ORCA – Online Research @ Cardiff

Applications of a Multivariate Hawkes Process to Joint Modeling of Sentiment and Market Return Events

Steve Y. Yang*†, Anqi Liu‡, Jing Chen‡, and Alan Hawkes§

†School of Business, Stevens Institute of Technology,
1 Castle Point on Hudson, Hoboken, NJ 07030
‡School of Mathematics, Cardiff University,
Senghennydd Road, Cardiff, CF24 4AG, UK
§School of Management, Swansea University Bay Campus,
Fabian Way, Swansea, SA1 8EN, UK

To investigate the complex interactions between market events and investor sentiment, we employ a multivariate Hawkes process to evaluate dynamic effects among four types of distinct events: positive returns, negative returns, positive sentiment and negative sentiment. Using both intraday S&P 500 return data and Thomson Reuters News sentiment data from 2008 to 2014, we find: a) self-excitation is strong for all four types of events at 15 minutes time scale; b) there is a significant mutual-excitation between positive returns and positive sentiment, and negative returns and negative sentiment; c) decay of return events is almost twice as fast as sentiment events, which means market prices move faster than investor sentiment changes; d) positive sentiment shocks tend to generate negative price jumps; and e) the cross-excitation between positive and negative sentiments is stronger than their self-excitation. These findings provide further understanding of investor sentiment and its intricate interactions with market returns.

Keywords: Point process; Hawkes process; Investor sentiment; Return jumps; News sentiment

JEL Classification: C32, C51, G14, G17

---

*Corresponding author. Email: syang14@stevens.edu

1.   Introduction and background

Price movements, trading behavior and investor decisions have always been important topics discussed in the finance literature. There is a whole gallery of associated questions such as how these affect market efficiency? and how are prices affected post significant market events? does feedback trading exist in certain scenario, etc. More fundamentally, what exactly drives trading and underlying price formation process? With the development of behavioral finance, particularly in the area of irrational trading, there exist two commonly accepted views: price-driven trading (e.g. Shleifer (2000) examining feedback trading) and news-driven trading (e.g. Tetlock (2007) looking at news sentiment premium affecting asset pricing). With the latter, news sentiment, an investor sentiment[1] proxy, has drawn great interest and attention from scholars and practitioners.

As Thaler (2005) argued, "Our attention is much more quixotic and capricious ... Instead, news functions more often as an *initiator* of a chain of events that fundamentally change the public's thinking about the market". The nature of the interaction of news and market events has long been speculated and inquired upon by both practitioners and academics. In this study, we consider two types of events: news sentiment and market returns; and both of them can move in two directions: positively and negatively. Typically, returns are considered as continuous events, which means that at a selected time frequency (say, 1-minute), price movements which lead to calculations of returns are deemed unchanged within that time interval (1-minute). Strictly speaking, the minute-on-minute price updates are discrete rather than continuous events for the time series. For news sentiment, the actual occurrence or arrival does not take place in a continuous fashion either. In light of the nature of occurrences and interactions of the different types of events, we consider them as discrete events and apply discrete point processes. Here, we aim to examine the interactions of the news sentiment and return events to gain insights about the cascading effects of extreme returns and elevated investor sentiment and how investors respond to and precipitate market events.

During the past decades, emerging studies on behavioral finance have explored behavior of investors, aiming to get better understanding of financial markets. For example, as an extension to the efficient market hypothesis, Lo (2004) proposed the concept of market evolution and raised the argument that the investor behaviors are corresponding to competition, adaptation, and natural selection in a market ecosystem. In fact, debates about market efficiency has never ceased for decades. However, the empirical evidence of the existence of market anomalies indicates that fundamental information cannot fully reflect financial market movements. The concept of noise trader risk highlighted a new view to explain market phenomena. Based on psychological findings about cognitive bias, De Long *et al.* (1990) argued that noise trader behaviors can be homogeneous. This viewpoint contradicted the hypothesis that noise trader risk cannot be systematic. They proposed a stochastic model to demonstrate that noise traders' misperception can drive the price deviation from its fundamental value. Moreover, such deviation can be large enough to cause limited arbitrage (Shleifer and Vishny 1997). These findings in behavioral finance provided evidence to show market inefficiency and encouraged recent research on investor sentiment and its market impact.

Motivation of this study rests on the argument that investor sentiment can have significant impact on market prices; at the same time, market prices can also have significant influence on investor sentiment as market prices move away from sideway patterns. Shleifer (2000) argued that investor sentiment is the cause of numerous market anomalies such as noise trader risk, limits of arbitrage, and closed-end fund puzzle. This claim has been echoed by a number of empirical studies showing predictive power of investor sentiment to financial market prices (Antweiler and Frank 2004, Tetlock 2007, Tetlock *et al.* 2008, Mo *et al.* 2016). On the other hand, other studies showed market price movements affect investor sentiment such as Antoniou *et al.* (2005), Salm

---

[1]Investor sentiment is defined broadly by Baker and Wurgler (2007a) as a belief about future cash flows and investment risks that is not justified by the facts at hand. Generally, positive (negative) sentiment indicates social belief of bull (bear) markets. Please see Section 4.1 for the precise sentiment data measurement.

and Schuppli (2010), and Arnold and Brunner (2015). Baker and Wurgler (2006) found the U-shaped pattern of trading behaviors which suggested that "investors demand both high growth and distressed firms when they are optimistic, or their speculative propensity is high, and avoid extremes when they are pessimistic, or their propensity to speculate is low".

Following these findings, we can have some good sense that investor sentiment events interact with market return events. However, what is still unclear is what dynamics exist among these types of critical market events. Would the interactions exist between the same types of events (e.g. returns and returns) or different types of events (e.g. returns and sentiment)? Or would it be more complicated? For instance, would the interaction of positive to positive returns be more active than negative to negative returns? Similarly, we need to consider the cross-event type interactions such as responses of positive returns to negative sentiment. These are not only interesting but also important questions as already shown in literature that these relations provide profound insights to form trading strategies and decisions, and even affect the market movements. For example, it is commonly known that negative movements in prices tend to lead to greater market negative reactions than positive ones to positive reactions. Therefore, it is necessary to understand the dynamics of these multiple events.

In this paper, we propose to examine four types of events: positive return, negative return, positive sentiment and negative sentiment. We adopt a methodology due to Buckle $et\ al.$ (2017) to identify such positive or negative events from the market return and news sentiment time series data[2] and Figure 1 demonstrates the dynamics of interactions among them. Intuitively, it reflects what we call "self-excitation" and "mutual-excitation" reactions, which are responses within the same and across different types of events, respectively. In theory, we can hypothesize any pair-wise interaction. One may argue that Hawkes processes only allow for 'positive excitation', for example, positive returns generate positive sentiment: this is often the case, but not necessarily always the case. Here we employ the usual multivariate Hawkes process in which information shocks from one type of events lead to causation of another type of events, by raising the intensity. The events that occur may have a negative connotation, e.g. a negative price jump, but the excitement is positive in the sense that the frequency of occurrence of those jumps may be increased (i.e. the intensity may be increased) following the occurrence of events of a different type, such as a positive news event. The interpretation of such a connection between these two types of event, if it is found to exist, is a question of behavioral finance theory. Note also that although Hawkes processes usually exhibit positive excitation in the sense described above, it is possible to exhibit inhibition (reducing the intensity) by using a non-linear Hawkes model: this is common in neural networks. Special cases of the usual linear Hawkes model can also exhibit inhibition (Chen $et\ al.$ (2018)). Note also that Figure 1 shows connections between positive and negative returns or between positive and negative sentiment events. Both are possible in theory, although our empirical results (see Table 2) in this paper show that the former does not seem to exist ($\alpha_{12} = 0.02$, $\alpha_{21} = 0.00$) but the latter does ($\alpha_{34} = 0.25$, $\alpha_{43} = 0.34$).

To further explain our motivation to form a comprehensive study of news sentiment and stock returns, we examine the pairwise interactions (both self-excitation and mutual-excitation) between any two types of events using a multivariate Hawkes (M-Hawkes) process. This is because each of these eight interactions has clear and significant market impact and needs to be well understood in order to explain investors' behaviors. We use intraday S&P 500 index return and Thomson Reuters News sentiment data and examine the following hypotheses (with two self-excitation and six mutual-excitation):

(i) H1: Self-excitation of market returns tend to cluster: positive returns lead to further positive returns and negative returns to negative returns. These effects could be caused by the fact

---

[2]Price (sentiment) event refers to the return (sentiment) that indicates a chain of movements in the same direction (e.g. positive or negative). In particular, we introduce the concept of running-jumps to identify a series of coherent price movements (see Section 3.2). In a similar way, positive sentiment jump would be a run of consecutive upward movements of the sentiment measure and negative sentiment jump would be a run of consecutive downward movements of the sentiment measure.

that investors chase positive returns to buy or negative returns to sell once they spot initial signs of positive (negative) returns.

(ii) H2: We also hypothesize similar clustering effects for news sentiment events (sentiment self-excitation): positive to positive sentiment and negative to negative sentiment. This is because sentiment is a reflection of social belief and it is not unusual to believe investors hold 'me-too' kind of mentality to respond to news shocks, especially towards to the release of extreme news articles (Mitra and Mitra 2011).

(iii) H3: Mutual-excitation between positive return and negative return can be predicted and we think they are often associated with bubbles and crashes, particularly, when market corrects quickly from price shocks (Daniel *et al.* 1998). and

(iv) H4: There may exist mutual-excitation between positive sentiment and negative sentiment due to natural shifting or even momentum of good news and bad news. It is worth noting again that our focus is to detect extreme events in both returns and sentiment and study their interactive dynamics. The point is that we believe these extreme events (events that have greater levels by their measures) would have more impact at both returns and sentiment in comparison to those moderate events.

(v) H5: Intuitively, there should be mutual-excitation from positive (negative) return to positive (negative) sentiment as financial news tell stories about preceding market situations. For example, if positive returns are spotted and investors start to chase positive returns, it is more likely to form a certain market belief so that more investors would join the queue to chase positive returns and build up on the positive sentiment. It is well perceived that negative signals are often associated with greater responses (see Brock *et al.* (1992)), therefore, we would expect negative return events, once in a run, would mutually excite negative sentiment.

(vi) H6: Mutual-excitation from positive (negative) sentiment to positive (negative) return is the most widely explored sentiment impact to financial markets. Positive news raises buying power in the market and price jumps up. Similarly, investors pull out investments due to negative news so that price drops immediately (see Gwilym *et al.* (2016) arguing the increased speculative demand that forms sentiment type of events would increase market returns). Similarly, one of the most recent reports published by the FEDs also provides evidence of predictive power of sentiment to stock returns and claim that positive sentiment stimulates quick positive returns while negative news stories generates relatively delayed responses in stock returns. It further states that "much of the delayed response to news occurs around the subsequent earnings announcement" (see Heston and Sinha (2016)).

(vii) H7: Regarding effects from positive (negative) return to negative (positive) sentiment, we do not expect significant results. There is no clear evidence in the literature to interpret this. However, it makes sense psychologically as it would be highly irrational that investors who are experiencing positive returns suddenly/immediately form strong negative belief that the market would go down; and continue to keep such negative belief to make it into market sentiment. However, we are indeed aware that after yielding positive returns for some time, the market may turn to produce negative returns, such as what we typically see when bullish and bearish markets take turns to govern the market. The question is more about the timing of the changing point and whether after the change, the opposite direction of return movements would form a negative sentiment. In other words, the exact reasons for this phenomenon are unclear and require more comprehensive studies.

(viii) H8: Finally, we expect to see mutual-excitation from positive (negative) sentiment to negative (positive) return, which is another type of news impacts to financial markets. This phenomenon indicates price correction of overreaction to news (De Bondt 1989).

To our best knowledge, this is the first paper to examine market returns and investor sentiment interactions at relatively high frequency level and the first paper utilizing a dynamic and non-
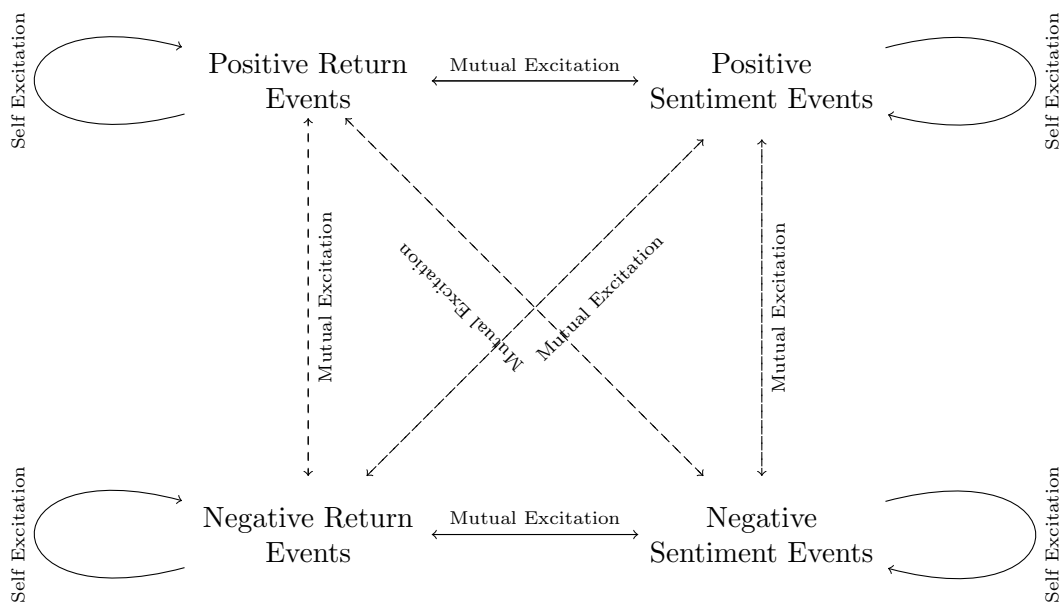
Figure 1.: Market and Sentiment Events Interaction Illustration

linear modeling framework for this purpose[3]. We focus on both self- and mutual-reactions of these events as illustrated in Figure 1. More specifically, we model significant jump events as a multivariate point process emphasizing the dynamics of investor sentiment shocks and market extreme returns. While there are several studies on sentiment-driven market events using lead-lag causality of investor sentiment to market return or volatility, we argued that investor sentiment and market return have a very complex relationship, and the existing econometric modelings of such relationships have limitations. In the recent literature, news sentiment has become a broadly adopted proxy of investor sentiment toward finance market and economic conditions (Tetlock 2007, Mian and Sankaraguruswamy 2012, Smales 2014a). In this study, we use news sentiment data from Thomson Reuters News Analytics (TRNA) as a proxy to measure investor sentiment (Healy and Lo 2011). We define extreme positive/negative returns and positive/negative sentiment shocks as the four distinct discrete events.

The primary contribution of our study resides in both technical and finance aspects. Technically, we introduce a multivariate Hawkes process to understand clustering dynamics of different discrete events based on the self-excitation and mutual-excitation between investor sentiment and market returns. Our approach, in particular, focuses on modeling clustering effects of jumps (extreme changes) in market returns and news sentiment. This is an important aspect of this kind of study because, on the one hand, we understand that the market needs volatility for price updates and trading incentives; and on the other hand, we believe that extreme movements in returns that subsequently forms sentiment need to be detected as they may have systematic impacts upon the financial markets. Technically, we build in a computationally effective jump detection method into our modeling for such purpose. Further, the multivariate modeling offers a great framework to fully capture the dynamics among the eight different pairs of interactions. We can clearly study the directional responses in each pair of events by recognizing the initiation events and response events. For example, if we compare Hypotheses H4 and H5, although both are examining the interactions between positive (negative) returns and positive (negative) sentiment, H4 takes positive (negative)

---

[3]The common methods used in the existing literature tend to apply linear modeling to regress positive or negative news sentiment represented by news texts onto a return variable. This is more of the modeling of redeeming sentiment as a feedback element to the return process. Such examples usually fall into the literature of feedback trading such as Merton (1980); Sentana and Wadhwani (1984); Shiller (1984); Tse and Tsui (2002); Antoniou *et al.* (2005); Chen *et al.* (2017) etc.

sentiment as the initiating events while H5 uses them as the response events. With the simplicity of Hawkes self- and mutual excitation, we will be able to achieve comprehensive understanding of sophisticated dynamics in a single modeling setup. Finally, there are several reasonings behind our choice of model design: 1) It is based on the belief that both return and sentiment are essential in studying the complex modern market as the intraday activities become more frequent and intense; 2) days with extreme returns and/or sentiment (we call them 'noisy days') are more likely to trigger clustering than those 'normal' trading days (we can call them 'quiet days'); 3) within a trading day, the more active returns (sentiment) update, the more likely we expect to see higher level of intra-day intensity of further return (sentiment) events within that day; and 4) after a 'noisy day', we expect to see a rise in the intensity of the next day due to overnight spillover effects.

More specifically, we find the following major interaction effects between the four types of events using S&P 500 index return and Thomson Reuters News sentiment data from 2008 to 2014. First, self-excitation is strong for all four types of events at 15 minutes time scale, and the mutual-excitation between return and news sentiment is not present at smaller time scale. We not only have found evidence for our hypotheses (H1 and H2 for self-excitation and H3 to H8 for mutual-excitation) but also suggest a time scale for this kind of studies. This, intuitively, makes sense as returns tend to update faster than sentiment and responses to returns could be instantaneous while formation of sentiment requires a longer period as a social belief of market participants towards the market.

Second, there is a significant mutual-excitation between positive return and positive sentiment, and negative return and negative sentiment. This is in line with what we hypothesized in hypotheses H4 and H5. On the sentiment side, we notice that the cross-excitation between positive and negative sentiment is even stronger than the self-excitation of these sentiment events. This provides new evidence for hypothesis H4 that has not been documented in the literature before. Third, decaying speed of return events is almost twice as fast as that of sentiment events, which means market prices move faster than investor sentiment changes. Lastly, positive sentiment shocks tend to generate negative price jumps and this is consistent with psychological theories and evidences listed in De Bondt (1989).

The rest of the paper is organized as follows. Section 2 reviews existing work on Hawkes processes and the applications in finance and current literature on sentiment market studies. We then formulate a multivariate Hawkes process and describe the model estimation methodology in Section 3. And then we describe both sentiment and return data in Section 4. Section 5 lays out the details of the model calibration, and Section 6 presents the modeling results and findings. We finally conclude the major findings and contributions in Section 7.

2.   Literature review

In this paper, we aim to understand interactions between investor sentiment and market return events through a multivariate Hawkes process. Hence, we focus on the literature of Hawkes processes and investor sentiment and its effect on market returns.

2.1.   *Hawkes process literature*

Hawkes processes form a class of multivariate point processes that were introduced by A.G Hawkes in two theoretical papers (Hawkes 1971a,b). The first practical application was to model occurrence of seismic events (Hawkes and Adamopoulos. 1973). The essential feature is that the occurrence of an event increases the probability that further events would occur: sometimes described as a property of contagion. Hawkes models are becoming more and more popular in finance. This popularity can be attributed to their great simplicity and flexibility, as anticipated by Bowsher (2007). Several generalizations have been proposed since, such as being enriched by endowing each event with a mark variable, thus obtaining a sequence of event times, components and marks;

and labeling events with different marks to have different effects on the future intensities (Hawkes 1972). In finance, marks can be used in order to model trades performed at different times with different volumes or a drawdown intensity (Fauth and Tudor 2012, Chavez-Demoulin and McGill 2012, Embrechts *et al.* 2011).

Bowsher (2007) observes the trading of General Motors shares over forty days on NYSE in 2000. After fitting a mutually-exciting model of two market events: changes in mid-price quotes and trades arrivals, they find the former approximately change every three minutes on average and the latter every thirty seconds. They also conclude that the cross-excitation between these two types of events were important but short-lived while the self-excitation terms are less significant. This suggested that trades increase the intensity of price changes of the share and the price changes also stimulate trades. Further, the decay of intensity after the event arrival is much faster for cross-excitation than the self-excitation terms, which is consistent with their general observation of the trades and quotes arrival speed and sequence.

Large (2007) suggests there are ten different types of events affecting the price changes of assets and focuses on the electronic limit order book of Barclays on LSE SETS over 22 trading days in January 2002. It identifies that four out of ten types of events are major evens shifting the pricing including buy moving ask, sell moving bid, bid between quotes and ask between quotes. Usually half-lives of excitation are less than one minute, except long-term effect of self-excitation could last between sixteen to forty minutes. Aggressive market orders are associated with the first two types of events and demand liquidity while aggressive limit orders are connected to the last two types of events and typically replenish liquidity following aggressive market orders. The resilient replenishment of liquidity tends to follow a shock less than 40% of the time and is equally likely to be at the bid rather than the ask with a half-life under 20s. Jaisson (2014) also suggests that the market order flow can be approximated as an unstable Hawkes process with a long flow memory based on the assumption that the price is a martingale and the impact of meta-orders is linear. This provides scope to incorporate a power law or square root law to further study the tail behavior of the order flow structure (The typical robustness test can follow Russell (1999)).

Bacry *et al.* (2013a) show the great simplicity with which Hawkes processes can reproduce significant high frequency phenomena including the signature plot and Epps effects, and provide empirical evidence through fitting Euro-bond and bund futures. Bacry *et al.* (2012) use a non-parametric estimation of the kernel shape of a symmetric process and find the decaying shape driven by the power law has a long memory. Bacry and Muzy (2014) model market behavior by a 4-dimensional Hawkes process at tick level (price changes up and down; market orders at best bid and best ask) that allows for reproducing microstructure noise that is often shown as either strong microscopic mean reversion or de-correlation of the increments. The kernel functions can be estimated from market data.

Chavez-Demoulin and McGill (2012) use Hawkes processes based on a generalized Pareto distribution to capture the excessive extreme losses and their clustering in the setting of a Value at Risk model (similar work includes Chavez-Demoulin *et al.* (2005), Herrera and Schipp (2013)). Further examples of the use of Hawkes processes in finance include Bauwens and Hautsch (2009) and Carlsson *et al.* (2007) regarding trading aspects, Embrechts *et al.* (2011) and Errais *et al.* (2010) utilising multivariate processes and affine models in credit risk, Aït-Sahalia and Saglam (2013) and Aït-Sahalia *et al.* (2010) in contagion of jumps. Particular forms of Hawkes processes called Epidemic-Type After Shock (ETAS) models are used in seismology.

Bacry *et al.* (2013b) show the functional Central Limit Theorem (CLT) can be obtained, extending a result from Hawkes and Oakes (1974), and suggest that a univariate Hawkes process on large timescales can be considered as an asymptotical Brownian motion. Jaisson and Rosenbaum (2015) show that, when the branching ratio (e.g. the integral of the kernel function) tends to be 1, a simple univariate Hawkes process asymptotically does not tend to Brownian motion but has the form of a Cox Ingersoll Ross (CIR) process while the biivariate model of Bacry *et al.* (2013a) can be represented as a Heston model. Karabash and Zhu (2015) provide CLT for marked Hawkes process and Zhu (2014) gives CLT for CIR process with Hawkes jumps. Jaisson and Rosenbaum

(2016) show that, if the regression kernel has a heavy power-law tail, after suitable rescaling, the process behaves asymptotically as a kind of integrated fractional CIR process, instead of the classical Brownian CIR process obtained with a light-tailed kernel.

Fulop *et al.* (2015) fitted a self-exciting process into the asset pricing model in order to capture co-jumps and jump clustering. They identified jump clustering during both the 1987 market crash and the 2008 global financial crisis; and further suggest that there is a learning effect of later crisis from the earlier one, which is reflected in the tail behavior of the return distribution.

2.2.   *Investor sentiment and market return literature*

Psychological evidence suggests that sentiment, emotion and mood play a key role in affecting investors when making financial decisions (Brown and Cliff 2004, Cohen and Kudryavtsev 2012, Hilton 2001, Nofsinger 2005). Barberis *et al.* (1998) developed a theory of investor sentiment to illustrate the impact of investor overreaction and underreaction to public information on generating on post-earnings announcement drift, momentum, long-term reversals and predictive power or scaled-price ratio. Daniel, Hirshleifer and Subrahmanyam further enriched the idea of investor sentiment with private information leading to overconfidence (Daniel *et al.* 1998, 2001). On the empirical front, a number of studies found different measures of investor sentiment significant in explaining asset price and volatility movements. Chopra et al. showed that prior losing portfolios significantly outperform prior winning portfolios by 5-10% annually during the next 5 years, validating the overreaction effect (Chopra *et al.* 1992). La Porta et al. also displayed evidence that the correction of the extreme investor sentiment tends to revert during earnings announcements when investors realize their initial beliefs were too extreme (Porta *et al.* 1997, Thaler 2005). These studies are instrumental in demonstrating the existence of investor sentiment along with its impact on financial markets.

The interaction between investor sentiment and market return has long been studied as a feedback effect between these two distinct market forces. Feedback mechanisms have been explored in the field of finance, mainly through the examination of its effects on price and volatility. Investor sentiment can also be quantified in the form of its feedback effects. Hirshleifer *et al.* (2006) presented a theoretical framework that justifies irrational investors to earn abnormal profits based on a feedback mechanism from stock prices to cash flows. Crude oil prices were found to contain feedback effects along with an inverse leverage impact with its implied volatility (Aboura and Chevallier 2013). Khanna and Sonti (2004) showed the feedback effect of stock prices on firm value through a herding equilibrium model and investigated the incentive for traders to conduct price manipulation. İnkaya and Okur (2014) estimated the volatility feedback effect rate using Malliavin calculus and suggested its predictability of large price declines. They showed that large feedback effect rate is a useful indicator for measuring market stability.

There is also empirical evidence that investors take actions based on price movement: a self-perpetuating pattern of investor's behavior is present in G7 stock markets and other international markets (Antoniou *et al.* 2005, Salm and Schuppli 2010). The effect of feedback trading was found to vary across business cycle (Chau and Deesomsak 2015) and the strongest influence was observed during periods of financial crisis with declining futures prices (Salm and Schuppli 2010). Hou and Li developed a regression model of feedback trading to analyze CSI300 stock returns and demonstrated that lagged index returns can predict market index return and conditional volatility (Hou and Li 2014). In addition, feedback trading was found to significantly influence exchange rate movements (Laopodis 2005). Using a theoretical framework, Arnold and Brunner (2015) showed that positive feedback trading causes price overreaction and the impacts of feedback trading would be dampened if news is incorporated into price in time. Mo *et al.* (2016) analyzed more than 12 million news articles and documented the presence of a significant feedback effect between news sentiment and market returns across the major indices in the US financial market.

Investor sentiment proxy was one type of the most widely used investor sentiment measurement in early studies. The rationale behind it is that the shock in investor sentiment can be observed

from the series of events resulted from that shock (Baker and Wurgler 2006). Some event-based proxies, including surveys, mood proxies, trade records, fundamental values, mutual fund flows, were verified to be associated with unexplained market features (Lee *et al.* 1991, Baker and Wurgler 2007b). However, Baker and Wurgler (2006) mentioned that confounding influence might induce potential bias between the proxies and investors' actual beliefs. For example, the survey results may not be trustworthy as people tend to misreport their behaviors. In addition, the explanation power of sentiment proxies to market activities was much weaker compared with fundamentals. To deduce the biases, Beer and Zouaoui (2013) combined sentiment proxies with fundamental or technical indicators to capture insights about market dynamics.

Investor sentiment analysis based on textual information, such as earning reports, news articles, and social media messages, is a breakthrough in measuring investor emotions and beliefs. Lexicon-based sentiment measurement has been initially investigated in the domain of computational methodologies for social mood analysis. Mishne *et al.* (2006) introduced a mood tracking framework that evaluates sentiment levels in blogosphere through a linear regression model of word frequencies. News is a widely used information source for investor sentiment evaluation. Tetlock (2007) examined mood categories of Wall Street Journal columns using the General Inquirer's Harvard IV-4 psychosocial dictionary. Following this study, Tetlock *et al.* (2008) applied the same approach to evaluate individual firm sentiment based on 350,000 news items on Dow Jones News Service and Wall Street Journal. These two studies proposed similar findings that media pessimism is strongly related to subsequent price drop. In another study, Antweiler and Frank (2004) designed a sentiment indicator from the view point of disagreement in news and confirmed that fluctuations of sentiment polarity raise trading volume. In a recent study, Dodds and Danforth (2010) defined sentiment level as average word scores referring to Affective Norms for English Words (ANEW) dictionary. They developed a large-scale text analysis framework to measure happiness expressed in song lyrics and blogs. Similar approaches have been applied to investor sentiment in recent years. Smales (2014b) argued that gold future returns reveal an asymmetric response to news releases since negative news has stronger impacts than positive news. Yang *et al.* (2015) demonstrated significant linear relationship of abnormal news sentiment to implied volatility of S&P 500 index in the following few days.

3. Methodology

3.1. *Multivariate Hawkes process*

Our study is based on the multivariate Hawkes process, and in this section we recall the essential features of this modeling framework. A point process (PP) is a random process for which any one realization consists of a set of isolated points in time. A comprehensive treatment of point processes is given in Brémaud (1981). In the following, we adapt the theoretical framework of Bowsher (2007) with consistent notations. Market events, such as changes in quoted prices or market flash crashes, can be described as realizations of an M-variate PP $\{T_i, Z_i\}_{i=1,2,\dots}$. Here T_i is the *occurrence time* of the ith event, and Z_i records the *type* of the ith event. In the following, we denote the *counting process* associated with $\{T_i, Z_i\}$ as $N(t) := (N_m(t))_{m=1}^M$, where $N_m(t)$ records the number of type m events that have occurred in $[0, t]$. The natural filtration generated by the PP $N(t)$ is denoted by $\{\mathcal{F}_t^N\}$. We take the approach of Bowsher (2007), and specify the PP through the vector conditional intensity process $\lambda(t) = (\lambda_m(t))_{m=1}^M$. Intuitively, $\lambda_m(t)$ can be understood as the conditionally expected number of type m events per unit time as the time interval shrinks to zero. We denote $\lambda(t)$ as the $(P, \mathcal{F}_t)-$intensity of $N(t)$, where P is the data generating process (DGP) and $\{\mathcal{F}_t\}$ is the filtration that we condition on.

The M-variate Hawkes process is defined via the conditional intensity vector

$(\lambda_1(t), \lambda_2(t), ..., \lambda_M(t))'$, where

$$\lambda_m(t) = \mu_m(t) + \underbrace{\sum_{j=1}^{k} \tilde{\lambda}_{mm}^{(j)}(t)}_{\text{self exciting}} + \underbrace{\sum_{q=1,q\neq m}^{M} \sum_{j=1}^{k} \tilde{\lambda}_{mq}^{(j)}(t)}_{\text{mutual exciting}}, \tag{1}$$

where $m = 1, 2, ..., M$. Here $\mu_m(t)$, a positive deterministic function in time, is the baseline intensity of type m events. The above unconvential form of the Hawkes model, taken directly from Bowsher (2007), is equivalent to the excitation kernel functions having the form of sums of k exponential terms. In addition, we allow the possibility that each of those components could spill over from the end of the previous day. Thus we have

$$\tilde{\lambda}_{mr}^{(j)}(t) = \underbrace{\pi_{mr}^{(j)} \tilde{\lambda}_{mr}^{(j)}(\tau_{d-1}) e^{-\rho_{mr}^{(j)}(t-\tau_{d-1})}}_{\text{intensity spillover from } \tau_{d-1}} + \int_{[\tau_{d-1},t)} \alpha_{mr}^{(j)} e^{-\beta_{mr}^{(j)}(t-u)} dN_r(u), \tag{2}$$

for $\tau_{d-1} < t \leqslant \tau_d$, $d = 1, 2, ...$, and $\tilde{\lambda}_{mr}^{(j)}(0) = 0$. Here $m \in \{1, 2, ..., M\}$ and $r \in \{1, 2, ..., M\}$. There is also the parameter constraints: $\alpha_{mr}^{(j)} \geqslant 0$, $\beta_{mr}^{(j)} > 0$, $\pi_{mr}^{(j)} \geqslant 0$, and $\rho_{mr}^{(j)} \geqslant 0$. Here, τ_d is defined as the time at the end of the dth trading day, using a clock that stops outside of trading hours. Hence, τ_{d-1} is both the beginning of the dth trading day and the end of the previous day. $\pi_{mr}^{(j)}$ is the fraction of the partial intensity $\tilde{\lambda}_{mr}^{(j)}(\tau_{d-1})$ at the end of the previous day that spills over into the next day, then declining exponentially with time.

3.2. *Identification of jumps in news sentiment and in return*

As stated earlier, we use news sentiment as a proxy to measure investor sentiment. In order to apply the multivariate Hawkes processes to study the interaction between news sentiment shocks and market extreme returns, we need to identify events in both news sentiment and return data series. Here we shall adopt the method developed in Buckle *et al.* (2017), which we briefly discuss in this subsection. We begin by discussing the well-known asset return series and then adopt a similar method for the news sentiment series.

Asset prices $p_{t,i\Delta}$ are observed on day t on a regular grid of times from 9:30 to 16:00 with a grid size Δ, for which we consider various values between 2 minutes and 15 minutes. For a given value of Δ, we calculate the series of log-returns: $r_{t,i} = \log(p_{t,i\Delta}) - \log(p_{t,(i-1)\Delta})$, is the return for the ith interval on day t, with $i = 1, 2, \ldots, N$ and $N\Delta = 390$ minutes, the duration of the trading day. In seeking jumps in the price series we look for values in the return series that are large relative to a measure of local volatility. Therefore, we define a sequence of *scaled returns*

$$r_{t,i}^* = \frac{r_{t,i} - med(r_t)}{\sqrt{MednRV_{t,N}/N}} \tag{3}$$

where $med(r_t)$ is the median return on day t. $MednRV_{t,N}$ is a jump-robust estimator of realized variation on day t based on a sum of rolling medians of n consecutive squared intra-day returns. This is defined as

$$MednRV_{t,N} = f_n \left(\frac{N}{N+1-n} \right) \sum_{i=(n+1)/2}^{N-(n-1)/2} (med(|r_{t,i-(n-1)/2}|...|r_{t,i}|...|r_{t,i+(n-1)/2}|))^2 \tag{4}$$

The factor f_n is equal to 1.62360, 1.74332 and 1.82184 for, respectively, $n = 5,7$ and 9. It is chosen so that $MednRV_{t,N}$ is an unbiased estimator of daily realized variation, $N(\sigma^2)$, in the event that returns were to be distributed as $i.i.d.N(0, \sigma^2)$. The use of running medians enables the omission of 2, 3 or 4 neighboring large returns from the realized variation estimate when using, respectively, $n = 5,7$ and 9. This prevents the realized variation estimate from becoming excessively inflated, and therefore reducing scaled returns and possibly preventing them from being identified as jumps. Consequently, we recommend using $n = 9$ if N is large and the return series has sections of high volatility. However, when Δ is as large as 15 minutes (and $N = 26$) we prefer a smaller n. For consistency when studying a range of Δ values we compromise by taking $n = 7$ and so using the realized variation estimator

$$Med7RV_{t,N} = 1.74332 \left(\frac{N}{N-6} \right) \sum_{i=4}^{N-3} (med(|r_{t,i-3}|...|r_{t,i}|...|r_{t,i+3}|))^2 \qquad (5)$$

and the scaled returns

$$r_{t,i}^* = \frac{r_{t,i} - med(r_t)}{\sqrt{Med7RV_{t,N}/N}} \qquad (6)$$

Running-jump is a more sensible measure of jump according to which a positive (or negative) jump consist of the accumulation of a series of positive (or negative) scaled returns. We exclude scaled returns that are less than 1 by replacing them with 0. This process is not applied to sentiment as scaled sentiment values are relatively small and a large number of them are 0.

Jumps are identified with accumulated scaled returns with absolute values greater than some critical value C which is chosen so to have enough data to sensibly fit our model. This should lead to an average of at least one jump per day, although several days may have no jumps, while we recommend the average number of jumps per day should be less than $N/10$, probably well less. About half the jumps are expected to be positive and about half negative.

Sentiment can be treated in a similar way with $s_{t,i}$, the change in sentiment over the ith interval on day t, being treated in exactly the same way as the log-returns $r_{t,i}$. A jump may be described as occurring instantaneously at an arbitrarily chosen point within the interval in which it occurs (say beginning, middle or end) provided that this is chosen consistently. However, sometimes a price jump and a sentiment jump may occur in the same interval, and then it should be decided randomly that one of them came first by some very small time difference.

3.3. *Model estimation: Maximum likelihood estimation of M-variate Hawkes process*

The $M-$variate Hawkes process allows the interaction of type q events with the intensity of type m events. In addition this model includes "intensity spillover effects", the effect of what is happening at the end of a day on what happens at the start of the next day, as indicated in Bowsher (2007). The log-likelihood of a multidimensional Hawkes process can be computed as the sum of the likelihood of each coordinate.

From Equation (17) of Bowsher (2007), we have the following representation of the log likelihood for the M-Hawkes process:

$$l(\theta) = \sum_{m=1}^{M} l_m(\theta_m), \qquad (7)$$

where $\theta = (\theta_1, \theta_2, ..., \theta_M)$ and

$$l_m(\theta_m) = \sum_{d=1}^{T/l} \underbrace{\left\{ \int_{A_d} (1 - \lambda_m(s; \theta_m)) ds + \int_{A_d} \log \lambda_m(s; \theta_m) dN_m(s) \right\}}_{\text{trading day decomposition}}, \quad (8)$$

and here $A_d = (\tau_{d-1}, \tau_d]$, T is the total trading time and the length of the trading day is $\tau_d - \tau_{d-1} = l$, so that T/l is the total number of trading days in the series. We make the assumption similar as Bowsher (2007) that θ_m are variation free, thus the maximization of $l(\theta)$ can be achieved through maximization of $l_m(\theta_m)$ individually. The above expression (8) decomposes the contributions of log likelihoods from different trading days, and this allows the recursive evaluation of the log likelihood across trading days.

The intermediate expressions $\lambda_m(s; \theta_m)$ are given in (1) and (2). Plugging (2) into (8) and after carrying out the integrations, we have

$$l_m(\theta_m) = T - \int_0^T \mu_m(s; \theta_m) ds + \sum_{d=1}^{T/l} \sum_{T_i^{(m)} \in A_d} \log \lambda_m(T_i^{(m)}; \theta_m)$$

$$- \sum_{r=1}^{M} \sum_{d=1}^{T/l} \sum_{j=1}^{k} \left\{ \frac{\pi_{mr}^{(j)}}{\rho_{mr}^{(j)}} (1 - e^{-l\rho_{mr}^{(j)}}) \tilde{\lambda}_{mr}^{(j)}(\tau_{d-1}; \theta_m) + \sum_{\tau_{d-1} \leqslant T_i^{(r)} < \tau_d} \frac{\alpha_{mr}^{(j)}}{\beta_{mr}^{(j)}} (1 - e^{-\beta_{mr}^{(j)}(\tau_d - T_i^{(r)})}) \right\}. \quad (9)$$

Note that the only difference of the above (9) and the equation (18) of Bowsher (2007) is to replace $\sum_{r=1}^{2}$ by $\sum_{r=1}^{M}$. In our application, we consider the case of $M = 4$, which means that there are 4 types of events.

For univariate point processes, the theoretical properties of MLE have been established in Ogata (1978). For multivariate PPs, there are very few theoretical results, but simulation evidence in Bowsher (2007) shows that the MLEs are well-behaved. Now the question is how we shall compute the above likelihood function and implement it through some numerical optimization schemes. Note that in (9), the entries $T_i^{(m)}$ are observed occurrence times of each event. Now it is important to understand how it is updating across trading days.

Due to intra-day seasonality phenomenon as observed in Russell (1999), it is recommended to use a piecewise linear function for the deterministic component of intensity $\mu_m(t) \equiv \mu_m(t; \gamma_m)$. This depends on the values of $\gamma_{m,i} > 0 (i = 1, ..., 8)$, which are the values of $\mu_m(t; \gamma_m)$ at $t = 0$, for $i = 1$, and $t = i - 1.5$ for $i > 1$: i.e. the knots of a linear spline at 9:30, 10:00, 11:00,..., 16:00: this function is the same for each day[4], and the $\gamma_{m,i}$ are parameters of the model to be estimated. The formula for this function is given below

$$\mu_m(t; \gamma_m) = \begin{cases} \mathbb{1}_{v(t) \in (0, 0.5]}[\gamma_{m,1} + 2v(t)(\gamma_{m,2} - \gamma_{m,1})] \\ \quad + \sum_{i=1}^{6} \mathbb{1}_{v(t) \in (i - 0.5, i + 0.5]}[\gamma_{m,i+1} \\ \quad + (v(t) - i + 0.5)(\gamma_{m,i+2} - \gamma_{m,i+1})], & \text{for } v(t) > 0 \\ \gamma_{m,1}, & \text{for } t = 0, v(t) = 0 \\ \gamma_{m,8}, & \text{for } t = 6.5, v(t) = 0. \end{cases}$$

---

[4]As mentioned, the returns data are expected to satisfy seasonality. It is not clear that the sentiment necessarily also possesses this property but, for the sake of simplicity, we assume that it behaves in a similar way.

where $t \in [0, 6.5]$ is the number of hours that have elapsed since the start of the trading day and $v(t) = 6.5(t/6.5 - [t/6.5])$.

On page 896 of Bowsher (2007), he mentions that numerical optimization of the log-likelihood is performed using the MaxBFGS algorithm with numerical derivatives in Ox (see documentation of this programming language by Doornik etc.) Please also check the working paper version of Bowsher (2002) for more details on estimation.

4.   Data

Market return data and financial news sentiment data are obtained from Thomson Reuters Tick History (TRTH) and Thomson Reuters News Analytics (TRNA) respectively. We collect data from January 2008 to December 2014, and formulate them to frequency in 2 minutes, 5 minutes and 15 minutes.

4.1.   *Investor sentiment data*

Thomson Reuters News Analytics (TRNA) is a structured database with over 80 metadata fields about financial news. It provides sentiment for each company mentioned in each news article. The sentiment is quantified as positive, negative and neutral probabilities so that we can customize the formula for our sentiment score. The fields we used for sentiment calibration in this study are listed below:

- datetime: The date and time of the news article.
- ric: Reuters Instrument Code (RIC) of the stock for which the sentiment scores apply.
- relevance: A real-valued number between 0 and 1 indicating the relevance of a piece of news to a stock. One news article may refer to multiple stocks. The stock with more mentions will be assigned a higher relevance.
- p^+, p^0, p^-: Positive, neutral, and negative sentiment probability (i.e., $p^+ + p^0 + p^- = 1$).

To evaluate the sentiment score for each stock mentioned in each news article, we calculate expectation of sentiment probabilities adjusted by relevance value (see Equation 10).

$$\text{Sentiment}(\text{ric}_i) = \text{relevance}(\text{ric}_i) \times [p^+(\text{ric}_i) \times 1 + p^-(\text{ric}_i) \times (-1)] \tag{10}$$

We collect news published in trading hours to match the time with market data. The investor sentiment of each 15-minute interval is the average of all news sentiment in that time period (see Equation 11).

$$s(t) = \frac{1}{N_t} \sum_{\text{ric}} \sum_{i=1}^{N_t} \text{Sentiment}(\text{ric}_i) \tag{11}$$

where the news ric_i is published during $[t-1, t)$, and N_t is the total number of news in $[t-1, t)$.

4.2.   *Market price data*

Stock market indices are proxies of equity market performance. We obtain 2 minutes, 5 minutes and 15 minutes intraday trading price $P(t)$ of S&P 500 Index from TRTH database and calculate

log-return (see Equation 12).

$$r(t) = \log \frac{P(t)}{P(t-1)} \tag{12}$$

5. *M*-variate Hawkes process calibration

The four types of events are modeled as an M-variate Hawkes process in which the arrival of events is affected by self-excitement and mutual-excitement from the other three events. We recap the models here, using single exponential kernels ($j = 1$ in Equations (1) and (2)).

$$\lambda_m(t) = \mu_m(t) + \sum_{r=1}^{4}[\pi_{mr}\tilde{\lambda}_{mr}(\tau_{d-1})e^{-\rho_{mr}(t-\tau_{d-1})} + \int_{[\tau_{d-1},t)} \alpha_{mr}e^{-\beta_{mr}(t-u)}dN_r(u)], \tag{13}$$

where $m = 1, 2, 3, 4$ denotes positive return, negative return, positive sentiment, and negative sentiment events respectively.

Initial values are important for parameter calibration through MLE. We first determine spline intensity by ignoring impacts of both self-exciting and cross-exciting. After some experimentation, we chose to base our analysis on 15-minute intervals and consequently decided to calculate baseline intensities using splines with 5 knots instead of the 8 knots described in Section 3.3. The first three time slots are 1.5 hours and the last one is 2 hours. Initial values for the iteration are shown in Table 1; final values are shown in Table 2.

$$\mu_m(t; \gamma_m) = \begin{cases} \sum_{i=1}^{3} \mathbb{1}_{v(t)\in(\frac{3}{2}(i-1),\frac{3}{2}i]}[\gamma_{m,i} \\ \qquad +(\frac{2}{3}v(t) - i + 1)(\gamma_{m,i+1} - \gamma_{m,i})] \\ \qquad +\mathbb{1}_{v(t)\in(4.5,6.5]}[\gamma_{m,4} + (\frac{v(t)}{2} - \frac{9}{4})(\gamma_{m,5} - \gamma_{m,4})], & \text{for } v(t) > 0 \\ \gamma_{m,1}, & \text{for } t = 0, v(t) = 0 \\ \gamma_{m,5}, & \text{for } t = 6.5, v(t) = 0. \end{cases}$$

Intuitively, the major contribution to arrival rates is from base-line intensity so that influence of both self-excitation and mutual-excitation should be lower than this level. In this model, α and π determine the shape of intraday and daily-spillover intensity excitation respectively. We set initial value of these parameters as 1.00, and expect the results to decrease and converge in the range of $(0, 1]$. The other two sets of parameters controlling the exponentially decaying memory are β and ρ. We apply constraints $1 < \rho < 25$ and $1 < \beta < 25$ based on the rationale that effective influence of excitation should be within 15 minutes to 1 day. The initial values for these parameters are $\beta = 15.00$ and $\rho = 15.00$, which is equivalent to memory length around 30 minutes.

	Positive return	*Negative return*	*Positive sentiment*	*Negative sentiment*
$\gamma_{m,1}$	2.84	2.63	0.68	0.89
$\gamma_{m,2}$	1.14	0.99	0.45	0.98
$\gamma_{m,3}$	0.40	0.42	0.69	1.00
$\gamma_{m,4}$	1.42	1.45	1.12	0.27
$\gamma_{m,5}$	0.10	0.27	1.24	1.50

Table 1.: Initial spline intensity

In loglikelihood maximization, we apply the Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm and set stopping criteria to achieve precision of 6 decimal points. The Hawkes process models

of positive return, negative return, positive sentiment, and negative sentiment take 61, 45, 60 and 71 iterations to converge respectively. Figure 2 shows parameter estimation convergence results for the four models. It is obvious that the most sensitive parameters are α and π, which control the shape of excitation. We also notice that the fitted spline intensity is very close to the initially calibrated non-homogeneous Poisson process. This confirms the convergence does happen, and the estimated model parameters are robust.
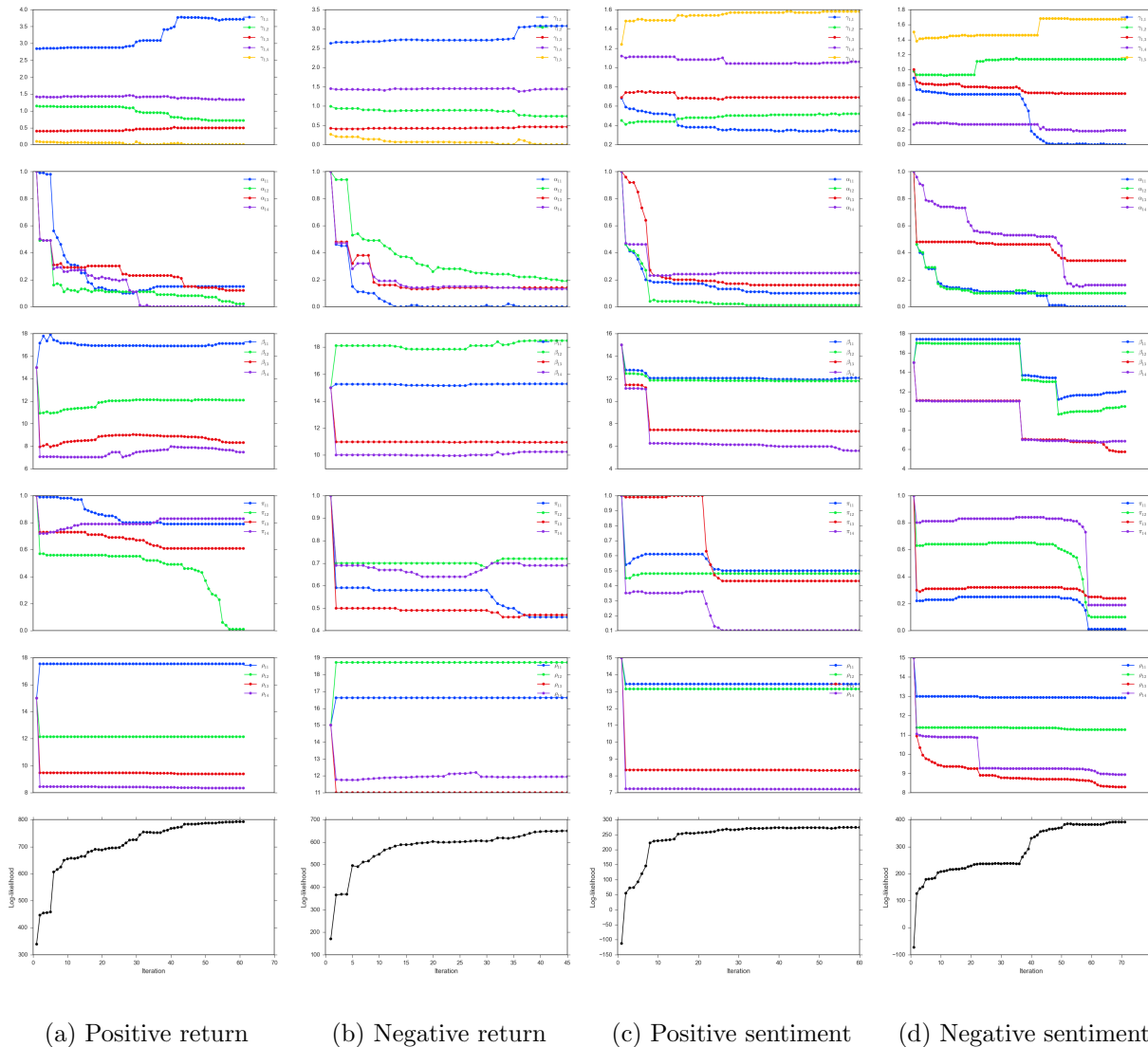

(a) Positive return   (b) Negative return   (c) Positive sentiment   (d) Negative sentiment

Figure 2.: Hawkes model parameter estimation

The M-variate Hawkes process model presents shock propagation in financial markets, especially interactions between market return and investors sentiment. We firstly verify that this model converges successfully (see Figure 2). With parameters changing and converging, likelihood of these four processes increase and reach to the maximum values. The calibration of MaxBFGS optimization is sensitive to initial value settings. We apply the following three steps to find out the initial values which can help us to obtain the best results. First, we determine the initial values for baseline intensity. We assume that the intensity is a non-homogeneous Poisson process, without any effects from self-excitement or mutual-excitement. Hence, α_{ij} and π_{ij} are set as 0, and β_{ij} and ρ_{ij} are set to large values. According to these settings, we get baseline intensity and apply that

into the second step in which we release the intraday excitement. In this step, the constraints are $\alpha_{ij} \geqslant 0$ and $1 \leqslant \beta_{ij} \leqslant 25$. Finally, we apply additional constraints $\pi_{ij} \geqslant 0$ and $1 \leqslant \rho_{ij} \leqslant 25$ to run a full optimization for all parameters. A special trait for the excitement that $\alpha_{ij} = 0$ is to fix $\pi_{ij} = 0$ and ρ_{ij} as a large number. The rationale behind this setting is that when the shocks have no intraday impacts to another type of event, the daily-spillover is meaningless.

6.   Results

In this section, we first present the results of detection of extreme events in sentiment and returns (see Section 6.1). Afterwards, we focus on our empirical results presented in Table 2, where MLE results are reported. We compare specific parameters for self-exciting and mutual-exciting in order to interpret causes and impacts of extreme events in Section 6.2. This also requires us to investigate size and half-life decaying of excitation (see Equation 14 and Equation 15).

$$\text{Intraday half-life: } T_{\frac{1}{2}} = \frac{\log 2}{\beta} \tag{14}$$

$$\text{Daily-spillover half-life: } T_{\frac{1}{2}} = \frac{\log 2}{\rho} \tag{15}$$

The second part of the results is to examine how these self- and mutual-excitation behave within and across trading days. As seen in Table 2, γs show the baseline intensity, αs and βs explain intra-day intensity and πs and ρs demonstrate inter-day intensity. Meanwhile, αs and πs control the shape of the exponential decay curves of the intensity functions while βs and ρs control the decay speed. All these parameters also suggest how these types of events interact with each other. Take α_{ij} as an example, they are parameters showing the significance of intra-day excitation impacts from event j to event i. When $i = j$, they are the cases for self-excitationexcitation or mutual-excitation otherwise. These detailed results by event types are in Section 6.3.

6.1.   *Extreme events in sentiment and in log-prices*

We identify investor sentiment shocks and extreme market returns using the running jumps identification method. According to occurrences of jumps, we calibrate M-variate Hawkes processes to model arrival of these events and obtain insights about investors' responses to environment changes. There are four major types of events that are used in our paper: positive price jumps, negative price jumps, positive sentiment shocks, and negative sentiment shocks. In the running-jump detection approach introduced in Section 3.2, we mention that a critical value C is predefined as a threshold. We notice that price moves are much more volatile than sentiment changes in terms of fatter tails for scaled return[5] (see Figure 3). The selected critical values are $C = 2.5$ and $C = 1.5$ for market return and news sentiment respectively based on the criteria of average 1 jump per day (see Figure 4).

We verify that return jumps present the U-shaped feature (see Figure 5a). The peaks of both positive and negative return jumps occur at 10:00AM and 15:15PM. Sentiment jumps do not show significant seasonality property (see Figure 5b). There are a large number of sentiment jumps at the beginning and the end of the day, while positive and negative sentiment shocks are active in different time periods in the middle of the day. In general, negative sentiment shocks appear frequently around the noon and positive sentiment jumps primarily occur during 13:30PM – 14:00PM.

---

[5]The gap of $[-1.00, 0.00)$ and $(0.00, 1.00]$ in scaled return is due to replacement of small jumps by zeros in running-jumps identification.
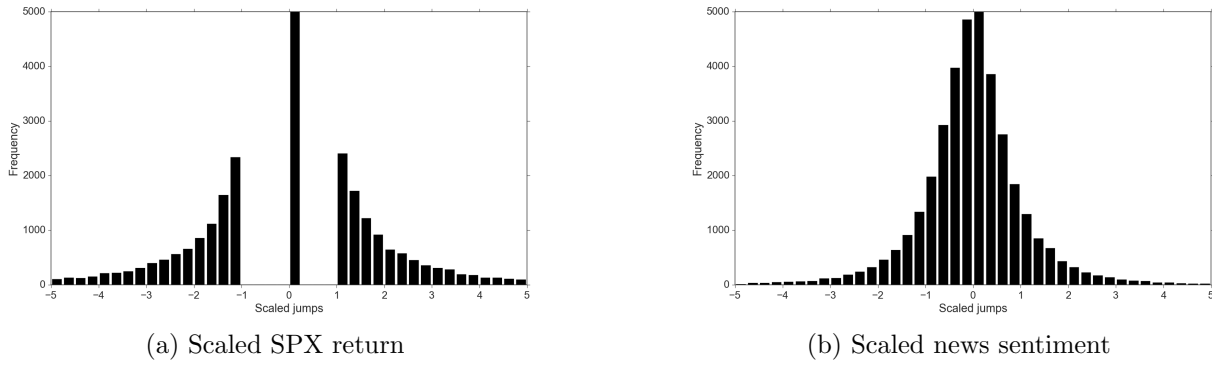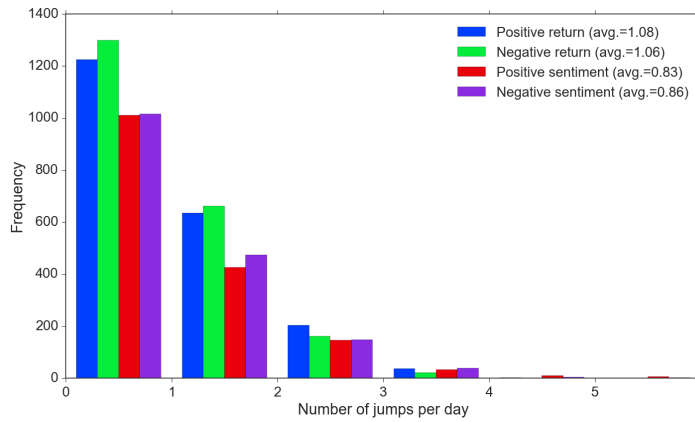
(a) Scaled SPX return

(b) Scaled news sentiment

Figure 3.: Histogram of scaled return and scaled sentiment

Figure 4.: Histogram of number of jumps per day

(a) SPX return jump intensity by time
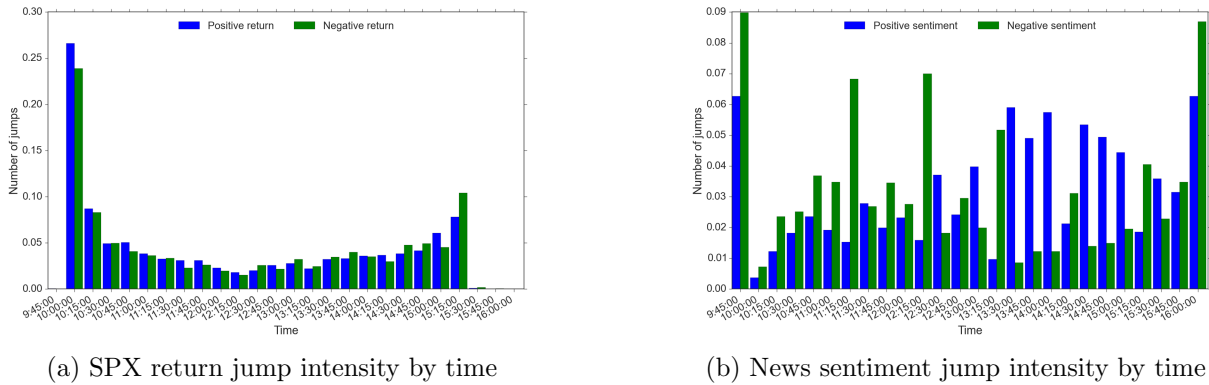
(b) News sentiment jump intensity by time

Figure 5.: Jump intensity of market return and news sentiment

6.2.   *General discussion of self- and mutual-excitation*

We observe that self-excitation is strong for all four types of events and both intra-day and inter-day. This confirms our hypotheses H1 and H2 due to clustering effects of occurrence of the same type of events.This exactly reflects the philosophy of Hawkes processes: the intensity of future events depend on the probability of previous events and information filtration up to date. The scale of intra-day intensity impacts α_{11}, α_{22} are 0.15 and 0.19 for positive and negative returns respectively, and 0.16 for both positive and negative sentiment events. If comparing them to mutual excitation terms, we can see that the self terms are generally greater than the cross terms apart from α_{34} and α_{43}. Thus one exception is the mutual-excitation between positive and negative sentiment appears to be even stronger than their self-excitation. For the roll-over day effects, self-excitation dominates

Positive return		Negative return		Positive sentiment		Negative sentiment	
$\gamma_{1,1}$	3.71	$\gamma_{2,1}$	3.08	$\gamma_{3,1}$	0.34	$\gamma_{4,1}$	0.00
$\gamma_{1,2}$	0.72	$\gamma_{2,2}$	0.74	$\gamma_{3,2}$	0.52	$\gamma_{4,2}$	1.14
$\gamma_{1,3}$	0.50	$\gamma_{2,3}$	0.46	$\gamma_{3,3}$	0.69	$\gamma_{4,3}$	0.68
$\gamma_{1,4}$	1.33	$\gamma_{2,4}$	1.44	$\gamma_{3,4}$	1.06	$\gamma_{4,4}$	0.19
$\gamma_{1,5}$	0.00	$\gamma_{2,5}$	0.00	$\gamma_{3,5}$	1.58	$\gamma_{4,5}$	1.67
α_{11}	0.15	α_{21}	0.00	α_{31}	0.10	α_{41}	0.00
α_{12}	0.02	α_{22}	0.19	α_{32}	0.01	α_{42}	0.10
α_{13}	0.12	α_{23}	0.14	α_{33}	0.16	α_{43}	0.34
α_{14}	0.00	α_{24}	0.13	α_{34}	0.25	α_{44}	0.16
β_{11}	17.13	β_{21}	15.27	β_{31}	12.07	β_{41}	11.98
β_{12}	12.09	β_{22}	18.49	β_{32}	11.80	β_{42}	10.45
β_{13}	8.30	β_{23}	10.96	β_{33}	7.32	β_{43}	5.73
β_{14}	7.46	β_{24}	10.26	β_{34}	5.58	β_{44}	6.85
π_{11}	0.79	π_{21}	$--$	π_{31}	0.50	π_{41}	$--$
π_{12}	0.01	π_{22}	0.72	π_{32}	0.48	π_{42}	0.10
π_{13}	0.61	π_{23}	0.47	π_{33}	0.43	π_{43}	0.24
π_{14}	$--$	π_{24}	0.69	π_{34}	0.10	π_{44}	0.19
ρ_{11}	17.53	ρ_{21}	$--$	ρ_{31}	13.44	ρ_{41}	$--$
ρ_{12}	12.14	ρ_{22}	18.72	ρ_{32}	13.14	ρ_{42}	11.27
ρ_{13}	9.38	ρ_{23}	11.00	ρ_{33}	8.32	ρ_{43}	8.30
ρ_{14}	$--$	ρ_{24}	11.94	ρ_{34}	7.20	ρ_{44}	8.94

Notes: α_{ij}, β_{ij}, π_{ij} and ρ_{ij} are parameters for impacts from event j to event i. When α_{ij} is zero, we ignore daily-spillover effects, marking as "$--$" for π_{ij} and ρ_{ij}.

Table 2.: M-variate Hawkes process model parameters
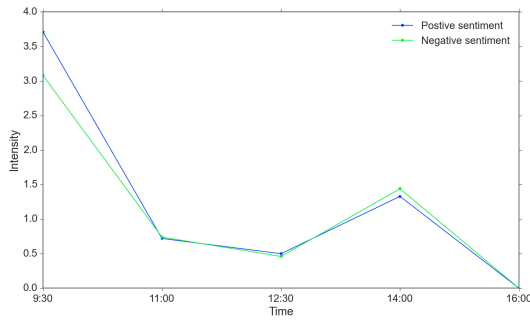
(a) SPX return jumps intensity spline
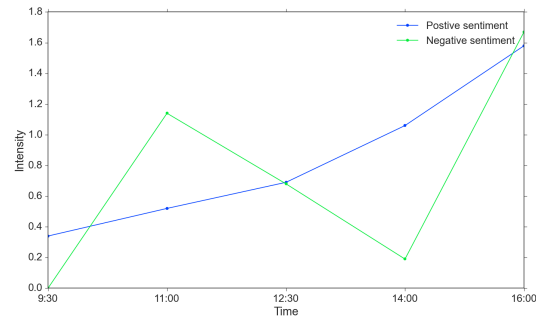
(b) News sentiment jumps intensity spline

Figure 6.: Spline intensity of return and sentiment

mutual-excitation for the return events, having larger π values, but the position is slightly reversed for the sentiment events with only two cross-excitation terms being less than the corresponding self-excitation terms while $\pi_{43} > \pi_{44}$ and $\pi_{31} > \pi_{32} > \pi_{33}$.

We take a closer look at the self- and mutual excitation, a few interesting results in relation to decay speed have emerged: 1) self-excitation decays much faster for positive and negative return events at both intra-day and inter-day levels. For example, among all excitation decays for positive returns, β_{11} overpowers β_{12}, β_{13} and β_{14}; while 2) mutual-excitation decays faster for positive

and negative sentiment events at both intra-day and inter-day levels. In this category, different mutual-excitation seem to have different level of impact on the decay speed, which depends on the initiating event type. In general, if both initiating and response events share the same nature (returns vs. returns or sentiment vs. sentiment), the excitation tends to be absorbed more quickly (e.g. $\beta_{31} > \beta_{34}$ and $\beta_{32} > \beta_{34}$ for one intra-day case).

We also detect significant mutual-excitation between positive return and positive sentiment, and negative return and negative sentiment (H5 and H6). These are well married to the argument of information responses that sequential responses to the movements in returns (sentiment) in one direction (positive or negative) would encourage the occurrence of the other type of events to move towards the same direction (positive or negative). The typical examples include speculative responses to market future returns or reactions to firms' earnings announcements etc. These are because traders and investors would either instantaneously or gradually form some kind of beliefs of market movements based on their interpretation of the information signals carried in these sequences of market events (see Brock *et al.* (1992) and Gwilym *et al.* (2016)). In addition, decaying speed of return is almost twice as fast as that of sentiment, which confirms the fact that market prices are more volatile and move faster than investor sentiment (see Heston and Sinha (2016)).

A special finding about market interaction between return and sentiment is that positive sentiment triggers negative price jumps (in H7) but not vice versa (in H8). One possible explanation for positive sentiment leading negative price jumps could be that investors have reverse mentality after experiencing a period of positive sentiment. For example, if a technical trader trading with header-and-shoulder pattern has experienced a fairly long period of a bull market and captured many heads, it is inevitable that he may become very cautious when he would hit the shoulders, especially the scenario that the market would suddenly drop into a bear market. For the fact that there is no evidence why negative price jumps lead to positive sentiment, we have managed to gain some insights from experienced traders claiming that typically no one would act on negative price jumps by naively assuming that the entire market would perceive it as something positive, especially when the market nowadays can operate at a much sophisticated level.

6.3.   *Discussion by event types*

In this section, we summarize the results from the view point of the four event types to further explain our findings:

- *Positive market return jumps*
  The seasonality of spline intensity is close to a U-shape (see Figure 6a). The value is the highest at the beginning of trading day, then it decreases to $\gamma_{1,3} = 0.50$ and increases until $\gamma_{1,4} = 1.33$. While the last intensity we capture for the end of trading is 0.00, which is indeed consistent with the fact that the occurrence of price jumps in the last few time slots is very rare (see Figure 5a). The third component of Equation 13 represents intraday excitation. We can split it into self-excitation and mutual-excitation. Self-excitation in positive return jumps is the strongest influence from all four events. There is mutual-excitement from positive news to positive return based on $\alpha_{13} = 0.12$. This validates our hypothesis H6 about sentiment impacts to return. In terms of these two lines of excitement, we notice that decaying of price shocks is much faster than that of sentiment shocks. Specifically, half-life of self-excitement is around 16 minutes versus 33 minutes from positive sentiment. The second component of Equation 13 represents daily-spillover in which π denotes the portion of influence that is passed from the previous trading day. As we do not detect intraday excitement from negative return and sentiment jumps, daily-spillover of these two events can be ignored. For the other two events, influence inherited from self-excitation is slightly higher than influence of positive news, and the half-life are 15 minutes and 29 minutes respectively.

- *Negative market return jumps*

The shape of spline intensity is the same as that of positive return jumps, presenting the U-shaped seasonality and a decreased intensity at the end of trading day (see Figure 6a). The strongest trigger of negative return shocks is also self-excitation with $\alpha_{22} = 0.19$. The same results for both positive and negative return jumps are consistent with clustered price shocks which is mentioned in hypothesis H1. The half-life of self-excitation decaying is around 15 minutes. We notice that both positive and negative sentiment jumps increase intensity of negative return. We cover this phenomenon in hypothesis H6 and H8. As the hypothesis H8 is not verified in the positive return jumps, we can conclude the asymmetric reactions to positive and negative news. Moreover, the strength of the two types of mutual-excitation are very close in terms of jump size, $\alpha_{23} = 0.14$ and $\alpha_{24} = 0.13$. The half-life decaying of mutual-excitation effect takes 25 minutes from positive sentiment and 26 minutes from negative sentiment respectively. In the daily-spillover part, mutual excitation from positive sentiment is much weaker than the other two. The half-life decaying time is 14 minutes for self-excitement, and 25 minutes and 23 minutes for positive and negative sentiment respectively.

- *Positive sentiment jumps*
  Sentiment jumps represent dramatic changes in investors attitude toward financial markets. In general, features of sentiment jumps intensity are different from return jumps. First, there is no U-shaped seasonality property for spline intensity of positive sentiment jumps. Instead, the intensity increases gradually from the beginning to the end of trading day (see Figure 6b). Second, $\alpha_{33} = 0.16$ shows self-excitation of positive sentiment events. However, the strongest trigger of events is the mutual-excitation from negative sentiment jumps rather than self-excitation. This type of mutual-excitation confirms hypothesis H4, indicating natural shifting of good news and bad news. Another cross-excitation trigger is positive return, showing that increasing market price leads to positive anticipation toward market returns (see hypothesis H5). Comparing with the two return jump processes, the effective time is longer for all three types of triggers. In detail, half-life decaying is around 23 minutes, 37 minutes, and 48 minutes for positive return, positive sentiment, and negative sentiment respectively. At last, around half of the influence from self-excitation and positive return is passed to the following trading day according to $\pi_{33} = 0.43$ and $\pi_{31} = 0.5$. The negative sentiment, albeit has strong influence during the same day, does not generate high impact to following days as π_{34} is only 0.10. Half-life of daily-spillover influence for positive return and sentiment are 20 minutes and 32 minutes.

- *Negative sentiment jumps*
  We find the U-shaped spline intensity in negative sentiment jumps, while it differs from return jumps in two ways (see Figure 6b). First, the beginning 1.5 hours are quiet for negative sentiment which is contrary to the top intensity of $\gamma_{1,1}$ and $\gamma_{2,1}$ for return jumps. The highest intensity appears at the second time interval, around 11:00AM, and the end of day. In negative sentiment jumps, we also notice the significant mutual-excitation from opposite sentiment events and the size is double that of self-excitation (see hypothesis H5). Half-life excitation decaying of positive and negative sentiment jumps last for 47 minutes and 39 minutes respectively. Another trigger is negative return jumps which verifies hypothesis H5. Its influence decays by half within 26 minutes. In this model, daily-spillover of positive sentiment is the strongest followed by self-excitation with half-life decaying of 33 minutes and 30 minutes.

7.   Conclusion

In this paper, we model financial market events as a multivariate point process emphasizing dynamics of investor sentiment shocks and market extreme returns. We find the following major in-

teraction effects between the four types of events, i.e. positive return event, negative return event, positive sentiment event, and negative sentiment event. First, both self-excitation and mutual-excitation are strong for all four types of events at 15 minutes time scale. We investigate interactions on higher time frequency (e.g. 2 minutes and 5 minutes) and find that the mutual-excitation between return and news sentiment is not present at smaller time scale. Secondly, there is a significant mutual-excitation between positive return and positive sentiment, and negative return and negative sentiment. Furthermore, the positive return only has cross-excitation effect from positive news. The mutual-excitation half-life lasts about 33 minutes, while the negative return has mutual-excitation effect from both positive sentiment and negative sentiment, and the mutual-excitation half-life lasts 25 minutes and 26 minutes respectively. Third, decaying speed of return events is almost twice as fast as that of sentiment events, which means market prices move faster than investor sentiment changes. While the positive return self-excitement half-life lasts around 16 minutes, the negative return self-excitation half-life lasts around 15 minutes: so they are in a similar range of decaying speed. Moreover, daily-spillover of excitation of return jumps is higher than that of sentiment jumps. One possible explanation for this phenomenon is the different movements of market and sentiment in off-trading hours. Generally, close market information has to be be kept and revealed until the next open market trading as no trading can be executed during off-trading hours. On the other hand, investor sentiment may be updated, or even be refreshed, during these time periods even though we do not record the changes in our models.

In addition, we observe that the news sentiment events spline intensity follows different patterns than the generally observed U-shaped intraday return patterns. While the positive sentiment jump intensity has an increase pattern from the beginning of the trading day to the end of the trading day, the negative sentiment follows more or less a U-shape pattern with some noise. Lastly, we observe that positive sentiment shocks tend to generate negative price jumps, and the cross-excitation between positive and negative sentiment is even stronger than their self-excitation of these sentiment events. There seems a contrarian effect at 15 minutes time scale, meaning positive news would trigger negative trading decisions at this time scale. News sentiment at 15 minutes time scale also may have the contrarian effect. As time moves on, the true effect will emerge at longer time scale. These two observations are not immediately obvious based on the current literature. Therefore, it requires further investigation. We suggest to investigate longer time scale jump events (e.g at 20 and 30 minute scales) to see whether such effect will continue, and then test the contrarian effect hypothesis.

8.   Acknowlegement

References

Aboura, S. and Chevallier, J., Leverage vs. feedback: Which Effect drives the oil market?. *Finance Research Letters*, 2013, **10**, 131–141.

Aït-Sahalia, Y. and Saglam, M., High frequency traders: Taking advantage of speed. *Working paper, Princeton University*, 2013.

Aït-Sahalia, Y., Cacho-Diaz, J. and Laeven, R.J., Modeling financial contagion using mutually exciting jump processes. Technical report, National Bureau of Economic Research, 2010.

Antoniou, A., Koutmos, G. and Pericli, A., Index futures and positive feedback trading: evidence from major stock exchanges. *Journal of Empirical Finance*, 2005, **12**, 219–238.

Antweiler, W. and Frank, M.Z., Is all that talk just noise? The information content of internet stock message boards. *The Journal of Finance*, 2004, **59**, 1259–1294.

Arnold, L.G. and Brunner, S., The economics of rational speculation in the presence of positive feedback trading. *The Quarterly Review of Economics and Finance*, 2015, **57**, 161–174.

Bacry, E., Dayri, K. and Muzy, J.F., Non-parametric kernel estimation for symmetric Hawkes processes. Application to high frequency financial data. *The European Physical Journal B-Condensed Matter and Complex Systems*, 2012, **85**, 1–12.

Bacry, E., Delattre, S., Hoffmann, M. and Muzy, J.F., Modelling microstructure noise with mutually exciting point processes. *Quantitative Finance*, 2013a, **13**, 65–77.

Bacry, E., Delattre, S., Hoffmann, M. and Muzy, J.F., Some limit theorems for Hawkes processes and application to financial statistics. *Stochastic Processes and their Applications*, 2013b, **123**, 2475–2499.

Bacry, E. and Muzy, J.F., Hawkes model for price and trades high-frequency dynamics. *Quantitative Finance*, 2014, **14**, 1147–1166.

Baker, M. and Wurgler, J., Investor sentiment and the cross-section of stock returns. *The Journal of Finance*, 2006, **61**, 1645–1680.

Baker, M. and Wurgler, J., Investor sentiment in the stock market. *The Journal of Economic Perspectives*, 2007a, **21**, 129–151.

Baker, M. and Wurgler, J., Investor Sentiment in the Stock Market. *The Journal of Economic Perspectives*, 2007b, pp. 129–151.

Barberis, N., Shleifer, A. and Vishny, R., A model of investor sentiment. *Journal of financial economics*, 1998, **49**, 307–343.

Bauwens, L. and Hautsch, N., Modelling Financial High Frequency Data Using Point Processes. *Handbook of Financial Time Series*, 2009, **1**, 953–979.

Beer, F. and Zouaoui, M., Measuring stock market investor sentiment. *Journal of Applied Business Research*, 2013, **29**, 51.

Bowsher, C.G., Modelling security market events in continuous time: Intensity based, multivariate point process models. *Journal of Econometrics*, 2007, **141**, 876–912.

Brémaud, P., *Point processes and queues.*, 1981 (Springer-Verlag, Halsted Press: Berlin/New York).

Brock, W., Lakonishok, J. and LeBaron, B., Simple technical trading rules and the stochastic properties of stock returns. *The Journal of finance*, 1992, **47**, 1731–1764.

Brown, G.W. and Cliff, M.T., Investor sentiment and the near-term stock market. *Journal of Empirical Finance*, 2004, **11**, 1–27.

Buckle, M., Chen, J. and Hawkes, A., Flash Crashes, Jumps and Running Jumps: A New Method for Jump Detection. *Working paper, SSRN*, 2017.

Carlsson, J., Foo, M.C., Lee, H.H. and Shek, H., High Frequency Trade Prediction with Bivariate Hawkes Process. *Stanford University Working Paper*, 2007.

Chau, F. and Deesomsak, R., Business cycle variation in positive feedback trading: Evidence from the G-7 economies. *Journal of International Financial Markets, Institutions and Money*, 2015, **35**, 147–159.

Chavez-Demoulin, V., Davison, A.C. and McNeil, A.J., A point process approach to value-at-risk estimation. *Quantitative Finance*, 2005, **5**, 227–234.

Chavez-Demoulin, V. and McGill, J., High-frequency financial data modeling using Hawkes processes. *Journal of Banking & Finance*, 2012, **36**, 3415–3426.

Chen, J., Dong, Y., Hou, W. and McMillan, D., Does feedback trading drive returns of cross-listed shares?. *Journal of International Financial Markets, Institutions and Money*, 2017, **Forthcoming**.

Chen, J., Hawkes, A. and Khashanah, K., A slightly depressing jump model: Intraday volatility pattern simulation. *Quantitative Finance*, 2018, **Forthcoming**.

Chopra, N., Lakonishok, J. and Ritter, J.R., Measuring abnormal performance: do stocks overreact?. *Journal of financial Economics*, 1992, **31**, 235–268.

Cohen, G. and Kudryavtsev, A., Investor rationality and financial decisions. *Journal of Behavioral Finance*, 2012, **13**, 11–16.

Daniel, K., Hirshleifer, D. and Subrahmanyam, A., Investor psychology and security market under-and overreactions. *the Journal of Finance*, 1998, **53**, 1839–1885.

Daniel, K.D., Hirshleifer, D. and Subrahmanyam, A., Overconfidence, arbitrage, and equilibrium asset pricing. *The Journal of Finance*, 2001, **56**, 921–965.

De Bondt, W.F., Stock price reversals and overreaction to news events: A survey of theory and evidence. In *A reappraisal of the efficiency of financial markets*, pp. 57–84, 1989, Springer.

De Long, J.B., Shleifer, A., Summers, L.H. and Waldmann, R.J., Noise trader risk in financial markets. *Journal of political Economy*, 1990, **98**, 703–738.

Dodds, P.S. and Danforth, C.M., Measuring the happiness of large-scale written expression: Songs, blogs, and presidents. *Journal of happiness studies*, 2010, **11**, 441–456.

Embrechts, P., Liniger, T., Lin, L. *et al.*, Multivariate Hawkes processes: an application to financial data. *Journal of Applied Probability*, 2011, **48**, 367–378.

Errais, E., Giesecke, K. and Goldberg, L.R., Affine point processes and portfolio credit risk. *SIAM Journal on Financial Mathematics*, 2010, **1**, 642–665.

Fauth, A. and Tudor, C.A., Modeling first line of an order book with multivariate marked point processes. *arXiv preprint arXiv:1211.4157*, 2012.

Fulop, A., Li, Y. and Yu, J., Self-Exciting, Jumps, Learning, and Asset Pricing Implications. *Singapore Management University Working Paper*, 2015.

Gwilym, O.A., Hasan, I., Wang, Q. and Xie, R., The Effects of Speculative Demand of Stock Returns. *European Financial Management*, 2016, **22**, 427–449 Doi:10.1111/eufm.12067.

Hawkes, A.G. and Adamopoulos., L., Cluster models for earthquakes: regional comparisons. *Bull. Int. Statist. Inst.*, 1973, pp. 454–461.

Hawkes, A.G., Point spectra of some mutually exciting point processes. *Journal of the Royal Statistical Society. Series B (Methodological)*, 1971a, pp. 438–443.

Hawkes, A.G., Spectra of some mutually exciting point processes with associated random variables. *Biometrika*, 1971b, **58**, 83–90.

Hawkes, A.G., Spectra of some self-exciting and mutually exciting point processes. In *Stochastic point processes*, edited by P.A.W. Lewis, pp. 261–271, 1972 (Wiley: New York).

Hawkes, A.G. and Oakes, D., A Cluster Process Representation of a Self-Exciting Process. *Journal of Applied Probability*, 1974, **11**, 493–503.

Healy, A.D. and Lo, A.W., Managing real-time risks and returns: the thomson reuters newsscope event indices. *Professor David J. Hand, Professor of Statistics, Imperial College, London; Chief Scientific Advisor, Winton Capital Management; and President, Royal Statistical Society*, 2011, p. 73.

Herrera, R. and Schipp, B., Value at risk forecasts by extreme value models in a conditional duration framework. *Journal of Empirical Finance*, 2013, **23**, 33–47.

Heston, S.L. and Sinha, N.R., News versus Sentiment: Predicting Stock Returns from News Stories. *Finance and Economics Discussion Series 2016-048. Washington: Board of Governors of the Federal Reserve System*, 2016, **048** Http://dx.doi.org/10.17016/FEDS.2016.048.

Hilton, D.J., The psychology of financial decision-making: Applications to trading, dealing, and investment analysis. *The Journal of Psychology and Financial Markets*, 2001, **2**, 37–53.

Hirshleifer, D., Subrahmanyam, A. and Titman, S., Feedback and the success of irrational investors. *Journal of Financial Economics*, 2006, **81**, 311–338.

Hou, Y. and Li, S., The impact of the CSI 300 stock index futures: Positive feedback trading and autocorrelation of stock returns. *International Review of Economics & Finance*, 2014, **33**, 319–337.

İnkaya, A. and Okur, Y.Y., Analysis of volatility feedback and leverage effects on the ISE30 index using high frequency data. *Journal of Computational and Applied Mathematics*, 2014, **259**, 377–384.

Jaisson, T., Market impact as anticipation of the order flow imbalance. *arXiv preprint arXiv:1402.1288*, 2014.

Jaisson, T. and Rosenbaum, M., Limit theorems for nearly unstable Hawkes processes. *The Annals of Applied Probability*, 2015, **25**, 600–631.

Jaisson, T. and Rosenbaum, M., Rough fractional diffusions as scaling limits of nearly unstable heavy tailed Hawkes processes. *The Annals of Applied Probability*, 2016, **26**, 2860–2882.

Karabash, D. and Zhu, L., Limit Theorems for Marked Hawkes Processes with Application to a Risk Model. *Stochastic Models*, 2015, **31**, 433–451.

Khanna, N. and Sonti, R., Value creating stock manipulation: feedback effect of stock prices on firm value. *Journal of financial markets*, 2004, **7**, 237–270.

Laopodis, N.T., Feedback trading and autocorrelation interactions in the foreign exchange market: Further evidence. *Economic Modelling*, 2005, **22**, 811–827.

Large, J., Measuring the resiliency of an electronic limit order book. *Journal of Financial Markets*, 2007, **10**, 1–25.

Lee, C., Shleifer, A. and Thaler, R.H., Investor sentiment and the closed-end fund puzzle. *The Journal of Finance*, 1991, **46**, 75–109.

Lo, A.W., The adaptive markets hypothesis. *The Journal of Portfolio Management*, 2004, **30**, 15–29.

Merton, R.C., On estimating the expected returns on the market. *Journal of Finance and Economics*, 1980,

8, 323–361.

Mian, G.M. and Sankaraguruswamy, S., Investor sentiment and stock market response to earnings news. *The Accounting Review*, 2012, **87**, 1357–1384.

Mishne, G., De Rijke, M. *et al.*, Capturing Global Mood Levels using Blog Posts.. In *Proceedings of the AAAI spring symposium: computational approaches to analyzing weblogs*, pp. 145–152, 2006 (California, USA).

Mitra, L. and Mitra, G., 1. In *Applications of news analytics in finance: A review*, pp. 1–39, 2011, John Wiley & Sons.

Mo, S.Y.K., Liu, A. and Yang, S.Y., News sentiment to market impact and its feedback effect. *Environment Systems and Decisions*, 2016, **36**, 158–166.

Nofsinger, J.R., Social mood and financial economics. *The Journal of Behavioral Finance*, 2005, **6**, 144–160.

Ogata, Y., The asymptotic behaviour of maximum likelihood estimators for stationary point processes. *Annals of the Institute of Statistical Mathematics*, 1978, **30**, 243–261.

Porta, R.L., Lakonishok, J., Shleifer, A. and Vishny, R., Good news for value stocks: Further evidence on market efficiency. *The Journal of Finance*, 1997, **52**, 859–874.

Russell, J.R., Econometric modeling of multivariate irregularly-spaced high-frequency data. *University of Chicago, Graduate School of Business working paper series*, 1999.

Salm, C.A. and Schuppli, M., Positive feedback trading in stock index futures: International evidence. *International Review of Financial Analysis*, 2010, **19**, 313–322.

Sentana, E. and Wadhwani, S., Feedback traders and stock return autocorrelations: evidence from a century of daily data. *Economics Journal*, 1984, **102**, 415–425.

Shiller, R.J., Stock prices and social dynamics. *Brookings Papers on Economic Activity*, 1984, **2**, 457–498.

Shleifer, A., *Inefficient Markets: An introduction to behavioural finance*, 2000 (Oxford University Press: Oxford).

Shleifer, A. and Vishny, R.W., The limits of arbitrage. *The Journal of Finance*, 1997, **52**, 35–55.

Smales, L.A., News sentiment and the investor fear gauge. *Finance Research Letters*, 2014a, **11**, 122–130.

Smales, L.A., News sentiment in the gold futures market. *Journal of Banking & Finance*, 2014b, **49**, 275–286.

Tetlock, P.C., Giving content to investor sentiment: The role of media in the stock market. *The Journal of Finance*, 2007, **62**, 1139–1168.

Tetlock, P.C., Saar-Tsechansky, M. and Macskassy, S., More than words: Quantifying language to measure firms' fundamentals. *The Journal of Finance*, 2008, **63**, 1437–1467.

Thaler, R.H., *Advances in behavioral finance*, Vol. 2, , 2005 (Princeton University Press: Princeton).

Tse, Y. and Tsui, A., A multivariate GARCH model with time-varying correlations. *Journal of Business Economic Statistics*, 2002, **20**, 351–362.

Yang, S.Y., Mo, S.Y.K. and Liu, A., Twitter financial community sentiment and its predictive relationship to stock market movement. *Quantitative Finance*, 2015, **15**, 1637–1656.

Zhu, L., Limit Theorems for a Cox-Ingersoll-Ross Process with Hawkes jumps. *Journal of Applied Probability*, 2014, **51**, 699–712.