



**Università Politecnica delle Marche**

---

FACOLTÀ DI ECONOMIA "GIORGIO FUÀ"  
Corso di Laurea Magistrale in Scienze Economiche e Finanziarie

## **Serie finanziarie ad alta frequenza e Flash Crash**

**High frequency financial series and Flash Crashes**

Relatore:  
**Ch.mo Prof. Giulio Palomba**

Tesi di Laurea di:  
**Aleandro Maria Rossignoli**

# Indice

<b>Introduzione</b>	<b>4</b>
<b>1 Analisi preliminare</b>	<b>7</b>
1.1 I dati . . . . .	11
1.2 Modelli di volatilità condizionale . . . . .	15
<b>2 Le Flash Crash</b>	<b>21</b>
2.1 La Flash Crash del Maggio 2010 . . . . .	21
2.2 High Frequency Trading e Flash Crash . . . . .	26
2.2.1 La velocità come chiave del successo . . . . .	29
2.2.2 Strategie operative degli HFTr e impatto sui mercati . . . . .	34
2.3 La regolamentazione del trading ad alta frequenza . . . . .	44
2.4 Mini Flash Crash . . . . .	49
2.5 Order flow toxicity . . . . .	54
<b>3 VPIN metric</b>	<b>58</b>
3.1 PIN vs VPIN . . . . .	59
3.2 La procedura di calcolo del VPIN . . . . .	62
3.3 La funzione di ripartizione del VPIN . . . . .	68

<i>INDICE</i>	3
3.4 Applicazione sui dati del FTSE 100 . . . . .	71
3.5 VPIN e Flash Crash . . . . .	75
3.6 Il VPIN nel campione completo . . . . .	83
3.7 L’Affidabilità dell’indicatore VPIN . . . . .	87
<b>4 Conclusioni</b>	<b>92</b>
<b>A Stime preliminari</b>	<b>94</b>
A.1 Modelli di volatilità condizionale . . . . .	94
A.2 Volatilità Realizzata . . . . .	99
<b>B VPIN</b>	<b>102</b>
B.1 Distribuzione di Burr . . . . .	102
B.2 Periodi di criticità e affidabilità del VPIN . . . . .	103

# Introduzione

Il crescente sviluppo tecnologico, nella prima decade del ventunesimo secolo, ha permesso la diffusione sui mercati finanziari di tecniche di negoziazione ad alta frequenza. Con l'*High Frequency Trading* ci si riferisce in particolare a quella modalità operativa incentrata su un'elevata velocità nella gestione delle informazioni di mercato e nella reazione alle stesse.

La rapida diffusione di questa particolare strategia di negoziazione, ha portato i mercati ad essere via via sempre più saturi di operatori, cosiddetti ad alta frequenza (*HFT*). Sebbene molti fossero i benefici riconosciuti all'attività svolta da tali soggetti, soprattutto in relazione al sostegno della liquidità dei mercati, nel Maggio del 2010 vennero alla luce drammaticamente anche tutti i problemi ad essa connessi. Ci si accorse infatti che il crescente grado di automazione dei mercati, poteva generare gravi situazioni di instabilità conosciute col nome di "Flash Crash" (letteralmente "*Crollo lampo*"). Il termine si diffuse in realtà a partire dal 6 Maggio 2010, giorno in cui il Dow Jones Industrial Average sperimentò una perdita di circa 1000 punti in poco meno di 10 minuti. Di fatto una Flash Crash non è altro che una crisi di liquidità, che si produce in periodi di tempo molto ristretti, a causa dell'elevato livello di reattività dei mercati alle tensioni, dovuto alla presenza massiccia

di operatori HFTr.

Seppur eventi come quelli del Maggio 2010 siano al giorno d'oggi rari, grazie all'elevato grado di regolamentazione e supervisione dei mercati, continuano comunque ad essere presenti crolli repentini dei prezzi di minore entità noti come "Mini Flash Crash".

In un contesto operativo del genere, analizzare serie finanziarie ad alta frequenza significa confrontarsi con la possibilità di incorrere in situazioni di volatilità estrema.

Si dispone in particolare di un campione di dati ad alta frequenza che descrive l'andamento delle quotazioni dell'indice azionario inglese FTSE 100, nel periodo storico compreso tra il 4 Gennaio 2011 e il 31 Maggio 2012. Per tale orizzonte temporale si è in possesso di osservazioni dei prezzi a differenti frequenze di rilevazione (che vanno dai 300 secondi ad 1 secondo). La cronaca ha portato alla luce il mese di Agosto 2011 come un periodo particolarmente turbolento per i mercati internazionali a causa delle tensioni legate alla diffusione della crisi del debito sovrano europea: in particolare l'indice UK, oggetto dell'analisi, subisce, nei primi giorni del mese, una delle perdite più marcate della propria storia fino a quel momento. La disponibilità di dati ad alta frequenza e di un campione comprendente un periodo storico con una volatilità particolarmente marcata, crea la possibilità di confrontarsi col tema delle Flash Crash. L'obiettivo principale è quello di cercare di costruire un modello utile alla previsione delle crisi di liquidità, e in particolare delle stesse Flash Crash.

Generalmente i dati intraday o a più alta frequenza vengono modellati attraverso processi stocastici in tempo continuo (dal più semplice moto geometrico

Browniano a traiettorie continue (Black and Scholes, 1973), ai più complessi processi del tipo “jump-diffusion” (Merton, 1976)). In quest’ambito la letteratura non offre ancora validi esempi su come tali modelli possano essere impiegati per prevedere fenomeni quali le Flash Crash, e questo potrebbe senz’altro costituire un interessante filone su cui indagare. In questo lavoro l’attenzione si è concentrata su un particolare modello in tempo discreto, prodotto da Easley et al. (2012b).

Lo studio sulla microstruttura dei mercati, sostenuto da quest’ultimi ha permesso di individuare, in un’elevata tossicità nel flusso degli ordini, la causa principale delle crisi di liquidità diffuse nei mercati. A partire da questo presupposto viene introdotta la possibilità di costruire un’indicatore, applicabile al contesto delle contrattazioni ad alta frequenza, che consenta di stimare effettivamente la tossicità presente sul mercato e quindi di prevedere la possibilità di crisi di liquidità.

Nel Capitolo 1 si tenta inizialmente di modellare le serie storiche finanziarie ad alta frequenza, a tutte le differenti frequenze, attraverso i modelli comunemente impiegati nell’analisi delle serie dei rendimenti giornalieri, quali i modelli GARCH e le sue varianti.

Nel Capitolo 2 si passa invece ad introdurre il problema delle Flash Crash, e si analizza il meccanismo che porta alla loro formazione nei mercati.

Nel Capitolo 3 infine si costruisce l’indicatore di tossicità di Easley et al. (2012b), e ne si constata la capacità previsiva sul campione di dati a disposizione. Si propongono poi delle varianti dello stesso, e si testa la loro affidabilità rispetto al modello originale.

# Capitolo 1

## Analisi preliminare

Lo studio delle serie storiche finanziarie è generalmente incentrato sulla costruzione di modelli ad eteroschedasticità condizionale, in grado di rappresentare correttamente il fenomeno della volatilità a grappoli (*volatility clustering*) tipico di queste. L'attenzione è rivolta alla volatilità dei rendimenti finanziari, in quanto è nel solo momento secondo che si osserva la persistenza. I modelli per la media condizionale sono invece considerati poco utili, poiché i valori previsti attraverso quest'ultimi, non fanno altro che convergere velocemente verso la media campionaria, definendo come migliore previsione per i rendimenti futuri la semplice media assunta in passato dagli stessi (*processi mean reverting*).

I modelli comunemente impiegati nel modellare le serie finanziarie, sono i cosiddetti GARCH (*Generalized AutoRegressive Conditional Heteroskedasticity*), i quali definiscono la varianza condizionale dei rendimenti in funzione dei propri valori passati, e dei residui del modello sulla media al quadrato. La costruzione di questi avviene attraverso la sti-

ma, col metodo della massima verosimiglianza, dei parametri del sistema di equazioni:

$$\begin{aligned}x_t &= \mu_t + \epsilon_t \\ \epsilon_t &= v_t \sqrt{h_t} & v_t &\sim i.i.d.(0, 1) \\ h_t &= \kappa + A(L)h_t + B(L)\epsilon_t^2\end{aligned}$$

dove  $v_t$  è un processo a media zero e varianza unitaria,  $\mu_t$  è un generico processo ARMA per i rendimenti, mentre  $A(L)$  e  $B(L)$  sono gli operatori ritardo rispettivamente della varianza e dei residui al quadrato.

Al modello originale di Bollerslev (1986), sono poi seguite molte varianti che permettono di modellare al meglio caratteristiche particolari delle serie dei rendimenti, quali l'eccesso di curtosi e le asimmetrie nella volatilità. Spesso si considerano distribuzioni leptocurtiche delle innovazioni, per spiegare la permanenza di effetti ARCH nel modello. Per modellare il comportamento asimmetrico delle quotazioni dei titoli, si utilizzano invece modelli come l'E-GARCH e il GJR-GARCH, capaci di rappresentare l'effetto leva, secondo cui scostamenti negativi dal valore atteso aumentano la volatilità più di quanto non facciano quelli positivi.

I modelli fin qui citati si adattano bene alle serie storiche dei rendimenti giornalieri, definiti sulla base dei prezzi di chiusura di giornate di negoziazione successive. Focalizzando però l'attenzione sul solo prezzo di chiusura, si finisce per trascurare un flusso importante di informazioni, relativo all'andamento che i prezzi hanno avuto all'interno delle singole giornate di negoziazione.



Il crescente sviluppo dei mercati finanziari e della tecnologia ad essi associata, ha permesso la diffusione di modalità di negoziazione che consentono di effettuare migliaia di scambi in pochissimi secondi, traendo profitti da differenziali di prezzo che si realizzano solo per pochi istanti. I mercati sono sempre più saturi di operatori che applicano strategie di *High Frequency Trading* (HFT): cioè di operatori che, servendosi di sofisticati strumenti software, mettono in atto negoziazioni ad alta frequenza., guidate da algoritmi matematici.

La crescente disponibilità di dati infragiornalieri a frequenze elevate, ha permesso di avere database più ricchi e complessi su cui ragionare, e su cui effettuare analisi più accurate di volumi, rendimenti e volatilità dei titoli azionari.

Nella classe dei processi *mean reverting* prima descritti rientrano anche i modelli FIGARCH, introdotti da Baillie et al. (1996). Questi modelli nascono dalla necessità di descrivere la presenza di una memoria lunga della volatilità riscontrata in molte serie finanziarie.

Ding et al. (1993) notarono infatti che le autocorrelazioni dei quadrati dei residui decrescono nel tempo più lentamente rispetto al caso esponenziale tipico dei modelli GARCH. In particolare gli shock sulla volatilità per i GARCH decadono rapidamente ad un tasso esponenziale, mentre quelli dei FIGARCH hanno un decadimento più lento con tasso iperbolico.

Questi modelli vengono utilizzati generalmente per modellare serie finanziarie con frequenze di rilevazione dei rendimenti più elevate rispetto a quella classica giornaliera. Il modello FIGARCH( $p, d, q$ ) è definito attraverso i parametri

del sistema:

$$\begin{aligned}\epsilon_t &= z_t \sqrt{h_t} \\ z_t &\sim i.i.d.N(0, 1) \\ \Phi(L)(1-L)^d \epsilon_t^2 &= \omega + [1 - \beta(L)]\nu_t\end{aligned}$$

dove  $0 < d < 1$ ,  $\nu_t \equiv \epsilon_t^2 - h_t$  e  $\Phi(L) \equiv [1 - \alpha(L) - \beta(L)]$ .

Se  $d = 0$  si ha un modello GARCH con caduta esponenziale della volatilità nel tempo, se  $d = 1$  si ottiene il modello IGARCH di Engle and Bollerslev (1986), mentre se  $0 < d < 1$  si ottengono strutture di memoria più flessibili. Restando in un contesto di rilevazioni dei prezzi ad alta frequenza, Barndorff-Nielsen and Shephard (2002) illustrarono la possibilità di utilizzare misure di variazione quadratica, per studiare le proprietà stocastiche dei rendimenti delle attività finanziarie. La principale misura impiegata è la Volatilità Realizzata ( “*Realized Volatility*”), la quale definisce una misura attendibile della volatilità di uno strumento sulla base delle effettive realizzazioni dei prezzi, e non delle previsioni calcolate tramite particolari modelli statistici. Definiti i rendimenti di un titolo come:  $r_\tau = 100(\ln P_\tau - \ln P_{\tau-1})$ , dove  $P_\tau$  è il vettore dei prezzi nella giornata di negoziazione  $t$ , la Volatilità Realizzata è calcolata come:

$$RV_t = \sqrt{\sum_{\tau=1}^{N_t} r_\tau^2} \quad t = 1, 2, \dots, T \quad (1.1)$$

dove  $N_t$  è il numero di rilevazioni che compongono la t-esima giornata di negoziazione. Occorre infatti, nell’implementazione del calcolo, definire un’o-

rizzonte temporale di riferimento. Nel caso di dati ad alta frequenza ha senso calcolare una volatilità realizzata giornaliera, piuttosto che mensile o annuale.

## 1.1 I dati

I dati a disposizione descrivono l'andamento dell'indice azionario inglese FTSE 100, che comprende le 100 società più capitalizzate quotate al London Stock Exchange, su un orizzonte temporale di 16 mesi, che va dal 4 gennaio 2011 al 31 Maggio 2012. Per tale arco di tempo si dispone di osservazioni dei prezzi a differenti frequenze di rilevazione, che sono, in ordine: cinque minuti, due minuti, un minuto, trenta secondi, dieci secondi, cinque secondi e un secondo.

Tipicamente quando si raccolgono dati ad alta frequenza, le varie transazioni sono spaziate nel tempo in maniera irregolare. La non omogeneità dei dati complicherebbe l'analisi, implicando la costruzione di modelli diversi da quelli tradizionali. Tuttavia il data-set disponibile, nonostante l'elevata frequenza di rilevazione, contiene tutte osservazioni equispaziate<sup>1</sup> ; per cui è possibile applicare le tecniche di analisi comunemente utilizzate per dati giornalieri.

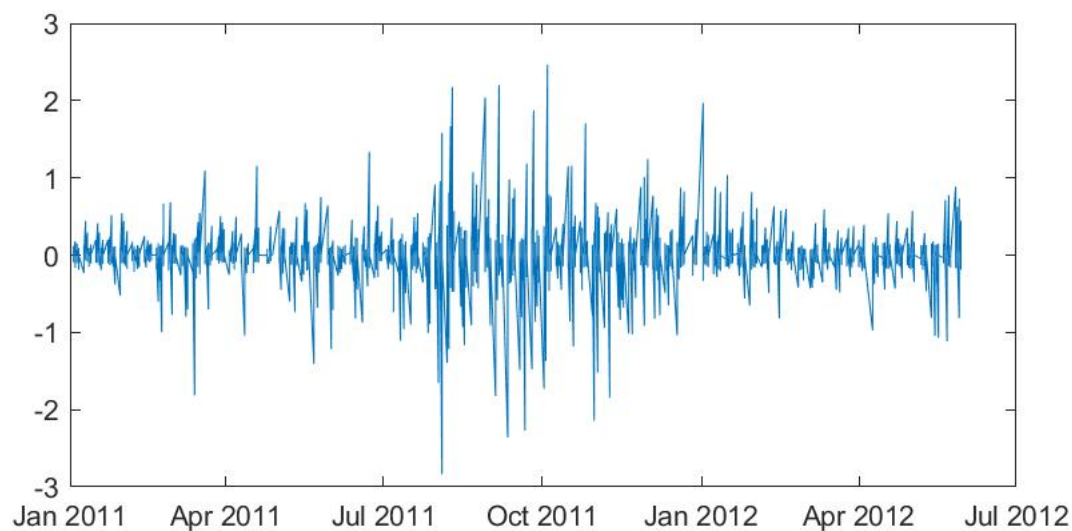
L'andamento dei rendimenti rappresentato in figura 3.3, è tipico delle serie finanziarie, contraddistinto da un valore atteso centrato all'incirca sullo zero e dal fenomeno della volatilità a grappoli.

La varianza aumenta vistosamente tra la fine del luglio 2011 e l'ottobre dello stesso anno, in corrispondenza di quella che è, in realtà, una caduta globale

---

<sup>1</sup>Originariamente si disponeva di rilevazioni transazione per transazione, le quali sono state poi aggregate per generare osservazioni equispaziate, a diverse frequenze.

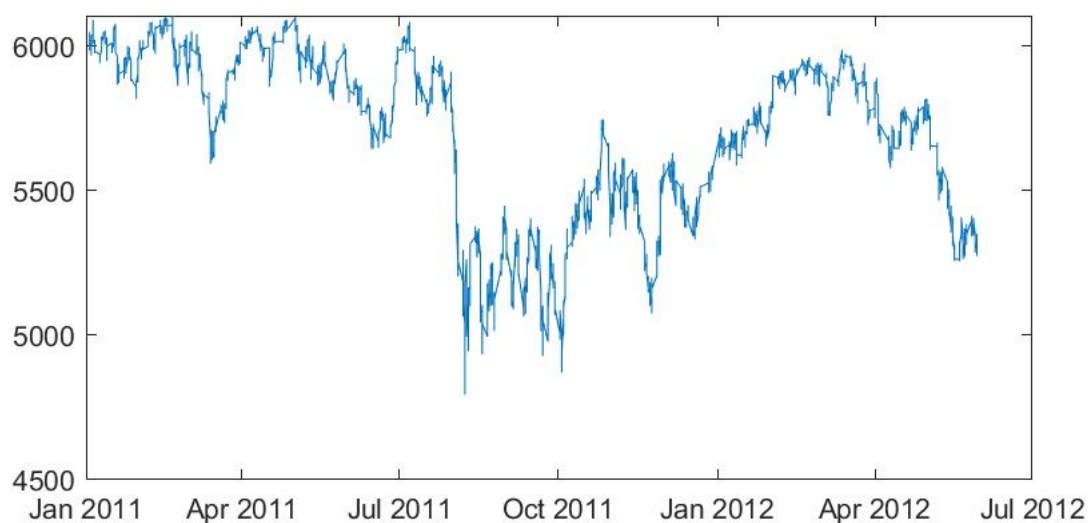
**Figura 1.1:** Serie dei rendimenti FTSE 100 (frequenza dati: 60 secondi)



dei mercati, osservata nell'estate del 2011, a seguito della diffusione dei timori per la crisi del debito sovrano europea.

La particolarità della serie in esame viene colta meglio osservando il grafico dei prezzi in figura 1.2. Si nota come l'indice azionario subisce un crollo

**Figura 1.2:** Serie dei prezzi FTSE 100

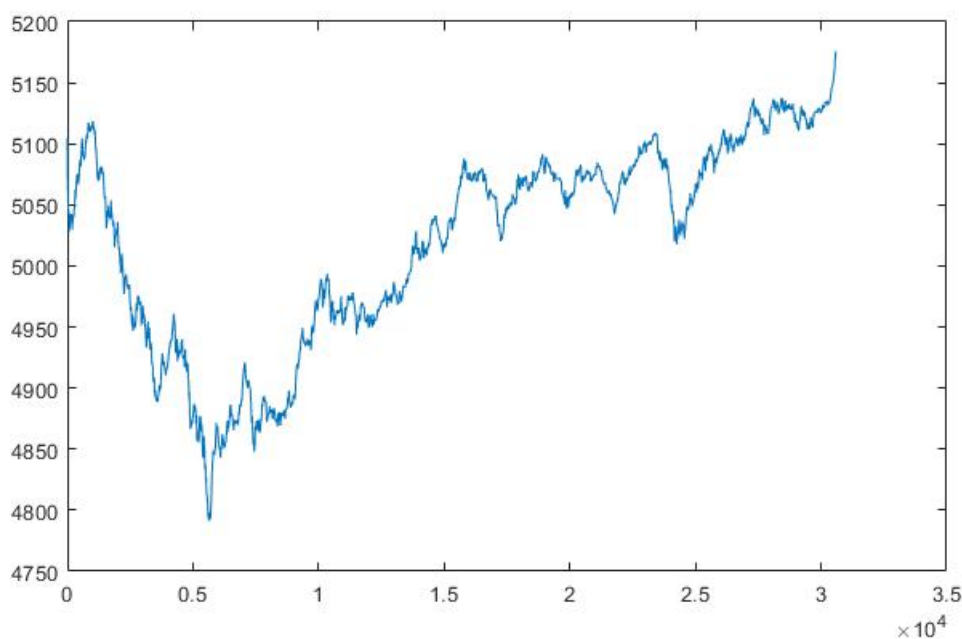


di circa il 23% del proprio valore in appena 8 giorni di continui e marcati

ribassi, tra il primo Agosto ed il 9 Agosto 2011. Soffermandosi sulle singole giornate di negoziazione, è possibile osservare come ci siano vari crolli di mercato improvvisi, seguiti da una rapida ripresa delle quotazioni. Più si aumenta la frequenza di osservazione del dataset, più il fenomeno risulta evidente. La figura 1.3 si riferisce alla giornata di negoziazione del 9 agosto 2011, composta da un totale di 30601 rilevazioni da un secondo. A partire dall'osservazione 1020 il prezzo, di 5118£, inizia a ridursi bruscamente fino all'osservazione 5668, raggiungendo il livello di 4791£: si tratta di un crollo della quotazione dell'indice FTSE 100 del 6.4% in un arco di tempo di circa un'ora.

La portata dell'evento appena descritto rimanda a un fenomeno generalmente noto col nome di "*Flash Crash*". Il termine venne coniato nel Maggio del 2010, quando il mercato borsistico americano crollò, facendo evaporare circa un trilione di dollari di capitale azionario in meno di 10 minuti, recuperando poi nelle successive ore di negoziazione solo parte della perdita.

La causa scatenante delle *Flash Crash* sarebbe il più delle volte un flusso di ordini tossico da parte di uno o più operatori del mercato, che poi innescerebbe la reazione immediata degli *High Frequency Trader* (HFT<sub>r</sub>), quindi movimenti vistosi di mercato in breve tempo. Un flusso di ordini è definito "tossico" quando seleziona avversamente i market maker, che sono ignari del fatto di fornire liquidità in perdita. Si tratta quindi di ordini che celano un'informazione significativa, di cui sono a conoscenza solo i trader che immettono l'ordine. Come la tossicità degli ordini aumenta, i *market maker*, che non sono altro che gli *high frequency trader*, devono confrontarsi con potenziali perdite ingenti e riducono quindi la propria esposizione o, in casi più estremi,

*Figura 1.3: Giornata di negoziazione del 9 Agosto 2011*

abbandonano la propria attività di market making. Questa vasta contrazione della liquidità sul mercato genera un sensibile aumento della volatilità, che esercita una forte pressione in vendita, facendo ridurre rapidamente i prezzi e originando la flash crash. I grandi investitori usano generalmente la volatilità come parametro di riferimento per misurare i rischi e per prendere posizione: se è bassa comprano, mentre se si alza troppo vendono. I loro acquisti nei buoni momenti riducono la volatilità, mentre le loro vendite nelle fasi negative la amplificano. Ecco allora che una volta salita la volatilità gli investitori iniziano a vendere in maniera immediata, per via dell'utilizzo sempre più diffuso di algoritmi automatici; di conseguenza, più vendono e più continua a crescere la volatilità, innescando un circolo vizioso di vendite che deprime rapidamente le quotazioni di mercato. L'elevato livello di automazione nei

mercati odierni sarebbe quindi, più che una causa scatenante, un meccanismo che consente il rapido diffondersi di perdite ingenti. L'argomento verrà approfondito in maniera dettagliata nel prossimo capitolo.

Lo scopo dell'analisi è quello di costruire un modello statistico che riesca a modellare in maniera efficace serie finanziarie ad alta frequenza, tenendo però in considerazione l'eventuale presenza di *Flash Crash*.

## 1.2 Modelli di volatilità condizionale

Inizialmente si cerca di analizzare la serie attraverso i modelli statistici comunemente utilizzati nell'analisi dei rendimenti giornalieri. Si procede quindi con la modellazione delle serie dei rendimenti attraverso la metodologia prevista da Box-Jenkins. Si costruisce a questo scopo un programma, attraverso il software MATLAB, che consenta di selezionare il miglior modello ARMA possibile, per ogni frequenza disponibile, sulla base della minimizzazione dei criteri informativi AIC e BIC.

Poichè si cerca inoltre di produrre modelli parsimoniosi, cioè con il minor numero possibile di parametri da stimare, si limita la ricerca del miglior modello tra quelli aventi componenti autoregressive e a media mobile al massimo fino al quinto ordine.

I risultati ottenuti sono illustrati nella tabella 1.1, dove, per ogni frequenza di osservazione si indicano, oltre al miglior modello ARMA selezionato, il valore dell'indicatore di bontà previsiva RMSE e i risultati del test ARCH sui residui, specificando il valore della statistica test e il p-value. L'indicatore RMSE si riduce all'aumentare della frequenza; con un numero crescente di

**Tabella 1.1:** Modelli ARMA

Frequenza	Migliori modelli	RMSE	ARCH Test	p-value
300s	AR(2)	0.1198	321.359	0
120s	ARMA(1,3)	0.0762	231.9611	0
60s	ARMA(5,5)	0.0522	171.6737	0
30s	ARMA(3,4)	0.0348	221.3885	0
10s	ARMA(5,5)	0.0152	$4.002 \cdot e^4$	0
5s	ARMA(5,2)	0.0095	$2.8771 \cdot e^4$	0
1s	ARMA(2,3)	0.0034	$4.4935 \cdot e^4$	0

dati elaborati cresce anche la bontà previsiva dei modelli. Come mostrato dal test ARCH, per ciascun modello stimato, i quadrati dei residui evidenziano sempre una forte autocorrelazione. La presenza di eteroschedasticità condizionale nelle serie dei rendimenti, rende necessario ricorrere a modelli GARCH. Il punto di partenza per tale modellizzazione restano i processi ARMA costruiti per la media condizionale.

Si procede quindi alla stima, per ogni frequenza di rilevazione, del miglior modello possibile per la varianza condizionale, sulla base del principio della minimizzazione dei criteri informativi AIC e BIC. Si prendono in esame a questo fine anche distribuzioni non normali dei rendimenti, ed eventuali effetti asimmetrici, colti dai modelli EGARCH e GJR-GARCH. Le statistiche della distribuzione dei rendimenti infatti, evidenziano, per tutte le varie frequenze, una certa asimmetria negativa, che suggerisce l'utilizzo di modelli asimmetrici.

Si considerano infine anche i modelli FIGARCH e la loro bontà di adattamento al campione dei dati.

L'indicatore RMSE è costruito in maniera differente rispetto a quello im-



piegato per i modelli ARMA. In generale il *root mean square error* è una misura dello scostamento dei valori effettivamente osservati rispetto a quelli previsti dal modello:

$$RMSE = \sqrt{\frac{1}{T} \sum_{t=1}^T (Y_t - \widehat{Y}_t)^2} \quad (1.2)$$

dove  $T$  è l'ampiezza della serie considerata,  $Y_t$  è la  $t$ -esima osservazione delle serie, mentre  $\widehat{Y}_t$  è la  $t$ -esima previsione di  $Y_t$  secondo il modello considerato. Per quanto riguarda i GARCH, la variabile d'interesse è la varianza. Il problema sta nello stabilire con quali valori effettivi confrontare le previsioni ( $\widehat{V}$ ) fornite dal modello. Essendo in generale il valore atteso dei rendimenti pari a zero, una buona indicazione del valore della volatilità al tempo  $t$ , è data dal valore assoluto dei rendimenti.

L'indicatore di bontà previsiva della (1.2) diventa allora:

$$RMSE = \sqrt{\frac{1}{T} \sum_{t=1}^T (|r_t| - \widehat{V}_t)^2}$$

Occorre precisare che la scelta dei modelli non viene fatta esclusivamente in funzione dei criteri informativi, ma considerando anche altri fattori quali la significatività dei parametri stimati, il livello di autocorrelazione dei residui, ed eventuali problemi del software nel giungere a convergenza nell'applicazione del metodo della massima verosimiglianza.

Infatti, al crescere della frequenza di rilevazione, i vari algoritmi di risoluzione previsti da MATLAB per la stima di massima verosimiglianza, non riescono a convergere, per vari modelli, a delle soluzioni accettabili. In particolare,

in relazione ai modelli EGARCH, applicati a frequenze elevate, la varianza condizionale stimata evidenzia outlier importanti, che non riescono ad essere interpretati neanche con l'inserimento di una variabile *dummy*, che annulli gli effetti di residui sballati sul modello.

**Tabella 1.2:** Modelli GARCH

Frequenza	Miglior modello	Distr.	RMSE	CI		LB p-value		
				AIC	BIC	5 lags	10 lags	20 lags
300s	EGARCH(2,2)	N	0.1090	$-6.36 \cdot e^4$	$-6.35 \cdot e^4$	0.5590	0.5251	0.1410
120s	GARCH(1,1)	t	0.0721	$-2.93 \cdot e^5$	$-2.93 \cdot e^5$	0.2291	0.4408	0.8177
60s	GJRGARCH(2,1)	t	0.0504	$-7.39 \cdot e^5$	$-7.39 \cdot e^5$	0.1790	0.4017	0.9121
30s	GARCH(2,1)	t	0.0342	$-1.83 \cdot e^6$	$-1.83 \cdot e^6$	0.0083	0.2058	0.3305
10s	GJRGARCH(2,1)	t	0.0156	$-7.34 \cdot e^6$	$-7.34 \cdot e^6$	0	0	0
5s	GJRGARCH(2,1)	t	0.0098	$-1.71 \cdot e^7$	$-1.71 \cdot e^7$	0	0	0
1s	GJRGARCH(2,1)	t	0.0034	$-1.14 \cdot e^8$	$-1.14 \cdot e^8$	0	0	0

La tabella 1.2 illustra i risultati ottenuti dall'analisi, mettendo in evidenza, in ordine: il miglior modello ottenuto dai dati, il tipo di distribuzione delle innovazioni utilizzato, il valore dell' RMSE, i valori dei criteri informativi AIC e BIC, e infine i p-value relativi al test di Ljung-Box sui residui standardizzati, da cui è possibile constatare la bontà del modello, cioè la capacità dello stesso di catturare l'eteroschedasticità presente nelle serie. Coerentemente a quanto visto per i modelli sulla media condizionale, la bontà previsiva cresce all'aumentare della frequenza di rilevazione. Allo stesso tempo, anche i criteri informativi si riducono al crescere della frequenza, indicando una sempre minore perdita di informazione dei modelli rispetto al processo generatore. Il modello previsto per la frequenza a 300 secondi, riesce a spiegare bene l'eccesso di curtosi connesso alla serie, conservando una distribuzione delle innovazioni gaussiana; questo perché si ha a che fare con un dataset relativamente più contenuto, composto da 36770 rilevazioni, che può essere accosta-

to a quelli normalmente impiegati nelle analisi dei rendimenti giornalieri. Al crescere della frequenza invece, l'autocorrelazione nei modelli con innovazioni gaussiane, porta a preferire quelli con una distribuzione dei residui secondo una  $t$  di student. Seppure la massimizzazione della log-verosimiglianza sia più onerosa e complessa dal punto di vista computazionale, tali modelli sembrano essere preferiti nell'analisi.

La bontà di adattamento ai dati di un generico modello GARCH, viene generalmente misurata studiando le autocorrelazioni dei residui standardizzati: più questi sono incorrelati, e più risultano normalmente distribuiti, migliore è il modello selezionato. Il test di Ljung-Box adoperato verifica l'assenza di autocorrelazione nei residui fino a un ordine di 5, 10 e 20 ritardi. Poiché l'ipotesi nulla del test afferma che tutte le autocorrelazioni fino al  $k$ -esimo ritardo sono nulle,  $p$ -value inferiori al livello di confidenza scelto (1%) portano a rifiutare tale ipotesi e a non considerare affidabili i risultati del modello. È necessario notare come al crescere della frequenza di rilevazione, cresca anche il grado di autocorrelazione nei residui dei migliori modelli individuati; tanto che in quelli con un numero di osservazioni superiori al milione i  $p$ -value assumono valori prossimi allo zero. Più il campione dei dati è numeroso, più i modelli proposti trovano difficoltà a spiegare l'eteroschedasticità nella serie dei rendimenti a disposizione. I risultati relativi alla stima dei FIGARCH

**Tabella 1.3:** Modelli FIGARCH

Frequenza	Miglior modello	Distr.	RMSE	CI		LB p-value		
				AIC	BIC	5 lags	10 lags	20 lags
120s	FIGARCH(0,0)	N	0.0726	$-2.36 \cdot e^5$	$-2.36 \cdot e^5$	0.0007	0.0093	0.0045
60s	FIGARCH(1,0)	N	0.0509	$-6.13 \cdot e^5$	$-6.13 \cdot e^5$	0.0006	0.0060	0.1345
30s	FIGARCH(1,0)	N	0.0345	$-1.52 \cdot e^6$	$-1.52 \cdot e^6$	0	0	0

sono illustrati invece nella tabella 1.3. Per molti modelli in realtà l'algoritmo risolutivo non riesce a convergere ad una soluzione accettabile. Inoltre si nota che, nei casi in cui si arrivi effettivamente ad una soluzione, questa non è affatto ottimale, in quanto evidenzia un elevato grado di autocorrelazione dei residui standardizzati.

I risultati dell'analisi proposta mostrano quindi che i tradizionali modelli GARCH non risultano particolarmente adatti alla modellazione delle serie finanziarie ad alta frequenza.

Le stime della varianza condizionale inoltre, quando possibili, sarebbero funzionali alla costruzione di modelli per l'*asset allocation* o il *risk hedging*, ma non sarebbero in alcun modo utili al fine di prevedere situazioni particolarmente turbolente dei mercati, quali le Flash Crash.

In altri termini, sebbene i modelli costruiti in questo capitolo consentano di prevedere efficacemente la volatilità futura di un titolo sul mercato, non sarebbero di supporto per le finalità della tesi, in quanto non permetterebbero di riconoscere ed anticipare, in tempo reale, picchi di volatilità significativi, dovuti a crisi di liquidità.

# Capitolo 2

## Le Flash Crash

### 2.1 La Flash Crash del Maggio 2010

Il 6 Maggio 2010 il mercato borsistico americano sperimentò uno dei periodi più turbolenti della sua storia. L'indice Dow Jones Industrial Average subì una contrazione di circa 1000 punti in poco meno di 10 minuti, facendo evaporare un trilione di dollari di capitale azionario. Nei successivi 20 minuti di negoziazione il mercato recuperò poi circa 600 punti di perdita (SEC and CFTC, 2010). Data la rapidità con cui il crollo del mercato si verificò, l'evento divenne famoso col nome di "Flash Crash".

Le autorità di vigilanza scoprirono che alla base del crollo c'era uno schema di *spoofing*. Lo *spoofing*<sup>2</sup> è una pratica illegale di manipolazione del mercato, che consiste nell'immissione di un ampio flusso di proposte di negoziazione al fine di orientare le contrattazioni, senza concludere l'operazione. Secondo questo schema, un soggetto che intenda acquistare uno strumento finanzia-

---

<sup>2</sup>Dal verbo inglese "to spoof" che significa falsificare, imbrogliare.

rio a un prezzo più basso di quello di mercato, immette, in una frazione di tempo notevolmente ridotta<sup>3</sup>, numerose proposte di vendita a un prezzo di poco superiore alla migliore offerta, in modo tale da generare una pressione ribassista sul titolo (data dai volumi di vendita elevati) senza che gli ordini immessi siano poi eseguiti (solitamente vengono cancellati). La pressione generata attraverso lo *spoofing* sfocia in un rapidissimo movimento delle quotazioni del titolo per via dell'elevato livello di automazione del sistema degli scambi, essendo i mercati saturi di operatori che negoziano tramite strategie di *High Frequency Trading* (HFT), cioè servendosi di sofisticati strumenti software con cui mettere in atto negoziazione ad alta frequenza guidate da algoritmi matematici.

Dal maggio 2010, il termine “Flash Crash” viene utilizzato generalmente per identificare una rapida caduta dei prezzi sui mercati elettronici, amplificata dal ritiro degli ordini azionari (Kenton, 2019).

Le autorità di vigilanza dei mercati americani, SEC e CFTC<sup>4</sup>, pochi mesi dopo l'accaduto redassero un report, col nome di “Findings regarding the market events of May 6, 2010”, in cui offrirono una spiegazione dettagliata del crash, attribuendone la causa principale erroneamente ad un algoritmo di vendita predisposto dalla società d'investimenti Waddell & Reed. Molto importante, si spiega, era il background nel quale si operava in quelle giornate: i mercati erano infatti molto volatili e sensibili ai mutamenti, a causa della diffusione della crisi del debito sovrano europea. In realtà lo *spoofing*, di nome Sarao, contribuì a renderli più nervosi del previsto (Levine, 2015).

---

<sup>3</sup>In particolare, lo spazio di un millisecondo, pari a  $10^{-3}$  secondi

<sup>4</sup>Rispettivamente la “Security Exchange Commission” e la “Commodity Futures Trading Commission”

Procedendo con ordine, l'algoritmo di *spoofing* di Sarao si attivò effettivamente dalle 11:17 a.m. alle 2:40 p.m. Occorre precisare che quest'ultimo non era un'*High Frequency Trader* (HFTr), ma piuttosto un trader di vecchia scuola, per cui continuò fino alle 2:45 p.m. a immettere falsi ordini manualmente, generando un forte disequilibrio nell'order book del mercato E-mini S&P. L'efficacia della sua azione, precisa la CFTC, si limita però al periodo antecedente l'effettiva flash crash, con perdite contenute; inoltre la pressione generata dai suoi ordini, in ogni caso non eseguiti, era pari a circa il 20-29% della pressione totale sul lato delle vendite del book. Quindi, è nel restante 71-80% dei venditori, i cui ordini furono effettivamente eseguiti, che deve essere ricercata l'effettiva causa del crollo del prezzo.

*Figura 2.1: SPX Flash Crash*



Alle 2:32 p.m. un complesso di fondi di investimento, la Waddell & Reed, a seguito dell'inusuale elevata volatilità e dell'assottigliamento della liquidità

sul mercato, iniziò un programma di vendita di 75000 contratti E-mini (dal valore di 4,1 miliardi) a copertura di una posizione esistente. A questo scopo predispose un algoritmo di vendita automatico che, per via delle stressate condizioni dei mercati, concluse il programma pianificato in appena 20 minuti (SEC and CFTC, 2010).

La pressione di vendita viene assorbita generalmente da 3 soggetti:

- gli High Frequency Trader (HFTr), nel ruolo di market maker.
- i compratori fondamentali.
- gli arbitraggisti, che traggono profitto acquistando futures per vendere azioni.

Tra le 2:32 p.m. e le 2:45 p.m. l'algoritmo del fondo vendette 35000 contratti dei totali. Durante lo stesso arco di tempo tutti i venditori fondamentali cedettero 80000 contratti a fronte dei 50000 in acquisto, generando uno sbilanciamento fondamentale di 30000 contratti.

Gli arbitraggisti scaricarono la pressione al ribasso dell'e-mini sul rispettivo indice azionario.

Gli HFTr d'altra parte costruirono posizioni long temporanee; una volta raggiunta una certa soglia predefinita, dovettero però alleggerire la propria esposizione cedendo contratti e ritirandosi dalla propria attività di market making. Se infatti in un primo momento gli HFTr avevano effettivamente operato come market maker, apportando così liquidità al mercato, subito



dopo cominciarono ad interrompere tale operazione. Per mantenere una posizione bilanciata, cominciarono ad affiancare i loro ordini di vendita a quelli degli altri venditori presenti sul mercato. Gli ordini di vendita degli HFTr indussero gli altri HFTr a fare lo stesso generando un fenomeno, conosciuto come “*hot potato trading*”<sup>5</sup>, in cui le controparti degli scambi erano operatori *high frequency* che continuavano a vendere.

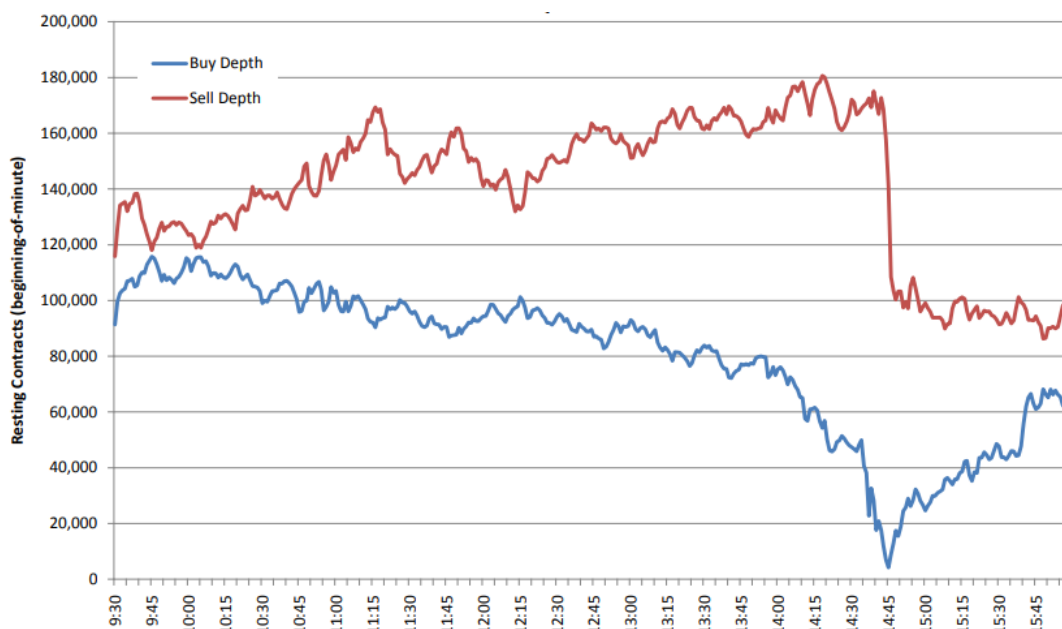
Si creò quindi una grave crisi di liquidità che, tra le 2:41 p.m. e le 2:45:28 p.m., spinse il mercato E-mini a cedere circa il 5-6%; la perdita si propagò poi rapidamente ai principali indici azionari US.

Sarebbe poi un meccanismo speculare a innescare una rapida ripresa, seppur parziale, delle quotazioni; una volta che il mercato fu sospeso tramite un *circuit breaker* del Chicago Mercantile Exchange, gli *Automated trader* si accorsero che il prezzo si trovava a livelli artificialmente troppo bassi, e procedettero così ad una serie di acquisti, inizialmente manuali, che complessivamente aiutarono il mercato a recuperare.

La figura 2.2 mostra l’entità della crisi di liquidità che si generò a ridosso dell’orizzonte temporale considerato. Si può notare come effettivamente lo *spoofing* di Sarao abbia influito sulla microstruttura del mercato, in quanto a partire da circa le 12:15 a.m. lo sbilanciamento tra ordini di acquisto e di vendita inizia ad accentuarsi, provocando una forte pressione ribassista. Tuttavia il crollo verificatosi nei 4 minuti tra le 2:41 p.m. e le 2:45 p.m. avvenne quando l’algoritmo di Sarao venne disattivato e il prezzo sarebbe dovuto teoricamente rimbalzare.

---

<sup>5</sup>Il termine, traducibile in italiano con “la negoziazione della patata bollente”, si riferisce appunto al ripetuto passaggio, tra gli operatori ad alta frequenza, degli squilibri del proprio magazzino titoli.

*Figura 2.2: E-mini market depth*

Questo non accadde, poichè l'attivazione dell'algoritmo di vendita automatico della Waddell & Reed, innescato dalla forte tensione sui mercati e dallo stesso *spoofing*, creò uno scompenso tale da generare la flash crash.

Di fatti Sarao scelse il giorno sbagliato per trarre profitti attraverso la propria attività manipolativa.

## 2.2 High Frequency Trading e Flash Crash

L'High Frequency Trading (HFT) è definito come il meccanismo tramite cui è possibile operare sui mercati attraverso l'impiego di software e hardware che, guidati da algoritmi matematici preimpostati, permettono di effettuare negoziazioni ad altissima frequenza. I soggetti che operano con questo tipo di tecnologia sono gli *High Frequency Trader*: generalmente si tratta di

grandi imprese d'investimento, per via dell'entità delle operazioni che vanno a svolgere e per via dei costi necessari per sostenere l'attività.

Tali aziende presentano alcune caratteristiche fondamentali:

- utilizzano programmi sofisticati di straordinaria velocità per la generazione ed esecuzione degli ordini.
- effettuano trading proprietario, cioè impiegano il capitale proprio nella negoziazione.
- usufruiscono del servizio di *co-location*, reso disponibile dalle piattaforme di negoziazione, attraverso cui collocano i propri server in prossimità fisica dei server della Borsa, al fine di minimizzare la latenza (il tempo che intercorre tra l'invio di un'ordine e l'esecuzione dello stesso).
- periodo di detenzione (*holding period*) delle posizioni molto breve. Prediligono infatti strumenti finanziari molto liquidi.
- generano numerosi ordini nell'unità di tempo (anche 5000 al secondo), molti dei quali, utilizzati per sondare il mercato, vengono cancellati.
- chiudono la giornata con posizioni piatte, ma hanno un *turnover* di magazzino elevato. Non detengono cioè, al termine della giornata di negoziazione, nessuna posizione rilevante, né lunga, né corta: il loro obiettivo è quello di mantenere una posizione bilanciata. Tuttavia nell'arco della giornata muovono comunque elevati volumi nelle negoziazioni, esponendosi sui vari strumenti finanziari solo temporaneamente.

L'evento del 6 Maggio 2010 e l'indagine ad esso connessa, rivelò al mondo della finanza l'esistenza concreta della presenza degli operatori ad alta fre-

quenza sui mercati; tanto che, in un primo momento, proprio questi furono ritenuti responsabili del crollo. L'HFT non viene considerato generalmente come la principale causa di flash crash poiché, in situazioni normali di mercato, questi sistemi non producono evidenti movimenti direzionali, ma anzi tendono a sostenere la liquidità dei titoli. Gli HFTr infatti inseriscono una grande quantità di ordini passivi, per guadagnare su margini di prezzo molto sottili. Solo in condizioni di forte stress del mercato, dovute a uno squilibrio nel flusso degli ordini, quest'attività di trading può amplificare i movimenti direzionali contribuendo ad aumentare la volatilità.

Se la presenza di HFTr non è condizione essenziale per il verificarsi di crolli verticali sui mercati, è pur vero che, sulla scia di quanto visto, risulta un fattore chiave nella fase di caduta dei prezzi, amplificando in modo esponenziale la velocità con cui il tutto avviene.

In un paper della Banca d'Italia, Puorro (2013a) sottolinea che: “Nel momento in cui viene inviato un ordine di vendita particolarmente grosso relativo ad uno strumento finanziario, aumenta la probabilità che il mercato possa diventare più sottile a causa dell'immediata cancellazione degli ordini in acquisto da parte degli Hftr meno capitalizzati e non in grado di gestire l'ordine, determinando una possibile discesa anomala dei prezzi e un aumento della volatilità di brevissimo periodo. Se tale movimento avviene in concomitanza di importanti livelli tecnici, è ragionevole ipotizzare l'attivazione di ordini automatici di *stop loss* che rafforzano il trend in atto. L'arrivo di ulteriore liquidità, contestualmente a tali ordini di chiusura, può provocare il ritiro degli ordini di negoziazione e la chiusura delle operazioni anche degli Hftr più capitalizzati che precedentemente avevano accumulato posizioni contro

il trend. L'uscita di tali Hftr e il tentativo di seguire il trend (in discesa) in atto da parte di altri Hftr spinge il prezzo dello strumento in questione in violenta accelerazione”.

Sia in periodi di calma, sia in quelli di turbolenza, gli HFTr hanno come obiettivo primario quello di detenere una posizione bilanciata del portafoglio di negoziazione, evitando di esporsi troppo *long* o *short*. Questo aspetto influenza molto spesso negativamente i mercati, portando a contrazioni rapide di liquidità difficili da gestire, come visto nel caso della flash crash del 2010, e come spiegato dallo stesso Puorro (2013a).

### 2.2.1 La velocità come chiave del successo

Come sottolinea la CONSOB: “L’HFT è una modalità operativa incentrata sulla velocità di acquisizione ed elaborazione delle informazioni di mercato, e di reazione a tali informazioni (Caivano et al., 2012)”.

Come si evince chiaramente da questa definizione, la caratteristica fondamentale che contraddistingue gli HFTr, è il possesso di un’elevata velocità operativa (da intendersi come rapidità nell’immissione, cancellazione, esecuzione, nonché modifica degli ordini inviati alle varie piattaforme di mercato). L’efficacia di tutte le strategie implementate da questi operatori, quindi la loro profittabilità, dipende esclusivamente dalla capacità di essere più rapidi degli altri partecipanti al mercato. Gli investimenti degli HFTr, molto spesso ingenti, sono quindi rivolti, in tal senso, a massimizzare la propria velocità di esecuzione degli ordini, al fine di massimizzare anche i profitti.

Dal punto di vista economico, non c’è alcuna nuova teoria innovativa, si trat-

ta semplicemente di cogliere i frutti del progresso tecnologico, per quanto costoso esso sia inizialmente. Gli HFTr cioè fondano la propria esistenza su un vantaggio competitivo di tipo tecnologico, senza il quale ricadrebbero semplicemente nella categoria di algo trading<sup>6</sup> (Puorro, 2013b).

Resta da chiarire cosa si intende a livello pratico con velocità, e verso cosa sono rivolti effettivamente gli investimenti degli HFTr al fine di raggiungerla. Per fare ciò, non si può prescindere dal ricorrere al termine di “latenza”. In informatica, il tempo di latenza, è il lasso di tempo che intercorre tra il momento in cui viene inviato un comando di lettura o scrittura a una unità periferica, e il momento in cui effettivamente inizia il trasferimento dei dati (Treccani, 1998). Si tratta quindi della misura della velocità di risposta agli input di un sistema elettronico.

Nel contesto economico di riferimento, può essere intesa semplicemente come il tempo necessario per la trasmissione degli ordini al mercato, ma ciò sarebbe riduttivo. La latenza interesserebbe infatti, non solo l’impulso iniziale, ma diversi momenti del processo di negoziazione:

- il tempo con cui gli operatori ricevono le informazioni dal mercato e le processano.
- il tempo tra la processazione del dato e l’invio effettivo dell’ordine al *broker*.

---

<sup>6</sup>Il trading algoritmico è un metodo di esecuzione degli ordini, che utilizza istruzioni di negoziazione automatiche programmate in precedenza (che coinvolgono variabili come tempo, prezzo e volume), con lo scopo di automatizzare le decisioni di investimento, eliminando la componente emotiva e comportamentale. L’HFT rientra nella famiglia del trading algoritmico, ma ciò che lo contraddistingue è l’elevata velocità di accesso al mercato e la breve durata del periodo di detenzione delle posizioni.

- il tempo tra la ricezione del dato dal *broker* e l'invio dello stesso alla piattaforma di negoziazione.
- il tempo che l'ordine impiega ad arrivare sul mercato una volta che viene rilasciato dal *broker*.
- il tempo tra la ricezione dell'ordine da parte del mercato e la divulgazione del dato ai partecipanti del mercato.

La velocità di un sistema di HFT sta quindi nel riuscire a coprire tutti i passi del processo d'investimento sopra elencati nel minor tempo possibile.

Se la chiave del successo è minimizzare la latenza e riuscire a svolgere tutte le fasi più rapidamente possibile, un HFTTr deve innanzitutto dotarsi di un apparato informatico che riesca a ricevere, analizzare e processare le informazioni del mercato in brevissimo tempo; deve inoltre servirsi di un *broker* con una tecnologia altrettanto avanzata, per gestire in maniera efficiente la trasmissione del segnale al mercato.

In un contesto operativo tecnologicamente avanzato, come quello appena descritto, gli impulsi elettrici inviati viaggino a velocità inferiori al millisecondo, praticamente impercettibili per l'uomo.

Supposto che tutti gli operatori ad alta frequenza siano dotati di apparati informatici estremamente avanzati, capaci di elaborare e trasmettere dati a velocità estreme, c'è un limite che tutti gli impulsi elettrici da essi inviati devono affrontare in ogni caso: il limite dello spazio fisico. In un mercato, dove è vero che sono presenti notevoli barriere all'ingresso, ma estremamente competitivo, guadagnare un millisecondo significa ottenere fette di mercato, ma soprattutto profitti superiori. In altri termini, fare in modo che il proprio

impulso elettrico impieghi il minor tempo possibile per essere attivo, significa guadagnare un rilevante vantaggio competitivo.

Ecco quindi che la vicinanza fisica al *server* di borsa permette ai trader ad alta frequenza di sfruttare un vantaggio, che apre possibilità di guadagno infinitamente superiori a quelle dei trader più lenti. In quest'ottica nasce quindi il servizio di *co-location*, già citato nel precedente paragrafo, tramite il quale le varie piattaforme di negoziazione offrono agli HFTr la possibilità di collocare i propri *server* in prossimità fisica dei *server* borsistici. La diffusione crescente di tale servizio è il frutto dell'interazione tra gli interessi delle piattaforme e quelli degli HFTr: se da una parte gli operatori ad alta frequenza guadagnano in termini di bassa latenza, le piattaforme di negoziazione possono contare sia sull'operatività degli stessi HFTr, che sono in grado di garantire ai partecipanti al mercato un'elevata liquidità, sia sugli elevati proventi incassati, garantitigli dal privilegio offerto a questi operatori. In ragione degli elevati introiti derivanti da tale business, molti canali di negoziazione borsistici, hanno ben presto avviato l'espansione dei propri *data-center*, per approfittare della crescente domanda del servizio.

In realtà molto spesso gli HFTr mettono in pratica una *multiple co-location*, cioè situano i propri server in più punti geografici, ognuno dei quali in prossimità fisica di varie Borse.

Risulta inoltre possibile per gli operatori ad alta frequenza usufruire di un servizio di *proximity central hosting*: in questo caso la locazione degli spazi, in cui disporre i propri *server*, viene offerta da soggetti terzi, estranei alle piattaforme di negoziazione. Solitamente gli HFTr usufruiscono di tale servizio posizionando i *server* in un unico punto strategico, che sia equidistante



dai canali di negoziazione su cui concentrare l'attività.

Sintetizzando quanto detto finora, gli HFTr sono operatori che, a fronte di un investimento corposo in tecnologia, ottengono profitti ingenti per il semplice fatto di essere più rapidi di tutti gli altri partecipanti al mercato. La loro attività non può prescindere dall'ottenimento di una bassa latenza nel processo di negoziazione, e dall'implementazione di servizi di *co-location*.

Il fenomeno del trading ad alta frequenza si è sviluppato progressivamente a partire dalla prima decade del ventunesimo secolo, fino a raggiungere livelli estremi intorno agli anni 2009-2010, quando nei mercati americani circa il 60% degli scambi era condotto da HFTr (Breckenfelder, 2019). Pur restando un fenomeno particolarmente radicato negli Stati Uniti, dove ad oggi la quota di scambi attribuibili agli HFTr tocca quasi il 50% delle negoziazioni, la sua diffusione ha interessato rapidamente tutto il mondo. Ad oggi l'HFT è un settore maturo e strutturato, in cui diverse importanti società d'investimento si dividono profitti sempre meno ingenti, per via delle crescenti regolamentazioni, sviluppatesi a seguito del diffondersi dell'attività e i problemi ad essa connessi (crash di mercato).

La CONSOB (2019) nella propria relazione annuale, evidenzia la portata del fenomeno sul mercato italiano: “Nel corso del 2018 il 32% circa dei controvalori scambiati sull'MTA è riconducibile a operatori registrati come high frequency traders (HFTrs), prevalentemente operatori esteri, [...]”. Scorrendo la stessa relazione, si sottolinea in seguito come il dato salga a circa il 63% se ci si riferisce ai soggetti registrati come HFTr ai sensi della nuova direttiva MiFID II. Carlini (2019a) mette in evidenza che nel 2018 i robot avrebbero gestito circa il 53% degli scambi delle azioni globali cash; si sta però in

questo caso considerando la più ampia fattispecie del trading algoritmo, e non semplicemente l'HFTr. Sempre nella stessa trattazione si specifica che l'algo trading nel complesso sarebbe in espansione: l'uso dei software sarebbe incentivato dalla crescente complessità dei listini di Borsa, che difficilmente possono essere gestiti dal solo operatore umano.

### 2.2.2 Strategie operative degli HFTr e impatto sui mercati

Nel precedente paragrafo si è parlato di quanto sia importante, per gli operatori ad alta frequenza, possedere una bassissima latenza nel processo di negoziazione, per essere competitivi ed ottenere un'elevata profittabilità.

Resta però da definire che tipo di attività viene svolta effettivamente dagli HFTr: cioè in che modo questi sfruttano la propria velocità e quali sono gli effetti della loro attività sui mercati.

La principale attività operativa degli HFTr si articola in:

- offrire liquidità ai mercati attraverso il market making, appropriandosi del *bid-ask spread*.
- catturare opportunità di arbitraggio.
- rispondere immediatamente, prima degli altri operatori, alle notizie di mercato (*trading on news*).
- implementare strategie manipolative del mercato.

Nella pratica, il ventaglio delle strategie implementabili è molto ampio: tale eterogeneità operativa rende ciascun HFTr differente dagli altri, complicando

spesso il processo di identificazione degli stessi da parte delle autorità di vigilanza.

L'attività maggiormente diffusa tra gli operatori ad alta frequenza è quella di *liquidity provider* (fornitore di liquidità). Molto spesso questa pratica è svolta senza assumere alcuna obbligazione formale verso il mercato, ma è anche possibile che gli HFTr offrano liquidità come *market maker* registrati. Operando in questo modo, gli HFTr traggono profitto in due modi:

- fornendo liquidità ai mercati appropriandosi dello *spread* (differenziale di prezzo) tra le quotazioni denaro (*bid*) e lettera (*ask*).
- ottenendo riduzione delle commissioni o rimborsi (*rebate*) sulle proprie operazioni, in quanto i canali di negoziazione favoriscono chi apporta liquidità, al fine di migliorare la qualità del mercato, la sua attrattività e quindi la profittabilità.

Nel primo caso gli HFTr agiscono come *market maker*: per ogni strumento negoziato espongono una quotazione di acquisto e una di vendita, garantendo ai vari operatori l'immediata esecuzione dell'operazione a quei prezzi. In questo modo, per ogni scambio gli HFTr, raccolgono lo spread tra il prezzo, più alto, al quale i partecipanti al mercato comprano, e il più basso, al quale questi vendono. Il secondo punto è legato invece a schemi incentivanti offerti, in ragione della crescente competitività tra canali di negoziazione alternativi, dai diversi mercati: quest'ultimi favoriscono infatti chi apporta liquidità, mentre sfavoriscono chi la assorbe. In particolare, gli operatori che immettono ordini al limite sono premiati, in quanto contribuiscono ad aumentare la robustezza e la profondità del *book*, attraverso un piccolo *rebate* (rimborso)

ogni volta che gli ordini vengono eseguiti. Coloro che invece inseriscono ordini al mercato saranno tenuti a pagare una commissione di un valore poco superiore ai *rebates*. Per molti HFTr tale meccanismo costituisce il business principale, data la capacità di operare su grandissimi volumi. In questo modo i canali di negoziazione riescono ad attrarre operatori<sup>7</sup>, mentre gli HFTr traggono importanti profitti su commissioni (Puorro, 2013b). Spesso, gli HFTr più avanzati, possono seguire anche una strategia mista, cercando di massimizzare i profitti appropriandosi, sia del *bid-ask spread*, sia dei *rebates* offerti dalle piattaforme. Tale strategia prende il nome di *Electronic Liquidity Provision* (ELP), e si basa sull'elevata potenza degli hardware, impiegati per negoziare simultaneamente moltissimi strumenti su più piattaforme (Bector et al., 2013).

Un altro naturale campo di applicazione dell'HFT è rappresentato dall'arbitraggio. Con tale termine ci si riferisce alla pratica attraverso cui un soggetto può trarre profitto senza correre alcun rischio. La possibilità di compiere arbitraggio si manifesta a seguito di inefficienze temporanee nel comportamento dei mercati: esiste cioè un'opportunità di arbitraggio quando una stessa attività è contrattata su due mercati a prezzi diversi o, nel caso in cui due attività siano perfettamente sostituibili tra loro, con identico *payout*.

Nel momento in cui queste opportunità si manifestano sui mercati, è chiaro che disporre di sistemi tecnologici ad alta frequenza consente di sfruttare l'arbitraggio prima di tutti gli altri operatori. Un HFTr è infatti in grado di individuare immediatamente gli squilibri presenti sui mercati e ad eseguire,

---

<sup>7</sup>I mercati ottengono la maggior parte dei loro introiti dalle commissioni richieste agli operatori sulle operazioni effettuate sulla piattaforma. Apportando nuovi operatori con la propria attività, gli HFTr, garantiscono ai mercati maggiori profitti.

in frazioni di secondo, gli ordini necessari per sfruttarne tutta l'ampiezza. Si parla in questo caso di arbitraggio da latenza, in quanto è il vantaggio tecnico, dato dalla bassa latenza, a consentire di sfruttare le opportunità di arbitraggio che si manifestano.

Un ulteriore sbocco operativo per gli HFT è rappresentato dalla possibilità di sfruttare gli effetti che le notizie economiche rilevanti hanno sulle quotazioni degli strumenti finanziari. La dotazione tecnologicamente avanzata consente a questi operatori di effettuare il “*trading on news*”: ogniqualvolta vengono pubblicati dati macroeconomici o notizie rilevanti, i software avanzati degli HFTr, sono in grado di associare *pattern*<sup>8</sup>, a strategie di trading predefinite, riuscendo così ad anticipare i movimenti dei prezzi, e quindi a trarre profitti ingenti con la negoziazione (Puorro, 2013b).

Le strategie fin qui descritte sono le più comunemente utilizzate dagli operatori ad alta frequenza. Tuttavia molto spesso gli HFTr finiscono per sfruttare la tecnologia avanzata della quale dispongono per operare strategie più aggressive e controverse, che possono finire per creare gravi scompensi alla qualità del mercato.

Tra queste attività troviamo il *flash trading*, un caso estremo dell'arbitraggio da latenza. Il *flash trading* è un servizio, offerto da alcune piattaforme di negoziazione, tramite cui quest'ultime offrono in prelazione agli HFTr gli ordini che, una volta arrivati sul mercato, non possono essere eseguiti e dovrebbero essere inoltrati ad altri mercati<sup>9</sup>.

---

<sup>8</sup>Serie di parole che statisticamente sono associate a un determinato effetto sull'andamento delle quotazioni di azioni e indici di mercato

<sup>9</sup>Quando un operatore invia per esempio un ordine al limite per  $n$  azioni ad un certo prezzo  $p$ , questo deve essere eseguito al miglior prezzo disponibile sui mercati (NBBO). Se il mercato che si fa inizialmente carico dell'ordine non ha il quantitativo sufficiente

Molto spesso il comportamento aggressivo degli operatori ad alta frequenza sfocia in vere e proprie strategie manipolative del mercato. Sono molte, in questo senso, le pratiche che gli HFTr possono mettere in atto, col fine di trarre in inganno gli operatori meno rapidi.

In generale, l'attività manipolativa consta di un momento iniziale, in cui gli HFTr osservano il comportamento dei mercati e li sondano attraverso la tecnica della *liquidity detection* (rilevazione della liquidità); c'è poi un momento successivo, in cui viene attuata una strategia adeguata al contesto operativo rilevato: generalmente si prevede di inserire elevate quantità di ordini al limite che indirizzino il mercato in un certo verso, per poi modificare o cancellare rapidamente tali ordini immessi, e trarre profitto da posizioni, intanto aperte, opposte a quelle iniziali.

Procedendo con ordine, è necessario definire l'attività preliminare di rilevazione della liquidità. Si tratta di una tecnica attraverso la quale gli HFTr inondano il mercato di piccoli ordini, al fine di ricercare importanti livelli di prezzo<sup>10</sup>, a cui solitamente sono posti ordini di *stop loss* e *take profit*<sup>11</sup>. Essendo quest'ultimi ordini al limite, quindi collocati nel *book* in maniera

---

per soddisfarlo, ed offre un servizio di flash trading, lo inoltra, per la quantità restante, all'HFTr, invece che ad un altro canale di negoziazione.

<sup>10</sup>Sono soglie di prezzo significative nell'analisi tecnica: coincidono spesso con particolari supporti e resistenze del titolo considerato. Al loro superamento generalmente viene associato un comportamento specifico degli operatori di mercato: ad esempio, il superamento al ribasso di una resistenza rilevante storicamente (oltre la quale il prezzo raramente è sceso) è un chiaro segnale di vendita del titolo.

<sup>11</sup>Gli *stop loss* sono ordini al limite impiegati per minimizzare le perdite, chiudendo automaticamente la posizione al raggiungimento di un determinato livello di prezzo, cui corrisponde la perdita massima che si intende sopportare. Uno *stop loss* associato ad una posizione lunga è un'operazione di vendita, viceversa uno *stop loss* associato ad una posizione corta è un'operazione di acquisto. Per *take profit* si intende invece un ordine al limite il cui scopo è quello di chiudere automaticamente in profitto la posizione al raggiungimento di un certo livello tecnico particolare, o di un certo livello assoluto di guadagno ottenuto dalla posizione.

latente, possono essere facilmente individuati dagli HFTr, i quali riescono a conoscere in anticipo: quali sono i livelli di prezzo su cui è posizionata la liquidità del mercato; e quale sarà il comportamento degli operatori al raggiungimento di determinate quotazioni. Il processo appena descritto è di fondamentale importanza in quanto, solo dopo aver sondato il mercato con piccoli ordini, disseminati su vari livelli di prezzo, risulta possibile comprendere le intenzioni negoziali degli altri operatori: l'HFTr potrà così simulare le condizioni di mercato ideali, alle quali sarà innescata la reazione dei *trader* oggetto di osservazione.

Una volta attuata la rilevazione della liquidità, agli HFTr non resta altro che scegliere la strategia migliore per trarre profitto dal mercato. La pratica più diffusa e raffinata è detta “ignition momentum”. Generalmente questa viene applicata a mercati in condizioni stabili, in quanto gli stessi sarebbero molto più reattivi in caso di variazioni significative del volume di negoziazione. Inizialmente l'HFTr inizia ad assumere una certa posizione immettendo un volume elevato di ordini al limite; in questa prima fase non si hanno però oscillazioni significative di prezzo. Successivamente manda al mercato un ordine con volumi elevatissimi, solo apparenti, con lo scopo di provocare la reazione degli operatori, i quali effettivamente iniziano a chiudere le proprie posizioni orientando il mercato nella direzione voluta dall'HFTr. A questo punto l'HFTr rastrella la liquidità del mercato traendo profitto dalla posizione aperta inizialmente.

Le diverse strategie manipolative degli HFTr sono tutte basate sostanzialmente su una pratica di tipo “ignition momentum”, anche se assumono nomi differenti, per via dei diversi aspetti operativi che le caratterizzano.

Lo schema resta sempre il medesimo: gli operatori ad alta frequenza inondano il *book* con ordini al limite, che vengono modificati e cancellati continuamente, in millesimi di secondo, con lo scopo di intuire il comportamento dei *trader* più lenti (prima) ed indurli a reagire a stimoli non reali (poi).

Una delle strategie più comuni è quella del *pinging*: si tratta di una pratica attraverso la quale un HFTr va alla ricerca di un grande operatore sul mercato, che si trovi in una fase di accumulo o di liquidazione della posizione, per poter contrattare con questo, a condizioni vantaggiose. La tattica risulta molto efficace nelle *dark pools*, in quanto su questi mercati negoziano generalmente i grandi operatori istituzionali <sup>12</sup>.

Un'ulteriore pratica manipolativa diffusa è quella dello *spoofing*, già trattata in maniera approfondita nel paragrafo 2.

Una pratica largamente condivisa tra gli HFTr è poi quella dello *stuffing*, la quale mira ad incrementare le opportunità di arbitraggi da latenza. Consiste nell'immissione in pochi secondi di un grande quantitativo di ordini, in modo tale da ingolfare la piattaforma di negoziazione e rallentare gli altri operatori di mercato, i quali sono impegnati a processare l'elevata mole di dati. Gli stessi ordini vengono poi rapidamente cancellati e l'HFTr sfrutta tutte le opportunità di arbitraggio che si sono venute a creare mentre gli altri trader sono occupati ad analizzare i dati.

Infine, gli HFTr, nello svolgimento dell'attività di supporto alla liquidità del mercato (*market maker*), possono adoperare una strategia di *smoking*. Pre-

---

<sup>12</sup>Le *dark pools* sono canali di negoziazione alternativi a quelli regolamentati, in cui non vengono diffuse informazioni relative agli ordini eseguiti e all'identità dei partecipanti agli scambi. Molto spesso i grandi operatori si muovono su tali canali, in modo tale da non creare gravi scompensi sui mercati regolamentati, nei quali, ordini di dimensioni considerevoli, potrebbero causare invece bruschi movimenti di prezzo.



supposto fondamentale per l'implementazione della stessa, è la capacità di essere più rapidi di tutti gli altri operatori presenti sul mercato trattato. Nella pratica, l'HFT si limita a immettere proposte di negoziazione migliori rispetto alle quotazioni offerte inizialmente, per poi modificare il prezzo, ripristinandolo ai livelli iniziali, una volta che gli operatori più lenti avranno già immesso ordini di acquisto o vendita.

Si è descritto finora come l'intensa attività degli operatori ad alta frequenza si componga di un ampio ventaglio di strategie adottabili, dalle più comuni, alle più aggressive e controverse. Non è chiaro però, a livello assoluto, quali siano stati i riflessi del trading ad alta frequenza sulla qualità dei mercati: diversi sono gli studi che ne mostrano gli effetti positivi, ma altrettanti sono quelli che ne sottolineano la dannosità.

Provando a fare chiarezza, si parte con l'illustrare il contributo positivo offerto dall'attività di HFT:

- aumento della liquidità, e contrazione del *bid-ask spread* medio, per via dell'intensa attività di *market making* svolta. Si sottolinea però che gli HFT gestiscono la liquidità in maniera strategica, e quindi tendono ad evitare la negoziazione verso controparti informate.
- riduzione dei costi di transazione.
- aumento dell'efficienza informativa dei prezzi, grazie all'attività di arbitraggio svolta.
- aumento dei collegamenti tra i vari mercati.

Tuttavia, è ragionevole sostenere, d'altro canto, che le pratiche manipolative del mercato, perseguite con una certa frequenza dagli HFTr, contribuiscano in maniera importante ad aumentare l'inefficienza dei mercati.

Tra gli effetti negativi troviamo:

- asimmetria informativa: nelle operazioni di *flash trading* gli HFTr dispongono, in via anticipata rispetto agli altri operatori, di informazioni chiave sulla pressione della domanda di mercato.
- selezione avversa: con le strategie di *pinging* in particolare, gli HFTr sfruttano la propria tecnologia avanzata, per trarre profitti a danno degli altri operatori.
- liquidità ombra (*Ghost Liquidity*): il termine è legato al fatto che nel book di negoziazione sarebbe indicata solo una liquidità apparente; cioè una liquidità che potrebbe svanire in brevi istanti al presentarsi di particolari condizioni di stress. Ciò sarebbe dovuto alla crescente diffusione nei mercati degli operatori ad alta frequenza, i quali, disponendo di strumenti informatici all'avanguardia, hanno la possibilità di inserire proposte di negoziazione fortemente mobili, le quali possono essere modificate o cancellate in millesimi di secondo. La presenza via via sempre più massiccia degli HFTr nei mercati moderni porterebbe ad avere maggiore incertezza sugli stessi. A questo proposito si cita uno studio, condotto da Caivano (2015) per conto della CONSOB, sull'impatto dell'HFT sulla volatilità dei prezzi dei titoli del mercato italiano, nel periodo tra il 2011 e il 2013: in questo si illustra come, un aumento di 10 punti percentuali del peso degli HFTr 'puri' sul volume totale de-

gli scambi, determini un aumento della volatilità *intraday* annualizzata compreso tra i 4 e i 6 punti percentuali.

- flash crash: rapido crollo del valore di uno o più strumenti finanziari, ingiustificato da ragioni economiche, seguito da un vigoroso rimbalzo. Nonostante questi crolli siano possibili anche in assenza di HFTr, la presenza di questi sul mercato amplificherebbe l'ampiezza e la rapidità dei movimenti di prezzo.

Breckenfelder (2019) propone un modello che analizzi come il comportamento degli HFTr influenzi effettivamente la qualità del mercato. I risultati dello studio mostrano come, quando nel mercato sono presenti più operatori ad alta frequenza, aumenti, l'utilizzo di strategie speculative da parte degli stessi. Questo perchè la crescente competizione, li spinge ad utilizzare pratiche più aggressive per riuscire ad ottenere buoni profitti. Il mercato diventa quindi meno liquido e la volatilità di breve termine aumenta. Se gran parte della letteratura sottolinea gli evidenti effetti positivi apportati dal trading ad alta frequenza in termini di liquidità ed efficienza del mercato, non si possono però trascurare i numerosi studi che mostrano invece problematiche evidenti insite nei mercati moderni. L'utilizzo di pratiche di mercato aggressive, apporta gravi danni agli altri operatori, e aumenta sensibilmente le possibilità di generare forti crisi di liquidità e crash di mercato. Inoltre, in relazione a mercati particolarmente tesi e nervosi, la sola presenza degli HFTr porterebbe ad amplificare i movimenti direzionali dei prezzi, aumentandone la volatilità, generando fluttuazioni estreme, come nel caso della Flash Crash del 2010.

## 2.3 La regolamentazione del trading ad alta frequenza

Il fenomeno del trading ad alta frequenza, per via del suo forte impatto sulla stabilità dei mercati, non poteva che essere sottoposto all'attenzione delle autorità di vigilanza. Dopo i fatti del 2010 risultò evidente che sarebbe stato necessario definire un perimetro normativo all'interno del quale gli HFTr potessero muoversi.

La problematica principale per i regolatori è quella di riuscire a distinguere tra l'uso inappropriato della tecnologia (pratiche manipolative), e la tecnologia stessa. Limitare l'attività di HFT di per sé, infatti potrebbe creare distorsioni di mercato ben più gravi, di quelle che si avrebbero con le eventuali pratiche aggressive degli operatori ad alta frequenza. Ecco quindi che la regolamentazione del fenomeno resta effettivamente poco chiara, non mostrando alcun orientamento definito, verso l'attività degli HFTr, i quali conserverebbero, nonostante l'ingente contenuto normativo introdotto, sufficiente spazio di manovra.

La prima risposta arrivò dagli Stati Uniti, Paese in cui il fenomeno del trading ad alta frequenza era ben radicato. L'evento del 6 Maggio 2010 mobilitò inizialmente due organi regolamentari fondamentali nel contesto americano, quali il SEC e la CFTC<sup>13</sup>. La reazione a livello normativo è però contenuta all'interno di una più ampia riforma dei mercati introdotta dal *Dodd Frank Act* (DFA). Si tratta in realtà di un intervento più complesso, voluto dal-

---

<sup>13</sup>Rispettivamente: la *Securities and Exchange Commission*, ente federale preposto alla vigilanza della borsa valori; e la *Commodity Futures Trading Commission*, un'agenzia indipendente che regola i mercati dei derivati statunitensi.

### 2.3. LA REGOLAMENTAZIONE DEL TRADING AD ALTA FREQUENZA<sup>45</sup>

l'amministrazione Obama, al fine di promuovere una completa regolazione del sistema finanziario US (Nigro and Romano, 2013). Tra le misure più importanti del DFA troviamo:

- la creazione di un istituto specifico per la supervisione della stabilità del sistema finanziario: il *Financial Stability Oversight Council* (FSOC), nelle cui competenze gravita il compito di individuare tutte le società che operano sui mercati come HFTr. L'identificazione avviene principalmente sulla base del criterio del volume di negoziazione, ma anche tenendo in considerazione l'entità del capitale proprio a disposizione delle varie società.
- l'introduzione della possibilità, da parte del SEC, di richiedere la redazione di report agli *hedge funds*<sup>14</sup>, contenenti tutte le informazioni necessarie per una corretta valutazione dell'attività del fondo, tra cui gli asset posseduti e il tipo di negoziazioni effettuate nella giornata.
- la creazione di una regolamentazione più stringente per il mercato delle *commodity*. La sezione 747 del DFA vieta in tali mercati la pratica di cancellazione dell'ordine, prima della sua esecuzione.

A supporto del DFA, nel 2011, entrava in vigore la così detta *Rule 13h-1* del "Securities Exchange Act", con la quale veniva definito un meccanismo in grado di monitorare l'attività di quei soggetti, definiti come *large trader*, che, negoziando quantità molto elevate, potessero influenzare negativamente i mercati. Nello specifico la legge stabilisce che i *large trader*, cioè tutti

---

<sup>14</sup>In quanto la maggior parte degli *hedge funds* operano sui mercati attraverso tecniche di trading ad alta frequenza.

quegli operatori che negoziano, giornalmente 2 milioni di azioni o 20 milioni di dollari, oppure che negoziano mensilmente 20 milioni di azioni o 200 milioni di dollari, devono identificarsi presso la Commissione per poter ottenere l'autorizzazione a operare sui mercati, e devono inoltre fornire una rendicontazione dettagliata della propria attività. Nel 2012 viene introdotto il "Consolidated Audit Trial" (CAT), un sistema di supervisione delle negoziazioni uniforme che ha il compito di registrare in maniera dettagliata tutte le operazioni (non solo gli ordini inseriti, ma anche tutti quelli modificati e cancellati) avvenute sui diversi canali di negoziazione americani. Nello stesso anno viene poi adottato un meccanismo di "limit up-limit down", che sostituisce il precedente sistema dei *circuit breakers*<sup>15</sup>. Con questo sistema si impedisce che le negoziazioni relative ad un titolo si concludano al di fuori di determinate fasce di prezzo, impostate in termini percentuali in relazione al valore delle quotazione del titolo nei 5 minuti precedenti.

La disciplina MiFID II, frutto del lavoro congiunto di ESMA<sup>16</sup> e Commissione Europea, è la prima normativa che cerca invece di regolamentare il fenomeno dell'HFT a livello europeo. Si tratta di una direttiva (2014/65/EU) emanata nel 2014, ma entrata effettivamente in vigore solo a partire dal 3 Gennaio 2018, il cui scopo è quello di costituire un mercato unico europeo dei servizi finanziari, nel quale siano assicurate la trasparenza e la protezione degli investitori. Occorre precisare che, in quanto direttiva europea, il recepimento delle nuove regole, diffuse attraverso questa, spetta ai singoli stati

---

<sup>15</sup>Il meccanismo tramite il quale vengono interrotte le negoziazioni temporaneamente, quando si è in presenza di un'eccessiva volatilità sul mercato.

<sup>16</sup>Autorità europea degli strumenti finanziari e dei mercati. Il suo compito è quello di supervisionare i mercati europei per garantire la stabilità del sistema finanziario.

### 2.3. LA REGOLAMENTAZIONE DEL TRADING AD ALTA FREQUENZA<sup>47</sup>

dell'Unione Europea, i quali ne prevedono il miglior adattamento in base alla rispettiva disciplina nazionale <sup>17</sup>. In concomitanza della direttiva 2014/65, il Consiglio Europeo ha emanato il Regolamento 596/2014 (cd MAR), applicato a partire dal 3 Luglio 2016, il quale definisce il quadro normativo in materia di informazioni privilegiate e manipolazione del mercato (Alagia, 2015).

All'art. 4 della direttiva MiFID II si definisce la tecnica di negoziazione ad alta frequenza, come una tecnica di negoziazione algoritmica<sup>18</sup> caratterizzata da: infrastrutture tecnologiche che minimizzano le latenze di rete o di altro genere; generazione, trasmissione o esecuzione degli ordini senza l'intervento umano; ed elevato volume infragiornaliero in relazione agli ordini (anche modificati o cancellati). All'art. 17 della stessa direttiva, vengono descritte le principali misure necessarie per tenere sotto controllo il trading ad alta frequenza:

- le imprese che effettuano trading ad alta frequenza devono notificarlo alle autorità del proprio Stato membro d'origine, nonché a quelle del canale di negoziazione su cui operano.
- gli stessi operatori, su richiesta di un'autorità competente dello Stato membro d'origine, devono fornire una descrizione sulla natura delle proprie strategie di negoziazione algoritmica.

---

<sup>17</sup>In altri termini, non trattandosi di un Regolamento Europeo (dove invece l'applicazione di tutte le norme negli Stati membri è immediata), i singoli Stati europei possono decidere in che modo perseguire le finalità imposte dalla direttiva.

<sup>18</sup>Già in precedenza si era sottolineato come l'HFT appartenga alla più ampia famiglia del trading algoritmico.

- gli HFTr, attraverso il sostegno di un modello approvato, devono conservare le registrazioni sequenziali e dettagliate di tutti gli ordini collocati (eseguiti, modificati e cancellati) e devono metterle a disposizione delle autorità se richiesto.
- gli HFTr che svolgono l'attività di *market making* devono farlo in maniera continuativa e in fasce orarie specifiche, al fine di fornire liquidità al listino in maniera regolare e prevedibile.
- le società che effettuano trading ad alta frequenza devono dotarsi di sistemi congrui a rispettare gli obblighi impostigli.

L'impressione generale è che la normativa europea sia più rigida di quella americana, e che i regolatori europei abbiano optato per un approccio più intensivo nella risoluzione del problema. (Woodward, 2017).

La crescente regolamentazione porterà probabilmente gli HFTr a rivedere le proprie strategie, spostando l'attenzione su nuovi mercati o su nuovi strumenti (criptovalute).

In generale si tratta come di un quadro normativo migliorabile. I migliori requisiti di trasparenza richiesti, non possono aumentare di per sé la capacità del sistema di resistere agli shock provocati dall'attività aggressiva degli operatori ad alta frequenza. Le autorità dovrebbero dotarsi di sistemi di controllo dei mercati più all'avanguardia, utilizzando gli stessi strumenti tecnologici degli HFTr. In questo senso quindi, le ingenti risorse a disposizione di quest'ultimi, potrebbero consentirgli di restare ancora un passo avanti rispetto ai regolatori.

Molto è stato fatto, soprattutto a livello europeo, ma una grande porzione



delle negoziazioni sui mercati continua ad essere gestita da operatori ad alta frequenza (circa il 50% degli scambi negli Stati Uniti). Questo significa che l'HFT è tutt'altro che un fenomeno in attenuazione. Più che una tecnica di negoziazione, sembrerebbe che il trading ad alta frequenza, sia una parte essenziale della struttura dei mercati, e che un'infrastruttura così fortemente tecnologica sia la causa primaria del verificarsi di eventi estremi come i flash crash.

La sensazione è che una regolamentazione più rigida, soprattutto sulle manipolazioni di mercato, abbia attenuato oscillazioni di mercato imponenti come quelle del 2010. Tuttavia le flash crash restano un fenomeno ricorrente nei mercati; crolli repentini e rimbalzi vengono registrati continuamente, e interessano sia mercati azionari che valutari; il fatto è che si tratta sempre più spesso di eventi ultra rapidi e di minore entità (che prendono il nome di *mini flash crash*), che passano inosservati ai meno esperti, ma non per questo sono meno rischiosi (Carlini, 2019b).

## 2.4 Mini Flash Crash

Oltre alle più note flash crash, occorre considerare che nei mercati sarebbero frequentemente presenti, eventi di crolli di minore entità e maggiore rapidità, definiti come “mini flash crash”. L'esistenza di questi sarebbe diventata nota solo dopo un'analisi condotta da Nanex (Nanex, 2011), che individuò le mini flash crash al verificarsi di tre determinate condizioni:

- si verifica un calo (crescita) del prezzo per almeno 10 volte consecutive prima di un'inversione.

- le variazioni di prezzo devono verificarsi entro 1,5 secondi.
- la variazione complessiva del prezzo deve superare lo 0.8%.

Considerando tutti gli eventi da comprendere all'interno di tale fattispecie, solo tra il 2006 e il 2011, ne vennero individuati circa 18000 (Johnson et al., 2012). Se si prende in esame poi una realtà come quella odierna, ancor più altamente tecnologica e automatizzata, è chiaro che le mini flash crash debbano essere considerate eventi assai frequenti.

Il meccanismo che porta a generare le mini flash crash coinvolge molto da vicino gli HFTr: sarebbero questi attraverso ordini al limite aggressivi detti "Intermarket Sweep Order" (ISO) a portare a situazioni estreme mercati poco liquidi (Golub et al., 2012).

Innanzitutto occorre definire concretamente la tipologia di ordini in questione: gli ISO sono ordini al limite che portano ad un immediata, automatica e completa esecuzione, su uno specifico mercato centrale (anche se altri presentano quote migliori) prescindendo dalla "order protection rule", che prevede l'esecuzione degli ordini al minor prezzo possibile sui vari trading venues. Si vogliono per esempio acquistare 100 azioni di una certa società: il miglior prezzo, di 1\$, si trova sul mercato X dove sono però disponibili solo 10 azioni; il mercato Y avrebbe le restanti 90 a un prezzo però di 1,10\$. Secondo la order protection rule il trader vedrebbe soddisfatto il proprio ordine solo per 10 azioni (quelle sul mercato X); immettendo invece un ordine di tipo ISO, il trader riuscirebbe ad ottenere immediatamente tutte le 100 azioni desiderate bypassando tale regola. Si tratta nella pratica di un ordine che rapidamente rastrella quante più azioni possibili sui mercati attraverso un algoritmo e che

risulta disponibile solo per trader istituzionali e market maker.

Golub et al. (2012), analizzando nello specifico le varie mini flash crash avvenute nel periodo 2006-2011, hanno evidenziato come l'aggressività degli ordini degli HFTr e la debolezza del "Regulation National Market System" (NMS), sarebbero le principali cause di questi eventi ricorrenti. La *Order Protection Rule* (norma essenziale del NMS), proteggerebbe infatti solo il livello più elevato del book di negoziazione: una volta che questo viene saturato, nulla impedisce agli ISO *order* di penetrare in maniera rapida ai livelli sottostanti del book, assorbendo gran parte della liquidità del mercato, e provocando brusche cadute delle quotazioni.

L'impiego di tale tipologia di ordini si lega a condizioni di mercato tese e incerte, in cui gli HFTr sono particolarmente reattivi. Questi infatti, captando la possibilità di un flusso informativo avverso, che li porterebbe a fornire liquidità in perdita, chiudono rapidamente le proprie esposizioni attraverso ordini ISO, e rastrellano la liquidità nell'order book, creando bruschi movimenti di mercato.

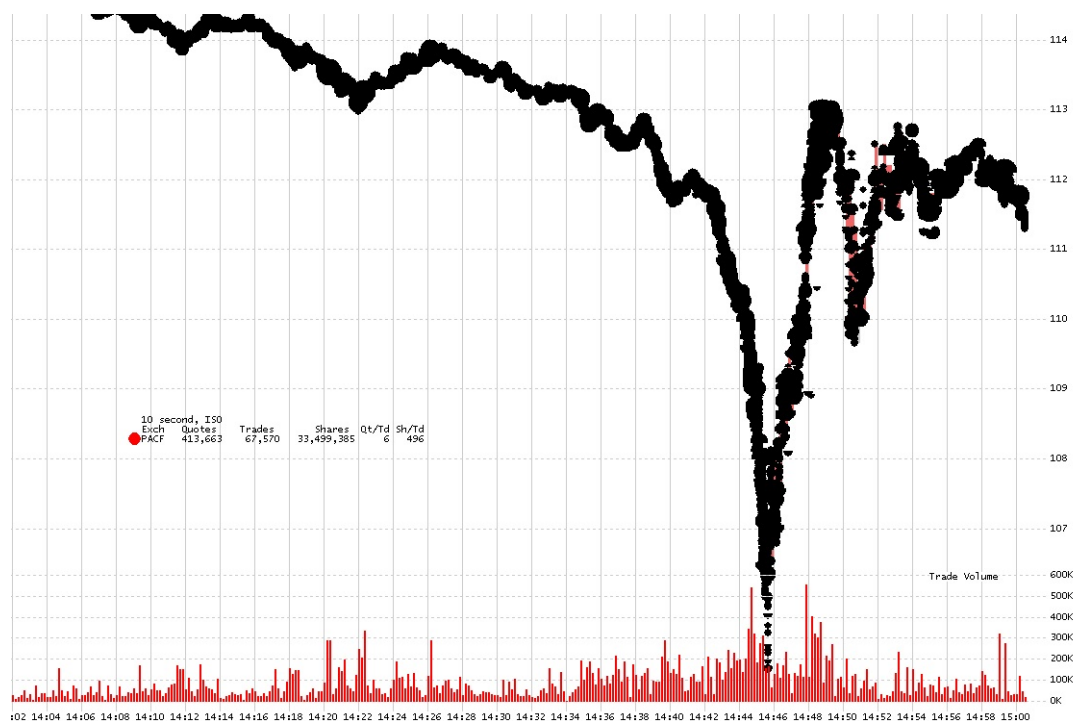
Puorro (2013b) illustra nel dettaglio come, nell'Aprile 2013, la reazione ad un *tweet* da parte degli operatori ad alta frequenza, generò una mini flash crash.

Nei momenti precedenti alla diffusione della notizia, il mercato del futures sullo S&P 500 è altamente liquido. Appena la notizia viene diffusa, gli HFTr reagiscono in maniera immediata ritirando i propri ordini e tutta la liquidità che garantiva robustezza al mercato. Al contempo, cercando di approfittare del nuovo orientamento assunto dal mercato (ribassista), inseriscono degli ordini in vendita che trovano come controparte gli operatori più lenti, i quali

nel frattempo non sono riusciti a ritirare le proprie proposte di negoziazione in acquisto. Aumentano quindi i volumi di negoziazione, con gli HFTr che rastrellano liquidità con ordini aggressivi, e si innesca un vero e proprio trend ribassista. Quando poi i numerosi ordini di *stop loss*, piazzati dai vari operatori, vengono innescati, si accelera il processo di discesa delle quotazioni. Una volta che la notizia del *tweet* non viene confermata, gli HFTr chiudono le proprie posizioni acquistando (in precedenza avevano infatti *shortato*) dagli operatori che, in ritardo, si apprestano a seguire il trend ribassista. La serie di rapidi acquisti ridà vigore alla quotazione lettera (prima rarefatta). A questo punto, vista la nuova pressione rialzista, i *trader* ritardatari si coprono dalle posizioni sbagliate assunte, rinforzando la risalita dei prezzi, i quali tornano ad attestarsi sui livelli iniziali.

Un crash di questo tipo può essere innescato ugualmente sia da uno scenario informativo avverso, che dalla presenza di particolari ordini con grossi volumi. In ogni caso però, il meccanismo di reazione degli HFTr, finisce per danneggiare i normali operatori presenti sul mercato. Tutti coloro che avessero deciso di vendere, non avrebbero infatti trovato sul mercato la liquidità necessaria, fino a pochi secondi prima presente sul mercato. Inoltre tutti gli operatori che di conseguenza, si trovano ad immettere ordini al mercato, finiscono per ottenere prezzi molto distanti dal valore ricercato a causa della rapidissima caduta dei prezzi. Infine diminuisce l'efficacia degli ordini di *stop loss*: l'attivazione degli stessi, infatti, è in questi casi economicamente dannosa, in quanto gli operatori sono costretti ad incamerare una perdita che non avrebbero subito, per via del rapido movimento successivo dei prezzi.

McInish et al. (2013) notarono come il fenomeno degli ISO order interessasse

*Figura 2.3: Ordini ISO nel 6 maggio 2010*

da vicino non solo le mini flash crash, ma anche crolli di mercato di dimensioni più rilevanti. Questi infatti, studiando la flash crash del 6 Maggio 2010, osservarono come il numero di ordini ISO in vendita, con volumi importanti, aumentò significativamente sia nella stessa giornata del crash che nei giorni precedenti. La figura 2.3 (Nanex, 2012) mostra appunto questo fenomeno. La giornata di negoziazione, come abbiamo affermato all’inizio del capitolo, era già molto turbolenta a causa delle tensioni sui mercati internazionali. Di fronte a mercati più volatili o meglio più “tossici”, gli HFTTr sono tutt’altro che fornitori di liquidità, ruolo svolto invece solitamente in condizioni di regolarità. Come il mercato iniziò, a seguito dello *spoofing*, a perdere liquidità, l’impiego degli ordini ISO da parte degli HFTTr aumentò, al fine di uscire dalle proprie posizioni esposte rapidamente, accelerando notevolmente il processo

di declino delle quotazioni. Si notò che circa il 65% del volume in vendita proveniva da ordini ISO.

La pratica d'immissione di ordini automatici e aggressivi è un fenomeno prettamente connesso all'operatività degli HFTr: il risultato del loro impiego può condurre quindi a rapidi crolli di mercato, come nel caso delle mini flash crash, ma anche, in casi più rari, in cui c'è una particolare interazione con le condizioni del mercato, a vere e proprie flash crash.

## 2.5 Order flow toxicity

Il crash del 6 Maggio 2010, come visto, si è originato dalla cattiva interazione tra una pratica di manipolazione del mercato e un algoritmo di vendita; è vero però che, probabilmente, in assenza di operatori ad alta frequenza, l'oscillazione dei prezzi sul mercato sarebbe stata più graduale e contenuta. Il comportamento degli HFTr, sintetizzando quanto detto finora, varia a seconda delle situazioni specifiche di mercato:

- in periodi di stress, con mercati particolarmente nervosi e volatili, gli HFTr amplificano le oscillazioni di prezzo, sottraendo liquidità al mercato attraverso ordini aggressivi ISO;
- in periodi di calma invece forniscono liquidità ai mercati riducendone la volatilità attraverso un'attività di market making con la quale scambiano grandi volumi; bisogna però considerare il fatto che si tratti in realtà di una "ghost liquidity", ovvero di una liquidità che potrebbe

scompare in qualsiasi momento, se gli HFTr captano la possibilità di selezione avversa.

Ciò che non cambia è il fatto che il loro comportamento, essendo sempre orientato alla massimizzazione del profitto, possa danneggiare spesso la stabilità dei mercati.

Ci sarebbe infatti uno stretto legame tra l'attività degli operatori ad alta frequenza e le flash crash.

Come evidenzia lo studio di Golub et al. (2012), gli HFTr sarebbero i principali responsabili delle crisi di liquidità note come mini flash crash.

D'altra parte, gli stessi, agendo attraverso pratiche manipolative del mercato, potrebbero innescare la reazione di altri HFTr e creare i presupposti per crash di mercato ben più marcati e radicati.

Inoltre la semplice presenza degli HFTr sui mercati sarebbe indispensabile per la formazione delle flash crash, in quanto l'elevato livello di automazione degli scambi che questi garantiscono facilita brusche cadute delle quotazioni. Resta però una domanda fondamentale a cui dare risposta: cosa spinge gli operatori ad alta frequenza ad utilizzare ordini aggressivi in certe circostanze, sottraendo improvvisamente liquidità ai mercati ed amplificando le variazioni dei prezzi?

Appurato cioè qual è il meccanismo che porta alla formazione delle crisi di liquidità e in particolare delle flash crash, ci si chiede quale sia effettivamente la causa scatenante delle stesse.

Easley, López de Prado and O'Hara (2012b), facendo riferimento all'evento del Maggio 2010, introdussero il concetto fondamentale di tossicità degli ordini (*order flow toxicity*). Come si afferma nel paragrafo 1.1, un flusso di

ordini si definisce “tossico” quando seleziona avversamente i market maker, che sono ignari del fatto di fornire liquidità in perdita.

In generale, i fornitori di liquidità nei mercati ad alta frequenza, cercano di non assumere posizione, bensì partecipano su entrambi i lati del book nel tentativo di massimizzare il turnover del proprio magazzino titoli (che in genere ruotano per almeno 5 volte a giorno). Quando il flusso di ordini è essenzialmente bilanciato gli HF market maker possono guadagnare molto da margini sottili su volumi ingenti; una volta che questo flusso diventa sbilanciato (unbalanced) verso un lato, questi operatori devono confrontarsi con la possibilità di subire perdite dovute alla selezione avversa: c'è la possibilità che diventino inconsapevolmente esposti all'azione di operatori in possesso di informazioni significative, i quali acquistano o vendono grandi quantità di titoli a un prezzo, quello offerto dai market makers, che è in realtà lontano dal proprio valore effettivo in virtù delle nuove notizie che non sono ancora note. Prima del crash del 6 Maggio si notò infatti che il volume sul mercato stava aumentando sensibilmente in maniera sbilanciata verso il lato delle vendite.

La stima della tossicità degli ordini (cioè della perdita attesa del trading verso controparti maggiormente informate) da parte degli HF market maker, diventa il fattore cruciale nel determinare il grado di partecipazione degli stessi al mercato. Se questi pensano che la tossicità sia troppo elevata, liquidano rapidamente le proprie posizioni e lasciano il mercato. Bisogna infatti pensare che gli HFTr che fungono da market maker, lo fanno solo al fine di ottenere un elevato margine di guadagno, pertanto, non essendo formalmente obbligati da nessuna istituzione ad esercitare tale attività, sono liberi di



abbandonare il mercato improvvisamente nel momento in cui lo ritengono più opportuno, creando gravi scompensi di liquidità.

# Capitolo 3

## VPIN metric

Easley et al. (2012b) produssero quindi una misura effettiva della tossicità del flusso di ordini, che consentisse di prevedere in qualche modo la possibilità del verificarsi di una Flash Crash. La misura in questione è definita come “*Volume-synchronized Probability of Informed Trading*”, o più brevemente VPIN. Per ogni periodo di tempo considerato, la VPIN *metric* è il rapporto tra il volume medio sbilanciato (*unbalanced*) sul volume totale del periodo. Nei periodi in cui ci sono molte negoziazioni basate su informazioni significative, il VPIN è molto elevato, i market maker sono nel lato sbagliato del mercato e accumulano o perdono magazzino titoli. Quando i prezzi iniziano effettivamente a muoversi, questi operatori subiscono delle perdite che, se si accumulano fino a determinati livelli, li costringono ad abbandonare le proprie posizioni ed uscire dal mercato, creando potenzialmente Flash Crash. Al crescere quindi della probabilità di trading informato con volumi significativi, cresce anche la probabilità che si verifichi una Flash Crash. Nel precedente capitolo si è descritto in maniera dettagliata la microscrittura di

una Flash Crash, cioè il meccanismo che porta alla formazione di improvvisi crolli di mercato, in contesti sempre più automatizzati, in cui gli High Frequency Trader recitano un ruolo fondamentale. Si è inoltre introdotto il concetto di tossicità degli ordini, definito da Easley et al. (2012b), in base al quale viene calcolato un nuovo indicatore di volatilità che prende il nome di VPIN (*“Volume-synchronized Probability of Informed Trading”*). L'utilizzo di tale misura di tossicità del mercato consentirebbe, a livello pratico, di prevedere il verificarsi di Flash Crash.

In questo capitolo si tratterà nel dettaglio la metodologia di implementazione del VPIN su dati ad alta frequenza. In particolare, a partire dal caso specifico del 6 maggio 2010, dal quale si deduce la bontà dell'utilizzo di tale indicatore, si cercherà di estendere il ragionamento e il calcolo al campione di dati a disposizione.

### 3.1 PIN vs VPIN

La creazione dell'indicatore VPIN, in realtà, è il frutto di un lungo processo di revisione e perfezionamento di un'altra misura di volatilità, introdotta da Easley, Kiefer, O'hara and Paperman (1996), definita come PIN (*“Probability of Informed Trading”*). Il modello PIN concepisce il trading come un gioco tra *liquidity provider* e *position taker* che si ripete ogni giorno. Le negoziazioni possono arrivare da trader informati o no. L'arrivo di ordini di acquisto o vendita da parte di trader non informati viene modellato attraverso due processi stocastici di Poisson indipendenti, con un tasso giornaliero di arrivo degli ordini Buy o Sell pari rispettivamente a  $\epsilon_B$  ed  $\epsilon_S$ . Il modello assume

inoltre che i possibili eventi di nuove notizie importanti avvengono ogni giorno con probabilità pari ad  $\alpha$ . I trader informati negoziano solo nei giorni in cui c'è un'informazione chiave, comprando se c'è una buona notizia (con probabilità  $1 - \delta$ ) o vendendo in caso di cattive notizie (con probabilità  $\delta$ ). Gli ordini che giungono dai trader informati seguono un processo di Poisson con un tasso giornaliero di arrivo pari a  $\mu$ . Sotto questo modello, la verosimiglianza (*likelihood*) di osservare B acquisti e S vendite in una singola giornata di negoziazione è pari a:

$$\begin{aligned} L(B, S|\theta) &= (1 - \alpha)e^{-\epsilon_b} \frac{(\epsilon_b^B)}{B!} e^{-\epsilon_s} \frac{\epsilon_s^S}{S!} \\ &\quad + \alpha\delta e^{-\epsilon_b} \frac{(\epsilon_b^B)}{B!} e^{-\epsilon_s + \mu} \frac{\epsilon_s + \mu^S}{S!} \\ &\quad + \alpha(1 - \delta)e^{-\epsilon_b + \mu} \frac{\epsilon_b + \mu^B}{B!} e^{-\epsilon_s} \frac{\epsilon_s^S}{S!} \end{aligned}$$

dove  $\theta = (\alpha, \delta, \mu, \epsilon_b, \epsilon_s)$  è il vettore dei parametri da stimare.

La funzione di verosimiglianza è una mistura di tre distribuzioni di probabilità di Poisson, ponderate per la probabilità di avere una giornata con “good news”  $\alpha(1 - \delta)$ , o una giornata con “bad news”  $\alpha\delta$ , o una giornata senza nuove informazioni  $(1 - \alpha)$ . Se si assume l'indipendenza tra le diverse giornate di negoziazione, la funzione di verosimiglianza su  $J$  giorni è semplicemente il prodotto tra le singole funzioni di verosimiglianza giornaliere:

$$L(M|\theta) = \prod_{j=1}^J L(\theta|B_j, S_j) \quad (3.1)$$

dove  $B_j$  ed  $S_j$  sono il numero di ordini buy e sell della  $j$ -esima giornata di negoziazione, ed  $M = [(B_1, S_1), \dots, (B_j, S_j)]$ . Si massimizza poi la funzione

(3.1) rispetto ai parametri di  $\theta$  sulla base delle varie rilevazioni  $M$ . Una volta stimati i parametri di interesse la *Probability of Informed Trading* è calcolata come:

$$PIN = \frac{\alpha\mu}{\alpha\mu + \epsilon_b + \epsilon_s}$$

Il PIN è espresso quindi semplicemente come il rapporto tra il numero di ordini informati sul totale degli ordini di negoziazione.

Occorre notare che risulta necessario ai fini della stima disporre, oltre che dei prezzi, dell'esatto numero di ordini di acquisto e vendita per ogni rilevazione; purtroppo questo non permette di implementare la stima del PIN per il campione di dati sul FTSE 100 di cui si dispone.

Se la procedura d'implementazione del PIN appare chiara e lineare dal punto di vista teorico, così non è a livello pratico. Per gli strumenti più attivi, con un elevato numero di negoziazioni giornaliere, si incorre infatti in problemi computazionali e si trova difficoltà a raggiungere una soluzione ottimale. L'utilizzo di questo indicatore è quindi stato presto abbandonato, in ragione di mercati via via sempre più automatizzati, in cui il numero di transazioni per strumento cresceva esponenzialmente. Il modello di costruzione del PIN prevede l'uso della massima verosimiglianza per stimare una serie di parametri non osservabili, che guidano il processo stocastico delle negoziazioni, per poi derivare l'indicatore a partire dalla stima di questi. Il passaggio alla logica del VPIN è fondamentale, in quanto permette il calcolo di un'indicatore affidabile di tossicità più adatto ai mercati moderni, in cui le transazioni avvengono ad alta frequenza, attraverso una stima analitica diretta e dunque senza l'impiego di tecniche numeriche intermedie per la stima di parametri

non osservabili (Abad and Yagüe, 2012).

## 3.2 La procedura di calcolo del VPIN

1. La procedura originale parte dall'identificazione di un'intervallo temporale di riferimento (*time bar*) rispetto al quale aggregare i dati.

Easley et al. (2012b) impiegano nel proprio studio una *time bar* di 1 minuto, rispetto alla quale rilevano il prezzo medio e il volume totale scambiato nell'intervallo scelto. Ciò che conta non è tanto il prezzo, quanto la variazione di prezzo registrata rispetto all'osservazione precedente.

Per ogni intervallo di tempo si dispone quindi della variazione di prezzo ( $\Delta P$ ) e del totale del volume scambiato: a questo punto si considera di avere per ogni *time bar*, invece che un'unica transazione con un determinato volume,  $n$  transazioni indipendenti, a cui associare la stessa variazione di prezzo<sup>19</sup>.

2. Il secondo passaggio, detto *volume bucketing*, rivoluziona il modo in cui i dati vengono elaborati ed ordinati. Se infatti genericamente si lavora sulle serie storiche attraverso un sistema temporale “clock-time”, cioè ordinando le varie osservazioni in base al tempo della rilevazione, la logica del VPIN porta a lavorare sulla base del “volume-time”: il campionamento dei dati è fatto sulla base del volume e non del tempo;

---

<sup>19</sup>Scelto un intervallo di riferimento di 1 minuto, si consideri per esempio la *time bar* dalle 9:04:01 alle 9:05:00. Si calcola la variazione di prezzo rispetto alla barra precedente semplicemente come  $\Delta P = P_t - P_{t-1}$ . Se il volume totale nell'intervallo è pari a 30000 transazioni e  $\Delta P = 0.01$ , si considera di avere nella barra temporale 30000 transazioni, ciascuna con una variazione di prezzo di 0.01.

cioè ogni osservazione non si riferisce più a un certo intervallo temporale di riferimento (per esempio 1 minuto), bensì a un determinato volume scambiato ( $n$  contratti). Campionare sulla base del volume significa suddividere la sessione di negoziazione in periodi con contenuto informativo comparabile, riducendo, in questo modo, l'impatto del *volatility clustering*, tipico delle serie finanziarie, presente nel campione.

Per implementare il campionamento per volume, si raggruppano le negoziazioni in maniera sequenziale in uguali *bucket* di volume di un'ampiezza  $V$ , scelta esogenamente. Ciascun *bucket* è quindi semplicemente un'insieme di più negoziazioni il cui numero totale risulta pari a  $V$ . Se le ultime negoziazioni, necessarie per completare il *bucket*, sono di un numero superiore rispetto a quello richiesto per ottenere esattamente  $V$ , quelle in eccesso sono attribuite al successivo *bucket*. Solitamente un *bucket* di volume per essere riempito necessita di più *time bar*, ma è anche possibile che il volume di una *time bar* sia sufficientemente grande da completare uno o più *bucket*.

$V$  è una grandezza fondamentale nel calcolo del VPIN, e viene determinata sulla base di un'altra variabile chiave:  $N$ , che costituisce il numero di *bucket* per giornata di negoziazione. La scelta operata da Easley et al. (2012b) è  $N = 50$ .

La procedura prevede di calcolare prima il volume medio giornaliero scambiato (*Average Daily Volume*), nell'intervallo temporale totale su cui si intende calcolare l'indicatore di tossicità (per esempio un anno e mezzo); dopo di che, scelto  $N$  opportunamente, si ottiene  $V$  sempli-

cemente come  $V = \frac{ADV}{N}$ .

Si procede infine al campionamento per volume, ottenendo le diverse osservazioni (*volume bar*) su cui costruire il VPIN<sup>20</sup>.

Se è vero che il livello totale di volume può segnalare la presenza di un flusso informativo importante, e quindi di tossicità, sul mercato, è importante anche identificare l'orientamento del segnale volumetrico per prevedere correttamente la direzione del cambiamento di prezzo. Una preponderanza di volume in acquisto (in vendita) suggerisce una tossicità che deriva da buone notizie (cattive notizie).

3. Il passaggio successivo consiste quindi nell'individuare per ogni intervallo di volume (*volume bar*), il volume in acquisto (*Buy volume*) e il volume in vendita (*Sell volume*).

Tuttavia, in questo caso, si torna a prendere come riferimento il volume scambiato nelle singole barre temporali iniziali che formano la barra di volume<sup>21</sup>: solo una volta distinto, per ogni *time bar*, il volume in acquisto da quello in vendita, si procede all'aggregazione dei risultati di tutte le barre temporali, per ottenere il volume totale in acquisto

---

<sup>20</sup>Riprendendo l'esempio in nota precedente, si considerano inoltre anche gli intervalli: 9:05:01-9:06:00 in cui si registrano 23000 transazioni con  $\Delta P = -0.02$ ; e 9:06:01-9:07:00 in cui avvengono 32000 scambi con un  $\Delta P = -0.01$ . Supposto che si scambi su quel mercato un ADV pari a 2.5 milioni di azioni si avrà una  $V = \frac{2.5M}{50} = 50000$ . A questo punto si formeranno dei nuovi intervalli di riferimento, *bucket* di volume: uno ogni 50000 transazioni registrate. In questo caso il primo *bucket* si ottiene considerando le prime due barre temporali; il volume restante di 3000 transazioni viene invece attribuito al *bucket* di volume successivo a cui contribuiranno la barra tra le 9:06:01-9:07:00 e le successive fino a raggiungere nuovamente altre 50000 transazioni.

<sup>21</sup>Con il *volume bucketing* si formano nuove osservazioni di riferimento, cioè le barre di volume (*time volume bar*); tuttavia si ricorda che ciascuna di queste, resta composta al suo interno ancora dalle varie barre temporali iniziali: ognuna delle quali caratterizzata da  $n$  transazioni con un  $\Delta P$  specifico.



e in vendita della *volume bar*. Questo passaggio fondamentale viene chiamato “Bulk classification”.

La distinzione tra volume in acquisto e in vendita è fatta in termini probabilistici utilizzando la distribuzione dei cambiamenti di prezzo, che viene standardizzata. Si sfrutta quindi nel calcolo una distribuzione normale standard.

Qualora nella singola barra temporale non ci sia alcuna variazione di prezzo, il volume in questa è diviso equamente tra  $Buy(B)$  e  $Sell(S)$ . Se invece i prezzi per esempio aumentano, il volume viene attribuito in maniera maggiore al  $Buy$  rispetto al  $Sell$ , con una ponderazione che dipende da quanto la variazione di prezzo è stata ampia rispetto alla distribuzione dei cambiamenti di prezzo.

Per ogni *volume bar* si ottiene quindi il volume totale in acquisto e in vendita come:

$$V_{\tau}^B = \sum_{i=t(\tau-1)+1}^{t(\tau)} V_i \cdot Z\left(\frac{P_i - P_{i-1}}{\sigma_{\Delta P}}\right)$$

$$V_{\tau}^S = \sum_{i=t(\tau-1)+1}^{t(\tau)} V_i \cdot \left[1 - Z\left(\frac{P_i - P_{i-1}}{\sigma_{\Delta P}}\right)\right]$$

dove  $t(\tau)$  è l'indice dell'ultima barra temporale inclusa nel  $\tau$ -esimo *bucket* di volume,  $Z$  è la funzione di ripartizione di una distribuzione normale standard, e  $\sigma_{\Delta P}$  è la stima della deviazione standard della variazione di prezzo tra barre temporali. Così il volume delle *time bar* con cambiamenti di prezzo positivi è principalmente classificato come in acquisto, mentre il volume di quelle con variazioni negative di prezzo

è principalmente in vendita. Maggiore è l'entità della variazione dei prezzi maggiore è lo sbilanciamento nel volume<sup>22</sup>.

4. Il successivo passaggio consiste nella determinazione, per ogni barra di volume, dell'*Order Imbalance* (OI), cioè dello sbilanciamento in termini assoluti tra volume in acquisto e volume in vendita. Sulla base dei valori di volume ottenuti tramite la *bulk classification*, si arriva quindi alla misura dello sbilanciamento semplicemente come:  $OI_\tau = |V_\tau^B - V_\tau^S|$ .
5. A questo punto, sulla base delle variabili fondamentali  $OI$ ,  $V$  ed  $N$ , è possibile costruire l'indicatore di tossicità VPIN come media degli sbilanciamenti negli ordini di negoziazione sulla lunghezza totale del campione.

$$VPIN = \frac{\sum_{\tau=1}^N OI_\tau}{NV}$$

Il VPIN viene aggiornato e quindi ricalcolato dopo ogni *bucket* (bar) di volume attraverso un processo di "rolling-window". Questo significa che il primo valore dell'indicatore viene prodotto a partire dal 50-esimo *bucket* (se  $N = 50$ ); quando poi il *bucket* #51 risulta riempito, viene cancellato il *bucket* #1, e viene prodotto un nuovo valore del VPIN sulla base dei *bucket* da #2

---

<sup>22</sup>Considerando l'esempio delle precedenti note, la prima barra di volume è composta dalle due barre temporali: 9:04:01-9:05:00 e 9:05:01-9:06:00. Per ciascuna di queste è necessario individuare il volume in acquisto e quello in vendita. Per la prima *time bar* per esempio si osserva un  $\Delta P = 0.01$ : tale variazione di prezzo standardizzata, viene applicata alla distribuzione normale  $Z(x)$ . A questo punto si avrà, per la barra temporale, un valore  $Z(\frac{\Delta P}{\sigma_{\Delta P}})$  che supponiamo pari circa a 0.8 e il corrispettivo valore  $1 - Z = 0.2$ . Il volume in acquisto della barra sarà dato semplicemente da  $0.8 \cdot 30000 = 24000$  mentre quello in vendita da  $0.2 \cdot 30000 = 6000$ . Lo stesso passaggio viene poi operato per la seconda barra temporale che forma la *volume bar*, con l'unica differenza che, in questo caso, avendo un  $\Delta P$  negativo ( $-0.02$ ), il volume di competenza (solo 20000 transazioni sulle 23000 totali) scambiato nella *time bar* sarà attribuito prevalentemente al volume in vendita.

a #51, e così via.

Mentre il modello PIN guarda solo al numero di acquisti e vendite per acquisire informazioni sulla struttura informativa sottostante al mercato, il VPIN dà invece rilevanza al volume delle negoziazioni. Questo è un aspetto assai rilevante per l'analisi: non si può ignorare per il calcolo della tossicità, l'intensità che viene posta nella negoziazione dai vari trader. Per questa ragione il VPIN tratta ogni operazione registrata come se fosse un insieme di più negoziazioni, ciascuna di una singola unità contrattuale: per esempio un'operazione per cinque contratti al prezzo  $p$  viene scomposta in cinque operazioni da un contratto ciascuno allo stesso prezzo  $p$ . In questo modo si pone una forte attenzione sull'intensità della negoziazione, aspetto chiave per monitorare il flusso informativo e la tossicità sul mercato.

Il VPIN può essere visto, in generale, come l'evoluzione del PIN, per due ragioni fondamentali:

- è uno strumento progettato per adattarsi bene ai mercati moderni e al contesto dell'HFT.
- riesce a produrre un'indicatore affidabile di tossicità del mercato senza incorrere nei problemi computazionali del PIN.

Inoltre uno dei pregi più grandi della metrica VPIN è dato dal fatto di lavorare in un'ottica di *volume-time*. Il campionamento per volume, sembrerebbe essere più che una valida alternativa alle specificazioni econometriche GARCH, per ridurre il fenomeno del volatility clustering presente nelle serie

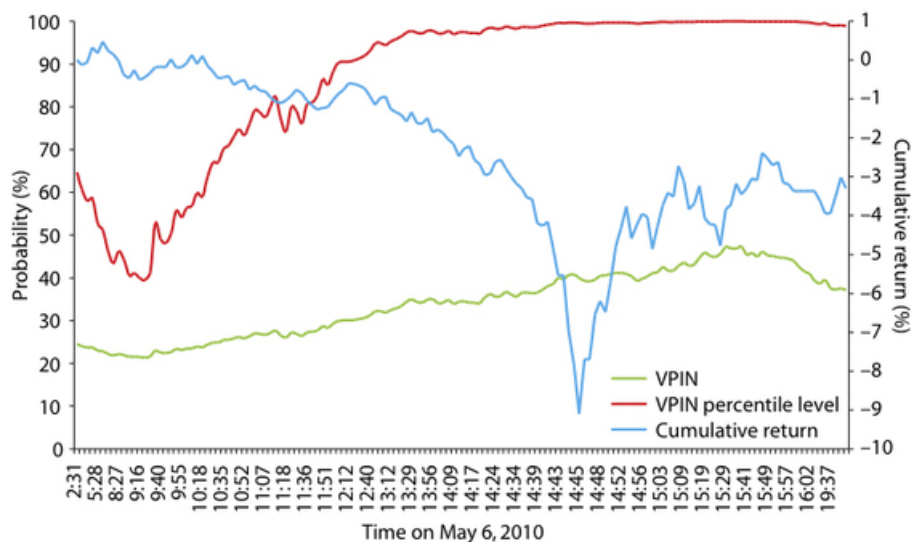
finanziarie. Mentre gli intervalli di tempo infatti non contengono quantità comparabili di informazioni, gli intervalli di volume, costruiti opportunamente, suddividono la sessione di negoziazione in periodi con contenuto informativo comparabile. Poichè ampi movimenti di prezzo sono associati ad elevati volumi, il campionamento per volume può essere considerato una proxy di un campionamento per volatilità. Il risultato è un insieme di osservazioni, la cui distribuzione è più vicina a quella di una normale e caratterizzata da una minore eteroschedasticità rispetto a quella che si avrebbe campionando in base al tempo (vedi Easley et al., 2012a).

Come visto nel capitolo primo, l'impiego di modelli GARCH non si adatta bene ai dati ad alta frequenza di cui si dispone, infatti molto spesso permane una significativa eteroschedasticità anche nei migliori modelli econometrici costituiti. Il modello del VPIN, descritto in questo paragrafo, può essere quindi nel nostro caso uno strumento necessario per riuscire a modellare i dati a nostra disposizione.

### 3.3 La funzione di ripartizione del VPIN

Analizzando nel dettaglio la metrica VPIN, in situazioni di crisi di liquidità indotte, si dovrebbe essere in grado di individuare un'aumento della tossicità del flusso di ordini sul mercato, quindi di anticipare possibili Flash Crash.

Easley et al. (2012b) notarono in primo luogo come il valore del VPIN fosse inusualmente alto la settimana antecedente il crash del 6 maggio 2010: questo avrebbe dovuto costituire già un segnale di allarme per i *market maker*. In secondo luogo osservarono come la situazione peggiorò notevolmente po-

**Figura 3.1:** VPIN metric il 6 maggio 2010

che ore prima dell'effettivo crash; in particolare, a partire dalle 11:55 a.m. del 6 Maggio, i valori assunti dal VPIN si collocavano nella coda del 10% della propria densità di probabilità, o in altri termini, superavano il 90% della propria funzione di ripartizione ( $CDF(VPIN)$ ). Più semplicemente, il VPIN stava assumendo valori estremi: a livello statistico equivale a dire che la probabilità che la variabile assumesse valori inferiori o pari al VPIN era del 90%, e cioè che  $CDF(VPIN) > 0.9$ .

La figura 3.1 evidenzia il deciso aumento della tossicità nel mercato, attraverso il valore del VPIN e della sua funzione di ripartizione. In particolare, il livello  $CDF(VPIN) = 0.9$ , sopra citato, risulta assai rilevante nell'analisi condotta: il superamento di tale soglia lascerebbe infatti presagire un'imminente brusca variazione delle quotazioni di mercato. La flash crash si verificò effettivamente poco dopo, una volta che i fornitori di liquidità si accorsero dell'elevato livello di tossicità e decisero di abbandonare il mercato.

Quindi, il raggiungimento di elevati livelli del VPIN, dovrebbero far presagire una maggiore volatilità dei prezzi imminente.

Questa tesi venne studiata esaminando nel dettaglio: la densità di probabilità dell'indicatore VPIN condizionata al valore assoluto dei rendimenti, e, viceversa, la densità dei rendimenti condizionata al valore del VPIN del periodo precedente.

In particolare Easley et al. (2012b) osservarono che:

- Considerando:  $P\left(VPIN_{\tau-1} \left| \left| \frac{P_{\tau}}{P_{\tau-1}} - 1 \right| \right.\right) \Rightarrow$  Quando i rendimenti sono elevati al tempo  $\tau$ , il valore del VPIN in  $\tau - 1$  raramente è esiguo.
- Considerando:  $P\left(\left| \frac{P_{\tau}}{P_{\tau-1}} - 1 \right| \left| VPIN_{\tau-1} \right.\right) \Rightarrow$  Quando il VPIN è basso, i rendimenti assoluti successivi sono anch'essi bassi. Quando invece il VPIN è elevato, la distribuzione condizionale dei rendimenti successivi è molto più dispersa.

Non potendo asserire che rendimenti elevati seguono un VPIN elevato, decisero di esaminare il comportamento della volatilità dei prezzi quando il VPIN restava all'interno di ciascun percentile del 5%. Ogni volta che il VPIN si muove da un intervallo di probabilità del 5% all'altro si computa il massimo rendimento assoluto tra ciascuno di due buckets intermedi, fino a che il VPIN non si sposta su un altro intervallo del 5%. Un'analisi di questo tipo è a un livello superiore, in quanto porta a considerare il massimo rendimento assoluto che si ha quando il VPIN resta all'interno di un certo intervallo di probabilità.

Il risultato di tale analisi condusse ad affermare che in realtà livelli elevati

di VPIN, devono essere mantenuti in maniera persistente, per generare ampi rendimenti assoluti.

Si è visto finora come sia importante, ai fini di prevedere correttamente una flash crash, considerare non solo i valori del VPIN, quanto quelli relativi alla sua funzione di ripartizione. In particolare è necessario che, una volta raggiunto un valore elevato di tossicità, che porta a superare la soglia di  $CDF(VPIN) = 0.9$ , il VPIN rimanga a tali livelli per un certo periodo di tempo.

A livello computazionale risulta quindi necessario, una volta calcolato il VPIN nell'intervallo temporale di riferimento, cercare di approssimare la sua densità di probabilità con una distribuzione nota, al fine di poter calcolare la funzione di ripartizione dello stesso, nell'intervallo di interesse.

Nella casistica a cui si fa riferimento Easley et al. (2011a) approssimano attendibilmente la distribuzione del VPIN con quella di una log-normale e utilizzano la funzione di ripartizione stimata per misurare il disequilibrio nel flusso degli ordini. Sarà necessario quindi anche per i dati presi in esame stimare una determinata densità di probabilità, sulla base dei valori del VPIN ottenuti, per determinare la  $CDF(VPIN)$  a cui fare riferimento, per individuare il livello di tossicità nel mercato considerato.

### **3.4 Applicazione sui dati del FTSE 100**

Si decide di seguire passo per passo la procedura di implementazione del VPIN proposta da Easley, Lopez de Prado e O'Hara, adattandola ai dati a disposizione del FTSE 100.

Ne segue che si opta per considerare nell'analisi il campione di dati con rilevazioni ogni minuto, mentre per la procedura di volume bucketing si sceglie un numero di bucket per giornata di negoziazione pari a  $N = 50$ . In relazione al volume delle negoziazioni, si prende come riferimento la variabile "ntrans", presente nel dataset originario, che definisce il numero di transazioni in millisecondi per ogni intervallo. Seppur tale grandezza potrebbe non esprimere un valore accurato della quantità di scambi per minuto, è da considerare comunque una *proxy* affidabile del volume.

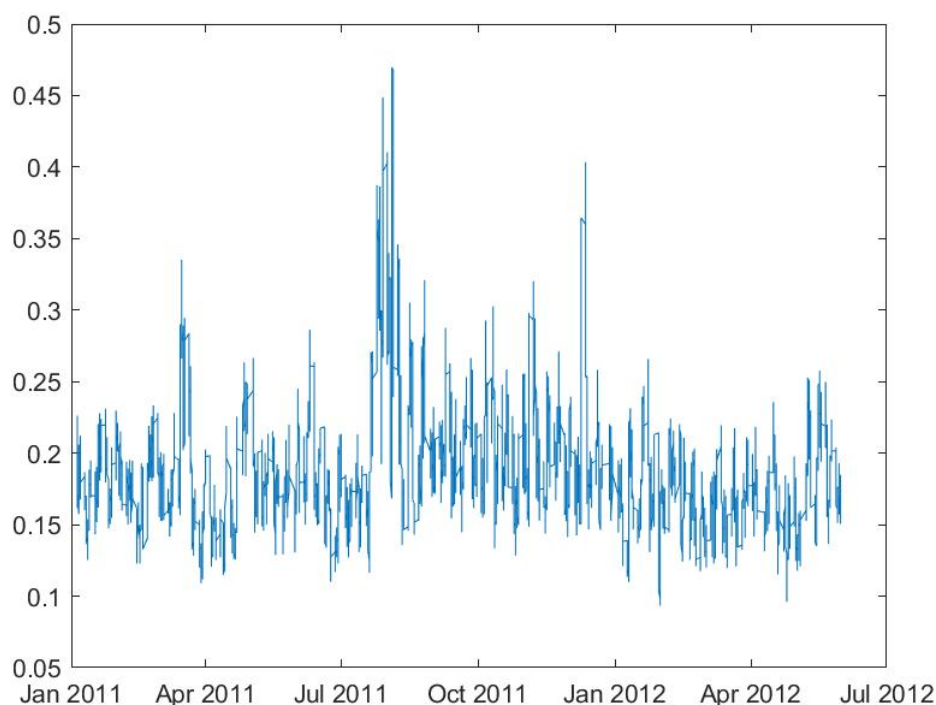
Una volta definite le variabili a cui fare riferimento, si può procedere con il calcolo dei *bucket* di volume, abbandonando quindi la dimensione originaria delle barre temporali per passare a quella delle barre di volume.

In primo luogo è necessario definire il volume medio giornaliero scambiato. Si individuano inizialmente quante rilevazioni costituiscono una singola giornata di negoziazione dopo di che, per ciascuna di queste, si calcola il numero di scambi totali effettuati, infine si passa a definire la media del volume scambiato in tutte le giornate nel campione di riferimento ottenendo la grandezza *ADV*. Supposto quindi  $N = 50$ , si ottiene  $V$  semplicemente come il rapporto tra *ADV* ed  $N$ .

A questo punto sulla base della serie dei prezzi, della serie dei volumi ("ntrans") e, definiti rispettivamente  $V$  ed  $N$ , è possibile implementare il calcolo del VPIN. A questo scopo si costruisce in MATLAB un algoritmo che, una volta inserite come input le quattro grandezze fondamentali prima definite, restituisca come output i valori del VPIN nel campione.

Essendo la serie dei volumi costituita da 182427 rilevazioni e individuato che una giornata di negoziazione è composta da 511 osservazioni, si ottengono



*Figura 3.2: VPIN sul FTSE 100 (frequenza dati: 60 secondi)*

17850 valori del VPIN, cioè 50 rilevazioni per ognuna delle 357 ( $\frac{182427}{511}$ ) giornate di negoziazione. Da questi valori ottenuti decidiamo però di eliminare le prime 50 osservazioni poichè costituite tutte da zeri, in quanto la procedura di calcolo dell'indicatore, come descritto nel paragrafo 3.2, produce la prima stima solo a partire dal cinquantesimo bucket. Si sta di fatto escludendo dall'orizzonte temporale di riferimento la giornata del 4 gennaio 2011, mentre si avranno, per ciascuna delle restanti giornate, cinquanta valori del VPIN. La figura 3.2 mostra l'andamento dell'indicatore di tossicità VPIN, calcolato relativamente all'indice azionario FTSE 100, sull'orizzonte temporale di riferimento del campione, che si estende dal 5 Gennaio 2011 al 31 Maggio 2012. Come si può notare bene dalla stessa figura, il valore del VPIN aumenta decisamente in prossimità di Agosto 2011. Non è assolutamente un caso, in

quanto, come anticipato nel primo capitolo, la prima metà di Agosto 2011 è nota come un periodo particolarmente turbolento per i mercati mondiali a causa della diffusione della crisi del debito sovrano europea. La cronaca in realtà porta alla luce la giornata di negoziazione dell'8 Agosto, appellandola come "Black Monday", in quanto, a seguito della diffusione delle notizie relative al downgrading del merito creditizio US, giunte nella notte di venerdì 5 Agosto, i principali mercati globali, oltre che ovviamente quelli americani, reagirono in maniera particolarmente brusca generando ampi crolli diffusi.

Soffermando l'attenzione sul mercato inglese, è nella giornata del 9 Agosto 2011 che l'indice azionario del FTSE 100 raggiunse l'apice del ribasso, crollando fino a quota 4800£, il livello più basso da Luglio 2010.

Considerando invece l'intervallo di tempo compreso tra l'ultima settimana del mese di Luglio 2011 e il 9 Agosto, il valore dell'indice, che si attestava inizialmente attorno quota 5900, subì in breve tempo una contrazione di più di venti punti percentuali.

Il fatto che, ad elevati valori del VPIN, corrisponda un periodo storico particolarmente volatile, è un importante segnale di conferma della bontà dell'indicatore calcolato. Questo aspetto tuttavia non è da considerarsi centrale nell'analisi svolta. L'attenzione deve essere piuttosto rivolta alle singole giornate di negoziazione che compongono il periodo di turbolenza del mercato. L'obiettivo principale è infatti quello di identificare un crollo improvviso del FTSE 100 (una flash crash) cercando poi di estendere il ragionamento esposto nel paragrafo 3.3 anche al caso in esame.

Una volta identificata effettivamente una flash crash nel campione di dati, si vuole esaminare il comportamento del VPIN e della sua funzione di riparti-

zione, nei periodi e nelle giornate di negoziazione antecedenti al crollo, per verificare la possibilità di predire in qualche modo una così eccessiva volatilità nei prezzi.

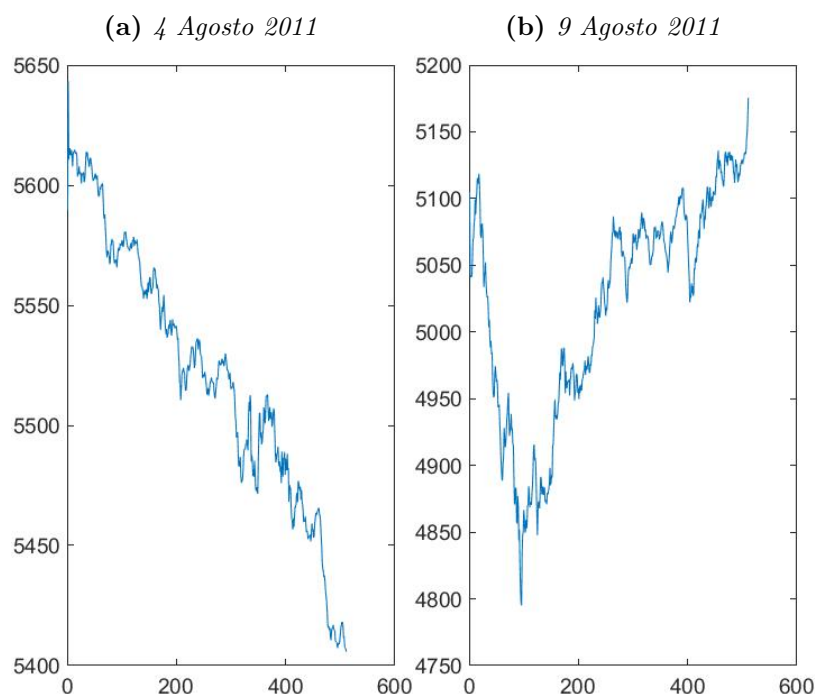
### 3.5 VPIN e Flash Crash

La discesa dei prezzi inizia effettivamente il primo giorno di Agosto 2011, quando l'indice azionario si trova ancora a una quota di circa 5900£, e raggiunge il minimo, di 4795£, solo nella giornata del 9 agosto. Esaminando nello specifico tale intervallo temporale è possibile notare come la caduta dei prezzi accelera in due punti precisi, che corrispondono a due specifiche giornate di negoziazione: quella del 4 Agosto e quella del 9 Agosto.

L'andamento dei prezzi, per entrambe le giornate, è illustrato in dettaglio dalla figura 3.3. Occorre tenere presente, come affermato in precedenza, che ogni giornata è costituita da un totale di 511 rilevazioni (una al minuto).

Nella giornata del 4 Agosto, si registra una perdita complessiva del 4.7%, con l'indice che crolla fino a quota 5405£. Tuttavia, pur trattandosi di un calo marcato, la discesa dei prezzi avviene senza registrare particolari salti o crolli improvvisi.

Differente è invece il caso del 9 Agosto. Solo nella prima ora e mezza di negoziazione, dopo un'apertura a 5104£, il FTSE 100 cede improvvisamente il 6.7%, per poi recuperare entro il termine della giornata tutto il terreno perso, attestandosi su una chiusura a quota 5175£. Si tratta questa volta di un rapido e ingente crollo del mercato, a cui segue un'altrettanto brusca ripresa: si è cioè in presenza di una flash crash. In particolare si registrano

**Figura 3.3:** Serie dei prezzi intra-day

una contrazione del 4.7% dell'indice FTSE 100 in poco più di trenta minuti, nonchè, poco dopo un'apparente breve ripresa, un'ulteriore caduta del 3.3% in appena venti minuti. È pur vero, bisogna precisare, che si tratta di un crash di minore rapidità rispetto a quello del 6 maggio 2010, quando si registrò invece una perdita di oltre il 5% in appena cinque minuti. Una volta individuata la Flash Crash nel campione, è necessario studiare il comportamento dell'indicatore VPIN nei momenti antecedenti il crollo, o meglio nelle giornate di negoziazione che precedono il 9 agosto. Coerentemente con la tesi formulata, occorre infatti riscontrare la presenza di un'elevata tossicità del flusso di ordini sul mercato, per constatare la possibilità di ampie oscillazioni dei prezzi, dovute a contrazioni di liquidità. A questo proposito è necessario non solo analizzare l'andamento del VPIN, ma anche e soprattutto quello

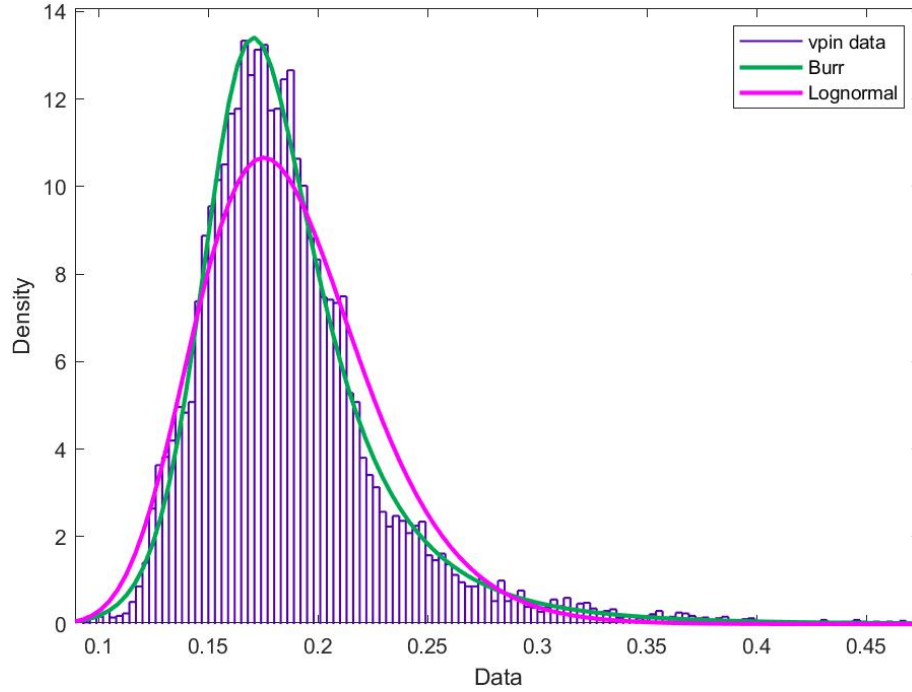
della sua funzione di ripartizione. In particolare, come descritto nel paragrafo 3.3, occorre verificare che prima del crash il VPIN abbia assunto valori estremi in maniera persistente: in termini statistici deve accadere cioè che  $CDF(VPIN) > 0.9$  per un certo orizzonte temporale. Per implementare tale metodologia di analisi occorre quindi individuare la funzione di ripartizione del VPIN. Easley et al. (2011a) approssimarono attendibilmente la distribuzione del VPIN con quella di una log-normale, ma nulla garantisce che la log-normale sia una buona distribuzione anche per i valori del VPIN da noi calcolati. È necessario, quindi, stimare la densità di probabilità che meglio si adatta al caso in esame. Per fare questo, occorre partire dalla costruzione della densità di probabilità effettiva dell'indicatore di tossicità.

Una volta costruita la densità di probabilità del VPIN, si procede a stimare differenti distribuzioni note al fine di individuare quella che meglio si adatta ai dati a disposizione.

Il risultato finale dell'analisi, effettuata considerando più di venti densità di probabilità possibili, conduce a una *distribuzione di Burr*.

La figura 3.4 mostra la bontà di adattamento di tale distribuzione ai valori del VPIN calcolati, mentre la distribuzione log-normale, impiegata invece nel lavoro originale di Easley et al. (2011a), non approssima attendibilmente gli stessi valori dell'indicatore.

Con distribuzione di Burr ci si riferisce in generale a una densità di probabilità di Burr di tipo XII, in quanto della stessa ne esistono più tipologie. In realtà la distribuzione di Burr comprende, si sovrappone o ha come casi limite, distribuzioni più comunemente utilizzate, come quelle a campana, le gamma, le loglogistiche, nonché la stessa log-normale. La densità di probabilità (PDF)

*Figura 3.4: Distribution fitting*

da noi considerata è una distribuzione di Burr di tipo XII, e viene definita, attraverso tre parametri strettamente positivi, solo per variabili casuali non negative, da un'espressione del tipo:

$$f(x|\alpha, c, \kappa) = \frac{\frac{\kappa c}{\alpha} \left(\frac{x}{\alpha}\right)^{c-1}}{\left(1 + \left(\frac{x}{\alpha}\right)^c\right)^{k+1}} \quad x > 0, \alpha > 0, c > 0, \kappa > 0,$$

La funzione di ripartizione (CDF) di tale distribuzione è invece:

$$F(x|\alpha, c, \kappa) = 1 - \frac{1}{\left(1 + \left(\frac{x}{\alpha}\right)^c\right)^k} \quad x > 0, \alpha > 0, c > 0, \kappa > 0,$$

dove  $c$  e  $\kappa$  sono gli “shape parameters” che determinano la forma della curva di probabilità, e  $\alpha$  è il parametro di scala che la contrae o l’allunga.

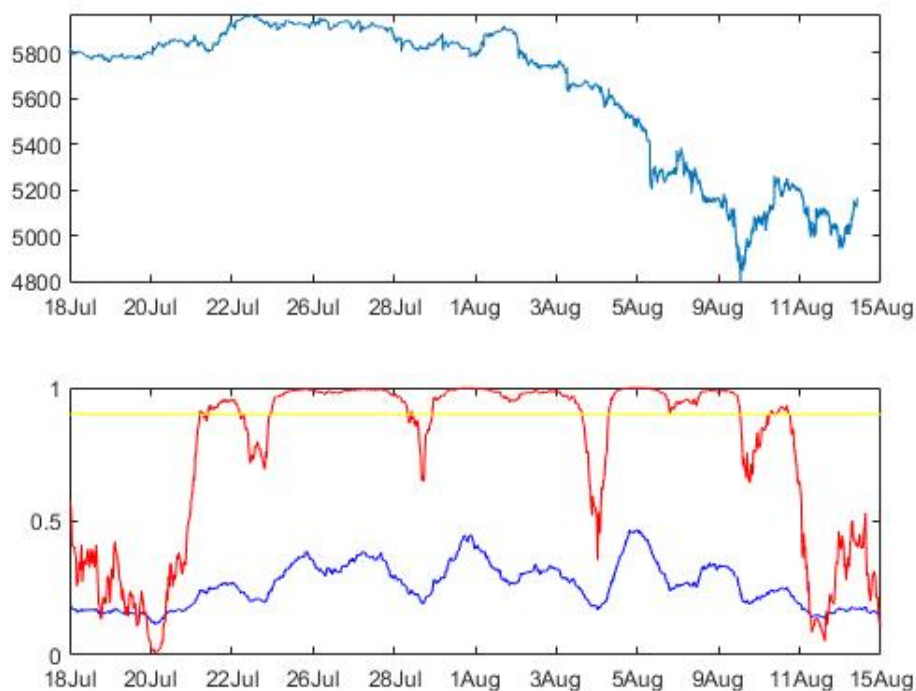
I parametri della distribuzione stimati, sulla base dei valori del VPIN, sono i seguenti:

- $\alpha = 0.1652$
  
- $c = 11.5603$
  
- $\kappa = 0.5450$

Si ricorda che si sta lavorando, in questo caso, sul campione di dati con frequenza a 60 secondi, ma lo stesso procedimento può essere applicato anche a tutti gli altri differenti campioni (appendice B). In particolare, si nota che, sviluppando il calcolo dell’indicatore di tossicità per tutte le altre differenti frequenze di rilevazione, la distribuzione che meglio approssima i valori del VPIN calcolati, resta sempre la *distribuzione di Burr*. Ciò che cambia è semplicemente il valore dei parametri che vengono stimati per ogni frequenza.

Una volta che si dispone, per ciascun valore dell’indicatore VPIN, dei corrispondenti valori della funzione di ripartizione  $CDF(VPIN)$ , per tutto l’orizzonte temporale oggetto di analisi, si è in grado di identificare un’eventuale tossicità del flusso degli ordini presente nelle varie giornate di negoziazione. Si focalizza ora l’attenzione sul particolare periodo di estrema volatilità, individuato a priori nel campione dei dati, e si studia il comportamento dell’indicatore VPIN in tale ristretto orizzonte temporale.

**Figura 3.5:** Prezzi, VPIN e CDF(VPIN) (frequenza dati: 60 secondi)



La figura 3.5 mostra, nell'arco di tempo compreso tra il 18 Luglio 2011 e il 12 Agosto 2011: nel quadrante in basso l'andamento del VPIN (in blu) e della sua funzione di ripartizione (in rosso) rispetto al livello chiave fissato  $CDF(VPIN)=0.9$  (indicato in giallo); nel quadrante in alto l'andamento della serie dei prezzi nello stesso periodo di tempo considerato. Prestando attenzione alla stessa immagine, si può notare come il CDF(VPIN) fornisce tre importanti segnali di allerta sul mercato, in quanto per tre volte supera in maniera persistente il valore fondamentale dello 0.9.

Una volta che l'indicatore si trova in una zona, che è possibile definire di tossicità, gli HFTr sostengono la liquidità solo per un periodo limitato di tempo: di lì a poco è molto probabile che gli stessi abbandonino il mercato, causando



una contrazione dei prezzi più o meno importante. Quando il CDF(VPIN) oltrepassa il livello dello 0.9 nella direzione opposta, la tossicità sta iniziando effettivamente a ridursi per via della chiusura delle posizioni da parte dei vari HFTr, esposti a selezione avversa.

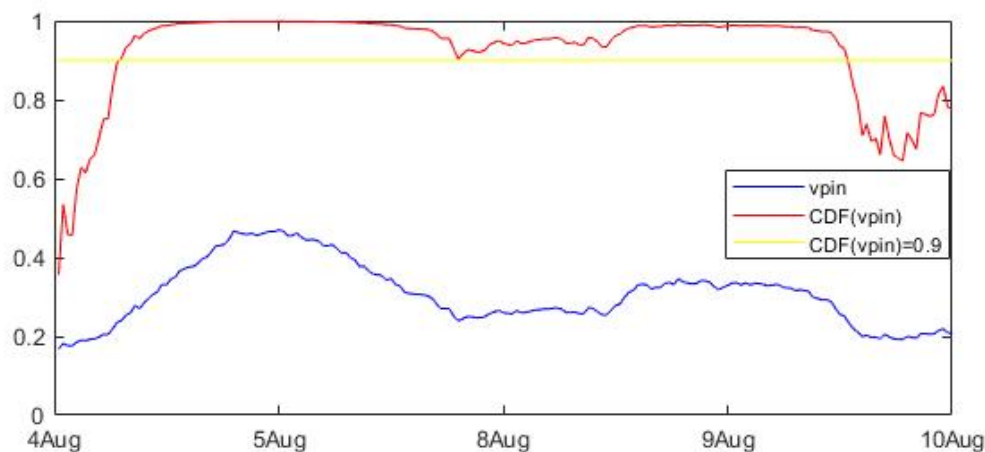
Nei momenti in cui la tossicità è persistentemente elevata, o nei momenti in cui, la stessa si riduce dopo esserlo stata, si possono registrare crash di mercato importanti, assimilabili o meno alla fattispecie delle Flash Crash. A partire dal 25 Luglio 2011, e fino al 9 Agosto 2011, il CDF(VPIN) si trova in maniera preponderante nella zona di allerta prima definita. In questo arco di tempo specifico l'indice azionario FTSE 100 registra una perdita di circa il 23% del proprio valore, passando da una quotazione di circa 5900£, a una di poco al di sotto delle 4800£.

Il 9 Agosto 2011, così come nelle tre giornate precedenti, il CDF(VPIN) si trova ai livelli massimi (vicino al limite superiore di uno): è possibile quindi affermare che l'elevata tossicità, valutata tramite l'indicatore, avrebbe consentito di prevedere la Flash Crash verificatasi poi effettivamente sul FTSE 100. In realtà il segnale di allerta è molto forte, in quanto come già mostrato nella figura 3.5, dal 25 Luglio l'indicatore si è mosso in modo quasi esclusivo nell'area di tossicità. Si decide di considerare le perdite massime registrate all'interno<sup>23</sup> delle singole giornate di negoziazione, in un intervallo di tempo pari ad un'ora.

Si confrontano poi le statistiche relative alle perdite massime a un'ora del periodo di estrema volatilità prima definito (dal 25 Luglio 2011 al 9 Agosto

---

<sup>23</sup>Significa che si escludono dall'analisi tutte le variazioni dei prezzi a cavallo tra una giornata di negoziazione e l'altra, considerando solamente ciò che avviene all'interno del giorno di contrattazione

**Figura 3.6:** Tossicità prima del 9 Agosto 2011 (frequenza dati: 60 secondi)

2011), con quelle rilevate invece nelle giornate di negoziazione del resto del campione di dati. I risultati, illustrati nella tabella 3.1, evidenziano come la

**Tabella 3.1:** Perdite massime a un'ora

Periodo	Media	1° quartile	Mediana	3° quartile	Perdita massima
25/07/2011-9/08/2011	-1.4311%	-1.7636%	-1.1345%	-0.6961%	-4.4137%
Mese precedente	-0.5376%	-0.7566%	-0.4455%	-0.3396%	-1.2751%
Resto del campione	-0.6282%	-0.7936%	-0.5632%	-0.3754 %	-2.4623%

tossicità elevata individuata dall'indicatore sfoci poi effettivamente in perdite significativamente superiori a quelle registrate, non solo nel mese precedente di negoziazione, ma anche in tutto il resto del campione<sup>24</sup>.

Da un punto di vista prettamente operativo, l'impiego dell'indicatore di tossicità permetterebbe ai partecipanti al mercato di uscire da quest'ultimo prima di subire perdite ingenti dovute a contrazioni di liquidità. Una volta che la funzione di ripartizione del VPIN entra nella zona di tossicità, è conveniente per un operatore finanziario abbandonare il prima possibile la propria

<sup>24</sup>Il quale comprende tutte le giornate di contrattazione tra il 4 Gennaio 2011 e il 31 Maggio 2012, escludendo quelle tra il 25 Luglio 2011 e il 9 Agosto 2011

esposizione; generalmente questo risulta possibile in quanto, una volta che il mercato inizia ad essere tossico, i *market maker* continuano comunque a sostenere la liquidità per un certo arco di tempo, soprattutto i più capitalizzati (Easley et al., 2011b).

Si è visto finora che l'utilizzo dell'indicatore VPIN ha permesso di individuare correttamente un periodo di elevata volatilità sui mercati, e di anticipare il verificarsi della Flash Crash del 9 Agosto 2011. Più in generale, il meccanismo consente di rilevare tutte quelle tensioni sui mercati che finiscono per sfociare in crisi di liquidità.

L'analisi iniziale conferma quindi sostanzialmente i risultati a cui erano giunti Easley et al. (2012b). Si è però focalizzata l'attenzione soltanto su una parte ristretta del campione di dati, individuata a priori come particolarmente turbolenta. Risulta invece necessario, al fine di fornire una trattazione completa, testare la bontà dell'indicatore di tossicità su tutto il dataset a disposizione.

### **3.6 Il VPIN nel campione completo**

Il possesso di diverse frequenze di rilevazione per i dati finanziari del FTSE 100, permette di approfondire l'analisi, studiando come varia il comportamento del VPIN al variare della frequenza.

Si vuole mostrare, in particolare, per ogni differente frequenza di rilevazione, quali e quanti sono i periodi di estrema volatilità individuati dall'indicatore, e se si assiste, sempre al variare della frequenza, a un fenomeno di sovrapposizione dei periodi critici nell'orizzonte temporale di riferimento del dataset. Si scelgono inoltre differenti soglie di significatività, oltre a quella originale

a 0.9, per definire la zona critica di tossicità. Questo permette di osservare come varia il processo di identificazione dei periodi critici al variare del grado di significatività della soglia del  $CDF(VPIN)$  scelta.

I risultati completi di tale analisi sono illustrati nell'appendice B. Se ne riassumono in questa sede i tratti salienti, a partire dal procedimento seguito, fino ai principali risultati raggiunti.

Fissata la soglia di significatività, si considerano tutti i valori della funzione di ripartizione del VPIN al di sopra di quest'ultima. Si crea quindi un vettore di 17800 elementi (numero totale di valori del VPIN) i quali assumono il valore di 1 quando il  $CDF(VPIN)$  supera tale soglia, e valore zero altrove. Si ricorda però che, come stabilito nel paragrafo 3.3, è importante che la tossicità rimanga a livelli elevati per un certo periodo di tempo; è necessario quindi trascurare tutti i casi isolati in cui il  $CDF(VPIN)$  varca la soglia di significatività in maniera temporanea. Al fine di scremare questi casi, si introduce una media mobile a 10 periodi: quando la media mobile assume valore 1, significa che il  $CDF(VPIN)$  è entrato nell'area di tossicità per almeno 10 rilevazioni consecutive. Ovviamente la scelta dei 10 periodi è compiuta in maniera del tutto soggettiva; in questo caso si sta considerando una tossicità persistente che interessa un arco temporale di circa un'ora e mezza<sup>25</sup>. Si individuano in questo modo tutti i vari periodi critici all'interno del dataset. Non resta poi che associare ad ognuno di questi le specifiche giornate di negoziazione a cui appartengono. La tabella 3.2 mostra la per-

---

<sup>25</sup>Si dispone infatti di 50 rilevazioni del VPIN (e quindi del  $CDF(VPIN)$ ) per giornata di negoziazione. Considerarne 10 consecutive, vuol dire prendere come riferimento un'orizzonte temporale di circa 96 minuti, supposto che le contrattazioni perdurino per 8 ore al giorno.

**Tabella 3.2:** Percentuale di criticità sul dataset completo

Frequenza	Valori sopra la soglia			Percentuale di criticità		
	0.90	0.95	0.99	0.9	0.95	0.99
300s	10.94%	5.40%	0.67%	7.84%	3.47%	0.28%
120s	8.91%	5.36%	1.40%	7.11%	4.32%	1.06%
60s	9.98%	5.19%	1.15%	7.46%	3.92%	0.88%
30s	9.12%	5.22%	1.02%	6.83%	3.98%	0.64%
10s	9.97%	5.22%	0.79%	7.12%	4.14%	0.34%
5s	10.16%	5.43%	0.51%	7.47%	3.9%	0.31%
1s	10.37%	5.71%	0.61%	7.67%	4.3%	0.24%

centuale di criticità rilevate sul totale del campione servendosi del metodo sopra descritto.

Come si evince dalla riduzione delle percentuali tra la prima macrocolonna e la seconda, la media mobile calcolata, ha permesso effettivamente di scremare i diversi casi isolati di valori del VPIN solo temporaneamente elevati. Escludendo la casistica a 300 secondi, si può notare che l'aumento della frequenza, a parità di livello di significatività, mantiene pressochè stabile la percentuale di criticità rilevate sul campione. Solo a livello 0.99 si ha una minima riduzione delle criticità rilevate.

Si può osservare d'altro canto che fissata la frequenza, l'aumento della soglia di significatività scelta, permette di ridurre la percentuale di criticità individuate: si può regionalmente pensare infatti che un livello di significatività più elevato garantisca un numero di segnali minore, ma con un'affidabilità superiore.

Spostando invece l'attenzione sui periodi effettivamente individuati dalla funzione di ripartizione del VPIN, si osserva che esiste un fenomeno di sovrapposizione degli orizzonti temporali considerati critici (si veda l'appendice B).

In particolare, esistono delle giornate critiche specifiche, che vengono individuate allo stesso modo da tutti i vari indicatori di tossicità a frequenze differenti. Si nota poi che, in molti casi, pur non essendoci una corrispondenza perfetta tra le varie giornate rilevate, esistono delle sovrapposizioni a livello di macro periodi: cioè i diversi VPIN, pur non indicando le stesse specifiche giornate critiche, ne indicano altre contigue. È possibile quindi individuare degli orizzonti temporali più generici che sono ritenuti tossici dalla maggior parte degli indicatori VPIN a tutte le frequenze. Questi sono indicati nella tabella B.2 in appendice.

Come già notato, relativamente alla tabella 3.2, l'aumento della soglia di significatività porta a ridurre il numero di criticità rilevate. Se a livello 0.9 sono diversi i periodi individuati, al livello più alto, di 0.99, gli unici periodi considerati critici sono quelli relativi alla fine di Luglio 2011 e l'inizio di Agosto 2011. Questo conferma in un certo senso l'ipotesi, prima definita, secondo cui una soglia di significatività più elevata porterebbe a segnali minori ma più importanti. Quello che accade è infatti che, a livello massimo considerato, siano forniti principalmente solo due segnali, i quali permettono di anticipare un Crash di mercato molto marcato.

Tuttavia una maggiore efficacia dell'indicatore VPIN, ad un livello di significatività più elevato, è stata supposta solo relativamente ai fatti dell'Agosto 2011, trascurando tutto ciò che accade nel resto del dataset: questo aspetto verrà quindi approfondito meglio nel prossimo paragrafo.

Si precisa inoltre che col termine efficacia si vuole intendere la capacità dell'indicatore di generare solo segnali importanti di tossicità: che siano cioè capaci di predire le perdite più rilevanti sul mercato. Con il termine affi-

dabilità (o attendibilità) invece ci si riferisce, più in generale, alla capacità dell’indicatore di funzionare correttamente: e quindi di riuscire a individuare le varie crisi di liquidità nel campione, che si sostanziano in perdite superiori alla media.

Per definire un’attendibilità a livello assoluto dell’indicatore è necessario perfezionare l’analisi fin qui effettuata, considerando anche ciò che accade in corrispondenza di tutti gli altri periodi identificati come tossici dal VPIN.

### **3.7 L’Affidabilità dell’indicatore VPIN**

Definire concretamente l’attendibilità del VPIN è qualcosa di complesso.

È noto infatti che questo indicatore è stato creato al fine di prevedere la possibilità di crolli di mercato repentini quali le Flash Crash. Non esiste però una definizione tecnica che permetta di identificare quest’ultime. A quale livello di perdita dovrebbe arrivare un ribasso delle quotazioni e in quanto tempo dovrebbe realizzarsi, per essere considerato una Flash Crash? Non esiste una risposta univoca a tale domanda.

Ecco quindi che definire la capacità del VPIN di identificare le Flash Crash risulta un processo non completamente oggettivo.

Si può però quantomeno cercare di verificare se effettivamente le situazioni di tossicità, individuate per mezzo del VPIN, si traducano effettivamente in crisi di liquidità che sfocino in drastiche riduzioni delle quotazioni.

Low et al. (2018) misurarono l’accuratezza dell’indicatore VPIN su diversi mercati internazionali in base al numero di crolli delle quotazioni registrati una volta che la CDF(VPIN) aveva superato una soglia di rilevanza definita

a priori (per esempio 0.9). Il ragionamento venne poi esteso considerando diversi orizzonti temporali entro i quali potessero manifestarsi effettivamente le perdite, ad esempio entro 5, 20, 60 giorni.

Si decide quindi di seguire, in parte, la metodologia di analisi proposta da Low et al. (2018) e di ragionare sulle perdite effettivamente registrate, nelle giornate di negoziazione successive al rilevamento di un'elevata tossicità sul mercato. In particolare, si sceglie di considerare le perdite massime registrate nella specifica giornata critica (individuata col procedimento esposto nel paragrafo 3.6) e nelle successive 4 giornate di contrattazione.

Nel momento in cui, all'interno di una specifica giornata, si va a ricercare la perdita massima, è necessario definire un'orizzonte temporale di riferimento entro cui quest'ultima debba realizzarsi. Essendo la nostra analisi basata sulla prevedibilità di crolli repentini nel mercato, si considerano perdite che si concretizzino in un arco di tempo compreso tra i 10 e i 30 minuti.

L'obiettivo principale è quello di confrontare l'andamento delle perdite massime che si verificano nelle giornate critiche (e nelle 4 successive), con quello delle perdite massime che si realizzano invece nel resto del campione di dati. Questo tipo di analisi viene condotta lasciando variare, per ogni differente VPIN <sup>26</sup>, il livello di significatività e l'arco temporale di riferimento delle perdite massime.

I risultati dello studio fin qui descritto sono mostrati dettagliatamente nelle tabelle in appendice B.

Per tutti i differenti VPIN e per tutti i livelli di significatività, considerare

---

<sup>26</sup>Si ricorda infatti che si sono calcolati 7 differenti VPIN: uno per ogni campione di dati disponibile



orizzonti temporali più ampi si traduce nel registrare perdite massime medie più elevate. Il discorso vale sia per i periodi critici che per quelli non critici. Tale risultato suggerisce che la maggior parte delle perdite massime registrate all’interno delle giornate di negoziazione del campione, impiegano del tempo per prodursi: cioè in occasione delle perdite più importanti i prezzi non decrescono solo per un breve lasso di tempo, ma continuano a farlo per periodi di almeno 30 minuti.

Mantenendo invece fisso l’arco temporale delle perdite registrate, si può notare come, in generale, l’aumento della soglia di significatività comporti un’aumento delle perdite massime medie rilevate. Il dato conferma a livello statistico l’ipotesi formulata nel paragrafo 3.6, in base a cui a livelli di significatività più elevati corrisponderebbe una maggiore efficacia dell’indicatore: i segnali inviati a livello 0.99 infatti predicono perdite massime medie superiori a quelle predette dai segnali inviati a livelli inferiori. Sono presenti però due eccezioni: per il VPIN a 300s e quello a 10s l’aumento della soglia di significatività, dal livello 0.95 a quello 0.99, non comporta effettivamente un aumento dell’efficacia.

Soffermandoci invece sull’obiettivo principale dell’analisi, i periodi critici, identificati per mezzo del VPIN, individuano, in media, perdite massime superiori di 2.08 volte rispetto a quelle dei periodi non ritenuti rilevanti. Questo ci permette di dire che l’indicatore di tossicità risulta affidabile anche una volta applicato sull’intero campione di dati.

Si vuole cercare però di stabilire a quale frequenza di calcolo l’indicatore VPIN potrebbe essere più attendibile.

Per i VPIN a 120, 60, 30 e 10 secondi, i valori assoluti delle perdite massime

medie rilevate nei periodi critici sono piuttosto stabili. A livello dei 300s le stesse perdite sono invece più contenute, mentre per i VPIN a 5 secondi e ad 1 secondo si registrano valori superiori: questo lascia pensare che il calcolo del VPIN con dati a frequenze più elevate permette di individuare crolli maggiormente significativi. Tuttavia ciò non significa che impiegare VPIN a 5 secondi o ad 1 secondo sia la scelta migliore. Per confermare una valutazione di questo genere occorre rapportare il livello delle perdite massime dei periodi critici con quello delle stesse nei periodi non critici: l'indicatore potrebbe infatti rilevare ottimamente situazioni di volatilità estrema, senza considerare però casi che sono comunque molto rilevanti<sup>27</sup>. Questo aspetto non può essere sottovalutato in quanto, trascurare alcune crisi di liquidità, seppur meno significative, può voler dire per un grande investitore, subire ingenti perdite. Si esegue quindi il calcolo per ogni differente VPIN. I ri-

**Tabella 3.3:**  $\frac{\text{Perdite max. medie periodi critici}}{\text{Perdite max. medie periodi non critici}}$

300s	120s	60s	30s	10s	5s	1s
1.921	2.188	2.249	2.203	1.914	2.011	2.074

sultati, illustrati nella tabella 3.3 suggeriscono che in realtà l'utilizzo di un VPIN a 60 secondi sembra essere la scelta migliore. La stessa frequenza dei dati adottata da Easley et al. (2012b) dovrebbe quindi essere da preferire rispetto alle altre.

Cosa si può dire invece riguardo al livello di significatività da adottare?

Rapportando sempre le perdite massime medie dei periodi critici con quelle

<sup>27</sup>I quali se tralasciati, vanno ad incrementare la media delle perdite massime, riducendo il rapporto in esame.

dei periodi stabili<sup>28</sup>, si scopre che in realtà la soglia di significatività a 0.9 risulta essere la migliore<sup>29</sup>.

In conclusione, lo studio compiuto relativamente alle perdite massime medie registrate nei periodi critici, mostra come il calcolo del VPIN con dati a frequenza 60 secondi e livello di significatività allo 0.9 sia quello maggiormente affidabile nel rintracciare crisi di liquidità sul mercato.

Aumentare la soglia di significatività nell'analisi della tossicità del mercato, significa invece ottenere un'indicatore migliore nel segnalare i movimenti di mercato più bruschi<sup>30</sup>, ma meno affidabile in via generale, poichè trascura dei segnali che possono essere comunque rilevanti per un investitore<sup>31</sup>.

---

<sup>28</sup>Si lascia variare in questo caso il livello di significatività.

<sup>29</sup>Il rapporto risulta pari a: 2.518 per la soglia a 0.9, 2.143 per quella a 0.95 e 2.085 per quella a 0.99.

<sup>30</sup>Come si è visto il VPIN a livello di significatività pari a 0.99, è più efficace nel rilevare perdite massime medie superiori a quelle degli altri livelli.

<sup>31</sup>Nel campione dei dati infatti, un VPIN a livello 0.99 ha permesso di individuare solo il crash di mercato più importante, trascurando invece pressochè tutti gli altri meno rilevanti

# Capitolo 4

## Conclusioni

Il brusco crollo dell'indice Dow Jones avvenuto il 6 Maggio 2010, ha portato alla luce dei problemi evidenti insiti nella struttura dei mercati moderni, sempre più saturi di operatori che negoziano attraverso tecniche di *High Frequency Trading*. La presenza massiccia di tali soggetti renderebbe i mercati molto più volatili soprattutto in periodi di particolari tensioni. Nonostante molto sia stato fatto a livello regolamentare, permane la possibilità di registrare importanti perdite repentine sui mercati, legate a crisi di liquidità indotte. L'inutilità dei tradizionali modelli di volatilità condizionale nella previsione di shock di questo genere, ci ha portato a focalizzare l'attenzione sul modello statistico proposto da Easley et al. (2012b). Tale modello, basato su ragionevoli assunzioni teoriche sul funzionamento dei mercati moderni, permette di individuare la possibilità di crisi di liquidità attraverso la costruzione dell'indicatore di tossicità VPIN. Ripercorrendo i passaggi fondamentali dello studio di Easley et al. (2012b) si è costruito l'indicatore VPIN per tutti i dati ad alta frequenza a disposizione. Si è voluto poi completare

l'analisi mostrando come varia il comportamento dello stesso indicatore al variare della frequenza di rilevazione e del livello di significatività scelto.

Occorre ricordare che i risultati ottenuti non possono essere presi per validi a livello assoluto in quanto sono molti gli elementi di soggettività introdotti, seppur in maniera del tutto ragionevole

L'esito dell'analisi conferma in realtà la bontà del VPIN originale. L'utilizzo di una frequenza dei dati a 60 secondi sembra effettivamente la soluzione migliore per ottenere un indicatore il più affidabile possibile. La stessa soglia di significatività, fissata a un livello di 0.9, sarebbe da preferire rispetto alle altre, sempre nell'ottica di una migliore attendibilità.

L'aumento del livello di significatività sopra lo 0.9 consentirebbe invece di ottenere un indicatore migliore nel rilevare le perdite più rilevanti sul mercato, ma meno affidabile in via generale, in quanto trascura segnali comunque importanti relativi alla tossicità del mercato.

Un VPIN a livello di significatività pari a 0.99 potrebbe essere utile qualora si vogliano prevedere solo movimenti estremi di mercato, assimilabili alla fattispecie delle Flash Crash. L'applicazione sul campione dei dati a disposizione ha mostrato infatti che l'unico periodo critico individuato attraverso un VPIN a livello 0.99 è quello che precede un crollo di mercato ingente, della portata di 6.4 punti percentuali.

Un VPIN a 60 secondi invece, con soglie di significatività inferiori, produce più segnali che non permettono di identificare in maniera specifica la possibilità di un crollo molto marcato. Il pregio principale in questo caso è però quello di allertare in maniera affidabile l'investitore ogni qual volta ci sia una crisi di liquidità imminente, a prescindere dall'entità della stessa.

# Appendice A

## Stime preliminari

### A.1 Modelli di volatilità condizionale

Si inseriscono le stime relative ai modelli sulla varianza condizionale per i campioni di dati a più alta frequenza.

In particolare si considerano in maniera sequenziale i dataset a 10 secondi, a 5 secondi e a 1 secondo, mostrando come i risultati ottenuti per i diversi modelli prodotti siano poco significativi.

Le stime sono prodotte attraverso il software MATLAB, il quale predispono per l'analisi della varianza condizionale, principalmente 3 tipologie di modelli: GARCH, EGARCH e GJR-GARCH. Per ogni modello viene mostrato il valore dei parametri stimati e dei criteri informativi.

I modelli sulla media su cui sono incentrate le stime della varianza condizionale, sono quelli elencati nella tabella 1.1 del capitolo primo.

Si evidenzia inoltre l'eventuale presenza di problematiche nelle stime:

- i modelli che riscontrano problemi di convergenza, e non giungono









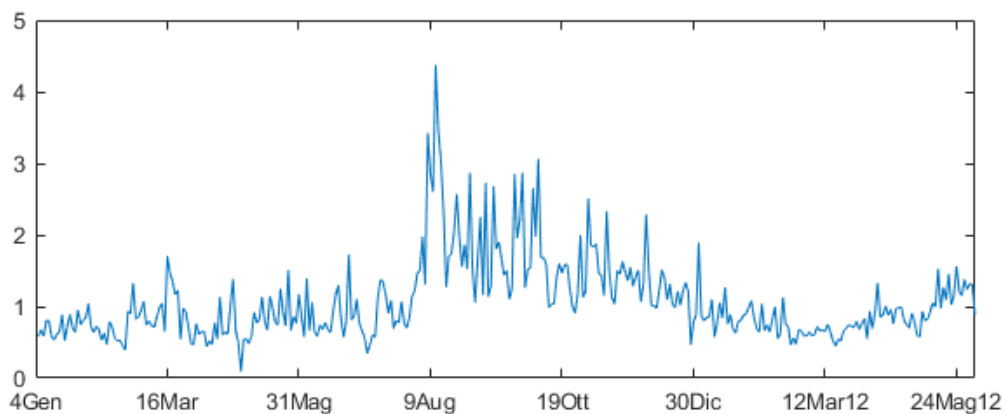
**Tabella A.4:** Stime dei modelli FIGARCH

Frequenza	Modello	Probl. Innov.	$\omega$	$d$	$\beta$	$\phi$	CI		
							AIC	BIC	
300s	-	NaN	-	-	-	-	-	-	
120s	FIGARCH(0,d,0)	✓	N	0.0016	0.1854	0	0	$-2.3599 \cdot e^5$	$-2.3597 \cdot e^5$
	FIGARCH(1,d,0)	✓	N	0.0016	0.1854	$2.23 \cdot e^{-9}$	0	$-2.3598 \cdot e^5$	$-2.3596 \cdot e^5$
60s	FIGARCH(0,d,0)	✓	N	$7.4486 \cdot e^{-4}$	0.2098	0	0	$-6.1293 \cdot e^5$	$-6.1291 \cdot e^5$
	FIGARCH(1,d,0)	✓	N	$7.4485 \cdot e^{-4}$	0.2098	$8.86 \cdot e^{-9}$	0	$-6.1293 \cdot e^5$	$-6.1291 \cdot e^5$
30s	FIGARCH(0,d,0)	✓	N	$3.6453 \cdot e^{-4}$	0.2184	0	0	$-1.5186 \cdot e^6$	$-1.5186 \cdot e^6$
	FIGARCH(1,d,0)	✓	N	$3.6454 \cdot e^{-4}$	0.2184	$3.0304 \cdot e^{-9}$	0	$-1.5186 \cdot e^6$	$-1.5186 \cdot e^6$
10s	-	NaN	-	-	-	-	-	-	
5s	-	NaN	-	-	-	-	-	-	
1s	-	NaN	-	-	-	-	-	-	

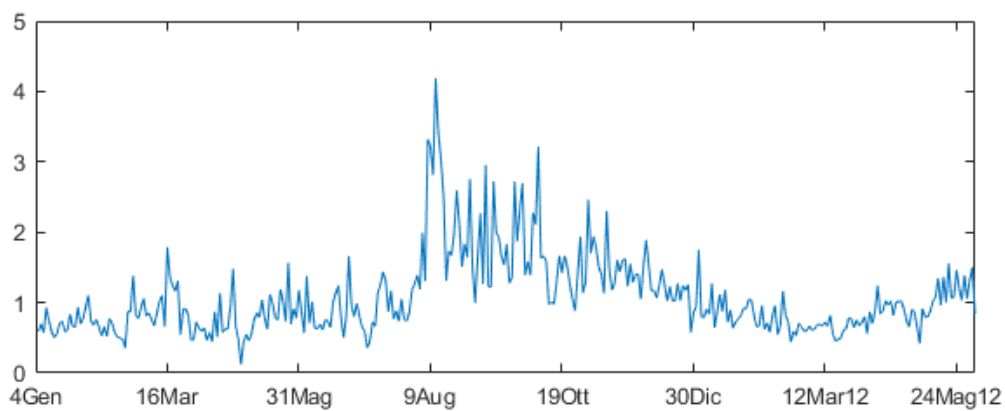
## A.2 Volatilità Realizzata

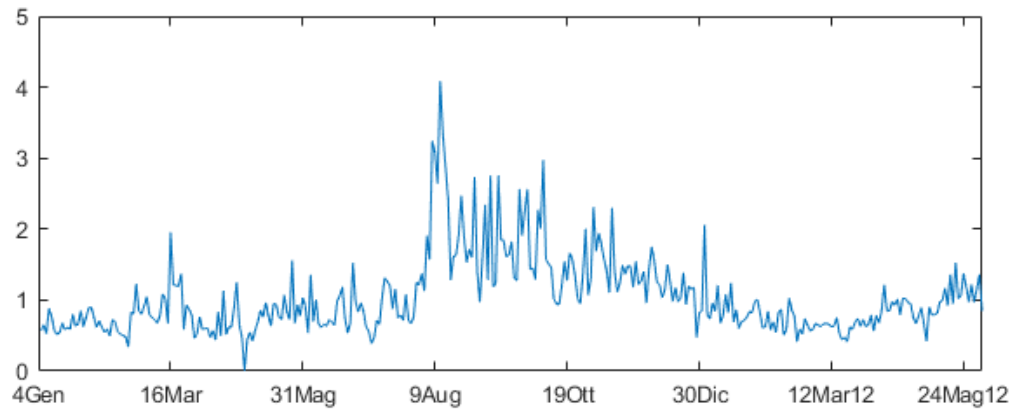
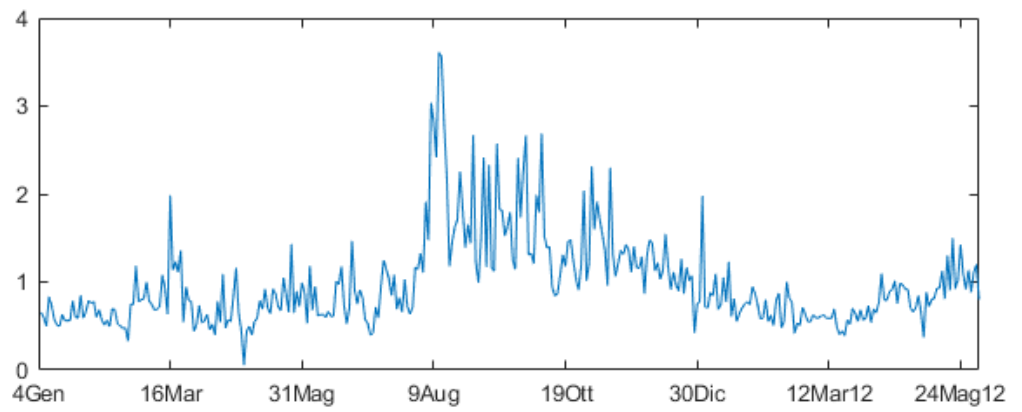
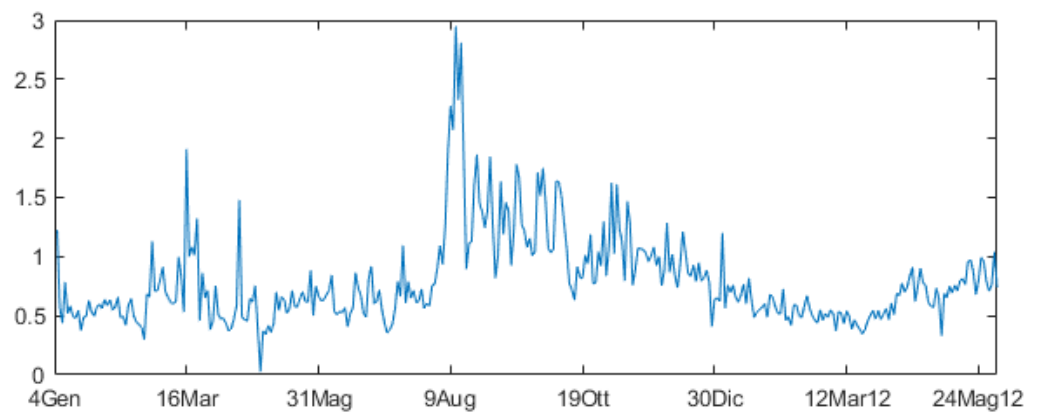
Si mostra in questa sede l'andamento delle Volatilità Realizzate, per ogni frequenza di osservazione disponibile, nell'orizzonte temporale di riferimento del campione. L'aumento della frequenza porta, da una parte a registrare valori della volatilità più contenuti, dall'altra ad evidenziare maggiormente le situazioni più turbolente. Si può notare inoltre che resta chiaramente in evidenza, per tutti i grafici, il Crash del 9 Agosto 2011.

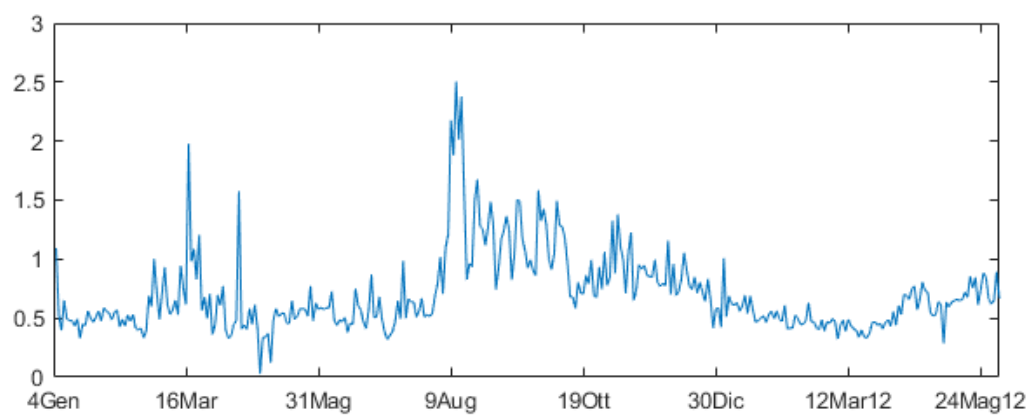
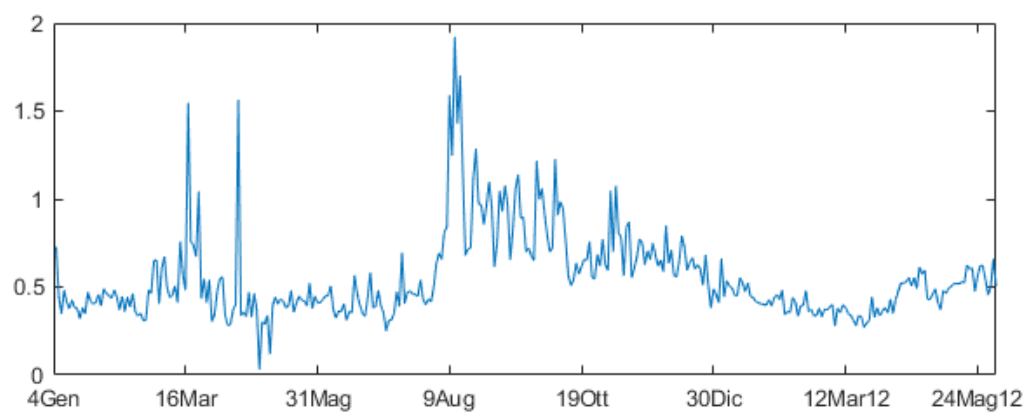
*Figura A.1: Volatilità Realizzata 300s*



*Figura A.2: Volatilità Realizzata 120s*



*Figura A.3: Volatilità Realizzata 60s**Figura A.4: Volatilità Realizzata 30s**Figura A.5: Volatilità Realizzata 10s*

*Figura A.6: Volatilità Realizzata 5s**Figura A.7: Volatilità Realizzata 1s*

# Appendice B

## VPIN

Si inseriscono in questa seconda appendice i risultati specifici relativi alla trattazione del capitolo 3 sull'indicatore di tossicità VPIN.

### B.1 Distribuzione di Burr

Vengono qui mostrate le stime dei parametri della distribuzione di Burr (cioè la densità di probabilità che si è visto approssima al meglio la distribuzione del VPIN in tutti i casi) per tutte le differenti frequenze di rilevazione.

*Tabella B.1: Parametri della distribuzione di Burr*

Frequenza VPIN	$\alpha$	$c$	$\kappa$
300s	0.6398	10.0024	1.3037
120s	0.2451	11.5375	0.4505
60s	0.1652	11.5603	0.5450
30s	0.1231	11.8078	0.5918
10s	0.0898	11.4951	0.7234
5s	0.0741	11.8478	0.6886
1s	0.0497	9.626	0.9583

## B.2 Periodi di criticità e affidabilità del VPIN

Si illustrano in questa sezione i vari periodi di criticità identificati attraverso il VPIN, per tutte le varie frequenze di rilevazione e per i differenti livelli di significatività fissati, nonché le perdite massime registrate in tali periodi. Più giornate consecutive critiche sono indicate con la scrittura: *[gg iniziale – gg finale]*.

Al di sotto di ogni tabella delle giornate critiche, sono indicate le sovrapposizioni perfette: ci si riferisce semplicemente alle giornate di negoziazione specifiche che sono individuate come critiche da tutte le diverse frequenze di rilevazione.

Si noti inoltre che sono specificate a parte le corrispondenze perfette considerando anche i dati a 300 e 120 secondi: questo perchè in realtà si tratta di frequenze al di sopra di quella ottimale indicata dal lavoro di Easley et al. (2012b), per le quali potrebbe non essere ideale il calcolo del VPIN.

Trascurando la necessità di un'esatta corrispondenza tra le specifiche giornate di contrattazione tossiche per poter parlare di una sovrapposizione, è possibile individuare delle sovrapposizioni generiche: si tratta cioè di tutti quegli orizzonti temporali generici in cui è riscontrata elevata tossicità dalla maggior parte dei differenti VPIN. In particolare, si individuano i periodi elencati nella tabella B.2.

Nell'ultima parte della sezione si illustrano invece dettagliatamente, per ogni frequenza a cui è stato calcolato il VPIN, le statistiche relative alle perdite massime a 10, 20 e 30 minuti, in tutte le giornate di negoziazione, e nelle 4 successive, individuate come critiche dall'indicatore. Queste statistiche, co-

struite per ogni differente livello di significatività, vengono poi confrontate con quelle relative alle perdite massime dei periodi non considerati invece critici dallo stesso indicatore. Tutti i risultati iscritti nelle tabelle delle perdite sono da considerare in termini percentuali (per esempio: -0.4 deve essere inteso come -0.4%).

**Tabella B.2:** Sovrapposizioni

0.9	0.95	0.99
[15-22] Marzo 11	[16-18] Marzo 11	-
[26-29] Aprile 11	-	-
[3-9] Maggio 11	-	-
[9-13] Giugno 11	-	-
[21-29] Luglio 11	[22-29] Luglio 11	[25,29] Luglio 11
Agosto 11	[1-9],[15-18] Agosto 11	[1-5] Agosto 11
9, 23 Settembre 11	9 Settembre 11	-
[6-11],[17-21] Ottobre 11	6,11 Ottobre 11	-
[4-10],[16-21] Novembre 11	[4-8] Novembre 11	-
[9-13] Dicembre 11	[9-12] Dicembre 11	-



**Tabella B.3:** Giornate critiche a livello 0.9

Orizzonte temporale	300s	120s	60s	30s	10s	5s	1s
Gen. 11	31	-	-	-	20,31	20	-
Feb. 11	[1-3],25	-	-	-	22,28	22	-
Mar. 11	[16-18]	[16-18]	[15-22]	[15-18]	[15-18]	[15-18]	[15-18]
Apr. 11	-	29	27	[26-29]	[26-29]	15,[26-29]	28,29
Mag. 11	-	3	3	3	3	3,9	6
Giu. 11	9	[9-13]	[10-13]	[10-13]	10	10,[17-20]	10,[17-20]
Lug. 11	1,[22-28]	[21-29]	[21-29]	[22-29]	1,[22-29]	1,[22-29]	1,12,[22-29]
Ago. 11	[1-10], 18,19	[1-10], [15-18], [24-26]	[1-10], [16-18], [24-26]	[1-9], [16-18], [24-26]	[1-9], [16-18]	[1-9], [16-18]	[1-9], [16-18]
Set. 11	5,[9-14], [22-27]	[9-12], 22,23	[9-12], 26,27	[9-12],26	9,[23-26]	7,9,23	23
Ott. 11	[7-11], 20	6,11	[6-11], 20,31	[6-11], 17,31	[6-12], 17,20,21	6,7,11,20,21	[5-7], 11,12, [20-25], [28-31]
Nov. 11	1, 2, [4-11], [16-18]	[4-8]	1,[4-8], 16,17,24	1,2,[4-8], [18-21], 24	2,[4-9], 21,24	[4-10], [18-21], 24	7,8,10, [17-21], 24,29
Dic. 11	2,12,13,20	[9-13],20	[9-12],20	5,[9-12]	[9-12]	[9-12]	6,[9-12]
Gen. 12	23	23	23	-	10,11	-	-
Feb. 12	9	-	-	-	-	-	-
Mar. 12	-	-	-	-	-	-	-
Apr. 12	-	-	-	-	-	4	4,5
Mag. 12	17,21	17	10,16,17,21	-	-	11	16

**Tabella B.4:** Sovrapposizioni perfette a livello 0.9

Sovrapp. perfette	Mar. 11	Giu. 11	Lug. 11	Ago. 11	Ott. 11	Nov. 11	Dic. 11
Tutte le frequenze	[16-18]	-	[22-28]	[1-9], 18	11	7, 8	12
Escludendo 300s e 120s	[15-18]	10	[22-29]	[1-9], [16-18]	6, 7, 11	7, 8, 24	[9-12]

**Tabella B.5:** Giornate critiche a livello 0.95

Orizzonte temporale	300s	120s	60s	30s	10s	5s	1s
Gen. 11	31	-	-	-	-	-	-
Feb. 11	25	-	-	-	-	-	-
Mar. 11	[16-18]	[16-18]	[16-18]	[16-18]	[16-17]	[16-17]	[15-18]
Apr. 11	-	-	-	[26-29]	[28-29]	-	-
Mag. 11	-	-	-	3	-	-	-
Giu. 11	9	10	-	10	-	-	17,20
Lug. 11	1, [22-25]	[22-29]	[25-29]	[25-29]	[25-29]	[25-29]	[22-29]
Ago. 11	[1-8],18	[1-9], [15-18]	[1-9], [16-18], 26	[1-9], [16-17], 24	[1-9], 16	[1-9], [16-17]	[1-9], [17-18]
Set. 11	[23-27]	9	9	9	9,23	9	-
Ott. 11	[7-11]	6	6,11	6,11	6,11,21	6,11,21	6,7,21,25,28
Nov. 11	2,[4-11], 18	7,8	[4-8]	1,7,8,24	[4-8]	7,8,24	-
Dic. 11	12	[9-12]	[9-12]	[9-12]	[9-12]	[9-12]	[9-12]
Gen. 12	-	-	-	-	-	-	-
Feb. 12	-	-	-	-	-	-	-
Mar. 12	-	-	-	-	-	-	-
Apr. 12	-	-	-	-	-	-	-
Mag. 12	-	-	-	-	-	-	-

**Tabella B.6:** Sovrapposizioni perfette a livello 0.95

Frequenze	Mar. 11	Giu. 11	Lug. 11	Ago. 11	Ott. 11	Nov. 11	Dic. 11
Tutte le frequenze	16, 17	-	25	[1-8]	-	-	12
Escludendo 300s e 120s	16, 17	-	[25-29]	[1-9]	6	-	[9-12]

**Tabella B.7:** Giornate critiche a livello 0.99

Orizzonte temporale	300s	120s	60s	30s	10s	5s	1s
Gen. 11	31	-	-	-	-	-	-
Feb. 11	-	-	-	-	-	-	-
Mar. 11	17	-	-	-	-	-	16
Apr. 11	-	-	-	-	-	-	-
Mag. 11	-	-	-	3	-	-	-
Giu. 11	-	-	-	-	-	-	-
Lug. 11	-	[25-27], 29	25,27,29	25,29	[25-27], 29	25,29	25,28,29
Ago. 11	-	[1-5],16	1,4,5	1,2,4,5	1	1	[1-3]
Set. 11	26	-	-	-	-	-	-
Ott. 11	-	-	-	-	-	-	-
Nov. 11	2	-	-	-	-	-	-
Dic. 11	12	-	12	12	[9-12]	-	-
Gen. 12	-	-	-	-	-	-	-
Feb. 12	-	-	-	-	-	-	-
Mar. 12	-	-	-	-	-	-	-
Apr. 12	-	-	-	-	-	-	-
Mag. 12	-	-	-	-	-	-	-

**Tabella B.8:** Sovrapposizioni perfette a livello 0.99

Frequenze	Mar. 11	Giu. 11	Lug. 11	Ago. 11	Ott. 11	Nov. 11	Dic. 11
Tutte le frequenze	-	-	-	-	-	-	-
Escludendo 300s e 120s	-	-	25, 29	1	-	-	-

*Tabella B.9: Perdite massime VPIN a 300s*

Livello di significatività	Statistiche	Periodi Critici			Resto del campione		
		10min	20min	30min	10min	20min	30min
0.9	Media	-0.436	-0.567	-0.656	-0.181	-0.228	-0.265
	1° quartile	-0.54	-0.692	-0.756	-0.303	-0.384	-0.445
	Mediana	-0.352	-0.486	-0.57	-0.187	-0.229	-0.253
	3° quartile	-0.253	-0.327	-0.386	0	0	0
	Perdita massima	-1.8	-2.731	-3.1	-1.103	-0.993	-1.147
0.95	Media	-0.473	-0.622	-0.72	-0.235	-0.298	-0.346
	1° quartile	-0.573	-0.738	-0.861	-0.338	-0.435	-0.514
	Mediana	-0.373	-0.516	-0.605	-0.238	-0.303	-0.349
	3° quartile	-0.284	-0.333	-0.388	-0.102	-0.146	-0.153
	Perdita massima	-1.8	-1.731	-3.1	-1.103	-0.993	-1.249
0.99	Media	-0.4	-0.51	-0.592	-0.32	-0.41	-0.476
	1° quartile	-0.53	-0.632	-0.712	-0.375	-0.511	-0.594
	Mediana	-0.591	-0.475	-0.37	-0.29	-0.356	-0.42
	3° quartile	-0.29	-0.39	-0.389	-0.2	-0.245	-0.282
	Perdita massima	-0.842	-0.913	-1.126	-1.8	-2.731	-3.1

*Tabella B.10: Perdite massime VPIN a 120s*

Livello di significatività	Statistiche	Periodi Critici			Resto del campione		
		10min	20min	30min	10min	20min	30min
0.9	Media	-0.513	-0.644	-0.752	-0.253	-0.315	-0.359
	1° quartile	-0.622	-0.742	-0.791	-0.365	-0.473	-0.5344
	Mediana	-0.433	-0.541	-0.635	-0.247	-0.318	-0.35
	3° quartile	-0.321	-0.395	-0.46	0	0	0
	Perdita massima	-1.773	-2.63	-3.22	-1.151	-1.4	-1.393
0.95	Media	-0.584	-0.752	-0.863	-0.294	-0.364	-0.419
	1° quartile	-0.722	-0.828	-1.1	-0.395	-0.521	-0.6
	Mediana	-0.498	-0.597	-0.669	-0.279	-0.348	-0.404
	3° quartile	-0.353	-0.461	-0.494	-0.184	-0.218	-0.247
	Perdita massima	-1.773	-2.63	-3.214	-1.151	-1.4	-1.393
0.99	Media	-0.787	-1.087	-1.281	-0.347	-0.429	-0.491
	1° quartile	-1.139	-1.504	-1.913	-0.422	-0.554	-0.64
	Mediana	-0.673	-0.818	-1.005	-0.323	-0.387	-0.453
	3° quartile	-0.442	-0.559	-0.666	-0.229	-0.272	-0.318
	Perdita massima	-1.773	-2.63	-3.214	-1.151	-1.4	-1.393

*Tabella B.11: Perdite massime VPIN a 60s*

Livello di significatività	Statistiche	Periodi Critici			Resto del campione		
		10min	20min	30min	10min	20min	30min
0.9	Media	-0.535	-0.662	-0.757	-0.212	-0.264	-0.3
	1° quartile	-0.639	-0.785	-0.871	-0.345	-0.41	-0.478
	Mediana	-0.469	-0.571	0.668	-0.232	-0.278	-0.306
	3° quartile	-0.353	-0.424	-0.482	0	0	0
	Perdita massima	-1.8	-2.731	-3.214	-1.176	-1.399	-1.393
0.95	Media	-0.637	-0.795	-0.899	-0.298	-0.369	-0.421
	1° quartile	-0.802	-0.923	-1.105	-0.399	-0.523	-0.609
	Mediana	-0.556	-0.635	-0.694	-0.296	-0.355	-0.41
	3° quartile	-0.421	-0.492	-0.567	-0.194	-0.23	-0.245
	Perdita massima	-1.8	-2.731	-3.214	-1.176	-1.399	-1.49
0.99	Media	-0.704	-0.968	-1.139	-0.37	-0.453	-0.514
	1° quartile	-0.995	-1.538	-1.728	-0.446	-0.581	-0.66
	Mediana	-0.512	-0.603	-0.694	-0.338	-0.402	-0.468
	3° quartile	-0.399	-0.504	-0.6	-0.242	-0.283	-0.322
	Perdita massima	-1.8	-2.731	-3.214	-1.542	-2.325	-2.085

*Tabella B.12: Perdite massime VPIN a 30s*

Livello di significatività	Statistiche	Periodi Critici			Resto del campione		
		10min	20min	30min	10min	20min	30min
0.9	Media	-0.579	-0.718	-0.81	-0.246	-0.302	-0.346
	1° quartile	-0.696	-0.857	-1.007	-0.366	-0.467	-0.529
	Mediana	-0.508	-0.603	-0.679	-0.261	-0.312	-0.354
	3° quartile	-0.384	-0.468	-0.515	0	0	0
	Perdita massima	-1.883	-2.778	-3.214	-1.176	-1.399	-1.393
0.95	Media	-0.606	-0.756	-0.864	-0.286	-0.35	-0.398
	1° quartile	-0.725	-0.857	-1.026	-0.399	-0.513	-0.593
	Mediana	-0.511	-0.603	-0.689	-0.29	-0.347	-0.387
	3° quartile	-0.388	-0.469	-0.539	-0.17	-0.195	-0.214
	Perdita massima	-1.883	-2.778	-3.214	-1.176	-1.399	-1.393
0.99	Media	-0.737	-0.991	-1.161	-0.379	-0.463	-0.524
	1° quartile	-1.1	-1.591	-1.833	-0.453	-0.595	-0.677
	Mediana	-0.512	-0.603	-0.694	-0.349	-0.411	-0.48
	3° quartile	-0.443	-0.511	-0.603	-0.251	-0.299	-0.331
	Perdita massima	-1.883	-2.778	-3.214	-1.542	-2.34	-2.101

**Tabella B.13:** Perdite massime VPIN a 10s

Livello di significatività	Statistiche	Periodi Critici			Resto del campione		
		10min	20min	30min	10min	20min	30min
0.9	Media	-0.566	-0.788	-0.768	-0.268	-0.313	-0.354
	1° quartile	-0.668	-0.798	-0.904	-0.403	-0.49	-0.561
	Mediana	-0.472	-0.559	-0.661	-0.255	-0.304	-0.331
	3° quartile	-0.346	-0.391	-0.466	0	0	0
	Perdita massima	-1.94	-2.778	-3.255	-2.071	-2.008	-2.449
0.95	Media	-0.658	-0.804	-0.912	-0.338	-0.398	-0.446
	1° quartile	-0.843	-0.946	-1.121	-0.439	-0.546	-0.634
	Mediana	-0.542	-0.629	-0.738	-0.318	-0.37	-0.421
	3° quartile	-0.4	-0.492	-0.556	-0.193	-0.225	-0.238
	Perdita massima	-1.94	-2.778	-3.255	-2.071	-2.008	-2.449
0.99	Media	-0.645	-0.779	-0.859	-0.44	-0.523	-0.589
	1° quartile	-0.73	-0.805	-0.94	-0.524	-0.63	-0.733
	Mediana	-0.517	-0.578	-0.674	-0.376	-0.449	-0.505
	3° quartile	-0.436	-0.483	-0.583	-0.265	-0.313	-0.348
	Perdita massima	-1.94	-2.458	-2.333	-2.071	-2.778	-3.255

**Tabella B.14:** Perdite massime VPIN a 5s

Livello di significatività	Statistiche	Periodi Critici			Resto del campione		
		10min	20min	30min	10min	20min	30min
0.9	Media	-0.621	-0.731	-0.801	-0.282	-0.319	-0.355
	1° quartile	-0.799	-0.907	-1.027	-0.418	-0.495	-0.547
	Mediana	-0.524	-0.588	-0.677	-0.255	-0.29	-0.324
	3° quartile	-0.362	-0.419	-0.487	0	0	0
	Perdita massima	-2.929	-3.461	-3.283	-2.452	-2.758	-2.704
0.95	Media	-0.722	-0.857	-0.938	-0.378	-0.432	-0.479
	1° quartile	-0.9	-0.984	-1.126	-0.482	-0.588	-0.657
	Mediana	-0.579	-0.631	-0.733	-0.332	-0.378	-0.43
	3° quartile	-0.431	-0.495	-0.567	-0.2	-0.232	-0.25
	Perdita massima	-2.929	-3.461	-3.283	-2.452	-2.758	-2.704
0.99	Media	-0.931	-1.096	-1.055	-0.492	-0.567	-0.63
	1° quartile	-1.148	-1.57	-1.246	-0.588	-0.655	-0.755
	Mediana	-0.714	-0.643	-0.709	-0.4	-0.49	-0.528
	3° quartile	-0.449	-0.586	-0.621	-0.279	-0.329	-0.36
	Perdita massima	-2.929	-3.461	-2.802	-2.452	-2.806	-3.283

*Tabella B.15: Perdite massime VPIN a 1s*

Livello di significatività	Statistiche	Periodi Critici			Resto del campione		
		10min	20min	30min	10min	20min	30min
0.9	Media	-0.683	-0.778	-0.85	-0.286	-0.327	-0.362
	1° quartile	-0.836	-0.921	-1.043	-0.413	-0.5	-0.538
	Mediana	-0.548	-0.626	721	-0.256	-0.296	-0.324
	3° quartile	-0.397	-0.483	-0.518	0	0	0
	Perdita massima	-3.188	-3.72	-3.289	-2.452	-2.758	-2.586
0.95	Media	0.792	-0.917	-0.975	-0.406	-0.461	-0.513
	1° quartile	-0.975	-1.077	-1.063	-0.504	-0.6	-0.695
	Mediana	-0.715	-0.662	-0.746	-0.349	-0.397	-0.464
	3° quartile	-0.453	-0.513	-0.58	-0.212	-0.253	-0.271
	Perdita massima	-3.188	-3.72	-3.289	-2.452	-2.758	-2.586
0.99	Media	-0.927	-1.115	-1.133	-0.492	0.559	-0.62
	1° quartile	-1.237	-1.53	-1.541	-0.591	-0.664	-0.756
	Mediana	-0.65	-0.636	-0.726	-0.399	-0.487	-0.521
	3° quartile	-0.431	-0.485	-0.475	-0.273	-0.322	-0.355
	Perdita massima	-3.188	-3.72	-3.289	-2.793	-2.758	-3.143

# Bibliografia

Abad, David and José Yagüe (2012) “From PIN to VPIN: An introduction to order flow toxicity,” *The Spanish Review of Financial Economics*, Vol. 10, No. 2, pp. 74–83.

Alagia, Ennio (2015) “Market abuse a trading ad alta frequenza,”  
URL: <http://www.giurisprudenzapenale.com/wp-content/uploads/2017/10/Market-Abuse-e-Trading-ad-alta-frequenza.pdf>.

Baillie, Richard T, Tim Bollerslev, and Hans Ole Mikkelsen (1996) “Fractionally integrated generalized autoregressive conditional heteroskedasticity,” *Journal of econometrics*, Vol. 74, No. 1, pp. 3–30.

Barndorff-Nielsen, Ole E and Neil Shephard (2002) “Econometric analysis of realized volatility and its use in estimating stochastic volatility models,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, Vol. 64, No. 2, pp. 253–280.

Bector, Raj, Anthony Marrato, and Chris Sparrow (2013) “The hidden alpha in equity trading: steps to increasing returns with the advanced use of information,” *Oliver Wyman*.



- Black, Fischer and Myron Scholes (1973) “The pricing of options and corporate liabilities,” *Journal of political economy*, Vol. 81, No. 3, pp. 637–654.
- Bollerslev, Tim (1986) “Generalized autoregressive conditional heteroskedasticity,” *Journal of econometrics*, Vol. 31, No. 3, pp. 307–327.
- Breckenfelder, Johannes (2019) “Competition among high-frequency traders, and market quality.”
- Caivano, Valeria (2015) “The impact of high-frequency trading on volatility,” *CONSOB Working Papers No.80*.
- Caivano, Valeria, Salvatore Ciccarelli, Giovanna Di Stefano, Marco Fratini, Giorgio Gasparri, Monica Giliberti, Nadia Linciano, and Isadora Tarola (2012) “Il trading ad alta frequenza-Caratteristiche, effetti, questioni di policy (High Frequency Trading-Definition, Effects, Policy Issues),” *CONSOB Discussion Papers*, No. 5.
- Carlini, Vittorio (2018) “Il lato oscuro dei listini: così gli algoritmi manipolano i mercati,” URL: <https://www.ilsole24ore.com/art/il-lato-oscuro-listini-cosi-algoritmi-manipolano-mercati-AE1qK3eE>.
- (2019a) “Borsa, il robot sostituisce l’uomo: ai software il 53% degli scambi,” URL: <https://www.ilsole24ore.com/art/borsa-robot-padrone-listini-decide-53percento-scambi--ACG4CdD>.

——— (2019b) “Franco svizzero, l’umano va in vacanza e il robot-trader fa «crash»,” URL: <https://www.ilsole24ore.com/art/franco-svizzero-l-umano-va-vacanza-e-robot-trader-fa-crash-AB504eSB>.

CONSOB (2019) “Relazione per l’anno 2018.”

Ding, Zhuanxin, Clive WJ Granger, and Robert F Engle (1993) “A long memory property of stock market returns and a new model,” *Journal of empirical finance*, Vol. 1, No. 1, pp. 83–106.

Easley, David, Marcos M Lopez De Prado, and Maureen O’Hara (2011a) “The microstructure of the “flash crash”: flow toxicity, liquidity crashes, and the probability of informed trading,” *The Journal of Portfolio Management*, Vol. 37, No. 2, pp. 118–128.

Easley, David, Marcos M López De Prado, and Maureen O’Hara (2012a) “The volume clock: Insights into the high-frequency paradigm,” *The Journal of Portfolio Management*, Vol. 39, No. 1, pp. 19–29.

Easley, David, Robert F Engle, Maureen O’Hara, and Liuren Wu (2008) “Time-varying arrival rates of informed and uninformed trades,” *Journal of Financial Econometrics*, Vol. 6, No. 2, pp. 171–207.

Easley, David, Nicholas M Kiefer, Maureen O’Hara, and Joseph B Paperman (1996) “Liquidity, information, and infrequently traded stocks,” *The Journal of Finance*, Vol. 51, No. 4, pp. 1405–1436.

- Easley, David, Marcos M López de Prado, and Maureen O'Hara (2011b) "The VPIN Flow Toxicity metric and liquidity crashes," URL: <https://www.youtube.com/watch?v=0vPMZXGHMpc>.
- (2012b) "Flow toxicity and liquidity in a high-frequency world," *The Review of Financial Studies*, Vol. 25, No. 5, pp. 1457–1493.
- Engle, Robert F and Tim Bollerslev (1986) "Modelling the persistence of conditional variances," *Econometric reviews*, Vol. 5, No. 1, pp. 1–50.
- Golub, Anton, John Keane, and Ser-Huang Poon (2012) "High frequency trading and mini flash crashes," *Available at SSRN 2182097*.
- Hamilton, James D. (1995) *Econometria delle serie storiche*: Monduzzi editore.
- Johnson, Neil, Guannan Zhao, Eric Hunsader, Jing Meng, Amith Ravindar, Spencer Carran, and Brian Tivnan (2012) "Financial black swans driven by ultrafast machine ecology."
- Kenton, Will (2019) "Flash Crash," URL: <https://www.investopedia.com/terms/f/flash-crash.asp>.
- Levine, Matt (2015) "Guy Trading at Home Caused the Flash Crash," URL: <https://www.bloomberg.com/opinion/articles/2015-04-21/guy-trading-at-home-caused-the-flash-crash>.
- Low, Rand Kwong Yew, Te Li, and Terry Marsh (2018) "BV–VPIN: Measuring the Impact of Order Flow Toxicity and Liquidity on International Equity Markets."

- Lucchetti, Riccardo (2015) “Appunti di analisi delle serie storiche.”
- McInish, Tom, James Upson, and Robert A Wood (2013) “The flash crash: trading aggressiveness, liquidity supply, and the impact of intermarket sweep orders,” *Financial Review*.
- Merton, Robert C (1976) “Option pricing when underlying stock returns are discontinuous,” *Journal of financial economics*, Vol. 3, No. 1-2, pp. 125–144.
- Nanex (2011) “Definition of a Mini-Flash-Crash,” URL: <http://www.nanex.net/Research/MiniFlashCrashDef/MiniFlashCrashDef.html>.
- (2012) “Flash Crash ISO Trades,” URL: <http://www.nanex.net/aqck2/3545.html>.
- Nigro, Casimiro Antonio and Giovanni Romano (2013) “La regolamentazione degli hedge funds negli USA: dal quasi-collasso di Long-Term Capital Management al Dodd-Frank act,” *V. SANTORO (a cura di), Crisi dei mercati finanziari: analisi e prospettive*, p. 187.
- Palomba, Giulio (2004) “Modelli ARCH.”
- Peter, Gember, Arndt Björn, Lutat Marco, and Uhle Tim (2011) “High-frequency trading,” *Available at SSRN*, Vol. 1858626.
- Puorro, Alfonso (2013a) *High frequency trading: una panoramica*: Banca d’Italia.

——— (2013b) “High Frequency Trading: Una Panoramica (High Frequency Trading: An Overview),” *Bank of Italy Occasional Paper*, No. 198.

SEC and CFTC (2010) “Findings regarding the market events of May 6, 2010,” *Report of the Staffs of the CFTC and SEC to the Joint Advisory Committee on Emerging Regulatory Issues*.

Treccani, Enciclopedia (1998) *Enciclopedia Treccani*.

Woodward, Megan (2017) “The Need for Speed: Regulatory Approaches to High Frequency Trading in the United States and the European Union,” *Vand. J. Transnat’l L.*, Vol. 50, p. 1359.