

LIBERA UNIVERSITÀ INTERNAZIONALE DEGLI STUDI SOCIALI

“LUISS - GUIDO CARLI”



FACOLTÀ DI ECONOMIA

Corso di laurea in Banca ed Intermediari Finanziari

**SOCIAL SENTIMENT ANALYSIS
E INVESTIMENTI AZIONARI**

RELATORE:

CHIARISSIMO PROF. EMILIO BARONE.

LAUREANDO:

NATALE DI TOMMASO

CORRELATORE:

CHIARISSIMO PROF. GAETANO CASERTANO

MATRICOLA:

665411

ANNO ACCADEMICO 2015-16

Indice

Capitolo 1...3

Introduzione alla Sentiment Analysis ed allo studio dei Big Data...3

- 1.1 Definizione Sentiment analysis...3
- 1.2 Big Data...4
 - Definizione ed utilizzo dei Big Data...4*
- 1.3 Utilizzo dei Social network e principali direzioni di ricerca...6
 - Numeri e diffusione dei social network...6*
 - Principale aree di utilizzo dei "social network data"...9*
- 1.4 Enunciazione tesi...13

Capitolo 2...15

Letteratura...15

- 2.1 Mutamento della sentiment analysis negli anni...15
 - Dai sondaggi telefonici agli Hashtag Trends...15*
- 2.2 Approccio Individuale...20
 - Presentazione Metodologia Sprenger- Welpé...20*
- 2.3 Approccio Indice...40
 - Presentazione Metodologia Bollen et al....40*
- 2.4 Critica...55
 - Principali errori e critiche e limitazioni dei modelli presentati...55*

Capitolo 3...63

Metodologia...63

- 3.1 Evoluzione della disciplina...63
 - Dalla accuratezza al rendimento atteso...63*
- 3.2 Bullishness Model...65
 - Presentazione del modello...65*
 - Twitter Bullishness e Google Bullishness...68*
 - Risultati...72*
 - Punti di forza e debolezza...84*

Capitolo 4...87

Applicazione e Modellizzazione...87

- 4.1 Applicazione del modello...87
 - Database tweets...87*
 - Twitter Bullishness Index...94*
 - Google Bullishness Index...98*
- 4.2 Strutturazione strategia...101
 - Utilizzo Bullishness Index e modellizzazione strategia di investimento...101*
- 4.3 Analisi Performance Investimento...107
- 4.4 Conclusioni finali...114

Riferimenti Bibliografici...118

Siti Web...122

Capitolo 1

Introduzione alla Sentiment Analysis ed allo studio dei Big Data

1.1 DEFINIZIONE SENTIMENT ANALYSIS

Lo studio della *sentiment analysis* ha ricoperto un ruolo sempre più importante negli ultimi 50 anni. Con il termine “sentiment analysis” è indicato lo studio della percezione o meglio l’opinione o l’umore della collettività in un determinato momento rispetto ad un determinato fenomeno di interesse.

L’analisi del *sentiment* è associata oggi praticamente a qualsiasi analisi di marketing o sondaggio a fini commerciali o qualitativi ma ha anche trovato larga applicazione nelle elezioni politiche in fase di votazione. L’applicazione di questa metodologia ai mercati finanziari, sia a livello di informazioni macro-economiche che come vedremo ad uso puramente speculativo e predittivo, sta muovendo però solo recentemente i primi passi. Come è facilmente intuibile considerato il vasto campo di utilizzo esistono svariate tecniche di *sentiment analysis* o *opinion mining*, nonostante questo però le modalità di svolgimento dell’analisi restano sostanzialmente simili in ogni sua sfaccettatura.

Gli approcci esistenti all’analisi del *sentiment* possono essere raggruppati in 4 categorie principali:

- *Spotting* di parole chiave
- Affinità lessicale
- Metodi statistici
- Tecniche di livello concettuale

Il primo approccio classifica il testo in categorie influenti basandosi sulla presenza di parole chiave come per esempio: contento, triste, impaurito e annoiato. L'approccio ad 'affinità lessicale invece non rileva solamente le parole influenzanti, ma inoltre assegna arbitrariamente alle parole una probabile affinità a emozioni particolari. I metodi statistici fanno leva invece su elementi tratti dal *Machine Learning* come analisi semantica latente, macchine a vettori di supporto, *bag of words* e orientazione semantica. Per estrapolare l'opinione in un contesto e ottenerne delle caratteristiche, sono usate le relazioni grammaticali delle parole. Le relazioni sono ottenute da un'analisi sintattica profonda del testo. A differenza delle tecniche puramente sintattiche, gli approcci a livello concettuale fanno leva sugli elementi della rappresentazione della conoscenza come le ontologie e le reti semantiche, e quindi, sono capaci di rilevare semantiche che sono espresse in maniera sottile.

Nei capitoli successivi verrà illustrata l'evoluzione di questa metodologia riportando dettagliatamente quelli che sono stati i modelli più apprezzati dal mondo accademico e professionale, ma prima ritengo necessario uno sguardo a quello che è il più ampio campo di utilizzo dei BIG DATA e soprattutto un'infarinatura di quelli che sono i maggiori campi di applicazione dei dati derivanti dai social network.

1.2 BIG DATA

Definizione ed utilizzo dei Big Data

La parola "Big Data" nella sua traduzione in italiano "Grandi Dati" non lascia molto all'immaginazione, ben diverso è capacitarsi dell'effettiva grandezza della mole di questi dati e del loro vastissimo campo di applicazione.

Non esiste una soglia di grandezza per stabilire la differenza fra "simple data" e "BIG DATA" ma soltanto un limite metodologico. BIG DATA è il termine usato per descrivere una raccolta di dati così estesa in termini di volume, velocità e varietà da richiedere tecnologie e metodi analitici specifici per l'estrazione di valore. Il progressivo aumento della dimensione dei *dataset* è legato alla necessità di analisi su un unico insieme di dati, con

L'obiettivo di estrarre informazioni aggiuntive rispetto a quelle che si potrebbero ottenere analizzando piccole serie con la stessa quantità totale di dati. Ad esempio, l'analisi per sondare gli "umori" dei mercati e del commercio, e quindi del trend complessivo della società e del fiume di informazioni che viaggiano e transitano attraverso Internet.

Il termine Big Data rappresenta anche l'interrelazione di dati provenienti potenzialmente da fonti eterogenee, quindi non soltanto i dati strutturati, come i database, ma anche non strutturati, come immagini, email, dati GPS, informazioni prese dai social network.

Secondo uno studio del 2001, l'analista Doug Laney aveva definito il modello di crescita come tridimensionale (modello delle "3V"): con il passare del tempo aumentano volume (dei dati), velocità e varietà (dei dati).

Il modello iniziale di Laney dava ai Big Data quindi le seguenti caratteristiche:

- volume: rappresenta la dimensione effettiva del dataset; l'ampio volume di dati che è possibile raccogliere oggi potrebbe apparentemente rappresentare un problema. In realtà quello del volume dei Big Data è un falso problema, in quanto *cloud* e virtualizzazione aiutano nella gestione del grosso volume di dati disponibili, semplificando i processi di raccolta, immagazzinamento e accesso ai dati.
- velocità: si riferisce alla velocità di generazione dei dati; si tende all'effettuare analisi dei dati in tempo reale o quasi;
- varietà: riferita alle varie tipologie di dati, provenienti da fonti diverse (strutturate e non).

Ovviamente il progresso tecnologico che si è osservato negli ultimi 15 anni ha stravolto quello che era il *set* informativo osservabile nel 2001 rendendo queste sole tre caratteristiche troppo restrittive per tale definizione. Oggi quando si parla di Big Data si è profondamente interessati anche a caratteristiche quali la variabilità ed alla complessità dei dati da dover modellizzare.

Infine per dare una un'idea della dimensione fisica "reale" dei Big Data osservabili oggi, essa è misurabile in diverse centinaia exabyte, una dimensione di memoria che corrisponde a circa a un milione di terabyte che a sua volta può essere tradotta in una capacità di memoria tale da immagazzinare circa 1400 cd musicali o 212 DVD di medie dimensioni. Per dare termine di paragone basta pensare che l'intera Biblioteca del Congresso di Washington comprende "solamente" 441 terabyte di dati e il colosso mondiale

del *colud Megaupload* chiuso nel 2012 conteneva comprensivamente 28 petabyte di dati. La mole odierna è una misura colossale, che graficamente è anche di difficile rappresentazione sia come confronto statico sia come evoluzione nel tempo.

Riporto di seguito due tabelle a scopo puramente illustrativo, nella tabella di sinistra 1.1 è raccolta l'evoluzione del volume dei dati disponibili in rete nel tempo, quella di destra la 1.2 una semplice tabella di conversione delle varie capacità di memoria.

Tabella 1.1

Anno	Quantità
1986	281 PetaByte
1993	471 PetaByte
2000	2.2 ExaByte
2007	65 ExaByte
2014	650 ExaByte
2016	1.04 ZettaByte

Tabella 1.2

Unità	Dimensione in Byte	In Gigabyte
KiloByte	10^3	-
MegaByte	10^6	-
GigaByte	10^9	1
TeraByte	10^{12}	1.000
PetaByte	10^{15}	1.000.000
ExaByte	10^{18}	1.000.000.000
ZettaByte	10^{21}	1.000.000.000.000

Secondo le stime aggiornate a giugno 2016 del colosso del settore tecnologico CISCO il volume dei dati disponibili nel 2020 ammonterà a 2.3 zettabyte, più che raddoppiando quindi il volume attuale.

1.3 UTILIZZO DEI SOCIAL NETWORK E PRINCIPALI DIREZIONI DI RICERCA

Numeri e diffusione dei social network

Dopo aver dato una breve sguardo ai *Big Data* ritengo necessario focalizzarsi su quello che sarà il soggetto principale della analisi di questo testo; i social network.

Una prima differenziazione risulta necessaria per comprendere perché ci si focalizzerà sui social network e non sui social media in generale.

I social media sono piattaforme che permettono di generare, rendere pubblici e condividere contenuti, che a loro volta sono stati creati da altri utenti. E' di facile comprensione quindi che i social network sono solo un sottoinsieme dei social media che posseggono

la caratteristica di essere una “rete sociale”. L’interconnessione fra gli utenti è una caratteristica necessaria ma non sufficiente affinché ci sia “network”, infatti rivestono un ruolo di pari importanza la possibilità della piena identificazione dei soggetti e soprattutto la possibilità di una comunicazione interattiva fra gli stessi.

Una volta data questa semplice regola distintiva, è più chiaro capire perché blog di portata globale o anche riferimenti mondiali come Wikipedia possano essere classificati solamente come “progetti collaborativi” e non come *social network*.

All’interno di questo sottoinsieme esiste un mondo in continua evoluzione, esistono in pratica social specifici per ogni passione, interesse o anche perversione. Per quanto riguarda però l’effettiva possibilità di utilizzare dati di un social network come *proxy* di quella che è l’effettiva compagine globale la scelta ricade solamente su tre grandi nomi: Facebook, Twitter e Google+.

Facebook è senza ombra di dubbio il più conosciuto ma anche il più “vecchio”, fondato nel 2004 per i primi anni è stato accessibile solamente ad una cerchia ristretta di studenti universitari americani, dopo ha avuto una crescita esponenziale che l’ha portato a registrare nel 2010 circa 100 milioni di utilizzatori e nel 2012 la strabiliante cifra di 1 miliardo di utenti attivi mensilmente.

Twitter nasce successivamente, nel 2006 e rispetto a Facebook è sempre stato caratterizzato da qualche restrizione in più ed una diversa filosofia di network che lo rende come vedremo in seguito molto più adatto per essere usato come fonte di *data mining*.

Infine Google+ è il network più recente, lanciato da Google nel 2011 e fortemente sponsorizzato dall’azienda di Mountain View tanto dal renderlo inscindibile dalla casella email. Google + si trova a metà strada fra la libertà di condivisione di Facebook e la strutturazione di hashtag e citazioni di Twitter.

Concludo questa introduzione con uno sguardo sui “numeri” della diffusione dei social media rapportandoli a quella che è la diffusione di internet in Italia e nel resto del mondo, evidenziando l’ormai impossibilità di poter scindere i due elementi.

A luglio 2013 nel mondo risultavano esserci circa 2,4 miliardi di utenti internet pari al 34,3% della popolazione mondiale in tale anno.

La tabella 1.3 sotto riporta la diffusione di internet nel mondo aggiornata al 2013, le stime per l’Italia nello stesso anno erano in linea con il resto dell’Europa ovvero con

circa 37.8 milioni di utenti a fronte di una popolazione di 60.2 milioni di persone con un tasso di penetrazione de 62% circa.

Scendendo nei dettagli si stima un tasso di penetrazione del 80.2% nella popolazione nella fascia di età fra gli 11 ed i 74 anni con circa 22 milioni di italiani che accedono tramite dispositivi mobili con una media giornaliera di 14 milioni di accessi e di circa 28.5 milioni di utenti italiani che accedono ad internet almeno una volta al mese

Tabella 1.3

Regioni del mondo	Popolazione (Stima 2012)	Utenti internet (ultimo agg. a lug. 2013)	Tasso penetrazione (% Popolazione)
Africa	1.073.380.925	167.335.676	15,6%
Asia	3.922.066.987	1.076.681.059	27,5%
Europa	820.918.446	518.512.109	63,2%
Medio Oriente	223.608.203	90.000.455	40,2%
America del Nord	348.280.154	273.785.413	78,6%
America Latina / Caraibi	593.688.638	254.915.745	42,9%
Oceania / Australia	35.903.569	24.287.919	67,6%
Totale nel mondo	7.017.846.922	2.405.518.376	34,3%

Passando ad analizzare i social network, in una analisi svolta nel 2013 Facebook risulta essere la piattaforma più utilizzata in termini di utenti complessivi con oltre un miliardo di utenti a livello globale ed un tasso di penetrazione nella fascia 16-65 anni del 51% seguita da Google+ con 500 milioni di utenti iscritti, 359 attivi su base mensile ed un tasso di penetrazione del 26% e da Twitter con 550 milioni di iscritti, 297 milioni attivi su base mensile ed un tasso di penetrazione del 22%.

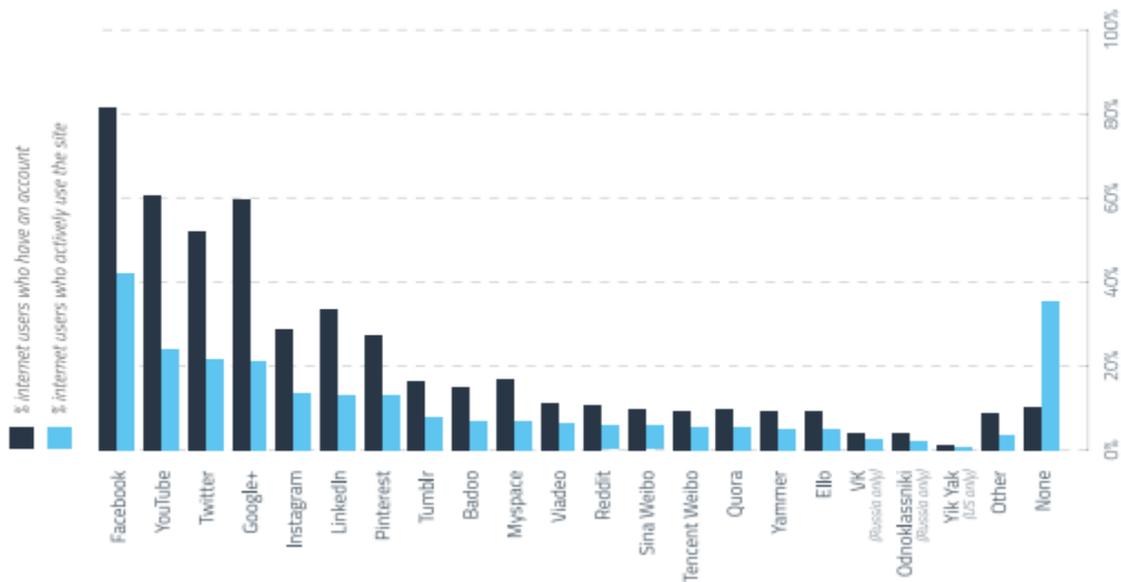
Il grafico 1.1 invece riporta le stime al 2015, come è possibile notare il tasso di penetrazione è notevolmente aumentato grazie anche all' avanzamento tecnologico ed alla diffusione degli smartphone portando Facebook, Twitter e Google+ ad un tasso di penetrazione rispettivamente superiore all' 80% ,50% e 60%.

Degno di nota però è il tasso di crescita degli utenti attivi che vedono al primo posto Twitter con un incremento annuo di circa il 35% medio annuo grazie ad un ingresso sempre più massiccio (79%) di fette di popolazione più mature che utilizzano il social

network su base quotidiana per il commento e la condivisione delle notizie della giornata, caratteristica distintiva di Twitter che rende il social network molto più proiettato verso l'informazione sociale che quella personale rendendolo particolarmente adatto alla rappresentazione di quello che può essere chiamato umore, percezione, opinione, parere o *sentiment* generale.

Il grafico 1.1 confrontato con i dati relativi al 2013 riportati precedentemente e letta con la consapevolezza che il tempo medio globale trascorso nel 2015 sui social network è stato di 47 minuti al giorno da una visione più chiara di quanto ormai sia indivisibile la parte di vita “virtuale” da quella reale e soprattutto quanto nel corso degli ultimi anni questa parte “virtuale” è cresciuta fino a radicarsi nella nostra società tanto da risultarne ormai un tutt'uno.

Grafico 1.1



Principale aree di utilizzo dei “social network data”.

Grazie alla loro crescente diffusione i social media sono stati oggetto di forte interesse sia in ambito accademico che commerciale e giornalistico.

Per quello che riguarda la ricerca scientifica la diversità delle aree di interesse che hanno prodotto studi basati sui social media è piuttosto marcata, si va dall' economia al marketing, dalla psicologia all'epidemiologia e dalla sociologia alle scienze politiche. Così come fatto precedentemente, nel presentare questi canali di diffusione è necessario fare una prima differenziazione in due grandi categorie di analisi: l'approccio classico e quello *bottom up*.

Quando si parla di approccio classico lo scopo della ricerca è quello di capire se la comunicazione sui social network è in grado di influenzare i comportamenti e le scelte dei propri utenti.

Questo tipo di analisi è fortemente spronato dalla evidenza empirica che sembra confermare l'effettività di tale processo, gli utenti dei social media tendono a recepire e diffondere in maniera straordinariamente confidente le notizie sui social media a differenza di quelle recepite attraverso i canali ufficiali. Esistono moltissimi esempi a conferma di tale tesi, si basti pensare a tutti i *rumors* sui prodotti più attesi, calciomercato e su gossip delle star che ormai provengono quasi esclusivamente dai profili Twitter o Facebook dei diretti interessati. Vale la pena ricordare un famoso caso domestico di "diffusione endogena" legato a *news* finanziarie che hanno avuto corso prima sui social network che sui canali ufficiali. Il 4 marzo del 2013 il giornalista Gad Lerner annunciava a mercati ancora aperti la cessione della rete televisiva La7 all'imprenditore Urbano Cairo con due ore di anticipo rispetto alla società dove in quel momento era ancora convocato il Consiglio di Amministrazione; nel momento in cui la notizia ufficiale fu data non ci fu in pratica nessuno *shock* nei corsi in quanto l'impatto della notizia era stato già assorbito.

Esistono innumerevoli esempi oltre a quelli delle *news* di quello che è definibile "passaparola generale"; in particolare gli utenti esprimono liberamente sui social network giudizi in relazione ad un determinato prodotto o una determinata marca. Questo aspetto svolge un ruolo cruciale nel valutare la *brand reputation* di un marchio o la *consumer satisfaction* in relazione ad un prodotto o servizio, soprattutto se avviene in concomitanza con la commercializzazione o la presentazione.

Questa possibilità di poter incidere sulla scelta degli utenti finali è stata sfruttata dalle aziende creando quello che chiamano "top-down". Le aziende condividono e pubblicizzano i propri prodotti sui social network in modo da innescare una reazione a catena di passaparola, garantendo al contempo la possibilità di interfacciarsi con gli utenti finali e

di rispondere alle loro domande; tutti fattori che permettono di consolidare quella che è la *brand reputation* e la *consumer satisfaction* in relazione ad un determinato prodotto. L'altro grande approccio è quello definito "bottom up", forse quello più interessante sul quale verterà principalmente la mia tesi. Ovvero i social media come moderna agorà in modo appropriato in modo da fornirci un aiuto prezioso per comprendere l'evolversi di fenomeni sociali complessi. I *social network* possono essere visti come una fonte di dati, di Big Data, che ci permettono, una volta analizzati con cura, di comprendere l'opinione di chi scrive e il cambiamento di essa rispetto a qualche avvenimento esterno. Questo è ciò che in letteratura viene chiamato *nowcasting* ovvero la possibilità di produrre previsioni sul presente identificando dinamiche di stadi che si relazionano in tempo reale. In questo filone di ricerca rientrano i sempre più numerosi studi che utilizzano i social media per far delle vere e proprie previsioni sfruttando ed aggregando la potenziale eterogeneità dei Big Data da essi derivanti. Su questa scia sono stati sviluppati ambiziosi progetti come l'*Open Source Indicators* ed il *Record Future* che hanno come obiettivo quello di sintetizzare indici che in tempo reale diano un'idea degli argomenti più trattati dall'intera popolazione di internet.

Passiamo ora invece alle effettive aree di applicazione dell'analisi scientifica che hanno utilizzato i social media per effettuare previsioni e soprattutto *nowcasting*.

La prima grande macro area è quella economica, non mi dilungherò molto in questa introduzione poiché gli interi prossimi capitoli parleranno approfonditamente di questo, però ritengo in ogni caso necessario uno sguardo su quelle che sono le sue sfaccettature più importanti.

L'analisi dei *social media* in questo caso è utilizzata sia per tematiche a livello micro quali possono essere quelle citate da Antweiler e Frank nel 2004 che dimostrano (oltre ad altri interessanti risultati) come il volume dei commenti pubblicati sui forum specializzati in questioni finanziarie sia un buon indicatore della volatilità dei mercati che a livello macro. Altri autori come vedremo nei prossimi capitoli hanno applicato questa analisi per la previsione dei movimenti futuri di indici azionari come il *Dow Jones Industrial Average* e il FTSE100 o anche il prezzo dell'oro ed il tasso di cambio per diverse valute.

A livello macro troviamo invece una curiosa applicazione creata dalla Banca d'Inghilterra utilizzando *Google Trend*, quindi i volumi delle stringhe di ricerca inserite su Google, per misurare tre diversi aspetti:

- Analizzare il mercato del lavoro osservando indicatori macroeconomici come il tasso di disoccupazione.
- Monitorare l'andamento del mercato immobiliare, osservando profili e dati di agenti immobiliari riuscendo a prevedere il volume di vendite.
- Misurare l'indice di fiducia dei consumatori, osservando una correlazione con le ricerche relative all'imposta sul valore aggiunto.

Passando all'area marketing, il più grande successo di questa categoria è senz'altro quello ottenuto dall'analisi di Asur e Huberman del 2011 nel quale gli autori dimostrano che è possibile prevedere in maniera fedele gli incassi al botteghino dei film in uscita basandosi solamente sul *sentiment* dei commenti postati sulle anteprime dei film stessi pubblicate nei giorni precedenti. La stessa metodologia è utilizzata a fini predittivi per produrre stime su eventi non prettamente legati agli utenti stessi (notte degli oscar, festival di San Remo, ecc.); questa forma di *nowcasting* è ritenuta affidabile soprattutto grazie alla diffusione dell'interattività fra la televisione e i social media che ha raggiunto il 38% in Europa ed il 52% in America del Sud.

Altro campo di applicazione del *sentiment* è nell'ambito delle scommesse sportive.

Uz Zaman et al. hanno per esempio tentato di identificare i risultati dei mondiali 2010 tramite Twitter, i risultati hanno previsto in numerosi casi previsioni azzeccate, e soprattutto mostrano che il *sentiment* generale era in linea con quello degli scommettitori.

Nel campo politico non sono mancati studi volti a prevedere colpi di stato o disordini, un famoso studio per esempio è stato fatto da Radinsky e Horvitz o Kalev. Gli autori mostrano che utilizzando un mix di Big Data sarebbe stato possibile prevedere le rivolte avvenute in Tunisia, Libia ed Egitto, la sostanziale stabilità dell'Arabia Saudita fino ad arrivare ad identificare con un raggio di 200km il nascondiglio di Osama Bin Laden.

Infine un'ultima area di applicazione dei *social network* è legata alle scienze mediche e naturali, in particolare dall'epidemiologia e la psicologia fino alla sismologia.

A partire dal 2009 Google ha iniziato a monitorare la ricerca delle parole chiave riguardanti particolari sintomi influenzali per determinare il tasso di diffusione delle malattie e

lo scoppio di epidemie. I volumi di ricerca infatti risultano positivamente correlati con la frequenza e la rilevazione clinica dei sintomi influenzali specifici. L'utilità dell'analisi di Twitter (Paul e Drezne ,2011) per le scienze mediche inoltre è stata confermata anche da lavori che hanno permesso non solo di individuare e geo-localizzare i sintomi delle malattie, ma anche di valutare i fattori di rischio in base ai comportamenti umani e misurare quali farmaci e terapie sono stati impiegate per curare le diverse malattie. Infine l'analisi di Twitter è applicata nella sismologia tutt'oggi per determinare l'epicentro dei fenomeni sismici in aree urbane in maniera molto più accurata rispetto ad altre tecniche, in zone particolarmente soggette a fenomeni sismici come il Giappone i *social media* sono costantemente monitorati dai governi in quanto si sono dimostrati essere più rapidi nell'identificare sismi di lieve entità rispetto alle classiche strutture di prevenzione.

1.4 ENUNCIAZIONE TESI

Dopo aver dato un'introduzione ai social media ed ai diversi utilizzi dei dati da loro estraibili è arrivato il momento di introdurre quella che sarà l'argomento dei prossimi capitoli: stima ed utilizzo della *sentiment analysis* nel settore dei mercati finanziari.

Quando parliamo di *sentiment analysis* nei mercati finanziari ci riferiamo ad una stima di quella che è la tendenza di investimento nel brevissimo termine della popolazione di investitori; in pratica così come è stato fatto negli anni passati in via prima cartacea e poi digitale, si stima quella che è l'intenzione della "collettiva" degli operatori di un determinato mercato o settore rispetto ad un determinato titolo o indice.

Come potremmo vedere approfonditamente nel prossimo capitolo ci è stata una grande evoluzione in questa metodologia. Questa è stata accompagnata soprattutto dall'evoluzione nel campo tecnologico che repentinamente ha cambiato i sondaggi telefonici con dati dei social media. Quello che si cercherà di determinare in questa tesi è se l'utilizzo di indicatori di *sentiment* rialzisti o ribassisti per poter aprire e chiudere posizioni all'interno di un portafoglio virtuale può condurre a risultati economicamente positivi. Verranno testate e commentate diverse applicazioni su un *dataset* raccolto personalmente da marzo 2016 a settembre 2016.

Concludo questa introduzione con le prime righe un articolo di Lisa Grossman del 2010 scritto immediatamente dopo la prima versione del “Metodo Bollen” di Johan Bollen e Huina Mao, che come avremo modo di vedere in seguito, è stato di grande impatto ma al contempo molto criticato sia dal mondo accademico che dall’ industria finanziaria.

LISA GROSSMAN SCIENCE 10.12.10 1:30 PM
**TWITTER CAN PREDICT THE
STOCK MARKET**



THE EMOTIONAL ROLLER COASTER captured on Twitter can predict the ups and downs of the stock market, a new study finds. Measuring how calm the Twitterverse is on a given day can foretell the direction of changes to the Dow Jones Industrial Average three days later with an accuracy of 86.7 percent.

“L’ottovolante emotivo catturato su Twitter in grado di prevedere gli alti e bassi del mercato azionario, un nuovo studio lo conferma. Misurare quanto la calma o l’euforia dell’universo di Twitter in un dato giorno può prevedere la direzione dei cambiamenti dell’indice Dow Jones Industrial tre giorni dopo con una precisione di 86,7 per cento.”

Capitolo 2

Letteratura

2.1 MUTAMENTO DELLA SENTIMENT ANALYSIS NEGLI ANNI

Dai sondaggi telefonici agli Hashtag Trends

Il market *sentiment* è monitorato con una varietà di metodi tecnici e statistici costantemente da decenni. Secondo l'interessante visione di Thomas Dorsey la gran parte del movimento globale di un singolo titolo è stato attribuito al *market sentiment*.

Una possibile evidenza empirica comune di questa tesi osservabile sul mercato può essere data dal fatto che spessissimo i titoli bancari hanno una correlazione altissima nelle fluttuazioni di brevissimo periodo, nonostante spesso le *news* che generano queste fasi di ribassi e rialzi collettivi interessino solo un singolo nome (non correlato in nessun modo agli altri) e non l'intero settore, a Wall Street è famosa la frase *trend is a friend*, in italiano "il trend è un amico".

Un particolare filone di letteratura economica collega ai risultati di finanza comportamentale i principi fondamentali di *asset pricing*. Autori come Barberis nel 1998 Barberis e Thaler nel 2003, e Baker & Würigler nel 2007 sostengono che i modelli comportamentali degli investitori al dettaglio hanno un impatto significativo sui rendimenti di mercato.

Almeno cinque principali approcci alla misurazione attenzione (*sentiment*) degli investitori sono noti oggi nella letteratura scientifica: misure basate sul mercato finanziario, indici a sondaggio, estrapolazione del *sentiment* tramite analisi testuali e di risorse specializzate on-line, volumi di ricerca *keyword* su internet, e di fattori non economici.

Secondo il primo approccio, l'attenzione degli investitori e quindi la "direzione" del *sentiment* può essere approssimata con particolari misure basate sul mercato finanziario.

Secondo Gervais et al. (2001) e Hou et al. (2009) il volume degli scambi è un buon indicatore per il *sentiment* degli investitori. Tanto più è alto (basso) il volume di scambio

su una particolare azione, tanto più questo porterà ad un apprezzamento (deprezzamento) del suo valore, ampliando quindi il comportamento dei cosiddetti *noise traders* (ed in particolare *positive feedback trader*) che tendono ad acquistare (vendere) titoli con rendimenti elevati (bassi).

Whaley (2001) e Baker & Würigler (2007) suggeriscono invece il *Chicago Board Options Exchange Volatility Index* (VIX) come misura alternativa al market sentiment in forma non direzionale, in maniera simile *Credit Suisse Fear Barometer* si basa sui prezzi di *zero premium collars* con scadenza a tre mesi. Il *Sentiment Indicator Acertus Markets* incorpora invece cinque variabili (in ordine decrescente di peso nell'indicatore): *price/earning ratio* (una misura di valutazioni del mercato azionario); *momentum* (una misura della psicologia del mercato); volatilità realizzata (una misura del rischio storico recente); i tassi obbligazionari ad alto rendimento (una misura del rischio di credito); e il *TED spread* (una misura del rischio finanziario sistemico). Ognuno di questi fattori fornisce una misura del sentiment del mercato attraverso una lente unica, e insieme possono offrire un indicatore più affidabile del sentiment del mercato rispetto agli indici precedenti.

In un ulteriore studio Kumar & Lee (2006) mostrano che il volume delle transazioni della sola clientela *retail* può essere un'ottima misura di *market sentiment*.

Lo studio dimostra che le operazioni di investitori al dettaglio "sono sistematicamente correlati - cioè, gli individui acquistano (o vendono) i titoli in massa" ovvero vi è un'evidenza che le transazioni tendono molto più ad essere simili che diverse per questa fetta di investitori. Infine l'*Initial Public Offering* (IPO) di una società può generare una grande quantità di informazioni che possono potenzialmente essere utilizzate come *proxy* per il sentiment degli investitori. Ljungqvist et al. (2006) e Baker & Würigler (2007) riportano i rendimenti del primo giorno di IPO e il volume generale dell'IPO come candidati più promettenti per prevedere l'attenzione (sentiment) degli investitori riguardo un particolare settore.

Tutte le misure basate sul mercato però hanno un uno svantaggio importante comune. In particolare, secondo Da et al. (2014): "Sebbene le misure di mercato hanno il vantaggio di essere prontamente disponibile ad una frequenza relativamente alta, hanno lo svantaggio di essere il risultato di equilibrio di molte forze economiche oltre che dei diversi

sentiment degli investitori. " In altre parole, non si può mai essere sicuri che un particolare indicatore basato sul mercato è stato trainato solamente dal grado di attenzione degli investitori.

Inoltre, alcuni indicatori possono lavorare in maniera pro-ciclica. Ad esempio, un elevato volume di scambi (secondo alcuni autori) può simulare un *sentiment* rialzista, con la conseguenza che questo induca il volume degli scambi diventerà ancora più alto.

Questo, a sua volta, porta alla stima di un *sentiment* ancora più rialzista, e così via.

Gli indicatori complessivi di mercato stanno giocando un ruolo molto importante nel misurare l'attenzione degli investitori, tuttavia, un investitore dovrebbe sempre cercare di fare in modo che nessun'altra variabile possa guidare il risultato.

La seconda *proxy* per il *sentiment* degli investitori può essere quella di utilizzare gli indici basati sui sondaggi. Tra gli indici più noti vanno ricordati quello dell'Università del Michigan *Consumer Sentiment Index*, il *Conference Board* di fiducia dei consumatori, e *UBS / Gallup Index* di ottimismo degli investitori. L'Università del Michigan *Consumer Sentiment Index* si basa su almeno 500 interviste telefoniche. L'indagine contiene una cinquantina di domande fondamentali. Tuttavia, il sondaggio si compone di soli cinque domande principali riguardanti le condizioni di lavoro, il lavoro e di reddito. Alle domande si può rispondere con solo tre opzioni: "positivo", "negativo" o "neutro". Tuttavia, secondo Da et al. (2014), utilizzare tali indici può portare a limitazioni significative. In primo luogo questo tipo di dati sono disponibili con frequenza settimanale o mensile. Allo stesso tempo, la maggior parte delle misure di *sentiment* alternative sono disponibili con frequenza giornaliera. In secondo luogo, l'incentivo per gli intervistati nel rispondere alle domande poste con cura ed in maniera veritiera è basso poiché tali sondaggi sono spesso rivolti a campioni di famiglie disinteressate ai fini dell'analisi. Per riassumere quindi, gli indici di *sentiment* basati sui sondaggi possono essere utili nel predire indicatori finanziari, tuttavia l'uso di tali indici ha svantaggi specifici e può essere limitativo in alcuni casi.

Sotto il terzo approccio, i ricercatori propongono di usare algoritmi di *text mining - sentiment analysis* per estrarre informazioni sullo stato d'animo degli investitori dai social network, piattaforme multimediali, blog, articoli di giornale, e di altre rilevanti fonti di dati testuali. Un filo di pubblicazioni (Barber & Odean (2008), Dougal et al. (2012) e Ahern & Sosyura (2015) segnala una significativa influenza degli articoli finanziari sul

comportamento dei prezzi dei titoli. È anche sorprendente che tali fonti di notizie come Wall Street Journal, New York Times o il Financial Times hanno una profonda influenza sul mercato indipendentemente dalla veridicità dell'informazione contenuta negli articoli. La forza dell'impatto può variare tra i diversi giornalisti anche all'interno di una particolare rivista o giornale (Dougal et al. (2012)). Tetlock (2007) suggerisce una misura di *sentiment* degli investitori contando il numero di parole "negative" in una famosa colonna di Wall Street Journal "al passo con il mercato". Zhang et al. (2011) e Bollen et al. (2011), Welpe e Sprenger (2010) introducono per la prima volta la possibilità di utilizzare dati di Twitter come una fonte estremamente importante di dati di *sentiment*, che aiuta a prevedere i prezzi delle azioni e volatilità. Il modo più utilizzato per analizzare l'influenza dei dati da piattaforme *microblog* sul comportamento dei prezzi delle azioni è la costruzione di particolari indici di monitoraggio dell'umore. Il modo più semplice sarebbe quella di contare il numero di parole "positive" e "negative" in ogni *tweet* pertinente e costruire un indicatore combinato base a questi dati. Nasseri et al. (2014) riporta il potere predittivo di *StockTwits* (piattaforma simile a Twitter specializzata sullo scambio di opinioni di trading legate) per quanto riguarda il comportamento dei prezzi delle azioni.

Karabulut (2013) riporta Facebook per essere una buona fonte di informazioni su l'umore degli investitori. Tuttavia, è importante notare che è relativamente più difficile da raccogliere tali tipo di dati (in molti casi un ricercatore ha bisogno di un software speciale). Inoltre, l'analisi di tali dati può anche richiedere profonda conoscenza *machine learning* e *data mining*.

La quarta importante fonte di informazioni sul *sentiment* degli investitori sono i volumi di ricerca su Internet. Questo approccio è supportato da risultati di Simon (1955), che conclude che le persone iniziano il loro processo decisionale di investimento con la raccolta di informazioni pertinenti. Pubblicamente sono disponibili i dati sui volumi di ricerca per la maggior parte dei motori di ricerca su internet partendo dal 2004. Da quel momento molti autori hanno dimostrato l'utilità di tali dati nel prevedere i rendimenti del mercato e gli stessi indicatori tradizionali di *sentiment* (Da et al. (2014), Preis et al. (2013), e Curme et al. (2014)). L'applicazione dei dati di ricerca su internet dà risultati promettenti nel risolvere diversi problemi finanziari. Da et al. (2014) e Dimpfl & Jank

(2015) indagano su un potere predittivo dei dati Google Trends per due misure di volatilità più popolari: volatilità realizzata (RV) e volatilità del mercato (VIX). Entrambi gli studi riportano la dipendenza positiva e significativa tra i volumi di ricerche su internet e le misure di volatilità. Bordino et al. (2012) e Preis et al. (2010) confermano a loro volta la capacità predittiva della quantità di *query* di ricerca per prevedere i volumi di scambio nei mercati azionari degli Stati Uniti. Secondo Bordino, "... i volumi di ricerche anticipano in molti casi i picchi di trading di uno o più giorni." Alcuni ricercatori trovano l'utilità dei dati Google Trends nel predire la volatilità nel mercato delle valute (Smith (2012)). I dati estraibili da Google Trends sono segnalati anche per essere un buon predittore per i flussi *day-by-day* di fondi comuni. Da et al. (2014) conclude che tale tipo di dati è possibile estrarre una misura di *sentiment* che "ha un significativo potere predittivo incrementale su base quotidiana per fondi sia azionari che obbligazionari." Un'altra promettente fonte di dati di ricerca su internet è la quantità di visite di pagine di Wikipedia relative a tematiche finanziarie.

In sintesi, l'utilizzo dei volumi di ricerca su internet è una promettente *proxy* per l'attenzione degli investitori. Tale tipo di dati di *sentiment* non richiede ulteriori informazioni da altre fonti e può essere utilizzato in studi scientifici in modo indipendente.

Infine (quinta fonte), si pone la possibilità che l'attenzione degli investitori può anche dipendere da alcuni fattori non economici. Ogni giorno molti eventi non economici (ad esempio notizie, meteo, stato di salute, ecc.) influenzano il nostro stato d'animo, che inconsciamente, influenza il livello della nostra avversione al rischio e quindi il comportamento di trading. Edmans et al. (2007) per esempio riporta l'influenza di eventi sportivi sul comportamento commerciale degli investitori. Gli autori riportano una forte evidenza di rendimenti azionari anormalmente negativi dopo le perdite in importanti eventi di calcio e *football*. L'effetto di perdita è valida anche dopo le partite internazionali cricket, rugby, basket.

Kaplanski & Levy (2010) studiano l'influenza di cattive notizie (disastri aerei, ferroviari, serial killer) sui prezzi delle azioni. Gli autori concludono che una cattiva notizia (per esempio su disastro aereo) può causare notevole calo rendimenti azionari (soprattutto per le società a bassa capitalizzazione e ad alta volatilità). L'evidenza che la quantità di minuti di luce solare in un giorno particolare influenza il comportamento di un

trader è presentato da Akhtari (2011) e Hirshleifer & Shumway (2003). Gli autori concludono che l' "effetto sole" è statisticamente significativo e robusto a diverse specifiche del modello.

L'influenza della temperatura sulle variazioni dei rendimenti azionari è trattata invece in Cao & Wei (2005). Secondo i risultati dello studio citato, non vi è una dipendenza negativa tra i cambiamenti di temperatura e rendimenti azionari su tutta la gamma di temperature (vale a dire i rendimenti sono più alto quando il clima è freddo). Anche l'attività geomagnetica secondo Robbiotti è candidata per avere un'influenza (negativamente correlata) su rendimenti azionari. Per riassumere, gli eventi non economici hanno una notevole influenza sul comportamento dei trader. Un investitore si dovrebbe aspettare rendimenti di mercato elevati in una giornata soleggiata, ma fredda, senza significative attività geomagnetiche, preferibilmente il giorno dopo una vittoria su un significativo evento sportivo nel quale però a livello globale non sia successo nessun disastro inaspettato. Nella maggior parte dei casi, tali dati devono essere trattati come supplementari nel misurare il *sentiment* degli investitori, ma non in maniera totalmente indipendente.

2.2 APPROCCIO INDIVIDUALE

Presentazione Metodologia Sprenger- Welp

Nel dicembre del 2010 il Dott. Tim O. Sprenger e la Proff.ssa Isabel Welp della TUM School of Management di Monaco pubblicano un *working paper* dal titolo *Tweets and trade: The Information Content of Stock Microblog*.

Questo *paper* spicca fra l'ampia categoria di lavori sull' *sentiment analysis* attraverso l'analisi testuale utilizzando Twitter per la sua innovatività e la possibilità di poter processare un elevato numero di informazioni in tempo reale.

L'architettura del modello è quella più diffusa in questo tipo di analisi, ovvero l'utilizzo della "popolarità" delle discussioni generate in tempo reale su internet per analizzare le performance di prodotti finanziari.

Questo lavoro può essere considerato come una naturale evoluzione del famoso studio svolto da Antweiler e Frank nel 2004, già citato diverse volte. Con l'evolversi dei social media e il boom dei social network negli anni l'utilizzo dei forum specifici (come *Yahoo Finance* nello studio del 2004) è andato calando, inoltre i social network come abbiamo visto sono presentano caratteristiche notevolmente più interessanti per questo tipo di analisi.

Twitter, è stato recentemente in prima linea in questo sviluppo. Alcuni commentatori hanno anche descritto le conversazioni su questa piattaforma come "la versione moderna dei trader che trattavano con aste alle grida" (Business Week (2009)).

Il social network Twitter è un servizio di *microblog* che consente agli utenti di pubblicare brevi messaggi con un massimo di 140 caratteri, i cosiddetti "tweet". Questi *tweets* sono visibili su un *public message board* del sito e sono ottenibili tramite varie applicazioni di terze parti (approfondiremo in seguito questo elemento fondamentale).

Gli utenti possono sottoscrivere (vale a dire, "seguire"), una selezione di autori preferiti o cercare i messaggi che contengono una chiave specifica parola (ad esempio, un simbolo di borsa).

La *timeline* pubblica si è trasformata in un flusso di informazioni in tempo reale che attualmente conta circa 90 milioni di messaggi al giorno di natura correlata o prettamente finanziaria. Molti di questi messaggi sono dedicati alla discussione di titoli di aziende quotate e idee di trading. Come risultato, ci sono investitori che possono attribuire il loro successo finanziario per le informazioni che trovano su siti di social media. Sono stati sviluppati sistemi di trading da professionisti del settore per avvisare gli utenti di opportunità di investimento (Bloomberg (2010)), e modelli da accademici cultori dell'analisi tecnica per prevedere *break-point* in serie temporali finanziarie (Vincent e Armstrong (2010)).

Pertanto, la comunità di investitori ha ideato applicazioni di terze parti come *Stock-Twits.com* che filtrano il flusso continuo di *tweets* e mettono in evidenza solo quelli di natura finanziaria in base ad un ordine di rilevanza gerarchica della fonte, qualcuno ha definito questo tipo di analisi come "il terminale Bloomberg per l'utente medio".

È interessante notare che una delle caratteristiche più frequentemente utilizzate (oltre a quella di sorgente di dati) fra i professionisti sui terminali di Bloomberg, che costando più di \$ 2.000 al mese, è il sistema di chat centralizzato che consente agli operatori di

comunicare tra loro in tempo reale. Twitter offre caratteristiche molto simili ed è disponibile gratuitamente. In realtà, Bloomberg ha implementato i messaggi di Twitter nei loro terminali, mentre NASDAQ ha fatto un passo in più, lanciando un'applicazione mobile che incorpora contenuti da *StockTwits* filtrandoli in base alla rispettiva rilevanza sui mercati.

È chiaro che da questo punto di vista Twitter diventa uno strumento potentissimo nelle mani delle persone influenti che hanno interesse a condividere le loro strategie.

Riporto in seguito la traduzione di una frase riportata sull'importante rivista TIME nel 2009: "La comunità di investitori attivi ed i traders giornalieri stanno condividendo opinioni e in alcuni casi informazione recepita dopo sofisticata ricerca sulle azioni, obbligazioni e altri strumenti finanziari su Twitter. L'istantanea diffusione dell'informazione unita alla rilevanza e la qualità della fonte hanno effettivamente il potere di spostare i prezzi delle azioni, questo fenomeno permette a Twitter di essere considerato per una specifica azione o strumento finanziario una sorgente di *news* simile a qualsiasi altra fonte non uffici".

Entrando nel dettaglio della ricerca degli autori improntano il loro studio principalmente sul lavoro di Antweiler e Frank (2004), i quali usarono metodi di classificazione di testo sofisticati per studiare il contenuto di informazioni sui forum specializzati *Yahoo Finance* e *Raging Bull* riguardanti le discussioni che interessavano le 45 società del *Dow Jones Media Industrial* e del *Dow Jones Internet Index*.

Secondo Antweiler e Frank quel volume di comunicazione poteva essere usato per prevedere volume e la volatilità delle transazioni. Lo studio da loro esposto ha però delle limitazioni innegabili: il periodo analizzato vicino agli anni 2000 include lo scoppio della bolla di Internet, le società *dotcom* con modelli di business insostenibili e valutazioni parzialmente irreali rappresentano una parte sostanziale del campione osservato. A differenza del lavoro di Antweiler e Frank i due autori utilizzano Twitter come fonte primaria di informazioni, nonostante i molti parallelismi possibili con i forum analizzati in precedenza, esistono caratteristiche distinte del social network che fanno pensare che questi risultati possano essere considerati come una generalizzazione dei e risultati precedenti nonostante siano stati ottenuti da una fonte diversa.

In primo luogo, la *timeline* pubblica di Twitter non ha bacheche categorizzate in *topic* separati per ogni società, questo può portare a una significativa attenzione alle informazioni non aggiornate finché non ci sono *news* più recenti, oltre alla possibilità di non dover cambiare continuamente discussione per trovare notizie riguardanti diverse società.

In secondo luogo, mentre i forum richiedono agli utenti di entrare attivamente o rispettare determinati canoni per partecipare ad una discussione riguardante un particolare argomento, Twitter viene utilizzato più come un mezzo di comunicazione informale e quindi neutro. In terzo luogo, i *microbloggers* hanno un forte incentivo a pubblicare informazioni preziose al fine di mantenere o aumentare la propria popolarità, il tasso di *retweet* (vale a dire, le citazioni da parte di altri utenti) e la loro platea di *follower*. Gli autori sostengono che questi incentivi forniscono alla *Twittersphere* con un meccanismo per pesare la qualità delle informazioni.

Accanto alle differenze nelle tipologie di bacheche internet, c'è un secondo aspetto che garantisce l'usabilità dei cosiddetti "stock microblog". La natura di Twitter rende aspetti precedentemente non disponibili informazioni parzialmente osservabili (ad esempio, *retweet* ed il numero di *follower* che visualizzano la notizia). Tuttavia, la ricerca scientifica non ha ancora esplorato questi meccanismi per la diffusione delle informazioni siano realmente utilizzabili in modo efficace. Così, non è chiaro se, su larga scala, i *microbloggers* producono informazioni preziose o semplicemente rappresentano l'equivalente online *noise traders* disinformati.

Pertanto, lo scopo degli autori è quello di indagare se e in che misura i *microblog* (Twitter in particolare) possono riflettere e influenzare gli sviluppi dei mercati finanziari. In particolare, per la comparabilità con relativa di ricerca (per esempio, Antweiler e Frank (2004)), il loro studio confronta la relazione tra la prevedibilità dei rendimenti, il volume degli scambi, e la volatilità del mercato con i *tweets* corrispondenti disponendo del *sentiment* dei messaggi *postati* sul social (vale a dire, tendenza rialzista "Bullishness" e ribassistica "Bearishness"). Inoltre, gli autori tentano di spiegare empiricamente il possibile meccanismo dietro l'aggregazione efficiente delle informazioni nei *microblog* (social network).

Secondo l'ipotesi efficiente mercato (EMH) i mercati finanziari sono "informativamente" efficienti, nel senso che i prezzi di mercato riflettono tutte le informazioni note.

La versione “semi forte” dell’EMH è ampiamente accettata e sostiene che i prezzi sono frutto di tutte le informazioni disponibili al pubblico e riflettono istantaneamente le nuove informazioni pubbliche prevenibili.

Tuttavia, un numero crescente di ricerche suggerisce che i mercati finanziari non sono sempre conformi con la EMH. Recenti studi hanno suggerito che le *news* particolarmente qualitative non si riflettono pienamente e immediatamente a prezzi di mercato. Tetlock, Saar-Tsechansky, e Macskassy (2008) hanno evidenziato che i prezzi delle azioni delle imprese reagiscono meno di quanto dovrebbero alle informazioni testuali incorporati in fatti di cronaca.

Gli autori sopra citati sostengono che i “rumors” stanno diventando sempre più la vera aspettativa del mercato e che le strategie di trading basate sul rapporto tra i *rumors* e le “*first call forecast*” permettono di guadagnare rendimenti anomali. Fonti di dati qualitativi, come quelli menzionati sopra, sono stati ampiamente trascurati nella letteratura finanziaria, forse perché i metodi di linguistica computazionale, come applicata in questo studio, che sono necessarie per elaborare le informazioni non appartengono propriamente al bagaglio culturale dei ricercatori economici.

Come già detto in precedenza gli autori di questo *working paper* si pongono come scopo principale quello di investigare la correlazione che c’è fra i rendimenti dei diversi titoli azionari e i *tweets* riguardanti le aziende interessate.

Prima di scendere nel dettaglio della metodologia credo sia necessario esporre le ipotesi che gli autori si impongono di testare:

Ipotesi A: Una crescita nel sentiment rialzista (bullishness) di una azione è associata ad un aumento dei rendimenti di quella determinata azione.

Ipotesi B1: L’ aumento del numero dei messaggi sui social network o sui forum specializzati è associato ad un aumento del volume di trading.

Ipotesi B2: Un aumento del numero di messaggi riguardanti un titolo azionario sui social network o forum specializzati è associato a rendimenti più alti per quello stesso titolo.

Ipotesi B3: Un aumento del numero dei messaggi sui social network e sui forum specializzati per un determinato titolo azionario comporta un aumento della volatilità per il medesimo titolo.

Ipotesi C1: Un aumento del disaccordo nei social network o forum specializzati comporta un aumento della volatilità.

Ipotesi C2: Un aumento del disaccordo all'interno dei social network e forum specializzati è associato ad un aumento del volume di scambi effettuati.

Molte delle ipotesi sopra citate sono state già testate all'interno del mondo accademico prima della stesura di questo *paper* con risultati negativi nella maggior parte dei casi.

Gli autori però, nella maggior parte degli studi, avevano testato le ipotesi solo sui forum specializzati e non su *social network* come Twitter.

Le motivazioni principali che hanno spinto il 2010 il Dott. Tim O. Sprenger e la Proff.ssa Isabel Welpé a rieseguire i test su queste ipotesi utilizzando Twitter anziché i forum specializzati sono le seguenti:

- A differenza dei forum specializzati che dividono le conversazioni in categorie ben fissate e divise, Twitter avendo una *timeline* unica si presta molto meglio a catturare una “naturale” conversazione sul mercato che potrebbe avvenire spontaneamente in qualunque altro luogo.
- Per poter usufruire di forum è necessario essere attivo, rispettare un determinato “regolamento” che talvolta può essere limitativo e non è possibile uscire al di fuori dell'argomento prestabilito. Twitter invece è molto più aperto e libero, non necessita di iscrizioni particolari e tutti possono dire la loro.
- Nei forum specializzati gli utenti non hanno particolari incentivi a scrivere informazioni di buona qualità in materia sintetica per aumentare la propria reputazione, anzi spesso in questi forum la reputazione aumenta in base al numero di messaggi che si carica sulla piattaforma. Su Twitter c'è tutta un'altra filosofia, esiste ugualmente un contatore di messaggi inseriti, ma la popolarità è direttamente proporzionale alla quantità di persone che apprezzano, condividono e commentano le tue pubblicazioni; questo fattore insieme alla formattazione massima di 140 caratteri (nel 2010 la dimensione massima del testo caricabile era ancora limitata a differenza di oggi) contribuiscono a fare in modo che gli utenti puntino a condividere solo i loro aggiornamenti migliori sintetizzando la notizia.

- Ultimo fattore, grazie all'architettura della piattaforma è possibile investigare anche sulla diffusione della notizia, inoltre l'utilizzo di *hashtag* come vedremo aiuta ad indicizzare le notizie in maniera del tutto naturale.

Prima di passare ai risultati di questo *working paper* ritengo necessaria un'esposizione di quella che è stata la metodologia utilizzata, evidenziando la peculiarità di questo lavoro che si posiziona accademicamente parlando più vicino al ramo dell'informatica e dell'analisi linguistica computazionale che a quello prettamente finanziario.

Nonostante i vantaggi di Twitter elencati in precedenza, questo *social network* oltre ad avere il più ampio tasso di inserimento nel mondo della *community* finanziaria dispone di strumenti interni ed esterni come le API, *Applications programming interfaces*, che lo rendono particolarmente adatto a questo tipo di analisi.

Come tutti sanno un *tweet* può essere composto da 140 caratteri, all'interno del testo possono essere utilizzate diverse formattazioni per menzionare un determinato utente oppure fare in modo che il *tweet* scritto faccia parte di un flusso di tutti strutturato per una determinata notizia. Lo strumento più utilizzato è l'*hashtag*, che in puri termini di formattazione testuale corrisponde ad un cancelletto # seguito dal tema interessato (es. #rio2016, #olimpiadi #brexit ..). Questa semplice formattazione permette agli utenti della *Twittersphere* di poter seguire le notizie ancorata a questi temi semplicemente cliccando sull'*hashtag* o cercandolo nella apposita barra di ricerca in modo da visualizzare tutti i *tweets* pubblicati dagli utenti che richiamano lo stesso *tag*.

Un meccanismo molto simile a questo è stato adottato per le discussioni e le notizie che riguardano prodotti finanziari quotati sostituendo il cancelletto con il simbolo del dollaro seguiti dalla sigla borsistica (per indicizzare le discussioni su Unicredit useremo \$UBG), un meccanismo che si è diffuso rapidamente e che permette al popolo degli investitori di Twitter di richiamare tutte le notizie su un determinato prodotto in maniera rapida e precisa.

Gli autori concentrano il loro studio su un periodo di 6 mesi che va dal 1 gennaio 2010 al 30 giugno 2010, in un momento di relativa calma dei mercati ormai lontani dalla crisi dei *subprime* del 2008. Durante questo periodo sono stati raccolti 249,533 *tweets* in lingua inglese contenenti *hashtag* che potessero ricondurli a discussioni di una società quotata facente parte dello SP100. La scelta di questo specifico *target* è motivata dal

fatto che l'indice SP100 rispecchia fedelmente quello che è lo spettro delle differenti tipologie di aziende ed industrie quotate americane.

Per classificare il tipo di testo, vista la mole ingente di dati, è stato utilizzato un classificatore testuale molto utilizzato nei lavori riguardanti la linguistica computazionale.

Il programma utilizzato per la classificazione è quello della *Weka Machine Learnings*, mentre il criterio è quello *Naive Bayesian Multinomiale*.

Il metodo *Naive Bayesian* è uno dei più usati per la classificazione testuale, è un algoritmo che, spiegato in maniera non dettagliata, utilizza il concetto fondamentale della probabilità bayesiana per determinare se un messaggio composto da diverse parole appartiene ad una determinata classe in base alla probabilità condizionata delle parole che lo compongono di appartenere alla classe. Questo metodo anche se poco avanzato rispetto ai nuovi sistemi di analisi per la linguistica computazionale, ha il vantaggio di essere chiaro e trasparente nella classificazione.

Per la programmazione del classificatore gli autori hanno inserito e classificato manualmente all'interno del programma 2500 *tweet* identificando manualmente il *sentiment* del messaggio; ogni stringa quindi veniva classificata in tre categorie di segnali:

Buy, Hold, Sell. Nella tabella 2.1 sotto è possibile vedere un esempio della classificazione manuale delle varie stringhe.

Dopo una prima selezione i risultati del campione si mostrano molto diversi da quelli ottenuti con lo studio dei forum specializzati, in particolare il campione è stato classificato con il 49.6% di messaggi "HOLD", 35.2% di "BUY" e 15.2% di "SELL". Questo risultato mostra come i social network diano una visione molto meno unidirezionale e più reale del *sentiment* generale rispetto ai dati ottenuti nei lavori di Dewally del 2003 e di Antweiler e Frank del 2004 che davano dei rapporti fra messaggi *bullish* e *bearish* rispettivamente di 7:1 e di 5:1.

Inoltre, la distribuzione più equilibrata di acquisto e segnali di vendita rispetto ai precedenti studi fornisce una maggiore quota di segnali di vendita nel *set* di dati principale (10,0% rispetto a solo il 1,3% nello studio di Antweiler e Frank). Questo permette agli autori di esplorare il contenuto informativo dei segnali separatamente.

La metodologia utilizzata per l'aggregazione dei *tweets* è la stessa utilizzata da Antweiler e Frank. Avendo necessità di dover processare centinaia di stringhe al giorno è obbligatorio utilizzare degli indici appositamente costruiti che ci permettano di aggregare le informazioni per poi poterle confrontare con gli obiettivi della ricerca.

Tabella 2.1

Sample Tweets from Training and Test Set including Classification

Tweets were randomly selected and are shown in their original format (before preprocessing).

Sample tweets* (training set)	Manual classification
RT @bampairtrading \$TGT Target Q4 Profits Surge http://bit.ly/ciQFjY	Buy
Great place to short \$X. Stop loss at 54.25. I am still short via puts from Friday HOD.	Sell
Big banks up or down with Bernanke's re-nomination? \$C \$BAC	Hold
\$DELL (Dell Inc) \$13.87 crossed its 1st Pivot Point Resistance #emppv #stocks http://empirasign.com/s/42f	Buy
Heinz Q3 EPS of 83c beats by 6c. Revenue of \$2.6B meets. \$HNZ #earnings http://bit.ly/avIHfH	Buy
Microsoft Corporation \$MSFT Not Moving. Docuware Integration In Microsoft Outlook: http://bit.ly/db66Ox	Hold
\$AXP looking strong here	Buy
\$BA Boeing Sees Sales Drop, Maintains 737 Output http://bit.ly/9kmvUa	Sell
Trader Bots has recently calculated a Neutral Overall Stock Prediction on \$TGT http://bit.ly/7k5H	Hold
I think if \$AMZN closes above 116 today! You could go long tomorrow.	Buy

Si definisce l'indice B_t come "Bullishness"; ovvero indice del *sentiment* collettivo rialzista al tempo t , con la formula 1.

$$(1) \quad B_t = \ln \frac{(1+M_t^{BUY})}{(1+M_t^{SELL})}$$

Dove M^{BUY} (M^{SELL}) rappresentano il volume di messaggi giornalieri di categoria “BUY” (“SELL”). Questa misura ha la particolarità di dare un peso solo relativo ai volumi dei messaggi di acquisto e di vendita poiché l’indicatore di *sentiment* è generato dalla trasformazione logaritmica del rapporto fra i due volumi giornalieri. Ci sono due problemi con questa misura di volume. In primo luogo, data la crescita dei social network come Twitter, il volume totale può non essere un indicatore stabile nel tempo. In secondo luogo, il volume dei messaggi può variare leggermente quindi presentare una serie fin troppo stabile. Pertanto, per ciascuna impresa, gli autori hanno calcolato anche un volume messaggio normalizzato rispetto al totale numero di messaggi al giorno. Mentre questo indicatore fornisce una misura analoga della quota relativa dei messaggi per ogni società, essa non riflette il volume assoluto. Il secondo indicatore, sempre in linea con lavoro svolto da Antweiler e Frank è quello che misura tasso di “accordanza”, ovvero quanto sono d’ accordo i messaggi in una specifica data. Con A_t gli autori indicano il tasso di “agreement” nelle raccomandazioni espresse rispetto ad una determinata società nel corso del periodo t .

$$(2) \quad A_t = 1 - \sqrt{1 - \left(\frac{M_{ct}^{Buy} - M_{ct}^{Sell}}{M_{ct}^{Buy} + M_{ct}^{Sell}} \right)^2}$$

È di semplice intuizione che qualora ci trovassimo di fronte ad una totalità di messaggi rialzisti (ribassisti) l’indicatore avrebbe il valore 1.

Anche dopo l’aggregazione dei singoli messaggi di indicatori quotidiani, ci sono giorni per alcuni titoli senza alcun *tweet*. Tuttavia, dato che il *set* di dati contiene una serie completa di entrambi i *tweets* associate alle variazioni di mercato dei prezzi per più del 80% di tutte le società con una frequenza giornaliera, l’influenza dei periodi di silenzio sui risultati è limitato.

Infine, poiché usiamo i dati finanziari dal NASDAQ e NYSE, gli autori si allineano con gli orari di negoziazione degli Stati Uniti (9:30 alle 16:00). Così, i messaggi inviati dopo che i mercati chiudono sono inclusi insieme con messaggi pre-mercato del giorno successivo nel calcolo dell’indice per il giorno seguente, perché questi messaggi non

possono aver influenzato le trattazioni avvenute a mercato aperto ma potrebbero influenzare quelle nelle ore del mercato aperto del giorno successivo.

Analizzando il *dataset* e confrontandolo con i dati relativi ai volumi delle transazioni di mercato si ottengono i primi segni (seppure ancora grezzi) di una correlazione fra il numero dei *tweets* e volume di scambi. Il grafico 2.1 mostra una forte correlazione ($r = 0.469$, $p=0.016$) fra le due variabili, in particolare il volume dei messaggi prevede l'aumento degli scambi nel periodo di gennaio e la caduta degli scambi agli inizi di febbraio quando il volume giornaliero dei *tweets* era diminuito, così come in aprile il comportamento delle due variabili è molto simile. Il risultato, in linea con gli altri articoli accademici riguardanti questa relazione, conferma già dalla fase iniziale di analisi la validità di Twitter come sostituto (migliore) dei forum specializzati.

Grafico 2.1



I risultati di questo *working paper* cercano di dare una risposta a tutte le ipotesi formulate in precedenza, come avremo modo di vedere secondo gli autori l'utilizzo di questo tipo di variabili riscontra risultati migliori nell'approccio non aggregato.

In una prima sezione gli autori esaminano la relazione fra *tweets* e mercato.

Attraverso la stima di correlazioni a coppie gli autori ottengono una prima indicazione sulla relazione che sussiste fra i prezzi di mercato ed i *tweets*.

La tabella 2.2 riassume tutte le statistiche ottenute in questa prima fase di analisi, dalla tabella è facile notare correlazione fra *bullishness* e rendimenti che mostra un'altissima significatività statistica ($r = 0.166$, $p = 0.002$), è semplice notare inoltre come affinando l'analisi ai soli rendimenti anormali i risultati sono altrettanto incoraggianti poiché anche se la correlazione stimata anche se più bassa mantiene la sua significatività. Analizzando i rendimenti *Market Index Adjusted* si evince una correlazione fra rendimenti anomali e l'indice di *bullishness* ad un livello molto simile a quello osservato precedentemente ($r = 0.156$, $p = 0$), così come con i rendimenti *Market Model Adjusted* i risultati sono altrettanto incoraggianti ($r = 0.147$, $p = 0$). Questi tre risultati non possono che supportare l'ipotesi A secondo la quale la crescita del *Bullish sentiment* è associata ad un aumento dei prezzi.

Inoltre così come analizzati in precedenza il volume di trading ed il volume di *tweet* registrati per i medesimi giorni presentano la correlazione più alta registrata pari circa al 44% ($r = 0.441$, $p = 0$) supportando quindi l'ipotesi B1, inoltre non si trova una correlazione forte fra volumi di *tweet* e rendimenti oppure fra volume di *tweet* e volatilità a supporto delle ipotesi B2 e B3.

Tabella 2.2

	Return	AR (simple)	AR (market model)	Traded volume	Volatility	Bullishne ss	Message volume
AR (simple)	0.757 0.00						
AR (market model)	0.658 0.00	0.917 0.00					
Traded volume	-0.044 0.00	0.004 0.65	0.029 0.00				
Volatility	-0.063 0.00	-0.016 0.08	-0.012 0.17	0.046 0.00			
Bullishness	0.166 0.00	0.156 0.00	0.147 0.00	0.126 0.00	-0.012 0.20		
Message volume	0.028 0.00	0.022 0.02	0.024 0.01	0.441 0.00	0.016 0.08	0.340 0.00	
Agreement	0.036 0.00	0.035 0.00	0.033 0.00	-0.113 0.00	-0.014 0.13	0.362 0.00	-0.016 0.08

Come era stato ipotizzato invece il volume di trading decresce all' aumentare del consenso nei *tweets* ($r = -0.113$ $p = 0$) in accordo con l'ipotesi C1 mentre infine la relazione fra tasso di consenso e volatilità risulta non statisticamente significativa anche a livella di confidenza elevati ($r = -0.014$ $p = 0.13$) mostrando in oltre un coefficiente di correlazione in valore assoluto prossimo a 0 in disaccordo con l'ipotesi C2.

Nel complesso, vale la pena notare che molte correlazioni tra *tweets* e mercato presentano caratteristiche più forte di relazioni stimate su variabili strettamente collegate al mercato, come ad esempio il rapporto tra volatilità e il volume degli scambi ($r = 0.046$, $p = 0.0$).

L' analisi continua con delle regressioni panel per testare la relazione fra volumi di *tweets* e rendimenti di mercato in corrispondenza delle ipotesi per testare se *tweets* possono essere usate come *proxy* per gli sviluppi del mercato.

La tabella 2.3 riassume i risultati ottenuti con questa tecnica per le ipotesi sopraelencate.

Tabella 2.3

	Return	AR ^{simple}	AR ^{market model}	Trading volume	Volatility
Bullishness	0.492*** <i>18.41</i>	0.347*** <i>17.67</i>	0.318*** <i>16.84</i>	-1.872*** <i>-3.65</i>	-0.593 <i>-1.27</i>
Message volume	0.002 <i>0.10</i>	0.014 <i>0.98</i>	0.012 <i>0.88</i>	10.798*** <i>28.68</i>	1.391*** <i>4.06</i>
Agreement (A)	-0.170*** <i>-4.15</i>	-0.127*** <i>-4.21</i>	-0.106*** <i>-3.66</i>	-4.644*** <i>-5.88</i>	-0.839 <i>-1.17</i>
Market return	0.066*** <i>16.17</i>	0 <i>0.06</i>	-0.003 <i>-1.08</i>	-2.057*** <i>-26.32</i>	-0.160* <i>-2.24</i>
R ²	0.052	0.028	0.026	0.104	0.002
F-value	168.1***	89.0***	81.1***	358.7***	5.5***

Notes: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, t-statistics in italics below the coefficients.

I risultati della regressione *panel* mostrano innanzitutto una forte relazione coerente con l'ipotesi A con tutte e tre le variabili utilizzate per i rendimenti. Quindi anche in questo caso, un incremento nella *bullishness* può essere usata come *proxy* per la tendenza rialzista degli investitori che si concretizza in un innalzamento dei corsi azionari.

Gli autori trovano inoltre supporti anche all' ipotesi B1, rafforzando quindi l'ipotesi che i messaggi crescano a al crescere dei volumi degli scambi, essendo entrambe le misure soggette a trasformazione logaritmica è possibile quindi identificare nel coefficiente una sorta di "elasticità" con coefficiente 10.798 significativamente diverso da 0 con un livello di confidenza del 99.9%. Osservando ancora la tabella è possibile notare come i risultati non siano coerenti con l'ipotesi B2 secondo cui il volume dei messaggi possa dare informazioni utili riguardo ai rendimenti dei rispettivi titoli, la situazione però è diversa per quanto riguarda l'ipotesi B3, in quanto si osserva che al crescere della volatilità del mercato aumenta il numero di *tweets postati* ($c = 1.391$, $p < 0.001$).

Per quanto riguarda invece l'ipotesi C1 e C2 si hanno risultati discordanti, con una prevalenza di correlazione negativa fra tasso di *agreement* e volume di negoziazioni ($c = -4.644$, $p < 0.001$).

L'analisi continua con delle regressioni di tipo temporale o *Time-sequencing*.

In questa sezione gli autori cercano di individuare il legame temporale fra le variabili, a differenza delle regressioni su base continua precedenti questo tipo di analisi utilizza dati "ritardati" per stimare gli effetti possibili sul mercato dei volumi di tweet.

In linea continuativa le ipotesi che saranno testate sono solo quelle non rigettate nell'

Tabella 2.4

X	Y	X->Y					Y->X				
		X-1	X-2	NWK	Market	F-value	Y-1	Y-2	NWK	Market	F-value
Bullishness	Return	0.006	-0.057*	0.435***	0.068***	97.2***	0.035***	0.003	0.101***	0.006***	41.5***
		0.002	-0.022*	0.091***	0.146***		0.091***	0.009	0.054***	0.033***	
Bullishness	Volume	0.332	-1.191*	-13.182***	-1.828***	200.5***	0.000	0.000	0.082***	0.009***	15.6***
		0.002	-0.008*	-0.046***	-0.066***		0.049	0.014	0.044***	0.049***	
Bullishness	Volatility	-0.047	-0.922*	-1.358	-0.128	3.0*	0.000	0.000	0.086***	0.008***	101.6***
		-0.001	-0.021*	-0.016	-0.016		0.004	0.008	0.046***	0.045***	
Messages	Return	-0.013	0.032	0.436***	0.067***	96.3***	0.006	0.003	0.348***	0.016***	95.0***
		-0.010	0.024	0.091***	0.144***		0.008	0.004	0.097***	0.047***	
Messages	Volume	5.777***	1.590***	-12.526***	-1.966***	287.2***	0.004***	0.001**	0.296***	0.025***	202.8***
		0.073***	0.020***	-0.044***	-0.071***		0.312***	0.057**	0.082***	0.072***	
Messages	Volatility	0.064	0.331	-1.315	-0.143*	2.0	0.001***	0.001**	0.342***	0.017***	101.6***
		0.003	0.014	-0.016	-0.018*		0.025***	0.015**	0.095***	0.049***	
Agreement	Return	0.024	0.047	0.436***	0.067***	96.3***	0.005*	0.000	0.013	0.000	1.5
		0.006	0.011	0.091***	0.145***		0.019*	0.002	0.011	0.003	
Agreement	Volume	-1.370	-2.022**	-13.164***	-1.830***	201.9***	0.000	0.000	0.013	0.000	1.0
		-0.006	-0.008**	-0.046***	-0.066***		-0.048	0.021	0.011	0.004	
Agreement	Volatility	-0.254	0.798	-1.329	-0.137	2.1	0.000	0.000	0.011	0.001	1.4
		-0.004	0.011	-0.016	-0.017		-0.008	0.015	0.009	0.006	

analisi precedente, i risultati ottenuti nella tabella 2.4 sono stati ottenuti ritardando rispettivamente le serie di uno o due giorni (in linea con il lavoro di Antweiler e Frank). Questa volta i risultati ottenuti non sono del tutto concordanti con le ipotesi precedentemente confermate. In particolare per l'ipotesi A, probabilmente la più importante in questo tipo di analisi.

La tabella nostra che l'effetto di un aumento nella *bullishness* ha praticamente un impatto quasi nullo sul giorno successivo di negoziazioni, inoltre il coefficiente associato a tale relazione con un *lag* di due giorni presenta un segno negativo completamente in disaccordo con i risultati fin ora trovati ($r = -0.057$, $p < 0.05$). Invertendo la causalità i rendimenti del giorno precedente influenzano il livello di "bullishness" nei messaggi il giorno successivo ($c = 0.035$, $p < 0.01$). Il coefficiente standardizzato mostra che, oltre all'alto livello di significatività statistica, l'effetto dei rendimenti sul indice *bullishness* dei messaggi è quattro volte superiore rispetto al caso inverso ($c = 0.091$, $p < 0.001$ vs $c = -0.022$ $p < 0.05$). Questo sta a significare che i rendimenti positivi influenzano molto più il *sentiment* rialzista nei *tweets* che vice versa.

Continuando l'analisi dei risultati il volume dei messaggi sembra essere un buon predittore del volume delle negoziazioni del giorno successivo o di due giorni dopo (Ipotesi B1), allo stesso modo però il la relazione sembra valere al contrario nel solo caso di due giorni di *lag*. Il volume dei messaggi sembra inoltre essere influenzato dall'alta volatilità, confermando che l'incertezza spinge gli investitori a scambiarsi informazioni sulle posizioni prese. La relazione opposta in ogni caso però non conferma l'ipotesi B3. In linea con le regressioni simultanee i risultati sulle ipotesi C1 riguardanti il tasso di *agreement*, in disaccordo invece per quanto riguarda l'ipotesi C2 dove ora è stata trovata un'evidenza che conferma la possibilità che il consenso fra i messaggi postati su Twitter sia una buona *proxy* per inferire un ribasso nel volume di transazioni nella giornata borsistica due giorni successivi all'osservazione ($c = -2.022$, $P = 0.01$).

Le regressioni temporali mostrano quindi che mentre solo in alcuni casi caratteristiche dei *tweets* danno informazioni utili per una previsione del mercato (specialmente il livello di *bullishness* per i rendimenti e il volume dei messaggi per il mercato), è molto più probabile che esista causalità invertita, cioè le caratteristiche del mercato influenzano le caratteristiche dei *tweets*.

Gli autori continuano la loro analisi focalizzandosi quindi solo su quelle relazioni che hanno mostrato un'evidenza empirica più forte nelle regressioni precedenti, quindi la relazione fra volume di trading e volume dei *tweets* e quella fra la *bullishness* ed i rendimenti.

Partendo dalla relazione fra i volumi di negoziazioni e *tweets*, a differenza di tutte le altre analisi in questo caso gli autori assegnano il volume dei *tweets* ad un determinato giorno come l'interno campione di *tweets* raccolti nelle 24 ore antecedenti l'apertura dei mercati. La motivazione alla base di questo approccio alternativo sta nel fatto che questa fascia oraria rispecchia in maniera più coerente il *set* informativo necessario per produrre una stima del *set* informativo disponibile. Anche in questo caso sono state utilizzate regressioni di tipo *panel* però a differenza di prima vengono incluse diverse variabili di controllo come ad esempio, il *quality spread* e il differenziale fra il tasso a pronti delle *Federal Reserve* ed i *Tesaurry bill* con scadenza a 10 anni, inoltre per catturare le anomalie di calendario sono state inserite delle *dummy* settimanali e delle altre *dummy* in precedenza dei giorni festivi.

Tabella 2.5

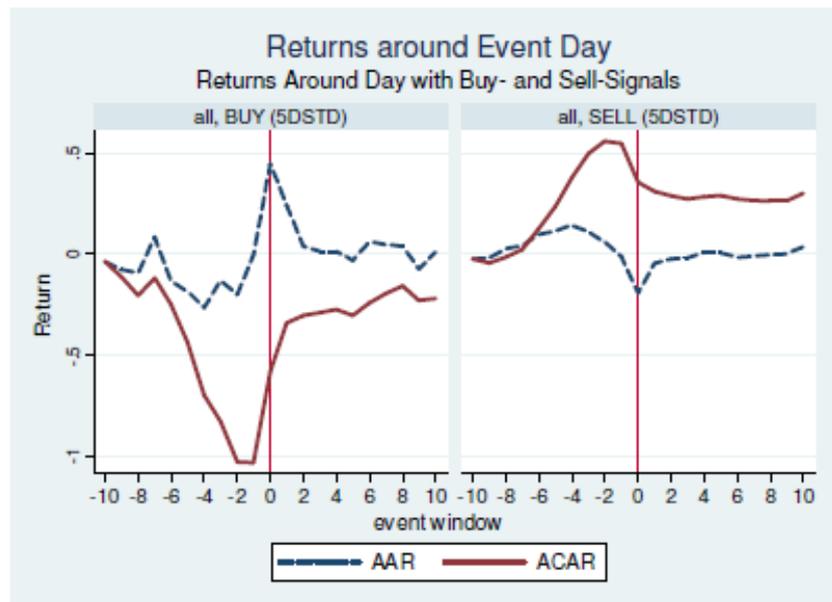
	Actual values	Standardized coefficients
Message volume	0.014***	0.019***
Agreement	-0.008	-0.004
Trading volume up yesterday	-0.266***	-0.059***
Trading volume down yesterday	0.181***	0.036***
Trading volume up last 5 days	-0.265***	-0.073***
Trading volume down last 5 days	0.087***	0.022***
Stock up yesterday	-1.575***	-0.017***
Stock down yesterday	0.201	0.002
Stock up last 5 days	-0.835***	-0.015***
Stock down last 5 days	-0.835***	-0.015***
Stock 5 day volatility	-0.001***	-0.010***
Market up yesterday	-9.750***	-0.064***
Market down yesterday	-2.076***	-0.017***
Market up last 5 days	2.220***	0.022***
Market down last 5 days	0.525	0.008
Market 5 day volatility	0.009***	0.047***
Federal funds rate	-0.329**	-0.010**
Quality spread	-0.497***	-0.025***
Term spread	-0.074	-0.005
Monday	-0.259***	-0.100***
Tuesday	-0.041***	-0.017***
Wednesday	-0.081***	-0.033***
Thursday	-0.081***	-0.032***
Holiday	-0.123***	-0.022***
Observations	11,736	
Adj. R ²	0.273	
F-value	188.6***	

Notes: * p<0.05, ** p<0.01, *** p<0.001.

La tabella 2.5 rileva degli aspetti chiave della relazione fra volume di messaggi ed *agreement*. Innanzitutto, mentre la direzione dei coefficienti è coerente con le analisi precedenti per entrambi gli indicatori, solo il volume dei messaggi mantiene la sua significatività statistica dopo aver introdotto tali variabili di controllo. Inoltre, con un confronto dei coefficienti standardizzati è possibile vedere che il volume dei messaggi è una variabile migliore ai fini previsionali rispetto a quelle di controllo per la stima dei volumi degli scambi del giorno successivo, ulteriore conferma della definitiva validità dell'ipotesi B1.

Per quanto riguarda invece le informazioni contenenti i *tweets* ed i rendimenti, i risultati fino ottenuti indicano che il *sentiment* dei messaggi può contenere informazioni non ancora riflesse dai prezzi di mercato. Tuttavia, la misura di *bullishness* aggregata non permette di analizzare distintamente le varie raccomandazioni per carpire quale di esse può racchiudere informazioni rilevanti rispetto alla massa. Gli autori per ovviare a questo problema svolgono una analisi approfondita di tipo *event study*).

Figura 2.1



La figura 2.1 mostra lo sviluppo della media dei rendimenti anomali intorno all' *event day*, dove *per event day* è considerato quel giorno in cui l'indice aggregato di bullishness eccede la media dell'indice 5 giorni precedenti di almeno 2 *standard deviation* (per approfondimenti su questa particolare imposizione rimando al lavoro di Tumarkin e Whitelaw riguardante la qualità dell'informazione disponibile online (2001)).

È semplice osservare come gli eventi di tipo *BUY* sono precedute da periodi estesi di rendimenti negativi. I rendimenti anomali sono negativi da 6 a 2 giorni prima delle raccomandazioni; risultato in linea con i risultati ottenuti nelle analisi precedenti. Un giorno prima dell'evento invece i rendimenti tornano a salire. Questo cambio di tendenze nei rendimenti è cruciale per l'analisi, poiché il cambio di tendenza consente di ottenere dei rendimenti anomali statisticamente significativi nel giorno successivo a quelli delle raccomandazioni. La media dei rendimenti anomali per il giorno successivo all' *event day* è di 45 punti base, una misura che eccede in molti casi le commissioni di *brokeraggio*, i traders quindi possono utilizzare queste raccomandazioni per entrare in una posizione il giorno successivo attendendosi un rendimento medio dello 0.45%.

Gli autori trovano caratteristiche simili per le raccomandazioni di tipo *SELL*. I rendimenti delle azioni raccomandate come da vendere salgono dal sesto giorno antecedente l' *event day* fino a due giorni antecedenti per poi raggiungere il picco massimo nel giorno precedente. Anche in questo caso la media degli *abnormal return* nei giorni successivi (-0.19%) è statisticamente significativa anche se meno profittevole che nel caso delle raccomandazioni di acquisto. Per sintetizzare questo *event study* mostra che i trader che utilizzano e condividono su Twitter le proprie informazioni di trading utilizzano una strategia di tipo *contrarian* con forti segnali di acquisto nelle raccomandazioni seguiti da giorni con ritorni anomali il giorno successivo (in linea con l'ipotesi A). Questo particolare risultato è in disaccordo con tutta la letteratura precedente sui forum *microblog*, dove gli utenti seguivano una strategia di trading semplicemente basata sul *momentum* e che nessuna informazione veniva fornita dalle raccomandazioni.

L'informazione sui social network quindi cambia ruolo, prima considerata come mero strumento identificativo della "popolarità" di un determinato strumento finanziario sul quale si voleva investire cavalcando l'onda, ad un ruolo più caratterizzante come *proxy* per prevedere il comportamento nei successivi giorni di negoziazione della massa di *noise trader* e non solo.

La ricerca degli autori continua con un ampio studio sulla diffusione delle notizie all'interno dei social network, rimando al *paper* principale gli approfondimenti su questa metodologia poiché non attinente con lo scopo finale e soprattutto l'applicazione empirica finale fulcro portante di questo testo.

Volendo riassumere i risultati di questo primo studio sulla relazione fra social network e mercati possiamo mettere in evidenza come prima cosa che i social network contengono informazioni valide che non sono del tutto incorporate negli indicatori di mercato e che la possibilità di poter seguire membri rilevanti e la possibilità di *retwettare* i loro suggerimenti fornisce ai social network un efficiente meccanismo di aggregazione dell'informazione. La tabella 2.6 riassume tutti i risultati ottenuti dagli autori su un campione di circa 250000 *tweets*. I risultati riguardanti la possibilità di sfruttare la *bullishness* dei messaggi sono incoraggianti per quanto riguarda i segnali di acquisto sulle relazioni ritardate.

Tabella 2.6

Analysis of tweet and market features	<i>Contemporaneous relationship</i>	<i>Lagged relationship</i>
<i>Bullishness</i>		
Increased bullishness of stock microblogs is associated with higher returns.	Yes	Yes
<i>Message volume</i>		
Increased message volume in stock microblogging forums is associated with an increase in trading volume.	Yes	Yes
Increases in message volume in stock microblogging forums are associated with higher returns.	No	No
Increased message volume in stock microblogging forums is associated with higher volatility.	Yes	No
Increased message volume in stock microblogging forums reduces information asymmetry indicated by lower spreads.	No	No
<i>Agreement</i>		
Increased disagreement among stock microblogs is associated with higher volatility.	No	No
Increased disagreement among stock microblogs is associated with an increase in trading volume.	Yes	(Yes)*
<i>Analysis of information diffusion</i>		
Users who consistently provide high quality investment advice have more influence in the microblogging forum (through retweets, mentions, or followers).		Yes
High quality pieces of investment advice are spread more widely than low quality pieces of advice.		No

*Only one of the two tweet feature lags helped explain the market feature with statistical significance.

In particolare i segnali “BUY” sono accompagnati da rendimenti anomali nei giorni successivi, rendimenti che in media eccedono i costi di transazione dei maggiori broker (gli autori ipotizzano dei costi nell’are 0.15% -0.20%).

La stessa relazione però non vale per i segnali di tipo “SELL” poiché nonostante manifestino un accompagnamento molto simile non sono accompagnati da rendimenti anomali tali da superare i costi di transazione.

Gli autori inoltre non escludono la possibilità che con diverse strategie di investimento e con algoritmi di classificazione più sofisticati sia possibile utilizzare questo tipo di *trading strategy* in maniera biunivoca.

Nel complesso inoltre vale la pena notare che molte correlazioni fra *tweet* e mercato sono più forti rispetto a caratteristiche del mercato stesso che sono state studiate ed imputate per anni nel mondo accademico finanziario, inoltre mentre certe caratteristiche dei *tweet data* sembrano contenere informazioni importanti per la previsione del mercato attraverso i coefficienti standardizzati è possibile notare che è molto più il mercato ad influenzare i *tweet data* che vice versa.

Concludendo quindi lo studio svolto dal Dott. Spenger e la Proff.sa Welppe ha trovato risultati contrastanti rispetto a quelli svolti nel passato dimostrando che i social network, ed in particolar modo Twitter, possono essere degli strumenti utili ai fini del trading.

Rispetto a studi cardini degli anni precedenti (come ad esempio Antweiler e Frank del 2004), la quantità dei messaggi con consigli di vendita e acquisto sembra più bilanciata e coerente con l’effettiva mole delle transazioni; ma soprattutto l’analisi mostra un cambio di strategia negli investitori amatoriali. Se prima la grande massa di investitori non professionali seguiva una strategia basata sul *momentum* ora invece sembra essere la *contrarian strategy* quella più adottata. A differenza degli studi precedenti si è riscontrata evidenza empirica che la qualità dei messaggi sia più ricercata della quantità e che i social network possano essere sicuramente utilizzati per recepire in maniera efficiente gli aspetti chiave dei mutamenti nel sentiment del mercato ed essere utilizzati come *proxy* per i futuri studi di finanza comportamentale.

2.3 APPROCCIO INDICE

Presentazione Metodologia Bollen et al.

In contrapposizione allo studio presentato nel paragrafo precedente, nello stesso anno il Prof. Johan Bollen dell'Università dell'Indiana e la Dott.ssa Huina Mao in collaborazione con Xiao Zeng presentano un approccio alternativo all' utilizzo dei social network per lo studio della finanza comportamentale.

La domanda di base che si sono posti gli autori in questo caso è molto simile a quella dei precedenti autori citati. Sapendo che la finanza comportamentale ha evidenziato quanto le decisioni economiche individuali possano essere influenzate dalle emozioni, si chiedono se è possibile applicare questo risultato su larga scala sintetizzando un indicatore d'umore della collettività e confrontarlo con le variazioni nei rendimenti del mercato nella sua interezza. Per fare questo anche in questo caso gli autori utilizzeranno Twitter per determinare se è possibile individuare una connessione fra un indicatore dello stato d'animo (*sentiment*) collettivo e i movimenti dell'*Dow Jones Industrial Average* nel tempo.

La previsione del mercato azionario ha sempre attirato molto l'attenzione del mondo accademico e finanziario e non solo. La modellistica riguardo previsione dei corsi azionari erano basate in un primo momento sulla teoria dei *random walk* poi in seguito sulla teoria dei mercati efficienti. Secondo quest' ultima teoria, come già detto in precedenza, i prezzi riflettono e sono guidati dalle informazioni nuove che giungono sul mercato, vale a dire quindi che i prezzi attuali o passati sono poco utili ai fini di previsione. Le *news* di mercato sono imprevedibili per definizione, i prezzi di mercato quindi seguiranno gli impatti che le *news* avranno su di essi man mano che queste si presenteranno, ma dal momento che le *news* sono imprevedibili per definizione anche i prezzi di mercato saranno imprevedibili seguendo un modello *random walk* che non può essere previsto con più del 50% di precisione.

Numerosi studi hanno mostrato evidenze empiriche contro queste due teorie interconnesse, in particolare diversi autori (articoli riportati in bibliografia) mostrano che i prezzi non seguono un *random walk* e che in particolare a anche se in misura minima certi movimenti possano essere previsti. Questo processo ha portato ad affinare anche la

teoria dei mercati efficienti creando una tripartizione del livello di efficienza del mercato.

L'intuizione degli autori subentra proprio in quest' ultima fase del processo di creazione dei prezzi, come riportato nell' introduzione di questo testo secondo ricerche recenti le *news* restano imprevedibili anche se utilizzano i social network possono essere estratti indicatori utili per prevedere i vari cambiamenti economici e commerciali in relazione a queste ultime. Gli autori seguendo questa linea guida e cercano di applicare lo stesso principio ai mercati finanziari. Secondo gli autori le *news* svolgono un ruolo essenziale nello sviluppo dei prezzi borsistici ma lo stato d'animo o *sentiment* può svolgere un ruolo altrettanto importante.

Dalla ricerca in campo psicologico sappiamo che le emozioni svolgono un ruolo importante nel processo decisionale umano, la finanza comportamentale ha confermato che le decisioni finanziarie sono significativamente guidate da emozioni ed umore. È quindi ragionevole pensare che a livello collettivo, il *sentiment* può guidare o deviare i valori di borsa nel breve periodo.

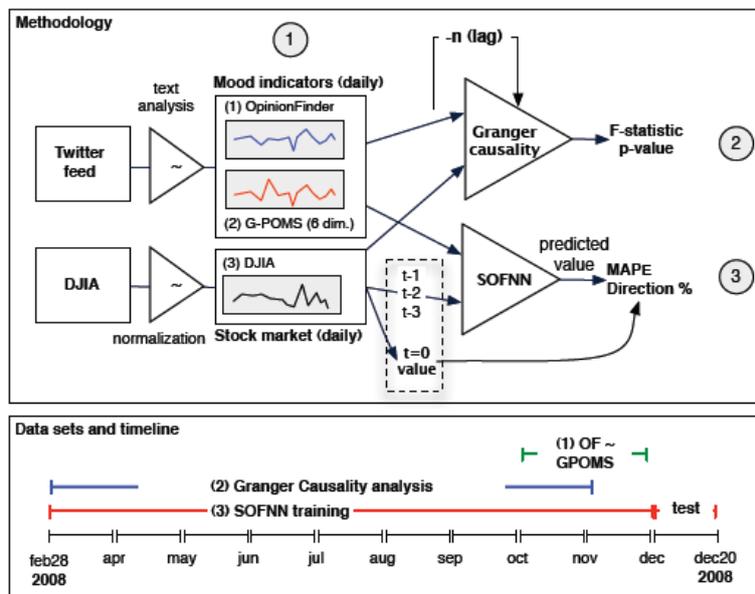
A differenza degli studi identificati in precedenza gli autori anche se condividono l'obiettivo finale ricercano un modello che abbia un legame facilmente identificabile con il mercato, sia affidabile e scalabile. Twitter come già riportato diverse volte in questo testo negli anni grazie alle sue qualità è diventato il protagonista e la fonte di molte di molte tecniche di analisi e monitoraggio *sentiment* che estraggono dai *tweets* una sintesi di tutto quello che è trattato in quel preciso momento nella *twittersphere*. Indicatori come *Pulse of Nation* sono monitorati negli USA costantemente per scoprire nuove tendenze ed interessi all' interno della nazione.

Nella loro ricerca gli autori cercheranno di capire se il *sentiment* collettivo, come espressione dell'aggregato di *tweets* giornalieri su larga scala, può essere usato per prevedere il mercato azionario, in particolare il *Dow Jones Industrial Average*. Il periodo di osservazione è stato dal 28 febbraio del 2009 al 19 dicembre dello stesso anno, per l'aggregazione e lo sviluppo dell'indicatore di *sentiment* hanno utilizzato due strumenti: il primo *OpinionFinder* che è un programma di analisi testuale dei *tweets* che restituisce un *output* bidimensionale sul *sentiment* giornaliero (positivo vs negativo) in modo da fornire una serie temporale dell'umore generale. Il secondo strumento, GPOMS, analizza in modo simile il contenuto del testo dei *tweets* per generare una serie temporale a

sei dimensioni su base giornaliera dello stato d'animo collettivo per fornire una visione più dettagliata delle variazioni del *sentiment* pubblico lungo una varietà di dimensioni diverse dalla semplice bidimensionalità di umore positivo o umore negativo. Le serie temporali sono sottoposte ad un'analisi statistica di tipo relazionale al *Dow Jones Industrial Average* per valutare la loro capacità di prevedere i cambiamenti dell'indice di mercato nel corso del tempo.

Il *dataset* raccolto nell' intervallo di tempo indicato racchiude 9,853,498 *tweets* in lingua inglese inseriti nel social network da circa 2.7 milioni di utenti. Ognuno dei *tweets* è stato catalogato e successivamente, una volta costituito l'intero *database*, gli autori hanno raggruppato i messaggi appartenenti alla stessa data filtrando quelli che erano facilmente identificabili come spam o pubblicità. Ai fini dell'analisi sono stati considerati solamente quei messaggi che includevano esplicite indicazioni che il messaggio contenuto fosse un'opinione propria dell'autore e non una citazione o altro (per esempio erano selezionati tutti i messaggi che contenevano stringhe del tipo "Credo che", "Mi sento di dire", "Devo dire", "Sono convinto" ecc.)

Figura 2.2



La metodologia può essere riassunta in tre fasi ed è illustrata dalla figura 2.2.

Il primo *step* consiste nell' analisi testuale dei *tweets* raggruppati in volumi giornalieri mediante i due strumenti presentati in precedenza, per ogni giorno quindi avremo due *output*, uno di tipo bidimensionale positivo / negativo ottenuto da *OpinionFinder* attraverso l'analisi del contenuto del testo e l'altro ottenuto da GPOMS che identifica attraverso l'analisi del testo sei modalità di espressione del contenuto testuale.

Da questo primo *step* quindi otteniamo 7 serie giornaliere che identificano l'"umore" della collettività, 6 ottenute tramite GPOMS ed una da *OpinionFinder*, ognuna delle quali ci evidenzierà le differenze nelle rispettive modalità di umore fra un giorno ed un altro, a queste viene aggiunta ovviamente la serie di riferimento finanziaria, il vettore dei prezzi di chiusura *Dow Jones Industria Average* (da ora in poi DJIA) per il medesimo arco temporale. Nel secondo *step* gli autori cercheranno di trovare la relazione esistente tra le modalità di output ottenute dagli analizzatori di testo ed i movimenti del DJIA mediante test di causalità di Granger nei quali verranno confrontati i valori del DJIA con gli output ottenuti dagli analizzatori testuale nei giorni precedenti.

Il terzo *step* invece gli autori utilizzano modello *Self Organizing Fuzzy Neural Network* per verificare l'ipotesi che l'accuratezza della stima di modelli di previsione DJIA può essere migliorato includendo misure di umore della collettività.

Fin da subito è possibile notare come a differenza del lavoro svolto da Spenger e Welpel nel paragrafo precedente questa volta i risultati ottenuti dall'analisi dei *tweets* sarà di per se di tipo aggregato senza la necessità di dover creare un indice per l'aggregazione, inoltre anche questo lavoro utilizza una mole di informazioni molto maggiore per sintetizzare gli indicatori anche su base giornaliera; elemento che nella metodologia del paragrafo precedente tendeva a essere troppo stringente poiché analizzando spesso i volumi di raccomandazioni di acquisto per i singoli titoli non superavano le 30 unità.

Così come per il paragrafo precedente, vista la rilevanza del lavoro nella disciplina dell'analisi del *sentiment* nei mercati finanziari, vale la pena spendere alcune pagine entrando nel dettaglio dell'analisi e soprattutto sulla validità statistica sottostante.

Innanzitutto ai fini della comprensione dei risultati finali risulta necessario approfondire le modalità di generazione delle serie temporali di *sentiment* attraverso i classificatori di opinione. *OpinionFinder* (OF) è un pacchetto *software* disponibile al pubblico per l'analisi del *sentiment* che può essere applicato per determinare la soggettività di ogni frase, vale a dire per identificare la parte emozionale (positiva o negativa) all'interno

delle frasi analizzate. OF è stato usato con successo per analizzare il contenuto emotivo di grandi volumi di *tweets* per determinare il rapporto di positività rispetto a quello di negatività dei *tweets* su un dato giorno. La serie storica risultante ha mostrato marcata correlazione con il *Consumer Confidence Index from Gallup* e la *Reuters / Università del Michigan Surveys*. Il sistema di identificazione è simile a quello della metodologia del paragrafo precedente, gli autori hanno utilizzato un insieme di parole chiave già ampiamente diffuso nell' utilizzo di questo classificatore per decretare l' emotività della frase analizzata. In particolare questo schema linguistico adottato conta 2718 contrassegnate come "positive" e 4912 parole contrassegnate come "negative", ogni parola inoltre presenta un attributo di tipo binario "Debole/Forte" a seconda che sia poco influente o caratterizzante nella frase. Un algoritmo determina per ciascun *tweets* se la stringa di messaggio sia classificabile come positiva o negativa a seconda del numero di parole chiave contenute ed il loro livello di caratterizzazione, etichettando quindi tale *tweet* come "positivo" o "negativo". Per ogni giorno quindi vengono calcolato il rapporto fra *tweets* con *sentiment* positivi e *tweets* con *sentiment* negativi.

Il grande difetto di questo semplice ma potente classificatore sta nella sua bidimensionalità, l' emotività umana è tutt' altro che bidimensionale poiché è facile immaginare che esistono sia in senso positivo che in senso negativi diversi livelli di intensità.

Per catturare la multidimensionalità dell' umore della collettività gli autori utilizzano oltre ad OF anche GPOMS che permette di misurare in 6 fasce diverse l' umore etichettate rispettivamente in *Calm, Alert, Sure, Vital, Kind, Happy*. I *cluster* per la distinzione della soggettività emotiva delle frasi derivano da uno strumento psicometrico ben rodato, il *Profile of Moods State*. Per rendere applicabile l' analisi dell' umore POMS a Twitter gli autori hanno ampliato il lessico dei termini da 72 a 964 utilizzando un' analisi delle parole che sono state digitate nel motore di ricerca Google nell' intero 2006. In particolare la validazione di questo schema lessico è il risultato di un' analisi di 2,5 miliardi di stringhe di ricerca effettuate Google composte da 4-grammi o 5-grammi, quindi in pratica circa un trilione di parole ottenute da siti web accessibili al pubblico.

Questa generalizzazione chiamata GPOMS ha permesso agli autori di catturare in maniera più precisa l' umore, mappando ogni *tweet* in base al suo livello POMS per poi essere diviso in base a dei *cluster* statistici in una delle 6 dimensioni di umore di *output*.

Infine per determinare una comparazione fra le serie temporali ottenute da OF e

GPOMS gli autori normalizzano i termini sotto forma di *z-score*, sulla base di media locale e deviazione standard all' interno di una finestra temporale di k giorni prima e dopo una particolare data. Per esempio lo *z-score* di una serie temporale X_t è data dalla formula 3 dove al denominatore troviamo la deviazione standard ed al numeratore troviamo la differenza dalla media calcolata per il periodo che va da $[t-k, t+k]$.

$$(3) \quad Z_{X_t} = \frac{X_t - \bar{x}(X_{t \pm k})}{\sigma(X_{t \pm k})}$$

La normalizzazione permette alle serie temporali di fluttuare intorno ad una media di valore 0 e di esprimere la *standard deviation* in misura unitaria, entrambe qualità estremamente apprezzabili ai fini della comparazione analitica e grafica. Dopo la sintesi degli indici gli autori come prima cosa testano la validità dei loro indicatori per prevedere l'umore generale della collettività. L'analisi solo in questo caso si focalizza su un sotto periodo di 3 mesi che va dal 5 di Ottobre al 5 di Dicembre, un periodo che include eventi socio culturali come le elezioni presidenziali e la festa del ringraziamento che dovrebbero avere un risvolto unidirezionale sull'umore della collettività, questi eventi possono essere usati per validare l'abilità degli indicatori generati da OF e GPOMS per la previsione dell'umore.

La figura 2.3 riporta la realizzazione delle serie temporali dei sette indicatori di umore normalizzati per il periodo interessato.

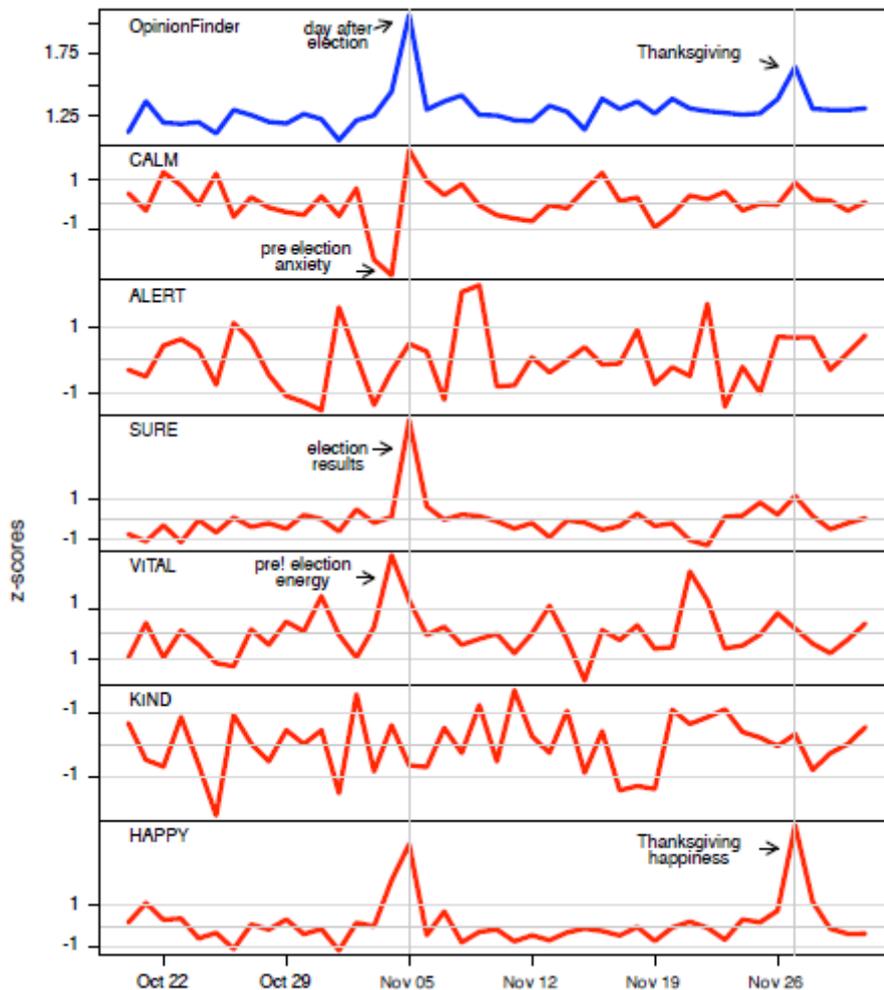
Il risultati ottenuti con il classificatore OF identificano in maniera evidente la risposta emotiva alle elezioni presidenziali del 4 novembre ed al ringraziamento il 27 novembre. In entrambi i casi la serie temporale porta degli shock, seppur brevi, di innalzamento del *sentiment* positivo per quelle date, dimostrano la validità palese del classificatore.

Per quanto riguarda il secondo strumento di classificazione, il GPOMS, le serie temporali risultanti rivelano una risposta più differenziata nei periodi riguardanti questi due eventi. Il giorno precedente alle elezioni presidenziali in particolare è caratterizzato da un crollo significativo nella serie che misura lo stato di *Calm* indicando un elevato livello di tensione in vista delle future elezioni. Il giorno delle elezioni è caratterizzato da un'inversione del medesimo indicatore, con un balzo verso l'alto degli indicatori *Vital*,

Happy e Kind. Il giorno dopo le elezioni gli indicatori *Calm, Sure, Vital, Happy* continuano ad attestarsi su livelli alti per poi tornare ad avere valori in linea con la media nei giorni successivi.

Nel periodo attorno all' altro evento di controllo, il Ringraziamento del 26 novembre, è possibile notare una risposta altrettanto differenziata degli indicatori, in particolare è riscontrabile un aumento sostenuto della modalità *Happy* in risposta ovviamente ad una contentezza generale dovuta alla festività, le altre modalità del GPOMS come vediamo subiscono variazioni di minore entità ed in ogni caso limitate al giorno stesso.

Figura 2.3



.Passando da un raffronto grafico ad uno quantitativo gli autori testano la correlazione fra le modalità di risposta di GPOMS e l'indicatore di OF attraverso delle regressioni multiple. Il modello di regressione è mostrato nella formula 4 dove i regressori X_i con $i=1,2,3,4,5,6$ rappresentano le modalità di output del classificatore GPOMS nell'ordine *Calm, Alert, Sure, Vital, Kind, Happy* ed Y_{OF} il valore dell'indice prodotto dal classificatore *OpinionFinder*.

$$(4) \quad Y_{OF} = \alpha + \sum_i^n \beta_i X_i + \epsilon_t$$

I risultati della regressione multipla sono riportati nella tabella sottostante. L'output ottenuto da OF è significativamente correlato con le modalità *Sure, Vital, Happy* ma non con *Calm, Alert e Kind*. Gli autori concludono questo primo aspetto della loro analisi ritenendo che anche se parte delle modalità espressa dal GPOMS tendono a comportarsi in maniera molto simile all'indice elaborato con OF (da notare la similitudine nella figura 2.3 della serie *Happy* ed OF), la multidimensionalità degli stati di umore di GPOMS fornisce una prospettiva sul *sentiment* della collettività che lo rende preferibile rispetto all'indicatore bidimensionale offerto da *OpinionFinder*.

Tabella 2.7

MULTIPLE REGRESSION RESULTS FOR OPINIONFINDER VS. 6 GPOMS
MOOD DIMENSIONS.

Parameters	Coeff.	Std.Err.	t	p
Calm (X_1)	1.731	1.348	1.284	0.20460
Alert (X_2)	0.199	2.319	0.086	0.932
Sure (X_3)	3.897	0.613	6.356	4.25e-08 ***
Vital (X_4)	1.763	0.595	2.965	0.004**
Kind (X_5)	1.687	1.377	1.226	0.226
Happy (X_6)	2.770	0.578	4.790	1.30e-05 **
Summary	Residual Std.Err	Adj.R ²	F _{6,55}	p
	0.078	0.683	22.93	2.382e-13

(p-value < 0.001: ***, p-value < 0.05: **, p-value < 0.1: *)

Dopo questa prima analisi di validazione delle modalità di espressione del *sentiment* della collettività gli autori si concentrano sull'obiettivo principale della loro ricerca, verificare se esiste una relazione fra le variazioni di *sentiment* e le fluttuazioni del DJIA. Per identificare questa relazione gli autori svolgono un test econometrico applicando la tecnica dell'analisi di causalità di Granger alle serie giornalieri ottenute da OF e GPOMS sulla serie storica delle fluttuazioni del DJIA calcolate con i prezzi di chiusura.

Il test di Granger testa il verso della relazione fra due variabili, il test si basa sull'assunzione logica che se una variabile X causa una variabile Y quindi un cambiamento in X deve avvenire sistematicamente prima di un cambiamento in Y . Ritardando di un *lag* predefinito la serie X sarà possibile testare la relazione con la serie Y . Si nota immediatamente che la serie X *laggata* mostra una correlazione con Y ma questa correlazione non ci indica la causalità.

L'analisi di causalità svolta dagli autori però non ha lo scopo di trovare il nesso causale fra le due variabili ma semplicemente per capire se una dei vettori della matrice X può essere utilizzato ai fini previsionali per prevedere le future variazioni di Y .

La serie delle variazioni de DJIA è ottenuta come differenza prima dei cambiamenti nel prezzo di chiusura del DJIA, in particolare $D_t = DJIA_t - DJIA_{t-1}$. In una prima analisi gli autori testano se le modalità di *sentiment* prodotte possono essere utilizzate per prevedere i cambiamenti nel DJIA confrontando la varianza espressa da due modelli lineari riportati nell'equazione 5 e 6. Il primo modello riportato nell'equazione 5 (L_1) usa solamente n *lag* della serie temporale dei DJIA ($(D_{t-1} \dots D_{t-n})$) per la previsione.

$$(5) \quad L_1 : D_t = \alpha + \sum_{i=1}^n \beta_i D_{t-i} + \epsilon_t$$

$$(6) \quad L_2 : D_t = \alpha + \sum_{i=1}^n \beta_i D_{t-i} + \sum_{i=1}^n \gamma_i X_{t-i} + \epsilon_t$$

Il secondo modello invece riportato nell'equazione 6 (L_2) inserisce oltre ai *lag* temporali del DJIA del modello precedente anche n *lag* tempora per GPOM e OF indicati in forma matriciale con X_t e quindi $X_{t-1} \dots X_{t-n}$.

Il test di casualità di Granger eseguito per entrambi i modelli è stato effettuato per i periodi fra il 28 febbraio ed il 3 novembre per evitare la distorsione negli indicatori di *sentiment* dovuta agli eventi pubblici citati in precedenza.

I risultati ottenuti dal test di Granger sono riassunti nella tabella 2.8. Il risultato principale riportato dalla tabella è senza dubbio la possibilità di rifiutare l'ipotesi nulla che le modalità di *sentiment* non causino le variazioni del DJIA per ognuno dei *lag* considerati con alti livelli di significatività. Dai risultati inoltre si evince che non tutte le modalità di *sentiment* sono utili ai fini previsionali. Il risultato largamente più elevato è quello ottenuto da *Calm* che da una significatività della relazione con *lag* da 2 a 6 giorni e *p-value*

minori del 5%, anche *Happy* risulta avere una relazione statisticamente significativa mentre le restanti 4 modalità del GPOMS non riscontrano alcun tipo di relazione

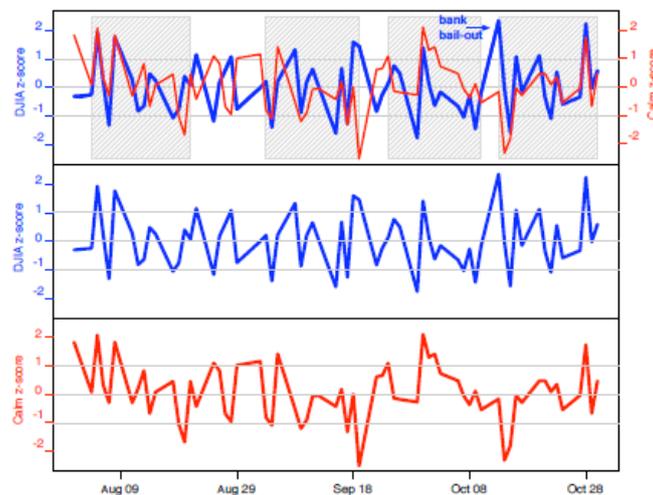
Tabella 2.8

Lag	OF	Calm	Alert	Sure	Vital	Kind	Happy
1 day	0.085*	0.272	0.952	0.648	0.120	0.848	0.388
2 days	0.268	0.013**	0.973	0.811	0.369	0.991	0.7061
3 days	0.436	0.022**	0.981	0.349	0.418	0.991	0.723
4 days	0.218	0.030**	0.998	0.415	0.475	0.989	0.750
5 days	0.300	0.036**	0.989	0.544	0.553	0.996	0.173
6 days	0.446	0.065*	0.996	0.691	0.682	0.994	0.081*
7 days	0.620	0.157	0.999	0.381	0.713	0.999	0.150

(p-value < 0.05: **, p-value < 0.1: *)

Il classificatore OF invece secondo il test di Granger non produce un indice di *sentiment* con relazione significativa se non per il *lag* di un solo giorno ed un livello di significatività basso del 10%. Per visualizzare la correlazione esistente fra le due variabili gli autori mostrano nella figura 2.4 la relazione grafica fra la serie storica normalizzata del DJIA e del vettore *Calm* normalizzato nella modalità di *lag* = 3, quindi in pratica il confronto è fra la realizzazione della performance borsistica D_t e l'indicatore di *Calm* ottenuta tre giorni prima ($Calm_{t-3}$) dall'analisi del *sentiment*.

Figura 2.4



Come è possibile vedere spesso le serie si sovrappongono ed in maniera molto frequente hanno variazioni nella stessa direzione. Cambiamenti nel valore di $Calm_{t-3}$ prevede una caduta o un aumento dell'indice DJIA di tre giorni dopo in maniera molto accurata.

Oltre all' elevato numero di volte in cui Cal_{t-3} riesce a tenere traccia del DJIA vale la pena notare quali sono i casi in cui esso non ci riesce. Il 13 ottobre per esempio fra i due indici c'è una discrepanza interessante, il DJIA ha un picco di tre deviazioni standard di rialzo mentre l'indice Cal_{t-3} resta relativamente piatto intorno la media fino al momento in cui ritorna a tracciare nuovamente le variazioni del DJIA. Questa anomalia può essere spiegata come risultato della manovra attuata dalla *Federal Reserve* nel giorno 13 ottobre per il salvataggio di una grande banca americana che ha inaspettatamente ed in maniera del tutto imprevedibile fatto schizzare il valore del DJIA. La deviazione fra i due valori in quella determinata data rafforza quindi la possibilità di modellare la previsione del mercato azionario con la *sentiment analysis* in quanto seppure questo evento effettivamente danneggia la *performance* di previsione era per definizione imprevedibile prima che si manifestasse, quindi sarebbe stato anomalo trovare *pattern* nella serie temporale di previsione.

Nonostante gli ottimi risultati ottenuti dall'analisi di causalità di Granger gli autori riconoscono l'esistenza di una connessione fra qualche modalità di *sentiment* il DJIA, riconoscono l'imposizione forte della relazione di tipo lineare dei modelli utilizzati.

La linearità delle relazioni sottostanti alle regressioni può essere un'ipotesi stringente soprattutto in vista del fatto che è molto più probabile che le relazioni fra le variabili di *sentiment* ed il mercato siano di tipo non lineare.

Per identificare la relazione non lineare fra le modalità di *sentiment* e le variazioni del DJIA, ma soprattutto però determinare la possibilità di utilizzare tali relazioni ai fini di previsione, gli autori costruiscono un modello *Self-organized Fuzzy Neural Network (SOFNN)*. In particolari utilizzano tale modello su due *data input* diversi per verificare poi l'output ottenuto. In particolare un primo modello SOFNN utilizza solamente le ultime tre osservazione dell'indice DJIA mentre il secondo modello utilizza sia le ultime osservazioni del modello precedente aggiungendo però anche delle combinazioni e permutazioni delle modalità di *sentiment* AF e GPOMS. Qualora le modalità di *sentiment* non risultino rilevanti ai fini della previsione si otterranno delle stime del DJIA non statisticamente diverse l'una dall'altra o con un secondo modello che addirittura produrrà una previsione peggiore del primo.

Gli autori hanno scelto un modello complesso come il SOFNN proprio perché le sue qualità sono state già apprezzate in passato per la decodifica di serie temporali con relazioni non lineari applicati ai mercati finanziari, in particolare il modello adottato è composto da una rete neurale a 5 strati capace di riorganizzarsi e aggiornarsi costantemente nel processo di apprendimento.

Per la previsione del valore DJIA al tempo t , gli autori scelgono, coerentemente con i risultati ottenuti nell'analisi lineare precedente, un *lag* massimo di tre giorni per le modalità di espressione del *sentiment* della collettività in quanto dalla tabella 2.8 si nota che passando da un *lag* di tre ad un *lag* di 4 giorni si perde significatività statistica nella relazione.

Figura 2.5



Inoltre tutti i valori vengono scalati linearmente $[0,1]$ per fare in modo che le variazioni di ogni variabile di input siano trattate uniformemente poiché appartenenti al medesimo range.

I modelli SOFNN per la previsione richiedono inoltre la stima di un discreto numero di parametri che può influenzare le performance del modello. Gli autori mantengono gli stessi parametri fra i due modelli per fare in modo che non ci siano delle performance di previsione diverse dovute ad una diversa calibrazione.

Il periodo di analisi è quello compreso dal 28 febbraio 2008 al 19 dicembre dello stesso anno, a sua volta tale lasso di tempo può essere scomposto in due sotto periodi: un periodo utile per la calibrazione del modello ai fini predittivi che va dal 28 febbraio al 28 novembre, mentre un periodo che va dal 1 dicembre al 19 dicembre scelto per testare l'effettiva abilità di previsione del modello. La scelta di questi due sotto periodi è stata dettata dalla marcata instabilità del DJIA nell'autunno 2008. Come mostra la figura 2.5

il primo sotto periodo di analisi è stato un periodo drammatico per l'indice poiché i corsi azionari hanno subito un sistematico calo di valore, mentre il secondo sotto periodo è caratterizzato da una volatilità minore di quello precedente e soprattutto dall' assenza di eventi socioculturali (elezioni presidenziali, ringraziamento).

Nonostante le informazioni già ottenute riguardo alle qualità previsionale delle modalità dall' analisi di causalità di Granger è possibile che una combinazione non lineare delle modalità possa essere utile ai fini della previsione del DJIA. Per esempio nell' analisi precedente *Happy* mostrava solo una relazione in presenza di *lag 6*, questo ci dice che può avere una relazione lineare con DJIA, ma allo stesso modo questa relazione non è detto che migliori la previsione in un modello SOFNN quando questa modalità è combinata con un'altra. Gli autori per rispondere a questo problema testano nel modello di predizione 7 combinazioni di variabili di input nel modello SOFNN.

La prima combinazione indicata con I_0 rappresenta il *set input* base utilizzato come riferimento anche nell'analisi precedente, quindi con i valori del DJIA al tempo t vengono prodotti utilizzando le osservazioni del medesimo indice al tempo $t-1$, $t-2$, $t-3$ che rispettando l'annotazione del testo è espresso come $DJIA_{t-3,2,1}$.

$$\begin{aligned} I_0 &= \{DJIA_{t-3,2,1}\} \\ I_1 &= \{DJIA_{t-3,2,1}, X_{1,t-3,2,1}\} \\ I_{1,2} &= \{DJIA_{t-3,2,1}, X_{1,t-3,2,1}, X_{2,t-3,2,1}\} \\ I_{1,3} &= \{DJIA_{t-3,2,1}, X_{1,t-3,2,1}, X_{3,t-3,2,1}\} \\ &\dots \end{aligned}$$

Rispettando l'annotazione $X_{1,t-3,2,1}$ rappresenta il valore della dimensione 1 (*Calm*) delle modalità di output GPOMS al tempo $t-3$, $t-2$, $t-1$. Seguendo la notazione è facile capire come le sette modalità di input non facciano altro che comprendere le combinazioni dei valori storici del DJIA e le modalità di output GPOMS al tempo $t-3$, $t-2$, $t-1$. Per esempio $I_{1,6}$ rappresenta il set di input che include per una certa data t le ultime tre osservazioni del DJIA, le ultime tre osservazioni della modalità 1 (*Calm*) e le ultime tre osservazioni della modalità 6 (*Happy*).

Inoltre per comparare le performance dei due selettori gli autori inseriscono I_{OF} come combinazione per una certa data t del *set* informativo delle ultime tre osservazione del DJIA e l'indice di *sentiment* positivo/negativo elaborato da *OpinionFinder*.

Tabella 2.9

DJIA DAILY PREDICTION USING SOFNN

Evaluation	I_{OF}	I_0	I_1	$I_{1,2}$	$I_{1,3}$	$I_{1,4}$	$I_{1,5}$	$I_{1,6}$
MAPE (%)	1.95	1.94	1.83	2.03	2.13	2.05	1.85	1.79*
Direction (%)	73.3	73.3	86.7*	60.0	46.7	60.0	73.3	80.0

La performance di previsione del modello è misurata attraverso la *Mean Absolute Percentage Error* (MAPE) e l'accuratezza di direzione durante il periodo di test che va come detto in precedenza dal 1 dicembre al 19 dicembre 2008. I risultati sono sintetizzati nella tabella 2.9.

I risultati sono sbalorditivi. Innanzitutto si nota immediatamente che l'aggiunta dell'indice OF al set informativo non migliora l'abilità di previsione rispetto alle sole ultime tre osservazioni del DJIA utilizzate come variabili di controllo (I_0), ulteriore conferma dei risultati ottenuti precedentemente col test si causalità.

In secondo luogo aggiungendo l'la modalità *Calm* si ha il risultato migliore in termini di accuratezza di previsione delle direzioni di variazione, l'incremento di performance del modello è evidente con un balzo del 13.3% in termini di accuratezza di previsione ed una riduzione di 12 *basis point* in termini di MAPE. Altro risultato interessante è dato dal fatto che introducendo nel *set* informativo le modalità 2,3,4, si ha un notevole peggioramento della performance di previsione; questo sta a significare che, come in precedenza nell' analisi di casualità di Granger, *Alert*, *Sure* e *Vital* non sono modalità di espressione del *sentiment* utile ai fini di previsione del DJIA. Infine è notevole come $I_{1,6}$ combinazione del set base più le modalità *Clam* e *Happy* mostri l'errore di previsione medio atteso (MAPE) più piccolo osservato (1.79%) ed un'accuratezza nella previsione della direzione alta (80%) ma comunque inferiore a quella ottenibile utilizzando il solo set informativo I_1 . Infine per testare la validità statistica del modello SOFNN ed in particolare della previsione con accuratezza all' 86.7% gli autori effettuano un test per verificare quale sia la probabilità che tale risultato sia puramente casuale e non dovuto alla bontà del modello. Essendo una previsione di accuratezza sulla direzione il tipo di variazione osservata ha una distribuzione binomiale, quindi in ogni ripetizione c'è una proba-

bilità del 50% di poter prevedere in maniera giusta la direzione presa. Secondo la distribuzione binomiale la probabilità di indovinare in maniera casuale nell' 86.7% dei casi su 15 giorni di borsa (sono solamente i 15 i giorni di apertura di borsa nel periodo di test) è pari allo 0.32%. Applicando questa percentuale all' intero periodo di applicazione del modello (28 febbraio -19 dicembre) otteniamo un valore atteso di 10.9 giorni le cui previsioni sono generate in maniera completamente casuale. Calcolando la probabilità che questi 10.9 giorni di casualità cadano proprio all' interno del periodo di test si ottiene un valore pari a $1-(1-0.0032)^{10.9} = 3.43\%$. Gli autori concludono quindi che la bontà del modello ai fini di previsione non è dovuta né al caso né alla particolare scelta del periodo di verifica.

Infine come ulteriore conferma gli autori eseguono un *f-test* sull' effetto lineare di entrambe le modalità di GPOMS *Calm* (X_1) e *Happy* (X_6) confrontando quindi il modello completo F ed il modello ridotto R che include solo *Calm* (X_1) come riportato di seguito:

$$\mathbf{F}: D_t = \alpha + \sum_{i=1}^n \beta_i D_{t-i} + \sum_{i=1}^3 \gamma_i X_{1,t-i} + \sum_{i=1}^3 \gamma_i X_{6,t-i} + \epsilon_t$$

$$\mathbf{R}: D_t = \alpha + \sum_{i=1}^n \beta_i D_{t-i} + \sum_{i=1}^3 \gamma_i X_{1,t-i} + \epsilon_t$$

Gli autori trovano un *p-value* di 0.66 e una *F-statistic* di 0.53 indicando che introducendo il set informativo X_6 all' interno del modello si ha un peggioramento della bontà di previsione rispetto al modello col il solo set informativo composto dalle n osservazioni passate del DJIA e le ultime tre della modalità *Calm*. Quest' ultimo test rappresenta un' ulteriore conferma al risultato ottenuto dal modello SOFNN, ovvero che la massima accuratezza di previsione in termini di MOPE ottenuto da un set informativo che includeva X_1 e X_6 è dovuta ad una relazione di tipo NON lineare fra queste due modalità di espressione del *sentiment* e le passate osservazione dell' indice di mercato.

Il lavoro di del Prof. Johan Bollen e la Dott.ssa Huina Mao è stato con ogni probabilità il pezzo di ricerca accademica sulla *sentiment analysis* che ha suscitato maggiormente l' attenzione dei media e del mondo della finanza in generale. Come vedremo nel prossimo paragrafo il lavoro non è esente da errori o limitazioni, ma allo stesso modo l' im-

patto mediatico che ha avuto accompagnato dal supporto e dalle raccomandazioni di autorevoli accademici con David Crandell e Michael Neslon e il supporto del premio Nobel per l' Economia Daniel Kahneman, ha innescato una scintilla nell' industria finanziaria inducendo ad una gemmazione di sezioni di ricerca specializzate in questa disciplina e dopo alcuni anni casi di hedge fund specifici.

2.4 CRITICA

Principali errori e critiche e limitazioni dei modelli presentati

I lavori presentati nei paragrafi precedenti identificato, a mio parere, la migliore espressione dei due filoni della ricerca scientifica sulla *sentiment analysis* del primo decennio del XXI secolo.

Dopo il 2010 c'è stata una gemmazione sia nell' industria che nel campo accademico di lavori di ricerca sull' argomento dovuta principalmente alle prove statistiche della validità della analisi del *sentiment* riportate come abbiamo avuto modo di vedere anche nei due lavori analizzati e soprattutto per la crescita esponenziale dei social network accompagnata dallo sviluppo tecnologico che nel giro di pochissimi anni ha stravolto le abitudini quotidiane integrandosi nella vita del cittadino medio praticamente in ogni parte del pianeta.

I due lavori presentati nei paragrafi precedenti, seppure siano stati punti di riferimento per lo sviluppo successivo della disciplina, non erano mancanti di errori e soprattutto limitazioni che vale la pena osservare prima di procedere nella prossima parte di questo testo, dove vedremo una applicazione empirica basata principalmente su un modello che cerca di rispondere proprio alle mancanze dei due modelli presentati fin ora.

Partendo dalla metodologia presentata dal Dott. Spenger e la Prof.sa Welpe principalmente oltre ad altri problemi la grande limitazione del volume di raccomandazioni disponibili. In particolare così come gli autori hanno riportato nel loro *paper*, segmentando la raccolta di messaggi su Twitter a quelli riguardanti uno specifico titolo azionario o prodotto finanziario i volumi che si riescono ad estrarre sono irrisori se paragonati all' effettiva mole di investitori *retail* presenti sul territorio. Proprio per questo motivo gli autori concentrano la loro ricerca sul mercato americano e principalmente sul solo

paniere di titoli più scambiati. Nonostante la scelta oculata dei titoli con il maggior numero di messaggi specifici scambiati nel paese col sistema borsistico più sviluppato del pianeta il volume dei *tweets* non garantisce la possibilità di poter svolgere valutazioni *intra-day* del *sentiment*. Secondo gli autori i social network, e Twitter in particolare, si prestano particolarmente bene a questo tipo di analisi poiché danno la possibilità di valutare il flusso delle informazioni in tempo reale e continuo non forzarle in tempo discreto come nel loro studio. La forzatura di raggruppamento è senza ombra di dubbio un altro grande limite della analisi. A differenza di quello che avviene nei mercati dei cambi la borsa valori è soggetta ad orari e giorni di apertura e chiusura in tutto il mondo, il flusso di informazioni fornite dai social network invece, sebbene con netti cali, non risente degli stessi ritmi. Per ovviare a questo problema gli autori raggruppano i messaggi *postati* su Twitter dopo le 4.00 PM con i messaggi appartenenti alla giornata successiva di trading, così come nei giorni di chiusura di borsa i messaggi vengono accorpati fino al prossimo giorno di apertura. Il problema principale di questo metodo sta nel fatto che spesso i messaggi seppur consentendo stringhe riconducibili dall'analizzatore (parlerò in seguito di questo) linguistico a raccomandazioni di tipo *BUY* o di *SELL* queste in una percentuale non irrilevante possono essere non messaggi "forward looking" ma piuttosto *report* di performance di propri investimenti nei giorni precedenti. Un altro filone di critica può essere rivolto al fatto che gli autori abbiano deciso di raccogliere dati per i singoli titoli, sintetizzarli in indicatori indice e confrontarli con misure di mercato aggregato per definizione quali per esempio la volatilità. Tralasciando la possibilità di non poter applicare tale studio alla maggior parte dei mercati visto la scarsa volume di *social data* disponibili negli altri mercati (ma anche all'interno dello stesso mercato USA) un altro aspetto chiave della limitatezza di questa analisi sta nella metodologia di selezione dei dati. La *Weka Machine* utilizzata per la ripartizione dei messaggi analizzati fra raccomandazioni di vendita o di acquisto è basata su un campione di circa 2500 inserimenti linguistici. Sebbene possano sembrare tante le frasi inserite va considerata l'estrema complessità della costruzione linguistica che può facilmente indurre una persona in difficoltà quindi ovviamente anche un classificatore. L'algoritmo Naïve Bayesian utilizzato ha un tasso di accuratezza stimato dell'81.5%. Questo vuol dire che su una selezione di 5 messaggi ci si attende che 1 messaggio sia stato classificato in maniera errata, va da sé che sui circa 250,000 messaggi classificati

circa 50,000 saranno stati classificati in maniera errata. È facile poter pensare quindi che l'indice alla base di tutta la metodologia per determinate osservazioni sia completamente distorto, in quanto un margine di errore pari quasi al 20% associato a poche osservazioni giornaliere con buone probabilità può distorcere completamente l'indice di *Bullishness* in un determinato giorno.

A differenza del lavoro criticato precedentemente che ha avuto impatto prevalentemente nel mondo accademico il lavoro del Professor J. Bollen e la Dott.ssa Mao ha avuto un impatto mediatico rilevante e diversi anni è stato il punto di riferimento al di fuori del mondo accademico quando si parlava di *sentiment analysis*.

Rispetto all' approccio Sprenger-Welpe questo lavoro è più coerente in termini di confronto di variabili, gli autori mettono a confronto e cercano di prevedere una variabile aggregata come il *Dow Jones Industrial Average* con degli indicatori di *sentiment* ottenuti da campioni aggregati della totalità dei messaggi *postati* su Twitter, inoltre come diretta conseguenza gli autori ampliano di molto il loro bacino di *tweets* rispetto allo studio discusso in precedenza, questo permette agli autori di rendere più robusti i risultati ottenuti agli occhi dei lettori.

Se da un lato il lavoro presenta diversi vantaggi dall' altro è stato pesantemente criticato dalla stampa specializza e da professionisti del settore. A differenza della limitatezza riscontrata nel lavoro sopra discusso in questo caso viene a mancare la chiarezza ed in alcuni casi la coerenza dei risultati.

I principali problemi riscontrati nel lavoro di Bollen sono principalmente tre:

- I risultati raggiunti sono frutto di dati molto difficili da reperire, poiché come diremo anche in seguito in maniera più dettagliata non sono dati pubblici indicizzati, errori metodologici a volte anche gravi commessi dagli autori possono comprometterne la validità del modello.
- Il fatto che i gli autori scelgano a priori l'intervallo temporale su cui testare il proprio modello (intervallo composto solamente da 15 giorni di osservazioni) a posteriori può indicare che questa scelta non sia casuale e che quel breve intervallo sia semplicemente l'arco temporale dove era possibile ottenere la migliore previsione possibile.

- La scelta della tanto pubblicizzata 87.6% di accuratezza di previsione farebbe di questo modello la scoperta più grande nel campo della finanza. Per quanto i due autori siano delle facce nuove all' interno del mondo finanziario è improbabile pensare che questi abbiano condiviso immediatamente un modello così profittevole con il mondo intero.

Partendo dalla prima critica, che è in comune con qualsiasi altro lavoro che utilizzi Twitter per la *sentiment analysis*, non esistono banche dati che raccolgono indicizzano l'intero flusso di *tweet* presenti sulla piattaforma per un anno intero. Questa irreperibilità dei dati verrà discussa ampiamente nell' ultimo capitolo del testo, in ogni caso però lo studio di Bollen presenta un'effettiva difficoltà di confutazione dovuta principalmente al fatto il *dataset* utilizzato non è stato reso pubblico ed è di difficile reperimento per altre vie. Sebbene questa impossibilità di reperire i dati ci lascia con il dubbio, gli autori commettono dei gravi errori tecnici di difficile contestazione.

Nei risultati del test di causalità di Granger riportati nella tabella 2.8 gli autori non praticano una correzione statistica essenziale in questi casi, ovvero il la correzione per molteplicità di verifica delle ipotesi.

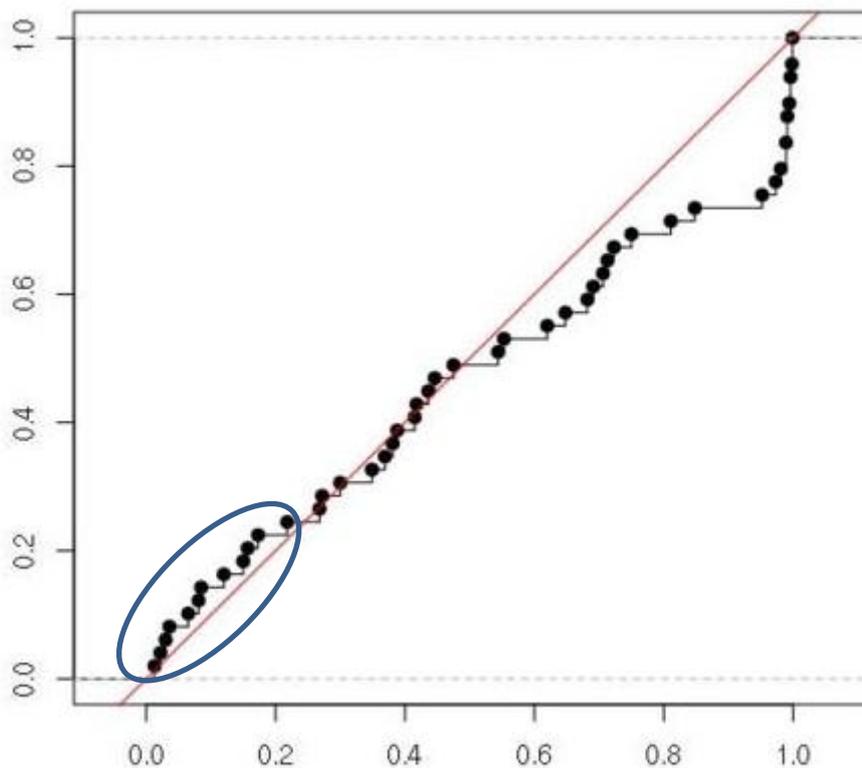
Il *p-value* come sappiamo rappresenta la probabilità di ottenere un risultato più estremo o più diverso da quello osservato posta come vera l'ipotesi nulla. In questo tipo di test però il *p-value* per poter essere ricondotto alla potenza del singolo test e non distorto dalla molteplicità dei test deve essere corretto usando la correzione di Bonferroni.

Nel test di causalità di Granger adottato dagli autori l'ipotesi nulla, che gli autori vogliono rifiutare, è che gli indicatori estratti da Twitter mediante i classificatori non abbiano relazione con i movimenti del *Dow Jones*, ovvero che siano effettivamente indipendenti.

Nell' esecuzione del test gli autori conducono in maniera simultanea 49 test di ipotesi, una volta ottenuti i *p-value* attribuiscono significatività statistica a tutti quelli inferiori ad una certa soglia con la classica denotazione in grassetto e * nel caso del 10%, ** per il 5% e *** per l'1%. Il grande errore commesso dagli autori sta nel fatto di prendere tali *p-value* non come estrazione di una distribuzione ma come valori definitivi. In caso

di verifica di ipotesi nulle ci si aspetta che i *p-value* risultati dal test siano uniformemente distribuiti, questo sta a significare che rieseguendo il test altre volte il risultato ottenuto in termini di *p-value* non sia altrettanto soddisfacente.

Al fine di comprendere quanto detto, riporto la distribuzione empirica dei *p-value* ottenuti dal test ipotesi riportati nella tabella 2.7. Sotto l'ipotesi della veridicità dell'ipotesi nulla dovrebbe essere uniforme, quindi in grafico distribuzione cumulato dovrebbe essere semplicemente la diagonale che congiunge gli estremi dell'intervallo ($y=x$), allo stesso modo è di facile intuizione capire che se come dicono gli autori l'ipotesi nulla può essere rigettata con alti livelli di significatività, dovrebbe esserci anche graficamente una distribuzione dei 49 *p-value* sostanzialmente differente da una distribuzione uniforme che nel grafico sottostante è riportato dalla linea rossa.



La distribuzione empirica ottenuta dagli autori è data riportando i 49 *p-value* della tabella 2.8. Se il test mostrasse casualità nelle variabili i *p-value* dovrebbero essere nella maggior parte dei casi estremamente a sinistra della linea rossa e non nei dintorni se non proprio sopra. Inoltre effettuando la correzione di Bonferroni per test multipli nessuno dei *p-value* della tabella risultano statisticamente significativi ad un livello del 10%. La conclusione da trarre non è di casualità fra le modalità di *sentiment* utilizzate ed i *lag*

nella analisi del *Dow Jones*, ma che esiste come si evince dal grafico qualche indicatore di *umore* che guarda nel verso giusto di previsione anche se questo non può essere accettato con livelli di significatività accettabili (graficamente visibili all' interno del cerchio blu).

La terza, e forse più pesante critica al modello, è rivolta alla parte previsionale e il tasso di accuratezza riportato.

Il primo grande errore sta proprio al numero (che risulta errato) ed all'enfasi riportata nella accuratezza della previsione. Come non è molto evidenziato nel *paper* questa accuratezza è stata testata sulla base di 15 giorni di negoziazione. Una accuratezza del 87.6% non è realizzabile in termini di variabili bimodali come quelle direzionali. L'ipotesi più plausibile è che il modello abbia prodotto stime direzionali corrette in 13 giorni su 15 con una accuratezza del 86.7% che con un errore di trascrizione è diventata 87.6%. Inoltre essendo l'intervallo di risposta composto solamente da 15 osservazioni la possibilità di ottenere un errore in più o in meno avrebbe fatto variare del 7% circa l'accuratezza del modello. Avendo un campione così piccolo l'errore standard per questo stimatore di accuratezza corrisponderebbe al 9% circa. È semplice a questo punto, anche solamente prendendo le cifre riportate dalla tabella 2.8, notare che non è improbabile che la stima dell'86.7% ottenuta sia dovuta semplicemente al caso poiché anche sottoponendo il coefficiente ad un banale test di ipotesi con significatività al 5% non si possa rifiutare l'ipotesi che tale coefficiente sia statisticamente diverso da quello ottenuto dal medesimo modello (73%) senza l'inclusione delle modalità di *sentiment*. Inoltre l'accuratezza ottenuta sembra essere poco coerente con il test di Granger.

Per testare questo confronto è stata svolta una simulazione Monte Carlo, nel dettaglio sono stati presi i rendimenti del DJIA del medesimo intervallo indicato nel *paper* (dal 28 febbraio al 3 novembre) e raffrontati con una serie storica generata in maniera del tutto casuale che può assumere i valori di 1 o -1 in base ad una estrazione casuale che assegna +1 (-1) se il giorno successivo borsistico il DJIA è salito (sceso) con probabilità del 86.7%, una volta ottenute le due serie queste vengono sottoposte ad un test di casualità di Granger con *lag* e viene riportato il *p-value*. Dopo 200 ripetizioni il massimo valore osservato è stato nell'ordine di $2e-05$. Comparare i *p-value* di modelli diversi è statisticamente sbagliato, ma qui sorge una grande contraddizione che è difficile da spiegare. Secondo la tabella 2.8 il *p-value* più basso ottenuto è pari circa al 0.013 in risposta

ad un modello Rete Neurale che può catturare le relazioni non lineari oltre che quelle lineari fra le variabili, come è possibile (se l'accuratezza di previsione è vera) che i risultati ottenuti dal test di Granger che catturano solo gli effetti lineari delle medesime variabili diano un *p-value* (per la precisione, il *p value* più alto ottenuto su 200 simulazioni) 650 volte inferiore. Infine al fine di utilizzare questa presunta relazione di casualità fra gli indicatori di *sentiment* tracciati da Twitter ed il DJIA il solo dato di una accuratezza del 86.7% non basta come ampiamente evidenziato nel *paper* gli autori non indagano sulla misura in cui il DJIA possa variare a seguito di una variazione nel *sentiment* ma solamente nella direzione, questo vuol dire che anche prevedendo in maniera corretta la direzione di variazione non si ha nessuna indicazione se l'effettiva variazione sarà in valore assoluto uguale ad 3 punti base o al 5%.

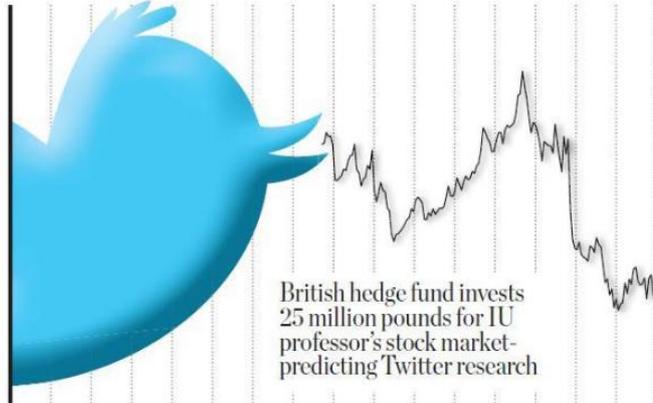
Inoltre (senza considerare che il risultato in evidenza è solamente quello migliore delle 8 strategie testate e non il valore medio ottenuto) per il livello di accuratezza del 86.7% è previsto un errore medio atteso di previsione del' 1.83%, questo valore confrontato con la deviazione standard giornaliera di un ETF del *Dow Jones Industrial* per il medesimo periodo che è pari al 2.51% sminuisce in maniera sostanziale tutto il fascino di questo modello strutturato.

Nonostante il lavoro di Bollen, Mao e Zeng non sia sgombro da errori e limitazioni il suo impatto mediatico è stato globale, i maggiori media come The Telegraph, The Daily Mail, USA Today, The Atlantic, Wired Magazine, Time Magazine, CNBC, CNN e svariati blog ne hanno parlato per molto tempo.

L'industria finanziaria ha rivolto sempre più attenzione verso la *sentiment analysis*.

Nel giro di pochi mesi sono stati tantissimi gli *hedge fund* che hanno investito sulla ricerca in questo campo. Si sono istituiti negli anni successi fondi di investimento, come Derwent Capital Market a Londra, basati prettamente su strategie di market-timing ancorate alla *sentiment analysis* attraverso l'utilizzo dei social media, ed in particolare Twitter.

A TRENDING #INVESTMENT



Capitolo 3

Metodologia

3.1 EVOLUZIONE DELLA DISCIPLINA

Dalla accuratezza al rendimento atteso

Gli articoli presentati nel precedente capitolo risalgono al 2010, dopo cinque anni gli autori del secondo lavoro presentato hanno elaborato una nuova metodologia, molto più pulita e di facile comprensione, che come vedremo è stata confermata ed apprezzata anche dalle autorità pubbliche.

Dal 2010 ad oggi l'attenzione sulla *sentiment analysis* nel settore finanziario ha avuto un andamento altalenante.

Come in molti casi nella storia, spesso il mercato è pronto ad investire in nuove tecnologie prima ancora che queste siano testate e rodiate. Sebbene molte istituzioni finanziarie abbiano fatto e facciano tutt'ora utilizzo della *sentiment analysis* come strumento di supporto per le strategie *market timing*, ci sono stati casi di hedge fund dedicati che hanno pubblicizzato in lungo ed in largo la loro innovazione nel campo della *sentiment analysis*, e una volta che si sono ritirati dal mercato hanno gettato del fango sull'intera disciplina.

Titolo dell' articolo sul Financial Times alla chiusura di Derwent Capital Market.

May 24, 2012 9:02 pm

Last tweet for Derwent's Absolute Return

By James Mackintosh, Investment Editor

Il caso Derwent Capital Market è stato forse il più eclatante. Il fondo costituito agli inizi del 2012 è stato chiuso a maggio dello stesso anno dal fondatore Paul Hawtin. La notizia della prematura chiusura di un fondo altamente specializzato in questa disciplina

ha fatto barcollare l'intera area di ricerca della *sentimen analysis* attraverso i *social network*. Non sono mancati titoli ironici di testate giornalistiche importanti e con la stessa intensità con cui si era acceso, l'interesse è andato calando per il periodo successivo. La notizia che però non ha avuto (ovviamente) lo stesso impatto è quella che Dewert Capital Market non ha chiuso battenti perché la strategia adottata non funzionava; nei circa due mesi in cui è stato aperto ha registrato un rendimento del 1.86% in linea con gli altri fondi in quel momento; ma per un cambio strategico di business passando da operatore diretto nei mercati a consulente per gli altri fondi senza operare in maniera diretta.

Negli anni successivi si sono create numerose start-up e servizi di consulenza di successo specializzati nella *sentimen analysis*, iSentium e Datameer per clienti *corporate* e StockTwits per *retail*, che hanno individuato forse il modello di business più efficace per questo tipo di ricerca.

La svolta successiva della disciplina si è avuta nel luglio del 2015 quando la Banca Centrale Europea ha condiviso nella sua raccolta di paper statistici la nuova versione del lavoro di Bollen et al. intitolata "Quantifying the effect of bullishness on international financial markets". Il prossimo paragrafo analizzerà dettagliatamente questo lavoro. Così come nel 2010, i risultati di questo lavoro hanno fatto in poco tempo comparsa sulle maggiori riviste specializzate e non solo, in seguito gli occhi attenti dell'industria finanziaria sulla quindi sulla *sentiment analysis*.

Articolo del 15 luglio 2015 di Lorcan Roche Kelly, Bloomberg.



Bloomberg Markets Markets Tech Pursuits Politics Opinion Businessweek

If Twitter Is Bullish, Maybe You Should Be Too

New study shows Twitter gets it first, and gets it right

3.2 BULLISHNESS MODEL

Presentazione del modello

In nuovo modello introdotto nel 2015 è in parte stato elaborato agli stessi autori del celebre lavoro *Twitter mood predict the stock market* del 2010. È però della Dott.ssa Huina Mao ricercatrice presso facoltà di *Computer Science* dell'Indiana University l'apporto maggiore coadiuvato dal Dott. Scott Counts, Ricercatore in Microsoft e dal Prof. Johan Bollen, ordinario di *Informatics and Computing* all' Indiana University.

Il lavoro intitolato *Quantifying the effect of bullishness on international financial markets* ha il vantaggio essere stato prodotto riconoscendo gli errori e le limitazioni dei precedenti studi sulla *sentiment analysis* effettuati dagli autori. Il modello introdotto come vedremo si pone a metà fra i due approcci analizzati nei capitoli precedenti, ottenendo risultati altrettanto validi.

Questo nuovo modello si posiziona, come i modelli presentati precedentemente, come applicazioni della teoria della finanza comportamentale, poiché come nei lavori precedenti, si cerca di dimostrare un legame che andrebbe contro la teoria dei mercati efficienti (EMH).

Secondo la forma forte teoria dei mercati efficienti, come già detto in precedenza, gli investitori operano come agenti razionali e i prezzi delle azioni quotate riflettono sempre tutte le informazioni nuove, passate e nascoste. La finanza comportamentale mette alla prova la EMH enfatizzando il ruolo essenziale svolto nelle interazioni economiche e nel comportamento degli investitori svolto dai fattori emotivi che possono far deviare gli individui dal modello di agenti razionali.

La finanza comportamentale è basata su due principali pilastri: il *sentiment* degli investitori, inteso come la componente emotiva che devia temporaneamente il comportamento dell'individuo da quello che avrebbe un investitore razionale, e i limiti all'arbitraggio; ovvero l'inesistenza opportunità di arbitraggio *risk free* che possano scommettere contro il comportamento razionale degli investitori. Queste due assunzioni permettono alla massa degli investitori di influenzare i prezzi di mercato con il loro

comportamento vista la limitatezza all'arbitraggio imposta dal grado di rischio non nullo per gli investitori sofisticati.

Come ho già riportato ampiamente nelle pagine precedenti di questo testo, nell'ultimo decennio l'avvento inesorabile dei social network e dei BIG DATA ci hanno permesso di avere accesso ad informazioni su una scala globale prima impensabili, sia come grado di dettaglio, sia come costi che come frequenza di aggiornamento.

Nei lavori analizzati fin ora abbiamo visto che gli autori utilizzano diversi approcci per la classificazione dei messaggi. Uno dei metodi più ampiamente apprezzati dalla letteratura è stato quello proposto dal Antweiler e Frank nel 2004 basato su un classificatore di tipo Naive Bayesian è stato utilizzato anche nella primo lavoro proposto nel capitolo due (Sprenger Welpel 2010) . Un altro metodo di classificazione ampiamente apprezzato è quello proposto da Tetlock nel 2007 che a differenza del metodo precedente non richiede una prima fase di preparazione del modello, ma utilizza una lista di parole negative determinata dall' Harvard Psychological Dictionary per la determinazione di un indice di pessimismo. La stessa lista viene poi corretta in un secondo momento da Loughran e McDonald nel 2011, poiché secondo gli autori molte delle parole con contenuto negativo generale sono applicate in senso forviante nei contesti finanziari, sviluppano quindi un elenco di 2337 parole che applicate allo stesso studio di Tetlock sovraperforma l'accuratezza di previsione ottenuta in precedenza. La classificazione del senso o dell'intenzione dietro i messaggi inseriti nei classificatori è un compito arduo, nonostante algoritmi sempre più complessi e la possibilità di un miglioramento dinamico durante l'analisi . La complessità delle strutture comunicazionali unane nonostante la continua implementazione di algoritmi e metodi *dictionary-based* è ancora lontana dall'essere perfetta. Secondo alcuni studi degli ultimi anni ((Das e Chen, 2007; Oh e Sheng, 2011; Pang e Lee, 2008) nell'utilizzo di programmi di classificazione è ritenuta accettabile una accuratezza di classificazione del 60-70%, va da sé che quindi una classificazione errata attesa del 30-40% può solo distorcere i risultati raggiunti mediante lo sviluppo di tale dati.

In contrasto con gli indicatori ottenuti in modo computazionale, le indagini "tradizionali" di *sentiment* degli investitori sono già ampiamente accettate del contesto dei dati finanziari. Ad esempio, *il Daily Sentiment Index (DSI)* e la relazione

settimanale *US Advisory Sentiment* report rilasciato dalla *Investors Intelligence* (II) coinvolge due indagini ben note di *sentiment* degli investitori.

La *Investor Intelligence* dal 1963 ha indagato e classificato opinioni dei lettori su bollettini di mercato, assegnando a questi tre categorie, vale a dire "bullish", "bearish" o "correzione" (neutrale). In poche parole, valutano se il sentimento è rialzista o ribassista sulla base delle parole precise usate da persone quando si risponde a domande specifiche.

Questa metodologia ha certamente il vantaggio di essere univoca e precisa, ma la raccolta di informazioni di questo tipo è molto costosa da condurre. Inoltre secondo diversi autori (Da et al, 2010; Singer, 2002) quello che può sembrare un punto di forza potrebbe effettivamente essere una debolezza; quando alle persone è chiesto esplicitamente il loro parere, esiste un'ampia varietà di fattori e pregiudizi sociali che possono distorcere la veridicità delle risposte date.

L'obiettivo degli autori è definire un indicatore che beneficia del vantaggio della tradizionale richiesta esplicita dei sondaggi (bullish vs bearish), ma che possa utilizzare l'ampia quantità di dati disponibile sui social network senza passare però attraverso un classificatore linguistico. L'indicatore proposto fa derivare il *sentiment* direttamente dai messaggi *tweettati* o ricercati dagli utenti, ma a differenza delle analisi precedenti, cerca di individuare pattern che identifichino direttamente l'indirizzo finanziario del messaggio *postato* o ricercato. Per ridurre al massimo l'ambiguità del *sentiment* sintetizzato, gli autori concentrano la loro attenzione sui soli due termini "Bullish" e Bearish, scelti perché ampiamente utilizzati in contesti finanziari e raramente in altre situazioni. Questi termini quindi riescono a massimizzare intrinsecamente l'accuratezza di identificazione nell'individuazione del *sentiment* rialzista (*bullish*) o ribassista (*bearish*) del messaggio inserito.

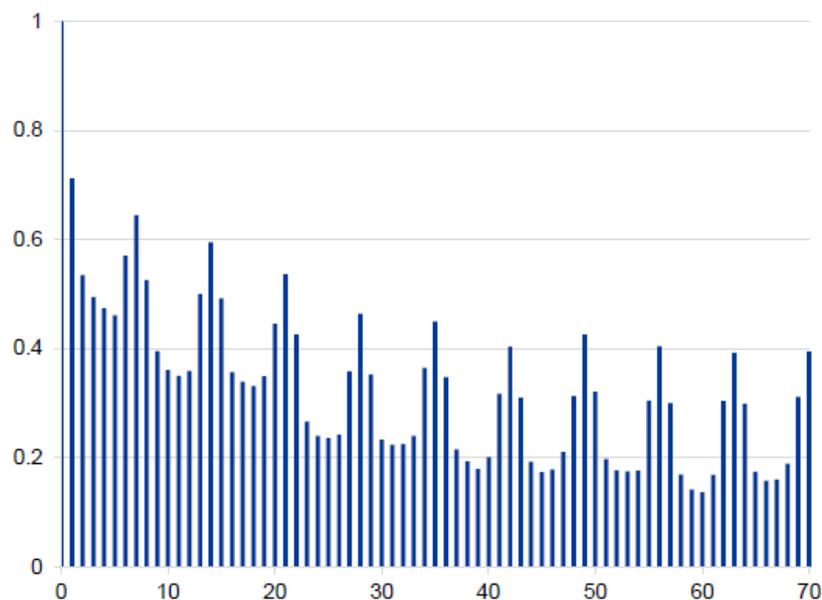
In questo lavoro gli autori utilizzano principalmente due fonti di dati: Twitter, raccogliendo i volumi di messaggi su base giornaliera e Google raccogliendo le stringhe di ricerca in volumi settimanali.

Dai dati viene estratto un indice, definito *Bullishness Index* che compara i volumi di messaggi rialzisti e ribassisti. Gli indici ottenuti vengono esaminati e comparati con i principali indicatori di *sentiment* ed infine vengono confrontati con gli indici azionari dei paesi finanziariamente più sviluppati per carpirne la possibile relazione.

Twitter Bullishness e Google Bullishness

In questa prima analisi gli autori si concentrano solamente sui dati ottenibili da Twitter. Come detto in precedenza l'analisi dei messaggi questa volta è molto più semplice e lineare. Individuando a priori parole chiave come *bullish* e *bearish* è possibile semplicemente identificare i *tweet* che li contengono, saltando tutta la fase di classificazione mostrata nei lavori precedenti. Gli autori quindi definiscono un *tweet* come rialzista se contiene la parola *bullish* e ribassista se contengono la parola *bearish*. Lo studio comprende i *tweet* che vanno da gennaio 2010 al dicembre 2012, avendo quindi un campione di circa 310,000 *tweet* rialzisti o ribassisti. La Grafico 3.1 mostra il grado di autocorrelazione del volume dei messaggi.

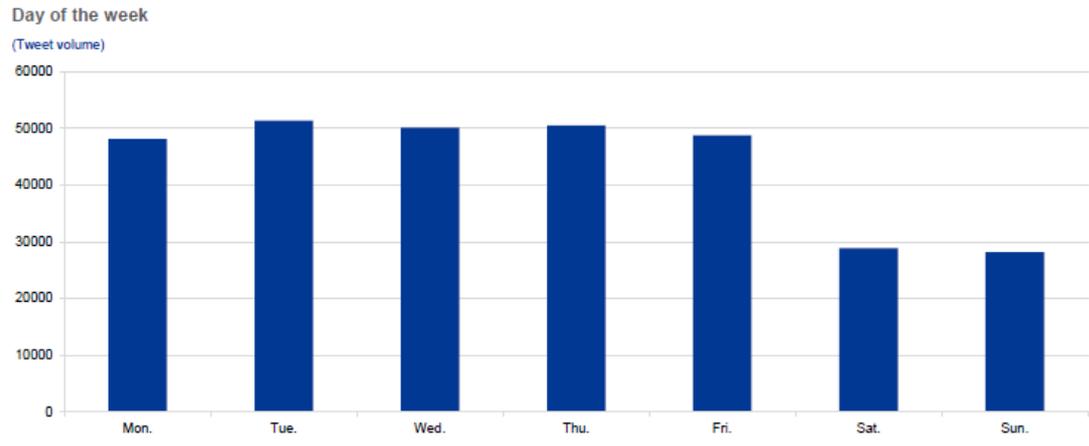
Grafico 3.1



Il grafico mostra sull'asse delle ordinate il coefficiente di autocorrelazione fra il volume di *tweet* al tempo t e lo stesso volume a ad un certo tempo $t-lag$, mentre sul grafico delle ascisse è espresso il numero di *lag*. Anche solo graficamente è semplice notare il forte grado di autocorrelazione settimanale che ci indica una differenza statistica significativa fra i volumi all'interno dei singoli giorni della settimana. La serie di *tweets* esibisce quindi una serie *patterns*, i giorni infrasettimanali da lunedì al venerdì

presentano volumi maggiori rispetto ai volumi osservabili nel fine settimana. Il grafico 3.2 riassume questa differenza riportando i volumi cumulati dei *tweets* sia rialzisti che ribassisti ottenuti nei diversi giorni della settimana.

Grafico 3.2



Dal grafico si evince una assoluta differenza in media per i giorni in cui i principali mercati finanziari sono aperti a quando questi sono chiusi, tale differenziazione è in linea ,oltre che col primo studio presentato in questo testo. Il numero di *tweets* rialzisti rispetto al totale raccolto è di circa il 69,4%, questo in linea con i risultati ottenuti da Oh e Sheng, è dovuto principalmente al fatto che esiste una distorsione da ottimismo negli investitori online riscontrata anche nei principali altri studi sull'argomento.

Una volta ottenuti i volumi giornalieri questi sono stati aggregati secondo l'ormai nota equazione di Antelwier e Frank del 2004 che attualmente, a 12 anni delle sua presentazione, risulta essere grazie alle sue qualità matematiche la misura più performante in termini di approssimazione di *sentiment* aggregati. La formula 7 riportata sotto non è altro che una riapplicazione della formula 1 riportata nel secondo capitolo. In questo caso però B_t e R_t indicano rispettivamente i volumi di *tweet* rialzisti e ribassisti ottenuti al tempo t .

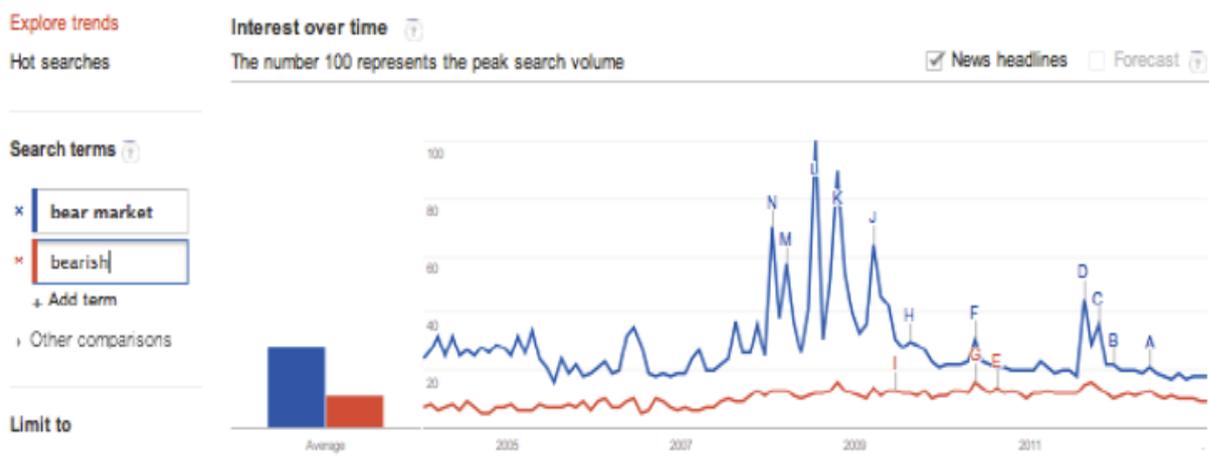
$$(7) \quad T_t^B = \ln \left(\frac{1 + \|B_t\|}{1 + \|R_t\|} \right)$$

In modo del tutto simile a T_t^B gli autori definiscono G_w^B che indica l'indice di Google Bullishness dai volumi settimanali delle *queries* di ricerca come mostra la formula 8 dove B_w indica il volume di *queries* settimanali rialziste e R_w le *queries* settimanali ribassiste.

$$(8) \quad G_w^B = \ln \left(\frac{1 + \|B_w\|}{1 + \|R_w\|} \right)$$

Il metodo di acquisizione dei dati in questo caso è stato leggermente diverso poiché a differenza dell'aggettivo *bearish* o *bullish* che può indicare l'intento del messaggio è insensato che un utente possa ricercare su Google il monogramma dell'aggettivo senza un riferimento. Con un suggerimento di *Google Hot Trend*, gli autori hanno rimpiazzato i termini *bearish* e *bullish* con "bear market" e "bull market". Questo cambiamento permette, come indicato dalla figura 3.1, di avere un netto miglioramento della reattività settimanale per il campione analizzato.

Figura 3.1



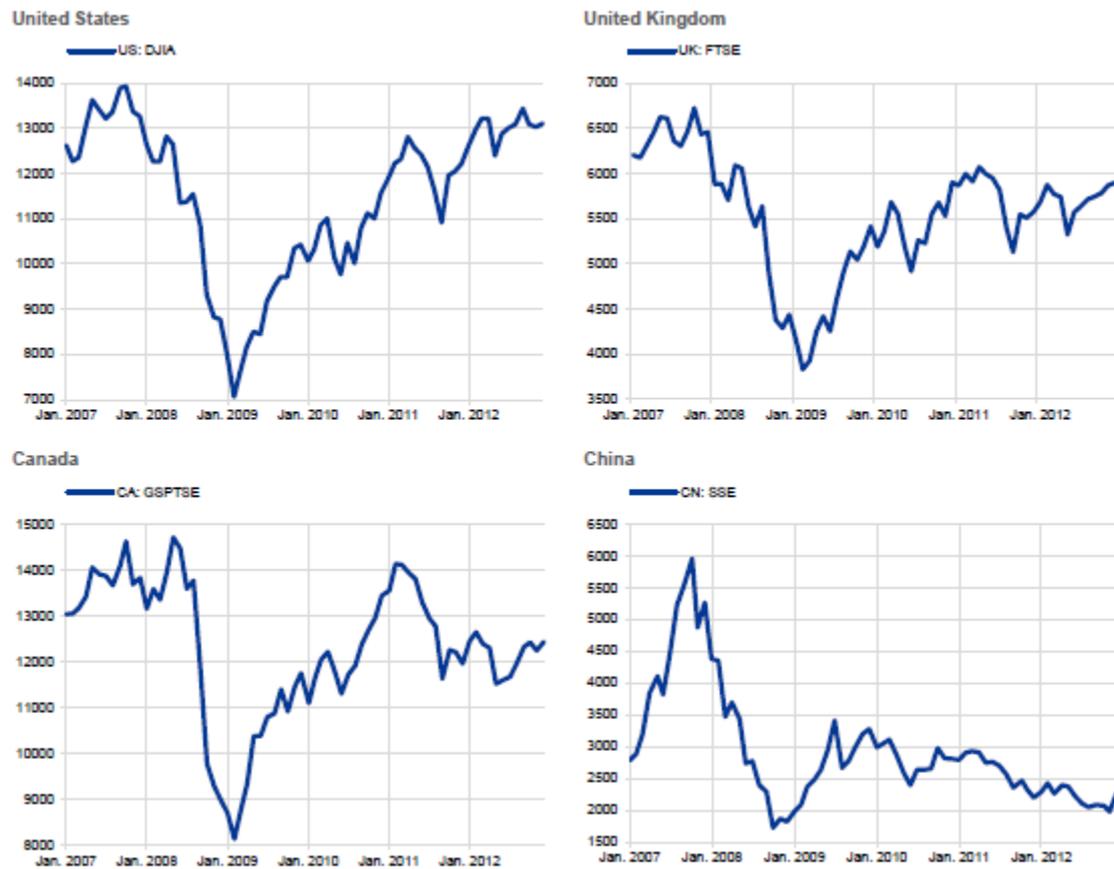
Come in ogni altro lavoro presentato in questo testo, gli indici di *sentiment* verranno confrontati con delle variabili di mercato per trovarne una relazione. I lavori presentati fin ora prendevano come misura di riferimento il *Dow Jones Industrial Average*, che come è noto si tratta di un indice *price weighted*, che sebbene non sia la *proxy* più adatta per mostrare l'andamento del mercato borsistico è senz'altro azzeccato per rispecchiare i mutamenti nel *sentiment* degli investitori. Gli autori al fine di aumentare la robustezza dei risultati introducono oltre al DJIA altri indici azionari di altre nazioni

ino modo da poter confrontare i risultati con gli indici rappresentativi delle nazioni finanziariamente parlando più sviluppate. I paesi sono stati selezionati in base alla capitalizzazione di mercato, in accordo con le informazioni riportate dalla World Bank, sono Stati Uniti d' America, Canada, Regno Unito e Cina. I primi tre paesi condividono un ampio radicamento di social network come Twitter e del motore di ricerca Google nella loro cultura oltre la lingua ufficiale anglofona, la Cina invece nonostante sia a tutti gli effetti uno dei paesi finanziariamente più sviluppati a livello globale, presenta una regolamentazione molto severa in termini di accesso ai contenuti internet oltre che una componente culturale molto diversa dalle altre tre nazioni in cluse nel campione, questo dovrebbe darle gli attributi necessari per essere considerata una variabile di controllo sul quale la metodologia non dovrebbe dare risultati positivi, in modo da irrobustire ulteriormente i risultati ottenuti.

Ogni nazione è rappresentata da un indice di mercato, per gli Stati Uniti abbiamo come già detto il *Dow Jones Industrial Average* che è un indice *price weighted* su 30 fra i titoli di borsa di società americane a più alta capitalizzazione di mercato. Per il Regno Unito il FTSE100, un indice *market weighted* che sintetizza i cambiamenti di capitalizzazione di mercato dei principali 100 titoli quotati alla *London Stock Exchange*. Per il Canada invece è stato selezionato il S&P/TSE, anche esso *market weighted*, che ha come componenti le principali aziende quotate alla *Toronto Stock Exchange*. Infine per la Cina è stato inserito il SSE Composite Index, anche quest' ultimo di tipo *market weighted*, ha come componenti le principali 50 società quotate alla *Shanghai Stock Exchange*.

Le profonde differenze fra i tre indici di paesi anglofoni ed il SSE Composite è marcata in una semplice rappresentazione delle serie storiche mensili degli ultimi anni. Nella figura 3.2 è possibile notare il comportamento a tratti diametralmente opposto dell' indice cinese rispetto agli altri, indicando ancora una volta l' estrema diversità non solo in termini di accesso alle piattaforme virtuali, social network, e motori di ricerca ma anche in termini di comportamento degli investitori e struttura di mercato.

Figura 3.2



Risultati

Le domande a cui gli autori cercano di dare una risposta sono essenzialmente tre:

1. Esiste un legame fra i due indici di *bullishness*? Sebbene uno sia generato dai post inseriti in un social network ed aggiornato quotidianamente e l'altro sia generato dal volume delle stringhe inserite nel motore di ricerca ed aggiornato su base settimanale entrambi sono un'espressione della attività online della compagine di investitori globali quindi dovrebbero esibire a grandi linee lo stesso *sentiment*.
2. Poiché l'indicatore Twitter Bullishness è considerabile un'espressione in tempo reale del *sentiment* del pubblico, aggiornandosi in maniera repen-

tina ai cambiamenti e le *news*, la sua relazione con i movimenti di mercato deve esse riscontrata in tempo reale o con qualche *lag* rispetto a quest'ultimo?

3. L' indicatore Google Bullishness settimanale che tipo di relazione ha con i rendimenti settimanali dei mercati? Esiste una relazione simultanea o *laggata*?

Partendo dalla prima domanda posta gli autori comparano gli indici di Bullishness ottenuti da Twitter e da Google per scoprire se estinte qualche tipo di relazione fra loro, e se eventualmente fosse riscontrata, scoprire il verso della causalità che li lega.

La prima distinzione che va fatta fra i due indici sta nella natura dei dati. I dati ottenibili da Twitter sono a flusso continuo, possono quindi essere divisi e raggruppati come si vuole, a differenza il volume delle *queries* di ricerca accessibili al pubblico sono disponibili solo sotto forma di volumi settimanali normalizzati. Per poter comparare i due indici, gli autori convertono su base settimanale l'indice T_t^B calcolando la media settimanale dell'indice per le 156 settimane di osservazione che vanno dal 9 gennaio 2010 al 29 dicembre 2012 e chiamano tale medie T_w^B .

Già una semplice regressione OLS fra le due variabili non *laggate* si trova il primo segnale di una relazione lineare statisticamente significativa fra i due indici ($r= 0.27$ $p= 0.0007$).

Per scendere più nel dettaglio della relazione gli autori eseguono una analisi di tipo VAR (Vector *Auto Regression*) che riesce a trovare l'effettiva relazione in entrambe le direzioni di causalità fra le variabili utilizzando dei *lag*. Si tratta essenzialmente di un modello statistico lineare che cattura le interdipendenze tra serie temporali multivariate ed è ampiamente utilizzato per convalidare e quantificare la prevedibilità degli indicatori finanziari. Il modello VAR utilizzato in questo caso è equivalente a prova di causalità serie vengono sottoposte ad un test Dickey-Fuller.

$$(9) \quad \Delta G_w^B = \alpha + \sum_{i=1}^4 \beta_i \Delta G_{w-i}^B + \sum_{i=1}^4 \chi_i T_{w-i}^B + \epsilon_w$$

Questo tipo di test serve per testare la presenza di un trend all' interno della variabile analizzato, la presenza di un trend indica un problema di non stazionarietà nel tempo che a sua volta rende la variabile non adatta al modello VAR, il problema però può essere risolto praticando una differenza prima che rende stazionaria la variabile. I risultati del test Dickey-Fuller confermano la non stazionarietà della serie Google Bullishness mentre la serie Twitter Bullishness risulta stazionaria con un livello di confidenza del 90%. Ai fini di non distorcere i risultati dell' analisi VAR gli autori utilizzano la differenza di primo ordine della variabile Google Bullishness chiamandola ΔG_w^B , inoltre tutte le variabili vengono normalizzate in modo da fornire risultati di più semplice interpretazione. La tabella 3.1 riassume i risultati ottenuti riportano i coefficienti stimanti ed i rispetti *p-value*. Essendo normalizzati i coefficienti misurano l' impatto dell' aumento di una *standard deviation* nella variabile indipendente sul cambiamento nel Google Bullishness Index durante la settimana *w*. Inoltre gli autori riscontrano che ε_w soddisfa le assunzioni di indipendenza, omoschedasticità e normalità.

Tabella 3.1

Bullishness	Coefficient	p-value
ΔG_{w-1}^B	-0.54	<< 0.01***
ΔG_{w-2}^B	-0.30	0.001***
ΔG_{w-3}^B	-0.21	0.02**
ΔG_{w-4}^B	0.009	0.91
T_{w-1}^B	0.18	0.03**
T_{w-2}^B	0.09	0.30
T_{w-3}^B	0.20	0.03**
T_{w-4}^B	0.10	0.20

$p \leq 0.001$: ***, $p \leq 0.05$: **, $p \leq 0.1$: *
Adjusted $R^2=0.23$, $F=6.69$ on df (8, 142), $p \leq 0.01$

Dalla Tabella 3.1 possiamo osservare che l'indice TBI ha un' influenza positiva statisticamente significativa sui cambiamenti della Google Bullishness nella settimana successiva. Un altro interessante risultato sta nel fatto che la relazione che lega la differenza

dei primi tre ordini ai cambiamenti dello stesso Google Bullishness è negativa e statisticamente significativa. In accordo con un'importante ricerca di Shapiro del 2001 nel campo della psicomotricità, è possibile motivare questa relazione negativa con la difficoltà dell'attenzione umana di focalizzare la propria attenzione su un solo argomento per un periodo prolungato, seguendo questa logica è possibile che gli utenti possano variare nel giro di 2-3 settimane il tema di ricerca su Google per poi ritornare a concentrarsi non appena qualcosa gli ricordi del suo interesse precedente. Infine sempre in riferimento alla stessa tabella si può notare che soltanto il 23% della varianza di ΔG_w^B può essere spiegata (R^2 corretto), indicando ancora una volta come nei lavori precedenti l'imprevedibilità di queste variabili e la difficoltà di poter fare previsioni utilizzando le sole variabili di input sintetizzate da Twitter o Google.

Infine gli autori riportano che invertendo la regressione gli autori non trovano alcuna relazione statisticamente significativa dell'effetto di ΔG_w^B su T_w^B . In poche parole, l'indice Twitter Bullishness può essere usato per prevedere la Google Bullishness ma non viceversa. Gli autori rimandano ai posteri questo tipo di ricerca che esce fuori dall'argomento del *paper*. Nonostante l'evidenza mostrerebbe una relazione di tipo sequenziale fra le due espressioni di *sentiment*, per esempio è possibile che un investitore recepisca una *news* attraverso i media (che hanno un'influenza quasi istantanea sui social network ed in particolare Twitter) e solo in un secondo momento approfondisca la news facendo ricerche attraverso i motori di ricerca.

Posto quindi che è il *Twitter Bullishness Index* a far variare il *Google Bullishness Index* gli autori passano ora alla seconda domanda di ricerca, trovare la relazione fra il *Twitter Bullishness Index* e le variabili di mercato.

Lo strumento applicato per trovare questa relazione è nuovamente un modello VAR per esaminare se l'indice sintetizzato da Twitter ha potere previsionale sui vari indici di mercato attraverso osservazioni con *lag* che va da 1 a 5 giorni. In prima battuta gli autori limitano l'area della loro analisi ai soli indici del mercato americano, in quanto oltre ad essere il paese con il sistema finanziario più sviluppato al mondo, l'America è anche il paese con il numero di utenti attivi su Twitter maggiore.

Come è possibile vedere nella tabella 3.2 gli autori utilizzano indici di mercato con un numero di componenti crescente in modo da poter osservare se l'espressione del *sentiment* sintetizzata attraverso i *tweets* riguarda solo la compagine di titolo più scambiati o

è ramificata anche in quelli con volumi minori. Per fare questo il T_t^B è stato confrontato con il *Dow Jones Industrial* che include come già detto 30 componenti, lo *Standard&Poors 500* che ne racchiude 500 e il *Russell 3000* suddiviso nelle due sue macro parti: il *Russell 1000* che misura i cambiamenti di valore delle 1000 compagnie con capitalizzazione maggiore (*large cap*) quotate nel mercato americano ed il *Russell 2000* che invece misura i cambiamenti nelle 2000 *small cap* più trattate.

Il volume di *tweets* giornalieri ai fini dell'indice è calcolato utilizzando il fuso orario di Greenwich (GMT + 0) dalle ore 00:00:00 alle ore 23:59:59 dello stesso giorno, invece i rendimenti degli indici azionari corrispondono alle variazioni che vanno dalle ore locali 09:30:00 alle 15:59:59 dello stesso giorno di negoziazione che corrisponde rimpostando l'orario al GTM+0 all'intervallo che va dalle 15:30:00 alle 22:00:00 dello stesso giorno. Infine in termini analitici il rendimento degli indici è stato calcolato con un regime capitalizzazione continua mediante la classica formula $R_t = \log(S_t^{CLOSE}) - \log(S_t^{OPEN})$. Il rendimento R_t è calcolato in questa maniera riflette le informazioni presenti solamente all'interno delle ore racchiuse nell'intervallo considerato del *Twitter Bullishness Index*, questo aiuta rispetto alla classica differenza logaritmica fra prezzi di chiusura di due giorni seguenti a fare in modo che il rendimento possa beneficiare attraverso i cambiamenti fra prezzo di chiusura del giorno precedente e di apertura del giorno successivo delle informazioni disponibili nell'intervallo di tempo post negoziazione poiché queste sono racchiuse nell'intervallo temporale con cui si è misurata la *Twitter Bullishness*.

$$(10) \quad R_t = \alpha + \sum_{i=1}^5 \beta_i R_{t-i} + \sum_{i=1}^5 \chi_i T_{t-1}^B + \sum_{i=1}^5 \delta_i Vol_{t-i} + \phi_i Exog_t + \epsilon_t$$

Gli autori per valutare il contributo di ogni nuovo predittore *Twitter Bullishness* inseriscono con variabili di controllo proditori già esistenti ampiamente accolti dal mondo accademico. In linea con precedenti ricerche, le variabili endogene del modello includono il prezzo delle azioni, così come il volume degli scambi, al fine di tener conto degli effetti di liquidità. Il volume degli scambi di registro viene *de-trendizzato* per garantire la stazionarietà. La terza variabile endogena è l'indice T_t^B . Le variabili esogene sono il *VIX* e il *Daily Sentiment Index*, e variabili *dummy* inserite per catturare i vari effetti da

calendario, compreso l'effetto gennaio. Tutte le variabili nel modello vengono ritardate fino a cinque giorni, che corrisponde a una settimana di trading. Il modello di regressione è riportato nell'equazione 10.

I risultati della regressione sono riportati nella tabella 3.2. La tabella riporta i coefficienti di regressione ed i rispettivi *p-value*. Come nel caso precedente i valori sono normalizzati quindi ogni coefficiente indica l'impatto dell'incremento di una deviazione standard avrebbe sui rendimenti giornalieri in misura di punti base. Gli errori di regressione ε_t del modello mostrano un quasi totale assenza di autocorrelazione ed una distribuzione molto vicina a quella normale.

Tabella 3.2

Bullishness	DJIA		SP500		Russell1000		Russell2000	
	Coeff.	p-value	Coeff.	p-value	Coeff.	p-value	Coeff.	p-value
Lag								
1	12.56	0.01***	10.98	0.05**	10.72	0.05**	11.02	0.05**
2	2.27	0.67	2.61	0.65	2.46	0.67	2.66	0.65
3	2.18	0.69	3.69	0.53	4.037	0.48	4.58	0.43
4	-7.81	0.15	-8.10	0.16	-9.99	0.08*	-10.28	0.08*
5	-1.12	0.80	-1.28	0.79	-1.35	0.77	-1.37	0.78

La prima colonna della tabella 2 elenca la stima di regressione per *il Dow Jones Industrial Average*. Osserviamo che un aumento di una deviazione standard in T_t^B nel giorno $t-1$ è seguito da un aumento di 12,56 *basis point* nei rendimenti DJIA il giorno seguente. Questo impatto è statisticamente significativo al livello di confidenza del 99%. Inoltre, rispetto alla media incondizionata dei rendimenti giornalieri DJIA durante il periodo di campionamento, cioè 3,46 punti base, un valore di 12,56 punti base è anche economicamente significativo.

Gli autori inoltre mettono a confronto T_t^B con il *Daily Sentiment Index* al di individuare le correlazioni contemporanee e paragonare l'effetto predittivo nei confronti rendimenti azionari. Il coefficiente di correlazione di Pearson tra la il DSI e T_t^B ($\gamma = 0,30$, $p < 0,01$) è statisticamente significativa, ma non è elevato. Gli autori evidenziano anche che un aumento di una deviazione standard nel DSI è seguito da soli 2,26 punti base di aumento nel giorno successivo di negoziazioni del DJIA, che non è economicamente significativo e solo marginalmente statisticamente significativo. Questo risultato suggerisce il *Twitter Bullishness Index* come una nuova *proxy* per il *sentiment* degli investitori,

e che sebbene sia correlato al DSI già esistente può essere un predittore di accurato dei cambiamenti nel mercato azionario di quanto lo non lo siano gli indicatori basati sui sondaggi.

La seconda, terza e quarta colonna della tabella riportano invece i coefficienti normalizzati stimanti con la medesima modalità per gli alti indici, anche in questo caso i coefficienti mostrano una relazione con $lag = 1$ statisticamente significativa fra l'indice di *Twitter Bullishness* ed i rendimenti dei vari indici di mercati con coefficienti positivi e molto simili. Inoltre per il Russell1000 ed il Russell2000 si è riscontrata una relazione negativa con $lag = 4$ statisticamente significativa per una soglia del 10%, questa evidenza indicherebbe che sebbene un incremento nella T_t^B sia seguito da un aumento nel giorno successivo di negoziazioni per gli indici Russell 1000 e Russell 2000, l'effetto è quasi completamente riassorbito dopo 4 giorni di negoziazione. Questo temporaneo scostamento con correzione è coerente con quanto teorizzato nel campo della finanza comportamentale nel 1990 da De Long, ovvero che l'irrazionalità economica dei *noise trader* può causare una deviazione del prezzo di un *asset* dal valore dei fondamentali in maniera solo temporanea, poiché in osservazioni più lunghe il prezzo tende sempre a ritornare al suo valore fondamentale.

Come annunciato precedentemente gli autori espandono l'area di applicazione della loro analisi al mercato inglese, canadese e cinese. Come già detto in precedenza il mercato cinese è inserito come variabile di controllo, gli autori si aspettano che poiché Twitter è ampiamente utilizzato in USA, UK e Canada mentre praticamente vietato in Cina la stima della relazione risultante dalla analisi di tipo VAR fra il *Twitter Bullishness Index* e questi due sotto campioni porti risultati sostanzialmente molto diversi.

A differenza del mercato americano, i mercati inglesi cadesi e cinesi hanno pochi *patterns* effettivamente dimostrati, per questo motivo oltre che l'impossibilità di reperire indicatori di *sentiment* basati sui sondaggi, gli autori propendono per un modello di tipo ridotto rispetto al precedente. L'equazione 11 riporta il generico modello utilizzato.

$$(11) \quad R_t = \alpha + \sum_{i=1}^5 \beta_i R_{t-i} + \sum_{i=1}^5 \chi_i T_{t-i}^B + \epsilon_t$$

per esaminare il potere previsionale dell'indice Twitter Bullishness ed i diversi indici di mercato. La tabella 3.3 riporta i coefficienti di regressioni per i vari indici, in particolare

così come nella tabella precedente, il coefficiente essendo normalizzato misura l'impatto che dell'incremento di una deviazione standard in T_t^B sui singoli rendimenti giornalieri degli indici misurati in punti base.

Tabella 3.3

Lag	US: DJIA		UK: FTSE		CA: GSPTSE		China: SSE	
	Coeff.	p-value	Coeff.	p-value	Coeff.	p-value	Coeff.	p-value
1	13.18	0.01*	17.98	0.0005**	14.08	0.001**	8.73	0.09*
2	1.30	0.81	-10.39	0.06*	-5.28	0.28	-3.18	0.571
3	3.03	0.57	11.11	0.04*	8.16	0.08	6.78	0.224
4	-8.79	0.10	-9.85	0.07*	-11.35	0.01*	-2.91	0.601
5	-2.31	0.60	-3.54	0.46	-1.799	0.64	-1.60	0.757

Uno dei risultati più interessanti leggibili in questa tabella è senz'altro quello del coefficiente del *Dow Jones Industrial Average*. Il coefficiente ottenuto attraverso il modello in forma estesa ed il modello ridotto è molto simile e significativo con un livello dell'1%. Questo risultato conferma l'effettiva utilità del *Twitter Bullishness Index* nella previsione dei futuri movimenti dell'indice DJIA in quanto, essendo il coefficiente di regressione molto simile sia nell'analisi col modello esteso che in quella con il modello ridotto, dimostra l'effettiva capacità di T_t^B di riuscire a dare informazioni non catturate in altri indicatori di *sentiment*. Rispettando i risultati ottenuti nei due modelli quindi, da un incremento di una deviazione standard nel *Twitter Bullishness Index* ci si attende un incremento di 13 punti base circa rendimento del DJIA il giorno successivo.

La tabella inoltre mostra che anche nella previsione del FTSE100 e del GSPTSE la relazione mostra un coefficiente statisticamente significativo in modulo anche maggiore di quello del DJIA, inoltre è marcata anche una relazione di tipo *reversal* su lag di ordine maggiore in linea con i risultati ottenuti precedentemente. È interessante notare come i risultati ottenuti siano sebbene simili molto meno robusti per il mercato cinese, in quanto in Cina il social network più popolare è Weibo mentre Twitter è solo raramente utilizzato. La relazione di lag 1 sebbene risulta positiva, con un incremento atteso di 8.74 punti base per unità di *standard deviation* in T_t^B , è statisticamente significativa solo al 10%, mentre non abbiamo significatività nelle relazioni con lag di ordine mag-

giore. Questo risultato può essere motivato dal fatto che sebbene Twitter non sia utilizzato nel paese, sappiamo che nei mercati finanziari moderni esiste un discreto livello di correlazioni fra le performance delle maggiori borse mondiali, quindi la relazione ottenuta che risulta in modulo pari circa alla metà della media delle relazioni ottenute negli altri mercati, potrebbe essere semplicemente dovuta al riflesso della leggera correlazione delle performance borsistiche nel periodo.

Dopo aver risposto anche alla seconda domanda di ricerca gli autori passano all'analisi dei dati ottenuti dai trend di ricerca di Google.

Come anticipato precedentemente in questo caso sia l'intervallo di ricerca che i termini chiave utilizzati cambiano leggermente per adattarsi al tipo di dati. Google offre al proprio pubblico la consultazione dei trend su base settimanale, per questo motivo gli autori ampliano il campione analizzato portandolo da gennaio 2007 a dicembre 2012 in modo da avere 313 osservazioni settimanali. Inoltre le *key word* analizzate per le ricerche sul motore cambiano da *bullish* e *bearish* a "bull market" e "bear market" per le lingue anglofone ed i rispettivi ideogrammi cinesi 牛市 e 熊市.

Una volta ottenuti i volumi settimanali, nonostante l'output sia scalato su base 100 per il pubblico, con l'equazione 8 si riesce a sintetizzare ugualmente la misura di Google Bullishness e definire il *Google Bullishness Index (GBI)* per le 313 osservazioni in entrambe le lingue.

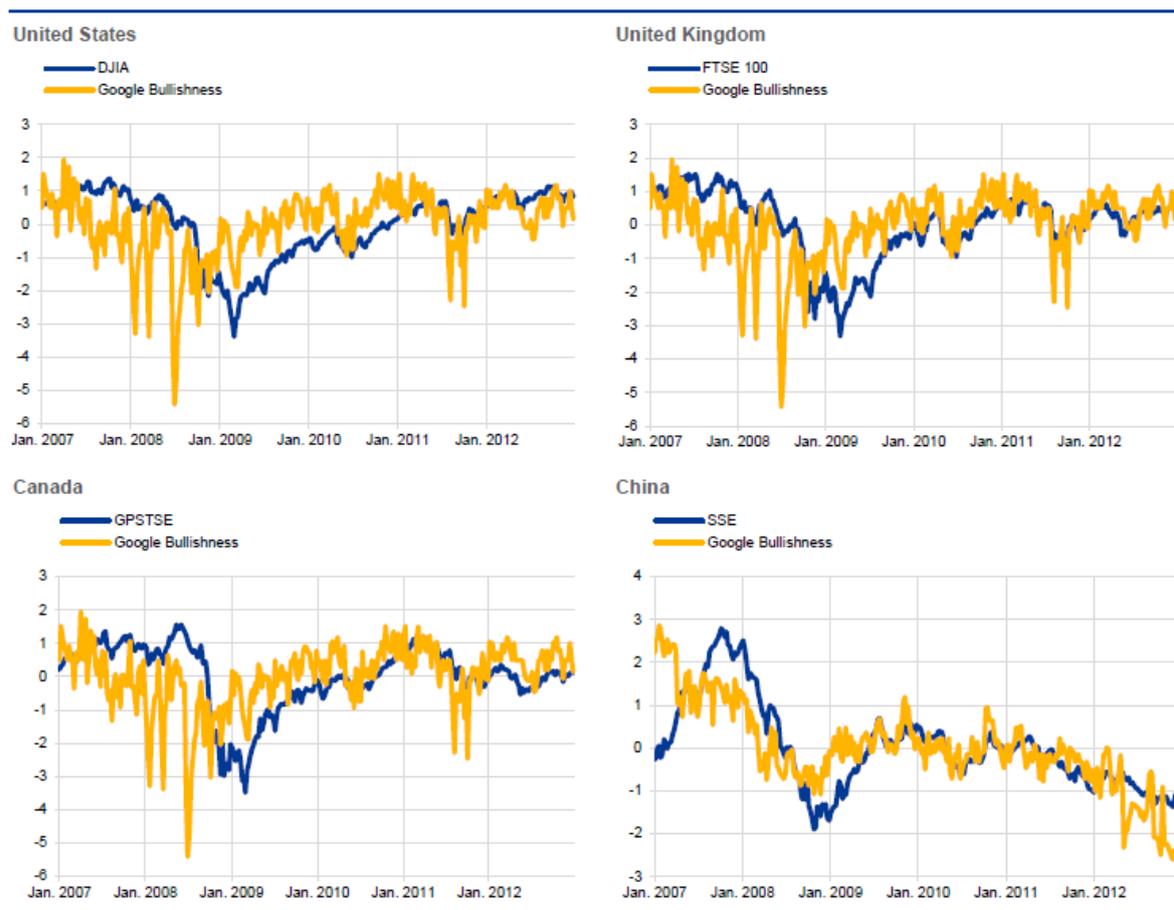
La figura 3.3 mostra i grafici degli indici finanziari analizzati comparati con il *Google Bullishness Index* sintetizzato dalle ricerche nella rispettiva lingua.

Anche prima di effettuare analisi di correlazioni è evidente come il comportamento del GBI sia in molti casi molto simile a quello dei vari indici. Stimando i coefficienti di correlazione lineare di Pearson fra l'indice G_w^B ed i mercati USA, UK, Canada e Cina si ottengono dei coefficienti rispettivamente del 30%, 38%, 23% e 65%, tutti statisticamente significativi con un livello di confidenza del 99%. La correlazione positiva fra il *Google Bullishness Index* e gli indici affiancata ad un semplice analisi grafica sembra mostrare che anche in questo caso G_w^B si muova in anticipo rispetto al mercato.

Ad esempio, *Google Bullishness Index* raggiunge il minimo a metà del 2008, prima della crisi dei mercati di fine 2008 e inizio 2009 negli Stati Uniti, Regno Unito e Canada. Allo stesso modo, l'indice G_w^B cinese ha raggiunto un picco all'inizio del 2007,

che precede il picco del mercato dei primi mesi del 2008. Successivamente, una tendenza al ribasso della *bullishness* è stata seguita da una tendenza al ribasso nel mercato fino al 2009. È sorprendente scoprire che il cinese *Google Bullishness Index* presente la più alta correlazione ($\gamma = 0,65$) con il mercato rispetto ai mercati degli altri tre paesi in esame, in cui Google è il motore di ricerca principale; in Cina, Google rappresenta meno del 15% del mercato della ricerca nel 2012, rispetto al oltre il 75% per Baidu, un *provider* locale poco conosciuto al di fuori della Repubblica Popolare Cinese. La forte correlazione positiva tra il mercato azionario cinese e l'indice di Google può essere attribuito al numero enorme di utenti con accesso ad internet nel paese (oltre 500 milioni nel 2012) il cui 15% corrisponde per approssimazione numericamente agli utenti con accesso al web nel intero Regno Unito.

Figura 3.3



Seguendo il filone metodologico attuato per Twitter gli autori non si fermano alla sola relazione di correlazione lineare di Pearson fra le variabili analizzate poiché questa non ci dà in nessuna indicazione sulla direzione della causalità fra le variabili. Adottando lo stesso modello VAR utilizzato in precedenza gli autori indagano sulla relazione sussistente fra il *Google Bullishness Index* e le variazioni settimanali degli indici analizzati. I risultati dell'analisi utilizzando un lag di relazione che va da 1 a 4 settimane sono riportati nella tabella 3.4. sia per la serie originaria G_W^B , che per quella ottenuta applicando una differenza di ordine 1 per renderla stazionaria ΔG_W^B .

Tabella 3.4

Bullishness	US:DJIA	UK: FTSE100	CA:GSPTSE	CN:SSE
ΔG_{W-1}^B	-21.48 (0.24)	18.36 (0.36)	3.84 (0.84)	4.91(0.87)
ΔG_{W-2}^B	6.65 (0.73)	23.68 (0.27)	16.09 (0.44)	20.0 (0.53)
ΔG_{W-3}^B	-19.92 (0.29)	0.14 (0.99)	1.83 (0.93)	-16.39(0.60)
ΔG_{W-4}^B	-17.71 (0.34)	8.40 (0.67)	-7.07 (0.71)	-25.84 (0.38)
G_{W-1}^B	-24.38 (0.32)	33.8 (0.26)	13.93 (0.64)	25.11(0.71)
G_{W-2}^B	35.87 (0.21)	9.26 (0.78)	24.54 (0.46)	47.40 (0.54)
G_{W-3}^B	-30.24 (0.29)	-32.76 (0.32)	-14.29 (0.66)	-63.20 (0.41)
G_{W-4}^B	18.28 (0.44)	8.14 (0.78)	-2.80 (0.92)	18.99(0.77)

Come è possibile notare in questo caso i coefficienti di regressione non risultano mai statisticamente significativi nonostante i livelli di correlazione lineare mostrata dalle variabili. Il motivo più probabile alla base di questa assenza di significatività statistica risiede nella frequenza delle osservazioni. In un mercato tecnologicamente avanzato come quello attuale le variazioni dovute dal *sentiment* degli investitori possono essere; così come abbiamo visto per Twitter, possono manifestarsi ed anche riassorbirsi all'interno di una settimana di negoziazioni. Questo spiegherebbe perché sebbene esista una correlazione lineare fra le variabili in esame, il *Google Bullishness Index* non abbia alcuna qualità predittiva nei confronti delle variazioni settimanali del mercato.

Invertendo il verso della causalità invece, l'analisi VAR mostra risultati con un'alta significatività statistica. I risultati sono coerenti con *modus operandi* che in finanza comportamentale viene chiamato *Positive FeedBack Trading Strategy*. In pratica questo comportamento comune degli investitori, scoperto da De Long già nel 1990 ma diventato famoso con le ricerche successive di Daniel Kahneman e Andrei Shleifer, prevede che una parte dei *noise trader* aspetti di avere un *feedback* dal mercato prima di investire, in pratica questo tipo di investitore compra quando il mercato è

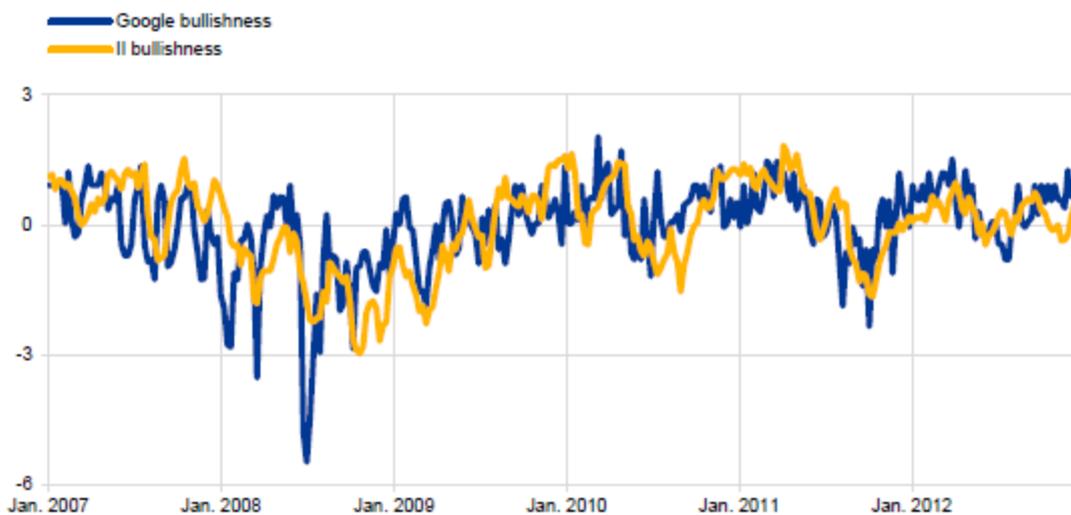
salito e vende quando il mercato è sceso facendosi governare dall'ottimismo quando il mercato si è dimostrato rialzista e dal pessimismo quando il mercato si è dimostrato ribassista.

Nonostante l'inutilità nel prevedere i rendimenti settimanali dei diversi indici di mercato, gli autori conducono un'ulteriore analisi per capire se il *Google Bullishness Index* potrebbe trasmettere informazioni predittive per il *sentiment* futuro degli investitori, piuttosto che i prezzi di mercato.

L'indice *US Advisor Sentiment Report of Investors Intelligence* è un indice di fiducia degli investitori ben accettato in finanza, che misura quale sono le aspettative del mercato in base al grado delle raccomandazioni dei maggiori intermediari dotati di una divisione di *Market Research* e servizi di *Financial Advisor*. Ogni raccomandazione ha un rating intrinseco che la identifica come rialzista, ribassista o neutra.

Sulla base di equazione (1) gli autori creano un indice alternativo chiamato "Bullishness II" che riassume le aspettative di breve periodo riguardo ai mercati finanziari (hai fini prettamente computazionali le raccomandazioni di tipo neutro non vengono utilizzate ai fini della composizione dell'indice). Gli autori mettono quindi a confronto *Bullishness II* con il *Google Bullishness Index*. Dalla figura 3.4 si evince immediatamente una forte relazione fra le due variabili che mostrano nel periodo osservato (per il solo caso del mercato americano) una correlazione molto forte ($\gamma = 0.54$, $p = 0,01$).

Figura 3.4



Come nei casi precedenti, più che per la relazione lineare gli autori sono interessati a determinare se esiste una casualità fra il *Google Bullishness Index* e *Bullishness II*.

Come nei casi precedenti si ricorre ad un modello VAR per determinare la relazione fra le serie delle differenze prime delle due variabili anziché le serie originali per ovviare al problema della non stazionarietà.

Sorprendentemente i risultati ottenuti, riassunti nella tabella 3.5, mostrano che nonostante esista una forte relazione lineare fra le due variabili non esiste una relazione *lag-gata* che abbia significatività statistica. Nonostante l'indice Google Bullishness mostri una significatività statistica come predittore, modello è poco utile ai fini previsionali in quanto presenta un R^2 corretto pari solo al 6%, il che conferma che è difficile prevedere le aspettative di mercato di breve periodo utilizzando questo tipo di dati.

Tabella 3.5

Predicting weekly II sentiment using Google bullishness

Lag	II bullishness		Google bullishness	
	Coeff.	p-value	Coeff.	p-value
1	0.08	0.18	0.18	0.002
2	0.005	0.93	0.19	0.002
3	-0.02	0.67	0.19	0.003
4	-0.06	0.27	0.002	0.980

Adjusted $R^2 = 0.06$, $F = 3.62$, $df (8, 299)$, $p = 0.0005$

Punti di forza e debolezza

Il modello presentato dalla Dott.sa Huina Mao ed il Dott. Scott Counts ed il Professor Johan Bollen riesce a prendere il meglio dei modelli presentati precedentemente e sviluppare una metodologia innovativa più chiara e soprattutto replicabile.

Gli autori risolvono in maniera molto semplice uno dei problemi che maggiormente gravano sull'intera disciplina della *sentiment analysis*, ovvero la poca accuratezza ed affidabilità dei classificatori linguistici.

La nuova misurazione del *sentiment* è frutto di un cambio di impostazione a monte dell'intera metodologia. Se nel modello del 2010 gli autori partivano da un campione che racchiudeva l'espressione dell'intera collettività per poi elaborarlo e filtrarlo, ora l'approccio è completamente diverso. Nel nuovo modello l'attenzione si focalizza solamente sugli investitori e non più sulla totalità degli utenti, inoltre la regola di selezione dei messaggi rialzisti e ribassisti è chiara e inconfutabile, e, sebbene possa essere ancora affinata, riesce a fare in modo che non vengano catturati all'interno della misurazione del *sentiment* i messaggi immessi da persone che non hanno nulla a che fare con il mercato. Il modello inoltre è risulta coerente in termini di tipologia di variabili, perché così come nella versione precedente del 2010 vengo confrontate espressioni della compagine di investitori con misure di mercato che riflettono l'andamento generale e non i singoli titoli. Gli autori sono riusciti a riapplicare con successo la metodologia più largamente apprezzata in termini di misurazione del *sentiment* aggregata nel mondo accademico, dall'ormai celebre articolo di Antweiler e Frank del 2004 sullo studio dei forum specializzati, ai social network senza che questo cambiamento possa inquinare la fonte dei messaggi e soprattutto evitando la possibile distorsione derivante dall'uso dei classificatori linguistici.

I risultati dello studio sono incoraggianti, a differenza degli studi precedenti questa volta non sia ha una sola accuratezza direzionale ma un vero e proprio rendimento atteso dovuto ai cambiamenti nel *sentiment* degli investitori. La significatività statistica e la robustezza dei coefficienti stimati propone ancora una volta la *sentiment analysis Twitter-based* come valido, se non più performante, predittore degli scostamenti temporanei dei prezzi dai fondamentali, inoltre a differenza degli studi precedenti che richiedevano basi di programmazione e di *machine learning*, questo metodo non è particolarmente complesso in nessuna delle sue fasi se non per l'iniziale acquisizione dei dati.

Sebbene il modello mostri numerosi vantaggi, ci sono diverse cose che non lo rendono perfetto. Innanzitutto lo studio mostra una correlazione tra il promettente predittivo *Twitter Bullishness Index* e le variazioni dei prezzi del mercato azionario, questo non offre alcuna informazione per quanto riguarda la causalità. Sebbene qualcuno possa pensare che le relazioni di tipo causale siano importanti solo ai fini modellistici, nelle applicazioni sono proprio queste le relazioni che fanno la differenza nei modelli che producono previsioni corrette e stime affidabili. Trovare questo tipo di relazioni causali dai Big

Data è un problema di ricerca impegnativo. Rispetto a qualsiasi altro fenomeno osservabile in natura che segue delle regole precise, il comportamento umano e le interazioni sociali non sono prevedibili con precisione. I Big Data forniscono un numero enorme di informazioni circa l'espressione umana in un determinato momento o in una determinata situazione, ampliare i risultati in entrambi le direzioni è ancora una sfida per la ricerca.

È probabile che in futuro ci si concentrerà sullo sviluppo di un quadro teorico, combinando metodi di progettazione sperimentale e algoritmi di apprendimento automatico per dedurre la relazione causale tra l'espressione del *sentiment* estratta da Twitter ed i mercati finanziari, ma per adesso questi risultati non ci sono ancora.

Nonostante però non esista un'evidenza empirica di relazione di causalità, a mio parere, i vantaggi che ho elencato in precedenza rendono questo modello il più adatto alla misurazione del *sentiment* nei mercati finanziari attraverso i *social network*, e quindi di conseguenza lo strumento con cui cercherò di dimostrare la mia tesi nei paragrafi successivi.

Capitolo 4

Applicazione e Modellizzazione

4.1 APPLICAZIONE DEL MODELLO

Dopo aver illustrato dettagliatamente la letteratura sottostante alla metodologia che userò nelle pagine successive, ripropongo le mie domande di ricerca in maniera più dettagliata ed esplicita.

D1) E' possibile utilizzare la *sentiment analysis* per effettuare investimenti a breve termine?

D2) In che modo si può costruire una strategia di investimento che sfrutti le evidenze empiriche analizzata fin ora?

D3) Se esiste una strategia profittevole, vale la pena davvero preferirla rispetto ad altre?

Per rispondere a queste domande ho raccolto circa 6 mesi di *tweets*, raggiungendo un archivio composta da circa 750.000 stringhe e li ho utilizzati per creare e valutare strategie di investimento con i principali indici anglofoni riportati nel paragrafo precedente. Nonostante esistano molti modelli validi per la stima del *sentiment* dei mercati ho ritenuto, in base a considerazioni prettamente personali, che il modello più adatto ai fini della mia tesi fosse quello presentato nel paragrafo precedente, ovvero Bullishness Model di Huina Mao, Johan Bollen e Scott Counts presentato nel luglio 2015 (MBC-2015)

Database tweets

Nei lavori analizzati fin ora gli autori non riportavano mai dettagliatamente la modalità di acquisizione dei *tweets*. Questa parte rispetto ad altre acquisizione di dati grezzi mostra diverse caratteristiche che la rendono più ostica.

Innanzitutto rispetto ai dati finanziari semplici (prezzi di chiusura, volume degli scambi, ecc.) per i quali esistono svariate fonti di accesso ufficiali e non solo, nel caso di Twitter non esiste un *database* universale accessibile al pubblico che raggruppi TUTTI i *tweets* globali. Twitter gestisce in maniera molto rigida gli accessi ai suoi dati, riconosciuti ormai globalmente come espressione della collettività. Entrando nel dettaglio ci sono diversi livelli di filtraggio di accesso.

Livello A – *Twitter Database* - Contiene tutti i *tweets* postati sul social network in qualsiasi momento, include sia messaggi pubblici che privati. Questo database è accessibile solo alla stessa società Twitter.inc.

Livello B – *Firehose*- Contiene la totalità *tweets* resi pubblici dagli utenti postati sul social network. L'accesso a questo livello è riservato solamente ad una manciata di aziende a livello globale (parliamo di aziende del calibro di Microsoft e Oracle)

Livello C – *GardenHose*- Rappresenta un campione pari al 10% dei *tweets* appartenenti al livello B. La quasi totalità delle API presenti online ed i Database a pagamento *retail* hanno accesso a questo livello.

Livello D – *Spritzer*- Rappresenta un campione di circa l'1% dei *tweets* contenuti nel livello B e contrassegnati come *hot trend*. Viene spesso utilizzato dai siti per informare in tempo reale di cose si parla in quel momento.

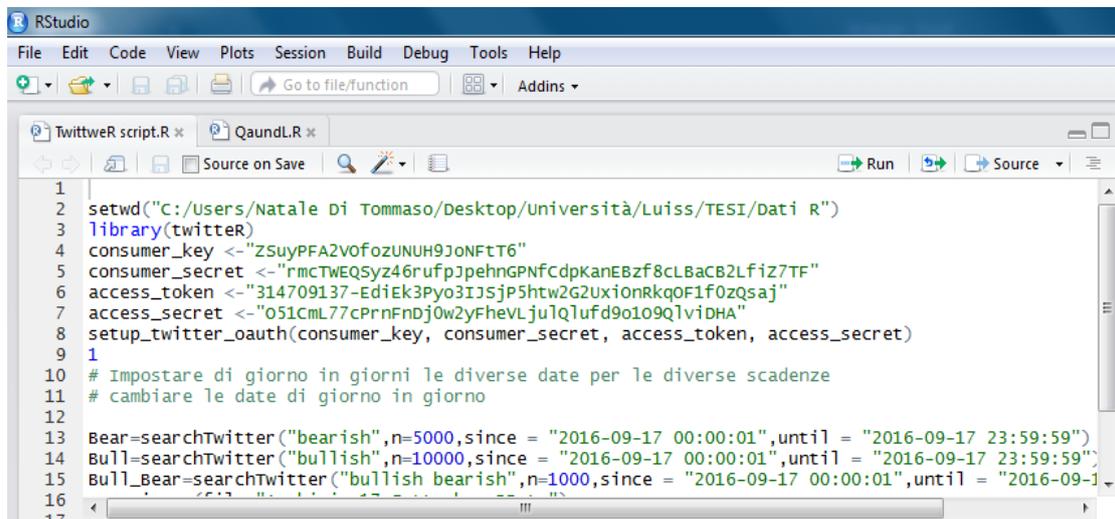
Non è chiaro che *tweets* siano inclusi in questi diversi flussi di dati e come viene effettuato il campionamento. È ugualmente possibile che l'API estragga campioni casuale di *tweets* dall'intera popolazione o che estragga soltanto le prime poche migliaia di *tweets* o ancora che estragga solo *tweets* da un particolare segmento del grafico di rete. Data l'incertezza è difficile per i ricercatori fare affermazioni circa la qualità dei dati che essi stanno analizzando, nonostante questo però poiché in ogni caso parliamo di Big Data (quindi di migliaia e migliaia di stringhe) è possibile che, per la teoria del campionamento statistico, la grande dimensione del campione estratto possa rappresentare una buona approssimazione della popolazione di *tweets*.

Seguendo l'approccio MBC-2015 per iniziare la mia analisi avevo bisogno della serie storica dei volumi dei *tweets* giornalieri contenenti le parole chiave *bullish* e *bearish*. In fase di costruzione del *dataset* quindi mi sono trovato come è possibile immaginare di fronte ad una raccolta dei dati non troppo semplice. Inizialmente la mia intenzione era quella di scaricare i volumi delle parole chiavi per le osservazioni dell'intero 2015. Dopo aver contattato le maggiori *tweet-data-bank* e aver ricevuto preventivi che andavano fra i 2500 ai 14400 euro per le sole due serie storiche richieste (o meglio il 10% di quelle richieste poiché il livello di accesso era di livello C) ho capito immediatamente che dovevo cambiare approccio.

Sebbene non esistono archivi storici gratuiti per questo tipo di dati, esistono diversi metodi che permettono di avere accesso ad un campione di livello C in maniera totalmente gratuita. Il rovescio della medaglia è che queste modalità di accesso hanno una memoria a breve termine. In pratica le API che offrono questo tipo di servizio si basano essenzialmente su componenti aggiuntivi (a volte a pagamento) di software a licenza libera, in particolare Python, R Studio e Google Drive. Questi programmi hanno accesso all'API ufficiale di Twitter (GET-API) che dà l'accesso ad un campione casuale ottenuto in base ad una determinata ricerca, limitato però ad un tetto massimo di volume pari a 10000 *tweets*. Questa limitazione consente di ottenere un campione di *tweet* non distorto solamente se ripetuta costantemente per ricerche che non hanno volumi di output maggiori di 10000. Fortunatamente nel nostro caso per entrambe le parole chiavi ricercate come vedremo i volumi giornalieri di output sono sensibilmente al di sotto di questa soglia nella gran parte dei casi. Per ottenere un database di partenza per la mia ricerca ho quindi raccolto quotidianamente mediante l'esecuzione di uno script in R-Studio i volumi di *tweets* contenenti le due *key words* di interesse ottenendo un *dataset* di circa 750,000 *tweets*. Inoltre a scopo precauzionale ho utilizzato la componente aggiuntiva *Twitter Archive Premium* di Google Drive per avere un'altra estrazione di controllo della medesima serie.

In accordo con quanto mostrato in MBC-2015 ho utilizzato come intervallo di tempo per la misurazione dei volumi giornalieri la fascia oraria dalle 00:00:01 alle 23:59:59 del GMT + 0.

Figura 4.1



```

1
2 setwd("C:/Users/Natale Di Tommaso/Desktop/Università/Luiss/TESI/Dati R")
3 library(twitter)
4 consumer_key <- "ZSuyPFA2VofozUNUH9J0NFTt6"
5 consumer_secret <- "rmcTWEQSyZ46rufpJpehnGPNFcdpKanEBzf8cLbaCB2LfiZ7TF"
6 access_token <- "314709137-EdiEk3Pyo3IJSjP5htw2G2Uxi0nRkqOF1f0zQsaj"
7 access_secret <- "051cML77cPrnFnDj0w2yFheVLju1q1ufd9o109Q1viDHA"
8 setup_twitter_oauth(consumer_key, consumer_secret, access_token)
9
10 # Impostare di giorno in giorni le diverse date per le diverse scadenze
11 # cambiare le date di giorno in giorno
12
13 Bear=searchTwitter("bearish",n=5000,since = "2016-09-17 00:00:01",until = "2016-09-17 23:59:59")
14 Bull=searchTwitter("bullish",n=10000,since = "2016-09-17 00:00:01",until = "2016-09-17 23:59:59")
15 Bull_Bear=searchTwitter("bullish bearish",n=1000,since = "2016-09-17 00:00:01",until = "2016-09-17 23:59:59")
16
17

```

Lo script che ho utilizzato, riportato nella figura 4.1, è stato ripetuto ogni giorno per i 6 mesi che vanno dal 9 marzo 2016 al 9 settembre 2016. L'output dello *script* basato sul pacchetto *Twitter* consiste nel conteggio dei volumi giornalieri di *tweets* contenenti le *key word* selezionate, rispettando come orario di riferimento quello GMT + 0.

Poiché R-Studio non implementa un'interfaccia grafica che ci permette di vedere realmente il contenuto dei vettori di stringhe, al fine puramente rappresentativo riporto nella figura sottostante l'output ottenuto da *Twitter Archiver* che essendo a livello di contenuti identico, è però in un formato Excel quindi più familiare e osservabile.

La figura 4.2 mostra una parte dei *tweets* della giornata di mercoledì 24 agosto 2016, in particolare il volume di *tweets* contenenti il termine *bearish* che osserviamo dalle 9:04 alle 9:24 del mattino. Come è possibile vedere per ogni messaggio, oltre ad essere riportato il testo sono riportate anche tutte le altre informazioni sensibili che possono aiutare l'utente a filtrare l'output ottenuto. Per ogni *tweet* quindi è disponibile l'ora esatta di pubblicazione, il nome dell'autore, la descrizione dell'autore, il testo, il numero di *follower* dell'autore, il numero di *retweet* e la zona di provenienza (le stesse informazioni sono disponibili anche su *Twitter*).

Sebbene il modello MBS 2015 non preveda l'utilizzo di informazioni aggiuntive rispetto ai volumi giornalieri è interessante notare l'alto grado di accuratezza delle informazioni ottenibili con programmi gratuiti alla portata di tutti.

Figura 4.2

Twitter Search Rule: bearish						User Details	
	Date	Full Name	Tweet Text	Followers	Retweets	Location	Bio
3468	8/24/2016 9:04:24	mbatrade	Oil down sharply on bets for bearish U.S. stockpile data	30	0	London, England	The New Generati
3469	8/24/2016 9:04:34	TeleTrade	EUR/USD: En-Route To 1.16; USD/JPY: En-Route To 97	2481	0	ČESKÁ REPUBLIKA	TeleTRADE - jedna
3470	8/24/2016 9:06:01	LabworxFX	EURUSD if we see a new high this is my prediction. http	108	0	Antwerpen, België	Pro Trader - Vladim
3471	8/24/2016 9:06:45	Shraddha Babla	Sundaram MF: Bearish on metals & technology stocks E	958	2		Senior Research A
3472	8/24/2016 9:07:33	Sanjay	Bearish engulfing on the weekly is a great buy signal#	341	0	England, United King	Speculator in FGBl
3473	8/24/2016 9:08:39	C Investment	Oil down sharply on bets for bearish U.S. stockpile data	800	0	Sweden	Trend following inv
3474	8/24/2016 9:09:28	Stock Exchange	#Financial Oil down sharply on bets for bearish U.S. stc	3529	0	Suisse	Financial News
3475	8/24/2016 9:13:01	TV_Trading Ideas	USOIL: Bearish Butterfly forU2026 \$USOIL https://t.co/5	5616	0	Chicago, IL	Best trading ideas
3476	8/24/2016 9:13:16	Andreas Cseh	RT @DavidinglesTV: Analysts haven't been this bearish	1999	5	Hong Kong, Singapo	Author of Quantum
3477	8/24/2016 9:13:51	Paul Ebeling	Sirius XM (NASDAQ:SIRI) Bearish Outlook - Live Tradin	1982	0	Nassau, The Baham	Tech Analyst: writ
3478	8/24/2016 9:14:26	TeleTrade	EUR/USD: En-Route To 1.16; USD/JPY: En-Route To 97	1672	0	POLSKA	TeleTrade - jeden z
3479	8/24/2016 9:14:33	TV_Trading Ideas	bearish \$GBPCHF https://t.co/OxSaa9j5zX	5616	0	Chicago, IL	Best trading ideas
3480	8/24/2016 9:15:08	TeleTrade	EUR/USD: En-Route To 1.16; USD/JPY: En-Route To 97	2002	0	Europe	TeleTrade - una dir
3481	8/24/2016 9:17:36	Edward Milewski	- Oil down sharply on bets for bearish U.S. stockpile da	279	0	London, Ontario, Cai	Philosopher and Cil
3482	8/24/2016 9:20:51	D Graef	Tellin_Fibs: MACD Bearish Signal Line Cross: This scan	3653	0		We are Paid Adver
3483	8/24/2016 9:22:12	TeleTrade	EUR/USD: En-Route To 1.16; USD/JPY: En-Route To 97	1737	0	Europe	TeleTrade - az egy
3484	8/24/2016 9:22:26	TeleTrade	EUR/USD: En-Route To 1.16; USD/JPY: En-Route To 97	136	0	ESPAÑA	TeleTrade - una de
3485	8/24/2016 9:22:54	Elliottwave Forecast	Is it time to be bullish or bearish the US Dollar? https://t.c	12102	0	Miami/London	Accurate & timely E
3486	8/24/2016 9:23:59	GreekStocks T2D	Gold bearish as long as price is below \$1,350. https://t.	473	0		Athens stock exch
3487	8/24/2016 9:24:02	Alexandros Yfantis	Gold bearish as long as price is below \$1,350. https://t.	2582	1	Greece	Portfolio Manager. I
3488	8/24/2016 9:24:26	TeleTrade	EUR/USD: En-Route To 1.16; USD/JPY: En-Route To 97	1706	0	Europe	Teletrade - μία από
3489	8/24/2016 9:24:28	TeleTrade	EUR/USD: En-Route To 1.16; USD/JPY: En-Route To 97	1706	0	Kyrenia, Cyprus	TeleTrade - μία από

La figura 4.1 oltre a darci una visione generale dei dati grezzi estrapolabili dall' API pubblica, mostra un problema che probabilmente non era presente (o era minimo) nel lavoro MBS 2015. Come è possibile leggere nella riga 3485 del grafico, indicata da una freccia blu, il messaggio sebbene rientri nel correttamente nel campione e venga conteggiato nell' indice presenta fra le altre parole anche il termine *bullish*. Questo significa che lo stesso messaggio viene conteggiato per il medesimo giorno sia al numeratore che al denominatore del rapporto dell'equazione 7. Contrariamente a come i meno attenti sono indotti a pensare, l'aumento di poche unità in entrambi i volumi di ricerca può distorcere l'indice. Sebbene l'incremento in termini assoluti sia identico, questo distorce il rapporto fra i volumi e quindi l'indice finale.

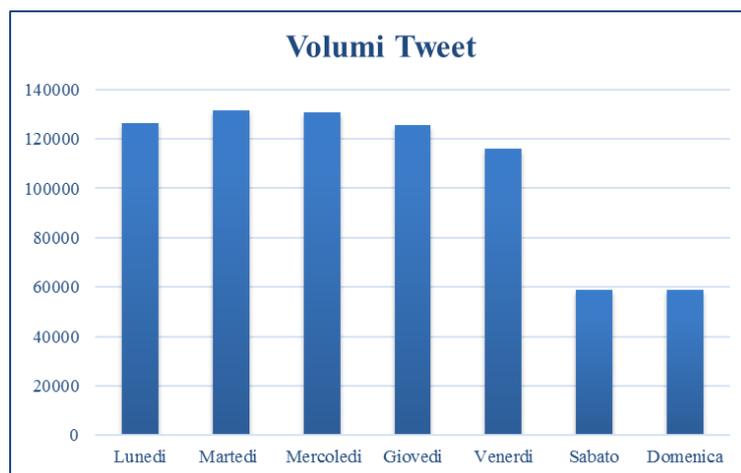
Per questo motivo quindi, a differenza del lavoro MBS 2015 ho deciso di “depurare” il campione da questa distorsione semplicemente sottraendo ad entrambi i volumi giornalieri la quantità di *tweets* che contenevano contemporaneamente entrambe le *key word* *bullish* e *bearish*.

Il database ottenuto alla fine dei sei mesi di osservazioni conta per la precisione 747,954 *tweets*, di cui 23,114 contengono entrambe le *key word* si ricerca.

La quantità di *tweets* raccolti è enorme se raffrontata a quella dello studio al quale fa riferimento questa applicazione. In 3 anni di osservazioni da gennaio 2010 a dicembre

2012 gli autori di MBS 2015 raccolgono circa 310,000 *tweets*, mentre il *database* che ho raccolto in soli 6 mesi di osservazioni è di circa 750,000, per approssimazione quindi proporzionalmente al tempo osservato, il volume di *tweets* contenente le *key word* è cresciuto di circa 14 volte nel corso degli ultimi quattro anni. Questa sostanziale differenza numerica è dovuta principalmente a due fattori: per prima cosa, come già ripetuto diverse volte in questo testo, al fatto che i *social network* in generale hanno avuto una crescita esponenziale negli ultimi anni; sia al fatto che Twitter rispetto ad altre piattaforme è il social network maggiormente utilizzato per condividere notizie di natura finanziaria. E' interessante notare però che nonostante il cambiamento radicale in termini di volumi, la proporzione fra i volumi giornalieri dei vari giorni settimanali sia rimasta quasi del tutto identica. Confrontando il grafico 4.1 con i grafici 4.2 e 4.3 si nota immediatamente come i vari *pattern* di autocorrelazione e diversità sostanziale fra i volumi nei giorni di apertura di borsa ed il weekend.

Grafico 4.1

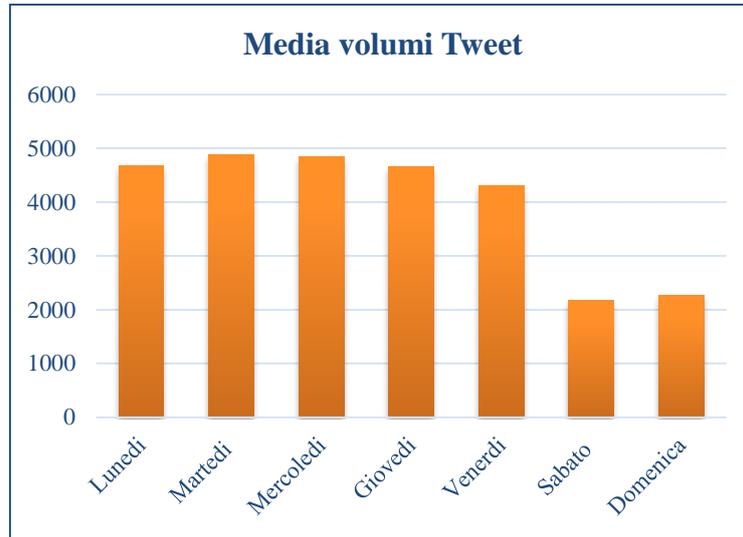


Un'ultima caratteristica interessante del database raccolto sta nella sua continuità nel tempo. Avendo un flusso medio giornaliero (festivi inclusi) di circa 3790 *tweet*, sarebbe possibile aspettarsi che questo cali nei periodi estivi, poiché a livello globale è il periodo in cui la maggior parte delle persone è in vacanza lontano dal lavoro e dal pc o smartphone.

Il grafico 3.5 ci mostra questo non accada. Il grafico riporta il volume cumulato di *tweets* raccolti fino a quella data, se ci fossero differenze significative per i volumi di

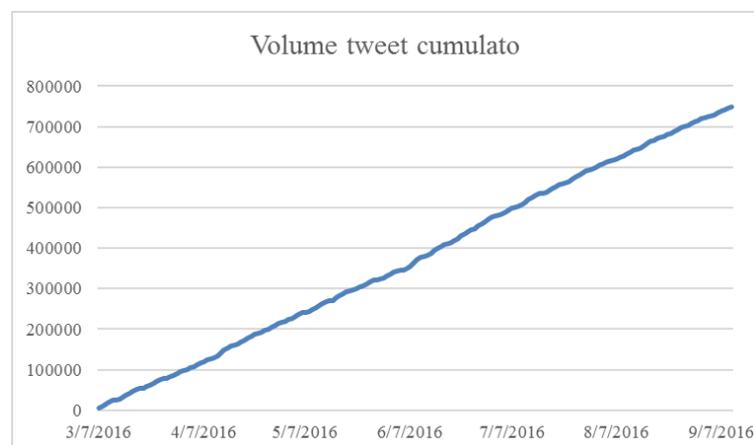
raccolta fra i vari mesi osservati, il grafico dovrebbe assumere un comportamento regolare approssimandosi ad una retta spezzata in caso di differenza sostanziale.

Grafico 4.2



Come è possibile vedere il flusso di informazioni è praticamente stabile per tutto il periodo analizzato, questo permette alla linea che traccia l'evoluzione dei volumi cumulati di seguire un movimento rettilineo con inclinazione costante. Inoltre anche se non è graficamente molto visibile, è presente un buon livello di autocorrelazione fra i volumi di *tweets* ottenuti dei medesimi giorni, come mostrato dalle lievi increspature dell'indicatore ripetute ad intervalli di 7 osservazioni.

Grafico 4.3



Tutte queste similitudini riscontrate fra il *dataset* raccolto e quello del modello di riferimento per questa analisi, non possono che confermare la bontà dei dati raccolti. Inoltre avendo una quantità di dati molto maggiore rispetto agli autori, sia in termini assoluti ma soprattutto in termini proporzionali (ma anche grazie all' opportuna correzione implementata), il *Bullishness Index* che ne verrà derivato sarà un' espressione migliore del *sentiment* collettivo degli investitori di quanto non lo sia stato nel modello di riferimento.

Twitter Bullishness Index

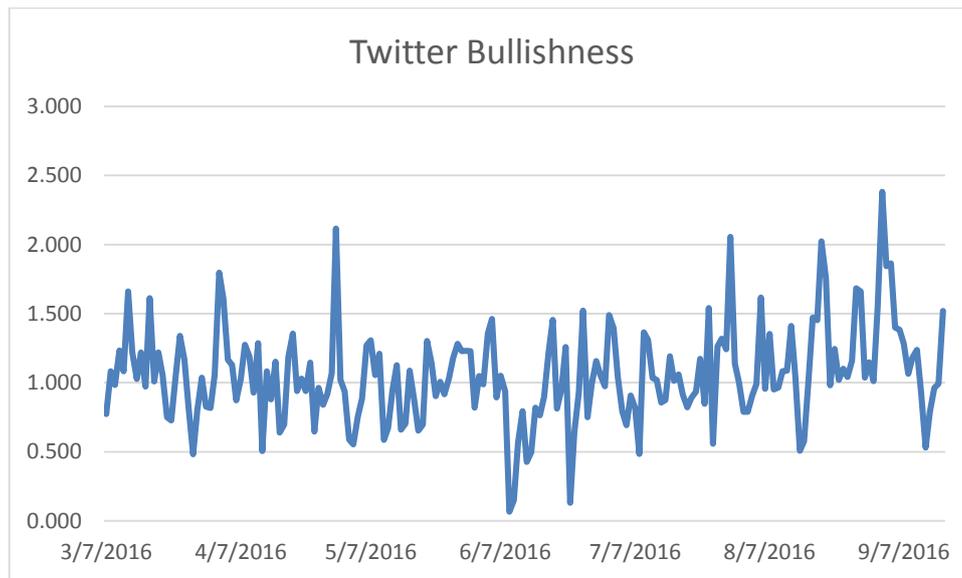
Continuando a seguire la metodologia presentata nel capitolo 3, viene stimato attraverso il *dataset* raccolto il *Twitter Bullishness Index* su base giornaliera.

La serie storica risultante ottenuta attraverso l' applicazione ripetuta dell' equazione 7 è riportata nel grafico 3.6. Prima di fare qualsiasi considerazione riguardo alle relazioni con gli indici di riferimento per questa applicazione è importante notare che la serie temporale raffigurata nel grafico 3.6 non è conformabile con quella risultante dai rendimenti dei mercati. Così come nei lavori presentati precedentemente, il flusso di *tweets* analizzati per estrarre il *Twitter Bullishness Index* non presenta giorni di interruzione come i mercati finanziari. La serie temporale quindi per essere relazionata a qualsiasi serie finanziaria necessita di una correzione per i giorni di chiusura del mercato, che prendono oltre ai giorni di chiusura settimanali quelli di chiusura nazionale specifica del mercato di riferimento. La serie temporale raffigurata nel grafico 4.4 presenta diverse caratteristiche interessanti. La media del valore dell' Twitter Bullishness Index per il periodo di applicazione considerato è pari a 1.057 con una *standard deviation* pari solamente a 0.349 mentre per il *dataset* raccolto in MBC 2015 i rispettivi valori di riferimento erano 0.853 e 0.507.

Sebbene se come abbiamo avuto modo di constatare diverse volte in questo testo, non esista ad oggi una teoria di previsione del *sentiment* è possibile dare diverse motivazioni per queste differenze di distribuzione. Il periodo analizzato nel *paper* di riferimento racchiude quello che analiticamente parlando è definibile come un cambio di tendenza, in particolare come è visibile anche dalla figura 3.3, in un primo sotto periodo assistiamo

ad una diminuzione generale del valore dei maggiori indici finanziari, mentre in un secondo periodo vediamo una fase di modesto rialzo che porta i corsi ad un livello simile a quello iniziale di osservazione.

Grafico 4.4

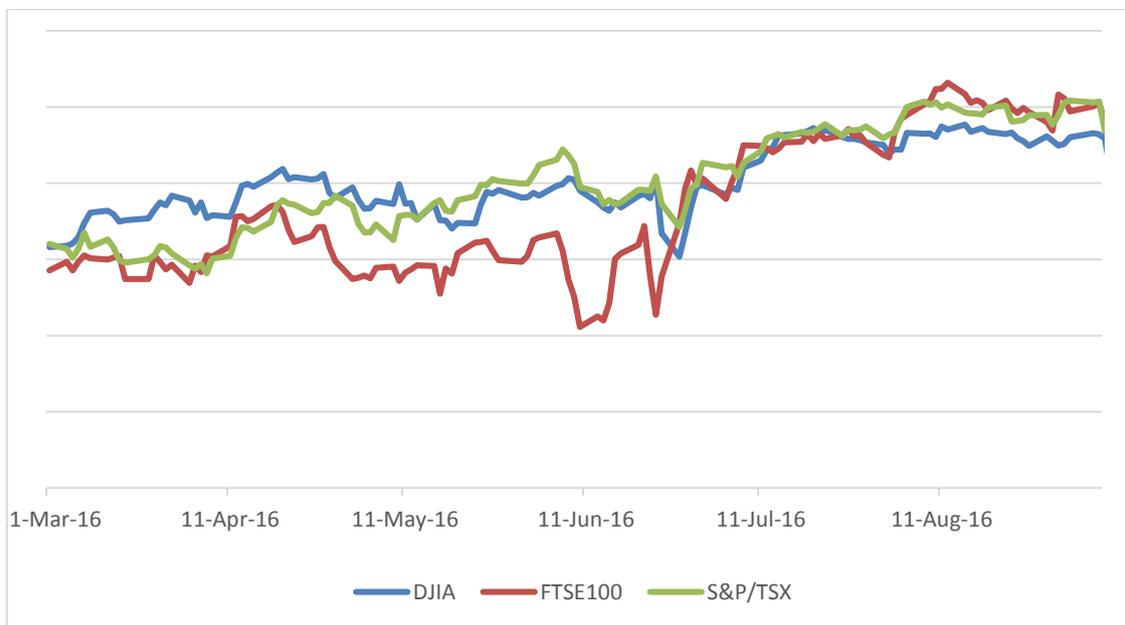


A differenza il periodo osservato nell'applicazione, come vedremo in seguito, è caratterizzato da un trend rialzista (seppure non costante) degli maggiori indici internazionali. Questa differenza potrebbe spiegare i discostamenti fra le due statistiche riscontrate. Poiché come è stato ampiamente discusso, esiste una correlazione positiva fra la misura del *sentiment* ed i movimenti degli indici azionari, è consono attendersi che in un campione caratterizzato da un rialzo dei mercati il *sentiment* misurato degli investitori sia più positivo (rialzista) e caratterizzato da meno variabilità, e che in un periodo caratterizzato da alti e bassi di mercato ci si attenda una misura di *sentiment* media più bassa ma caratterizzata da una variabilità maggiore.

Prima di passare alla rettifica del *Twitter Bullishness Index*, risulta necessario introdurre gli indici che utilizzeremo per vagliare le domande di ricerca poste. In accordo con i risultati ottenuti da MBC 2015 ho scelto gli indici che hanno mostrato una relazione migliore con il *Twitter Bullishness Index*. Gli indici scelti sono quindi lo S&P/TSE della *Toronto Stock Exchange*, il FTSE100 della *London Stock Exchange* ed il *Dow Jones Industrial Average* per il mercato USA. Le performance di questi indici per il periodo di

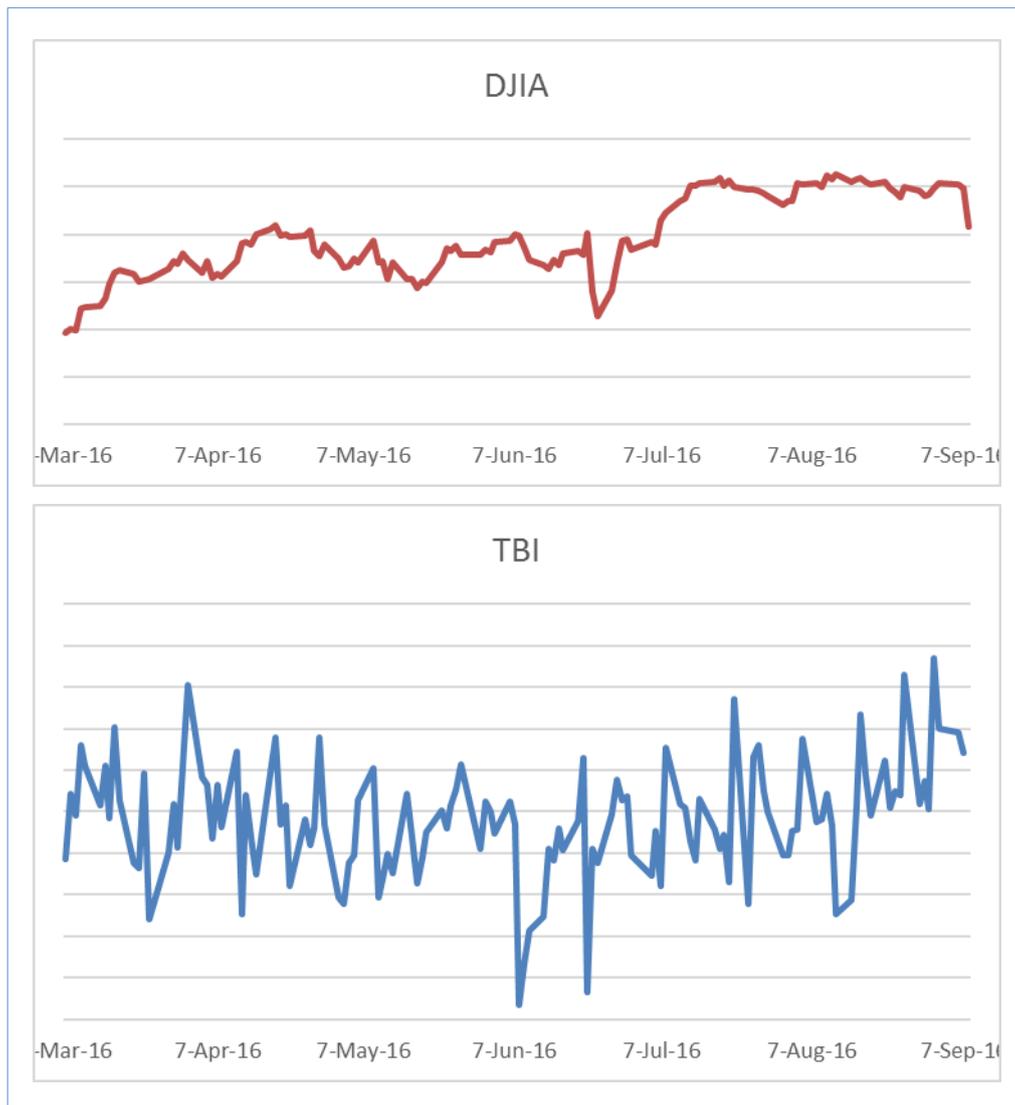
applicazioni sono riassunte nel grafico 4.5 sottostante. Come già anticipato, è possibile notare la tendenza generale positiva del periodo di riferimento, caratterizzato anche da tratti bassa volatilità, come per esempio agosto che secondo Standard&Poor's è stato il mese con la volatilità più bassa registrato negli ultimi 20 anni. Sebbene esista una tendenza globale che accomuna le variazioni di questi tre indici (in particolar modo il DJIA e S&P/TSE) è importante notare che questo non indebolisce l'analisi, perché come vedremo trattandosi di strategie trading ad orizzonte giornaliero queste possono dare risultati diametralmente opposti nonostante la grande similitudine nel comportamento.

Grafico 4.5



Il passaggio successivo quindi consiste nel ricalcolare il *Twitter Bullishness Index*. Rispettando la metodologia presentata ad inizio del secondo capitolo, facendo riferimento al fuso orario GMT+0 vengono incorporati nel giorno di negoziazione i *tweets* compresi fra le 00:00:00 e le 23:59:59, mentre la totalità dei *tweets postati* in giorni di chiusura dei mercati viene accorpata alla quantità osservata nel primo giorno di negoziazione successivo. Questo passaggio che potrebbe sembrare immediato in realtà richiede comunque attenzione. Sebbene globalmente sia diffusa la tendenza delle borse ad essere chiuse il sabato e la domenica, esistono feste nazionali che influenzano i singoli mercati.

Grafico 4.6



Questo problema appare evidente già da un confronto iniziale delle serie storiche dove riscontriamo per il periodo di riferimento ,130 osservazioni per il DJIA e 129 per gli altri due indici. Sebbene ad occhio la discrepanza sembrerebbe limitata solo ad una osservazione in realtà è molto maggiore.

Esistono festività come il 4 luglio negli Usa o il *Victoria Day* il 23 maggio o il *Canada Day* il 1° luglio, o ancora il giorno dopo Pasqua in UK, quest' anno il 28 marzo, o le diverse *Summer e Spring Bank Holiday* che rendono le serie temporali se affiancate completamente sfalsate. Per questo motivo verranno calcolati dal *Twitter Bullishness Index* tre indici differenti *ad hoc* che tengano conto delle date di chiusura dei diversi mercati.

A fini puramente illustrativi il grafico 4.6 riporta l'indice DJIA per il periodo di riferimento ed il *Twitter Bullishness Index* (TBI) ricalcolato con *lag* 1 rispetto ai giorni di apertura di mercato.

Sebbene i due grafici non siano direttamente paragonabili, è possibile notare una certa concordanza nella direzione. Analizzando i dati è possibile osservare una correlazione pari a 17,58 percento fra i due indicatori. Questo dato non deve però essere immediatamente associato ad un'inefficienza previsionale di TBI poiché come è stato ampiamente detto in precedenza, non esiste una relazione di causalità fra queste due famiglie di variabili, ed ai fini di investimento come vedremo siamo più interessati gli shock nell'indicatore rispetto che all'indicatore stesso. È interessante notare come nel periodo di giugno (caratterizzato dalle maggiori turbolenze nel periodo di riferimento) si tocchino i picchi inferiori dell'*Twitter Bullishness Index* in quanto in diversi giorni si è osservato un netto cambiamento fra il volume dei messaggi rialzisti e quelli ribassisti portando il rapporto vicino all'unità e quindi l'indicatore vicino a 0.

Google Bullishness Index

Il *Google Bullishness Index* mostrato nella metodologia esposta del *paper* di riferimento, nonostante esprima una relazione lineare rilevante con i maggiori indici di mercato, non mostra alcuna qualità predittiva nei confronti di questi ultimi.

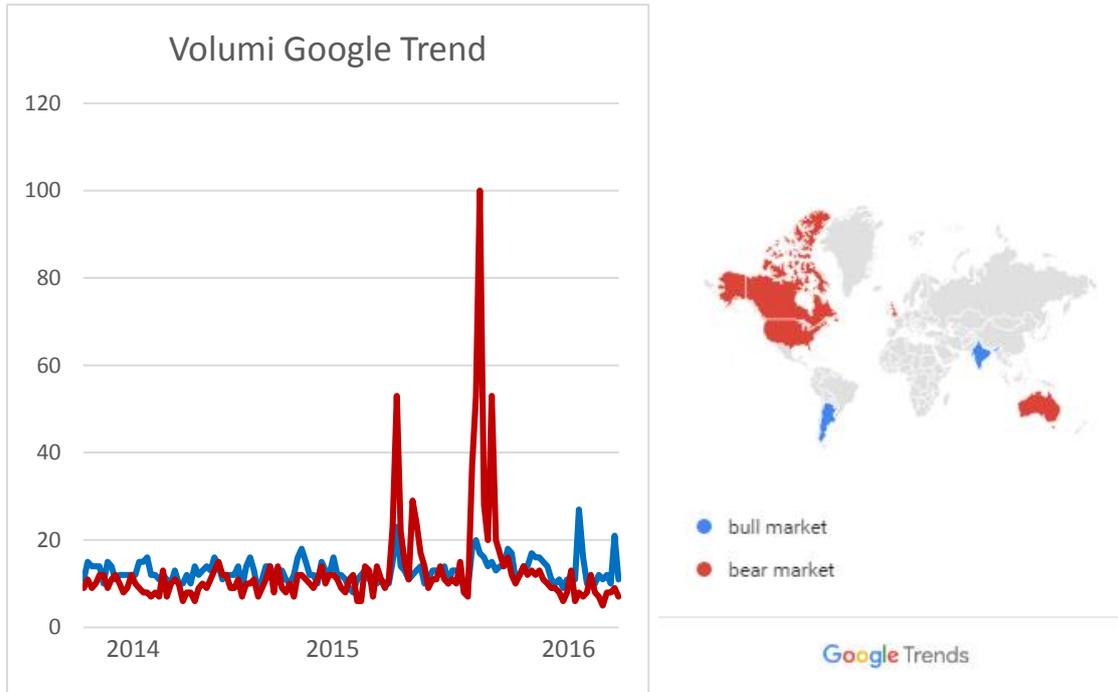
Seguendo i risultati riportati in MBS 2015 questa misura di *sentiment* potrebbe essere usata come variabile di controllo. Se i risultati del *paper* di riferimento fossero confermati, applicando la stessa strategia di investimento basata sul *sentiment* ad entrambi gli indicatori si dovrebbero ottenere performance nettamente differenti.

L'applicazione dell'indice sarà limitata al solo DJIA in quanto, rispettando i risultati del *paper*, è l'indice anglofono che mostra una relazione statisticamente significativamente migliore (la distribuzione di *p-value* riportati nella tabella 3.4 ha una media minore).

Al fine di rendere i due campioni, osservazioni settimanali per il *Google Bullishness Index* e giornalieri per il *Twitter Bullishness Index* (per il DJIA si hanno 130 giornate di apertura dei mercati nell'orizzonte temporale osservato) ho considerato un intervallo che comprende 130 osservazioni settimanali che portano il periodo di osservazione dal 16 febbraio 2014 al 18 settembre 2016.

La figura 4.3 mostra le serie temporali prese da Google Trends dei volumi di ricerca normalizzate ad un tetto di 100 per l'intervallo di riferimento.

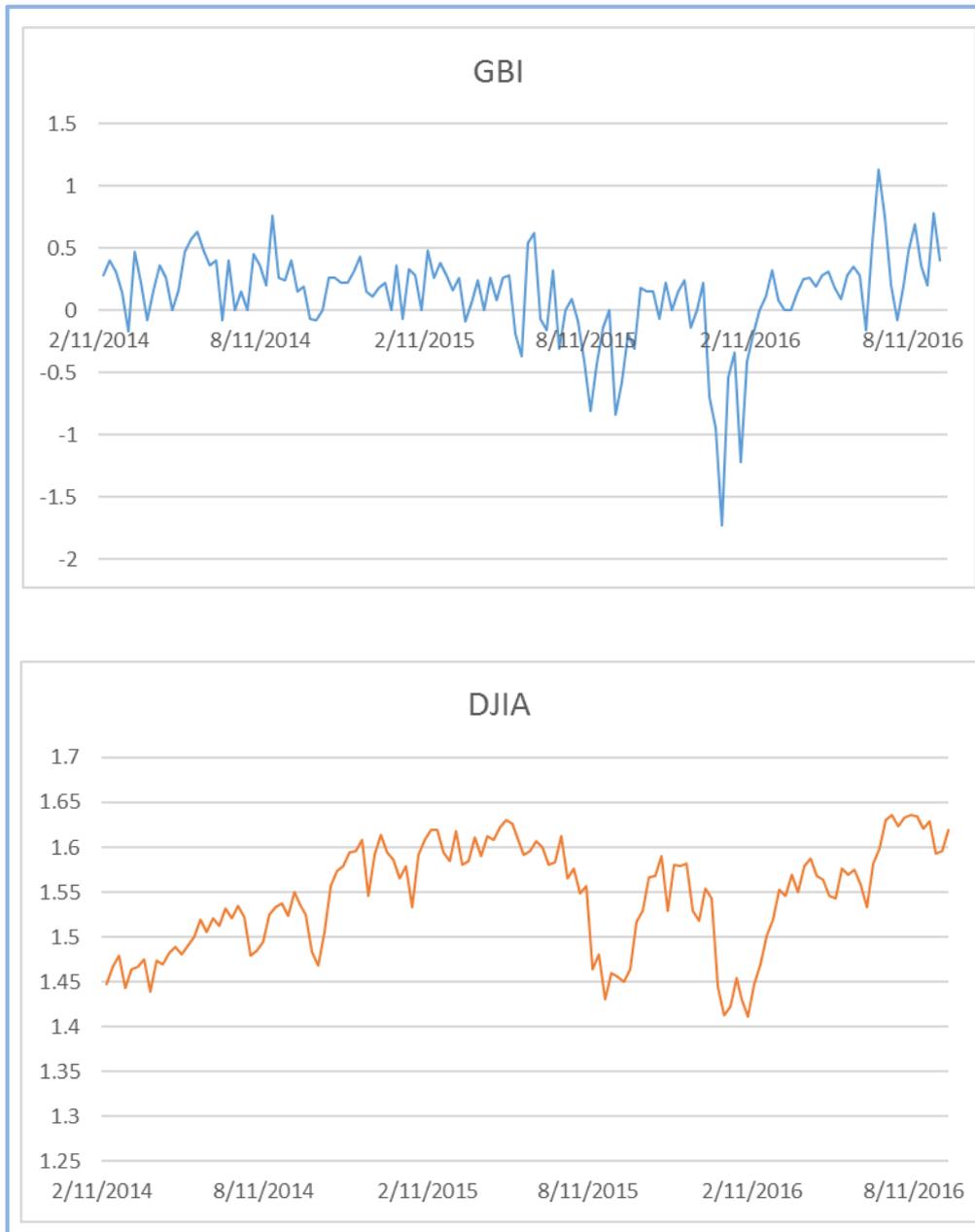
Figura 4.3



La normalizzazione consiste nel dividere i volumi di ricerca per il volume massimo osservato all'interno del campione, una volta effettuata la divisione i valori vengono espressi in unità percentuali, questo vuol dire che l'osservazione maggiore di coefficiente 1 ottenuta come divisione per sé stessa, sarà pari a 100 mentre tutte le altre avranno un coefficiente minore approssimato ad un numero intero.

Questa trasformazione non distorce in nessun modo l'indicatore di *sentiment* aggregato ottenuto con l'equazione 7, poiché esso si basa sulla proporzione fra i volumi e non sui relativi moduli. Dalla figura 4.3 è visibile anche la localizzazione delle ricerche, è interessante notare come sebbene la maggior parte delle ricerche riguardino paesi anglofoni o con un bilinguismo che comprende la lingua inglese, esistano volumi di ricerca rilevanti anche in paesi terzi.

Grafico 4.7



Proseguendo l'analisi quindi si calcola il *Google Bullishness Index* per il periodo considerato. Rispetto al caso precedente questo passaggio è molto più semplice ed automatico, poiché sebbene esistano differenze nelle festività nazionali, l'indice raccoglie i volumi settimanali, direttamente confrontabile con i rendimenti settimanali di qualsiasi indice di mercato.

Così come nel paragrafo precedente, il grafico 4.7 riporta la serie temporale del' *Google Bullishness Index* affiancata a quella del *Dow Jones Industrial Average*.

Anche in questo caso, così come nel *paper* di riferimento è apprezzabile una certa concordanza fra i movimenti delle due variabili. Il coefficiente di correlazione di Pearson infatti è pari a 0.502 per il periodo considerato.

4.2 STRUTTURAZIONE STRATEGIA

Utilizzo Bullishness Index e modellizzazione strategia di investimento

Una volta scaricati i dati e fatto le opportune correzioni per i rispettivi indici inizia la vera fase sperimentale di questa tesi. Sebbene ci siano diversi riferimenti sull'effettiva possibilità di utilizzo della *sentiment analysis* nelle strategie di investimento come ad esempio il lavoro di Oh e Sheng del 2011 non esiste un lavoro che utilizza questa metodologia per creare una strategia di investimento.

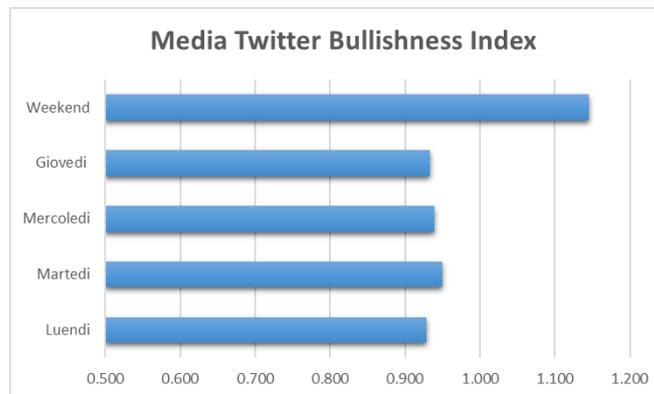
Per modellare una strategia di investimento ho pensato inizialmente di suddividere il campione in due sotto periodi. Uno contenente i primi due mesi di osservazioni, un altro di applicazione dei successivi 4 mesi.

Come primo passo ho analizzato dettagliatamente l'indice ottenuto per cercare di identificare qualche *patterns* fra le varie osservazioni.

Effettuando un raggruppamento infragruppo fra le varie misure di *sentiment* raggruppato per isoli giorni antecedenti ad i giorni di negoziazione si nota una prima differenza. Il Grafico 4.8 riporta la media dell'indice nei diversi giorni. Si osserva come la misura di *sentiment* aggregato nel fine settimana tende ad avere una media maggiore di quanto lo sia nei diversi giorni. Questo fenomeno è a parer mio ricollegabile a due principali fattori. Il primo, la misura aggregando le informazioni ottenute durante l'intero *weekend* risente della differenza "fisiologia" fra il volume di *tweets* rialzisti e quelli ribassi distorcendo l'indice verso l'alto. Una seconda, è possibile che essendo la misura del *sentiment* sintetizzata da un accorpamento di tre gironi diversi, questa sia maggiore in modulo rispetto a quella degli altri giorni proprio perché essa rispecchia l'aspettativa di rendimento per tre giorni e non per uno.

Come è possibile notare però non esistono differenze in media rilevanti fra i diversi giorni, inoltre è interessante osservare come in tutti i giorni l'indici si assesti ad un valore medio di circa 0.93.

Grafico 4.8



Questo valore indica, attraverso l'inversa dell'equazione 7 un rapporto fra il volume di messaggi *bullish* ed i messaggi *bearish* pari circa a 2.53.

Dopo aver appurato che non esistono differenze in media del TBI rilevanti per i diversi giorni, ho tentato di identificare l'esistenza di una relazione fra il valore dell'indice di riferimento ed i valori di borsa.

Ogni osservazione del *Twitter Bullishness Index* viene quindi affiancata ad una variabile *dummy* che riporta un coefficiente binario [0,1] che identifica cambiando valore se l'osservazione è seguita da un giorno in cui l'indice di riferimento ha avuto uno un rendimento negativo o positivo. I risultati ottenuti (come per ogni applicazione in questo paragrafo riporto graficamente i soli risultati ottenuti per il DJIA se questi sono condivisi con gli altri indici) non sono incoraggianti e sono riassunti nella tabella 4.1.

Tabella 4.1

	Osservazioni	Media TBI	Media DJIA	Dev.St
Positivi	26	1.00581	0.427%	0.258532
Negativi	17	0.96586	-0.392%	0.263223
Totale	43	0.98165	0.1030%	0.258012

Dalla tabella è evidente come non esista una differenza statisticamente significativa fra le due classi di osservazioni, inoltre si nota come la variabilità fra le classi sia simile fra loro e simile a quella della distribuzione totale. Inoltre conducendo un test di Fischer

classico sulla varianza fra i due gruppi si ha l'output seguente che conferma l'impossibilità di rifiutare l'ipotesi che le due variabili abbiano la stessa varianza.

	<i>Positivi</i>	<i>Negativi</i>
Media	0.965862466	1.005812
Varianza	0.066838577	0.069286
Osservazioni	26	17
gdl	25	16
F	0.964673365	
P(F<=f) una coda	0.455445086	
F critico una coda	0.483304806	

Questo risultato aggiunto al precedente non permettono di trovare le caratteristiche necessarie nelle variabili per costruire una strategia di trading basata su un valore soglia dell'indice di *sentiment*. Qualora i centroidi delle due classi di osservazioni avessero espresso una diversità apprezzabile sarebbe stato possibile costruire attraverso una analisi discriminante di Fischer un valore soglia che potesse essere usato come riferimento per un eventuale strategia di trading.

Per eliminare completamente questa strategia al dal paniere delle possibili applicazioni, ho effettuato una riclassificazione del campione completo di osservazioni (quindi sia il periodo di osservazione che quello di applicazione). Ho deciso di inserire nel set informativo tutte le informazioni disponibili in modo da osservare le differenze che osserverei fra le diverse classi durante l'ultimo giorno di applicazione del modello.

Per evidenziare ulteriormente le differenze ho diviso i valori del *Twitter Bullishness Index* rispetto al rendimento ottenuto dal indice di mercato di riferimento in 3 classi.

Come primo *step* ho calcolato la media dei rendimenti dell'indice di mercato nei giorni di borsa con chiusura positiva ed in quelli con chiusura negativa. Come secondo *step* ho suddiviso in tre classi, *DOWN*, *NEUTRAL*, *UP*. La prima classe indica le osservazioni seguite da rendimenti negativi inferiori alla media della classe, l'ultima invece le osservazioni di rendimenti positivi maggiori della media della classe, la classe *NEUTRAL* invece comprendeva le osservazioni restanti. Se fosse possibile applicare una strategia di questo tipo, sarebbe consono aspettarsi che avendo un set informativo completo esistessero differenze in media rilevanti fra queste classi appositamente costruite. La tabella

4.2 riporta i risultati ottenuti per il DJIA (i risultati sono simili per gli altri due indici di mercato).

Tabella 4.2

	<i>Neutral</i>	<i>Down</i>	<i>Up</i>	<i>Totale</i>
Conteggio	82	19	29	130
Media	0.97633	0.84841	0.985714	0.959725
Standard Dev.	0.26998	0.403641	0.195064	0.280916
Max	1.73783	1.609438	1.354767	1.737834
Min	0.42528	0.065926	0.481756	0.065926

Guardando la seconda riga della tabella 4.2 si nota immediatamente come non esista una differenza in media significativa fra le diverse classi. Sebbene l'ordine delle medie dell'indice rispecchia la relazione esistente con la categoria non è possibile dire che esistono differenze statisticamente significative fra le classi, la classe *Down* esprime una media inferiore di circa 0.14 rispetto alle altre due che sono praticamente identiche. Si nota però che questa differenza non è statisticamente significativa anche senza effettuare test poiché la differenza espressa corrisponde a circa un terzo della standard *deviation* per la medesima classe. Questa evidenza conferma l'impossibilità di applicare una strategia basata sulla *cluster analysis* e la determinazione dei valori soglia per modellare strategie di investimento.

I risultati ottenuti fin ora mi hanno indotto a cambiare completamente approccio per trovare una strategia di investimento basata sulla grandezza degli indicatori.

Abbandonando l'approccio precedente, ovvero di identificare *patterns* in un sotto periodo iniziale per poi modellarli in una strategia di investimento da adottare nelle restanti osservazioni, ho deciso di modellizzare una strategia sulla base delle evidenze empiriche più rilevanti ottenuti negli studi riportati in questo testo.

Ne capitolo 2, durante la presentazione dei risultati ottenuti da Sprenger e Welpé ho riportato nella figura 2.1 l'evidenza empirica ottenuta dagli autori effettuando un *event study*. Come ho riportato, la metodologia utilizzata è tratta dal lavoro più datato di Tomaini e White law (2001). Rimando al paragrafo 2.2 per le evidenze riportate dall'*event study* in questione, in questa sede però voglio evidenziarne le caratteristiche di base. Gli

autori impongono come regola di identificazione dell'*event day*, per massimizzare la probabilità di conferma della tesi, una caratteristica precisa: affinché un'osservazione possa rientrare nell'evento (oltre le imposizioni sui volumi di *tweets* non applicabili nel mio caso), l'indice di *sentiment* di riferimento deve eccedere di almeno due deviazioni standard la media delle osservazioni di tale indice calcolata sulle ultime 5 giornate di apertura di borsa per essere identificato come *BUY*, o ovviamente essere inferiore della stessa misura per essere identificato come *SELL*. Questa regola opportunamente affinata, può essere riapplicata al campione osservato per determinare anziché l'inserimento o meno in un *event study* per inserire o meno un ordine di acquisto o di vendita sul riferimento dell'indice.

Dai lavori di Bollen e Mao invece è emersa una relazione robusta anche se di tipo non causale fra il *Twitter Bullishness Index* ed il rispettivo mercato. Sebbene la non causalità questa relazione non possa essere utilizzata in modelli multifattoriali, l'evidenza empirica riscontrata può essere utilizzata per modellare una strategia di trading che la incorpori. La tabella 3.3 mostra, come già detto, il valore atteso in termini di *basis point* del rendimento dei diversi indici di mercato nei giorni successivi che si attende in seguito ad uno *shock* pari ad una deviazione standard nell'indice di mercato.

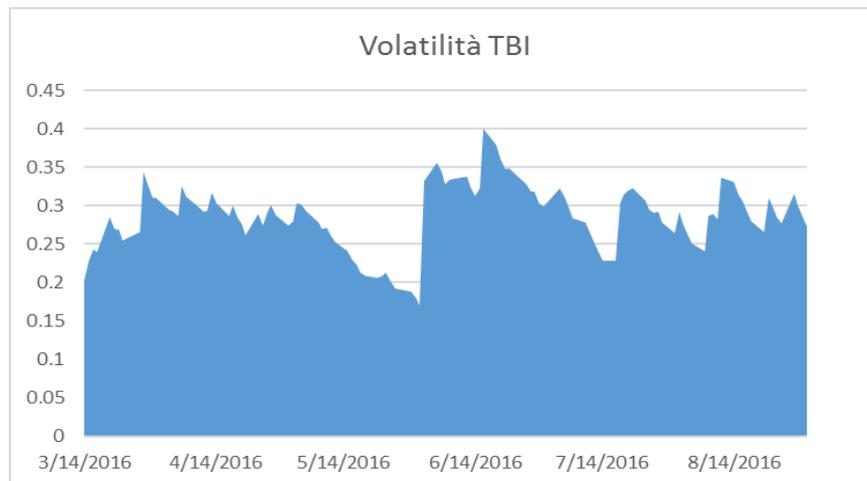
La strategia di investimento che tesserò sul campione di richiama entrambe le metodologie citate applicando un modello *exponential weighted moving average* per la stima della volatilità σT_t^B con equazione 12, mentre la strategia di investimento è riassunta nella formula 13.

$$(12) \quad \sigma^2 T_t^B = \lambda \sigma^2 T_{t-1}^B + (1 - \lambda) \left(T_t^B - \sum_{n=1}^5 \frac{T_{t-n}^B}{5} \right)^2$$

La stima della volatilità del *Twitter Bullishness Index* necessita di un modello che si adatti prontamente alle variazioni dell'indice T_t^B , rispettando la struttura infrasettimanale dei volumi osservati nei paragrafi precedenti, al fine di definire la volatilità vengono usate le sole ultime 5 osservazioni per un parametro $\lambda = 0.90$.

Il grafico 3.12 riporta la stima della volatilità ottenuta mediante l'applicazione dell'equazione 12 per il *Twitter Bullishness Index* di DJIA.

Grafico 4.9



Dal grafico 4.9 è visibile come il livello volatilità del TBI non sia costante durante il periodo osservato, inoltre si nota come i picchi di volatilità dovuti a *shock* nel *sentiment index* tendono a riassorbirsi nel corso della successiva settimana di negoziazione.

L'idea di base per modellare la strategia di trading è proprio basata su questa evidenza. La metodologia proposta in MBC2015 ci permette di estrarre un indicatore che rientra nella famiglia di strumenti per il *nowcasting*, ovvero che può essere usato per sfruttare la relazione a brevissimo termine con il fenomeno sottostante a cui è legato. Poiché la relazione è basata sugli *shock* e non sul modulo della misura di *sentiment*, la regola di selezione di Tumarkin e Whitelaw necessita di un accorgimento. Anziché selezionare le osservazioni in base alla differenza fra il valore corrente e quello della media, in maniera molto più coerente con la teoria esposta nel paragrafo 3.2, ho deciso cambiare il criterio di selezione in quanto non siamo interessati alla differenza fra il valore osservato e la media settimanale, ma piuttosto alla differenza fra le ultime due osservazioni e la volatilità giornaliera osservata nell'ultima settimana.

La strategia che chiameremo S viene applicata sull'osservazione T_t^B dei rispettivi indici di mercato, ed in base alle condizioni della formula 13 definisce se entrare *LONG*, ovvero in una posizione lunga sull'indice di riferimento per il successivo giorno di negoziazione. Oppure *SHORT*, quindi entrare in una posizione corta sull'indice di mercato, oppure per esclusione *WAIT* cioè non effettuare nessuna operazione di trading per quella data.

$$(13) \quad S(T_t^B) = \begin{cases} \text{LONG} & \text{se } T_t^B - T_{t-1}^B > 2 * \sigma T_{t-1}^B \\ \text{SHORT} & \text{se } T_t^B - T_{t-1}^B < -2 * \sigma T_{t-1}^B \\ \text{WAIT} & \text{se } |T_t^B - T_{t-1}^B| < 2 * \sigma T_{t-1}^B \end{cases}$$

Nel prossimo paragrafo vedremo la performance di investimento per questa strategia rispetto ai diversi mercati inclusi nella applicazione. Inoltre la strategia S sarà applicata, con una piccola opportuna calibrazione, anche al *Google Bullishness Index* per verificare se esista davvero una differenza di performance significativa che rispecchi la differenza in termine di potere previsionale evidenziata dalla letteratura.

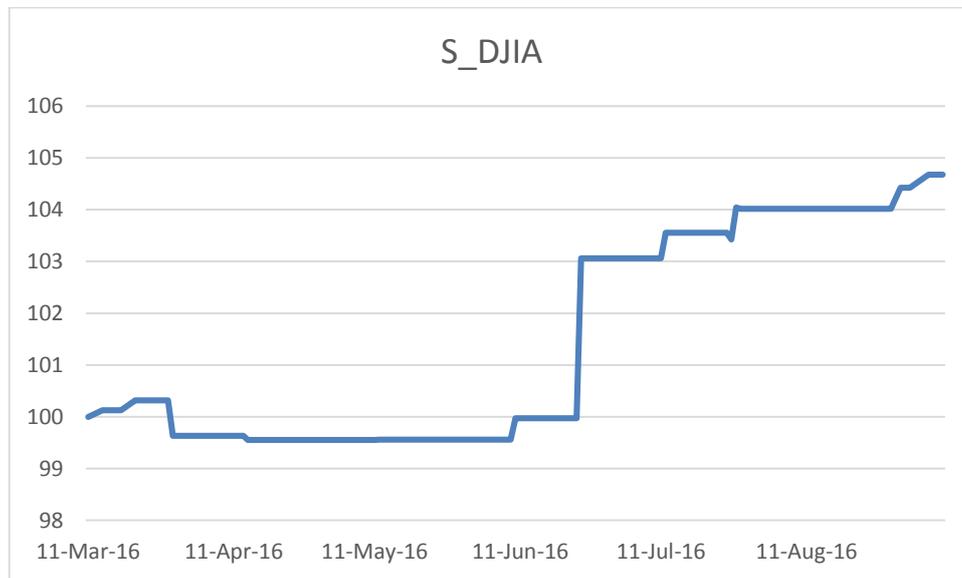
4.3 ANALISI PERFORMANCE INVESTIMENTO

La strategia S riportata nel paragrafo precedente viene applicata ai tre mercati analizzati ed al *Google Bullishness Index* come variabile di controllo qualitativa in quanto i risultati ottenuti non sono quantitativamente confrontabili se non per il numero di osservazioni. Indicherò da ora in poi la strategia S applicata per esempio sul FTSE100 attraverso la notazione sintetica $S_{FTSE100}$, mentre per differenziare l'applicazione che utilizza i volumi di ricerca di Google utilizzerò la notazione S_{Google} .

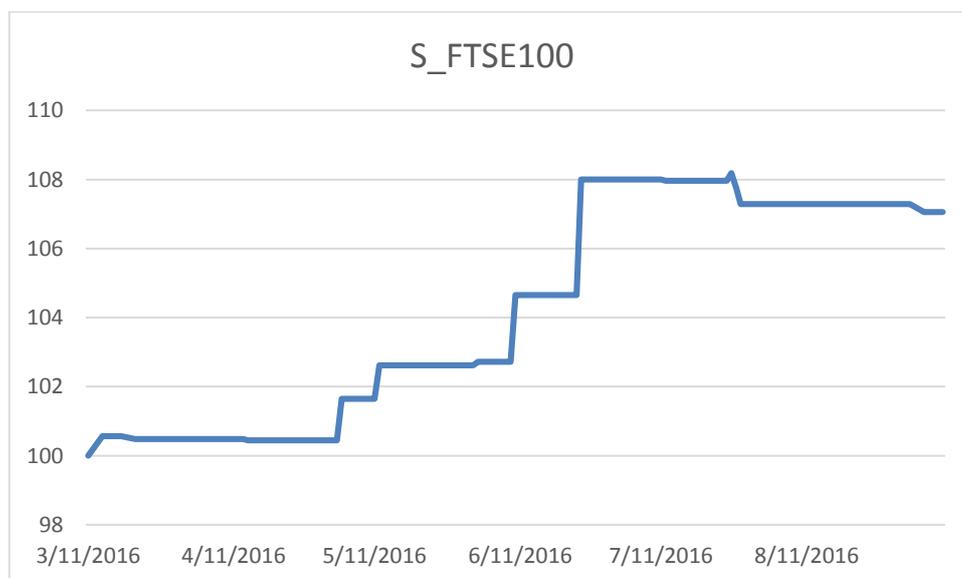
Per dare una raffigurazione più facilmente interpretabile nella figura 3.8a sono riassunti le performance ottenute applicando la strategia S ai diversi indici partendo da un investimento iniziale di 100€ e considerando l'ipotesi che non esistano costi di commissione.

Il grafico della figura 4.4a riguarda le performance di S_{DJIA} . Si nota immediatamente come la capacità predittiva del modello conduca ad un risultato economico positivo.

Il rendimento registrato nel semestre di osservazione dalla strategia di trading sul *Dow Jones Industrial Average* è pari al 4.68% ottenuto tramite posizioni in 14 giornate di negoziazione, 10 delle quali hanno mostrato rendimenti positivi, esprimendo una accuratezza di previsione del 71,4%.

Figura 4.4a

La figura 4.4b riporta le performance della $S_{FTSE100}$, questa applicazione ha dato in termini di rendimento assoluto la performance migliore, realizzando un rendimento semestrale del 7,05%, questo dato però è accompagnato da una accuratezza di previsione più bassa

Figura 4.4b

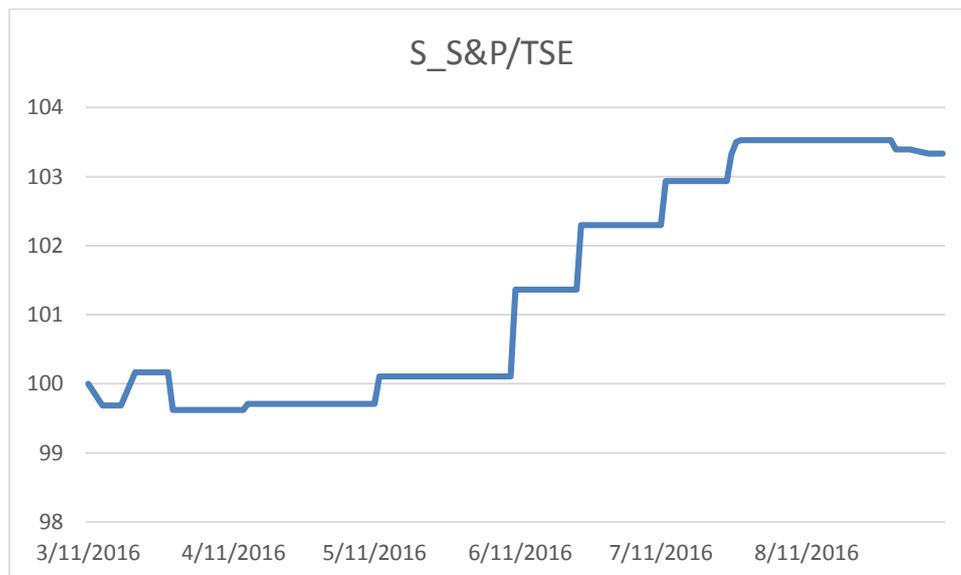
La $S_{FTSE100}$ nel periodo osservato realizza in tutto 13 posizioni giornaliere sull'indice FTSE100, 8 delle quali si dimostrano profittevoli in termini economici, realizzando quindi nel periodo analizzato un'accuratezza di previsione del 61,5%.

La figura 4.4c invece riporta le performance di $S_{S\&P/TSE}$, anche in questo caso la strategia ha riportato nei 6 mesi di osservazioni un rendimento positivo pari al 3,34%, inferiore alle altre due applicazioni sopra elencate. Il livello di accuratezza in questo caso si pone a metà fra le applicazioni illustrate precedentemente, realizzando una previsione esatta nel 69,3% delle 13 volte in cui è entrato in posizioni sull'indice.

Prima di riportare i risultati ottenuti applicando la S_{Google} , è possibile apprezzare la similitudine dei grafici finora commentati. Questa deriva dal fatto che nonostante le variazioni degli indici siano diverse, ed esistano giorni di chiusura particolare che vanno a modificare la serie storica (quindi anche la deviazione standard), la gran parte della serie storica del *Twitter Bullishness Index* calcolato per i sei mesi di applicazione è condivisa. Questo implica che in molti casi le differenze rilevate nelle performance siano riconducibili alle differenze nelle variazioni fra gli indici di mercato piuttosto che alle differenze fra i vari TBI ricalcolati.

Infine la figura 4.4d riporta la performance ottenuta da S_{Google} ai fini di un confronto puramente qualitativo.

Figura 4.4c



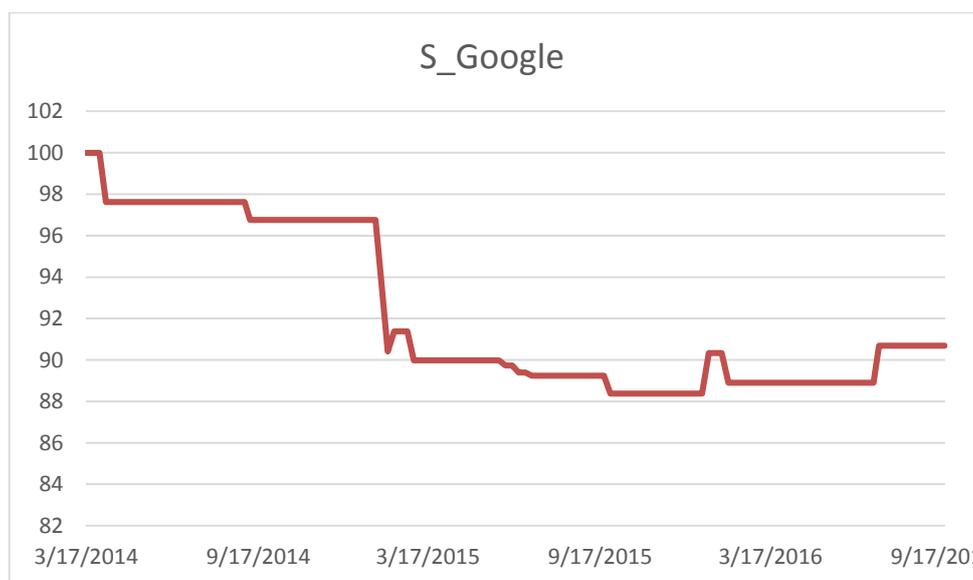
Sebbene il periodo di riferimento non sia lo stesso, l'applicazione della strategia attraverso il *Google Bullishness Index* è stata modellata in modo da avere un numero di osservazioni utili ai fini di investimento identico alle tre sopra elencate.

L'unica differenza tecnica nelle due strategie sta nel fatto che questa, ai fini del calcolo della volatilità, consideri nel modello EWMA riportato della formula 12 una media che comprende le ultime 4 osservazioni e non 5 (coerentemente con la letteratura riportata nel paragrafo precedente).

L'applicazione della attraverso la misurazione del *sentiment* misurato settimanalmente non riporta un risultato economico positivo. Il rendimento ottenuto nell'arco di tempo analizzato è pari al -9,31%, con un'accuratezza di previsione di solo il 23,07%. Sebbene questi risultati non siano direttamente confrontabili vista la grande differenza nel periodo di osservazione è apprezzabile la coerenza fra questa applicazione e le evidenze empiriche riportate nel *paper* di riferimento Mao Bollen e Counts del 2015.

A fini puramente illustrativi, ho pensato di riproporre la S_{Google} in una maniera diversa rispetto a quanto fatto precedentemente. Come evidenziato sia nel *paper* di riferimento che all'inizio di questo paragrafo esiste una forte relazione fra il *Google Bullishness Index* e le variazioni di mercato (in questa applicazione il DJIA) settimanali osservate in assenza di *lag*.

Figura 4.4d



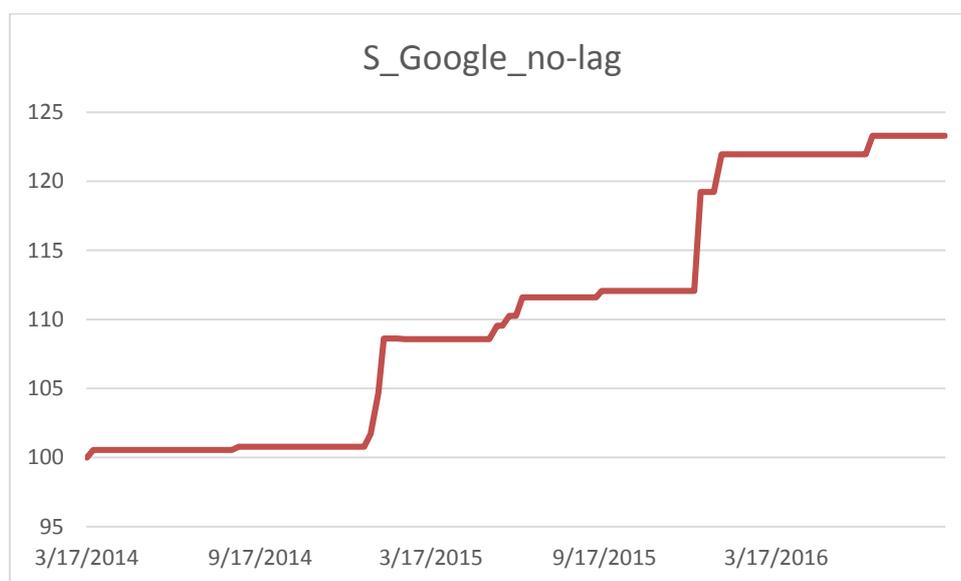
Nella figura 4.5 ho riportato le performance del risultato ottenuto da S_{Google} azzerando il *lag* fra l'osservazione della misura del *sentiment* e l'investimento ($S_{Google-no_lag}$). Sebbene con i dati accessibili al pubblico ad oggi sia materialmente impossibile osservare la formazione dei volumi di ricerca in tempo reale, è interessante notare che qualora avessimo accesso a questo tipo di dati sui volumi ricerche di Google, i risultati che otterrebbe questa strategia tenderebbero molto di più ad avvicinarsi a quelli ottenuti nella figura 4.5 che in quella 4.4d.

La figura 4.5 evidenzia un risultato economico diametralmente opposto rispetto a quello ottenuto in precedenza, il rendimento di questa applicazione virtuale è del 23,9% con una strabiliante accuratezza di previsione del 92,3%.

Concludo questa analisi con una tabella 4.3 che riassume i risultati ottenuti da questa strategia di trading attraverso posizioni *long* e *short* sui diversi indici di mercato.

I risultati ottenuti nella prima sezione della tabella riguardano le varie applicazioni della strategia S sui diversi indici, la seconda parte invece riporta fini prettamente comparativi i risultati ottenuti dai diversi indici nello stesso periodo (marzo 2016 – settembre 2016) rispettando i criteri sopra citati, come tasso *risk free* è stato usato il LIBOR a 6 mesi osservato agli inizi di marzo 2016.

Figura 4.5



Dalla prima riga è possibile notare come la strategia di trading ottenga risultati inferiori in termini di rendimento puro rispetto agli indici su cui opera.

Questo risultato però, come qualsiasi altro rendimento di natura finanziaria, non va letto ed interpretato singolarmente.

Tabella 4.3

<i>T=6 mesi</i>	<i>Strategia S</i>			<i>Benchmark</i>		
<i>rf=0.892%</i>	DJIA	FTSE100	S&P/TSE	DJIA	FTSE100	S&P/TSE
Rendimento	4.677%	7.056%	3.333%	6.267%	9.808%	6.146%
Rend.Giornaliero	0.037%	0.056%	0.026%	0.049%	0.078%	0.049%
σ	3.301%	4.077%	1.934%	6.804%	11.173%	5.685%
σ Giornaliero	0.302%	0.382%	0.166%	0.604%	0.995%	0.506%
DSR	0.698%	0.642%	0.642%	4.463%	7.347%	3.385%
DSR Giornaliero	0.062%	0.057%	0.057%	0.396%	0.654%	0.302%
Media Positivi	0.554%	1.157%	0.482%	0.420%	0.713%	0.389%
Media Negativi	-0.230%	-0.200%	-0.261%	-0.418%	-0.711%	-0.420%
Accuratezza	71.429%	61.538%	69.231%	-	-	-
Sortino Ratio	5.42	9.60	3.80	1.20	1.21	1.55
Sharpe Ratio	1.147	1.512	1.262	0.790	0.798	0.924

In termini di rischio tutte le applicazioni hanno prodotto una volatilità dei rendimenti inferiore rispetto ai *benchmark* di riferimento, $S_{FTSE100}$ e $S_{S\&P/TSE}$ in particolare esprimono una volatilità nel periodo di riferimento circa 3 volte inferiore rispetto a quella del mercato di riferimento. Continuando l'analisi, il *down side risk* calcolato per le strategie risulta essere nettamente inferiore rispetto agli indici di mercato. Questa sostanziale differenza è deducibile dalle differenze in media fra le giornate a rendimenti positivi e negativi che sussistono fra i rendimenti ottenuti attraverso S_{DJIA} , $S_{FTSE100}$ e $S_{S\&P/TSE}$ e quelli ottenuti dai rispettivi *benchmark*. Per esempio i rendimenti giornalieri positivi (negativi) ottenuti da $S_{FTSE100}$, sebbene sia l'applicazione della strategia che riporti l'accuratezza minore, sono in media 0.44% superiori (0.511% inferiori in modulo) rispetto alle stesse classi osservate nel FTSE100, questa differenza notevole ci dice che in pratica la $S_{FTSE100}$ sia nel caso in cui restituisce una previsione corretta, ma soprattutto

nel caso in cui la previsione si rivela sbagliata, genera in media rendimenti migliori di quelli del *benchmark* di riferimento.

Infine la tabella riporta lo *Sharpe ratio* ed il *Sortino ratio* calcolati utilizzando il tasso LIBOR con scadenza a 6 mesi osservato a marzo 2016, nel momento in cui è partita l'applicazione. Anche in questo caso le misure di *performance risk-adjusted* danno un valore nettamente superiore ad S_{DJIA} , $S_{FTSE100}$ e $S_{S\&P/TSE}$ piuttosto che ai *benchmark* di riferimento.

I valori di queste due misure di performance indicano che a parità rischio, misurato nello *Sharpe Ratio* con la *standard deviation* e con il *down side risk* nel *Sortino Ratio*, i risultati ottenuti dalle tre applicazioni sono migliori in termini di rendimento in eccesso al tasso *risk free* considerato.

È infine interessante notare come entrambi i rapporti calcolati per S_{DJIA} , $S_{FTSE100}$ siano nettamente superiori e diversi fra loro rispetto ai benchmark di riferimento che presentano sia uno *Sharpe Ratio* che un *Sortino Ratio* molto simile. Questa diversità è motivata prevalentemente da due fattori. Il primo, già citato in precedenza, la differenza in media fra le realizzazioni positive e soprattutto negative, più marcata per la $S_{FTSE100}$ rispetto che per il S_{DJIA} . Il secondo, per il diverso tasso di accuratezza di previsione; maggiore per la S_{DJIA} , che rende più vicine le due misure di *performance risk adjusted* di quanto lo sarebbero se presentassero qualità previsionali identiche.

4.4 CONCLUSIONI FINALI

I risultati ottenuti dall'applicazione della strategia *S* sono incoraggianti, soprattutto se confrontati utilizzando come benchmark gli stessi indici di mercato utilizzati come riferimento della strategia. L'applicazione proposta nel capitolo precedente dovrebbe indurmi a dare una risposta affermativa decisa a tutte e tre le domande di ricerca che mi sono poste durante questa tesi sperimentale.

Nonostante abbia svolto io stesso ogni passo di questa ricerca che ha impegnato la gran parte del mio tempo negli ultimi mesi, non riesco a frenare la mia modalità di osservazione critica che da sempre mi ha accompagnato, quindi risponderò alle domande di ricerca evidenziando i dubbi e le debolezze che le accompagnano nonostante sia stato io stesso l'artefice dei risultati.

Prima di rispondere alla **D1**, nella quale mi chiedevo se è possibile utilizzare la *sentiment analysis* per modellare una strategia di investimento a breve termine, preferisco rispondere prima alle domande successive in modo da fornire poi una risposta più completa e motivata.

Nella seconda domanda di ricerca, **D2**, mi chiedevo in che modo potevano essere sfruttate le evidenze empiriche per creare una strategia di trading. La strategia *S* che ho proposto nella applicazione risponde solo in parte alla domanda, sebbene conduca comunque a risultati positivi. Come ho motivato nel capitolo precedente nel paragrafo di strutturazione, ogni dettaglio della strategia *S* è motivato da un'evidenza empirica o una deduzione logica. Ci sono però evidenze riscontrate nella letteratura, come ad esempio la relazione fra l'indice di *sentiment* e i rendimenti con $lag > 1$ di tipo *contrarian*. che questa strategia non considera.

Nella tabella 3.3 i coefficienti riportati mostrano l'evidenza della relazione fra il TBI e le osservazioni del FTSE con lag da 1 a 4 e per il S&P/TSE con lag 1 e 4. La strategia *S* che ho adottato nella applicazione tiene conto della relazione di lag 1 mediante una sorta di test delle ipotesi ripetuto che ne catturi solo le variazioni statisticamente più rilevanti. Nulla è modellizzato per gli altri lag .

È possibile riscontrare un'evidenza che questa mancata implementazione influisca negativamente sulle performance in casi in cui la Strategia *S* segnali una posizione *LONG* o

SHORT in due o più giorni successi di negazione. Dando uno sguardo più attento ai grafici di performance 4.4 è possibile notare che nelle date 26-27-28 luglio tutte e tre le applicazioni S_{DJIA} , $S_{FTSE100}$ e $S_{S\&P/TSX}$ segnalavano delle posizioni LONG o SHORT sui rispettivi indici di mercato. Le posizioni prese successive alla prima (quindi per i giorni 27-28 luglio) si sono dimostrate sbagliate 3 volte su 6, in particolare per la $S_{FTSE100}$ risultano errate 2 volte su 2. Questo risultato, a mio parere, indica che la strategia S'è un buon punto di partenza in quanto ci restituisce risultati soddisfacenti ma per un utilizzo reale può essere migliorata per tener conto delle relazioni fra il *sentiment index* e le variazioni di mercato con *lag* maggiore di 1.

In merito alla domanda di ricerca **D3**, nella quale mi chiedevo se davvero ci fosse un interesse economico nel prediligere strategie di trading basate sulla *sentiment analysis* ad altre tradizionali, la tabella 4.3 sembra dare una risposta positiva considerando il rendimento medio della strategia pari al 5.03%. I risultati ottenuti sono sicuramente positivi, soprattutto in termini di performance *risk adjusted*.

La tabella però si basa sulla assunzione semplificatrice che non esistano costi di commissione per l'immissione degli ordini. Nella vita reale, tranne che per casi particolari, purtroppo non è così. È facilmente intuibile immaginare come l'introduzione di *transaction fees* possa notevolmente peggiorare le performance della strategia, in quanto in tutte e tre le applicazioni, indipendentemente dall'entità, la commissione dovrebbe essere pagata 13 o 14 volte. In linea con la maggior parte dei modelli teorici che ho studiato fin ora ho deciso arbitrariamente di non includere i costi di commissione, non nascondo però che qualora un giorno dovessi scegliere come investire il mio denaro rifarei i conti la maniera accurata prima di proiettarmi verso una strategia del genere.

Dopo aver risposto eloquentemente alle **D2** e **D3** posso finalmente rispondere alla domanda principale della tesi. La *sentiment analysis* può essere utilizzata per modellare strategie di investimento a breve termine? La mia risposta è **SI**, ma probabilmente è un investimento più adatto ad investitori istituzionali che *retail*.

Per motivare questa risposta non mi dilungherò ancora sull'effettiva utilità previsionale della *sentiment analysis* ai fini di investimento poiché è già riportata in ogni pagina di questo testo. La motivazione che mi induce a identificare un *target* istituzionale piuttosto che *retail* per questo tipo di investimenti è legata principalmente alla difficoltà di acquisizione, manipolazione ed osservazione dei *sentiment data*.

Per raccogliere i circa 750.000 *tweet* che ho utilizzato nella mia applicazione, non disponendo di algoritmi o programmi automatizzati né di particolari *skill* in campo informatiche, ho dovuto dedicare circa mezz'ora al giorno, tutti i giorni, dei miei ultimi sei mesi al *download*, la pulizia e l'archiviazione dei dati. Ammesso che esistano persone con capacità informatiche molto più avanzate delle mie, che riescano ad automatizzare il processo tramite un semplice click o facendo a meno anche di quello, il problema principale risiede nella fonte stessa di questo tipo di dati. L'archivio che ho costruito è generato da API che hanno un accesso di livello C, quindi rappresenta un'estrazione casuale dal database completo di solamente il 10% dei *tweets* pubblici per il periodo analizzato, ed è il livello massimo di informazioni ottenibili dalla clientela *retail*.

Anche accettando l'ipotesi forte che la totalità degli investitori condivida le proprie intenzioni finanziarie su Twitter, questo vuol dire che la misura di *sentiment* che ho ottenuto non è l'espressione della totalità degli investitori attivi su Twitter, ma di un campione casuale (risultante da un'estrazione non ben definita) contenente informazioni sul 10% di coloro che decidono di rendere pubblici i propri messaggi. Anche senza avere numeri alla mano è semplice immaginare che in realtà questo campione non sia sempre e per forza adatto a stimare la propensione rialzista o ribassista della totalità degli investitori.

Per quando riguarda i volumi delle ricerche di Google la situazione non migliora. A livello *retail* sono disponibili solo i volumi settimanali normalizzati delle *queries* di ricerca. Anche in questo caso a livello *corporate* invece esistono accessi privilegiati che consentono di osservare i volumi in maniera più dettagliata arrivando fino alle osservazioni orarie prive di normalizzazione.

Solo pochi clienti *corporate* possono avere ad un accesso di livello superiore, avendo a disposizione BIG DATA più fluidi, che non solo consentano di ottenere una *proxy* migliore del *sentiment* dell'intera collettività, ma permettano soprattutto di osservare le mutazioni dello stesso in tempo reale ed operare di conseguenza.

Possiamo immaginare l'aumento di performance ottenibile accorciando quanto più possibile il *lag* temporale fra osservazione dei cambiamenti nel *sentiment* e l'immissione di ordini di acquisto o vendita osservando il caso limite del *Google Bullishness Index No-Lag* riportato nei paragrafi precedenti.

Sebbene non esistano certezze sul futuro, ritengo molto probabile che la *sentiment analysis* possa essere sviluppata ed implementata nel *set* di strumenti standard disponibili per gli investitori istituzionali per modellare strategie di investimento alternative.

È anche vero però che progetti come l'attuale *Google Clouds Sentiment Analysis Tools*, lanciato lo scorso agosto, mi fanno ben sperare che in un futuro prossimo possa esistere un livello di informazione di maggiore qualità sui *social data* accessibile anche a livello *retail*. In tal caso ovviamente le considerazioni fatte in questo paragrafo diverrebbero superflue, poiché esisterebbero informazioni di elevata qualità accessibili a chiunque abbia voglia e intenzione di modellare strategie per il *nowcasting* nei mercati finanziari.

Riferimenti Bibliografici

Pag 1 – Erik Cambria e Amir Hussain, *Sentic Computing: Techniques, Tools, and Applications*, Springer, 2012.

Pag. 5 – Lorenzo Pascucci, *Big Data: cosa sono*, webmasterpoint.org.

Pag. 5 – *Economist, Data, data everywhere*, economist.com.

Pag. 8 – *Evoluzione dei fenomeni sociali attraverso la rete* – Careon 2011

Pag 10 – *Social Media e Sentiment Analysis*. Curini -Iacus - 2012

Pag 13 -Asur Huerban , *Predicting the future with the social media*- 2001

Pag 13 - Uz Zaman – *Extraxting aggregate prediction from twitter*- 2012

Pag 14 – Paul Drezde – *You are what you tweet:analyzing tweets for public healt* - 2011

Pag 14 - Pang, Bo; Lee, Lillian (2008). "4.1.2 Subjectivity Detection and Opinion Identification". *Opinion Mining and Sentiment Analysis*. Now Publishers Inc.

Pag 15 - Thomas Dorsey, *Point and Figure Charting, Sentiment has a "66% influence on the overall movement of an individual stock"*

Pag. 16,18,19 - Da, Zhi; Engelberg, Joseph; Gao, Pengjie (2014-10-17). "The Sum of All FEARS Investor Sentiment and Asset Prices"

Pag 15 - Hou, Kewei; Xiong, Wei; Peng, Lin (2009-01-16). "A Tale of Two Anomalies: The Implications of Investor Attention for Price and Earnings Momentum". Rochester, NY: Social Science Research Network

Pag 15- Barberis, Nicholas; Shleifer, Andrei; Vishny, Robert W. (1997-02-01). "A Model of Investor Sentiment". National Bureau of Economic Research.

Pag 16 - "Consumer Confidence Index® | The Conference Board". www.conference-board.org. Retrieved 2016-04-26.

Pag 16 - Whaley, Robert E (2000-03-01). "The Investor Fear Gauge". *The Journal of Portfolio Management*

Pag 16 -Barber, Brad M.; Odean, Terrance (2008-04-01). "All That Glitters: The Effect of Attention and News on the Buying Behavior of Individual and Institutional Investors"

Pag 17 - Brown, Gregory W.; Cliff, Michael T. (2005-01-01). "Investor Sentiment and Asset Valuation". *The Journal of Business*

Pag 18 - Zhang, Xue; Fuehres, Hauke; Gloor, Peter A. (2011-01-01). "Predicting Stock Market Indicators Through Twitter "I hope it is not as bad as I fear"

Pag 19 - Peterson, Richard (2016-03-21). *Trading on Sentiment: The Power of Minds Over Markets*

Pag 20 - Dimpfl, Thomas; Jank, Stephan (2012-06-06). "Can Internet Search Queries Help to Predict Stock Market Volatility?"

Pag 20 - Hirshleifer, David; Shumway, Tyler (2003-01-01). "Good Day Sunshine: Stock Returns and the Weather". *The Journal of Finance*.

Pag 20 - Cao, Melanie; Wei, Jason (2005-06-01). "Stock market returns: A note on temperature anomaly"

Pag 20 - Robotti, Cesare; Krivelyova, Anya (2003-10-01). "Playing the Field: Geomagnetic Storms and the Stock Market"

Pag[20-40] Sprenger, Timm O. and Welppe, Isabell M., *Tweets and Trades: The Information Content of Stock Microblogs* (November 1, 2010). Available at SSRN: <http://ssrn.com/abstract=1702854> or <http://dx.doi.org/10.2139/ssrn.1702854>

Pag 21 - BusinessWeek, 2009, *StockTwits may change how you trade*, BusinessWeek, Online Edition (author Max Zeledon), February 11

Pag 21 - Bloomberg, 2010, *Hedge Fund Will Track Twitter to Predict Stock Moves*, Bloomberg, Online Edition (author Jack Jordan), December 22.

Pag 21 - Vincent, Arnaud, and Armstrong Margaret, 2010, *Predicting break-points in trading strategies with Twitter*, Working paper, École Nationale Supérieure des Mines de Paris.

Pag[22,27,29,34,38, 67, 85] - Antweiler, Werner, and Frank, Murray Z., 2004, *Is all that talk just noise? The information content of internet stock message boards*, *Journal of Finance* 59, 1259-1294.

Pag 24 - Malkiel, Burton G., 2003, *The efficient market hypothesis and its critics*, *The Journal of Economic Perspectives* 17, 59-82.

Pag 24 - Tetlock, Paul C., Saar-Tsechansky, Maytal, and Macskassy, Sofus, 2008, *More than words: Quantifying language to measure firms' fundamentals*. *The Journal of Finance* 63, 1437-1467

Pag 24 - Fama, Eugene F., 1991, *Efficient capital markets: II*, *Journal of Finance* 46, 1575–1617.

Pag 25 – O'Connor, Brendan, Balasubramanyan, Ramnath, Routledge, Bryan R., and & Smith, Noah, 2010, *From Tweets topolls: Linking text sentiment to public opinion time series*, *International Conference on Weblogs and Social Media*, 122-129, Washington, DC

Pag [29,67] - Singer, E. (2001), "The Use of Incentives to Reduce Nonresponse in Household Surveys", in Groves, R.M., Dillman, D.A., Eltinge, J.L. and Little, R.J.A. (eds.), *Survey Nonresponse*, pp. 163-178.

Pag 37 - Tumarkin, Robert, and Whitelaw, Robert F., 2001, *News or noise? Internet postings and stock prices*, *Financial Analysts Journal* 57, 41-51.

Pag 41-53 - *Twitter mood predicts the stock market* J Bollen, H Mao, X Zeng *Journal of Computational Science* 2 (1), 1-8

Pag 41 - Fama, Eugene F, e. a. (1969) *International Economic Review* 10, 1–21.

Pag 41- Fama, E. F. (1991) *Journal of Finance* 46, 1575–617.

Pag 41- H.Cootner, P. (1964) *The random character of stock market prices* (MIT)

Pag [42,84]- Kahneman, D & Tversky, Amos (1979) *Prospect Theory: An Analysis of Decision under Risk*. (*Econometrica*), pp. 263–291

Pag 42- Dolan, R. J. (2002) *Science* 298, 1191–1194.

Pag 44 -O'Connor, B, Balasubramanyan, R, Routledge, B. R, & Smith, N. A. (2010) *From Tweets to Polls: Linking Text Sentiment to Public Opinion Time Series*.

Pag 45,48- Brants, T & Franz, A. (2006) *Web 1t 5-gram version 1*, (*Linguistic Data Consortium, Philadelphia*), Technical report

Pag 46 -Frey, B. S. (2008) *Happiness: A Revolution in Economics*. (The MIT Press).

Pag 46 - Granger, C. W. J. (1969). "Investigating Causal Relations by Econometric Models and Cross-spectral Methods". *Econometrica*.

Pag 46 -Riloff, E, Wiebe, J, & Wilson, T. (2003) *Learning subjective nouns using extraction pattern bootstrapping*. (*Association for Computational Linguistics, Morristown, NJ, USA*), 25–32.

Pag 46 –Pang, B & Lee, L. (2008) *Foundations and Trends in Information Retrieval* 2, 1–135.

Pag 47 - McNair, D, Heuchert, J. P, & Shilony, E. (2003) *Profile of mood states*.Bibliography 1964–2002. (*Multi-Health Systems*).

Pag 48 -Norcross, J. C, Guadagnoli, E, & Prochaska, J. O. (2006) *Journal of Clinical Psychology* 40, 1270 – 1277.

Pag 50- Mao, H, Zeng, X.-J, Leng, G, Zhai, Y, & Keane, A. J. (2009) *IEEE Transaction on Power System*. 24, 1080–1090.

Pag 51 - Zhu, X, Wang, H, Xu, L, & Li, H. (2008) *Expert Syst. Appl.* 34, 3043–3054.

Pag 52 -A Self-Organizing Fuzzy Neural Networks H. S. LIN, X. Z. GAO, XIANLIN HUANG, AND Z. Y. SONG

Pag [52, 55, 74] - Granger, C. W. J. (1969), "Investigating Causal Relations by Econometric Models and Cross-Spectral Methods", *Econometrica*, Vol. 37, pp. 424-38.

Pag 52 - *NEURAL NETWORKS ,A Comprehensive Foundation Second Edition* Simon Haykin

Pag 66- 88, 90 - *Quantifying the Effects of Online Bullishness on International Financial Markets* Huina Mao , Scott Counts , and Johan Bollen

Pag 66 Shiller, R. J. (2006), *Irrational Exuberance*, Crown Business

Pag[67,70] - Oh, C. and Sheng, O. R. L. (2011), "Investigating predictive power of stock micro blog sentiment in forecasting future stock price directional movement", *ICIS 2011 Proceedings*, pp. 57-8.

Pag[67,74] - Tetlock, P. C. (2007), "Giving Content to Investor Sentiment: The Role of Media in the Stock Market", *The Journal of Finance*, Vol. 62, pp. 1139-68, June .

Pag 68 - Da, Z., Engelberg, J. and Gao, P. (2009), "The Sum of All FEARS: Investor Sentiment and Asset Prices", SSRN eLibrary.

Pag 74 - Gilbert, E. and Karahalios, K. (2010) "Widespread worry and the stock market" in *Proceedings of the fourth international AAAI conference on weblogs and social media*, Vol. 2, pp. 229-47.

Pag 78 - Shapiro, K. (ed.), (2001), *The Limits of Attention: Temporal Constraints in Human Information Processing*, Oxford University Press, New York.

Pag [79 ,84] De Long, B., Shleifer, A., Summers, L.H. and Waldmann, R. J. (1990a), "Noise Trader Risk in Financial Markets", *Journal of Political Economy*, Vol. 98, No 4, pp. 703-738.

Pag 84 - *Inefficient Markets: An introduction to behavioural finance* A Shleifer - 2000

Pag 87- *Six Provocations for Big Data* danah boyd Kate Crawford

Pag. 101- *Il Mercato Azionario Italiano: Efficienza e Anomalie di Calendario* – Emilio Barone - 1990

Pag -105 - Tumarkin, R. & Whitelaw, R. (2001) "News or noise? Internet postings and stock prices" *Journal of Financial Analyst* 57(3) 41-51

Pag 115 - Frank Alphonse Sortino, Stephen Satchell *Managing Downside Risk in Financial Markets: Theory, Practice and Implementation*, Butterworth-Heinemann, 2001, ISBN 0-7506-4863-5

Pag. 119 - https://cloud.google.com/prediction/docs/sentiment_analysis

Siti Web

- Pag. 18, 45 - <http://www.sca.isr.umich.edu/>
- Pag. 22- <http://stocktwits.com/>
- Pag. 27 - <http://www.cs.waikato.ac.nz/ml/weka/>
- Pag. 42 - <http://www.ccs.neu.edu/home/amislove/twittermood/>
- Pag. 45 - <http://www.cs.pitt.edu/mpqa/opinionfinderrelease/>
- Pag. 45 - www.gallup.com/poll/122840/Gallup-Daily-Economic-Indexes.aspx
- Pag. 59 - https://en.wikipedia.org/wiki/Bonferroni_correction
- Pag. 62 <http://www.telegraph.co.uk/technology/twitter/8755587/Twitter-becomes-latest-tool-for-hedge-fundmanagers.html>
- Pag. 62 - <http://www.dailymail.co.uk/sciencetech/article-2036499/Twitter-used-hedge-fund-managers-predict-share-prices.html>
- Pag. 62 - http://usatoday30.usatoday.com/money/perfi/stocks/2011-05-03-wall-street-traders-mine-tweets_n.htm
- Pag. 62 - <http://www.theatlantic.com/technology/archive/2010/10/predicting-stock-market-changes-using-twitter/64897/>
- Pag. 62- <https://www.wired.com/2010/10/twitter-crystal-ball/>
- Pag. 62 - <http://techland.time.com/2011/03/24/let-twitter-tell-you-where-to-invest-your-money/>
- Pag. 64 - <http://www.datameer.com/company/datameer-blog/tag/twitter-sentiment-analysis/>
- Pag. 64- <http://isentium.com/>
- Pag. 64 - <http://www.bloomberg.com/news/articles/2015-07-15/if-twitter-is-bullish-maybe-you-should-be-too>
- Pag. 70 - <http://www.trade-futures.com/dailyindex.php>
- Pag. 71- <http://www.cboe.com/micro/vix/vixintro.aspx>
- Pag. 80 - <https://passport.weibo.com/visitor/visitor?>
- Pag. 81- <https://www.google.it/trends/hottrends>
- Pag. 82- <http://www.baidu.com/login.php>
- Pag. 88 - <http://www.followthehashtag.com/>
- Pag. 89- <https://chrome.google.com/webstore/detail/twitter-archiver/pkan-pfekacaojdncfjadedbgbbphi>
- Pag. 114- <http://www.global-rates.com/interest-rates/libor/libor.aspx>
- Pag. 114 - https://it.wikipedia.org/wiki/Indice_di_Sortino
- Pag. 119 - https://cloud.google.com/prediction/docs/sentiment_analysis

LIBERA UNIVERSITÀ INTERNAZIONALE DEGLI STUDI SOCIALI

“LUISS - GUIDO CARLI”



FACOLTÀ DI ECONOMIA

Corso di laurea in Banca ed Intermediari Finanziari

**SOCIAL SENTIMENT ANALYSIS
E INVESTIMENTI AZIONARI**

RIASSUNTO

RELATORE:

CHIARISSIMO PROF. EMILIO BARONE.

LAUREANDO:

NATALE DI TOMMASO

CORRELATORE:

CHIARISSIMO PROF. GAETANO CASERTANO

MATRICOLA:

665411

ANNO ACCADEMICO 2015-16

Introduzione

Quando parliamo di *sentiment analysis* nei mercati finanziari ci riferiamo ad una stima di quella che è la tendenza nel brevissimo termine della popolazione di investitori istituzionali e non; in pratica così come è stato fatto negli anni passati in via prima cartacea e poi digitale, si stima quella che è l'intenzione della "collettiva" degli operatori di un determinato mercato o settore rispetto ad un determinato titolo o indice. Come possiamo vedere approfonditamente nel prossimo capitolo ci è stata una grande evoluzione in questa metodologia accompagnata soprattutto dall'evoluzione nel campo tecnologico che repentinamente ha cambiato i sondaggi telefonici con dati dei social media. Quello che si cercherà di determinare in questa tesi è se l'utilizzo di indicatori di *sentiment* rialzisti o ribassisti per poter aprire e chiudere posizioni all'interno di un portafoglio virtuale. Verranno testate diverse e commentate diversi metodi su dati raccolti personalmente da marzo 2016 a settembre 2016.

Letteratura

Il market *sentiment* è monitorato con una varietà di metodi tecnici e statistici costantemente da decenni. Secondo l'interessante visione di Thomas Dorsey la gran parte del movimento globale di un singolo titolo è stato attribuito al *market sentiment*.

Una possibile dimostrazione empirica di questa tesi osservabile sul mercato può essere data dal fatto che spessissimo i titoli bancari hanno una correlazione altissima nelle fluttuazioni di brevissimo periodo, nonostante spesso le news che generano queste fasi di ribassi e rialzi collettivi interessino solo un singolo nome e non l'intero settore, a Wall Street è famosa la frase "trend is a friend", in italiano "il trend è un amico".

Un particolare filone di letteratura economica collega i risultati di finanza comportamentale, cambiamenti di attenzione degli investitori sui mercati finanziari, e dei principi fondamentali di *asset pricing*. Autori come Barberis nel 1998 Barberis e Thaler nel 2003, e Baker . & Würigler nel 2007 sostengono che i modelli comportamentali degli investitori al dettaglio hanno un impatto significativo sui rendimenti di mercato.

Almeno cinque principali approcci alla misurazione attenzione (*sentiment*) degli investitori sono noti oggi nella letteratura scientifica: misure basate sul mercato finanziario, indici a sondaggio, estrapolazione del *sentiment* tramite analisi testuali e di risorse specializzate on-line, volumi di ricerca keyword su internet, e di fattori non economici.

In questa tesi ho concentrato la mia ricerca al terzo approccio sopra elencato. Dopo aver studiato gran parte della bibliografia sull' argomento ho dato grande importanza alla presentazione e alla critica di tre lavori che ritengono principali per questa particolare disciplina.

Il primo lavoro è stato quello presentato del dicembre del 2010 dal Dott. Tim O. Sprenger e dalla Prof.ssa Isabel Welp della TUM School of Management di Monaco, un *working paper* dal titolo "Tweets and Trade: The Information Content of Stock Microblog".

Questo *paper* spicca fra l' ampia categoria di lavori sulla *sentiment analysis* attraverso l' analisi testuale utilizzando Twitter per la sua innovatività e la possibilità di poter processare un elevato numero di informazioni in tempo reale.

L' architettura del modello è quella più diffusa in questo tipo di analisi, ovvero l' utilizzo della "popolarità" delle discussioni generate in tempo reale su internet per analizzare le performance di prodotti finanziari. Nella versione estesa della tesi ho analizzato nel dettaglio i risultati di questo studio che rappresenta, a parer mio, una delle migliori applicazioni di questo tipo di analisi a singoli titoli e non indici aggregati. Durante questo periodo sono stati raccolti 249,533 *tweet* in lingua inglese contenenti hashtag che potessero ricondurli a discussioni di una società quotata facente parte dell' SP100. La scelta di questo specifico target è motivata dal fatto che l' SP100 rispecchia fedelmente quello che è lo spettro delle differenti tipologie di aziende ed industrie quotate americane.

Per classificare il tipo di testo, vista la mole ingente di dati, è stata utilizzato un classificatore testuale molto utilizzato nei lavori riguardanti la linguistica computazionale.

Il programma utilizzato per la classificazione è quello della *Weka machine learnig* , mentre il criterio è quello Naive Bayesian Multinomiale.

Per la programmazione del classificatore gli autori hanno inserito e classificato manualmente all' interno del programma 2500 *tweet* identificando manualmente il *sentiment* del messaggio; ogni stringa quindi veniva classificata in tre categorie di segnali: *Buy, Hold, Sell*.

I messaggi una volta classificati vengono aggregati tramite l' equazioni 1-2 e confrontati con le variazioni dei titoli azionario di riferimento. Le tempistiche di aggregazione in questa analisi seguono gli orari di chiusura del NYSE (es- al informazioni disponibili al giorno t vengono raccolte dalle 16.31 del giorno $t-1$ fino alle 16.30 del giorno t).

I risultati riguardanti la possibilità di sfruttare la *bullishness* dei messaggi sono incoraggianti per quanto riguarda i segnali di acquisto sulle relazioni ritardate. In particolare i segnali "BUY" sono accompagnati da rendimenti anomali nei giorni successivi, rendimenti che in media eccedono i costi di transazione dei maggiori broker (gli autori ipotizzano dei costi nell' are 0.15% -0.20%). La stessa

relazione però non vale per i segnali di tipo “SELL” poiché nonostante manifestino un accompagnamento molto simile non sono accompagnati da rendimenti anomali tali da superare i costi di transazione. Gli autori inoltre non escludono la possibilità che con diverse strategie di investimento e con algoritmi di classificazione più sofisticati sia possibile utilizzare questo tipo di *trading strategy* in maniera bidirezionale.

I due modelli successivi sono presentati praticamente dagli stessi autori a distanza di 5 anni uno dall'altro, il primo nel 2010 il secondo nel 2015.

Il Prof. Johan Bollen dell'Università dell'Indiana e la Dott.ssa Huina Mao in collaborazione con Xiao Zeng presentano un approccio alternativo all'uso dei social network per lo studio della finanza comportamentale. La domanda di base che si sono posti gli autori in questo caso è molto simile a quella dei precedenti autori citati. Sapendo che la finanza comportamentale ha evidenziato quanto le decisioni economiche individuali possano essere influenzate dalle emozioni, si chiedono se è possibile applicare questo risultato su larga scala sintetizzando un indicatore d'umore della collettività ed applicarlo anziché sui singoli titoli sull'intero mercato. Per fare questo anche in questo caso gli autori utilizzeranno Twitter per determinare se è possibile individuare una connessione fra un indicatore dello stato d'animo collettivo e i movimenti dell'*Dow Jones Industrial Average* nel tempo.

La differenza di approccio con gli autori citati in precedenza è netta, questa volta lo strumento utilizzato come *proxy* del *sentiment* degli investitori non è rivolto solamente ai trader che condividono le proprie informazioni, ma all'umore di tutta la collettività.

Il metodo è leggermente più complesso di quello precedente, anche in questo caso i *tweet* vengono convertiti in base a due classificatori d'umore, uno bidimensionale *OpinionFinder*, e GPOMS, analizza in modo simile il contenuto del testo dei *tweet* per generare una serie temporale a sei dimensioni su base giornaliera dello stato d'animo collettivo per fornire una visione più dettagliata delle variazioni del *sentiment* pubblico lungo una varietà di dimensioni diverse dalla semplice bidimensionalità di umore positivo o umore negativo.

Queste serie vengono analizzate congiuntamente a quella del DJIA, in un primo sotto periodo si indaga la relazione e si stimano i coefficienti per la calibrazione di un modello, in un secondo sotto periodo si testa la bontà previsionale e l'accuratezza di quest'ultimo. Il principale risultato di questo lavoro è sintetizzabile in un'accuratezza della previsione di previsione del 86.7% nella direzione del Dow Jones utilizzando un modello di sintesi del *sentiment* generale da Twitter. Questo risultato come è possibile vedere nella critica al modello non è scarso da errori sia di battitura che teorici.

Il terzo lavoro riguarda gli autori sopracitati, questa volta però l'approccio è diverso. Il lavoro intitolato *Quantifying the effect of bullishness on international financial markets* ha il vantaggio essere stato prodotto riconoscendo gli errori e le limitazioni dei precedenti studi sulla *sentiment analysis* effettuati dagli autori. Il modello introdotto come vedremo si pone a metà fra i due approcci analizzati nei capitoli precedenti, ottenendo risultati altrettanto validi. Questa volta la ricerca non sarà estesa alla totalità dei tweet inseriti sui social ma ai soli *tweet* contenenti i termini “bearish” e “bullish”. L'obiettivo degli autori è definire un indicatore che beneficia del vantaggio della tradizionale richiesta esplicita dei sondaggi (bullish vs bearish), ma che possa utilizzare l'ampia quantità di dati disponibile sui social network senza passare però attraverso un classificatore linguistico. L'indicatore proposto fa derivare il *sentiment* direttamente dai messaggi *tweettati* o ricercati dagli utenti, ma a differenza delle analisi precedenti, cerca di individuare pattern che identifichino direttamente l'indirizzo finanziario del messaggio *postato* o ricercato. Per ridurre al massimo l'ambiguità del *sentiment* sintetizzato, gli autori concentrano la loro attenzione sui soli due termini “Bullish” e Bearish, scelti perché ampiamente utilizzati in contesti finanziari e raramente in altri contesti, quindi che riescono a massimizzare intrinsecamente l'accuratezza di identificazione nell'individuazione del *sentiment* rialzista (*bullish*) o ribassista (*bearish*) del messaggio inserito.

I risultati ottenuti sono incoraggianti, la tabella sottostante riporta i coefficienti di regressione ottenuti utilizzando il *Twitter Bullishness Index* come regressore sugli indici dei maggiori mercati anglofoni.

Lag	US: DJIA		UK: FTSE		CA: GSPTSE		China: SSE	
	Coeff.	p-value	Coeff.	p-value	Coeff.	p-value	Coeff.	p-value
1	13.18	0.01*	17.98	0.0005**	14.08	0.001**	8.73	0.09*
2	1.30	0.81	-10.39	0.06*	-5.26	0.26	-3.16	0.571
3	3.03	0.57	11.11	0.04*	8.16	0.08	6.78	0.224
4	-8.79	0.10	-9.85	0.07*	-11.35	0.01*	-2.91	0.601
5	-2.31	0.60	-3.54	0.46	-1.799	0.64	-1.60	0.757

Questi coefficienti normalizzati indicano l'impatto che si avrebbe in termini di punti base di rendimento nell'indice di mercato a seguito di uno shock di una *deviazione standard* nel *Twitter Bullishness Index*.

Gli autori sviluppano anche un modello, con metodologia simile, basato però sul volume di ricerca su Google definendo un *Google Bullishness Index*. Il modello però basandosi su osservazioni settimanali

non benefica della rapidità di adattamento nel misurare il sentiment che invece è possibile attraverso i dati che si estraggono da Twitter.

Applicazione

Al fine di creare una strategia di investimento basata sulla sentiment analysis ho raccolto i tweet contenenti il termine *bearish* e *bullish* da marzo 2016 a settembre 2016, analizzando in totale 747,954 tweet. I dati si sono mostrati molto simili a quelli presentati nella terza metodologia esposta nella sezione della letteratura. Gli indici scelti sono quindi lo S&P/TSE della *Toronto Stock Exchange*, il FTSE100 della *London Stock Exchange* ed il *Dow Jones Industrial Average* per il mercato USA. Dopo aver svolto differenti analisi sul campione ottenuto ai fini di individuare *cluster* nella distribuzione dell'indice affiancata a quella degli indici di mercato, ho scelto di sviluppare una strategia che include tutte le maggiori evidenze empiriche riscontrate nella letteratura presentata. Ne capitolo 2 (riassunto qui nella sezione letteratura), durante la presentazione dei risultati ottenuti da Sprenger e Welpke ho riportato nella figura 2.1 l'evidenza empirica ottenuta dagli autori effettuando un *event study*. Come ho riportato, la metodologia utilizzata è tratta dal lavoro più datato di Tumarkin e Whitelaw (2001). Rimando al paragrafo 2.2 per le evidenze riportate dall'*event study* in questione, in questa sede però voglio evidenziarne le caratteristiche di base. Gli autori impongono come regola di identificazione dell'*event day*, per massimizzare la probabilità di conferma della tesi, una caratteristica precisa: affinché un'osservazione possa rientrare nell'evento (oltre le imposizioni sui volumi di *tweet* non applicabili nel mio caso), l'indice di *sentiment* di riferimento deve eccedere di almeno due deviazioni standard la media delle osservazioni di tale indice calcolata sulle ultime 5 osservazioni per essere identificato come *BUY*, o ovviamente essere inferiore della stessa misura per essere identificato come *SELL*. Questa regola opportunamente affinata, può essere riapplicata al campione osservato per determinare anziché l'inserimento o meno in un *event study* per inserire o meno un ordine di acquisto o di vendita sul riferimento dell'indice.

Dai lavori di Bollen e Mao invece è emersa una relazione robusta anche se di tipo non causale fra il *Twitter Bullishness Index* ed il rispettivo mercato. Sebbene la non causalità questa relazione non possa essere utilizzata in modelli multifattoriali, l'evidenza empirica riscontrata può essere utilizzata per modellare una strategia di trading che la incorpori. La tabella 3.3 mostra, come già detto, il valore

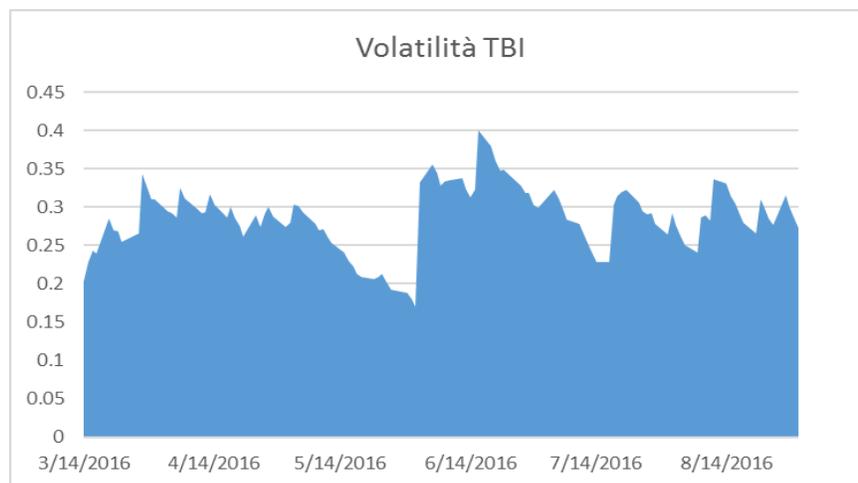
atteso in termini di *basis point* del rendimento dei diversi indici di mercato nei giorni successivi che si attende in seguito ad uno *shock* pari ad una deviazione standard nell' indice di mercato.

La strategia di investimento che tesserò sul campione di richiama entrambe le metodologie citate applicando un modello *esponenziale weighted moving average* per la stima della volatilità σT_t^B con equazione sottostante, mentre la strategia di investimento è riassunta nella formula nella pagina successiva.

$$\sigma^2 T_t^B = \lambda \sigma^2 T_{t-1}^B + (1 - \lambda) \left(T_t^B - \sum_{n=1}^5 \frac{T_{t-n}^B}{5} \right)^2$$

La stima della volatilità del *Twitter Bullishness Index* necessita di un modello che si adatti prontamente alle variazioni dell'indice T_t^B , rispettando la struttura infrasettimanale dei volumi osservati nei paragrafi precedenti, al fine di definire la volatilità vengono usate le sole ultime 5 osservazioni per un parametro $\lambda = 0.90$.

Il grafico riporta la stima della volatilità ottenuta mediante l'applicazione dell'equazione 12 per il *Twitter Bullishness Index* di DJIA.



Dal grafico 3.12 è visibile come il livello volatilità del TBI non sia costante durante il periodo osservato, inoltre si nota come i picchi di volatilità dovuto a shock nel sentimento tendono a riassorbirsi nel corso della successiva settimana di negoziazione.

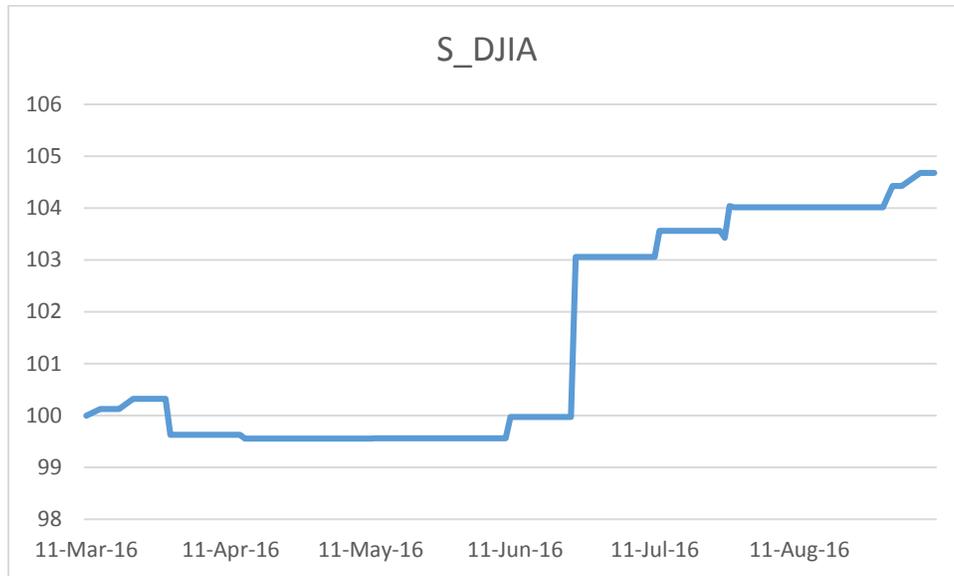
L'idea di base per modellare la strategia di trading è proprio basata su questa evidenza. La metodologia proposta in MBC2015 ci permette di estrarre un indicatore che rientra nella famiglia di strumenti per il *nowcasting*, ovvero che può essere usato per sfruttare la relazione a brevissimo termine con il fenomeno sottostante a cui è legato. Poiché la relazione è basata sugli *shock* e non sul modulo della misura di *sentiment*, la regola di selezione di Tumarkin e Whitelaw necessita di un accorgimento. Aniché selezionare le osservazioni in base alla differenza fra il valore corrente e quello della media, in maniera molto più coerente con la teoria esposta nel paragrafo 3.2 del testo, ho deciso cambiare il criterio di selezione in quanto non siamo interessati alla differenza fra il valore osservato e la media settimanale, ma piuttosto alla differenza fra le ultime due osservazioni e la volatilità giornaliera osservata nell' ultima settimana.

La strategia che chiameremo S viene applicata sull' osservazione T_t^B dei rispetti indici di mercato, ed in base alle condizioni della formula 13 definisce se entrare *LONG*, ovvero in una posizione lunga sull' indice di riferimento per il successivo giorno di negoziazione. Oppure *SHORT*, quindi entrare in una posizione corta sull' indice di mercato, oppure per esclusione *WAIT* cioè non effettuare nessuna operazione di trading per quella data.

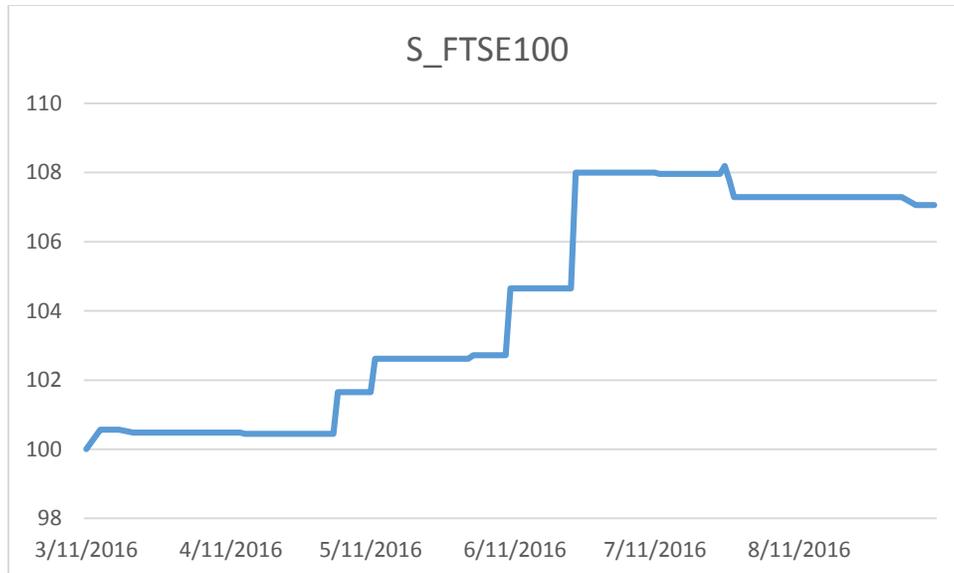
$$S(T_t^B) = \begin{cases} LONG & se \quad T_t^B - T_{t-1}^B > 2 * \sigma T_{t-1}^B \\ SHORT & se \quad T_t^B - T_{t-1}^B < -2 * \sigma T_{t-1}^B \\ WAIT & se \quad |T_t^B - T_{t-1}^B| < 2 * \sigma T_{t-1}^B \end{cases}$$

Nel prossimo paragrafo vedremo la performance di investimento per questa strategia rispetto ai diversi mercati inclusi nella applicazione. Inoltre la strategia S sarà applicata, con una piccola opportuna calibrazione, anche al *Google Bullishness Index* per verificare se esista davvero una differenza di performance significativa che rispecchi la differenza in termine di potere previsionale evidenziata dalla letteratura.

La strategia S riportata nel paragrafo precedente viene applicata ai tre mercati analizzati. Il primo grafico riguarda le performance di S_{DJIA} . Si nota immediatamente come la capacità predittiva del modello conduca ad un risultato economico positivo.



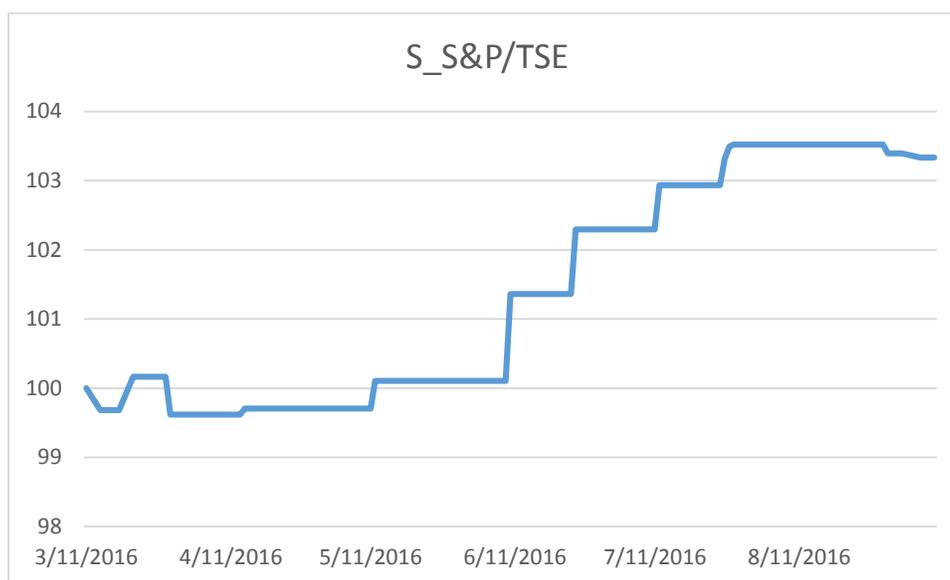
Il rendimento registrato nel semestre di osservazione dalla strategia di trading sul *Dow Jones Industrial Average* è pari al 4.68% ottenuto tramite posizioni in 14 giornate di negoziazione, 10 delle quali hanno mostrato rendimenti positivi, esprimendo una accuratezza di previsione del 71,4%.



Il secondo grafico riporta le performance della $S_{FTSE100}$, questa applicazione ha dato in termini di rendimento assoluto la performance migliore, realizzando un rendimento semestrale del 7,05%, questo dato però è accompagnato da una accuratezza di previsione più bassa. La $S_{FTSE100}$ nel periodo osservato realizza in tutto 13 posizioni giornaliere sull' indice FTSE100, 8 delle quali si dimostrano

profittevoli in termini economici, realizzando quindi nel periodo analizzato un'accuratezza di previsione del 61,5%.

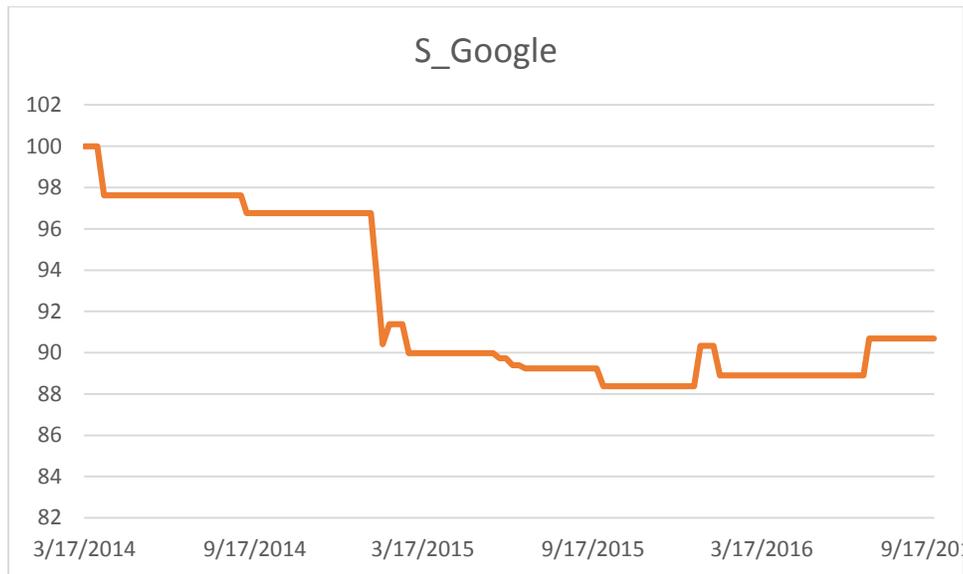
Infine il terzo grafico invece riporta le performance di $S_{S\&P/TSE}$, anche in questo caso la strategia ha riportato nei 6 mesi di osservazioni un rendimento positivo pari al 3.34%, anche se inferiore alle altre due applicazioni sopra elencate. Il livello di accuratezza in questo caso si posiziona a metà fra le applicazioni illustrate precedentemente, realizzando una previsione esatta nel 69.3% delle 13 volte in cui è entrato in posizioni sull'indice.



Prima di riportare i risultati ottenuti applicando la S_{Google} , è possibile apprezzare la similitudine dei grafici fin ora commentati. Questa deriva dal fatto che nonostante le variazioni degli indici siano diverse, ed esistano giorni di chiusura particolare che vanno a modificare la serie storica (quindi anche la deviazione standard), la gran parte della serie storica del *Twitter Bullishness Index* calcolato per i sei mesi di applicazione è condivisa. Questo implica che in molti casi le differenze rilevate nelle performance siano riconducibili alle differenze nelle variazioni fra gli indici di mercato piuttosto che alle differenze fra i vari TBI ricalcolati.

Coerentemente con la teoria riportata nei capitoli precedenti, l'applicazione della strategia attraverso la misurazione del *sentiment* misurato settimanalmente non riporta un risultato economico positivo. Il rendimento ottenuto nell'arco di tempo analizzato è pari al -9,31%, con un'accuratezza di previsione di solo il 23,07%. Sebbene questi risultati non siano direttamente confrontabili vista la grande differenza

nel periodo di osservazione è apprezzabile la coerenza fra questa applicazione e le evidenze empiriche riportate nel *paper* di riferimento Mao Bollen e Counts del 2015.



Concludo questa analisi con una tabella che riassume i risultati ottenuti da questa strategia di trading attraverso posizioni *long* e *short* sui diversi indici di mercato.

I risultati ottenuti nella prima sezione della tabella riguardano le varie applicazioni della strategia S sui diversi indici, la seconda parte invece riporta finì prettamente comparativi i risultati ottenuti dai diversi indici nello stesso periodo (marzo 2016 – settembre 2016) rispettando i criteri sopra citati, come tasso *risk free* è stato usato il LIBOR a 6 mesi osservato agli inizi di marzo 2016. Dalla prima riga è possibile notare come la strategia di trading ottenga risultati inferiori in termini di rendimento puro rispetto agli indici su cui opera esprimendo un rendimento medio de 5.02%.

Questo risultato però, come qualsiasi altro rendimento di natura finanziaria e non, non va letto ed interpretato singolarmente.

In termini di volatilità tutte le applicazioni hanno prodotto una volatilità dei rendimenti inferiore rispetto ai *benchmark* di riferimento, $S_{FTSE100}$ e $S_{S\&P/TSE}$ in particolare esprimono una volatilità nel periodo di riferimento circa 3 volte inferiore rispetto a quella del mercato di riferimento. Continuando l’analisi della volatilità, il *Down Side Risk* calcolato per le strategie risulta essere nettamente inferiore rispetto agli indici di mercato.

<i>T=6 mesi</i>	Strategia S			Benchmark		
<i>rf=0.892%</i>	DJIA	FTSE100	S&P/TSX	DJIA	FTSE100	S&P/TSX
Rendimento	4.677%	7.056%	3.333%	6.267%	9.808%	6.146%
Rend.Giornaliero	0.037%	0.056%	0.026%	0.049%	0.078%	0.049%
σ	3.301%	4.077%	1.934%	6.804%	11.173%	5.685%
σ Giornaliero	0.302%	0.382%	0.166%	0.604%	0.995%	0.506%
DSR	0.698%	0.642%	0.642%	4.463%	7.347%	3.385%
DSR Giornaliero	0.062%	0.057%	0.057%	0.396%	0.654%	0.302%
Media Positivi	0.554%	1.157%	0.482%	0.420%	0.713%	0.389%
Media Negativi	-0.230%	-0.200%	-0.261%	-0.418%	-0.711%	-0.420%
Accuratezza	71.429%	61.538%	69.231%	-	-	-
Sortino Ratio	5.42	9.60	3.80	1.20	1.21	1.55
Sharpe Ratio	1.147	1.512	1.262	0.790	0.798	0.924

Questa sostanziale differenza è deducibile dalle differenze in media fra le giornate a rendimenti positivi e negativi che sussistono fra i rendimenti ottenuti attraverso S_{DJIA} , $S_{FTSE100}$ e $S_{S\&P/TSE}$ e quelli ottenuti dai rispettivi *benchmark*. Per esempio i rendimenti giornalieri positivi (negativi) ottenuti da $S_{FTSE100}$, sebbene sia l' applicazione della strategia che riporti l' accuratezza minore, sono in media 0.44% superiori (0.511% inferiori in modulo) rispetto alle stesse classi osservate nel FTSE100, questa differenze notevole ci dice che in pratica la $S_{FTSE100}$ sia nel caso in cui restituisce una previsione corretta ,ma soprattutto nel caso in cui la previsione si rivela sbagliata, genera in media rendimenti migliori di quelli del *benchmark* di riferimento.

Infine la tabella riporta lo *Sharpe Ratio* ed il *Sortino Ratio* calcolati utilizzando il tasso LIBOR con scadenza a 6 mesi osservato a marzo 2016, nel momento in cui è partita l' applicazione. Anche in questo caso le misure di *performance risk-adjusted* danno un valore nettamente superiore ad S_{DJIA} , $S_{FTSE100}$ e $S_{S\&P/TSE}$ piuttosto che ai *benchmark* di riferimento.

I valori di queste due misure di performance indicano che a parità rischio, misurato nello *Sharpe Ratio* con la *standard deviation* e con il *down side risk* nel *Sortino Ratio*, i risultati ottenuti dalle tre applicazioni sono migliori in termini di rendimento in eccesso al tasso *risk free* considerato.

È infine interessante notare come entrambi i rapporti calcolati per S_{DJIA} , $S_{FTSE100}$ siano nettamente superiori e diversi fra loro rispetto ai benchmark di riferimento che presentano sia uno *Sharpe Ratio* che un *Sortino Ratio* molto simile. Questa diversità è motivata prevalentemente da due fattori. Il primo, già citato in precedenza, la differenza in media fra le realizzazioni positive e soprattutto negative, più marcata per la $S_{FTSE100}$ rispetto che per il S_{DJIA} . Il secondo, per il diverso tasso di accuratezza di

previsione; maggiore per la S_{DJIA} , che rende più vicine le due misure di performance *risk adjusted* di quanto lo sarebbero se presentassero qualità previsionali identiche.

Conclusioni finali

Se dovessi dare una risposta alla mia principale domande di ricerca “La *sentiment analysis* può essere utilizzata per modellare strategie di investimento a breve termine?” La mia risposta sarebbe **SI**, ma probabilmente è un investimento più adatto ad investitori istituzionali che *retail*.

Per motivare questa risposta non mi dilungherò ancora sull’ effettiva utilità previsionale della *sentiment analysis* ai fini di investimento poiché è già riportata in ogni pagina di questo testo. La motivazione che mi induce a identificare un target istituzionale piuttosto che *retail* per questo tipo di investimenti è legata principalmente alla difficoltà di acquisizione, manipolazione ed osservazione dei sentiment data. Per raccogliere i circa 750.000 tweet che ho utilizzato nella mia applicazione, non disponendo di algoritmi o programmi automatizzati, ho dovuto dedicare circa mezz’ora al giorno, tutti i giorni, dei miei ultimi sei mesi al download, la pulizia e l’ archiviazione dei dati. Ammesso che esistano persone con capacità informatiche molto più avanzate delle mie, che riescano ad automatizzare il processo tramite un semplice click o facendo a meno anche di quello, il problema principale risiede nella fonte stessa di questo tipo di dati. L’archivio che ho costruito è generato da API che hanno un acceso di livello C, quindi rappresenta un estrazione casuale dal database completo di solamente il 10% dei tweet pubblici per il periodo analizzato, ed è il livello massimo di informazioni ottenibili dalla clientela *retail*. Anche accettando l’ipotesi forte che la totalità degli investitori condivida le proprie intenzioni finanziarie su Twitter, questo vuol dire che la misura di sentiment che ho ottenuto non è l’espressione della totalità degli investitori attivi su Twitter, ma di un campione casuale contenente informazioni sul 10% di coloro che decidono di rendere pubblici i propri messaggi. Anche senza avere numeri alla mano è semplice immaginare che in realtà questo campione non sia sempre e per forza adatto a stimare la propensione rialzista o ribassista della totalità degli investitori.

Per quando riguarda i volumi delle ricerche di Google la situazione non migliora. A livello *retail* sono disponibili solo i volumi settimanali normalizzati delle *queries* di ricerca. Anche in questo caso a livello corporate invece esistono accessi privilegiati che consentono di osservare i volumi in maniera più dettagliata arrivando fino alle osservazioni orarie prive di normalizzazione.

Solo pochi clienti corporate possono avere ad un accesso di livello superiore, avendo a disposizione BIG DATA più fluidi, che non solo consentano di ottenere una *proxy* migliore del *sentiment* dell’intera

collettività di investitori, ma permettano soprattutto di osservare le mutazioni dello stesso in tempo reale ed operare di conseguenza.

Sebbene non esistano certezze sul futuro, ritengo molto probabile che la *sentiment analysis* possa essere sviluppata ed implementata nel *set* di strumenti disponibili per gli investitori istituzionali per modellare strategie di investimento alternative.

Inoltre progetti come l'attuale *Google Clouds Sentiment Analysis Tools*, lanciato lo scorso agosto, mi fanno ben sperare che in un futuro prossimo possa esistere un livello di informazione di maggiore qualità sui *social data* accessibile a chiunque abbia voglia di usarla.