# INDICE

**Capitolo primo:**
**Introduzione..... 3**

**Capitolo secondo:**
**Dati ad alta frequenza e microstruttura del mercato..... 9**
2.1 Dati ad alta frequenza..... 9
2.2 Modalità di raccolta e gestione dei dati..... 11
2.3 Microstruttura. I diversi meccanismi di scambio..... 15
2.3.1 Mercato limit order..... 16
2.3.2 Mercati floor..... 18
2.3.3 Mercato dei dealers..... 19
2.3.4 Aste e altri meccanismi di compensazione..... 21
2.3.5 Contrattazione..... 22
2.4 Errori di Microstruttura..... 23
2.4.1 Scambi non sincronizzati..... 24
2.4.2 Bid-Ask Spread..... 25
2.4.3 Discretezza dei prezzi..... 26

**Capitolo Terzo:**
**Varianza integrata e suoi stimatori..... 29**
3.1 Volatilità integrata e variazione quadratica..... 29
3.2 Stima della volatilità integrata..... 31
3.2.1 Varianza realizzata..... 32
3.2.2 Stimatore di Fourier..... 34
3.3 Gli effetti dell'errore..... 37
3.4 Stimatori alternativi..... 41
3.4.1 Sottocampionamento..... 41
3.4.2 Correzione per l'autocorrelazione..... 48
3.5 Effetti degli schemi di campionamento..... 51
3.6 Confronto tra le diverse tecniche..... 53

**Capitolo Quarto:**
**Simulazioni..... 55**
4.1 Disegno sperimentale..... 55
4.2 Implementazione..... 58
4.3 Risultati..... 60

**Capitolo Quinto:**
**Applicazioni su dati reali..... 77**
5.1 I dati..... 77
5.2 Analisi descrittiva..... 79
5.3 Stima della volatilità..... 81

**Capitolo Sesto:**
**Conclusioni..... 83**

**Bibliografia..... 85**

# CAPITOLO PRIMO

# INTRODUZIONE

L'analisi e la misura della volatilità hanno attirato un grande interesse negli ultimi anni, soprattutto da quando è aumentata la disponibilità dei dati in nostro possesso. In presenza di dati ad alta frequenza uno dei problemi principali riguarda appunto la stima della volatilità dei rendimenti di un'attività finanziaria. Una prassi frequente consiste nello stimare la volatilità attraverso la somma dei rendimenti al quadrato. Questo approccio è giustificabile da un punto di vista teorico, come vedremo meglio nel capitolo 3, ma presenta dei problemi nelle applicazioni reali dovuti alla presenza di errori addebitati al modo di operare dei mercati finanziari (microstruttura).

Nel seguito di questo lavoro spiegherò come questo approccio porti a dei risultati non corretti e presenterò alcuni stimatori che cercano di ovviare agli inconvenienti dovuti all'errore di microstruttura.
Per fissare le idee, supponiamo che il processo generatore del logaritmo dei prezzi di un'attività finanziaria sia il seguente:

$$dp(t) = \mu(t)dt + \sigma(t)dW(t)$$

dove $W(t)$ è un moto Browniano standard (vedi capitolo 3), $\mu(t)$, noto come il coefficiente drift, e $\sigma(t)$, nota come la volatilità istantanea del processo, possono a loro volta seguire un processo stocastico a tempo continuo. La quantità di interesse è la variazione quadratica o volatilità integrata (IV), calcolata in un determinato periodo di tempo, tipicamente un giorno, definita da

$$IV = \int_0^t \sigma^2(t)dt.$$

Lo stimatore usuale di IV, come menzionato precedentemente, è quello basato sulla somma dei rendimenti al quadrato ed è chiamato volatilità realizzata (RV). Nel caso in cui si consideri un intervallo giornaliero, e si osservi il processo a tempi intragiornalieri $k_i$ per $i = 0, \dots n$, RV è data da

$$RV_t = \sum_{i=1}^{n} (p_i - p_{i-1})^2.$$

Si può far vedere che RV, all'aumentare della frequenza di campionamento converge in probabilità a IV e che quindi è uno stimatore consistente di IV, cioè

$$plim \sum_{i} (p_i - p_{i-1})^2 = \int_0^t \sigma^2(t)dt.$$

Jacod e Protter (1998) sono riusciti a determinare anche la distribuzione asintotica di RV. Questi risultati indicano quindi che è meglio scegliere la frequenza di campionamento più grande possibile. In questo modo, però, si incorre in un errore causato dalla microstruttura del mercato e ci si trova ad affrontare un *trade-off* tra accuratezza, ottenuta attraverso un'alta frequenza di campionamento, ed errore di microstruttura. Quindi la volatilità realizzata può risultare uno stimatore non robusto di IV in presenza di errore di microstruttura.

Un approccio diverso alla stima della volatilità integrata è stato recentemente introdotto da Malliavin e Mancino (2002), i quali utilizzano la trasformata di Fourier per costruire un nuovo stimatore chiamato appunto stimatore di Fourier, definito da

$$F = \frac{\pi^2}{S} \sum_{s=1}^{S} \left( \hat{a}_s^2(dp) + \hat{b}_s^2(dp) \right)$$

dove $\hat{a}_s$ e $\hat{b}_s$ sono le stime dei coefficienti di Fourier, calcolati a partire dai prezzi dopo averli normalizzati nell'intervallo $[0, 2\pi]$ e $S = n/2$ dove $n$ è il numero di osservazioni intragiornaliere.

Tale stimatore invece risulta robusto rispetto alla presenza di errore dovuto alla microstruttura che tipicamente si manifesta alle alte frequenze di campionamento.

Ribadiamo nuovamente che in presenza di errore di microstruttura RV calcolata sui rendimenti osservati è uno stimatore non soddisfacente.
Solitamente, il meccanismo utilizzato per superare il problema indotto dalla microstruttura del mercato, è quello di non considerare una grande quantità di dati disponibili campionando meno frequentemente. Statisticamente però questo non è il modo più corretto di procedere infatti è risaputo che più dati si hanno a disposizione e più accurata è la stima. Inoltre campionare su un orizzonte temporale più lungo, riduce solo l'impatto della microstruttura procurando una perdita di informazione che potrebbe essere rilevante per l'analisi. Quindi una buona soluzione sarebbe quella di cercare uno stimatore soddisfacente utilizzando tutti i dati a disposizione. In questa direzione Zhang *et al.* (2005) hanno proposto uno stimatore della volatilità realizzata $\left(RV_t^{(TTSE)}\right)$ che combina due quantità basate su due scale temporali diverse :

- $RV_t^{(avg)}$, la media delle volatilità realizzate calcolate su una scala temporale "lenta", ad esempio considerando i prezzi ogni 5 o 15 minuti
- $RV_t$, la volatilità calcolata usando tutti i dati a disposizione (scala temporale "veloce"), opportunamente pesata in base ad un fattore che dipende dal numero delle osservazioni

$$RV_t^{(TTSE)} = RV_t^{(avg)} - \frac{\bar{n}}{n} RV_t$$

Soluzioni alternative erano già state introdotte precedentemente da Zhou (1996) che, nel caso di $\sigma$ costante, ha proposto un approccio che corregge lo stimatore della volatilità tenendo conto della struttura di dipendenza dei rendimenti. Un ulteriore contributo in questa direzione viene fornito da Hansen e Lunde (2006), che hanno studiato lo stimatore di Zhou, estendendolo al caso di $\sigma$ variabile nel tempo. Lo stimatore di Zhou e le sue

estensioni sono tuttavia non consistenti. Questo significa che quando la frequenza di campionamento aumenta lo stimatore non converge alla quantità IV. Recentemente però Barndorff-Nielsen *et al.* (2006) hanno proposto uno stimatore consistente basato sempre sulla correzione per l'autocorrelazione, definita da

$$RV_{t,q}^{(BHLS)} = RV_t + \sum_{q=1}^{Q} k\left(\frac{q-1}{Q}\right)\left(\gamma_q + \gamma_{-q}\right)$$

dove $k$(x) per x$\in$[0,1] è una funzione Kernel tale che $k$(0)=1 e $k$(1)=0, $\gamma_q$ è la funzione di autocovarianza al ritardo q.

In questa tesi ho studiato alcuni stimatori della volatilità integrata, utilizzando diversi meccanismi generatori dei dati, differenti schemi di campionamento e vari errori nella microstruttura del mercato.
Il lavoro è stato organizzato come segue.
Nel capitolo 2 parlerò dei cosiddetti dati ad alta frequenza e dei meccanismi della microstruttura del mercato spiegando i motivi che inducono a registrare errore. Nel capitolo 3 introduco alcuni stimatori della volatilità integrata. Particolare attenzione viene dedicata alle loro proprietà anche in presenza di errore dovuto alla microstruttura del mercato sia questo IID o presenti una qualche forma di dipendenza. Nel capitolo 4 presento i risultati di una simulazione Monte Carlo per la verifica delle proprietà di alcuni degli stimatori trattati attraverso un implementazione con il programma Ox. Ho utilizzato diversi schemi di campionamento, 1,5 e 15 minuti, e due modelli per la volatilità uno che tiene conto della memoria lunga e l'altro della presenza di salti (*jumps*). I risultati vengono proposti in termini di errore quadratico medio e *bias*. Infine il capitolo 5

offre una breve illustrazione empirica con dati reali del mercato borsistico italiano. Il capitolo 6 conclude.

# CAPITOLO SECONDO

# DATI AD ALTA FREQUENZA E MICROSTRUTTURA DEL MERCATO

## 2.1. I DATI AD ALTA FREQUENZA

Una delle conquiste più importanti negli ultimi anni consiste nello studio dei cosiddetti dati ad alta frequenza, ovvero osservazioni registrate in tempo reale sui mercati. A differenza degli studi tradizionali, nei quali si tende a considerare dati misurati ad intervalli equispaziati, nel caso dell'alta frequenza si registra ogni singola transazione che avviene sul mercato. La possibilità di sfruttare questa enorme mole di informazione costituisce un indubbio vantaggio in quanto si possono ricavare stime della volatilità più

accurate. Tali stime si riflettono su gran parte della moderna teoria finanziaria: dal prezzaggio di opzioni alla previsione della volatilità intragiornaliera, dal calcolo del *value at risk* alla gestione della liquidità.

Se fino a poco tempo fa le applicazioni si limitavano a considerare al massimo osservazioni giornaliere, si sta diffondendo negli ultimi anni la pratica di sfruttare la massima informazione possibile utilizzando i dati *tick-by-tick*. Ogni *tick* è un'unità logica di informazione, come una quotazione (cioè proposte di vendita e di acquisto che non vanno a buon fine) o un prezzo di una transazione; e ogni giorno d'apertura del mercato genera centinaia e migliaia di *ticks*.

Utilizzando i dati giornalieri, si suole considerare il prezzo di chiusura come indicativo dell'intera giornata di contrattazioni; ma ciò non è necessariamente vero, infatti, tralasciando i movimenti intragiornalieri del prezzo, si rischia di trascurare informazioni rilevanti.

Tuttavia utilizzare questo tipo di dati pone alcuni problemi. In primo luogo notiamo che le osservazioni così rilevate non necessariamente sono equispaziate: tra un osservazione e l'altra può intercorrere un lasso di tempo che oscilla da pochissimi secondi a qualche minuto. Al contrario è immediatamente evidente che, utilizzando dati giornalieri, tale problema non si pone.

In questo capitolo mi occupo quindi di alcuni degli aspetti dei dati ad alta frequenza. Per fare questo cercherò dapprima di spiegarne la loro gestione per poi passare in rassegna i diversi meccanismi che generano gli scambi ed infine illustrare la formazione degli errori di microstruttura.

## 2.2 MODALITA' DI RACCOLTA E GESTIONE DEI DATI

I dati ad altissima frequenza sono raccolti e forniti da organismi di vario tipo, spesso identificabili con l'istituzione che gestisce e regola il singolo mercato nazionale; nel caso italiano, essi sono gestiti direttamente dalla Borsa italiana S.p.A. Tali dati si presentano sotto forma di database in formato ASCII che riporta le seguenti informazioni:

- Data della transazione
- Orario della transazione
- Prezzo in euro al quale avviene lo scambio
- Volume scambiato
- Controvalore in euro

Una tipica giornata di contrattazioni sul mercato italiano supera abbondantemente le 1500 osservazioni. Si capisce subito come ciò risulti costoso sia dal punto di vista dello stoccaggio che da quello del trattamento dei dati: un database di un anno consterebbe di circa 600.000 rilevazioni.
Se il problema del trattamento dei dati si avverte già sul mercato italiano, su quello statunitense è molto più grande: non solo le transazioni sono molto più numerose, ma l'insieme delle informazioni rilevate risulta molto più ampio. Il principale fornitore di dati per il mercato statunitense è il New York Stock Exchange (NYSE) stesso, che cura il database TAQ (*Trades and Quotes*). Tale database riporta, oltre alle transazioni che effettivamente hanno luogo (*trades*), anche le quotazioni *bid* e *ask*, cioè le proposte di vendita e di acquisto che non vanno a buon fine (quotes). Nel database TAQ, oltre ai dati relativi al NYSE, sono rilevati anche dati di altri mercati, tra cui il *National Association of Securities Dealers Automated Quotations*

(NASDAQ), *l'American Stock Exchange* (AMEX) e il *Chicago Board Options Exchange* (CBOE).

La modellistica tradizionale che si è soliti utilizzare per l'analisi delle serie storiche finanziarie è stata sviluppata quando ancora le osservazioni giornaliere rappresentavano il massimo grado di informazione che si potesse riuscire a gestire. Tale modellistica è quindi applicabile solo nel caso in cui i dati siano equispaziati e, come abbiamo visto, questo non è il caso delle osservazioni ad alta frequenza. I primi studi in tale ambito si sono concentrati sul cercare di recuperare la modellistica tradizionale, trasformando le osservazioni ad alta frequenza in osservazioni equispaziate. L'idea più naturale in questo senso è quella di registrare i dati ad intervalli di tempo predefiniti; ad esempio ogni minuto o ogni secondo. Questo approccio comporta però alcuni problemi: se infatti si vogliono rilevare osservazioni ogni minuto può succedere che non ci siano contrattazioni esattamente ogni minuto nel mercato. Per esempio, si può registrare la prima transazione dopo esattamente un minuto dall'apertura del mercato, la seconda ad un minuto e 48 secondi (1,8 minuti) e la terza dopo 2 minuti e 36 secondi (2,6 minuti).

Vi sono varie modalità di intervento possibili.

Una prima idea potrebbe essere quella di utilizzare la transazione più vicina, *previous tick*; nell'esempio menzionato poco sopra, per l'osservazione corrispondente al termine dei 2 minuti, si sarebbe dovuta prendere la seconda transazione.

Altrimenti, si può considerare la media aritmetica delle osservazioni immediatamente precedente e seguente; quindi si sarebbe dovuto prendere la media della seconda e la terza transazione come osservazione dei 2 minuti.

Un'ulteriore possibilità è quella di considerare una media ponderata con gli intervalli di tempo che intercorrono tra le transazioni e il termine del

minuto considerato. Nel caso suddetto per il termine dei 2 minuti si dovrebbe considerare:

$$x_2 = \frac{[102 \cdot (2 - 1{,}8) + 104 \cdot (2{,}6 - 2)]}{[(2 - 1{,}8) + (2{,}6 - 2)]}$$

Menzioniamo infine il metodo dell'interpolazione lineare, consistente nel tracciare una retta per le due transazioni più prossime a quella mancante e prendere come valore quello appartenete a tale retta in corrispondenza dell'istante considerato. In pratica ciò si traduce nel risolvere la proporzione:

$$(x_2 - 102) : (2 - 1{,}8) = (104 - 102) : (2{,}6 - 1{,}8)$$

E' da notare che la media ponderata è la metodologia che considera il numero più ampio di fattori, ma non è necessariamente superiore alle altre in termini di risultato. Quello che vale la pena osservare è che procedere ad accorpare i dati per renderli equispaziati comporta comunque, per quanto alta possa essere la frequenza di campionamento, una perdita di informazione. Abbiamo già rilevato che, disponendo di dati equispaziati, possiamo sfruttare la modellistica preesistente. Tuttavia, se si aggiunge a questo indubbio vantaggio l'inconveniente relativo alla perdita di informazione, questa strategia può rivelarsi non ottimale. Una delle possibilità che sta riscuotendo maggior successo in letteratura consiste nel modellare i tempi che intercorrono tra una transazione e l'altra (le cosiddette *durations*) tramite un nuovo processo introdotto da Engle e Russel (1998) e noto con l'acronimo di modello ADC (*Autoregressive Conditional Duration*).

Si è descritto come rendere i dati equispaziati, ma questo non è l'unico schema di campionamento che viene utilizzato nella pratica.
Alcuni autori ( vedi ad esempio, Oomen (2005) Hansen e Lunde (2006) McAleer e Medeiros (2007)) nei loro studi sulla volatilità integrata, hanno infatti utilizzato diversi schemi di campionamento per la costruzione dei rendimenti intragiornalieri quali *calendar time sampling* (CTS), *transaction time sampling* (TrTS), *business time sampling* (BTS) e *tick time sampling* (TkTS).
Supponiamo che in un dato giorno t suddividiamo l'intervallo $[0,1]$ in $n$ sottointervalli e definiamo l'insieme $\Lambda = \{k_0, \ldots\ldots, k_n\}$ dove $0 = k_0 < k_1 < \ldots . k_n = 1$. La lunghezza del i-esimo sottointervallo è data da

$$\delta_{i,n} = k_i - k_{i-1}$$

tale lunghezza tende a zero quando il numero di osservazioni $n$ aumenta.

- CTS si usa quando gli intervalli sono equispaziati nel tempo di calendario cioè $\delta_{i,n} = 1/n$ per ogni $i$ e richiede la costruzione di un prezzo artificiale a partire dal prezzo "grezzo" utilizzando il metodo *previous tick* o il metodo dell'interpolazione temporale, sopra menzionati. Quest'ultimo però, quando si campiona ad altissime frequenze, presenta uno spiacevole inconveniente: infatti quando $n$ tende ad infinito la varianza realizzata tende in probabilità a zero. Sostanzialmente ciò significa che la variazione quadratica di una retta è zero. Benché questo sia un risultato limite, esso suggerisce che tale metodo non è adatto per la costruzione dei rendimenti intragiornalieri ad altissime frequenze.
- Quando i prezzi vengono registrati ad ogni m-esima transazione si utilizza TrTS. Ad esempio, se m=5, significa scegliere i $k_i$ ogni cinque transazioni.

- Se i tempi di campionamento $k_0, k_1, \dots \dots, k_n$ sono tali per cui $IV_{i,t} = IV_t/n$ siamo nel caso di BTS (dove $IV_{i,t}$ è la varianza integrata dell'intervallo $i$ nel giorno $t$ e $IV_t$ è la varianza integrata del giorno $t$, quantità che verranno analizzate nel capitolo successivo).
- L'ultimo schema TkTS si utilizza quando i prezzi sono registrati ad ogni loro cambiamento.

Mentre i $k_i$ sono osservabili nei casi CTS, TrTS e TkTS, nel terzo caso sono latenti poiché sono definiti dalla volatilità che non è osservabile. Risultati empirici suggeriscono che BTS può essere approssimato da TrTS. Gli effetti dei diversi schemi di campionamento sulla stima della volatilità integrata saranno discussi nel capitolo sucessivo.

## 2.3 MICROSTRUTTURA. I DIVERSI MECCANISMI DI SCAMBIO (TRADING)

Abbiamo visto nella sezione precedente come si formano e si gestiscono i dati ad alta frequenza, più tale frequenza è alta e più esiste la possibilità di incorrere negli errori indotti dai meccanismi di scambio, la cosiddetta microstruttura del mercato. Infatti ci si trova davanti ad un *trade-off* tra accuratezza, che è teoricamente ottimizzata utilizzando la più alta frequenza possibile, ed errore di microstruttura che si presenta attraverso *bid-ask spread*, scambi non sincronizzati, assenza di scambi e discretezza dei prezzi.

Prima di tutto bisogna quindi capire cosa si intende per microstruttura del mercato, ossia quel ramo dell'economia e della finanza che si occupa di come avvengono gli scambi nel mercato.
O'Hara (1995) definisce la microstruttura del mercato come "lo studio del processo e dei risultati dei beni scambiati sotto uno specifico insieme di regole. Mentre la maggior parte dell'economia esclude i meccanismi di scambio, la teoria della microstruttura si focalizza su come specifici meccanismi di trading influenzano il processo di formazione dei prezzi."

Qualunque sia il meccanismo utilizzato, l'evento che noi chiamiamo *trading*, inteso come esecuzione o soddisfazione di un ordine, costituisce solo un accordo preliminare. Questo accordo avvia le procedure di compensazione e pagamento che alla fine si traducono nel trasferimento di titoli e fondi. I processi sono di solito automatici e di *routine* e gli agenti del mercato raramente si preoccupano dei dettagli. E' importante, comunque, che essi richiedano una qualche sorta di relazione pre-esistente tra le parti. Stabilire un accordo di compensazione è sia senza costi che istantaneo e potrebbe creare una barriera all'entrata per i potenziali compratori o venditori non precedentemente conosciuti nel mercato.
Di seguito vengono discussi i principali meccanismi di scambio.

### 2.3.1 Mercato limit order

La maggior parte dei mercati continui dei titoli hanno almeno un libro elettronico dei limit order. Un *limit order* è un ordine che specifica una direzione, quantità e prezzo, e pone un limite nel farlo, ossia si può comprare a non più (o vendere a non meno) di uno specifico prezzo.

In un tale mercato gli ordini arrivano casualmente nel tempo e il loro prezzo viene paragonato con quello degli ordini già presenti, in modo da verificare se c'è un *matching*. Se questo si verifica, allora lo scambio avviene al prezzo stabilito dal primo ordine. L'insieme dei *limit orders* presenti nel sistema che non vengono eseguiti costituiranno il libro. Visto che i *limit orders* possono essere cancellati o modificati in qualsiasi momento, il libro viene detto dinamico. Questo mercato è solitamente trasparente, infatti il libro è visibile a tutti i partecipanti. Lo straordinario livello di trasparenza di cui attualmente godono gli agenti del mercato (traders) è un fenomeno recente: le regole del NYSE storicamente proibivano la rivelazione del libro il quale non è stato disponibile fino a Gennaio 2002.

Un mercato potrebbe avere più di un libro, ciascuno guidato da un diverso *broker* o un'altra entità. Un *broker* è un intermediario finanziario che ricerca e acquista, per conto del cliente, nel mercato di riferimento, il prodotto che offre il miglior rapporto qualità-prezzo.

*Limit order books* potrebbero anche essere usati in congiunzione con altri meccanismi. Quando tutti gli scambi di un azione si svolgono attraverso un singolo libro il mercato si dice CLOB, *consolidated limit order book.* Il CLOB è utilizzato per gli *stocks* scambiati nella maggior parte dei mercati asiatici ed europei.

C'è un meccanismo di regole sulla priorità che governa la sequenza in cui gli ordini sono eseguiti. La prima priorità è il prezzo, cioè un ordine a vendere prezzato a 100 sarà eseguito prima di un ordine prezzato a 99. Il tempo di solito è la seconda priorità, a parità di prezzo gli ordini sono eseguiti in ordine di arrivo nel mercato. Benché queste regole di priorità possano sembrare ovvie raramente vengono applicate.

Un *trader* potrebbe desiderare che un ordine venga eseguito "al mercato", cioè al miglior prezzo disponibile. Se la quantità dell'ordine è più grande

della quantità disponibile al miglior prezzo del libro, si dice che l'ordine "camminerà nel libro", raggiungendo esecuzioni parziali a prezzi progressivamente peggiori fino a che l'ordine sarà completamente soddisfatto. Ciò potrebbe condurre ad esecuzioni a prezzi assai peggiori di quello che il *trader* aveva pensato al tempo della richiesta. Fatti di questo tipo possono essere evitati. Nel sistema Euronext non è permesso che un ordine di mercato "cammini nel libro"; esso dovrà essere eseguito per la quantità assegnata al miglior prezzo disponibile e la rimanenza aggiunta al libro con tale prezzo.
In un tale mercato compratori e venditori interagiscono direttamente, usando principalmente i *brokers* come tramite per i loro ordini.

### 2.3.2 Mercati floor

Il consolidamento degli interessi di scambio è importante in quanto accresce la probabilità che le controparti si incontrino una con l'altra. Prima che il mercato elettronico permettesse la virtualizzazione degli scambi tale consolidamento poteva avvenire solo fisicamente. In un mercato *floor*, i numerosi compratori e venditori sono rappresentati da un numero molto piccolo di *brokers* che negoziano e concludono accordi bilaterali *face-to-face*. Questi *brokers* sono spesso chiamati membri e i mercati sono storicamente organizzati come cooperative. I membri agiscono sia come agenti, rappresentando gli ordini dei clienti agli altri membri, sia come principali, ponendosi dall'altro lato degli ordini dei clienti. La combinazione di queste due funzioni, quindi, soffre di un conflitto di interessi. Un *broker* che intende agire come controparte all'ordine del suo cliente non ha interesse nel rappresentare l'ordine agli altri nel mercato, poiché potrebbe offrire un miglior prezzo. Per questa

ragione il “commercio duale” è sia espressamente vietato che fortemente regolato.

Nonostante il comportamento dei *brokers* possa apparire caotico e rumoroso, il *floor trading* è solitamente un processo ordinato infatti:

- ✓ mediante i segnali di mano viene velocemente comunicata la caratteristica chiave dell’ordine;
- ✓ le azioni ingannevoli sono vietate, come le offerte ad un prezzo più basso delle offerte correnti degli altri membri con il tentativo di trovare un venditore disposto a commerciare ad un prezzo inferiore;
- ✓ i prezzi delle transazioni sono riportati velocemente e diffusi pubblicamente;
- ✓ controversie ed errori sono risolti velocemente.

I mercati *floor* hanno conosciuto il loro massimo splendore nel ventesimo secolo ma adesso hanno perso la loro importanza a causa della diffusione dei mercati elettronici.

### 2.3.3 Mercato dei dealers

Un *dealer* è un operatore di mercato che investe risorse proprie per la composizione di un proprio portafoglio. È visto come uno speculatore perché acquista strumenti finanziari quando sono sottovalutati dal mercato e li rivende quando vanno a rialzo.

Tra *dealer* e *broker* esistono due differenze importanti:

1. Prima tra tutte sicuramente è il concetto di rischio, che nel caso del *broker* ricade in capo al cliente che gli commissiona un determinato ordine, mentre nel caso del *dealer* la responsabilità ricade pienamente in capo all'intermediario stesso

2. Il ruolo del *broker* è poi anche per sua natura differente da quello del *dealer*: il *broker* si configura per lo più come un operatore in cui si limitano a confluire domanda ed offerta di strumenti finanziari, ma è di per sé un attore passivo nell'ambito delle contrattazioni in quanto non può eseguire proposte sue sul mercato ma si limita ad eseguire quelle dei clienti. Il *dealer* invece è visto più come uno "speculatore" poiché sulla base delle informazioni in suo possesso sfrutta opportunità offertegli dal mercato acquistando titoli quando sono sottovalutati e rivendendoli quando sono a rialzo.

Accade spesso comunque che un *broker* integri anche la figura del *dealer*.

Uno scambio in un mercato dei *dealers*, come il mercato FX (*Foreign Exchange*), tipicamente comincia con una chiamata del cliente al *dealer*. Il *dealer* quota il prezzo *bid* e il prezzo *ask*, al che il cliente potrebbe comprare al prezzo di domanda del *dealer*, vendere al prezzo di offerta del *dealer* o non fare niente. Questo presume che cliente e *dealer* abbiano una relazione pre-esistente. Tale relazione, oltre a stabilire il quadro per la compensazione ed il pagamento, gioca un ruolo più significativo, infatti il comportamento e la storia degli scambi del cliente potrebbero rivelare desideri o informazioni non espressi dal cliente e potrebbero quindi influenzare i termini dello scambio che il *dealer* offre.

La relazione *dealer*-cliente coinvolge la reputazione dimostrata e prolungata dalle ripetute interazioni. La reputazione del *dealer* dipende dalla sua buona volontà di quotare sempre un “ragionevole” prezzo *bid* e *ask*. La reputazione del cliente è basata sulla sua frequente accettazione dei termini di scambio proposti dal *dealer*.

Una vasta clientela potrebbe avere relazioni con molti *dealer*; questo forma le basi per la competizione che allevia il potere contrattuale del dealer. Piccoli clienti al dettaglio, però, spesso non hanno una tale quantità di relazioni e quindi hanno un basso potere contrattuale.

Il mercato dei *dealers* è di solito caratterizzato da un basso grado di trasparenza: i *dealers* forniscono quotazioni solo in risposta alle richieste dei clienti e queste non sono pubblicamente visibili.
Oltre alle interazioni *dealer*-cliente sono ugualmente importanti gli scambi tra *dealer*. Gli ordini in arrivo ad un dato *dealer* sono raramente bilanciati, poiché nella realtà vi è sempre un eccesso di domanda o di offerta. In tal caso, il dealer nell'accontentare i clienti potrebbe rimanere in una posizione indesiderata e quindi cercherà di vendere o comprare nel mercato *interdealer*.

### 2.3.4 Aste e altri meccanismi di compensazione

Quando ci sono molteplici compratori e venditori concentrati in un luogo nello stesso momento, lo scambio non ha bisogno di essere coordinato. Gli agenti si contatteranno l'un l'altro sequenzialmente concludendo accordi bilaterali. Possono presentarsi, però, risultati economicamente inefficienti. Un'altra considerazione da fare è che se l'affare è condotto dai *brokers* per conto dei clienti e i prezzi degli scambi sono riportati pubblicamente, i clienti vedrebbero i lori scambi eseguiti a prezzi peggiori del miglior prezzo disponibile nel mercato. Questo è improbabile se si vuole promuovere confidenza nel *broker* o nel meccanismo.
Un singolo prezzo permette di evitare il problema.
Questo è generalmente implementato con un asta bilaterale a prezzo singolo. Le curve di domanda ed offerta sono costruite dalla lista delle domande e delle offerte. Prezzi, quantità, e identità dei commercianti sono di solito determinati massimizzando il volume di commercio fattibile.
L'asta bilaterale è ampiamente usata nei mercati dei titoli. Per titoli con esiguo interesse commerciale la maggior parte degli scambi avviene usando

aste periodiche. Le aste bilaterali sono di solito usate per aprire la sessione di mercati continui (*Euronext, Tokio Stock Exchange, NYSE..*) ma di frequente vengono usate anche alla loro chiusura.
L'esperienza ci suggerisce che un aspetto particolarmente importante è la scadenza. Come si può notare da e-Bay, la maggior parte delle offerte avvengono poco prima della chiusura dell'asta. Per scoraggiare i clienti ad attendere fino all'ultimo istante, l*'Euronext market* fornisce un tempo di arresto casuale. Questo introduce incertezza nella strategia dell'ultimo istante e così ne scoraggia il suo uso.

### 2.3.5 Contrattazione

Alcune interazioni di scambio di titoli potrebbero ricordare l'usuale situazione cliente-venditore nei mercati dei beni in cui il negoziante fissa il prezzo e il cliente decide o meno di comprare. Benché il prezzo quasi certamente deve essere vincolato da forze di competizione (l'accesso dei clienti a fornitori alternativi o beni sostitutivi), l'interazione è essenzialmente un gioco di contrattazione.
Nello scambio di titoli i clienti al dettaglio e i loro *dealers* possono trovarsi in una situazione simile. Un cliente al dettaglio in U.S. che desidera comprare o vendere un obbligazione contatterà il suo *broker* e richiederà il prezzo a cui il *broker* vorrebbe comprare o vendere. Il *broker* annuncia il prezzo e il cliente può decidere se effettuare uno scambio oppure no.
In termini economici, questa situazione si traduce nella teoria dei giochi. In un gioco standard a piena informazione, un agente (l'allocatore) propone una divisione del pay-off totale e gli altri agenti (i beneficiari) o accettano o rifiutano la proposta. Se il beneficiario accetta, entrambi i giocatori ricevono il pay-off proposto, se rifiuta, entrambi i giocatori ricevono zero.

Il beneficiario razionale accetta ogni proposta che gli da un pay-off diverso da zero, e conoscendo ciò, l'allocatore razionale tiene per se stesso quasi tutto il pay-off totale. In pratica, i beneficiari spesso rifiutano le proposte percepite come ingiuste e questo costringe gli allocatori a disciplinare la loro avarizia. Quando il pay-off totale è conosciuto dall'allocatore ma non dal beneficiario quest'ultimo non può valutare così facilmente l'equità della proposta. Questa incertezza favorisce l'allocatore, in questo caso, il *dealer*. Per questa ragione, nell'interazione *dealer*/cliente, i regolatori americani hanno cercato di promuovere la pubblicazione dei prezzi degli scambi.

Un'altra situazione di contrattazione standard si presenta nel *LiquidNet*, un sistema di scambio per gli investitori istituzionali americani. Le istituzioni anonimamente inseriscono quantità e direzione (vendere o comprare) degli scambi desiderati. Il sistema LiquidNet cerca tra tutte le inserzioni e quando trova un match, contatta il venditore e il compratore e li mette in una stanza di incontro (*matching room*) virtuale dove essi possono contrattare sul prezzo. Il protocollo di contrattazione essenzialmente permette al beneficiario di rifiutare la proposta iniziale dell'allocatore e di suggerirne un'altra, e cosi via.

## 2.4 ERRORI DI MICROSTRUTTURA

Dopo tale rassegna sui vari meccanismi di scambio che possono verificarsi nei mercati dei titoli concentriamo l'attenzione su alcune delle cause dell'errore di microstruttura.

### 2.4.1 Scambi non sincronizzati

L'effetto degli scambi non sincronizzati (*asynchronous trades*) o dell'assenza di scambi (*no trading*) si presenta quando le serie temporali, di solito prezzi di beni, sono erroneamente misurate ad intervalli di tempo spaziati uniformemente. Per esempio, i prezzi giornalieri dei titoli quotati nei giornali finanziari sono di solito prezzi di chiusura. Tali prezzi generalmente non vengono registrati ogni giorno nello stesso istante, ma riferendoci a quelli come prezzi giornalieri abbiamo implicitamente e incorrettamente assunto che sono ugualmente spaziati nell'intervallo delle 24 ore. Tale assunzione può creare una previsione errata nei cambiamenti di prezzo.

In particolare l'effetto di assenza di scambi induce errori nei momenti singoli e congiunti dei rendimenti.

Per esempio supponiamo che i rendimenti di due stock A e B siano indipendenti ma A è scambiato meno frequentemente di B. Se arrivano news poco prima della chiusura del mercato è probabile che il prezzo di B rifletta questa informazione piuttosto che quello di A semplicemente perché A potrebbe non registrare alcun scambio dopo l'arrivo delle news. Naturalmente anche A risponderà a queste informazioni, ma il fatto che risponda con un ritardo induce cross-correlation spuria tra i rendimenti giornalieri di A e B quando sono calcolati con i prezzi di chiusura. Questa risposta ritardata indurrà anche autocorrelazione spuria nel rendimento di A: durante il periodo di assenza di scambi il rendimento osservato di A è zero, mentre quando A viene scambiato il suo rendimento osservato torna al rendimento medio, questa mean-reversion crea correlazione seriale negativa nei rendimenti di A.

In generale l'assenza di scambi non influisce nella media dei rendimenti osservati ma se il titolo ha un rendimento previsto diverso da zero ne

aumenta la loro varianza. Inoltre l'avere un rendimento previsto diverso da zero induce correlazione seriale negativa.

### 2.4.2 Bid-ask spread

Una delle caratteristiche più importanti che gli investitori cercano in un mercato è la liquidità, cioè l'abilità di comprare o vendere quantità di un titolo velocemente, anonimamente e con un impatto sul prezzo relativamente piccolo. Per mantenere la liquidità molti mercati organizzati usano i *market maker*, individui disposti a comprare o vendere ogni volta che il pubblico lo desidera. Per fornire liquidità i *market maker* si accordano sui prezzi, comprano ad un prezzo *bid* ($\mathbf{p_b}$) e vendono al più alto prezzo *ask* ($\mathbf{p_a}$). Questa abilità di comprare a prezzi bassi e vendere a prezzi alti è la risorsa primaria dei *market maker* e benché il *bid-ask spread* (cioè $\mathbf{p_a}$-$\mathbf{p_b}$) è raramente più grande di 1 o 2 *ticks*, su un gran numero di scambi possono guadagnare abbastanza da compensare il loro servizio.
La presenza del *bid-ask spread* complica la questione in diversi modi.
Invece di un prezzo ce ne sono tre : il prezzo *bid*, il prezzo *ask* e il prezzo della transazione che non non necessariamente è il prezzo *bid* o quello *ask* né un valore tra i due. Inoltre, quando arrivano vendite o acquisti casuali nel mercato, il prezzo può rimbalzare avanti e indietro tra i due prezzi creando volatilità spuria e correlazione nei rendimenti anche se il valore economico del titolo non è cambiato.
Per spiegare l'impatto del *bid-ask spread* sulle caratteristiche delle serie temporali di un titolo, Roll (1984) ha proposto un semplice modello.
Sia $p(t)$ il logaritmo del prezzo in assenza di effetti dovuti alla microstruttura e $I(t)$ una variabile indicatrice che vale $I(t) = 1$ se il prezzo osservato è *ask* e $I(t) = -1$ se il prezzo è *bid*. La sequenza $I(t)$ è formata da

variabili casuali IID con $\Pr(I(t) = 1) = \Pr(I(t) = -1) = {}^{1}/_{2}$ e $\xi$ rappresenta lo spread. Allora il nuovo prezzo contaminato sarà

$$p^*(t) = p(t) + \frac{\xi}{2} I(t) \tag{1}$$

Così che

$$\begin{aligned} dp^*(t) &= dp(t) + \frac{\xi}{2}\big(I(t) - I(t-1)\big) \\ &= dp(t) + \frac{\xi}{2} dI(t) \end{aligned} \tag{2}$$

Poiché $dp(t)$ e $dI(t)$ sono indipendenti la varianza, la covarianza, e l'autocorrelazione del primo ordine dei rendimenti dell'attività finanziaria in presenza di errore possono essere facilmente calcolati da (2) e sono

$$var\big(dp^*(t)\big) = \sigma^2 \quad (t) + \frac{\xi^2}{2} \tag{3}$$

$$Cov\big(dp^*(t), dp^*(t-1)\big) = -\frac{\xi^2}{4} \tag{4}$$

$$Corr\big(dp^*(t), dp^*(t-1)\big) = -\frac{{}^{\xi^2}/_{4}}{\sigma^2(t) + {}^{\xi^2}/_{2}} \leq 0 \tag{5}$$

Quindi si nota come $dp^*(t)$ esibisce volatilità spuria e autocorrelazione negativa come risultato di *bid-ask bounce* e, più grande è lo *spread* più grande sarà tale volatilità.

Poiché un dato *spread* implica un' auotocovarianza del primo ordine di $-\xi^2/4$, la logica potrebbe essere rovesciata così che un dato coefficiente di autocovarianza e un dato valore di $p$ implicano un determinato valore di $\xi$; infatti risolvendo per $\xi$ l'equazione (4) si ottiene

$$\xi = 2\sqrt{-cov[dp^*(t), dp^*(t-1)]}$$

Roll stima lo spread degli stocks dei mercati NYSE e AMEX usando rendimenti dal 1963 al 1982, e trova che lo spread medio è circa 0.298% usando rendimenti giornalieri e 1,74% usando rendimenti mensili. Più recentemente Nielsen e Frederiksen (2006) hanno studiato lo spread di sei stocks scambiati frequentemente e tre indici di stock utilizzando rendimenti giornalieri e il metodo di stima di Roll per il periodo di tempo 1.1.1995-30.6.2004. I risultati trovati risultano solo un po' più bassi dei primi trovati da Roll in particolare corrispondono a $\xi \in (0.0019, 0.0108)$.

### 2.4.3 Discretezza dei prezzi

Come è ben noto i prezzi si muovono in unità discrete, la grandezza esatta di queste unità può essere imposta da un regolatore o da uno scambio, o può presentarsi come una convenzione del mercato. Nel mercato americano i titoli sono quotati in 1/8 di dollaro ciò vuol dire che i prezzi vengono registrati secondo multipli di questa quantità sebbene il loro valore in realtà è diverso. E' chiaro intuire come questo provochi un errore nella serie dei prezzi e quindi dei rendimenti. La discretezza dei prezzi è inoltre importante per *bid-ask spread* poiché il minimo incremento di prezzo, o "*tick size*", pone un limite inferiore nella grandezza del *bid-ask spread*.

# CAPITOLO TERZO

# VARIANZA INTEGRATA E SUOI STIMATORI

## 3.1 VOLATILITA' INTEGRATA E VARIAZIONE QUADRATICA

Innanzitutto bisogna fare una piccola premessa sui modelli utilizzati per stimare la volatilità. Se nella maggior parte della letteratura si sono utilizzati i modelli GARCH e loro varianti nella modellazione dei dati ad alta frequenza, qui utilizzeremo un secondo filone che si riferisce ai modelli a volatilità stocastica (SV), nei quali si pone che la volatilità sia generata da un processo stocastico cosicché anche l'equazione per la varianza ha un termine d'errore.

Supponiamo quindi che il logaritmo del prezzo di una data attività finanziaria sia descritto dalla seguente equazione differenziale stocastica

$$dp(t) = \mu(t)dt + \sigma(t)dW(t) \qquad con\ \ t > 0 \tag{6}$$

dove la media del processo $\mu(\cdot)$ e la volatilità istantanea del processo $\sigma(\cdot) > 0$ sono assunti essere indipendenti dal moto Browniano $W(\cdot)$. Un importante caratteristica di tale modello è la gaussianità condizionatamente all'insieme di informazione disponibile

$$p(t)|\int_0^t \mu(s)d(s), \sigma^{2*}(t) \sim N\left(\int_0^t \mu(s)ds, \sigma^{2*}(t)\right)$$

dove

$$\sigma^{2*}(t) = \int_0^t \sigma^2(s)ds = IV_t \tag{7}$$

è chiamata ***volatilità integrata*** (o varianza integrata) ed è l'oggetto di interesse.

Un altro importante oggetto è la ***variazione quadratica*** del processo $[p(t)]$, definita da

$$[p(t)] = p^2(t) - 2\int_0^t p(s-)dp(s) \tag{8}$$

equivalentemente da

$$[p(t)] = plim \sum_{j=1}^{n} \left(p(s_j) - p(s_{j-1})\right)^2 \qquad (9)$$

dove $0 = s_0 < s_1 < .... < s_n = t$ e il limite è dato per $max_j |s_j - s_{j-1}| \to 0$ quando $n \to \infty$.

Sotto alcune condizioni di regolarità molto generali è stato evidenziato da Andersen e Bollerslev (1998) e Barndorff-Nielsen e Shepard (2001) che

$$[p](t) = \sigma^{2*}(t) \qquad (10)$$

per il modello (1). Per lo scopo di questo lavoro notiamo che questo implica che l'oggetto di interesse, $\sigma^{2*}(t)$, può essere stimato sia direttamente tramite modelli parametrici sia non parametricamente tramite variazione quadratica.

## 3.2 STIMA DELLA VOLATILITA' INTEGRATA

Nei prossimi paragrafi illustreremo due diversi metodi di stima della volatilità integrata basati su un campione di rendimenti ad alta frequenza. Primo, descriveremo l'approccio più conosciuto, la volatilità realizzata poi lo stimatore di Fourier che stima $\sigma^{2*}(t)$ attraverso appunto la trasformata di Fourier.

Per rendere più chiara la notazione facciamo alcune assunzioni semplificatrici. Assumiamo che sia disponibile solo un giorno di osservazioni e chiamiamo $p_i$ le osservazioni intragiornaliere del logaritmo

del prezzo dell'attività finanziaria. Nell'analisi empirica che è stata sviluppata in questo lavoro per il calcolo della volatilità integrata, $\sigma^{2*}(t)$, sono stati ovviamente considerati $t = 1, \ldots\ldots T$ giorni e utilizzati dati intragiornalieri del processo dei prezzi. In questa sezione invece, abbiamo fissato $T = 1$ e considerato la stima della volatilità solo in un periodo di tempo.

### 3.2.1 Varianza realizzata

Supponiamo di avere a disposizione $n$ osservazioni intragiornaliere del logaritmo del prezzo ad intervalli di tempo $k_0, k_1, \ldots k_n$ per cui la varianza integrata su ogni sottointervallo è definita da:

$$\sigma_{i,t}^{2*} = \int_{k_{i-1}}^{k_i} \sigma^2\,(s) d(s) = IV_{i,t} \qquad i = 1, \ldots\ldots n \tag{11}$$

Usando gli $n$ prezzi si costruiscono i rendimenti intragiornalieri

$$r_i = p_i - p_{i-1} \qquad i = 1, \ldots\ldots n$$

Da cui usando la (9) la variazione quadratica può essere stimata da:

$$\hat{\sigma}_{RV,n}^{2*} = \sum_{i=1}^{n} r_i^2 = RV_t. \tag{12}$$

Tale quantità è chiamata ***varianza realizzata*** nel giorno $t$ del processo $p(\cdot)$. Andersen e Bollerslev (1998) e Andersen *et al.* (2003) hanno dimostrato che è uno stimatore consistente della volatilità integrata. Il risultato di consistenza non richiede che le osservazioni siano equispaziate regolarmente ma solo che la massima distanza tra le osservazioni tenda a zero. Barndoff-Nielsen e Shepard (2002) oltre che a confermare i risultati di consistenza hanno derivato la distribuzione asintotica di RV come:

$$n^{\frac{1}{2}} \cdot \frac{1}{\sqrt{2IQ_t}}(RV_t - IV_t) \xrightarrow{d} N(0,1) \qquad per\ n \longrightarrow \infty$$

dove IQ, la quarticità integrata, è definita come:

$$IQ_t = \int_0^1 \sigma^4\,(t)dt.$$

Inoltre Barndoff-Nielsen e Shepard, sotto l'assunzione di assenza di errore dovuto alla microstruttura, hanno mostrato che la quarticità integrata è consistentemente stimata dalla quarticità realizzata, che è definita come:

$$RQ_t = \frac{n}{3}\sum_{i=o}^{n} r_{t,i}^4.$$

E quindi sono riusciti a dimostrare che

$$n^{\frac{1}{2}} \cdot \frac{1}{\sqrt{\frac{2}{3}RQ_t}}(RV_t - IV_t) \xrightarrow{d} N(0,1). \qquad (13)$$

La bontà dello stimatore RV è governata dalla scelta di $n$. Per esempio se gli scambi avvengono durante tutte le 24 ore del giorno come nel caso dei mercati dei cambi, scegliere rendimenti ogni minuto corrisponde a $n = 1440$, ogni 5 minuti corrisponde a $n = 288$ e ogni 15 minuti a $n = 96$. Scegliere un numero più grande di osservazioni intragiornaliere migliora la precisione dello stimatore ma allo stesso tempo lo rende più vulnerabile agli effetti della microstruttura del mercato, discussi nel primo capitolo, e quindi vedremo di seguito come lo stimatore diventi inconsistente e necessiti così di alcune modifiche.

### 3.2.2 Stimatore di Fourier

Questo stimatore fu suggerito da Malliavin e Mancino (2002) e in seguito applicato da Barucci e Reno (2002a, 2002b). Il metodo di Fourier richiede solo che la variazione quadratica (8) o (9) sia limitata. Il metodo è basato sulla trasformata di Fourier definita da

$$\int_0^{2\pi} \sigma^2(s)ds = 2\pi a_0(\sigma^2) \tag{14}$$

dove

$$a_o(\sigma^2) = \lim_{S\to\infty} \frac{\pi}{2S} \sum_{s=1}^{S} (a_s^2(dp) + b_s^2(dp)) \tag{15}$$

e $S$ rappresenta il numero di coefficienti di Fourier dati da

$$a_s(dp) = \frac{1}{\pi}\int_0^{2\pi} \cos(st)dp(t) \qquad s \geq 1$$

$$b_s(dp) = \frac{1}{\pi}\int_0^{2\pi} \sin(st)dp(t) \qquad s \geq 1$$

Quindi le $n$ osservazioni intragiornaliere spaziate irregolarmente ai tempi $k_1, \dots \dots, k_n$ cioè nell'intervallo $[k_1, k_n]$ hanno bisogno di essere normalizzate nell'intervallo $[0, 2\pi]$, chiamiamo allora i punti temporali rinormalizzati come:

$$\tau_i = \frac{2\pi(k_i - k_1)}{k_n - k_1} \qquad con\ i = 1, \dots \dots n$$

Barucci & Reno hanno derivato la seguente approssimazione per la stima dei coefficienti di Fourier

$$\hat{a}_s(dp) = \frac{p(\tau_n) - p(\tau_1)}{\pi} - \frac{1}{\pi}\sum_{i=2}^{n} p(\tau_{i-1})(\cos(s\tau_i) - \cos(s\tau_{i-1}))$$

$$\hat{b}_s(dp) = \frac{1}{\pi}\sum_{i=2}^{n} p(\tau_{i-1})(\sin(s\tau_i) - \sin(\tau_{i-1})) \qquad s \geq 1$$

Per ottenere lo stimatore di volatilità integrata, sostituiamo queste due ultime equazioni in (15) e usiamo la (14):

$$\hat{\sigma}_{F,S}^{2*} = \frac{\pi^2}{2} \sum_{s=1}^{S} \left( \hat{a}_s^2(dp) + \hat{b}_s^2(dp) \right) = FR_t \qquad (16)$$

La consistenza di tale stimatore è stata provata in assenza di errore di microstruttura da Malliavin e Mancino(2005), in particolare $FR_t$ converge in probabilità alla quantità $\int_0^{2\pi} \sigma^2(s)ds$. Recentemente in un lavoro di Mancino e Sanfelici (2007) è stato inoltre provato che lo stimatore di Fourier non ha bisogno di correzioni per essere robusto rispetto ad errori del mercato autocorrelati con una struttura a media mobile del primo ordine.

La bontà dello stimatore è controllata dalla scelta di $S$, cioè il numero di coefficienti di Fourier da includere nella stima, che è legato alla scelta di $n$ per la varianza realizzata da $S = n/2$. Scegliere un $S$ grande allora corrisponde alla scelta di un ottimo campionamento ($n$ grande) per la varianza realizzata, cioè scegliere 1 minuto piuttosto che 5 o 15 minuti. Bisogna notare che se tra le $S$ frequenze si includono solo quelle basse nello stimatore di Fourier allora l'errore, che si registra alle alte frequenze, viene ignorato. Quindi scegliendo un $S$ piccolo lo stimatore di Fourier è poco influenzato dall'errore di microstruttura del mercato.

## 3.3 GLI EFFETTI DELL'ERRORE

Supponiamo ora di osservare il logaritmo del prezzo con errore. Denominando $p(t)$ il prezzo che realmente osservo e $p^*(t)$ il vero prezzo, l'errore è dato da:

$$\varepsilon_{t,i} \equiv p_{t,i} - p^*_{t,i} \tag{17}$$

L'errore $\varepsilon_t$ riassume un insieme di diversi effetti dovuti alla microstruttura del mercato che possono essere suddivisi in tre gruppi. Primo, $\varepsilon$ rappresenta gli errori inerenti il processo di scambio: *bid-ask bounce*, discretezza dei prezzi e arrotondamento, scambi in diversi mercati o *networks*, ecc. Secondo, ε cattura gli effetti dovuti all'informazione: differenze nelle dimensioni dello scambio o nel contenuto informativo dei prezzi, risposta graduale ad un blocco commerciale, ecc. Terzo, ε comprende errori di misura o di registrazione dei prezzi.
Segue che i veri rendimenti giornalieri sono definiti da:

$$r^*_{t,i} = p^*_{t,i} - p^*_{t,i-1} \tag{18}$$

Mentre quello che osservo è

$$r_{t,i} = p_{t,i} - p_{t,i-1}$$

Sostituendo a $p_{t,i}$ e a $p_{t,i-1}$ le loro rispettive equazioni (ricavate da (17)) risulta

$$r_{t,i} \equiv r_{t,i}^{*} + \varepsilon_{t,i} - \varepsilon_{t,i-1} = r_{t,i}^{*} + \nu_{t,i}$$

Quindi la varianza realizzata sarà

$$RV_t = \sum_{i=1}^{n} r_{t,i}^{2*} + 2\sum_{i=1}^{n} r_{t,i}^{*}\,\nu_{t,i} + \sum_{i=1}^{n} \nu_{t,i}^{2}$$

Si può mostrare che, condizionatamente ai veri rendimenti

$$E(RV_t|r^*) = RV_t^* + 2nE\left(\varepsilon_{t,i}^2\right) \qquad (19)$$

$$Var(RV_t|r^*) = 4nE\left(\varepsilon_{t,i}^4\right) + O_p(1) \qquad (20)$$

Tale risultato suggerisce che $RV_t$ in presenza di errore non è uno stimatore corretto della vera volatilità realizzata. Per un valore di $n$ elevato, la varianza realizzata potrebbe non avere nulla a che fare con i veri rendimenti. Infatti, nel momento primo si riferisce al termine d'errore, $E(\varepsilon^2)$, e nel momento secondo a $E(\varepsilon^4)$. Inoltre, si può osservare dalla (19) che $RV_t$ ha un *bias* positivo la cui grandezza aumenta linearmente con la grandezza del campione, cioè $n$.

E' interessante notare come tale analisi, oltre a rivelare la distorsione dello stimatore $RV$, ci consegna uno stimatore consistente per la varianza del termine d'errore. Infatti si può far vedere che:

$$\widehat{E\left(\varepsilon_{t,i}^2\right)} = \frac{1}{2n} RV_t \qquad (21)$$

Consideriamo ora, come in Bandi e Russel (2005), la seguente assunzione riguardante l'errore di microstruttura:

**Assunzione 1**

**a)** L'errore $\varepsilon_{t,i}$ ha media zero ed è un processo stocastico stazionario in covarianza

**b)** La varianza di $\nu_{t,i} = \varepsilon_{t,i} - \varepsilon_{t,i-1}$ è $O(1)$

Sotto tale assunzione, Bandi e Russel (2005) hanno mostrato che

$$RV_t \xrightarrow{q.c} \infty \qquad per \qquad n \to \infty$$

Consideriamo inoltre una struttura IID per l'errore

**Assunzione 2**

**a)** L'errore, $\varepsilon_{t,i}$, ha media zero ed è indipendente ed identicamente distribuito

**b)** L'errore è indipendente dal processo del prezzo

**c)** La varianza di $\nu_{t,i} = \varepsilon_{t,i} - \varepsilon_{t,i-1}$ è $O(1)$

Sotto l'assunzione 2 Zhang *et al*. (2005), di seguito ZMA (2005), hanno mostrato che

$$n^{-\frac{1}{2}}\left[RV_t - IV_t - 2nE(\varepsilon_{t,i}^2)\right] \xrightarrow{d} 2\left[E(\varepsilon_{t,i}^4)\right]^{\frac{1}{2}} \cdot N(0,1)$$

Nelle applicazioni pratiche, anche campionando alla più alta frequenza possibile, il numero di osservazioni risulta sempre finito. Questo introduce un errore dovuto alla discretizzazione, tale che

$$RV_t \overset{d}{\approx} IV_t + \underbrace{2n\left(\varepsilon_{t,i}^2\right)}_{\substack{bias\ dovuto \\ all'errore}} + \underbrace{[\underbrace{4nE\left(\varepsilon_{t,i}^4\right)}_{\substack{dovuto \\ all'errore}} + \underbrace{\frac{2}{n}\int_0^1 \sigma_t^4\, dt}_{\substack{dovuto\ alla \\ discretizzazione}}]^{\frac{1}{2}}}_{varianza\ totale} \cdot N(0,1)$$

dove $\overset{d}{\approx}$ significa, a meno di un opportuno fattore, convergenza in distribuzione. Così il *bias* dovuto all'errore che è dell'ordine $O(n)$ prevale sulla vera volatilità integrata che è dell'ordine di $O(1)$.

Infine consideriamo il caso in cui l'errore non sia IID, l'assunzione 2 sarà quindi modificata come segue:

**Assunzione 3**

**a)** L'errore, $\varepsilon_{t,i}$, è un processo stocastico stazionario *strong mixing* di media nulla. Inoltre $E[\left(\varepsilon_{t,i}\right)^{4+k}] < \infty$ per qualche $k > 0$

**b)** L'errore è indipendente dal processo dei prezzi

**c)** La varianza di $\nu_{t,i} = \varepsilon_{t,i} - \varepsilon_{t,i-1}$ è $O(1)$

Sotto l'assunzione 3 Ait-Sahalia et. Al (2006), di seguito AMZ (2006), hanno mostrato che

$$RV_t \overset{d}{\approx} IV_t + \underbrace{2n\left(\varepsilon_{t,i}^2\right)}_{\substack{bias\ dovuto \\ all'errore}} + \underbrace{[\underbrace{4n\Omega}_{\substack{dovuto \\ all'errore}} + \underbrace{\frac{2}{n}\int_0^1 \sigma_t^4\, dt}_{\substack{dovuto\ alla \\ discretizzazione}}]^{\frac{1}{2}}}_{varianza\ totale} \cdot N(0,1)$$

dove

$$\Omega = Var\left[\left(\varepsilon_{t,i} - \varepsilon_{t,0}\right)^2\right] + 2\sum_{i=1}^{\infty} Cov[\left(\varepsilon_{t,i} - \varepsilon_{t,0}\right)^2, \left(\varepsilon_{t,i+1} - \varepsilon_{t,i}\right)^2]$$

## 3.4 STIMATORI ALTERNATIVI

Nella letteratura esistente sono stati suggeriti diversi metodi per la riduzione del *bias*

- Tecniche di sottocampionamento
- Metodi di correzione dell'autocorrelazione

### 3.4.1 Sottocampionamento

Abbiamo appena argomentato che la volatilità realizzata stima una quantità sbagliata in presenza di errore. Questo problema peggiora quando le osservazioni sono campionate a frequenze molto alte. Infatti i rendimenti sono ancor di più contaminati da errore di microstruttura e diventano meno rappresentativi della vera quantità da stimare (cioè $IV$). Lungo questa linea di discussione, l'opinione generale nelle applicazioni finanziarie è di non campionare così spesso almeno quando usiamo la volatilità realizzata. Una scelta abbastanza condivisa è considerare l'intervallo di campionamento nel *range* tra i 5 e i 30 minuti. Questa procedura è chiamata *sparse*

*sampling*. Ciò significa ridurre il valore di $n$ da 1440 osservazioni ogni minuto a 288, se per esempio si effettua un campionamento ogni 5 minuti. Facendo questo abbiamo il vantaggio di ridurre la grandezza del *bias* $2nE[\varepsilon^2]$. Nonostante l'errore si riduca quando campioniamo ad intervalli più grandi, quello che si può notare è che la varianza dovuta alla discretizzazione è aumentata. Ci troviamo di fronte ad un *trade-off* tra *bias* e varianza.

Definiamo come prima cosa un nuovo insieme di osservazioni $\Lambda^{(sparse\ )}$contenente $n^{(sparse\ )}$ prezzi campionati a tempi equidistanti. Ovviamente $\Lambda^{(sparse\ )}$è un sottoinsieme di $\Lambda$ e di conseguenza $n^{(sparse\ )} < n$. Definiamo allora il nuovo stimatore come

$$RV_t^{(sparse\ )} = \sum_{i=1}^{n^{(sparse\ )}} r_{t,i}^2 \tag{22}$$

Basandoci sui risultati di ZMA (2005) e AMZ (2006), sotto l'assunzione 2 o 3

$$RV_t^{(sparse\ )} \overset{d}{\approx} IV_t + \underbrace{2nE\left(\varepsilon_{t,i}^2\right)}_{\substack{bias\ dovuto \\ all'errore}}$$

$$+ \underbrace{[\underbrace{4n^{(sparse\ )}E(\varepsilon_{t,i}^4)}_{\substack{dovuto \\ all'errore}} + \underbrace{\frac{2}{n^{(sparse\ )}}\int_0^1 \sigma_t^4\, dt}_{\substack{dovuto\ alla \\ discretizzazione}}]^{\frac{1}{2}}}_{varianza\ totale} \cdot N(0,1)$$

La scelta di $n^{(sparse\ )}$ non è casuale. Infatti è possibile determinare una frequenza di campionamento ottimale minimizzando l' errore quadratico medio di $RV^{(sparse\ )}$:

$$MSE = 2n^{(sparse\ )}E\left(\varepsilon_{t,i}^2\right) + 4n^{(sparse\ )}E\left(\varepsilon_{t,i}^4\right) + \left[8RV_t^{(sparse\ )}E\left(\varepsilon_{t,i}^2\right) - 2V\left(\varepsilon_{t,i}^2\right)\right] + \frac{2}{n^{(sparse\ )}}IQ_t^{(sparse\ )}$$

Si può far vedere che l'$n^{(sparse\ )}$ ottimale è dato da

$$n^{*(sparse\ )} \approx \left\{\frac{IQ_t}{4[E\left(\varepsilon_{t,i}^2\right)]^2}\right\}^{\frac{1}{3}}$$

Abbiamo visto come si può trarre beneficio usando dati campionati meno frequentemente per quanto riguarda il *bias* ma come questo porti ad un aumento della varianza totale. Presentiamo ora un modo alternativo di affrontare il problema. Invece di selezionare, arbitrariamente o ottimamente, un sottocampione il metodo si basa nel selezionare un numero di sottocampioni da quello originale e poi fare la media degli stimatori derivati dai sottocampioni.

In quanto segue prima presenteremo un insieme di notazioni per poi studiare la volatilità realizzata in un contesto con più insiemi di osservazioni; infine mostreremo come eliminare il *bias* dello stimatore usando due diverse scale temporali, una combinazione di un singolo insieme di osservazioni con più insiemi di osservazioni.

Supponiamo che l'intero insieme di informazioni $\Lambda = \{k_0 \dots \dots k_n\}$ sia suddiviso in H sottoinsiemi disgiunti $\Lambda^{(h)}$ con h=1,….,H; in altre parole

$$\Lambda = \bigcup_{h=1}^{H} \Lambda^{(h)} \qquad dove \qquad \Lambda^{(h)} \cap \Lambda^{(l)} = \emptyset \quad \forall h \neq l$$

Il modo più naturale per selezionare gli H sottoinsiemi è di cominciare con $k_{h-1}$ e poi scegliere le altre osservazioni ogni H cioè

$$\Lambda^{(h)} = \{\, k_{h-1}, k_{h-1+H}, k_{h-1+2H} \dots \dots \} \;\; per\; h = 1, \dots \dots H$$

Sia $n^{(h)}$ il numero di osservazioni in ogni sottoinsieme. Definiamo la varianza realizzata per ogni insieme h come

$$RV_t^{(h)} = \sum_{i=1}^{n^{(h)}} r_{t,i}^2 \tag{23}$$

Quindi facendo la media otteniamo lo stimatore

$$RV_t^{(avg)} = \frac{1}{H} \sum_{h=1}^{H} RV_t^{(h)} \tag{24}$$

Come visto in precedenza combinando i due tipi di errore, quello dovuto al *bias* e quello dovuto alla discretizzazione, dai risultati di ZMA (2005) risulta che

$$RV_t^{(avg)} \overset{d}{\approx} IV_t + \underbrace{2\bar{n}E\left(\varepsilon_{t,i}^2\right)}_{\substack{bias\ dovuto \\ all'\ errore}}$$

$$\underbrace{+\left[\underbrace{4\frac{\bar{n}}{k}E(\varepsilon_{t,i}^4)}_{\substack{dovuto \\ all'errore}} + \underbrace{\frac{4}{3\bar{n}}IQ_t}_{\substack{dovuto\ alla \\ discretizzazione}}\right]^{\frac{1}{2}}}_{varianza\ totale} \cdot N(0,1)$$

dove

$$\bar{n} = \frac{1}{H}\sum_{k=1}^{H} n^{(h)} = \frac{n-k+1}{k}$$

Tale quantità è stata definita poiché in generale le $n^{(h)}$ osservazioni non necessariamente sono uguali all'interno di ogni sottoinsieme *h*.

Come possiamo notare lo stimatore $RV^{(avg)}$ è ancora uno stimatore non corretto della variazione quadratica del vero processo dei rendimenti. Per quanto riguarda il *bias* invece possiamo notare che è comunque uno stimatore migliore di *RV* calcolata sull'intero insieme di informazioni disponibili, poiché $\bar{n} < n$. Per bilanciare la coesistenza di *bias* e varianza possiamo cercare di ottimizzare la scelta di $\bar{n}$ minimizzando l'MSE da cui risulta che

$$\bar{n}^* = \left[\frac{IQ_t}{6E(\varepsilon^2)^2}\right]^{\frac{1}{3}}$$

Di seguito cercheremo di migliorare lo stimatore ottenuto adottando una correzione per l'errore. Come dimostrato in precedenza da (21), $E(\varepsilon^2)$ può essere stimato consistentemente, di conseguenza il *bias* di $RV^{(avg)}$ può

essere stimato consistentemente da $2\bar{n}E[\varepsilon^2]$. Uno stimatore di IV aggiustato per l'errore può essere allora ottenuto come

$$RV_t^{(TTSE)} = \underbrace{RV_t^{(avg)}}_{\substack{slow\ time \\ scale}} - \frac{\bar{n}}{n} \underbrace{RV_t}_{\substack{fast\ time \\ scale}} \tag{25}$$

combinando due diverse scale temporali. Tale stimatore è chiamato ***Two Time Scales Estimator*** **(TTSE)**. Sotto l'assunzione 2 e se il numero di sottocampioni è ottimamente selezionato, cioè $H = cn^{2/3}$, ZMA (2005) hanno mostrato che

$$RV_t^{(TTSE)} \overset{d}{\approx} IV_t + \frac{1}{n^{1/6}} \underbrace{+ [\underbrace{8c^{-2}E(\varepsilon_{t,i}^4)}_{\substack{dovuto \\ all'errore}} + \underbrace{\frac{4c}{3} IQ_t}_{\substack{dovuto\ alla \\ discretizzazione}}]^{\frac{1}{2}}}_{varianza\ totale} \cdot N(0,1)$$

dove, nel caso di osservazioni equidistanti,

$$c = \left\{ \frac{1}{12E\left[(\varepsilon_{t,i}^2)\right]^2} IQ_t \right\}^{\frac{1}{3}}$$

In piccoli campioni, possiamo perfezionare la scrittura come segue:

$$RV_t^{(TTSE;adj)} = \left(1 - \frac{\bar{n}}{n}\right)^{-1} RV_t^{(TTSE)} \tag{26}$$

Entrambi gli stimatori in (25) e (26) sono derivati sotto l'assunzione di errore IID.

Per prendere in considerazione la possibilità di errore dipendente, AMZ (2006) proposero uno stimatore alternativo basato anch'esso sull'utilizzo di due diverse scale temporali. Tutti i risultati seguenti sono quindi derivati sotto l'assunzione 3.

AMZ (2006), come nel caso precedente, hanno proposto uno stimatore calcolato come media di stimatori ottenuti su sottocampioni costruiti su ogni j-esima osservazione, chiamato stimatore volatilità realizzata media al ritardo J definito da

$$RV_{t,J} = \frac{1}{J}\sum_{i=0}^{n-J}\left(r_{t,i+J} - r_{t,i}\right)^2 \tag{27}$$

Poi gli autori proposero una generalizzazione di TTSE combinando due diverse scale temporali J e K

$$RV_t^{(TTSE)} = \underbrace{RV_{t,K}}_{\substack{slow\ time \\ scale}} - \frac{\bar{n}^{(K)}}{n^{(J)}} \underbrace{RV_{t,J}}_{\substack{fast\ time \\ scale}} \tag{28}$$

dove $1 \leq J \leq K \leq n$, $K = o(n)$, $\bar{n}^{(K)} = \frac{n-K+1}{K}$ e $n^{(J)} = \frac{n-J+1}{J}$. Lo stimatore in (28) diventa TTSE (25) quando J=1 e K→∞ per n→∞. Come nel caso precedente una correzione per piccoli campioni è data da

$$RV_t^{(TTSE,adj\ )} = \left(1 - \frac{\bar{n}^{(K)}}{n^{(J)}}\right)^{-1} RV_t^{(TTSE)} \tag{29}$$

AMZ (2006) hanno mostrato che per determinate scelte di J e K lo stimatore è consistente; il range ragionevole di valori è 5 minuti ≤ K ≤ 9 minuti e 5 secondi ≤ J ≤ 2 minuti.

Infine hanno mostrato che

$$RV_t^{(TTSE)} \overset{d}{\approx} IV_t + \frac{1}{n^{1/6}} + \underbrace{[\ \underbrace{c^{-2}\xi^2}_{\substack{dovuto \\ all'errore}} + \underbrace{\frac{4c}{3} IQ_t}_{\substack{dovuto\ alla \\ discretizzazione}}\ ]^{\frac{1}{2}}}_{varianza\ \ totale} \cdot N(0,1)$$

dove c è una costante e

$$\xi^2 = 16 \cdot Var\left(\varepsilon_{t,i}\right)^2 + 32 \sum_{i=1}^{\infty} Cov\left(\varepsilon_{t,0}, \varepsilon_{t,i}\right)^2$$

### 3.4.2 Correzione per l'autocorrelazione

Stimare consistentemente la variazione quadratica in presenza di errore dovuto alla microstruttura è, in un certo senso, simile ad effettuare correzioni per le autocorrelazioni che sono frequentemente usate nella letteratura delle serie temporali per la stima della varianza e della covarianza di processi stazionari (vedi per esempio, Newey e West (1987)). Di conseguenza, è naturale adattare una simile tecnica per il caso preso in analisi. Zhou (1996) fu il primo a considerare l'uso dei metodi kernel per trattare il problema dell'errore dovuto alla microstruttura nei dati ad alta frequenza. Egli propose uno stimatore nel caso di errore indipendente che tenesse conto di una correlazione seriale del primo ordine.

$$RV_t^{(Zhou)} = RV_t + \sum_{i=1}^{n} r_{t,i}\, r_{t.i-1} + \sum_{i=1}^{n} r_{t,i}\, r_{t,i+1} \qquad (30)$$

Come si può notare la (30) richiede $r_0$ e $r_{n+1}$ che sono rendimenti fuori dell'intervallo di campionamento. Hansen e Lunde (2006) risolvono questo problema riscalando la covarianza con il termine $n/(n-q)$, dove $q$ è il ritardo oltre il quale la funzione di autocovarianza non è significativa, per compensare i valori mancanti ed ottenendo come stima per la quantità teorica

$$\gamma_q \equiv \sum_{i=1}^{n} r_{t,i} r_{t,i+q}$$

la quantità empirica (per $q > 0$)

$$\hat{\gamma}_q = \frac{n}{n-q} \sum_{i=1}^{n-q} r_{t,i}\, r_{t,i+q}$$

Allora, definendo per $q < 0$ $\hat{\gamma}_q \equiv \hat{\gamma}_{|q|}$, lo stimatore (30) diventa

$$RV_{t,1}^{(HL)} = RV_t + 2\frac{n}{n-1} \sum_{i=1}^{n-1} r_{t,i}\, r_{t,i+1} \qquad (31)$$

Hansen e Lunde Hanno dimostrato che sotto l'assunzione 2 lo stimatore è corretto rispetto alla quantità $RV^*$ ma non consistente ed inoltre se applicato in assenza di errore provoca un aumento della varianza di circa tre volte della varianza di RV.

Più in generale Hansen e Lunde hanno considerato uno stimatore robusto a correlazioni seriali di ordini superiori

$$RV_{t,q}^{(HL)} = RV_t + 2\sum_{q=1}^{Q} \hat{\gamma}_q \tag{32}$$

L'aver però riscalato la covarianza verso l'alto ha provocato un aumento della varianza dello stimatore; per questa ragione Hansen e Lunde (2005) considerano uno stimatore basato sulla correzione di Bartlett

$$RV_{t,q}^{(HL,Bartlett\ )} = RV_t + 2\sum_{q=1}^{Q} \left(1 - \frac{q}{Q+1}\right) \hat{\gamma}_q \tag{33}$$

dove

$$Q = \left\lceil \left(\frac{4n}{100}\right)^{\frac{2}{9}} \right\rceil$$

Lo stimatore (33), però, è anch'esso non consistente.

Recentemente Barndorff-Nielsen *et al*. (2006), di seguito BHLS (2006), hanno proposto lo stimatore

$$RV_{t,q}^{(BHLS)} = RV_t + \sum_{q=1}^{Q} k\left(\frac{q-1}{Q}\right)\left(\gamma_q + \gamma_{-q}\right) \tag{34}$$

Dove $k$(x) per x$\in$[0,1] è una funzione Kernel tale che $k$(0)=1 e $k$(1)=0.

BHLS (2006) mostrarono che, se $Q=cn^{2/3}$, allora lo stimatore si distribuisce asintoticamente una mistura di gaussiane con un tasso di convergenza $n^{1/6}$. I loro risultati rimangono robusti anche in presenza di errore dipendente.

## 3.5 EFFETTI DEGLI SCHEMI DI CAMPIONAMENTO

Come discusso nel primo capitolo, ci sono diversi metodi di campionamento per i rendimenti intragiornalieri, e la scelta dello schema di campionamento può avere una forte influenza nelle proprietà statistiche della varianza realizzata. Fino ad ora la nostra discussione non ha tenuto in considerazione tale problema. Il primo contributo in questa direzione è stato di Oomen (2005). A differenza della letteratura esistente Oomen ha proposto un processo con *jump* per modellare i prezzi ad alta frequenza e ha distinto tra diversi schemi di campionamento includendo *calendar time sampling, business time sampling, e transaction time sampling*. Il risultato principale portato da questo articolo è che transaction time sampling è superiore al solito utilizzo di *calendar time sampling* poiché si ottiene il più basso MSE.

Oomen (2006) ha esteso il modello precedente per studiare gli effetti di una correzione di primo ordine per il *bias* su diversi schemi di campionamento. La sua correzione è in linea con quella proposta da Zhou (1996) e Hansen e Lunde (2006). Però, i suoi risultati sono stati derivati assumendo per i prezzi un processo con *jump* ed errore IID. In questo lavoro Oomen mostra che la correzione riduce significativamente il *bias* e ha più effetto nel *transaction time sampling* piuttosto che in *calendar time*. Inoltre per un numero uguale di rendimenti campionati lo stimatore corretto per l'errore

raggiunge un errore quadratico medio più basso nei rendimenti campionati su scala temporale *transaction* piuttosto che in quella *calendar*.

Infine Griffin e Oomen (2006) introdussero un nuovo modello per distinguere l'effetto del *tick time* e *transaction time sampling*. I principali risultati del lavoro sono:

- l'utilizzo del metodo *tick time sampling o transaction time sampling* porta a stime della volatilità equivalenti per livelli alti dell'errore di microstruttura, risulta invece migliore l'utilizzo del *tick time sampling* per errori piccoli.
- quando si considera la correzione di primo ordine per il *bias*, il metodo *transaction time sampling* in termine di risultati di stima è sempre il migliore.

## 3.6 CONFRONTO TRA LE DIVERSE TECNICHE

Riassumiamo nella tabella seguente i risultati principali in termini di proprietà degli stimatori in presenza di errore. La correttezza è intesa rispetto alla quantità $RV^*$, ovvero la volatilità realizzata calcolata sui veri prezzi.

| **Metodo** | **Corretto** | **Consistente** | **Noise** |
|---|---|---|---|
| RV | No | No | Dipendente/ IID |
| FR | Si | Si | Dipendente/ IID |
| $RV_{sparse}$ [1] | No | No | - |
| $RV_{avg}$ | | No | |
| $RV_{ttse}$ | Si | Si | IID |
| $RV_{ttse}$ | Si | Si | Dipendente |
| $RV_{hl}$ | Si | No | Dipendente/ IID |
| $RV_{bhls}$ | Si | Si | Dipendente/ IID |

[1] Lo stimatore *sparse* non necessariamente richiede assunzioni nell'errore di microstruttura

# CAPITOLO QUARTO

# SIMULAZIONI

## 4.1 DISEGNO SPERIMENTALE

In questa sezione descriverò lo scenario utilizzato per esaminare il *bias* e l'errore quadratico medio di alcuni degli stimatori analizzati in precedenza. Attraverso esercizi di simulazione ho cercato di far luce su quale metodo stimasse la volatilità integrata in maniera più accurata utilizzando dimensioni diverse del campione.

Ho utilizzato come meccanismo generatore dei prezzi il modello

$$dp(t) = \sigma(t)dW_1(t) \tag{35}$$

ed ho assunto come processo per la volatilità $\sigma(t)$

Modello A $$dln\sigma^2(t) = \alpha\big(\beta - ln\sigma^2(t)\big)dt + \gamma dW_d(t) \qquad (36)$$

Modello B $$d\sigma^2(t) = \alpha\big(\beta - \sigma^2(t)\big)dt + \sigma(t)\gamma dW_2(t) + \kappa(t)dq(t) \qquad (37)$$

dove $W_1(\cdot)$ e $W_2(\cdot)$ sono moti Browniani standard indipendenti. Entrambi i modelli utilizzati per la volatilità assicurano, sotto condizioni di regolarità, che la volatilità non possa diventare negativa.

Nel modello A, $W_d(\cdot)$ rappresenta un moto Browniano frazionario di ordine *d* indipendente da $W_1(\cdot)$ che può essere rappresentato dall'integrale

$$W_d(t) = \int_0^t \frac{(t-s)^2}{\Gamma(d+1)} dW(s) \qquad t > 0$$

dove $W(\cdot)$ è un moto Browniano standard.

Il meccanismo generatore per il processo di volatilità nel modello A è un processo logaritmico Ornstein-Uhlenbeck, guidato da un moto Browniano frazionario che permette al processo di esibire memoria lunga, caratteristica verificata anche empiricamente.

Nel modello B la memoria lunga è stata sostituita da jumps, rappresentati dal processo $q(t)$. In particolare, il processo (37) è un modello CIR (Cox-Ingersoll-Ross) con l'aggiunta di un processo positivo di jumps. L'arrivo di jumps segue un processo di Poisson con intensità $\lambda_0 + \lambda_1\sigma^2(t)$, cioè l'arrivo dei jumps dipende dalla volatilità attraverso il parametro $\lambda_1$. La grandezza dei jumps, $\kappa(t)$, è distribuita come un esponenziale di media $\mu$.

Attraverso la discretizzazione di Eulero ho effettuato un esperimento di simulazione per il processo della volatilità e dei prezzi. In letteratura esistono due diverse alternative per farlo:

i. si replicano tante traiettorie tenendo fisso il giorno, ossia si calcola la volatilità in un solo giorno per diverse serie dei prezzi, come Zhangh *et al*.(2005);
ii. si replica una traiettoria in tanti giorni e quindi si calcola la volatilità in ogni giorno per la singola serie dei prezzi, come Nielsen e Frederiksen (2006).

Ho deciso di intraprendere la seconda strada ed ho considerato l'intero arco delle 24 ore come giornata di scambio ottenendo cosi *n* osservazioni giornaliere, dove il valore di *n* cambia a seconda della frequenza di campionamento ed ho ripetuto questa procedura per T=100 giorni.

Per valutare la performance dei diversi metodi di stima ho calcolato per ogni metodo e per ogni giorno l'errore statistico relativo

$$\pi_t = \frac{\hat{\sigma}_t^{2*} - \sum_{i=1}^{n}\left(p_{t,n} - p_{t,n-1}\right)^2}{\sum_{i=1}^{n}\left(p_{t,n} - p_{t,n-1}\right)^2} \qquad (38)$$

dove $\hat{\sigma}_t^{2*}$ rappresenta lo stimatore utilizzato e $\sum_{i=1}^{n}\left(p_{t,n} - p_{t,n-1}\right)^2$ è la volatilità nel giorno $t$ calcolata sui prezzi registrati ad ogni minuto assunta come *benchmark* per la volatilità integrata. Considero per i confronti le quantità *bias* e RMSE definite come

$$Bias = \bar{\pi} = \frac{1}{T}\sum_{t=1}^{T}\pi_t \qquad (39)$$

$$RMSE = \left( \frac{1}{T} \sum_{t=1}^{T} \pi_t^2 \right)^{1/2} \tag{40}$$

Bisogna notare che poiché dividiamo per la volatilità in (38), il *bias* è un *bias* relativo e non assoluto.

## 4.2 IMPLEMENTAZIONE

Cercherò ora di procedere per passi alla spiegazione di tutte le procedure utilizzate.

Ho creato una routine con il software Ox che mi potesse generare dati dai due modelli sopra indicati. Per simulare il moto browniano frazionario ho invece utilizzato il software statistico R seguendo la letteratura trovata in proposito (Marinucci & Robinson (1999)).

Per la scelta dei valori dei parametri da utilizzare per la simulazione del processo di volatilità mi sono basata su risultati di stime precedenti, in particolare sulle stime di Andersen *et al*. (2002) ed Eraker (2004).

I valori dei parametri per la simulazione del modello A in (36) sono ispirati all'analisi dello S&P 500 nel lavoro di Andersen *et al*. (2002) che trova come valori dei parametri stimati $(\hat{\alpha}, \hat{\beta}, \hat{\gamma}) = (0.0062, -1, 0.0374)$. In particolare ho usato $\alpha \in \{0, 0.0062\}, \beta = -1, \gamma \in \{0.0374, 0.1122\}$. Inoltre c'è evidenza empirica che le serie finanziarie della volatilità esibiscono memoria lunga e quindi ho deciso di considerare come valori per il parametro di memoria lunga del moto browniano frazionario $d = \{0.30, 0.45\}$, scelta giustificata da studi precedenti di Andersen *et al.* (2001)

e Andersen *et al.* (2001,2003). I valori iniziali sono stati scelti come $p(0) = \ln(100)$ e $\sigma(0) = e^{\beta}$.

Per il modello B, mi sono basata sui risultati ottenuti da Eraker (2004) sulla stima della volatilità dello S&P 500, il quale trova i seguenti parametri stimati $(\hat{\alpha}, \hat{\beta}, \hat{\gamma}) = (0.023, 0.943, 0.137)$. Gli stessi valori sono stati utilizzati per il modello in (37). Per il processo dei jumps, Eraker (2004) trova le stime $(\hat{\mu}, \hat{\lambda}_0, \hat{\lambda}_1) = (1.53, 0.002, 1.298)$ mentre nella mia simulazione ho usato $\mu \in \{0.7515, 1.53\}$, $\lambda_0 = 0.002$ e $\lambda_1 \in \{0.1298, 2.596\}$. Come valori iniziali sono stati usati $p(0) = \ln(100)$ e $\sigma(0) = \beta$.

Una volta generati i modelli per la volatilità li ho utilizzati per simulare prezzi ogni minuto dal modello (35) per un totale di 1440 osservazioni al giorno, ho poi effettuato un operazione di selezione dal campione originale estraendo i prezzi ogni 5 e 15 minuti. Infine ho introdotto un errore di microstruttura a media nulla e varianza 0.005^2 e uno con una struttura AR(1) con coefficiente autoregressivo pari a 0.5.

Ho quindi calcolato i seguenti stimatori per ogni frequenza di campionamento e in presenza/assenza di errore:

- Varianza realizzata, lo stimatore più utilizzato nelle applicazioni reali
- Lo stimatore di Fourier realizzato usando S=720 che corrisponde a prezzi ogni minuto, S=144 corrispondente ai prezzi ogni 5 minuti e S=48 corrispondente ai prezzi 15 minuti
- Lo stimatore $VR^{(avg)}$ come media degli stimatori calcolati sui sottoinsiemi, tutti di uguale dimensione, realizzati utilizzando i prezzi scelti a partire dalla prima osservazione e pescandoli ogni 5 o 15 osservazioni poi partendo dalla seconda osservazione e cosi via in modo da utilizzare tutti i dati
- Lo stimatore $VR^{(TTSE)}$ottenuto da due scale temporali diverse una slow e una fast; per quanto riguarda la scala temporale più lenta ho

usato i rendimenti a 5 e 15 minuti per quella veloce i rendimenti a 1 minuto

- Infine lo stimatore $VR_{t,1}^{(HL)}$ utilizzato solo per i rendimenti contaminati da errore di microstruttura del mercato; la scelta di questo stimatore non a caso è stata effettuata perché ho inserito un errore con una dipendenza seriale del primo ordine.

Nel prossimo paragrafo metterò a confronto i diversi stimatori in termini di *bias* e RMSE utilizzando le tre diverse frequenze di campionamento.

## 4.3 RISULTATI

Nel calcolare gli errori ho considerato come vero processo generatore dei dati quello a frequenza di campionamento 1 minuto quindi la "vera" volatilità utilizzata per confrontare i diversi stimatori è quella calcolata sui prezzi ad 1 minuto. E' ovvio aspettarsi che *bias* e RMSE della volatilità realizzata per i modelli in assenza di errore e a frequenza di campionamento 1 minuto siano zero visto che la consideriamo come *benchmark*.

Ogni modello è stato analizzato in assenza di errore, con errore IID e con errore AR(1); i risultati sono stati rappresentati in termini di *bias* (prima riga di ogni stimatore) e RMSE (seconda riga di ogni stimatore) per ognuno dei suddetti casi.

Vale la pena ricordare che nell'analisi ci siamo trovati di fronte a due tipi di errore, quello di microstruttura volutamente inserito e quello dovuto alla discretizzazione. Quindi nei modelli in assenza di errore da microstruttura il *bias* riscontrato è da imputare tutto alla discretizzazione mentre nei

modelli con errore di microstruttura una parte del *bias* è dovuta appunto a tale errore e una parte alla discretizzazione del modello.

Come si può notare per tutti i modelli analizzati, in assenza di errore di microstruttura, campionare alla più alta frequenza possibile, in accordo con i risultati teorici, è la cosa migliore infatti si ottengono bias e RMSE più piccoli. Sempre dalla teoria sappiamo che la volatilità realizzata e lo stimatore di Fourier sono stimatori consistenti in assenza di errore e che la costruzione degli altri stimatori è stata appunto effettuata per superare il problema di microstruttura. In conclusione nei modelli privi di errore usare lo stimatore volatilità realizzata o lo stimatore di Fourier è equivalente; qui si è comunque scelto di utilizzare anche gli altri per verificare se il loro comportamento non si discostasse di molto dal vero valore della volatilità. Infatti non possiamo sapere con certezza se nella realtà i prezzi che osserviamo siano quelli reali o se comunque c'è un errore quindi l'utilizzo degli stimatori creati per i prezzi con errore dovrebbe comunque funzionare in modo adeguato anche in assenza di errore.

Aggiungendo l'errore le cose cambiano notevolmente.

Dobbiamo ora fare una distinzione tra il modello A ossia quello guidato da moto Browniano frazionario e il modello B quello che considera la presenza di *jumps* nel processo della volatilità.

Nel primo caso per tutti i modelli presi in considerazione ci troviamo di fronte ad un aumento cospicuo dei valori di *bias* e RMSE, eccetto per lo stimatore TTSE, l'aumento può essere spiegato dal risultato visto al capitolo precedente (formula 19). Tale risultato, infatti, ci fornisce l'ordine di grandezza del *bias* ossia $2nE\left(\varepsilon_{t,i}^2\right)$. Se consideriamo, per esempio, la stima della volatilità realizzata del primo giorno del modello 1) in assenza di errore, 9.6972e-005, e la stessa stima per il modello con errore, 0.069466, notiamo che l'aumento riscontrato corrisponde all'incirca alla quantità menzionata sopra, la differenza è da imputarsi alla

discretizzazione. Ovviamente aumentando la frequenza di campionamento tale aumento si riduce infatti sappiamo che l'errore di microstruttura alle alte frequenze diminuisce, è per questo motivo che la maggior parte delle analisi empiriche viene effettuata con rendimenti a 15 minuti per evitare di incorrere in tale problema. L'unico stimatore, come dimostrato dalla teoria, corretto per la quantità $VR^*$ in tale ambito è TTSE, *Two Time Scales Estimators* infatti otteniamo per ogni modello di tipo A *bias* e RMSE più piccoli.

Esaminando ora il caso con errore AR(1) troviamo una situazione alquanto ambigua. Nonostante l'utilizzo dello stimatore HL costruito effettivamente per tenere in considerazione una correlazione del primo ordine, lo stimatore migliore risulta comunque TTSE. In particolare, lo stimatore HL è il peggiore sia in presenza di errore IID sia in presenza di errore generato da un processo autoregressivo. La sola eccezione si ha nel caso in cui la frequenza di campionamento è di 1 minuto, in quanto è stato possibile confrontarlo solo con gli stimatori VR e FR a causa della costruzione del nostro disegno sperimentale.

Dopo una rassegna sull'utilizzo del miglior stimatore per il modello di tipo A ci concentriamo sulle differenze apportate dall'utilizzo di diversi valori dei parametri. Aumentando $\gamma$ aumenta il *bias* quasi ovunque eccetto per gli stimatori a 15 minuti in assenza di errore e si registra un impatto sulla variabilità di tutti gli stimatori infatti si ha un aumento del RMSE ovunque. Inoltre quando rendiamo il processo molto persistente ossia poniamo $\alpha=0$ troviamo il *bias* più alto.

Poniamo ora la nostra attenzione ai modelli simulati con la presenza di *jumps,* cioè quelli di tipo B. Per i modelli in assenza di errore aumentando la grandezza dei jumps, ossia $\mu$, si indebolisce la performance degli stimatori. Infatti *bias* e RMSE aumentano quando poniamo $\lambda_1=0$ al contrario aumentando l'intensità della volatilità ($\lambda_1=1{,}298$) al variare di $\mu$

*bias* e RMSE diminuiscono. Questo succede anche nel caso in cui si fa variare l'intensità dell'arrivo dei jumps ma solo per la parte relativa alla volatilità, cioè $\lambda_1$ mantenedo fisso il valore di $\mu$. Aumentare infatti tale quantità e fissare $\mu$=0.7515 implica che bias e RMSE aumentano mentre fissando $\mu$=1.53, al contrario, l'aumento dell'intensità dei jumps porta ad una diminuzione di bias e RMSE. Come si può vedere dalle tabelle sottostanti è il modello 12) che si comporta in maniera piuttosto anomala, situazione imputabile probabilmente al caso. Anche per il caso con errore di microstruttura dobbiamo fare due distinzioni a proposito della variazione dei parametri. Nel modello 9) e 11) dove varia $\mu$ e $\lambda_1$=0 l'aumento di bias e Rmse è dello stesso ordine di grandezza come nel caso dei modelli 10) e 12).

In generale per i modelli in presenza di jumps l'aver inserito un errore di microstruttura ha provocato un aumento generale di bias e RMSE che diminuisce con l'aumento della frequenza di campionamento come riscontrato precedentemente. Anche in questo caso il miglior stimatore risulta essere TTSE sia nel caso di errore IID sia nel caso sia generato da un processo AR(1). Inoltre, come per i modelli guidati da moto browniano frazionario, lo stimatore HL costruito effettivamente per affrontare il problema di correlazione non apporta risultati soddisfacenti.

In conclusione possiamo affermare che Two Times Scales Estimators sembra essere lo stimatore preferibile in diverse situazioni.

## 1) Modello A con α=0.0062 β=-1 γ=0.00374 *d*=0.3

| | 1minuto | 5minuti | 15minuti |
|---|---|---|---|
| RV | 0<br>0 | -0.0045113<br>0.069037 | **-0.0048038**<br>0.14877 |
| FR | 1.0680e-006<br>0.0015937 | -0.0041054<br>0.069069 | -0.0054223<br>0.14885 |
| $RV^{(avg)}$ | | **0.0021487**<br>**0.062859** | -0.013073<br>**0.12420** |
| TTSE | | -0.19585<br>0.20662 | -0.079073<br>0.14687 |

Se inserisco errore di microstruttura

| | **1minuto** | **5minuti** | **15minuti** |
|---|---|---|---|
| **RV** | 607.77<br>617.04 | 122.70<br>125.19 | 41.814<br>43.184 |
| **FR** | 609.01<br>618.31 | 124.00<br>126.50 | 42.918<br>44.349 |
| **$RV^{(avg)}$** | | 121.29<br>123.20 | 40.129<br>40.736 |
| **TTSE** | | **-0.46497**<br>**4.4798** | **-0.45533**<br>**1.6482** |
| **HL** | -606.85<br>616.98 | -124.16<br>127.58 | -43.027<br>45.311 |

Se inserisco errore AR(1)

| | **1minuto** | **5minuti** | **15minuti** |
|---|---|---|---|
| **RV** | 405.27<br>411.32 | 155.55<br>160.58 | 56.303<br>58.052 |
| **FR** | 405.73<br>411.79 | 159.01<br>162.04 | 57.769<br>59.573 |
| **$RV^{(avg)}$** | | 156.67<br>159.21 | 53.875<br>54.756 |
| **TTSE** | | **75.414**<br>**76.979** | **26.791**<br>**27.365** |
| **HL** | -202.42<br>206.49 | -153.75<br>157.64 | -58.300<br>60.966 |

## 2) Modello A con α=0.0062 β=-1 γ=0.1122 *d*=0.3

| | 1minuto | 5minuti | 15minuti |
|---|---|---|---|
| RV | 0<br>0 | -0.0045035<br>0.069043 | **-0.0047056**<br>0.14883 |
| FR | 1.1014e-006<br>0.0015934 | -0.0041009<br>0.069075 | -0.0053264<br>0.14892 |
| $RV^{(avg)}$ | | **0.0021700**<br>**0.062870** | -0.012999<br>**0.12425** |
| TTSE | | -0.19583<br>0.20660 | -0.078999<br>0.14687 |

Se inserisco errore di microstruttura

| | **1minuto** | **5minuti** | **15minuti** |
|---|---|---|---|
| **RV** | 633.05<br>642.46 | 127.80<br>130.34 | 43.553<br>44.960 |
| **FR** | 634.35<br>643.78 | 129.15<br>131.70 | 44.703<br>46.174 |
| **$RV^{(avg)}$** | | 126.33<br>128.27 | 41.799<br>42.418 |
| **TTSE** | | **-0.48154**<br>**4.6658** | **-0.47014**<br>**1.7103** |
| **HL** | -632.12<br>642.46 | -129.31<br>132.80 | -44.802<br>47.156 |

Se inserisco errore AR(1)

| | **1minuto** | **5minuti** | **15minuti** |
|---|---|---|---|
| **RV** | 422.12<br>428.25 | 164.11<br>167.20 | 58.647<br>60.454 |
| **FR** | 422.60<br>428.74 | 165.63<br>168.72 | 60.172<br>62.032 |
| **$RV^{(avg)}$** | | 163.17<br>165.75 | 56.116<br>57.015 |
| **TTSE** | | **78.547**<br>**80.140** | **27.909**<br>**28.500** |
| **HL** | -210.86<br>215.04 | -160.15<br>164.14 | -60.722<br>63.495 |

### 3) Modello A con α=0 β=-1 γ=0.0374 *d*=0.3

| | 1minuto | 5minuti | 15minuti |
|---|---|---|---|
| RV | 0<br>0 | -0.0045229<br>0.069049 | **-0.0048512**<br>0.14878 |
| FR | 1.2464e-006<br>0.0015938 | -0.0041164<br>0.069081 | -0.0054711<br>0.14887 |
| $RV^{(avg)}$ | | **0.0021397**<br>**0.062855** | -0.013092<br>**0.12420** |
| TTSE | | -0.19586<br>0.20663 | -0.079092<br>0.14687 |

Se inserisco errore di microstruttura

| | **1minuto** | **5minuti** | **15minuti** |
|---|---|---|---|
| **RV** | 780.58<br>785.82 | 157.51<br>159.27 | 53.685<br>54.975 |
| **FR** | 782.17<br>787.42 | 159.22<br>161.01 | 55.119<br>56.491 |
| **$RV^{(avg)}$** | | 155.72<br>156.82 | 51.558<br>51.928 |
| **TTSE** | | **-0.58874**<br>**5.6570** | **-0.54654**<br>**2.0822** |
| **HL** | -779.05<br>785.11 | -159.22<br>161.94 | -55.097<br>57.456 |

Se inserisco errore AR(1)

| | **1minuto** | **5minuti** | **15minuti** |
|---|---|---|---|
| **RV** | 520.61<br>524.06 | 202.25<br>204.34 | 72.240<br>73.790 |
| **FR** | 521.20<br>524.65 | 204.17<br>206.28 | 74.129<br>75.753 |
| **$RV^{(avg)}$** | | 201.20<br>202.77 | 69.178<br>69.723 |
| **TTSE** | | **96.881**<br>**98.047** | **34.405**<br>**34.845** |
| **HL** | -259.71<br>262.41 | -197.46<br>202.72 | -74.777<br>77.558 |

## 4) Modello A con α=0 β=-1 γ=0.1122 *d*=0.3

| | 1minuto | 5minuti | 15minuti |
|---|---|---|---|
| RV | 0<br>0 | -0.0045160<br>0.069055 | **-0.0047561**<br>0.14885 |
| FR | 1.2848e-006<br>0.0015935 | -0.0041127<br>0.069088 | -0.0053785<br>0.14894 |
| $RV^{(avg)}$ | | **0.0021604**<br>**0.062865** | -0.013019<br>**0.12424** |
| TTSE | | -0.19584<br>0.20661 | -0.079019<br>0.14688 |

Se inserisco errore di microstruttura

| | 1minuto | 5minuti | 15minuti |
|---|---|---|---|
| RV | 823.29<br>829.01 | 166.13<br>168.02 | 56.619<br>57.986 |
| FR | 824.97<br>830.70 | 167.93<br>169.85 | 58.133<br>59.589 |
| $RV^{(avg)}$ | | 164.24<br>165.42 | 54.381<br>54.788 |
| TTSE | | **-0.61963**<br>**5.9638** | **-0.57055**<br>**2.1881** |
| HL | -821.70<br>828.30 | -167.91<br>170.79 | -58.082<br>60.563 |

Se inserisco errore AR(1)

| | **1minuto** | **5minuti** | **15minuti** |
|---|---|---|---|
| **RV** | 549.10<br>552.85 | 213.31<br>215.56 | 76.188<br>77.836 |
| **FR** | 549.72<br>553.47 | 215.34<br>217.61 | 78.178<br>79.902 |
| **$RV^{(avg)}$** | | 212.19<br>213.88 | 72.962<br>73.553 |
| **TTSE** | | 102.17<br>103.41 | **36.290**<br>**36.762** |
| **HL** | -273.93<br>276.87 | -208.27<br>211.75 | -78.849<br>81.799 |

## 5) Modello A con α=0.0062 β=-1 γ=0.00374 *d*=0.45

| | 1minuto | 5minuti | 15minuti |
|---|---|---|---|
| **RV** | 0<br>0 | -0.0045129<br>0.069026 | **-0.0048408**<br>0.14874 |
| **FR** | 1.0715e-006<br>0.0015937 | -0.0041051<br>0.069058 | -0.0054549<br>0.14883 |
| **RV$^{(avg)}$** | | **0.0021393**<br>**0.062849** | -0.013108<br>**0.12419** |
| **TTSE** | | -0.19586<br>0.20663 | -0.079108<br>0.14687 |

Se inserisco errore di microstruttura

| | 1minuto | 5minuti | 15minuti |
|---|---|---|---|
| **RV** | 597.69<br>606.82 | 120.66<br>123.12 | 41.123<br>42.475 |
| **FR** | 598.92<br>608.07 | 121.94<br>124.41 | 42.208<br>43.620 |
| **RV$^{(avg)}$** | | 119.28<br>121.17 | 39.463<br>40.061 |
| **TTSE** | | **-0.45636**<br>**4.4055** | **-0.44918**<br>**1.6243** |
| **HL** | -596.77<br>606.72 | -122.10<br>125.47 | -42.323<br>44.577 |

Se inserisco errore AR(1)

| | 1minuto | 5minuti | 15minuti |
|---|---|---|---|
| **RV** | 398.56<br>404.52 | 154.94<br>157.93 | 55.370<br>57.091 |
| **FR** | 399.01<br>404.98 | 156.38<br>159.36 | 56.813<br>58.589 |
| **RV$^{(avg)}$** | | 154.08<br>156.59 | 52.982<br>53.850 |
| **TTSE** | | **74.170**<br>**75.719** | **26.346**<br>**26.911** |
| **HL** | -199.04<br>203.03 | -151.20<br>155.03 | -57.342<br>59.963 |

## 6) Modello A α=0.0062 β=-1 γ=0.1122 *d*=0.45

| | 1minuto | 5minuti | 15minuti |
|---|---|---|---|
| RV | 0<br>0 | -0.0045084<br>0.069009 | **-0.0048163**<br>0.14876 |
| FR | 1.1130e-006<br>0.0015933 | -0.0040999<br>0.069042 | -0.0054242<br>0.14885 |
| $RV^{(avg)}$ | | **0.0021419**<br>**0.062838** | -0.013103<br>**0.12420** |
| TTSE | | -0.19586<br>0.20662 | -0.079103<br>0.14687 |

Se inserisco errore di microstruttura

| | **1minuto** | **5minuti** | **15minuti** |
|---|---|---|---|
| **RV** | 602.05<br>611.01 | 121.54<br>123.97 | 41.427<br>42.778 |
| **FR** | 603.28<br>612.26 | 122.82<br>125.25 | 42.519<br>43.929 |
| **$RV^{(avg)}$** | | 120.15<br>122.01 | 39.752<br>40.340 |
| **TTSE** | | **-0.45509**<br>**4.4368** | **-0.45124**<br>**1.6368** |
| **HL** | -601.10<br>610.88 | -122.98<br>126.31 | -42.634<br>44.897 |

Se inserisco errore AR(1)

| | **1minuto** | **5minuti** | **15minuti** |
|---|---|---|---|
| **RV** | 401.47<br>407.33 | 156.07<br>159.02 | 55.778<br>57.495 |
| **FR** | 401.93<br>407.79 | 157.52<br>160.46 | 57.231<br>59.003 |
| **$RV^{(avg)}$** | | 155.21<br>157.70 | 53.371<br>54.227 |
| **TTSE** | | **74-720**<br>**76.259** | **26.540**<br>**27.102** |
| **HL** | -200.47<br>204.40 | -152.30<br>156.10 | -57.772<br>60.405 |

## 7) Modello A con α=0 β=-1 γ=0.0374 *d*=0.45

| | 1minuto | 5minuti | 15minuti |
|---|---|---|---|
| RV | 0<br>0 | -0.0045242<br>0.069037 | **-0.0048869**<br>0.14876 |
| FR | 1.2478e-006<br>0.0015937 | -0.0041158<br>0.069069 | -0.0055024<br>0.14885 |
| $RV^{(avg)}$ | | **0.0021306**<br>**0.062845** | -0.013126<br>**0.12419** |
| TTSE | | -0.19587<br>0.20663 | -0.079126<br>0.14688 |

Se inserisco errore di microstruttura

| | 1minuto | 5minuti | 15minuti |
|---|---|---|---|
| RV | 763.76<br>768.87 | 154.12<br>155.84 | 52.532<br>53.789 |
| FR | 765.31<br>770.44 | 155.79<br>157.53 | 53.934<br>55.280 |
| $RV^{(avg)}$ | | 152.38<br>153.44 | 50.446<br>50.806 |
| TTSE | | **-0.57411**<br>**5.5373** | **-0.53684**<br>**2.0419** |
| HL | -762.25<br>768.13 | -155.79<br>158.45 | -53.925<br>56.245 |

Se inserisco errore AR(1)

| | **1minuto** | **5minuti** | **15minuti** |
|---|---|---|---|
| **RV** | 509.40<br>512.77 | 197.90<br>199.95 | 70.688<br>72.208 |
| **FR** | 509.98<br>513.35 | 199.78<br>201.84 | 72.537<br>74.131 |
| **$RV^{(avg)}$** | | 196.88<br>198.42 | 67.690<br>68.222 |
| **TTSE** | | **94.802**<br>**95.951** | **33.664**<br>**34.094** |
| **HL** | -254.09<br>256.71 | -193.21<br>196.40 | -73.182<br>75.911 |

## 8) Modello A con α=0 β=-1 γ=0.1122 *d*=0.45

| | 1minuto | 5minuti | 15minuti |
|---|---|---|---|
| RV | 0<br>0 | -0.0045199<br>0.069020 | **-0.0048631**<br>0.14878 |
| FR | 1.2902e-006<br>0.0015934 | -0.0041108<br>0.069053 | -0.0054723<br>0.14887 |
| $RV^{(avg)}$ | | **0.0021331**<br>**0.062834** | -0.013122<br>**0.12420** |
| TTSE | | -0.19587<br>0.20663 | -0.079122<br>0.14688 |

Se inserisco errore di microstruttura

| | **1minuto** | **5minuti** | **15minuti** |
|---|---|---|---|
| **RV** | 771.11<br>776.32 | 155.60<br>157.35 | 53.042<br>54.329 |
| **FR** | 772.68<br>777.89 | 157.28<br>159.05 | 54.457<br>55.823 |
| **$RV^{(avg)}$** | | 153.85<br>154.94 | 50.933<br>51.299 |
| **TTSE** | | **-0.57414**<br>**5.5919** | **-0.54051**<br>**2.0631** |
| **HL** | -769.57<br>775.53 | -157.28<br>159.97 | -54.445<br>56.799 |

Se inserisco errore AR(1)

| | **1minuto** | **5minuti** | **15minuti** |
|---|---|---|---|
| **RV** | 514.32<br>517.75 | 199.81<br>201.90 | 71.374<br>72.919 |
| **FR** | 514.90<br>518.33 | 201.71<br>203.81 | 73.241<br>74.859 |
| **$RV^{(avg)}$** | | 198.79<br>200.37 | 68.344<br>68.888 |
| **TTSE** | | **95.726**<br>**96.900** | **33.990**<br>**34.429** |
| **HL** | -256.51<br>259.15 | -195.07<br>198.31 | -73.901<br>76.681 |

## 9) Modello B con $\lambda_0$=0.0002 $\lambda_1$=0 $\mu$=0.7515

| | 1minuto | 5minuti | 15minuti |
|---|---|---|---|
| RV | 0<br>0 | 0.0050310<br>0.077579 | -0.010267<br>0.14681 |
| FR | -0.00018744<br>0.0013214 | 0.0042118<br>0.077563 | **-0.0078400**<br>0.14725 |
| $RV^{(avg)}$ | | **-0.0015862**<br>**0.058941** | -0.0086579<br>**0.11944** |
| TTSE | | -0.20159<br>0.21002 | -0.075325<br>0.14095 |

Se inserisco errore di microstruttura IID

| | 1minuto | 5minuti | 15minuti |
|---|---|---|---|
| RV | 113.24<br>113.47 | 22.783<br>22.959 | 7.4660<br>7.5918 |
| FR | 113.54<br>113.77 | 23.084<br>23.270 | 7.6869<br>7.8289 |
| $RV^{(avg)}$ | | 22.561<br>22.602 | 7.4760<br>7.4896 |
| TTSE | | **-0.28728**<br>**0.89705** | **-0.13999**<br>**0.37563** |
| HL | -113.37<br>113.81 | -22.883<br>23.215 | -7.4854<br>7.7559 |

Se inserisco errore AR(1)

| | **1minuto** | **5minuti** | **15minuti** |
|---|---|---|---|
| **RV** | 75.403<br>75.527 | 29.639<br>29.830 | 9.9860<br>10.150 |
| **FR** | 75.517<br>75.642 | 29.999<br>30.193 | 10.343<br>10.539 |
| **$RV^{(avg)}$** | | 29.081<br>29.146 | 10.016<br>10.038 |
| **TTSE** | | **13.800**<br>**13.891** | **4.9221**<br>**4.9520** |
| **HL** | -37.978<br>38.288 | -28.855<br>29.207 | -9.9468<br>10.301 |

## 10) Modello B con $\lambda_0$=0.0002 $\lambda_1$=1,298 $\mu$=0.7515

| | 1minuto | 5minuti | 15minuti |
|---|---|---|---|
| RV | 0<br>0 | -0.010232<br>0.078634 | -0.023381<br>0.12946 |
| FR | -0.00021953<br>0.0015955 | -0.0096554<br>0.078879 | -0.023502<br>0.13583 |
| $RV^{(avg)}$ | | **-0.0048665**<br>**0.056487** | **-0.018160**<br>**0.10870** |
| TTSE | | -0.20487<br>0.21246 | -0.084827<br>0.13668 |

Se inserisco l'errore di microstruttura

| | **1minuto** | **5minuti** | **15minuti** |
|---|---|---|---|
| **RV** | 1.0377<br>9.2825 | 0.20229<br>1.9487 | 0.029427<br>0.48767 |
| **FR** | 1.0388<br>9.2903 | 0.20475<br>1.9661 | 0.031220<br>0.50734 |
| **$RV^{(avg)}$** | | 0.19739<br>1.8261 | **0.051558**<br>0.64191 |
| **TTSE** | | **-0.21014**<br>**0.21917** | -0.084287<br>**0.13780** |
| **HL** | -1.0588<br>9.5541 | -0.24856<br>2.1187 | -0.091817<br>0.63043 |

Se inserisco errore AR(1)

| | **1minuto** | **5minuti** | **15minuti** |
|---|---|---|---|
| **RV** | 0.68250<br>6.0897 | 0.24469<br>2.3298 | 0.059864<br>0.77402 |
| **FR** | 0.68293<br>6.0914 | 0.24470<br>2.3231 | 0.060295<br>0.77778 |
| **$RV^{(avg)}$** | | 0.25422<br>2.3395 | -0.072244<br>0.82882 |
| **TTSE** | | **-0.082282**<br>**1.1192** | **-0.039923**<br>**0.43104** |
| **HL** | -0.36370<br>3.2888 | -0.27453<br>2.3635 | -0.12179<br>0.88977 |

## 11) Modello B con $\lambda_0$=0.0002 $\lambda_1$=0 $\mu$=1.53

| | 1minuto | 5minuti | 15minuti |
|---|---|---|---|
| RV | 0<br>0 | -0.0059228<br>0.073111 | -0.029983<br>0.14840 |
| FR | 0.00021371<br>0.0016235 | -0.0055826<br>0.073281 | -0.029478<br>0.15169 |
| $RV^{(avg)}$ | | **-0.0049222**<br>**0.060456** | **-0.019114**<br>**0.12756** |
| TTSE | | -0.20492<br>0.21360 | -0.085781<br>0.15253 |

Se inserisco l'errore di microstruttura

| | 1minuto | 5minuti | 15minuti |
|---|---|---|---|
| RV | 113.12<br>113.35 | 22.723<br>22.869 | 7.4741<br>7.6067 |
| FR | 113.32<br>113.55 | 22.920<br>23.068 | 7.7153<br>7.8687 |
| $RV^{(avg)}$ | | 22.611<br>22.662 | 7.4593<br>7.4792 |
| TTSE | | **-0.21260**<br>**0.84061** | **-0.14875**<br>**0.36250** |
| HL | -112.85<br>113.21 | -22.748<br>23.058 | -7.5519<br>7.8466 |

Se inserisco errore AR(1)

| | **1minuto** | **5minuti** | **15minuti** |
|---|---|---|---|
| **RV** | 75.485<br>75.633 | 29.034<br>29.228 | 9.7280<br>9.9246 |
| **FR** | 75.556<br>75.704 | 29.325<br>29.517 | 10.041<br>10257 |
| **$RV^{(avg)}$** | | 29.255<br>29.334 | 9.9420<br>9.9739 |
| **TTSE** | | **13.958**<br>**14.046** | **4.8430**<br>**4.8839** |
| **HL** | -37.623<br>37.845 | -27.926<br>28.264 | -9.7532<br>10.167 |

## 12) Modello B con $\lambda_0$=0.0002 $\lambda_1$=1.298 $\mu$=1.53

| | 1minuto | 5minuti | 15minuti |
|---|---|---|---|
| RV | 0<br>0 | -0.0015345<br>0.062808 | **-0.0071939**<br>0.14164 |
| FR | -0.00014941<br>0.0017698 | 0.0019735<br>0.064897 | -0.0097460<br>0.14310 |
| $RV^{(avg)}$ | | **-0.0025786**<br>**0.053761** | -0.011317<br>**0.11265** |
| TTSE | | -0.20258<br>0.20958 | -0.077984<br>0.13654 |

Se inserisco errore di microstruttura

| | 1minuto | 5minuti | 15minuti |
|---|---|---|---|
| RV | 1.4778<br>10.184 | 0.29174<br>2.0113 | 0.093109<br>0.70651 |
| FR | 1.4779<br>10.201 | 0.29446<br>2.0118 | 0.093459<br>0.71637 |
| $RV^{(avg)}$ | | 0.28137<br>1.9442 | 0.084841<br>0.67385 |
| TTSE | | **-0.21418**<br>**0.24195** | **-0.080344**<br>**0.13875** |
| HL | -1.4915<br>10.331 | -0.30060<br>2.0107 | -0.10363<br>0.68273 |

Se inserisco errore AR(1)

| | **1minuto** | **5minuti** | **15minuti** |
|---|---|---|---|
| **RV** | 0.97864<br>6.7234 | 0.37243<br>2.6106 | 0.11612<br>0.89344 |
| **FR** | 0.97935<br>6.7299 | 0.37755<br>2.6304 | 0.11659<br>0.90378 |
| **$RV^{(avg)}$** | | 0.35683<br>2.4758 | 0.11463<br>0.86547 |
| **TTSE** | | **-0.038896**<br>**1.1210** | **-0.017284**<br>**0.42293** |
| **HL** | -0.51479<br>3.5775 | -0.38402<br>2.6775 | -0.12348<br>0.82455 |

# CAPITOLO QUINTO

# APPLICAZIONI SU DATI REALI

## 5.1 I DATI

In questa sezione ho applicato gli stimatori utilizzati nel capitolo precedente ad una serie storica finanziaria reale: l'indice di Borsa MIB30. I dati utilizzati in questo studio provengono dalla Borsa Valori di Milano. Si tratta di dati infragiornalieri a frequenza 1 minuto per il periodo dal 1 Settembre 2001 al 28 Febbraio 2002. Si hanno quindi a disposizione 123 giorni di dati ad alta frequenza con un numero di osservazioni diverse all'interno di ciascun giorno. Il primo passo dell'analisi, per una questione di implementazione computazionale, è stato quello di creare una serie avente numero uguale di osservazioni all'interno di ogni giorno. Prima di procedere ad un'analisi della stima della volatilità mi sembra doveroso dare una descrizione dell'indice analizzato.

Il MIB30 è un indice di borsa che racchiude le quotazioni delle trenta principali società che hanno maggior peso nell'economia Italiana, descrivendo l'andamento complessivo dei titoli più rappresentativi

selezionati in base all'indicatore ILC (Indice di Liquidità e Capitalizzazione) più alto.

Una prima analisi grafica dell'indice nel periodo considerato evidenzia un andamento che oscilla tra 30000 e 33000 punti indice ad eccezione del mese di Settembre. In quell'anno infatti si sono verificati gli attacchi terroristici alle Torri Gemelle di New York che hanno portato al crollo di tutte le Borse Valori mondiali. Nel suddetto mese l'indice ha registrato un continuo calo fino a toccare il suo valore minimo, 23541 punti indice, in data 21 Settembre 2001 per poi risalire gradualmente verso i valori precedenti l'attentato.

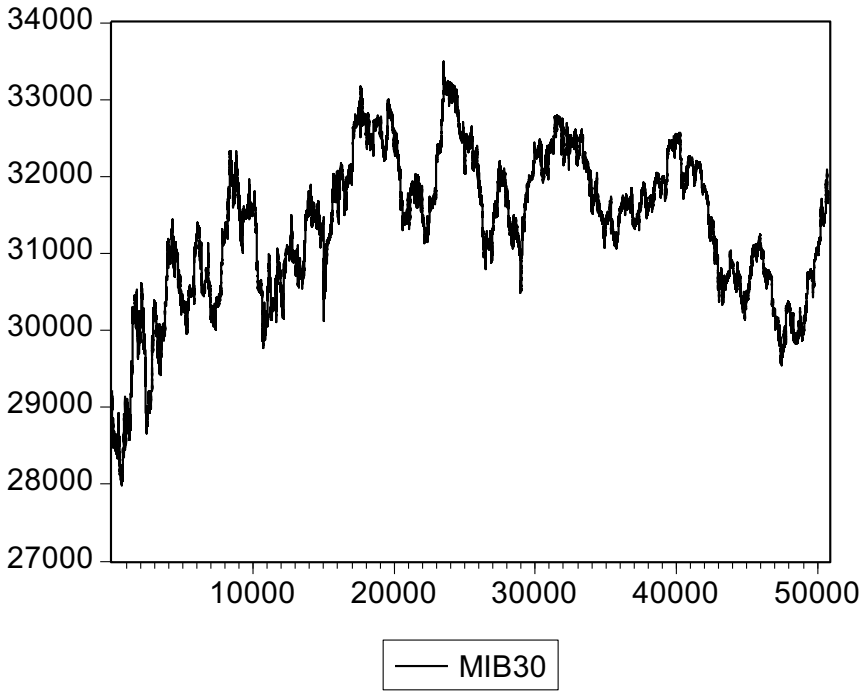


**Figura 1: Andamento del valore assunto dall'indice MIB30 dal settembre 2001 al febbraio 2002, dati infragiornalieri**

Il mese di settembre potrebbe portare ad una distorsione della stima in quanto è un mese anomalo rispetto all'andamento della serie. Ho quindi eliminato tale mese ed eseguito il resto dell'analisi sul campione ridotto.

## 5.2 ANALISI DESCRITTIVA

L'analisi dei prezzi presenta alcune peculiarità statistiche, quali la non stazionarietà. Per ottenere informazioni più 'fini' sulle caratteristiche statistiche delle attività finanziarie, si preferisce solitamente lavorare sui rendimenti.

La distribuzione di frequenza dei rendimenti del MIB30, raffigurata in seguito, consente di mettere in evidenza alcune caratteristiche salienti che ricorrono nello studio dei mercati finanziari.

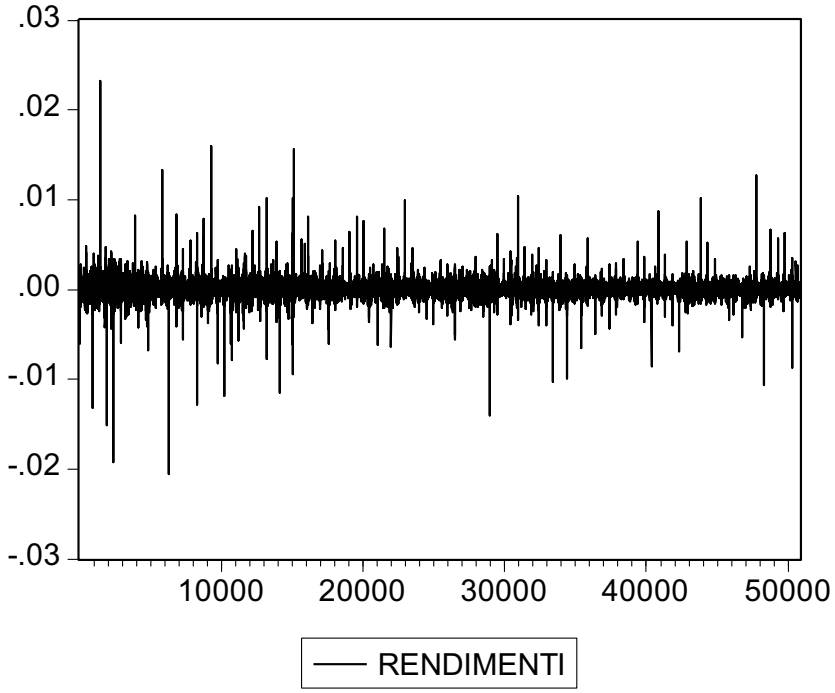


**Figura 2: Rendimenti del MIB30**

Per un analisi descrittiva più dettagliata riportiamo di seguito il relativo istogramma unitamente al calcolo di misure sintetiche della distribuzione dei rendimenti.

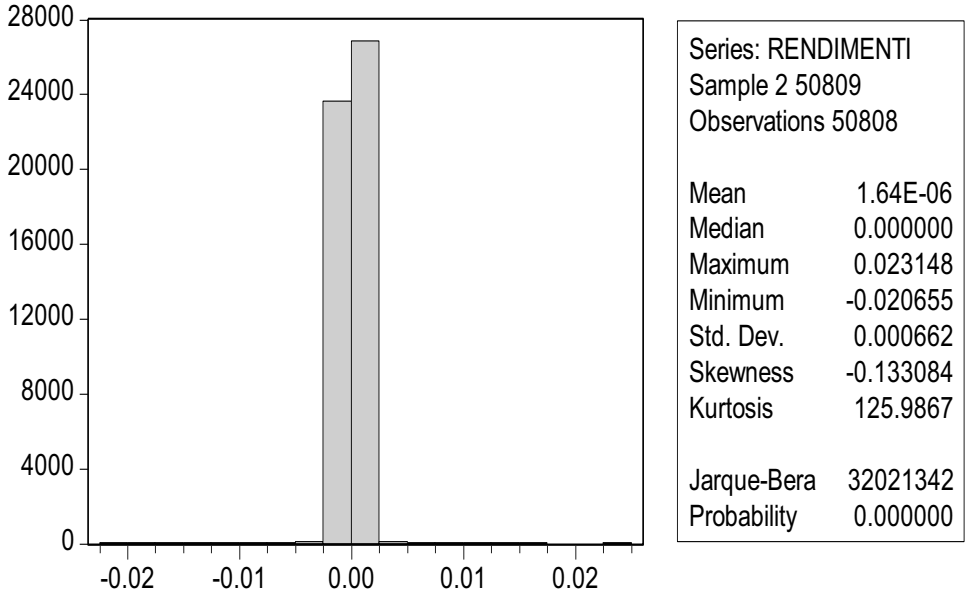


**Figura 3: Istogramma dei rendimenti intragiornalieri del MIB30**

Analizzando la serie e guardando gli indici che misurano l'allontanamento dalla distribuzione normale si può rilevare un valore di asimmetria pari a -0.133084, ovvero una presenza di asimmetria negativa. La curtosi della distribuzione come è naturale aspettarsi è maggiore di tre ciò significa che la distribuzione è leptocurtica, quindi con code più pesanti della normale.

Per verificare la plausibilità o meno dell'ipotesi generatore gaussiano si utilizza un usuale verifica di ipotesi statistica basata appunto sulle caratteristiche di asimmetria e curtosi. Questa è la statistica test di Jarque-Bera.

Nel caso descritto, il valore osservato di tale test è molto elevato e il relativo p-value indica un inevitabile rigetto dell'ipotesi nulla di normalità.

Procediamo ora nel paragrafo seguente ad un'analisi sulla nostra reale quantità di interesse, la volatilità.

## 5.3 STIMA DELLA VOLATILITA'

In questa sezione verranno presentati i risultati delle stime della volatilità integrata utilizzando diversi stimatori. Sottolineiamo quindi che lo scopo di questo lavoro è quello di verificare la *performance* di diversi stimatori per la volatilità in termini di *bias* e RMSE.

In linea con quanto effettuato per le simulazioni abbiamo come prima cosa effettuato un analisi sulla serie dei prezzi ad 1 minuto e poi selezionato un sottocampione a 5 minuti e un altro a 15 minuti. Abbiamo quindi calcolato per ogni frequenza di campionamento i relativi stimatori:

- Volatilità realizzata
- Fourier
- Media delle volatilità realizzate nei sottocampioni ($VR^{(avg)}$)
- Two Times Scales Estimators
- Hansen & Lunde

La costruzione dei suddetti stimatori è già stata ampliamente descritta in precedenza quindi di seguito saranno riportati solamente i relativi *bias* e RMSE calcolati come dalle formule (39) e (40). Anche in questo caso è stato considerato come *benchmark* per la 'vera' volatilità quella calcolata sui prezzi ad 1 minuto di conseguenza il *bias* e RMSE risultano zero.

| | 1minuto | 5minuti | 15minuti |
|---|---|---|---|
| **RV** | 0<br>0 | 0.084874<br>0.28454 | 0.019478<br>0.31622 |
| **FR** | -0.00040468<br>0.0024248 | **0.082697**<br>0.28580 | **0.016563**<br>0.31855 |
| **$RV^{(avg)}$** | | 0.11017<br>0.23258 | 0.093968<br>0.27653 |
| **TTSE** | | -0.087887<br>**0.22279** | 0.027949<br>**0.26142** |
| **HL** | 0.091104<br>0.21944 | 0.13299<br>0.34584 | 0.086209<br>0.48742 |

**Tabella 1: Bias e RMSE per stime della volatilità sull'indice MIB30**

Dall'analisi dei risultati contenuti nella tabella si evince che aumentando la frequenza di campionamento tutti gli stimatori presentano una diminuzione del *bias*, in accordo con la teoria a riguardo. A causa della costruzione del nostro campione possiamo effettuare un confronto adeguato solo per le frequenze di campionamento a 5 e 15 minuti.

Per tali frequenze se consideriamo come criterio il *bias* lo stimatore migliore è quello di Fourier ma quando si utilizza come indicatore per il paragone RMSE la scelta ricade su TTSE.

In conclusione possiamo confermare i risultati trovati in precedenza attraverso l'esercizio di simulazione ed utilizzare lo stimatore TTSE in quanto preferibile in diverse situazioni.

# CAPITOLO SESTO

# CONCLUSIONI

In questo lavoro ho affrontato il problema della stima della volatilità integrata di attività finanziarie sfruttando dati infragiornalieri ad altissima frequenza.

Come ho cercato di spiegare nel capitolo 2, i dati ad altissima frequenza sono soggetti ad errori dovuti alla microstruttura del mercato, ovvero l'insieme dei meccanismi che regolano gli scambi.

Di un certo rilievo sono gli errori dovuti a *bid-ask spread*, scambi non sincronizzati o assenza di scambi e discretezza dei prezzi.

E' prassi comune stimare la volatilità integrata tramite la volatilità realizzata, ovvero la somma dei rendimenti al quadrato nell'arco di un intervallo di tempo giornaliero.

Tale stima è soddisfacente nel caso di assenza di errore di microstruttura, può invece incontrare degli inconvenienti in presenza di errore.

In particolare più aumenta la frequenza di campionamento e più tale quantità si discosta dalla vera quantità da stimare poiché aumenta la possibilità di incorrere nell'errore di microstruttura.

Per ovviare a questi inconvenienti ho studiato altri stimatori robusti rispetto all'errore. In particolare ho studiato stimatori che cercano di tener conto

dell'errore sotto diverse specificazioni (ad esempio errori IID o errori auto correlati).

Per verificare le caratteristiche di questi stimatori ho condotto un esperimento di simulazione basandomi su differenti specificazioni di modelli a tempo continuo per la volatilità dei prezzi. In questo senso ho considerato un modello con moto Browniano frazionario ed uno con jumps.

I risultati delle simulazioni sembrano fornire evidenza verso la scelta dello stimatore TTSE robusto a errori di microstruttura. Inoltre a conferma di quanto riportato in letteratura, ho riscontrato che aumentare la frequenza di campionamento in presenza di errore provoca un aumento del *bias*, ma in assenza di errore è la cosa migliore da fare in quanto ci restituisce una stima più accurata per la volatilità.

I risultati delle applicazioni reali, infine, sembrano confermare quanto trovato in sede di simulazione.

# BIBLIOGRAFIA

Ait-Sahaila Y., Mykland A. e Zhang L. (2005), 'How often to sample a continuos-time process in the presence of market microstructure Noise', *The review of Financial Studies,* 18, 351-416.

Ait-Sahaila Y., Mykland A. e Zhang L. (2006), 'Ultra high frequency volatility estimation with dependent microstructure noise', *Working Paper 11380, National Bureau of Economic Research.*

Andersen T.G., Benzoni L. e Lund J. (2002), 'An empirical investigation of continuous-time equity return models', *Journal of Finance*, 57, 1239-1284.

Andersen T.G. e Bollerslev T. (1998), 'Answering the skeptics: yes, standard volatility models do provided accurate forecasts', *International Economic Review,* 39, 885-905.

Andersen T.G., Bollerslev T., Diebold F.X. e Ebens H. (2001), 'The distribution of realized stock return volatility', *Journal of Financial Economics* 61, 43-76.

Andersen T.G., Bollerslev T., Diebold F.X. e Labys P. (2001), 'The distribution of realized stock return volatility', *Journal of the American Statistical Association,* 96, 42-55.

Andersen T.G., Bollerslev T., Diebold F.X. e Labys P. (2003), 'Modelling and forecasting realized volatility', *Econometrica 71,* 579-625.

Amihud Y., Mendelson Haim (1987), 'Trading mechanism and stock returns: an empirical investigation', *Journal of Finance,* 42, 533-553.

Bandi F.M. e Russel J.R. (2003), 'Microstructure noise, realized volatility and optimal sampling', *Journal of the American Association,* 96, 42-55.

Bandi F.M. e Russel J.R. (2005) 'Realized covariation, realized beta, and microstructure noise', *Working paper*, *Graduate School of Business, The University of Chicago.*

Barndoff-Nielsen O.E. e Shepard N. (2001), 'Non-Gaussian Ornstein-Uhlenbeck-based models and some of their uses in financial economics', *Journal of the Royal Statistical Society S, B,* 63,167-241.

Barndoff-Nielsen O.E., Hansen P.R., Lunde A. e Shepard N. (2002), 'Econometric analysis of realized volatility and its use in estimating stochastic volatility models', *Journal of the Royal Statistical Society, B,* 64,253-280.

Barndoff-Nielsen O.E., Hansen P.R., Lunde A. e Shepard N. (2006), 'Designing realized Kernels to measure the ex-post variation of equity prices in the presence of noise', *Working paper, Stanford University.*

Barucci E. e Renò R. (2002a), 'On measuring volatility and the GARCH forecasting performance', *Journal of international Financial Markets, Institutions and Money,* 12, 183-200.

Barucci E. e Renò R. (2002b), ‘On measuring volatility of diffusion processes with high frequency data’, *Economics Letters*, 74, 371-378.

Eraker B. (2004), ‘Do stock prices and volatility jump? Reconciling evidence from spot and option prices’*, Journal of Finance,* 59*,*1269-1300.

Engle R. F. e Russell J. R. (1998), ‘Autoregressive conditional duration: a new model for Irregularly spaced transaction data, *Econometrica*, 66, 1127-1162

Griffin J.E. e Oomen R.C.A. (2006) ‘Sampling returns for realized variance calculation: tick time or transaction time?, *Forthcoming Econometrics Reviews.*

Glosten Lawrence R.,(1987) ‘Components of the bid-ask spread and the statical properties of transaction prices’, *Journal of Finance,* 42, 1293-1307.

Hansen P. e Lunde A. (2005), ‘A realized variance for the whole day based on intermittent high-frequency data’, *Journal of Financial Econometrics,* 3, 525-554.

Hansen P. e Lunde A. (2006), ‘Realized variance and market microstructure noise’, *Journal of Business & Economic Statistics,* 24, 127-218.

Hasbrouck J. (2004), ‘Empirical market microstructure: Economic and statistical perspectives on the dynamics of trade in securities markets’, *Lecture notes*, *Stern School of Business, new York university.*

Jacod J. e Protter P. (1998), ‘Asymptotic error distribution for the Euler method for stochastic differential equation’, *Annals of Probability,* 26, 267-307.

Madhavan A. (2000), ‘Market micristructure: a survey’, *Journal of Financial Markets,* 3, 205-258.

Malliavian P. e Mancino M.E. (2002), ‘Fourier series method for measurement of multivariates volatilities’, *Finance and Stochastics,* 6, 49-61.

Malliavian P. e Mancino M.E. (2005), ‘A Fourier transform method for nonparametric estimation of volatility’, *Working paper, Department of Mathematics for Decision, University of Firenze.*

Mancino M.E e Sanfelici S. (2007), ‘Robustness of Fourier estimator of integrated volatilità in the presence of microstructure noise’, *Computational statistics & data analysis,*52, 2966-2989*.*

Marinucci D. e Robinson P.M. (1999), ‘ Alternative forms of fractional Brownian motion’, *Journal of Statistical Planning and Inference*, 80, 111-122.

McAleer M, e Medeiros M.C (2007), ‘ Realized volatility: A review’, *Working Paper.*

Newey W.K. e West K.D. (1987), ‘A simple semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix, *Econometrica,* 55, 703-708*.*

Nielsen M.O. e Frederiksen P.H. (2006), ‘Finite sample accuracy and choice of sampling in integrated volatility estimation’*, Working Paper, Cornell University.*

Oomen Roel C.A., (2006) ‘Properties of bias-corrected realized Variance Under Alternative Sampling Schemes’, *Journal of Financial Econometrics,* 3, 555-577*.*

Oomen Roel C.A., (2006) ‘Properties of realized Variance Under Alternative Sampling Schemes’, *Journal of Financial Business and Economic Statistics,* 24, 219-237*.*

O’Hara M. (1995), ‘*Market Microstructure Noise’*, Cambridge, MA: Blackwell.

Roll R. (1984), ‘A simply implicit meausure of the effective bid-ask spread in a efficient market’, *Journal of Finance*, 39, 1127-1139.

Zhang L., Mykland P.A. e Ait-Sahalia Y. (2005), ‘A Tale of Two Svales: Determining Integrated Volatility with Noisy High-Frequency Data’, *Journal of the American Statistical Association,* 100,1394-1411.

Zhou B. (1996), “ High frequency data and volatility in foreign-exchange rates’, *Journal of Business and Economics Statistics,* 14, 45-52.

# RINGRAZIAMENTI

Non ci posso ancora credere ho finito!!!! Se sono arrivata a questo punto significa che ho terminato la mia tesi ma anche l'università….e purtroppo la vita da studente! Ormai pensavo fosse un traguardo irraggiungibile invece come sempre ce l'ho fatta, da sola, con fatica ma tanta soddisfazione. Tutti avete sempre avuto grande fiducia in me anche se io come sempre non credevo nelle mie capacità, non mi resta quindi che ringraziarvi di cuore.

In particolare al primo posto devo mettere la mia cara mammina (come la chiamo io), Ketty, per avermi sempre sostenuto e appoggiato in ogni mia scelta in tutti questi anni universitari, mio fratello, Manuel, perché insieme a lei non ha mai smesso di credere in me e nelle mie capacità e mia nonna che non mi ha mai fatto mancare la sua presenza.

La mia amica di una vita, di viaggi, di divertimenti, di chiacchierate….la mia sorellina Carla, grazie per avermi fatto conoscere il Venezuela, per avermi fatto "scappare" da Mestre, Padova,…dall'Italia, gracias por todos.

La mia compagna … in questi ultimi due anni universitari, Eris, per avermi sopportata quando scleravo, per avermi coccolata nei momenti tristi, per essermi stata vicina in tutti questi mesi di tesi (lo so, è stata dura!!!!)

I miei amici di Mestre: Lara, Andrea, Diego, Luca, Chiara, Arcangelo, Giorgio, Consu, Enrica per le belle serate tra bevute e chiacchierate….(ho fatto anche la rima). Un posticino particolare per Morris che ha sempre

creduto in me e mi ha accompagnato in tutti questi anni tra difficoltà e felicità.

I miei nuovi amici di Padova, i ragazzi della Goito, come poter dimenticare le notti passate a studiare tra caffè e sigaretta ma anche le feste e le bevute in piazza.

I miei colleghi dell'università, compagni di studio in questi ultimi due anni ma anche di feste, spritz, compleanni: AleRich, Sumy, Luc, De luchi, Oneda, Giulia, Stefano, Fabio 28, Manuela, Ilaria, Riccardo, Damiano, Roberto, Ombretta, Silvia e Angela.

Infine ma non per minor importanza un ringraziamento particolare va al mio prof. Silvano Bordignon e a Davide Raggi.

GRAZIE A TUTTI DI CUORE