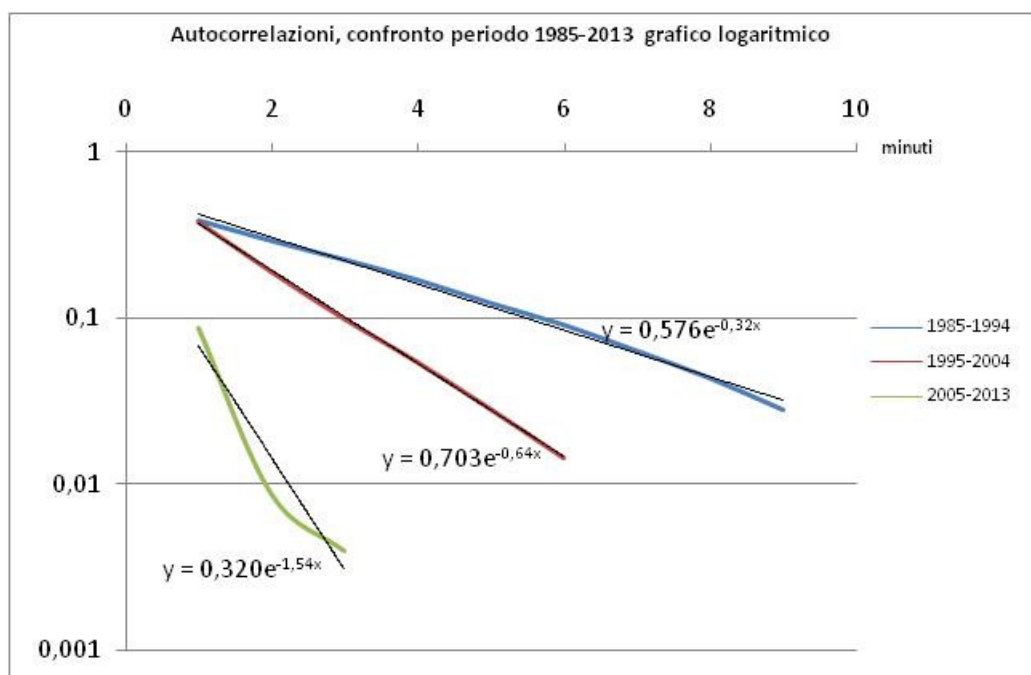


Figura 1.7 Grafico log-log del precedente (le code sono tagliate in prossimità dello zero).

Dato che l'autocorrelazione è legata alla possibilità di arbitraggio, ovvero la possibilità di ottenere profitto senza assumersi rischio, risulta chiara anche una evoluzione del mercato verso una situazione che offre sempre minore possibilità di arbitraggio.

La possibilità di arbitraggio, come si vedrà nel capitolo 3, è strettamente legata alla buona riuscita delle strategie di trading. Maggiore è l'arbitraggio che il mercato offre, maggiore è la riuscita di una strategia di trading legata allo sfruttamento delle autocorrelazioni. La medesima analisi è stata fatta relativamente allo studio della funzione di autocorrelazione dei valori assoluti, in questo caso quello che si è potuto notare è stato un aumento nelle varie decadi della persistenza della volatilità.

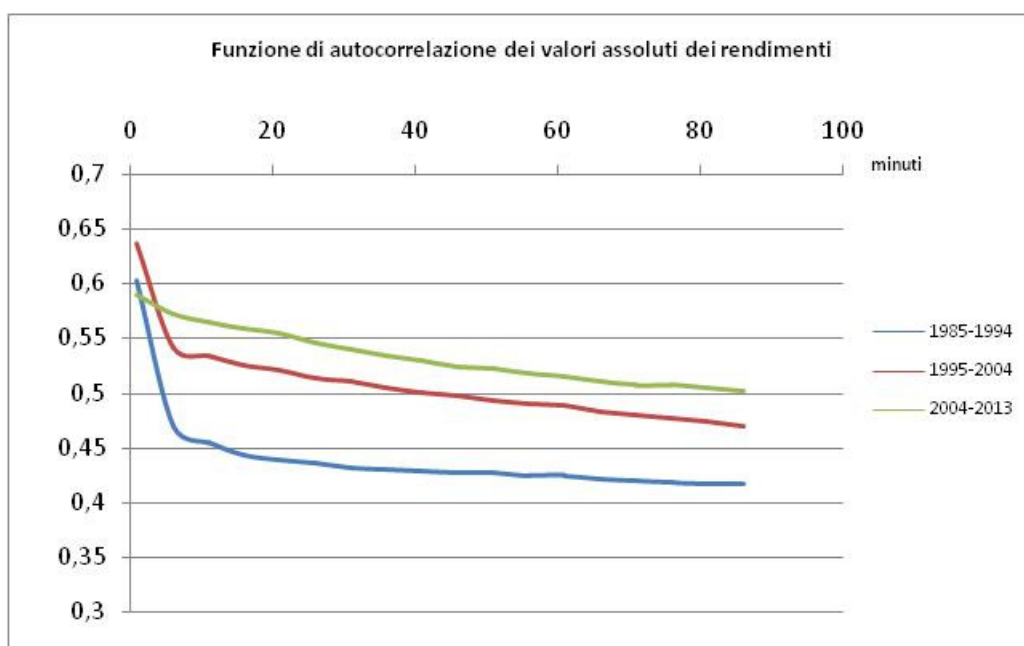


Figura 1.8 Confronto fra la funzione di autocorrelazione calcolata con serie di dati di

La funzione di autocorrelazione dei valori assoluti come si è detto ha valori molto maggiori della funzione di autocorrelazione semplice, e anche in questo caso si può notare una variazione progressiva dell'andamento dagli anni '80, ove presenta un minore effetto di memoria nella volatilità, rispetto agli anni recenti ove questo effetto di memoria è molto più marcato e consistente.

Capitolo 2 LO SCALING

In questo capitolo si studia un modello che sia compatibile con le evidenze empiriche analizzate nel capitolo 1, ad eccezione delle correlazioni, seguendo le idee in [5]. Nella prima parte si fisserà la notazione con una piccola introduzione sui processi stocastici, per poi definire rigorosamente le basi matematiche del modello ed in particolare l'idea di scaling, pilastro fondamentale della trattazione.

2.1 PROCESSI STOCASTICI

Ogniquale volta l'abbondanza di gradi di libertà di un processo non ne permetta l'analisi deterministica, a causa del numero troppo elevato di equazioni del moto da risolvere, si vede necessario trattare il processo dal punto di vista statistico.

Per descrivere un processo casuale, che chiameremo $R_\tau(t)$ a valori reali, dove $R_\tau(t)$ rappresenta il processo che genera un ritorno, si utilizzerà una funzione di densità di probabilità (PDF) $P(r_\tau, t)$ con le seguenti caratteristiche [13]:

-La probabilità che una qualsiasi variabile casuale, abbia valori compresi fra a e b è data dall'espressione

$$P_{R_\tau(t)}(a < r_\tau < b) = \int_a^b P(r_\tau, t) dr \quad 2.1$$

-La probabilità è sempre positiva,

$$P_{R_\tau(t)}(a < r_\tau < b) \geq 0 \quad \forall r_\tau \in R; \quad \forall a, b \quad 2.2$$

-La probabilità è normalizzata ad uno

$$\int_{-\infty}^{+\infty} P(r_\tau, t) dr = 1 \quad \forall t \quad 2.3$$

Si definisce la distribuzione gaussiana di varianza σ^2 e media m attraverso la:

$$N_{R_\tau(t)} = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(r_\tau - m)^2}{\sigma^2}\right) \quad 2.4$$

La Normale è la distribuzione alla quale converge la distribuzione della somma di una serie di variabili casuali indipendenti e identicamente distribuite (i.i.d).

Il Teorema del Limite Centrale [11] afferma infatti che “qualora un fenomeno casuale, sia nel suo complesso conseguenza di un grande numero di piccole, indipendenti cause, esso è descritto da una distribuzione gaussiana.” Il teorema può essere dimostrato a partire dalla stabilità delle distribuzioni e dalla fattorizzazione che vale per le rispettive funzioni caratteristiche in quanto somma di variabili iid. L’importanza di questo teorema deriva dal fatto che rende possibile l’introduzione di un semplice processo di scaling secondo la:

$$p(r_\tau) = \frac{1}{\tau^D} g\left(\frac{r_\tau}{\tau^D}\right) \quad 2.5$$

Dove τ rappresenta il parametro di scala, D l’esponente di Hurst del processo, g la funzione di scaling e r_τ il ritorno relativo all’intervallo di tempo τ .

Si riprende ora il concetto di ritorno aggregato, perché utile nel seguito. Il ritorno aggregato di durata $\tau = n\theta$, o semplicemente il ritorno di durata τ , come definito in 1.12, è uguale alla somma degli $n-1$ ritorni semplici di durata θ compresi nell’intervallo di tempo a cui il ritorno aggregato fa riferimento.

I ritorni aggregati, per costruzione, sono somma di ritorni elementari. Proprio per questo fatto, ovvero essere somma di variabili casuali indipendenti, ci si aspetta che valga il teorema del limite centrale [11]. Se vale il teorema del limite centrale, la funzione di scaling g deve essere Gaussiana e avere un esponente di Hurst di valore pari a $1/2$.

Tuttavia il teorema del limite centrale ha come ipotesi che le variabili su cui si somma siano fra loro indipendenti. Invece, a causa delle correlazioni non nulle già discusse, che coinvolgono i ritorni analizzati, esse non possono essere considerate come variabili indipendenti. Bisognerà quindi, come in [14], estendere la formulazione anche al caso di variabili non statisticamente indipendenti.

Per fissare e chiarire la notazione, da qui in avanti, i ritorni aggregati indicati fino ad ora il più genericamente possibile con $r_{n\theta}(t) = r_\tau(t)$, dove con τ si indica l’intervallo di tempo su cui è calcolato il ritorno, multiplo n della durata elementare θ , e con t il tempo nell’ intraday al quale è calcolato il ritorno, potranno essere scritti per brevità di notazione come $r_\tau(\tau) = r_\tau$, qualora t e τ coincidano. Ciò avviene se i ritorni sono calcolati a partire dal tempo di apertura del mercato, condizione necessaria per lo scaling inhomogeneo, ovvero lo scaling in cui il parametro di scala λ dipenda da τ non come semplice potenza. Con questa notazione, r_τ indica il ritorno aggregato di n ritorni elementari di durata θ , *calcolato a partire dall’orario di apertura dei mercati*. Si pone l’attenzione sul fatto che in questo modo, per ragioni relative appunto allo scaling inhomogeneo, vale il fatto che il ritorno aggregato $r_\tau(t)$ relativo all’intervallo τ e calcolato al tempo t coincide con $r_\tau(\tau)$, dato che $t = \tau$

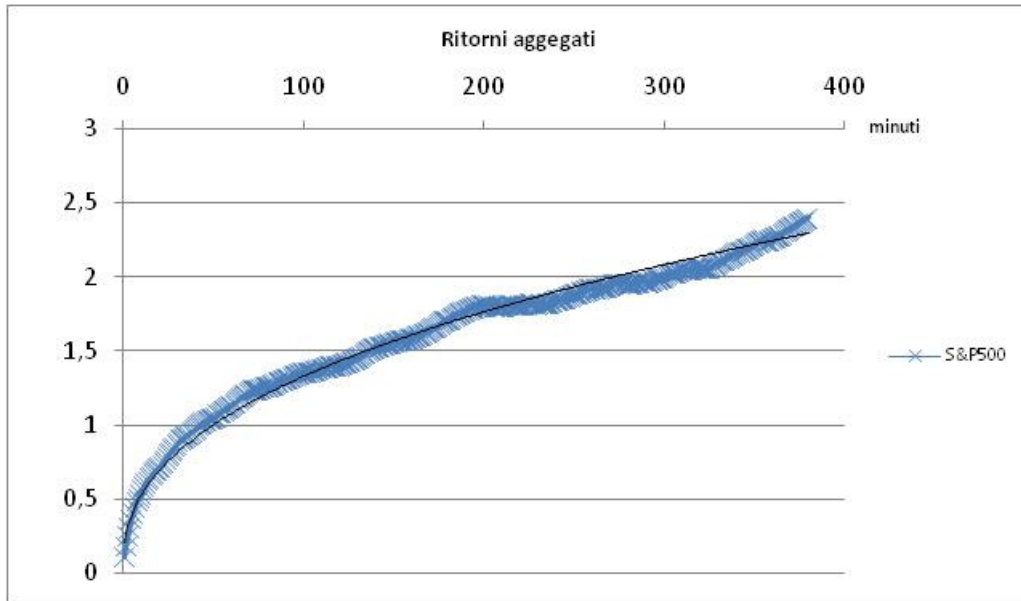


Figura 2.1 S&P500 : Ritorni aggregati relativi alla finestra intraday

2.2 EVIDENZE DI SCALING

Si considerano innanzitutto i fatti che sono alla base dell'applicabilità della teoria dello scaling [15].

Analizzando il momento secondo, è possibile fittare il grafico delle prime tre ore della giornata, ovvero fino a $\tau = t_m = 200$ minuti, ottenendo un andamento che segue la

$$m_2^e(\tau, \tau) = \frac{1}{M} \sum_{l=1}^M (r_\tau^l(\tau))^2 \quad 2.6$$

con τ che è compreso nella finestra intraday e M uguale al numero di diverse giornate borsistiche di cui è composto l'ensemble. Il fit dei dati fino a $t_m = 200$ minuti restituisce un valore di D pari a 0.360 e $m_2^e(1,1) = 5.2 * 10^{-7}$.

Richiamando la 1.15 si generalizza la definizione di momento secondo con la definizione di momento q -esimo relativo ad un ritorno aggregato di intervallo $n\theta = \tau$ al tempo t con la:

$$\begin{aligned} m_q^e(t, n\theta) &= \frac{1}{M} \sum_{l=1}^M |r_\theta^l(t) + r_\theta^l(t - \theta) + \dots + r_\theta^l(t - (n-1)\theta)|^q \\ &= \frac{1}{M} \sum_{l=1}^M |r_\tau^l(t)|^q = m_q^e(t, \tau), \end{aligned} \quad 2.7$$

dove l'indice l rappresenta il giorno, M rappresenta il numero totale di giorni nella serie dati presa in esame, t corre nella finestra dell'intraday, ed n è definito come in (1.12).

Si vuole ora verificare che l'esponente di Hurst D non dipenda dalla scala temporale τ sulla quale si vanno a calcolare i ritorni aggregati. Se ciò avviene, si può ipotizzare che le PDF empiriche di ritorni aggregati di diverse durate temporali non rappresentino processi incompatibili e diversi l'uno dall'altro, ma possano essere fatti convergere in un solo processo.

Si procede quindi variando l'indice q nel range $[0;2]$, si calcola qD dalle distribuzioni empiriche e si crea un grafico rappresentante i punti di coordinate $(q;qD)$

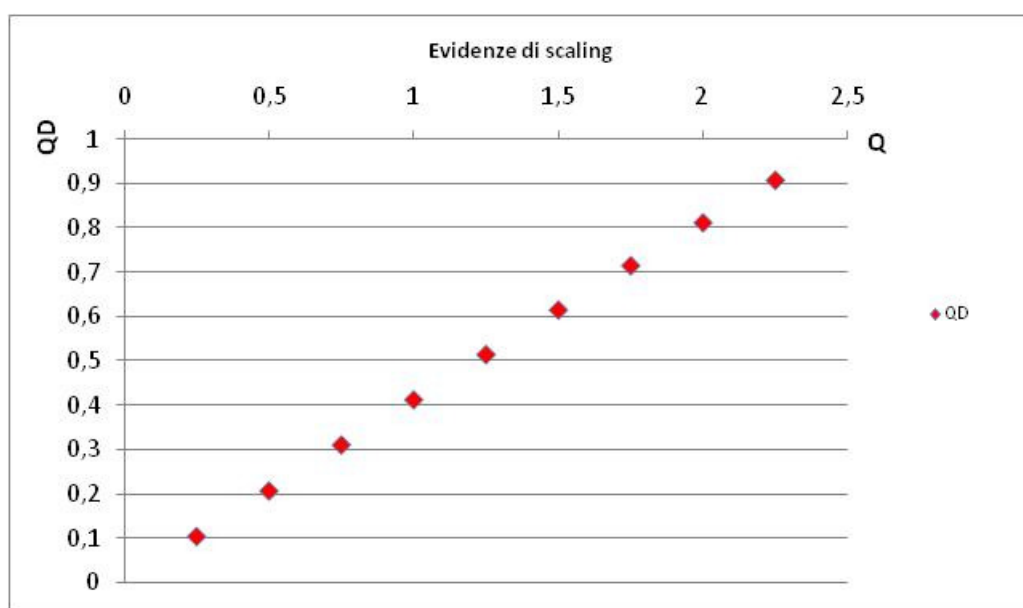


Figura 2.2 momento secondo empirico

Se vi è scaling deve valere la:

$$m_q^e(\tau, \tau) \sim \tau^{qD} \tag{2.8}$$

Con D esponente di Hurst non dipendente da q .

I valori qD sono stati ottenuti considerando la distribuzione empirica dei momenti q -esimi per i ritorni e fittandola con una funzione a potenza ottenendo per ogni valore di q il corrispondente valore empirico di qD derivante dall'esponente del fit. Il comportamento lineare che si vede in figura 2.2 suggerisce in maniera evidente che l'esponente di Hurst D effettivamente non dipende da q e dunque che le PDF di ritorni aggregati di diverse durate temporali non rappresentano processi incompatibili e diversi l'uno dall'altro, ma possono essere fatti collapsare in un solo processo, sotto opportune ipotesi di scaling.

2.3 SCALING

Si presentano ora i risultati della formalizzazione del modello adottato, ideato da Baldovin *et al.*, rimandando a [5];[14];[15] per tutti i passaggi della derivazione teorica completa. Il modello in questione è basato sulla teoria dello scaling e riproduce le evidenze empiriche discusse in precedenza.

L'ipotesi robusta su cui si basa il modello è la validità della simmetria di scaling. Questa simmetria fa convergere le distribuzioni dei ritorni aggregati riferiti a differenti tempi scala ad un'unica distribuzione, non gaussiana, opportunamente riscalata di una quantità λ_τ , che può avere un comportamento semplice a potenza secondo la (2.5) [*scaling omogeneo nel tempo*] o non avere un comportamento semplice a potenza [*scaling inomogeneo nel tempo*] e seguire, come vedremo, la 2.23.

Inoltre, come già accennato, la volatilità all'interno dell'intraday presenta un tipico andamento ad U, ovvero decresce fino a raggiungere un t_m pari a 200 minuti, per poi ricrescere fino a chiusura mercato. Questo indica come già osservato una non stazionarietà durante l'intraday.

Le ipotesi di partenza sono, riassumendo:

1. *La non stazionarietà durante l'intraday*
2. *L'assenza di correlazioni lineari*
3. *Il comportamento di Scaling*

Vi è una largo insieme di *funzioni di scaling* $w(r)$ che soddisfino queste ipotesi, e seguendo il teorema di Schoenberg [16] si trova che la più generale forma che sia valida secondo le ipotesi sia

$$w(r) = \int_0^\infty \rho(\sigma) \frac{e^{-\frac{r^2}{2\sigma^2}}}{\sqrt{2\pi\sigma^2}} d\sigma, \quad 2.9$$

con ρ una misura positiva e normalizzata nell'intervallo $]0, +\infty[$

La PDF di n ritorni aggregati si ricava dal prodotto di distribuzioni gaussiane di differenti ampiezze a_t

$$p_{R_1(1), \dots, R_1(\tau)}(r_1(1), \dots, r_1(\tau)) = \int_0^\infty d\sigma \rho(\sigma) \prod_{t=1}^{\tau} \frac{\exp\left(-\frac{r_1^2(t)}{2\sigma^2 a_t^2}\right)}{\sqrt{2\pi\sigma^2 a_t^2}}, \quad 2.10$$

con

$$a_t = [t^{2D} - (t-1)^{2D}]^{\frac{1}{2}}. \quad 2.11$$

La funzione di Scaling, come proposto in [5], è data da:

$$g(r) = \int_0^\infty \rho(\sigma) \frac{e^{-\frac{r^2}{2\sigma^2}}}{\sqrt{2\pi\sigma^2}} d\sigma. \quad 2.12$$

Per una corretta parametrizzazione della funzione di scaling g , si è deciso di utilizzare per σ^2 una *inverse gamma density* in maniera tale che con questa scelta $\rho(\sigma)$ assuma la forma:

$$\rho(\sigma) = \frac{2^{1-\frac{\alpha}{2}} \beta^\alpha}{\Gamma\left(\frac{\alpha}{2}\right) \sigma^{\alpha+1}} e^{-\frac{\beta^2}{2\sigma^2}}, \quad 2.13$$

con $\beta > 0$ parametro di scala e $\alpha > 0$ parametro di forma, e Γ funzione gamma di Eulero. Ne segue che la funzione di scaling abbia così la forma di una distribuzione T di Student.

$$g(r) = \frac{\Gamma\left(\frac{\alpha+1}{2}\right)}{\sqrt{\pi}\Gamma\left(\frac{\alpha}{2}\right)} \frac{1}{\beta} \left(1 + \frac{r^2}{\beta^2}\right)^{-\frac{\alpha+1}{2}}. \quad 2.14$$

Questa distribuzione presenta per grandi r un decadimento a potenza con esponente $\alpha + 1$; mentre la scala della sua ampiezza è rappresentata da β .

L'esistenza di una definita simmetria di scaling con D fissato implica un comportamento power-law per il momento dei ritorni aggregati secondo la:

$$\overline{|r_1(1) + \dots + r_1(\tau)|^q} = \overline{|r_1(1)|^q} \tau^{qD}. \quad 2.15$$

Con la notazione utilizzata si intende una media di ensemble su tutti gli M giorni di cui è composta la serie di dati, con τ che come sempre corre nella finestra intraday.

Con la scelta per $\rho(\sigma)$ data da 2.13 l'equazione 2.15 prende la forma:

$$\overline{|r_1(1) + \dots + r_1(\tau)|^q} = \frac{\Gamma\left(\frac{q+1}{2}\right) \Gamma\left(\frac{\alpha-q}{2}\right) \beta^q}{\sqrt{\pi}\Gamma\left(\frac{\alpha}{2}\right)} \tau^{qD}. \quad 2.16$$

Risulta importante una corretta calibrazione dei parametri α, β, D , in quanto questi determinano rispettivamente la $\rho(\sigma)$ e le a_i che compaiono nell'equazione della PDF

2.10. Per fare ciò si procede utilizzando questo protocollo: per prima cosa viene calibrato D, e di seguito si utilizza D per calibrare α e β , tramite un collasso dei ritorni su un'unica distribuzione. Utilizzando l'analisi dei momenti q-esimi dei ritorni aggregati, si mettono in relazione i momenti empirici dati dalla

$$\langle |r_1(1) + \dots + r_1(\tau)|^q \rangle = \frac{1}{M} \sum_{l=1}^M |r_1^l(1) + \dots + r_1^l(\tau)|^q \quad 2.17$$

Dove con la notazione $\langle \ \rangle$ si vuole denotare una media fatta su quantità empiriche, negli M giorni che compongono la serie dei dati.

Da qui, si esegue una regressione lineare fra il logaritmo dei momenti empirici e q stesso. Si costruisce un grafico in cui in ordinata si ha $qD(q)$ ed in ascissa q. Si nota un comportamento lineare fra queste due variabili e dunque una sostanziale indipendenza di $D(q)$ da q.

Il risultato del fit fornisce per D un valore di $D=0,403$ come rappresentato nella figura

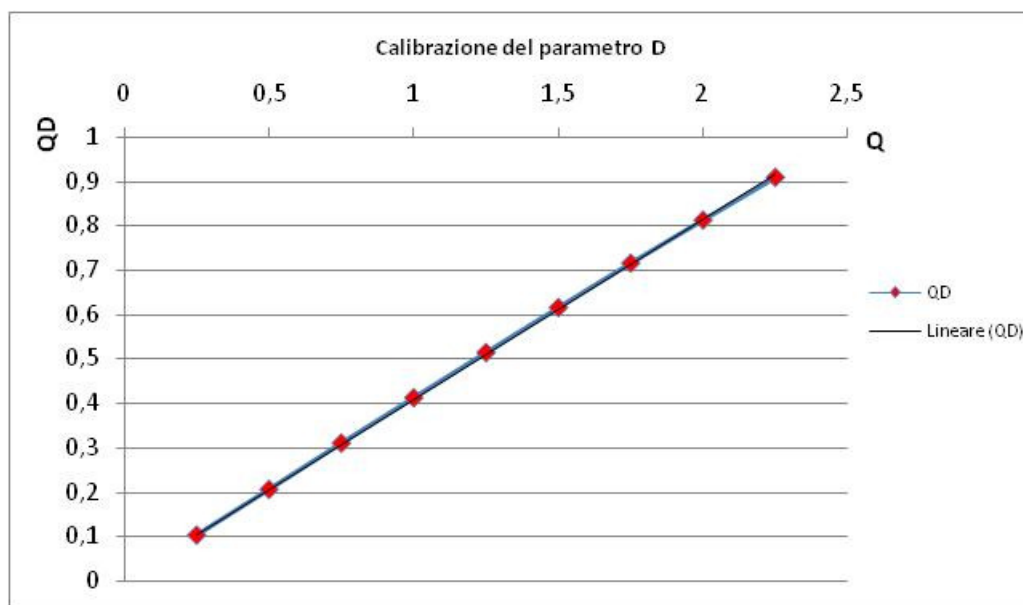


Figura 2.3 Fit del momento secondo

L'analisi è stata effettuata per q compreso fra 0,25 e 2,25, perché per valori superiori è evidente un comportamento multiscaling. L'analisi procede utilizzando D così ottenuto e l'equazione 2.17 con $q = 1$ per ricavare un valore per $\langle |R_1(1)| \rangle$. Si può così determinare tramite la equazione 2.16 β come funzione di α . Si fa poi collassare la PDF empirica dei ritorni marginali per vari indici dell'intervallo τ , e si esegue un Fit con l'equazione 2.14 che fornisce la forma algebrica della funzione di scaling in

funzione di (r, α, β) , con β dipendente da α come definito sopra. I risultati del fit forniscono $\alpha = 3.39$; $\beta = 0.34 * 10^{-3}$; $D = 0.403$

In figura 2.4 e 2.5 si vedono le distribuzioni prima del collasso e dopo il collasso, con la funzione di fit.

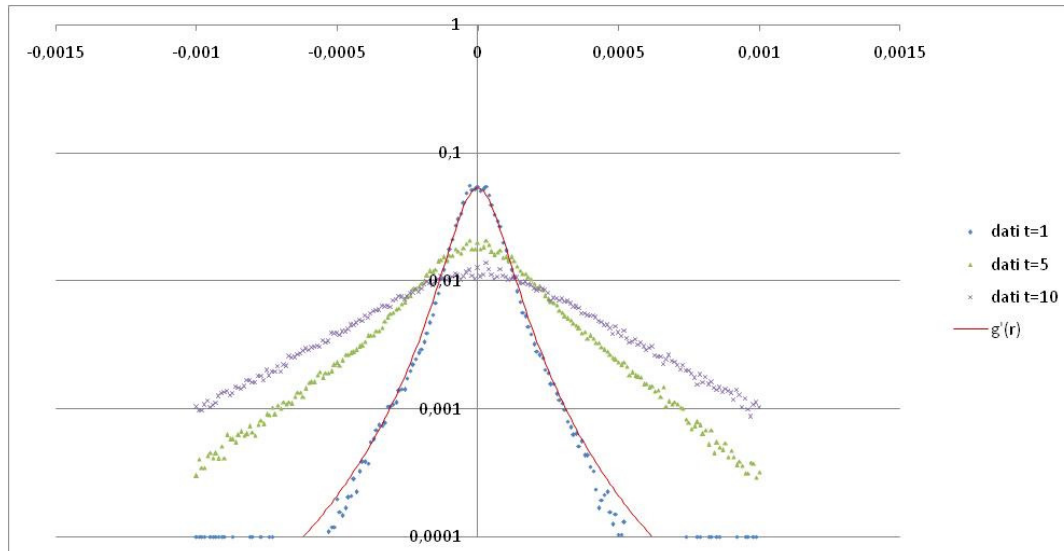


Figura 2.4 Calibrazione: Fit della distribuzione empirica con la distribuzione g per calcolare i parametri α, β . t espresso in minuti.

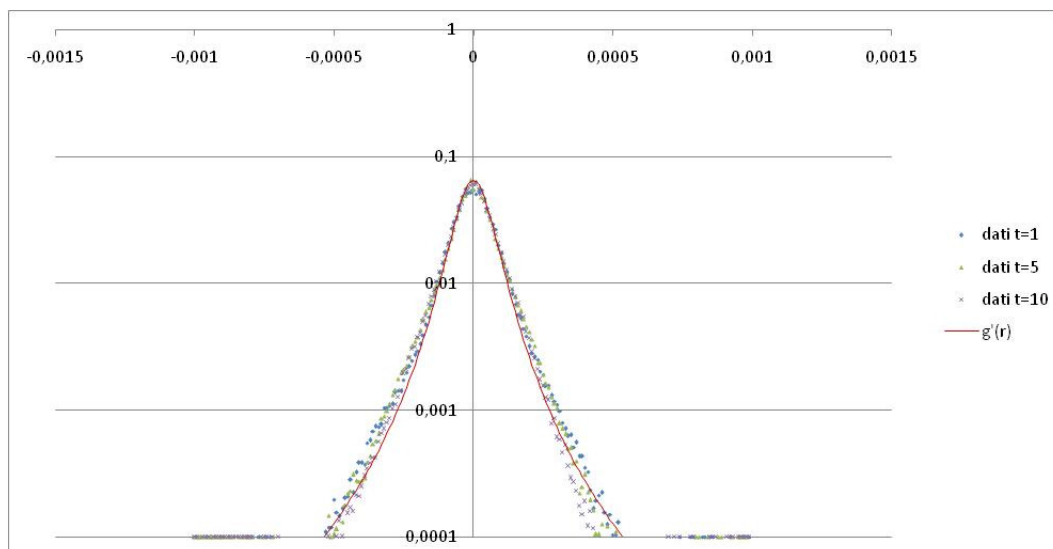


Figura 2.5 Collasso delle distribuzioni empiriche della figura precedente. t espresso in minuti.

2.3.1 Estensione del modello alla sessione pomeridiana

Il modello così formulato può essere esteso anche alla sessione pomeridiana, con una opportuna generalizzazione.

Innanzitutto si sostituisce $\sigma \rightarrow \frac{\sigma}{\beta}$.

$$p_{R_1(1), \dots, R_1(\tau)}(r_1(1), \dots, r_1(\tau)) = \int_0^\infty d\sigma \rho'(\sigma) \prod_{t=1}^{\tau} \frac{\exp\left(-\frac{r_1^2(t)}{2\sigma^2 a_t^2 \beta^2}\right)}{\sqrt{2\pi\sigma^2 a_t^2 \beta^2}} \quad 2.18$$

$$\rho'(\sigma) = \frac{2^{1-\frac{\alpha}{2}}}{\Gamma\left(\frac{\alpha}{2}\right)} \frac{1}{\sigma^{\alpha+1}} e^{-\frac{1}{2\sigma^2}} \quad 2.19$$

In questo modo risulta evidente il ruolo di β come parametro di scala. Durante la mattina la volatilità decresce, fino a toccare un minimo dopo approssimativamente 200 minuti (13:20 ora di New York, $=t_m$), per poi tornare ad aumentare fino a fine giornata borsistica, dove manifesta un massimo relativo. Per rendere ragione di questo comportamento si assumerà valida l'equazione per ρ' , e si introdurrà una dipendenza temporale per l'esponente di Hurst D e per il parametro di scala β .

$$D \rightarrow D_t \quad \beta \rightarrow \beta_t$$

$$a_t = [t^{2D_t} - (t-1)^{2D_t}]^{\frac{1}{2}} \quad 2.20$$

$$p_{R_1(1), \dots, R_1(\tau)}(r_1(1), \dots, r_1(\tau)) = \int_0^\infty d\sigma \rho'(\sigma) \prod_{t=1}^{\tau} \frac{\exp\left(-\frac{r_1^2(t)}{2\sigma^2 a_t^2 \beta_t^2}\right)}{\sqrt{2\pi\sigma^2 a_t^2 \beta_t^2}} \quad 2.21$$

Per mantenere invariato il comportamento mattutino e modificare come voluto il comportamento pomeridiano, si utilizzeranno le relazioni:

$$D_t = \begin{cases} D_m & \text{se } 1 \leq t \leq t_m \\ D_p & \text{se } t_m \leq t \leq 360 \end{cases}$$

$$\beta_t = \begin{cases} \beta_m & \text{se } 1 \leq t \leq t_m \\ \beta_p & \text{se } t_m \leq t \leq 360 \end{cases}$$

Dove si introducono le nuove notazioni, andando a suddividere D e β in un $D_m = D$ relativo alla mattina (calcolato in precedenza), D_p relativo al pomeriggio; $\beta_m = \beta$ relativo alla mattina e β_p relativo al pomeriggio. La simmetria di scaling assume la forma:

$$p_{R_{\tau(\tau)}}(r_{\tau}, \tau) = \frac{1}{\lambda(\tau, t_m)} g' \left(\frac{r_{\tau}}{\lambda(\tau, t_m)} \right), \quad 2.22$$

con

$$\lambda(\tau, t_m) = \left(\sum_{t=1}^{\tau} a_t^2 \beta_t^2 \right)^{1/2} = \begin{cases} \beta_m \tau^{D_m} & 1 \leq \tau \leq t_m \\ \left[\beta_m^2 (t_m)^{2D_m} + \beta_p^2 [\tau^{2D_p} - (t_m)^{2D_p}] \right]^{1/2} & t_m \leq \tau \leq 380 \end{cases}$$

e

$$g'(r) = \frac{\Gamma\left(\frac{\alpha+1}{2}\right)}{\sqrt{\pi}\Gamma\left(\frac{\alpha}{2}\right)} (1+r^2)^{-\frac{\alpha+1}{2}}. \quad 2.23$$

La correttezza dell'approccio adottato è suffragata dal fatto che il modello riproduce le evidenze empiriche dei dati a disposizione. Vale pertanto la:

$$\overline{|r_1(1) + \dots + r_1(\tau)|^q} = \frac{\Gamma\left(\frac{q+1}{2}\right) \Gamma\left(\frac{\alpha-q}{2}\right)}{\sqrt{\pi}\Gamma\left(\frac{\alpha}{2}\right)} [\lambda(\tau, t_m)]^q. \quad 2.24$$

Per la calibrazione si procede come già descritto per i parametri β_m, D_m, α . Rimangono da calibrare i parametri β_p, D_p , per i quali si procede come segue: imponendo la condizione di continuità delle PDF si può definire β_p :

$$\beta_p = \frac{[t_m^{2D_m} - (t_m - 1)^{2D_m}]^{1/2}}{[t_m^{2D_p} - (t_m - 1)^{2D_p}]^{1/2}}. \quad 2.25$$

In questo modo resta solo da calibrare D_p , per il quale si procede come in considerando $t > t_m$.

La calibrazione così effettuata restituisce i valori di $\alpha = 3.39$; $\beta_m = 0.34 * 10^{-3}$; $D_m = 0.403$; $D_p = 1.4$; $\beta_p = 2.7 * 10^{-3}$

Capitolo 3 IL TEMPO IN FINANZA

Il concetto di tempo in finanza è per certi aspetti molto particolare, perché non sempre è legato al tempo fisico che regola il passare dei giorni e l'evolvere delle nostre azioni. I mercati aprono, il tempo fisico scorre, si arriva a metà giornata, il tempo fisico scorre, si arriva alla chiusura, e il tempo fisico continua a scorrere, sempre uguale. Il tempo fisico non presenta nel suo scorrere asimmetrie o disomogeneità. Invece, la non stazionarietà riscontrata nell'intraday, suggerisce che il tempo della finanza non necessariamente debba coincidere con il tempo fisico. Una possibile alternativa al tempo fisico può essere quella di definire un tempo in ragione del volume delle transazioni che progressivamente si realizzano. Ad esempio, un quesito su cui interrogarsi è come scorra il tempo in finanza durante la notte. E più in generale, se si prende un intervallo di tempo fisico qualsiasi, quale sarà il corrispondente intervallo di tempo finanziario, e come poterlo definire.

In questo capitolo si cerca di rispondere a queste domande.

Fino ad ora è stata analizzata la finestra temporale dell'*intraday* ed è stata utilizzata un'ottica di ensemble, in questo capitolo si esce dall'*intraday* e ci si spinge nell'*interday*, andando a considerare ritorni calcolati con τ relativi a più giorni. Nell'*interday* si utilizzerà un approccio di media lungo la serie temporale e medie su singola storia invocando stazionarietà ed ergodicità [4], [9]. L'obiettivo è provare se esiste un ponte comune fra l'*intraday* e l'*interday*, dato che per entrambi valgono leggi di scaling.

Inoltre ci si occuperà di verificare la reciproca compatibilità fra la trattazione intraday e la trattazione interday, per poi definire il concetto di tempo in finanza. Inoltre, si definirà la funzione per il parametro di scala λ_τ che a partire dal tempo fisico, permetta di calcolare il relativo valore del tempo finanziario. Si stabilirà così una relazione biunivoca fra i due tempi.

3.1 Tempo e scaling

Poiché in finanza la distribuzione dei ritorni è l'oggetto su cui si focalizza l'attenzione, il fatto che valga uno scaling inhomogeneo (con λ_τ), per le PDF dei ritorni, suggerisce in modo naturale che λ_τ possa costituire una scala dei tempi. I rendimenti, perché valga la validità dello scaling inhomogeneo, devono pertanto essere calcolati tutti a partire dal momento di apertura del mercato.

Si inizia andando a calcolare i λ_τ relativi ai ritorni intraday. Sono state fatte collassare le distribuzioni dei ritorni a $\tau = 1, 2, 4, 8, 16, 32, 64, 128, 150, 200, 250, 300, 350, 380$ minuti e sono stati calcolati i parametri di scaling [figure 3.5; 3.6] λ_τ tali da soddisfare la semplice formula di scaling

$$P_{R_\tau}(r_\tau) = \frac{1}{\lambda_\tau} g' \left(\frac{r_\tau}{\lambda_\tau} \right), \quad 3.1$$

dove la funzione g' è data dalla 2.23

Questo fatto suggerisce che sia possibile utilizzare la λ_τ di scala come un misuratore del tempo in finanza. Il λ_τ di scala è stato trovato imponendo che fosse minima la quantità Ds , definita come il valore assoluto della differenza fra la probabilità cumulativa teorica e la probabilità cumulativa della distribuzione empirica. Questo è utile nell'applicazione del Test di Kolmogorov (Appendice A) per verificare la compatibilità fra una distribuzione empirica ed una distribuzione teorica.

Attraverso la A2 del test di Kolmogorov (Appendice A) si ricava in funzione di Ds un valore che, qualora inferiore a 0.01, indica la non compatibilità della distribuzione cumulativa empirica con la distribuzione cumulativa teorica con un indice di confidenza pari al 99%. Di seguito vengono riportati i valori per λ_τ ricavati a partire dalle distribuzioni empiriche ed i relativi risultati del test di kolmogorov, tutti più di un ordine di grandezza superiore alla soglia.

τ (min)	1	2	4	8	16	32	64
λ_τ	0.00034	0.00056	0.00086	0.00126	0.00183	0.00261	0.00363
[A2]	0.44	0.49	0.50	0.50	0.50	0.48	0.49

τ (min)	128	150	200	250	300	350	380
λ_τ	0.0050	0.00649	0.00696	0.00761	0.00892	0.0104	0.0116
[A2]	0.50	0.48	0.5	0.46	0.43	0.44	0.44

Si vede che la relazione che lega tempo fisico al tempo finanziario segue una legge a potenza della forma della:

$$\lambda_\tau = a\tau^b \quad a = 0,0004 ; b = 0,537 \quad 3.2$$

Di seguito vengono riportate le varie distribuzioni empiriche ricavate dai dati dei ritorni aggregati su intervalli che iniziano all'apertura da frequenza del minuto a 128 minuti.

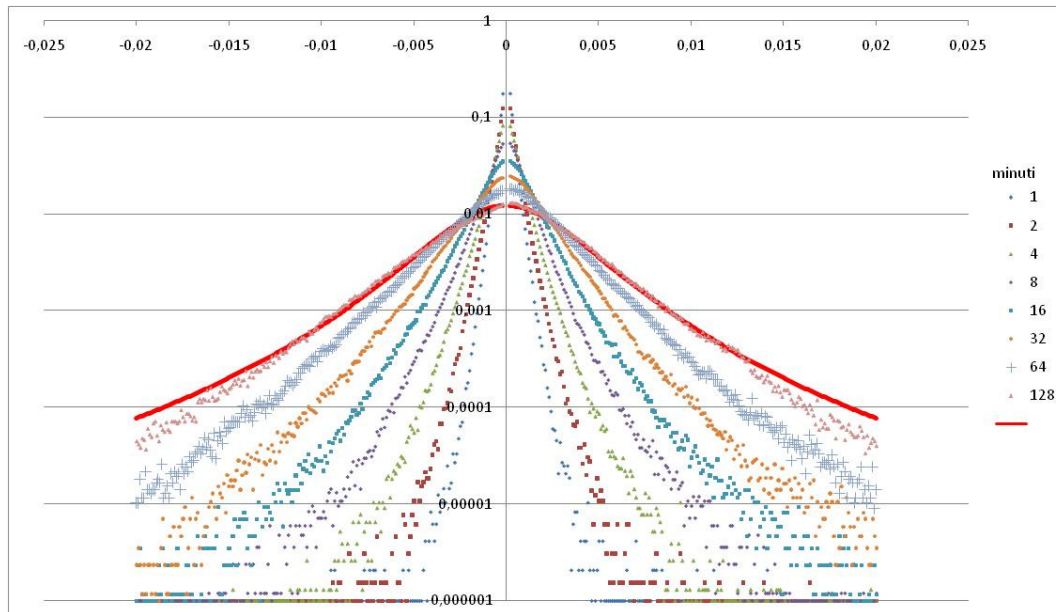


Figura 3.1 Distribuzioni empiriche dei ritorni aggregati per $\tau=1,2,4,8,16,32,64,128$ minuti

3.2 Oltre l'intraday: ritorni notturni e giornalieri ed unificazione

Utilizzando lo stesso formalismo finora adottato, è possibile estendere le proprietà di scaling anche ai ritorni interday. Questi sono caratterizzati dal fatto che superano la durata del giorno borsistico, ovvero escono dalla finestra intraday. Un esempio possono essere i ritorni notturni e quelli giornalieri.

Per tutti i ritorni interday non vale più l'approccio di ensemble finora utilizzato per i ritorni *intraday*, che ricordiamo sono un processo marcatamente non stazionario. Per i ritorni *interday* si ipotizzerà l'ergodicità e la stazionarietà, e si utilizzeranno medie lungo le serie temporali. Tutti i ritorni *interday* possono essere descritti dalla medesima funzione di scaling g .

Si analizzano ora i ritorni a $380min$ (pari come specificato a 380 minuti di contrattazione, da apertura a chiusura mercato) calcolato utilizzando quanto detto al paragrafo precedente, i ritorni *overnight* (notturno), e *dayday* di durata pari ad un giorno. E' chiaro che la procedura descritta ora può essere riprodotta per qualsiasi altro ritorno interday multiplo di un numero intero di giorni

Il ritorno notturno riveste particolare interesse in quanto tutti i ritorni interday comprendono il ritorno *overnight*. Infatti un ritorno interday pari a due giornate borsistiche è dato dalla somma di due ritorni $380min$ (dall'apertura alla chiusura) e di due ritorni *overnight*. Quindi risulta importante conoscere la durata in tempo fisico del ritorno *overnight* visto che rientra poi in tutti i calcoli relativi ai ritorni interday.

La procedura adottata per trovare il corretto parametro di scala $\lambda_{notturmo}$, λ_{380min} , λ_{dayday} consiste nel fare un fit delle distribuzioni empiriche con la funzione di scaling risolvendo numericamente le

$$P_{R_{notturmo}}(r_{notturmo}) = \frac{1}{\lambda_{notturmo}} g\left(\frac{r_{notturmo}}{\lambda_{notturmo}}\right) \quad 3.3$$

$$P_{R_{380min}}(r_{380min}) = \frac{1}{\lambda_{380min}} g\left(\frac{r_{380min}}{\lambda_{380min}}\right) \quad 3.4$$

$$P_{R_{dayday}}(r_{dayday}) = \frac{1}{\lambda_{dayday}} g\left(\frac{r_{dayday}}{\lambda_{dayday}}\right) \quad 3.5$$

Dove $\lambda_{notturmo}$ sono i particolari valori di λ di scala relativo al ritorno aggregato dell'overnight, 380min intraday e dayday e g è dato dalla 2.14.

Il ritorno notturno, ovvero il ritorno aggregato dell'overnight, è calcolato considerando l'ultima quotazione della giornata e la prima della giornata successiva, mentre il ritorno giornaliero è calcolato considerando la quotazione di apertura di mercato e la quotazione di apertura del giorno successivo ed il ritorno intraday 380min è calcolato considerando la quotazione di apertura e quella di chiusura di uno stesso giorno.

$$r_{notturmo}^l = \ln(S_a^l) - \ln(S_c^{l-1}) \quad 3.6$$

$$r_{giornaliero}^l = \ln(S_a^l) - \ln(S_a^{l-1}) \quad 3.7$$

$$r_{380min}^l = \ln(S_c^l) - \ln(S_a^l) \quad 3.8$$

Dove S_a e S_c indicano rispettivamente il prezzo di apertura e quello di chiusura.

Effettuando il test di Kolmogorov (Appendice A) si ritrova che la funzione di scaling g e le curve empiriche sono compatibili, oltre il 99%.

I risultati sono:

$$\lambda_{380min} = 0.0116; \quad \lambda_{notturmo} = 0.0058; \quad \lambda_{dayday} = 0.0130.$$

Ora si procede a verificare la coerenza e la compatibilità dei risultati.

Risulta infatti aperto il problema di definire la durata dell'intervallo overnight relativa al tempo fisico.

Per fare questo, si procede in due maniere parallele, la prima [A] che sfrutta quanto ottenuto con l'ensemble nell'intraday, la seconda [B] nell'interday. Ciò che si spera di trovare, è una stima del tempo fisico del ritorno notturno che sia la medesima in entrambi i casi, in maniera tale che vi sia un ponte fra le due scale di tempo, e le due trattazioni: interday ed intraday.

[A] *Si utilizza la 3.2, parametrizzata nella finestra temporale dell'intraday, ed invertita, per calcolare il tempo fisico relativo al ritorno notturno. Il risultato è una durata pari a 120 minuti.*

[B] *Sono state considerati intervalli di una notte, un giorno borsistico (380 minuti di contrattazioni), un giorno completo (giorno borsistico più notte), 2 notti e un giorno, due giorni completi, etc. fino a 15 giorni completi. Si è quindi cercato il valore del tempo fisico τ espresso in minuti della notte $[T_{night}]$ e del giorno borsistico $[T_{day}]$ che meglio adattasse i λ_τ empirici alla 3.2. Questo significa che se una notte dura un tempo T_{night} ed il giorno un tempo T_{day} , dove T_{night} e T_{day} corrispondono ad un tempo fisico di lunghezza arbitraria calcolato in minuti, allora la giornata completa composta da notte e giorno dovrebbe durare un tempo pari a $T_{full\ day} = T_{night} + T_{day}$. Ovviamente questo risultato non è imposto come condizione necessaria, ma si risolve variando i parametri di T_{day} e T_{night} in maniera da trovare i valori che rendano compatibili le distribuzioni empiriche con la funzione di scala λ_τ (3.2).*

Il numero di minuti relativo a T_{day} e T_{night} necessario per calcolare il parametro di scala λ_τ è stato ottenuto tramite :

$$T_{N\ giorni, M\ notti} = (M \cdot notti * T_{night}) + (N \cdot giorni * T_{day})$$

Dove in questo contesto N è il numero di giorno e M il numero di notti. Ovviamente M può valere al più $M = N \pm 1$. Questi, una volta definita la durata della notte e del giorno, forniscono il corretto valore di minuti fisici a cui ogni λ_τ fa riferimento, come si può notare dalla figura 3.4. Si nota che si vedono i punti procedere a coppie di due per valori di minuti superiori ai 500, ovvero la durata complessiva di una giornata totale, inquanto sono state considerate N giorni ed N notti, e N giorni e $N+1$ notti.

Come detto sono state trovate le durate della notte e del giorno in maniera da ottimizzare la compatibilità delle distribuzioni empiriche con la distribuzione teorica nella quale il tempo fosse espresso dalla relazione di cui sopra.

Il risultato ottenuto è Notte = 121 minuti, Giorno = 380 minuti. Il che è in ottimo accordo con la durata effettiva dell'intraday che è di 380 minuti, dato ottenuto ovviamente da una analisi diretta, e della notte calcolata al punto [A].

Questo è rappresentativo del fatto che vi sia un ponte fra la trattazione interday e la trattazione intraday, essendo i risultati ottenuti relativamente alle stesse quantità fra loro compatibili ed in ottimo accordo.

La funzione λ_τ (3.2), risulta quindi capace di fornire la corretta stima della durata in tempo finanziario λ_τ del ritorno calcolato su un intervallo di durata τ a partire dal momento di apertura di mercato. Di seguito il grafico delle probabilità cumulative, che si utilizzano coerentemente con quanto riportato in Appendice A, campionando fino a 64 giorni:

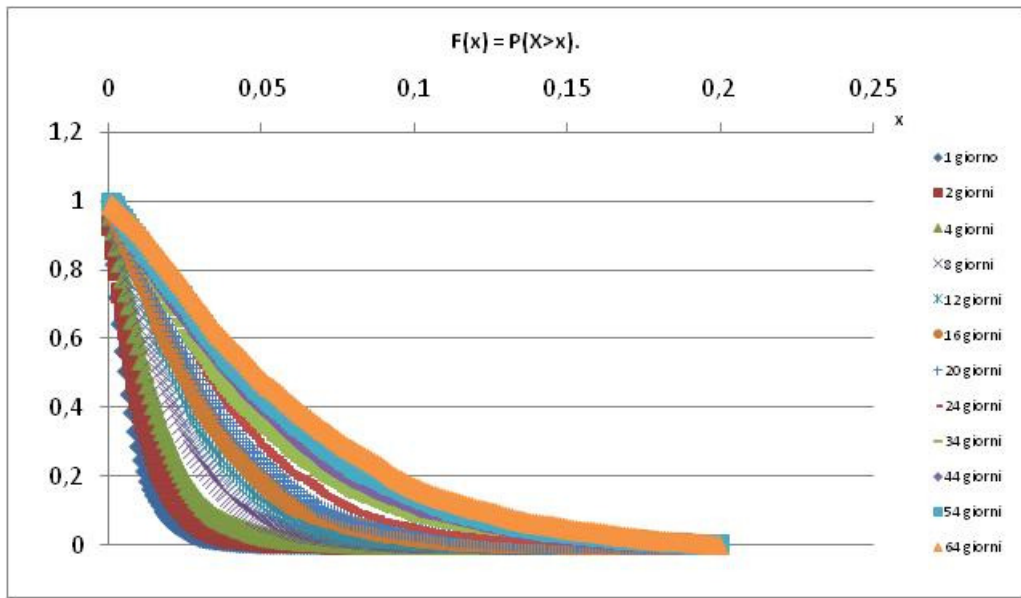


Figura 3.3 Distribuzioni cumulative dei ritorni aggregati per τ fino a 64 giorni

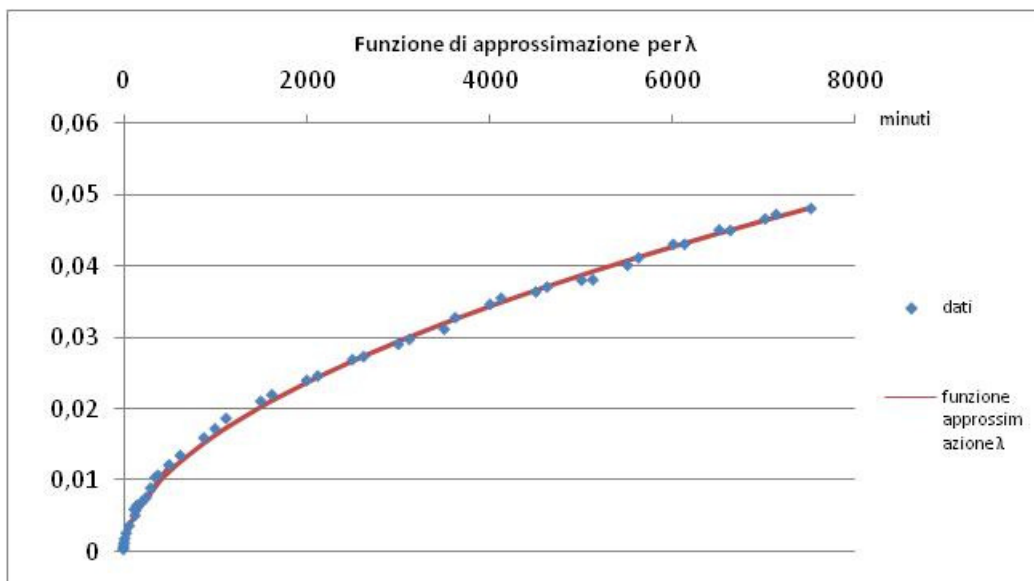


Figura 3.4: Funzione di approssimazione per λ : unione della trattazione intraday con la trattazione interday.

Questa particolare definizione di tempo, che trae la propria origine teorica nella formulazione universale dello scaling, risulta la più generale possibile per definire il tempo in maniera parallela e complementare al tempo fisico.

L'idea che ne rappresenta la base è che in un sistema che evolve nel tempo -nel quale grandezze calcolate considerando medie di ensemble per i ritorni intraday e medie temporali per i ritorni interday- l'idea di tempo stessa possa essere legata alla evoluzione temporale di queste grandezze.

Nel sistema che siamo andati ad analizzare la distribuzione di probabilità evolve rispetto al tempo fisico in maniera diversa fra giorno e notte: durante il periodo di apertura dei mercati evolve velocemente, mentre durante il periodo di chiusura dei mercati evolve più lentamente. Dunque l'evolvere di λ_τ non è più legato al tempo fisico, quanto piuttosto ad un tempo finanziario definito dall'evoluzione del parametro di scala λ_τ . Questo vuol dire che il tempo relativo al quale si ha evoluzione stabile e costante dei mercati finanziari non sia il tempo fisico, ma sia il tempo finanziario. Questo, come detto, è possibile ottenerlo a partire dallo scaling delle funzioni di distribuzione dei ritorni, che per i diversi intervalli temporali su cui sono calcolati forniscono i relativi valori di λ_τ . È l'evoluzione di λ_τ che permette di formulare un concetto di tempo differente dal tempo fisico e che rappresenta un tempo proprio di evoluzione universale del mercato.

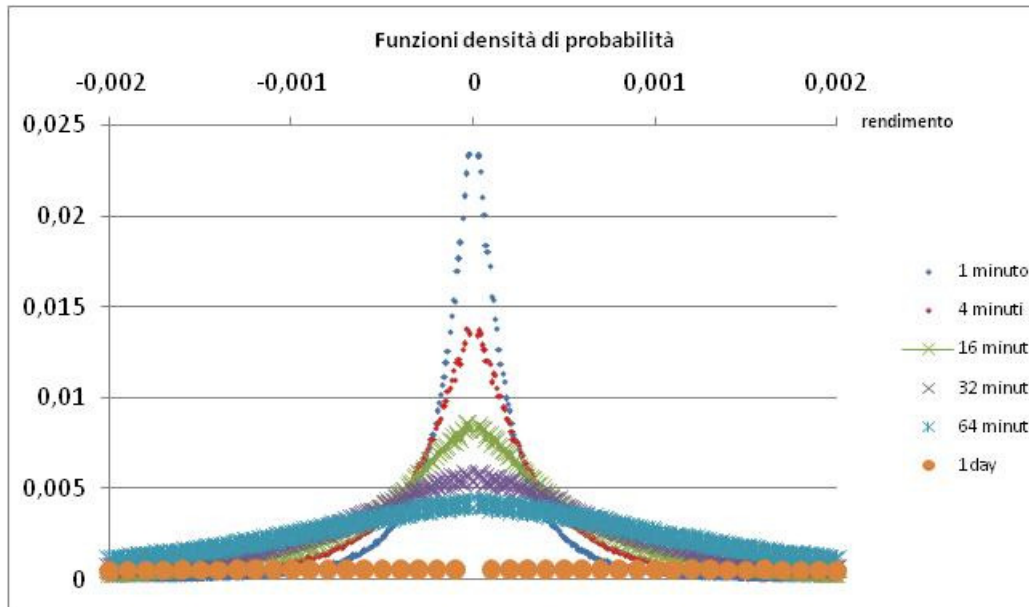


Figura 3.5 Densità di probabilità per ritorni fino ad 1 giorno

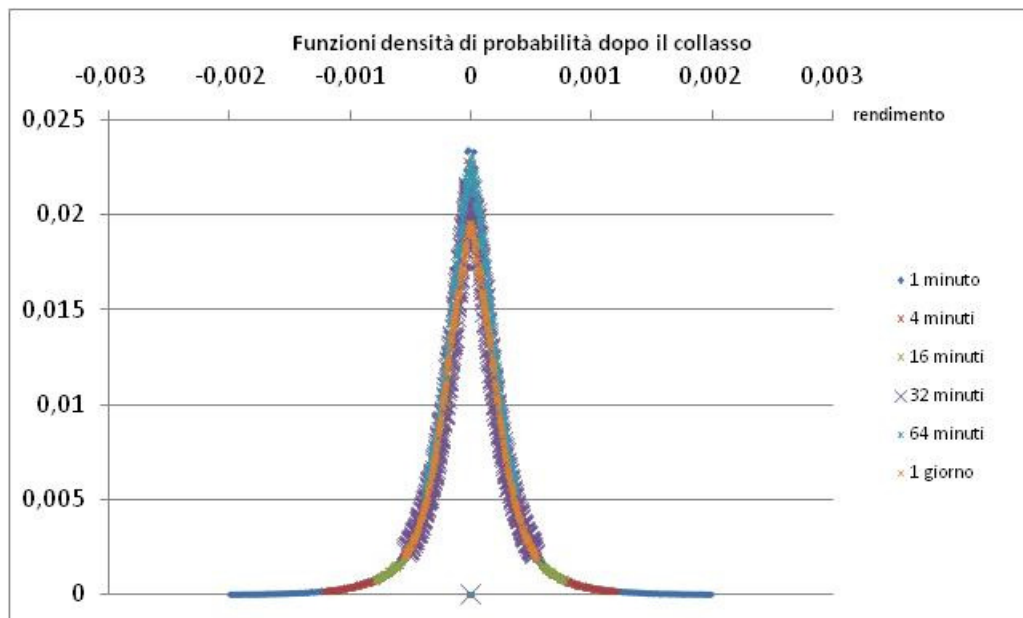


Figura 3.6 Funzione densità di probabilità per ritorni fino ad 1 giorno, dopo il collasso

Capitolo 4 TRADING

Il modello costruito nel capitolo 2 sulla base dello scaling considera per ipotesi le correlazioni lineari uguali a zero. Tuttavia i mercati reali, come visto in 1.5, presentano correlazioni lineari basse ma non nulle. Sono queste correlazioni lineari, su scala di tempo molto piccola (pochi minuti), che aprono la strada alla possibilità di arbitraggio. Nello spirito di un approccio di finanza quantitativa si utilizzano delle strategie di trading basate sulle correlazioni lineari per vedere quanto queste influiscano sui mercati reali, e per testare la applicabilità del modello definito nel capitolo 2.

4.1 FUNZIONE QUANTILE

La funzione quantile rappresenta lo strumento per mettere in pratica una prima strategia di trading per l'asset preso in considerazione. Esso è dato dalla soluzione numerica della equazione 4.2, ed è composto da due rami, uno positivo ed uno negativo. Dato un certo valore di Q , con $0 < Q < \frac{1}{2}$, il prezzo predetto dal modello si trova sopra il quantile positivo, o rispettivamente sotto il quantile negativo, in $(1-2Q)$ dei casi totali. Per questo motivo la funzione quantile può essere un utile strumento per identificare i trend, conseguenza della correlazione lineare non nulla, e fonte di possibile profitto.

Per calcolare la funzione quantile, si utilizzerà quanto già ricavato sulla PDF dei ritorni aggregati:

$$p(r_\tau, \tau) = \frac{\Gamma\left(\frac{\alpha+1}{2}\right)}{\sqrt{\pi}\Gamma\left(\frac{\alpha}{2}\right)} \frac{1}{\lambda(\tau, t_m)} \left[1 + \left(\frac{r_\tau}{\lambda(\tau, t_m)} \right)^2 \right]^{-\frac{\alpha+1}{2}}, \quad 4.1$$

con $\lambda(\tau, t_m)$ dato dall'equazione 2.22. Una volta definito il valore di Q voluto, si ricava $r_{min,\tau}(Q)$ risolvendo numericamente l'equazione

$$Q = \int_{-\infty}^{r_{min,\tau}} dr p(r, t). \quad 4.2$$

Visto la parità della funzione di scaling $g(r)$, il livello superiore del quantile si ricava a partire da quello inferiore semplicemente con un cambio di segno, ovvero:

$$r_{max,\tau}(Q) = -r_{min,\tau}(Q). \quad 4.3$$

I valori limite superiore ed inferiore dell'asset sono ricavati tramite esponenziazione:

$$S_\tau = S_0 e^{\sum_{t=1}^{\tau} r_1(t)}.$$

4.4

E' inoltre possibile, come è stato fatto in questo lavoro, tralasciare l'ultima espressione che prevede il passaggio attraverso il prezzaggio dell'asset, e lavorare direttamente a partire dai ritorni aggregati calcolati dal momento di apertura del mercato. Ne segue che i due rami del quantile non rappresenteranno più il prezzo minimo ed il prezzo massimo da superare, ma il ritorno aggregato minimo e il ritorno aggregato massimo che devono essere superati dal ritorno aggregato calcolato a partire dall'istante di apertura del mercato e generato dalla quotazione in tempo reale dell'asset.

La scelta di $0 < Q < \frac{1}{2}$ risulta arbitraria e relativa alla strategia che si vuole applicare, tuttavia per ogni scelta di Q , $r_{max,\tau}(Q)$ e $r_{min,\tau}(Q)$ rappresentano le barriere di prezzo che hanno la caratteristica di contenere il ritorno aggregato relativo all'intervallo τ con probabilità di $(1-2Q)$.

Tutta l'informazione necessaria per la costruzione della funzione quantile è contenuta nella parametrizzazione della PDF, non è pertanto necessaria alcuna informazione ulteriore.

Di seguito vengono riportati i grafici delle barre per quantili a 0.05; 0.10; 0.25.

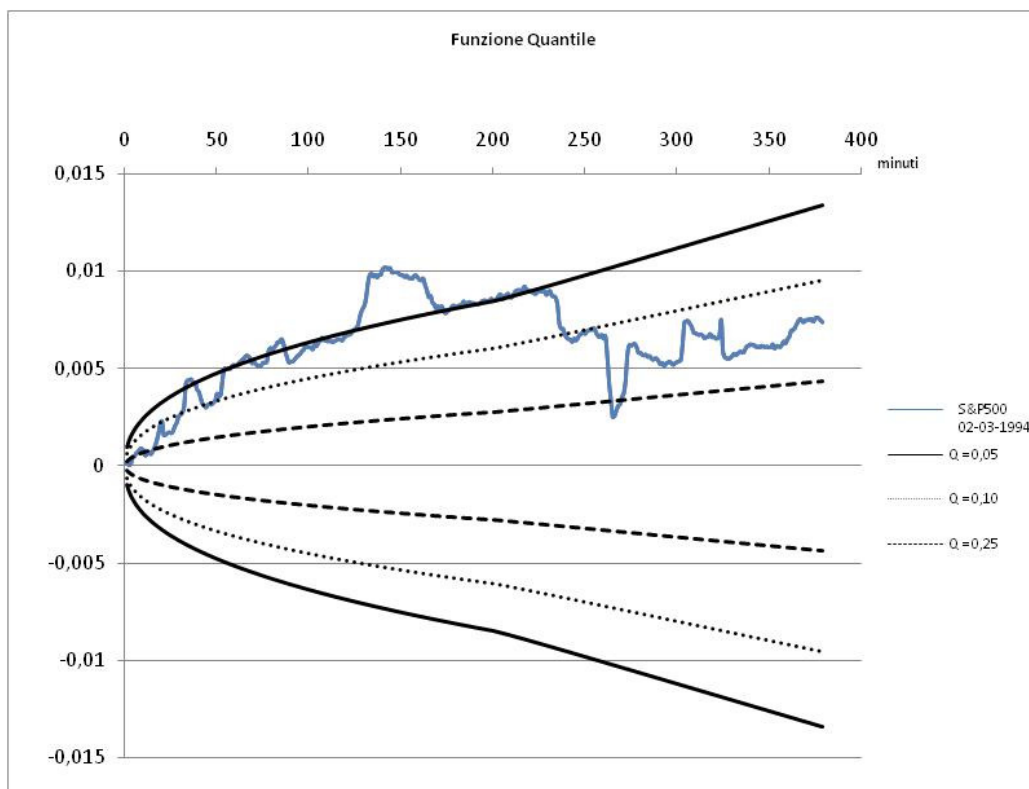


Figura 4.1 Funzione Quantile

4.2 STRATEGIE DI TRADING

Sulla linea di [7], è possibile utilizzare quanto visto fino ad ora per definire strategie operative di trading. Una strategia di trading consiste in una serie di regole operative dalle quali si ricavano segnali di apertura o chiusura di posizioni. Una posizione può essere lunga o corta. Aprire una posizione lunga, o “andare lungo” a seguito di un segnale al tempo t significa acquistare un asset al prezzo $S(t)$, chiudere una posizione lunga a seguito di un segnale al tempo t significa vendere un asset al prezzo $S(t)$. Aprire una posizione corta a seguito di un segnale al tempo t significa vendere un asset al prezzo $S(t)$, mentre ricoprire la posizione significa ricomprare l’asset precedentemente venduto. Non è possibile operare istantaneamente nel mercato, e in sede di simulazione delle strategie si è adottato la filosofia di essere quanto più aderenti a quella che può considerarsi una vera implementazione del modello al mercato reale, considerando un tempo di ritardo fra segnali ed ordini pari ad un minuto, che corrisponde alla frequenza di campionamento dei dati.

Dunque, operando con dati campionati alla frequenza del minuto, a seguito del segnale di entrata nel mercato il prezzo di entrata sarà ritardato di un minuto rispetto al momento in cui si è registrato il segnale, mentre se si opera con dati a frequenza di dieci minuti, il momento di entrata nel mercato avverrà con un ritardo di dieci minuti rispetto al momento in cui è stato generato il segnale. Ovviamente si apre una posizione lunga o corta se si ipotizza rispettivamente che il prezzo salirà o scenderà. Altrimenti, semplicemente si resterà fuori dal mercato, ovvero non si parteciperà alle contrattazioni.

Come già detto si è scelto di operare su S&P 500 a frequenza massima, con dati campionati ogni minuto. In questo modo si hanno a disposizione per ogni giornata borsistica 380 realizzazioni. Si eseguono le calibrazioni e si calcolano le barriere di quantile, e a partire da queste si sviluppano le diverse strategie descritte nel seguito.

Si è ipotizzato che non vi siano costi di transazione per ogni operazioni, in quanto operando con piccoli capitali questi costi rappresentano la parte maggioritaria di spesa, tuttavia quando si opera con capitali di molti ordini di grandezza superiore ai costi di transazione questi diventano pressoché trascurabili.

Una volta messa a punto una strategia vi sono due modi fondamentalmente diversi di testarla a partire da una serie di dati a disposizione. Il primo consiste nell’utilizzare tutta l’informazione presente nei dati per approntare i parametri necessari all’operatività del modello, per poi andare a testare il modello sui dati stessi che sono serviti a definire i parametri. Il secondo modo consiste nel suddividere i dati a disposizione in due intervalli, in maniera tale che il primo intervallo serva a fornire la calibrazione necessaria al modello mentre il secondo intervallo serva a verificare le capacità del modello. In quest’ottica è anche possibile un modello che utilizzi gli ultimi N periodi di tempo disponibili per determinare i parametri necessari al modello

per funzionare, in maniera che il modello lavori sempre con parametri ricavati da dati quanto più aggiornati possibile.

In questa sede sono stati ricavati da una prima serie di dati i parametri della calibrazione necessari per la definizione del modello nel capitolo 2, che poi sono stati utilizzati per testare le strategie di trading su tempi successivi.

Come già accennato in apertura, le correlazioni lineari svolgono un ruolo fondamentale nel successo di una strategia di trading: infatti l'obiettivo principale è quello di riuscire a trovare una serie di regole che identifichino la presenza di un movimento ben definito e stabile del mercato.

Come già visto le correlazioni lineari, manifestano un decadimento esponenziale all'aumentare dell'intervallo di tempo su cui sono calcolate, e comunque diventano irrilevanti dopo i dieci minuti. Dal momento che i dati a disposizione sono per loro natura discreti, è possibile calcolare le correlazioni lineari solo con multipli del tempo $\theta = 1$ minuto. Risulta evidente che maggiore è la dimensione delle correlazioni lineari maggiore è la probabilità che si verifichi un trend reale. Si è già visto come siano cambiate le correlazioni lineari nell'arco dei trent'anni analizzati, e risulta altresì chiaro che molto probabilmente una strategia che faccia largo uso di queste ultime e che testata nel mercato degli anni '80 - '90 dia in questo periodo buoni risultati possa non mantenerli nelle decadi successive, a causa proprio di una evoluzione sostanziale del mercato.

Questo fatto indica che il mercato tende ad evolvere per diventare più efficiente e quindi minimizzare le possibilità di arbitraggio.

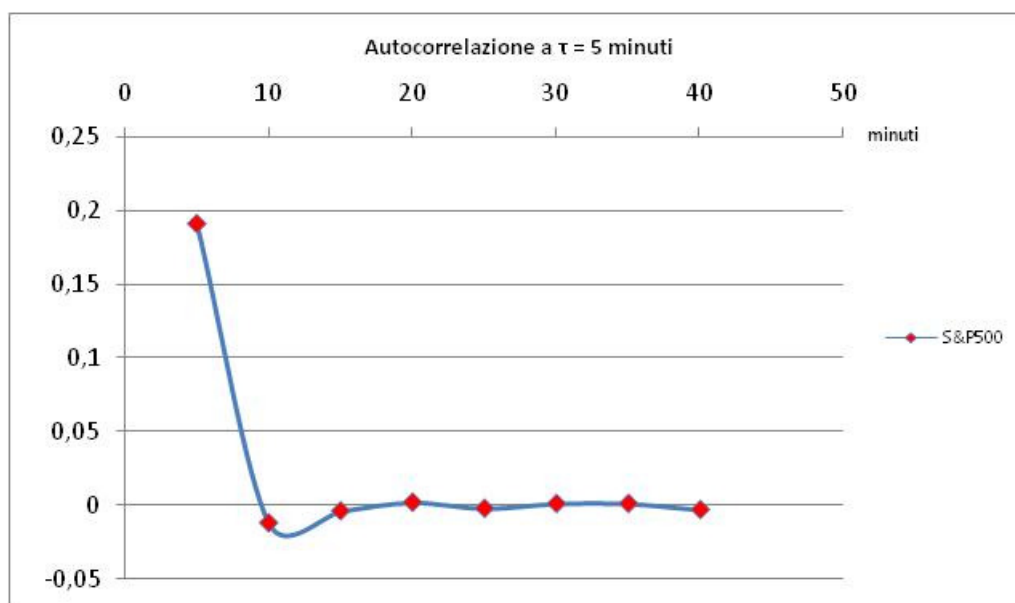


Figura 4.2 Funzione di Autocorrelazione a 5 minuti

4.3 STRATEGIE BASATE SUI QUANTILI

Di seguito verranno descritte varie strategie, in ordine di complessità.

4.3.1 Strategia barriere Q Semplici (SQS)

Una volta calcolate le barriere di quantile, una barriera superiore ed una barriera inferiore, per ogni tempo nella finestra temporale dell'intraday, si può procedere a stabilire una strategia operativa. A causa delle correlazioni lineari non nulle, il superamento positivo o negativo di una di queste barriere, può significare la presenza di un trend. Con il termine "superamento positivo" si intende il verificarsi simultaneo di queste due condizioni: il valore del ritorno aggregato dell'asset al tempo $(\tau-1\theta)$, dove per τ e θ si ricorda la notazione introdotta in 1.12, deve essere inferiore al valore della barriera di quantile positivo per poi divenire maggiore al tempo successivo (τ) . Con il termine "superamento negativo" si intende il verificarsi simultaneo di queste due condizioni: il valore del ritorno aggregato dell'asset al tempo $(\tau-1\theta)$ deve essere maggiore del valore della barriera di quantile negativo per poi divenire minore al tempo successivo (τ) .

Una volta che è avvenuto il superamento di una di queste due barriere si aprirà una posizione.

Le posizioni verranno poi chiuse nel momento in cui il rendimento aggregato dell'asset passerà in senso opposto la barriera precedentemente oltrepassata, o, se questo non avviene, a prezzo di chiusura del mercato. In questa strategia non è stata volutamente considerata alcuna possibile evoluzione overnight. Ovviamente la strategia ha successo se una volta superate le barriere di quantile l'asset continua nel suo trend. Nel seguito, per semplicità, $R_S(t)$ indica il ritorno aggregato dell'asset S relativo all'intervallo di tempo t calcolato come sempre a partire dall'apertura di mercato.

Dunque, se non vi sono posizioni aperte:

- (1) Si acquista se $R_S(t) > Q^+(t)$ e $R_S(t-1) < Q^+(t-1)$
- (2) Si vende se $R_S(t) < Q^-(t)$ e $R_S(t-1) > Q^-(t-1)$

A seguito dell'apertura di una posizione,

- (1) Si acquista se $R_S(t) > Q^-(t)$ e $R_S(t-1) < Q^-(t-1)$
- (2) Si vende se $R_S(t) < Q^+(t)$ e $R_S(t-1) > Q^+(t-1)$
- (3) A fine giornata, se vi sono ancora posizioni aperte, si chiudono.

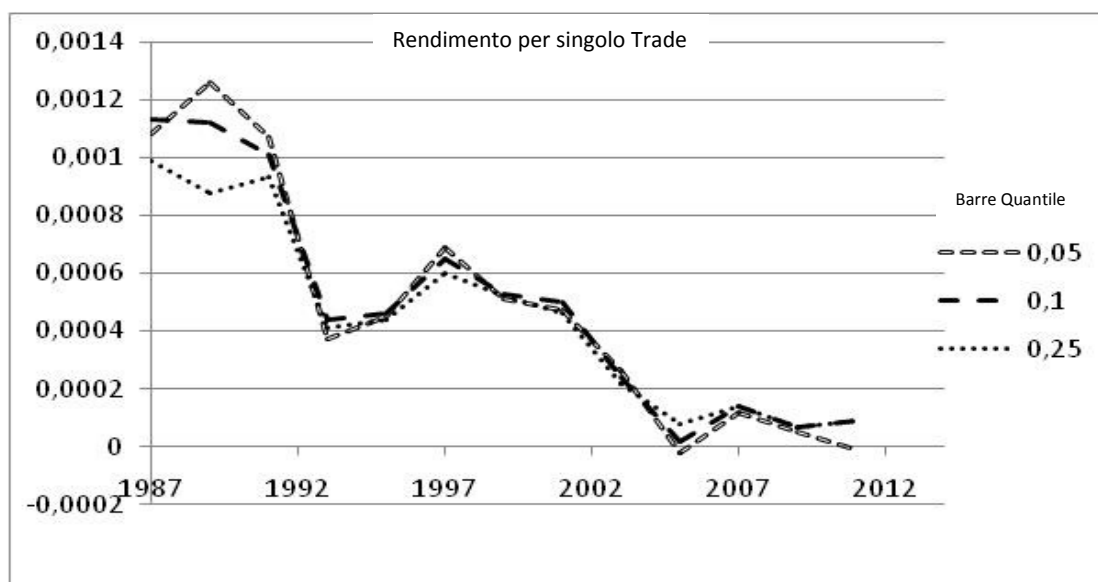
La scelta di quale dei vari quantili possibili (0.25; 0.10; 0.05) usare è ovviamente non imposta dal modello, tuttavia ci si può aspettare che utilizzando quantili più stretti (0.25), ovvero tali che la probabilità di avere al proprio interno l'asset sia del 75%, il

numero di operazioni cresce, mentre utilizzando quantili più larghi (0.05), ovvero tali che la probabilità di avere al proprio interno l'asset sia del 95%, il numero di operazioni diminuisce.

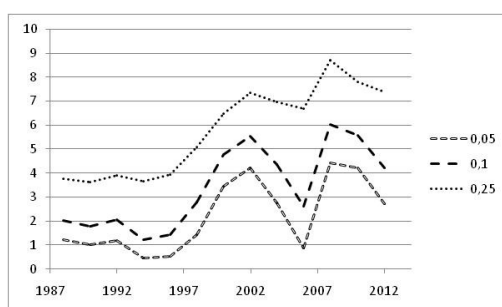
I risultati vengono esposti secondo la seguente logica: Viene posta innanzitutto l'attenzione sul ritorno per singolo trade, in maniera da avere un risultato universale confrontabile fra diverse tipologie di strategie a maggiore o minore numerosità di trade, e a seguire si pone l'attenzione sul numero effettivo di trade. In questa maniera è possibile discriminare una strategia piuttosto che un'altra tenendo conto ad esempio dei costi di negoziazione per eseguito.

Ovviamente il profitto totale di una strategia si ottiene moltiplicando il profitto per singolo trade per il numero di trade complessivi.

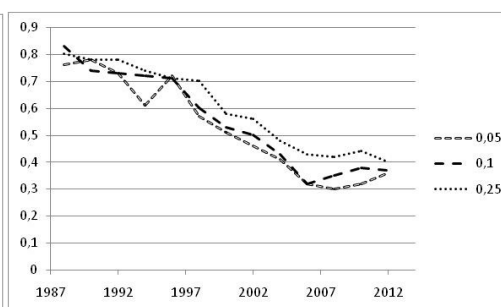
Di seguito i risultati :



Numero di trade medi a giornata borsistica



Frazione di giornate concluse con rendimento positivo



Commento alla strategia SQS

La strategia SQS rappresenta il primo e più semplice tentativo di implementazione della teoria di scaling al mercato. La strategia funziona in presenza di trend dall'andamento definito e stabile. Infatti in presenza di correlazioni lineari molto lunghe nella scala temporale, e di valore sensibile, questa strategia riesce ad ottenere i migliori risultati.

Il maggior inconveniente della strategia è rappresentato dal fatto che molto spesso a segnali di acquisto al tempo t , seguono segnali di ricopertura delle posizioni ai tempi immediatamente successivi. Questo comportamento deriva dal fatto che per l'arbitrarietà della scelta della barriera di quantile quando l'asset supera la suddetta barriera, negli istanti di tempo immediatamente successivi staziona in un intorno del valore della barriera, manifestando oscillazioni che spesso lo fanno ricadere oltre il limite della barriera stessa.

I dati, filtrati andando a selezionare solamente i trend effettivi, quelli ovvero che hanno manifestato guadagni maggiori, rendono conto di un discreto rapporto segnale rumore (25%). È ovvio che se si considerassero i costi di transazione questo provocherebbe la non applicabilità in pratica di questa strategia.

4.3.2 Strategia barriere Q a Bande divise (SQD)

Per ovviare al maggior problema della strategia precedente, ovvero l'insorgenza di un gran numero di segnali di acquisto / vendita in prossimità del passaggio delle barriere, si può procedere come segue:

L'idea di base consiste nel dividere ogni barra di quantile in due diverse barre [rispettivamente Q_{ramo+} e Q_{ramo-}], molto vicine fra loro e modulate o da una distanza standard o da una funzione dipendente dal valore medio della volatilità, calcolata come media di ensemble (sulla serie di dati a disposizione per la calibrazione) per ogni valore di t all'interno della giornata borsistica. In riferimento alla barra positiva di quantile, essa verrà suddivisa in due rami, uno superiore ed uno inferiore. Si acquista al superamento del ramo superiore, e si vende al superamento opposto del ramo inferiore. Lo stesso considerando la barra di quantile negativa.

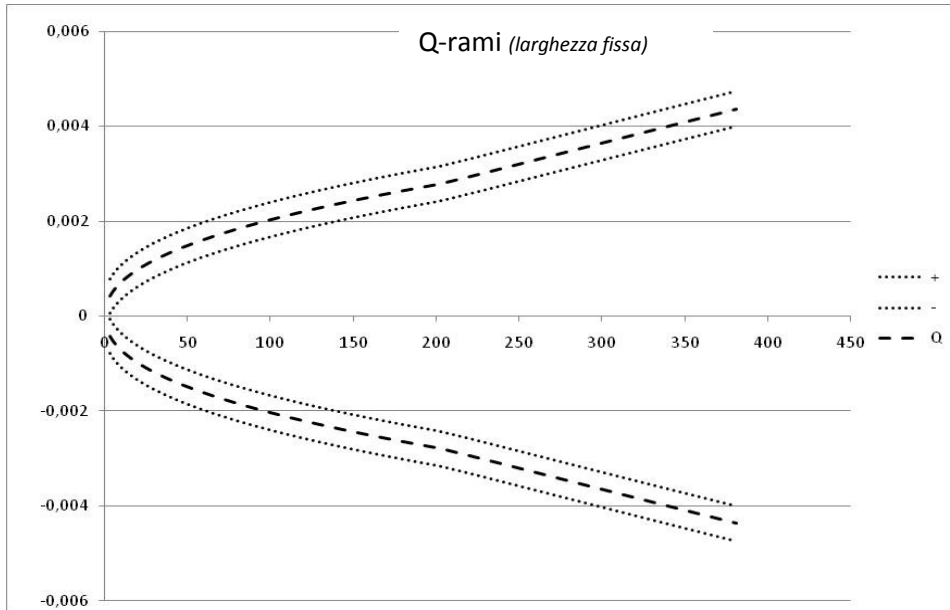


Figura 4.3 Funzione Quantile con Q-rami

La strategia si dimostra efficace nel far diminuire il numero di trade, come riporta il grafico seguente:

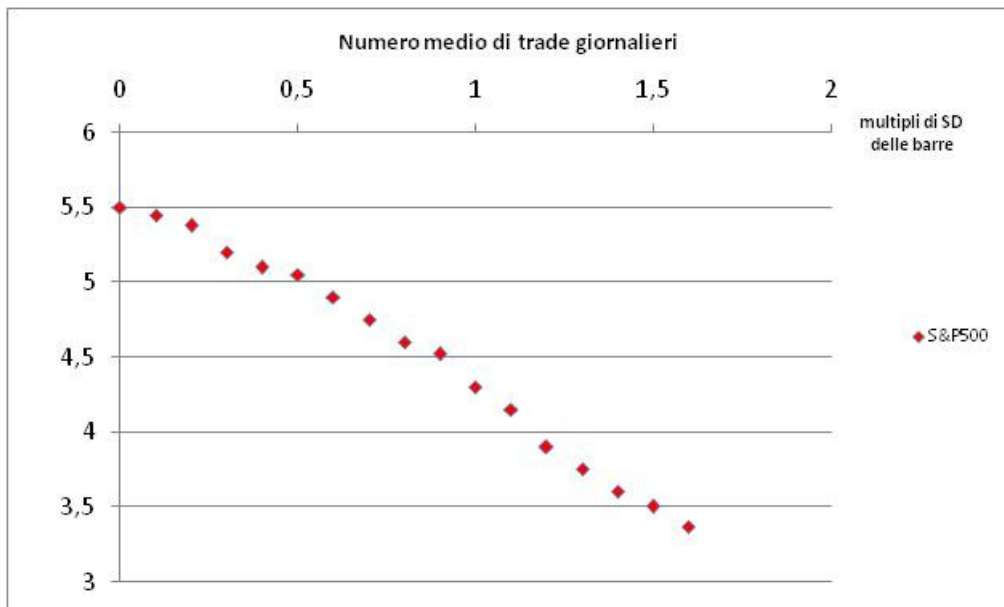


Figura 4.4 Numero medio di trade giornalieri con strategia a Qrami in funzione della larghezza delle bande (in unità di St.Dev)

In sintesi, la strategia segue le seguenti regole:

- $Q^{++}(t)$ indica Il ramo superiore della barra di quantile positiva
- $Q^{+-}(t)$ indica Il ramo inferiore della barra di quantile positiva
- $Q^{-+}(t)$ indica Il ramo superiore della barra di quantile negativa
- $Q^{--}(t)$ indica Il ramo inferiore della barra di quantile negativa

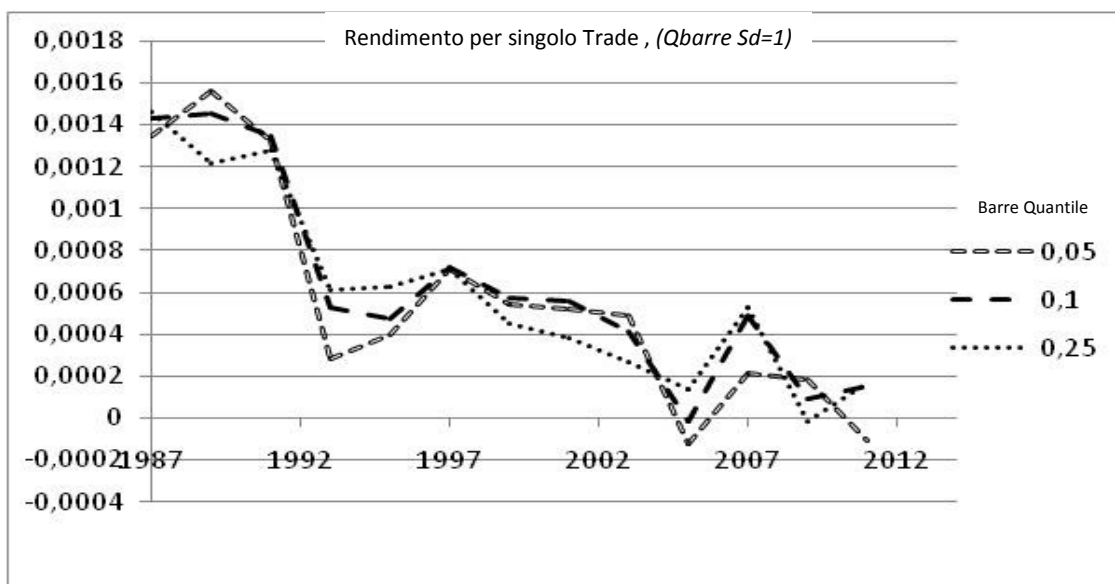
Se non vi sono posizioni aperte:

- (3) Si acquista se $R_S(t) > Q^{++}(t)$ e $R_S(t-1) < Q^{++}(t-1)$
- (4) Si vende se $R_S(t) < Q^{--}(t)$ e $R_S(t-1) > Q^{--}(t-1)$

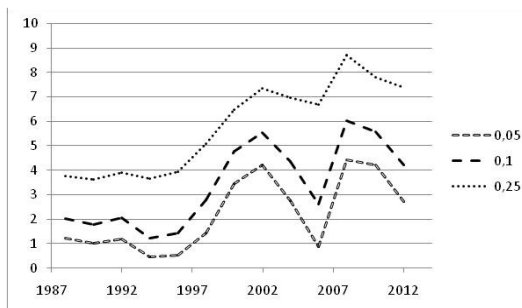
A seguito dell'apertura di una posizione,

- (4) Si acquista se $R_S(t) > Q^{-+}(t)$ e $R_S(t-1) < Q^{-+}(t-1)$
- (5) Si vende se $R_S(t) < Q^{+-}(t)$ e $R_S(t-1) > Q^{+-}(t-1)$
- (6) A fine giornata, se vi sono ancora posizioni aperte, si chiudono.

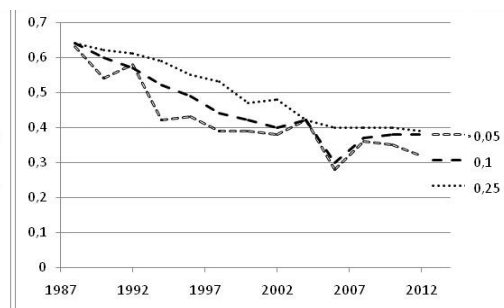
Di seguito i risultati:



Numero di trade medi a giornata borsistica



Frazione di giornate concluse con rendimento positivo



Commento alla strategia SQS

Come si può notare la strategia risulta effettivamente efficace nel diminuire considerevolmente il numero di trade. Inoltre dato che la larghezza delle bande può essere scelta arbitrariamente come multiplo della deviazione standard questo può essere un ulteriore fattore di adattamento del modello a situazioni eterogenee di mercato nelle quali si abbia un maggior o un minor peso dei costi di transazione per eseguito, in maniera tale da lasciare una possibilità di calibrazione volta a massimizzare il profitto.

4.3.3 Strategia barriera Q a Bande di ST.DEV (SQV)

Un ulteriore sviluppo rispetto all'idea del paragrafo precedente consiste nell'utilizzare anziché la volatilità calcolata con media di ensemble, che dunque risulta essere fissata e non dipendente dalla particolare evoluzione della giornata borsistica, una approssimazione istantanea di quest'ultima.

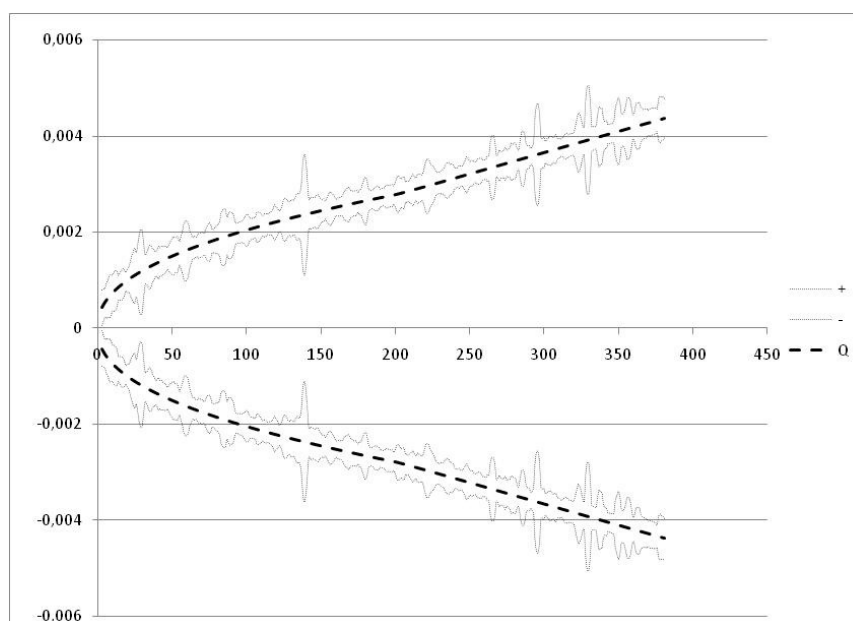


Figura 4.5 funzione quantile a Bande di ST. DEV.

Risulta così possibile eliminare momentanei eccessi di volatilità che possano generare segnali spuri. La strategia ha fornito risultati pressoché uguali alla strategia precedente, per cui si omette e si cita solo per completezza

4.3.4 Strategia a Barre di quantile disaccoppiate

La strategia a barre di quantile disaccoppiate prevede di utilizzare una differente percentuale di quantile fra la barra di quantile superiore e la barra di quantile inferiore, in maniera tale che il modello sia più sensibile a segnali selettivi di acquisto o di vendita, relativamente a diverse percentuali di quantile. E' pertanto possibile variare le barre di quantile o arbitrariamente, per seguire una strategia soggettiva, o ricercando correlazioni con quantità desumibili dal modello (ritorni overnight, volatilità del giorno precedente etc.) in maniera da ottimizzare il rendimento della strategia.

numero trade per gg

	Q- 0,25	Q- 0,1	Q- 0,05
Q+ 0,25	5,53	5,23	4,95
Q+ 0,1	5,06	4,76	4,48
Q+ 0,05	4,94	4,64	4,36

ricavo (punti base) singolo trade

	Q- 0,25	Q- 0,1	Q- 0,05
Q+ 0,25	0,72	0,73	0,73
Q+ 0,1	0,75	0,76	0,76
Q+ 0,05	0,73	0,73	0,73

Di seguito il grafico con riportata la per la barra di quantile positiva un valore $Q+=0.25$, mentre per la barra di quantile negativa un valore $Q-=0.05$

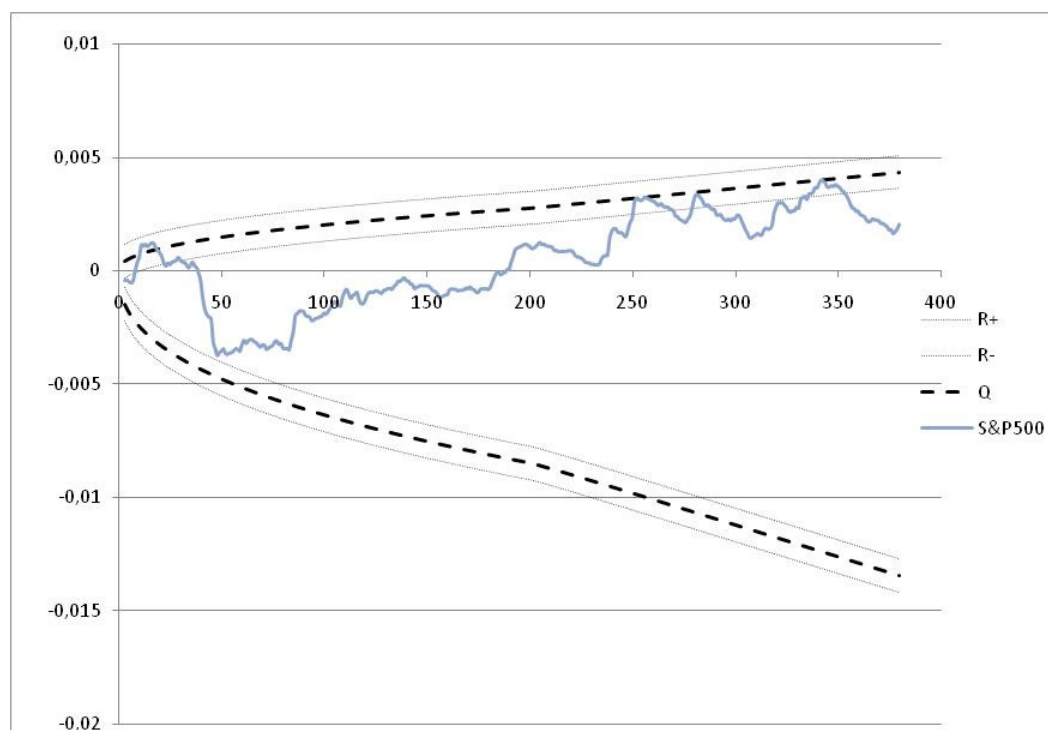


Figura 4.6 funzione quantile a Barre disaccoppiate

4.3.5 Strategia con take profit condizionato

Questa strategia si può applicare come complemento ad una qualsiasi delle strategie analizzate in questo paragrafo, e consiste in una presa di profitto condizionata al raggiungimento di una soglia definita in base ad una seconda barra di quantile modulata in base al profitto atteso. In pratica sia che si sia entrati lunghi sia che si sia entrati corti nel mercato, si esce dalla posizione al raggiungimenti di un margine, definito al momento dell'entrata nel mercato. E' anche possibile definire un livello oltre al quale non continuare una certa strategia, condizionata ad aumenti anomali della volatilità del titolo e rendimenti di segno contrario rispetto a quelli favorevoli alla condizione di entrata nel mercato.

Le barre di quantile che determinano l'uscita dal mercato (condizionatamente alla precedente entrata avvenuta al superamento della barra standard di quantile) sono barre che contengono al loro interno una elevatissima parte dei dati, ovvero il 99,6%-99,8% e l'uscita dal mercato al raggiungimento di questa soglia evita di perdere quanto ipoteticamente guadagnato mediante il buon successo della strategia a seguito di correzioni di mercato vicino alla chiusura, spesso frequenti a causa del rientro nelle posizioni di chi ha operato con strategie di leva finanziaria.

In particolare si è notato che questa particolare strategia fornisce i risultati migliori in periodi di forte incertezza e volatilità. Facendo riferimento a complemento della strategia SQS, le regole da seguire sono le seguenti:

Se non vi sono posizioni aperte:

- | | |
|--------------------|---|
| (5) Si acquista se | $R_S(t) > Q^+(t)$ e $R_S(t-1) < Q^+(t-1)$ |
| (6) Si vende se | $R_S(t) < Q^-(t)$ e $R_S(t-1) > Q^-(t-1)$ |

A seguito dell'apertura di una posizione,

- | | |
|----------------------|---|
| (7) Si riacquista se | $R_S(t) > Q^-(t)$ e $R_S(t-1) < Q^-(t-1)$ oppure se |
| | $R_S(t) > Q_{A+}(t)$ |
| (8) Si rivende se | $R_S(t) < Q^+(t)$ e $R_S(t-1) > Q^+(t-1)$ oppure se |
| | $R_S(t) < Q_{A-}(t)$ |

A fine giornata, se vi sono ancora posizioni aperte, si chiudono

.Il grafico che segue rappresenta il titolo S&P500 con barre di quantile e rami di quantile.

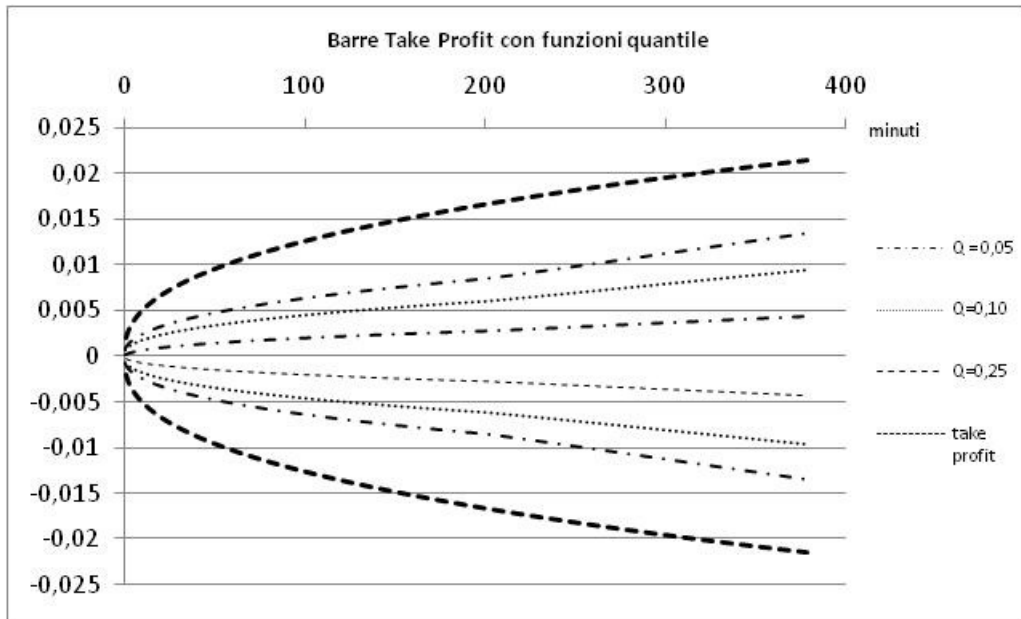
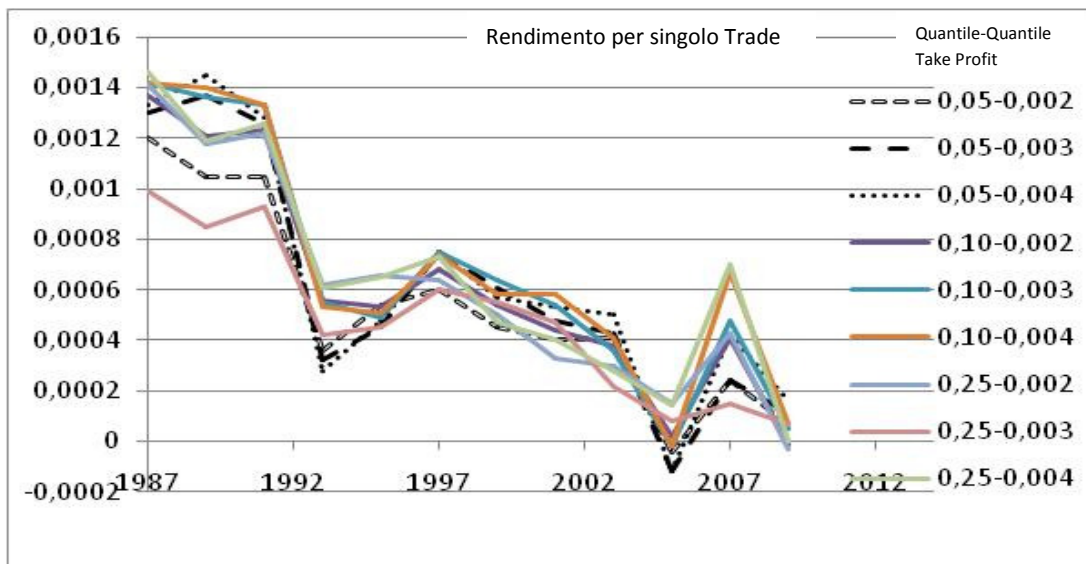


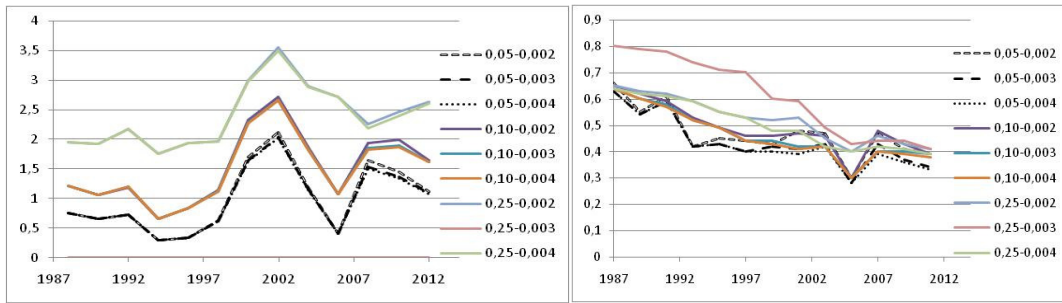
Figura 4.7 funzione quantile a con bande di take profit

Di seguito i risultati:



Numero di trade medi a giornata borsistica

Frazione di giornate concluse con rendimento positivo



Commento alla strategia SQS

La strategia è particolarmente utile per mettersi al riparo da brusche inversioni di mercato o nei periodi in cui la volatilità è molto elevata. Altresì è utile qualora i mercati attendano notizie di carattere macroeconomico o che comunque abbiano un forte impatto destabilizzante.

4.4 STRATEGIE TRADING VELOCI

Il principale svantaggio delle strategie fino ad ora analizzate risiede nella relativa lentezza di un eventuale cambio di strategia. Se il mercato presenta un andamento monotono lento, principalmente crescente o decrescente ma fondamentalmente coerente nell'arco di tempo dell'intraday, le strategie trattate fino ad ora rappresentano un ottimo strumento di profitto. Tuttavia il "tallone d'Achille" risulta essere la lentezza tra il cambio di trend del mercato e il suo sfruttamento per ottenere buoni profitti (o per limitare le perdite). Maggiore è la volatilità di mercato maggiore la possibilità di utilizzare questa volatilità se ci si sposta seguendo il mercato, come suggerito dall'analisi delle correlazioni. E' invece probabile che quando ci si imbatte in un trend di medio o breve periodo, dove il medio o breve periodo è definito dalla scala delle correlazioni, questo si esaurisca prima di poter essere sfruttato.

La scala di tempo che definisce la durata di un trend nel medio-breve periodo è quello appunto determinato dalla correlazione lineare, non superiore ai quindici minuti. E' quindi necessario che la strategia possa avere tempi simili, per accorgersi del cambiamento di trend, uscire da una eventuale posizione contraria ed entrare in una nuova posizione che segua il mercato.

Le strategie di trading trattate in questo paragrafo differiscono da quelle trattate nei paragrafi precedenti in quanto utilizzano il segnale generato dalla barriera di quantile come segnale di entrata nel mercato e di attivazione della strategia di trading veloce. Una volta entrati si cerca di seguire il mercato per sfruttare istantaneamente sia i trend al rialzo sia quelli al ribasso.

Va fatto notare l'importanza del segnale iniziale dato dal superamento delle barriere di quantile in quanto vengono selezionati solo quelle realizzazioni dell'ensemble che presentano maggiori livelli di autocorrelazione e volatilità, condizioni necessarie alla buona riuscita delle strategie di trading.

4.4.1 Strategia Trading attivo con densità di correlazione

Questa strategia lavora su una finestra temporale che va dal minuto al tempo massimo di 15 minuti e permette di riconoscere le zone in cui vi è una densità di autocorrelazione elevata, come segnale di inizio di un trade.

Prima di una formalizzazione, si procede con un esempio: si ipotizzi una finestra di 15 minuti, con campionamento ad 1 minuto, che generi i seguenti rendimenti, di cui si riporta nelle stringhe che seguono solamente il segno positivo o negativo:

- 1) (+ + - + + + + - - - + + + +)
- 2) (+ - + + - + - + + - + + - + +)

Come risulta chiaro, le due stringhe hanno stesso numero di segnali positivi e negativi, ma una differisce dall'altra in quanto manifesta una maggiore autocorrelazione.

Per quantificare l'autocorrelazione vengono definite le funzioni:

$$E(r_1(t)) = \sum_{i=1}^T 1 * segno[r_1(t_i)] * segno[r_1(t_{i+1})]$$

$$F(r_1^+(t)) = \sum_{i=1}^T 1 * segno[r_1^+(t_i)]$$

$$G(r_1^-(t)) = \sum_{i=1}^T 1 * segno[r_1^-(t_i)] \quad 4.5$$

Dove con $r_1^+(t)$ e $r_1^-(t)$ si intendono i ritorni condizionati al fatto di essere di segno positivo o negativo e T rappresenta il numero di ritorni che vengono presi in considerazione per formulare la strategia.

In base alla finestra temporale T scelta (in questo caso T=15), la funzione E assume valori compresi fra [-T ; +T]. L'estremo inferiore corrisponde alla condizione di minima autocorrelazione, l'estremo superiore alla condizione di massima autocorrelazione.

L'idea della strategia consiste nel riconoscere regioni ad alta autocorrelazione, sempre operando all'interno della finestra temporale stabilita dalle analisi precedenti sulla funzione di autocorrelazione e sul suo decadimento, e al raggiungimento delle condizioni sotto elencate si genera il segnale operativo di acquisto-vendita.

Se non vi sono posizioni aperte:

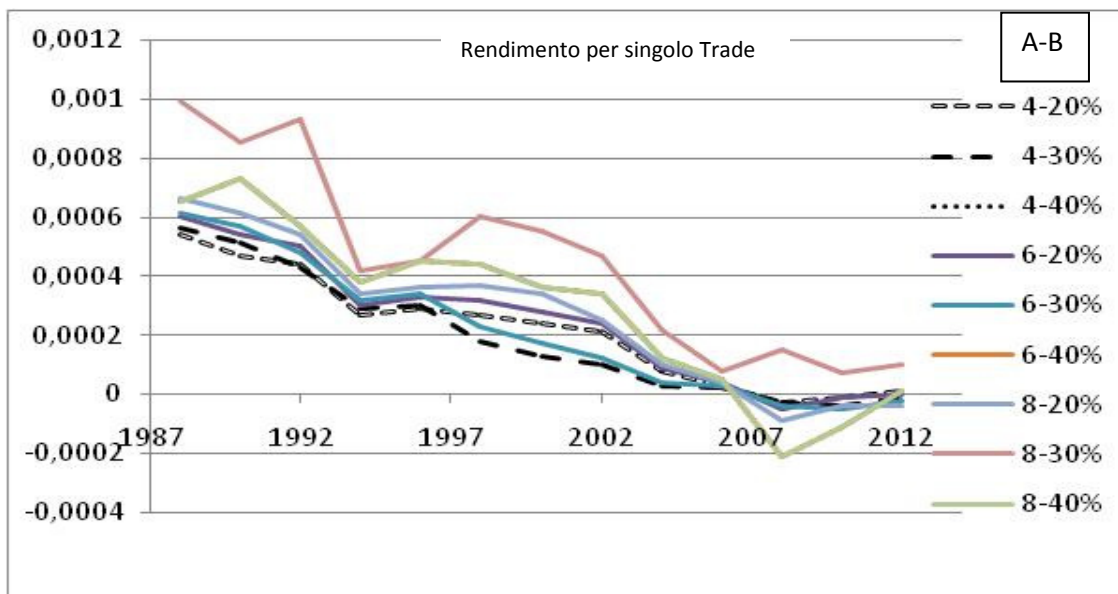
Si acquista se	$E_R(t) > A$ e $F_R(t) > G_R(t) * (1 + B)$
Si vende se	$E_R(t) > A$ e $G_R(t) > F_R(t) * (1 + B)$

A fine giornata, se vi sono ancora posizioni aperte, si chiudono.

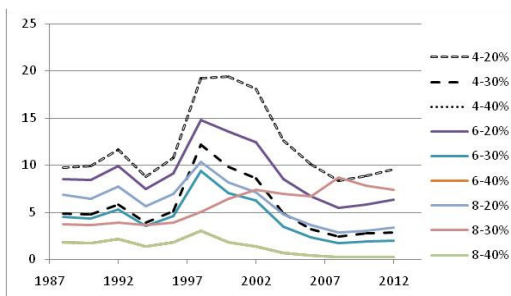
I parametri A e B identificano la tipologia di strategia, se più aggressiva o più cauta.

Si è impostato una finestra temporale di 10 minuti, con E che possa variare fra [-10 ; +10] ed un valore di A di [4;6;8] , ovvero selezionando regioni ad alta autocorrelazione, e valori di B che vadano dal 20% al 40%.

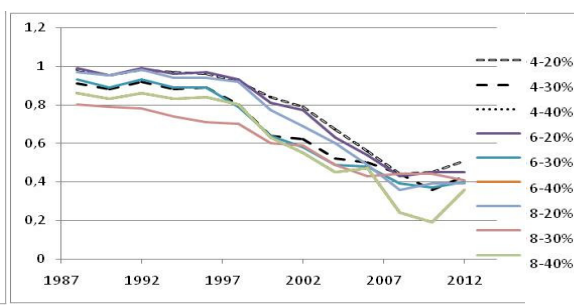
Di seguito i risultati:



Numero di trade medi a giornata borsistica



Frazione di giornate concluse con rendimento positivo



Commento alla strategia

Questa strategia che trae la sua ragione teorica oltre che su idee di scaling anche sullo sfruttamento immediato delle di autocorrelazioni, permette rispetto alle altre strategie di aumentare di molto il numero di trade per singola giornata borsistica lasciando pressoché invariato rispetto alle altre strategie il rendimento per singolo trade. Ne deriva un considerevole aumento del profitto totale per giornata borsistica, nel caso in cui soprattutto i costi di transazione non siano troppo elevati. Come si può notare la strategia non è molto sensibile, salvo alcuni casi alla scelta di A e B, che in quanto parametri esterni dati in base all'approccio aggressivo o meno risultano abbastanza trascurabili, quantomeno nel range suggerito.

Conclusioni

In questo lavoro di tesi si è voluto portare un contributo alla analisi e alla modellizzazione delle serie finanziarie. Per fare ciò, è stata utilizzata la serie dell' S&P 500 che va dal 1985 al 2013 campionata alla frequenza di 1 minuto. Ci siamo concentrati sul problema di ricercare un corretto approccio di analisi dei dati, creare un modello quanto più vicino alle evidenze empiriche, definire e caratterizzare l'idea di tempo in finanza e formulare delle strategie di trading che si basino sulla modellizzazione derivante dalla simmetria dello scaling e che diano una misura di alcune limitazioni di questo modello .

Si è partiti nel Capitolo 1 dall'analisi ad alta frequenza di fatti empirici,

- il comportamento periodico della volatilità nell'arco dell'intraday, sintomo si una assenza di stazionarietà per scale di tempo inferiori al giorno,
- le correlazioni lineari semplici dei ritorni molto minori delle correlazioni dei ritorni in valore assoluto, e la loro evoluzione nel tempo,
- la non gaussianità delle PFD dei ritorni,
- l'evidente comportamento di scaling.

In questo modo sono stati evidenziati dei fatti stilizzati ai quali aderire nella formulazione del modello matematico.

Per quanto riguarda la modellizzazione delle serie finanziarie nella finestra temporale dell'intraday si è proceduto con un approccio di ensemble. Studiando infatti le serie storiche si è notato che queste presentano un pattern caratteristico di evoluzione della volatilità durante l'intera giornata di contrattazioni, che poi si ripete pressochè uguale i giorni seguenti. Proprio la non stazionarietà dell'intraday associata alla periodicità appena accennata, getta le basi per una nuova formulazione della stazionarietà interday, rendendo giustificato il fatto di considerare le singole giornate come realizzazioni diverse dello stesso fenomeno, e di conseguenza rendere possibile applicare l'approccio di ensemble.

Nel capitolo 2 è stato evidenziato inoltre un robusto comportamento di scaling. Si è passati a definire, in accordo con le evidenze empiriche un modello che fosse coerente con la simmetria di scaling nella sua versione in omogenea nel tempo. Questo modello fornisce la base matematica su cui costruire l'idea di tempo in finanza.

Nel capitolo 3 si è affrontato il tema di cosa sia il tempo in finanza, provando a fornire una possibile interpretazione di tempo e proponendo λ_τ come misuratore del trascorrere del tempo in finanza, ottenendo così una grandezza che facesse da ponte tra la finestra intraday e la finestra interday. Infatti il tempo in finanza è stato uno dei maggiori obiettivi di indagine di questo lavoro, sia perché rappresenta un campo di forte interesse per un fisico, sia perché non aveva ancora trovato una formalizzazione

sostanzialmente diversa dal cosiddetto tempo fisico e dalla definizione di tempo in ragione del volume delle transazioni che progressivamente si realizzano. L'idea di partenza è stata di considerare il tempo nell'ottica dello scaling. Sono state considerate le PDF empiriche di ritorni aggregati su intervalli temporali di diversa durata, dai minuti alle settimane, come fotografia del trascorrere del tempo fisico. Grazie allo scaling le PDF empiriche dei ritorni sono state fatte tutte collassare nella medesima funzione al variare del parametro di scala λ_τ , sia nella finestra intraday, dove vale l'ensemble, sia nella finestra interday, dove si suppone stazionarietà ed ergodicità. Riuscire a riunire sotto un'unica trattazione queste due scale di tempo avvalora l'universalità del processo. Operativamente sono stati ricavati tutti i parametri di scala per i ritorni aggregati con intervalli di tempo dal minuto alle due settimane e da questo si è ricavato una funzione di tempo che definisse operativamente il tempo in finanza.

Nel capitolo 4 si è proceduto alla formulazione di strategie operative di trading, volte soprattutto a verificare la applicabilità del modello intraday ad un mercato caratterizzato da correlazioni lineari non strettamente nulle. Si è poi proceduto ad una definizione operativa di trading che permettesse una massimizzazione dei ricavi a partire da quanto trovato sullo scaling, e dalla evidenza di autocorrelazione non strettamente nulla dei ritorni. Si è considerato sia il ricavo totale che il ricavo normalizzato per numero di operazioni, essendo in questo senso importante nell'operatività reale la considerazione dei costi di transazione. I risultati ottenuti dai modelli sono fortemente dipendenti dalla funzione di correlazione dei ritorni del particolare periodo storico preso in esame: è evidente una progressiva diminuzione delle correlazioni ed una progressiva minor efficacia dei modelli di trading. Questo è indice che nel tempo il mercato tende a ridurre sempre di più le possibilità di arbitraggio, diventando più efficiente.

Appendice A Il test di Kolmogorov-Smirnov

Il test di Kolmogorov-Smirnov rappresenta un utile strumento per determinare se la distribuzione cumulativa empirica ottenuta da una serie di dati e la distribuzione cumulativa teorica siano compatibili.

Sia $P_{>,th}(x)$ la distribuzione cumulativa teorica che si vuole testare e $P_{>,e}(x)$ la distribuzione cumulativa empirica. Viene definita la distanza D fra la distribuzione cumulativa teorica e la distribuzione cumulativa empirica come il valore massimo delle loro differenze prese in valore assoluto.

$$[A1] \quad D = \max_x |P_{>,th}(x) - P_{>,e}(x)|$$

D è una quantità random che dipende dalla serie x_k , e la sua distribuzione è conosciuta quando si ha un numero di dati caratterizzanti la distribuzione di probabilità cumulativa sufficientemente elevato, e, fatto importante, è indipendente dalla particolare forma di $P_{>,th}(x)$.

La distribuzione cumulativa di D è data da:

$$[A2] \quad P_{>}(D) = Q_{KS}(\sqrt{N}D) \quad Q_{KS}(u) = 2 \sum_{n=1}^{\infty} (-1)^{n-1} \exp(-2n^2u^2)$$

Questa formula significa che se $P_{>,th}(x)$ rappresenta la corretta distribuzione, allora la distanza fra la distribuzione cumulativa empirica e la distribuzione teorica dovrebbe decadere per grandi N come $\frac{u}{\sqrt{N}}$.

Dunque si procede come segue: si ricava dai dati la quantità D dalla quale si deduce, utilizzando l'equazione sopra, la quantità $P_{>}(D)$. Se si trova un numero piccolo, ad esempio 0.01, allora l'ipotesi che le due distribuzioni siano la stessa è molto improbabile. L'ipotesi può essere rigettata ad un livello di confidenza del 99%.

Il test di Kolmogorov-Smirnov perché fornisce una risposta universale indipendentemente dalla forma della distribuzione presa in esame.

Bibliografia

- [1] L. Bachelier, “Theorie de la speculation” , Annales Scientifiques de l’Ecole Normale Superieure, 1900
- [2] B. Mandelbrot, “The variation of certain speculative prices” Journal of Businnes, 1963
- [3] Mantegna R. and Stanley H:E. , Introdution to Econofisics. Correlation and complessity in finance, Cambridge University press. Cambridge.
- [4] R. Cont, *Empirical properties of asset returns: stylized facts and statistical issue* , Quantitative finance, pp 223-236, 2001
- [5] F. Baldovin, D. Bovina, F. Camana, A. Stella; *Modeling the non-markovian,non stationary scaling dynamics of financial markets;* Econophysics of order-driven markets, pp239-252, 2011
- [6] K.Bassler, J.McCauley, G.Gunaratne, *Nonstationary increments, scaling distributions, and variable diffusion processes in financial markets;* PNAS pp.17287-17290, 2007
- [7] L. Kadanoff, Statistical Physics, Statics, Dynamics and Renormalization. Singapore: World Scientific, 2005
- [8] F. Baldovin, F. Camana, M.Caporin, M.Caraglio, A. Stella; *Ensemble properties of high frequency data and intraday rules;* Quantitative Finance, 2014
- [9] J.Voit , *The Statistical Mechanics of Financial Markets*, Springer, 20057
- [10] Jean-Philippe Bouchaud, Marc Potters; *Theory of Financial Risk and Derivative Pricing;* Cambridge University press
- [11] B. Gnedenko, A.Kolmogorov, *Limit distribution for sums of indipendent randon variables.* Reading: Addison Wesley, 1954
- [12] B. Mandelbrot e J. Van Nesse, *Fractional brownian motions, fractional noises and application,* The SIAM Review, pp 422-437, 19687
- [13] P.Levy, *Calcul des probabilités.* Paris: Gauthiers-Villars 1925
- [14] F.Baldovin e A. Stella; *Anomalous scaling due to correlations. Limit teorems and self similar processes;* Journal of Statistical Mechanics, 2010

- [15] Marco Zamparo, Fulvio Baldovin, Michele Caraglio, Attilio Stella;
Scaling simmetry, renormalization, and time series modeling: the case of financial asset dynamics. Physical Review, 2013
- [16] I. Schoenberg, “Metric spaces and completely monotone functions”,
Annales of Mathematics, n. 30, 1938
- [17] F.Black e M.Scholes; *The price of options and corporate liabilities*;
Journal of political Economy, pp 637- 654
- [18] Giovanni L. Vasconcelos; *A guided Walk down wall street: an introduction to Econophysics*; 2004