Stock Price Jumps and News Sentiment: A Case of Investor Overreaction

Article *in* The Journal of Prediction Markets · January 2017

DOI: 10.5750/jpm.v11i2.1355

Ritu Yadav
Indian Institute of Management, Rohtak

The Journal of Prediction Markets 2017 Vol 11 No 2 pp 20-

STOCK PRICE JUMPS AND NEWS SENTIMENT: A CASE OF INVESTOR OVERREACTION

A Vinay Kumar[*]
Professor of Finance
Indian Institute of Management, Lucknow

Ashwani Kumar
Professor of IT and Systems
Indian Institute of Management, Lucknow

Ritu Yadav
Post-Doctoral Fellow
Indian School of Business

ABSTRACT

The research paper develops an understanding on how news based sentiment capture investor behaviour reflected in price jumps in stock markets. It compares the impact on two models of stock price jumps; the non-parametric model proposed by BNS and the wavelet based method. The study is also a perspective on the semi strong form of market efficiency

Using the high frequency data from the stock and options market along with the actual high frequency news data from Bloomberg, the two alternative methodologies of jumps have been tested. In addition, options trades have been simulated to see whether profits can be earned from the news sentiment captured by jumps.

Methodologically, jumps based on wavelets were found to be better related with the news sentiment compared to the BNS method. Also, the news sentiment based jumps were found to present opportunities in the simulated trades that could be exploited for earning profits suggesting that investors overreact.

The paper uses an innovative method for computation of the news based sentiment. To the best of our knowledge, the paper is the first to evaluate jumps and news sentiment using the actual news data. A perspective on the semi strong form of market efficiency is presented, that too by departing from the event study based models.

Keywords: BNS jumps, Wavelet Jumps, News Sentiment, Overreaction, Behavioural Finance, Market Efficiency

---

[*] Corresponding author: Prof A Vinay Kumar,Professor of Finance, Indian Institute of Management, Lucknow. Email: vinay@iiml.ac.in. This research is funded by Indian Institute of Management, Lucknow, India

1 INTRODUCTION

The study of discontinuities which are characterised as jumps attracted many research endeavours concerning option pricing, volatility modelling and risk earlier. For instance, Merton (1976) presents a strong case for options pricing with jumps and implications to specification errors due to exclusion of jumps. Besides Cox and Ross (1976), and Eraker et al.(2003) argued that incorporating jumps is essential to price options. While Jorion (1989) argues the presence of discontinuities in the form of jumps in the foreign exchange markets and the need to model them. Similarly authors like Maheu and McCurdy (2004) and Andersen et al., (2007) argued the need for incorporating jumps to estimate volatility of an asset. Also, jump risk and its impact is reported to be a significant component in pricing (Yan, 2011). Most of these studies consider jumps as an additional process to a well behaved log normal process. It is argued that both of these process in conjunction would lead to superior modelling results in the context of their research (Ait Sahalia (2002), Ait Sahalia (2004) Cont and Tankov (2009). Needless to say all of these studies adhere to the expected utility framework with rational investors.

While a lot of continuing research effort on modelling jumps is being witnessed, there seems to be sparse research on understanding jumps as case of investor sentiment and behaviour. Unlike the expected utility framework the notion of an investor following trends is popular in behavioural finance. Lakonishok et al. (1994) examined selling and buying strategies based certain popular measures of overvaluation and undervaluation. The study reports that due investor overestimation the strategies make money. Daniel at al (1998) reports the optimism bias of investors where investors give undue value to their anticipated price. Also post earnings announcement the stock price drifts gradually for over 60 days in the expected direction. Positive announcement will lead the stock price increases and negative announcement will lead decreases in the months (Bernard and Thomas (1990). Jumps in isolation could be a way to understand investor overreaction and the inclination to follow a trend which is contrary scenario to the efficient markets. Understanding jumps in isolation would be a better way to capture trend based trading and market sentiment. Investors at the back of hot news would follow a sort of herd and try and be part of the trend thus creating a fairly large jump in the price. Broadly jumps are believed to be caused by the unexpected arrival of news. News arrival evokes and shapes the sentiment in the market and sometimes results in these extreme changes. It is only germane that some news dramatically changes investor sentiment, resulting in market actions of extreme buying or selling. The question is; does news arrival represented by sentiment lead the jumps? Also, if jumps are identified; the jump direction should match the polarity or positivity or negativity of the news/sentiment. Does the polarity match with the jump direction? The paper studies these questions and provides a methodological alternative to studying and understanding the semi strong form of market efficiency. In the semi strong of form of market efficiency, market prices should instantaneously adjust to the public news arrival. The traditional finance studies use event studies to understand the semi strong form of market efficiency. However, the event studies have a selectivity bias. The present study proposes a much better method of understanding the semi strong form of market efficiency. The study explains the behavioural dynamics of the agents while they trade and brings out the sentiment associated with their trading. Understanding jumps with public news could throw some light on Barker's (1998) question of how information gets subsumed into prices as the efficient market hypothesis is unable to answer such questions.

The empirical analysis in the study includes high frequency data of 8 stocks traded on the National Stock Exchange of India and the relevant high frequency news related to the stocks provided have been extracted from the Bloomsberg terminal. The two methods for jump identification, namely the maximum overlap discrete wavelet transformation (MODWT) suggested by Sun and Meinl (2012) and the non-parametric method suggested by Barndorff-Nielsen and Shepherd (2006) have been used; and a new method for the news sentiment analysis has been evolved for the purpose. It was found that the news sentiment leads jumps and there is a significant match in polarity and the jump direction. The performance of wavelet based jump identification seems to be better than the non-parametric method suggested by Barndorff-Nielsen and Shepherd (2006) (hereafter referred to as the BNS method). Also, there are profitable trading opportunities that arise in simulated trades for jumps identified by news sentiment which in turn means that investors have overreacted. This work will throw open many future research possibilities, e.g. learning and replication of jumps using machine learning method. To this best of our understanding we have not seen any work that modelled news sentiment and jumps to elucidate investor behaviour.

2 LITERATURE REVIEW

News based sentiment analysis gained a lot of significance since the time Wysocki (1998) studied Yahoo Finance message board's relevance on stock prices movement. News based sentiment literature over the years developed either as Dictionary based methods or as machine learning based methods. Market news is increasing available in various forms including the high frequency wire news services like Bloomberg, Reuters and Newswire18.

Dictionary based news sentiment is computed using the word lists that are preselected with a semantic sentiment assigned to them (Stone et al.,1962; Loughran and McDonald, 2011). The assumption is that the news text sentiment reflects markets sentiment fairly well. These words when found in the just arrived news article can be immediately translated to their sentiment equivalent (Ritu et al. ,2013). Assigning semantic sentiment is done manually (Das and Chen ,2001) or using well-established sentiment lexicons (Tetlock

,2007). Computational linguistics methodologies can also be used in determining the sentiment.

Wysocki (1998) used the message board of Yahoo finance and computed the sentiment by counting the words. The sentiment thus computed was related to trading volume and abnormal stock returns where he found significant correlation. Devitt and Ahmad (2007) extracted market sentiment related features through the words and sentences using textual analysis. The area opened up discussion on the appropriateness, given the context and the efficiency of the dictionary based news sentiment computation. Most of the earlier works used online blogs, news articles, message boards or micro-blogging sites to derive the sentiment; however, there seems to be little work done when it comes to high frequency wired news service that most financial institutions use. The area offers immense opportunities in this context. In the past, sentiment analysis works have been conducted to explore correlation and predictability of news sentiment and financial market movements (Devitt and Ahmad, 2007; Tetlock, 2007; Schumaker et al,2012). News sentiment has been found to be correlated with market movements (Ahmad, 2007; Tetlock, 2007), as well as holding significant predictability on the assets (Schumaker et al, 2012). However, to the best of our knowledge, this is the first study conducted to explore correlation between market jumps and news sentiment.

On the other hand, jump identification literature has grown over the recent years. The primary purpose of these studies was to understand the structure of asset returns so that a suitable model can be developed to understand returns and volatility of assets. The assets returns are usually characterized in two main parts; the continuous diffusive part and the jump part. A limited selection of the papers in the area that made significant contribution include Aït-Sahalia (2002), Huang and Tauchen (2005), Barndorff-Nielsen and Shephard (BNS: 2006), Cont and Mancini (2007), Jiang and Oomen (2008), Lee and Mykland (2008), Tauchen and Todorov (2008), and Aït-Sahalia and Jacod (2009a, b).

Jumps can be extracted using the wavelets based frequency decomposition. There are several ways to do it. For the purpose of the study wavelet based multi-scale approach as developed by the referred studies such as Sun and Meinl (2012), and Fan and Wang (2007) is used.

Relating news with price, volumes and volatility was taken up by earlier studies. Huang (2007, 2015) studied macroeconomic news announcements and its relationship with volatility and news. The papers report that jumps are prevalent in news days rather than non-news days. However, the papers use macroeconomic news only which is quite restrictive.

3 METHODOLOGY AND DATA

This paper identifies jumps using two methods; first, the jump diffusion price model with noise as is being computed in the referred studies such as Wei (2012), Boudt et al. (2012), and Bollerslev et al. (2013); and the other is the wavelet based multi-scale approach as developed by the referred studies such as Sun and Meinl (2012), and Fan and Wang (2007). A price jump is a significant discontinuity in the price process that occurs due to new and unexpected market information. In jump diffusion price model, log-price process X_t is a combination of latent log price process $\overline{X_t}$ and a noise process ε_t^X:

$$X_t = \overline{X_t} + \varepsilon_t^X \tag{1}$$

The latent process $\overline{X_t}$ is supposed to be a Brownian semi-martingale process with finite activity jumps. This means that $\overline{X_t}$ can be decomposed into its constituent parts that consist of drift, diffusive variance and jump components as shown in equation 2 below

$$d\bar{X}_t = \mu_t dt + \sigma_t dW_t + K_t dq_t \tag{2}$$

Where μ_t is the drift term with continuous and locally finite variation sample path, σ_t is strictly positive spot volatility process and W_t is a standard Brownian motion. The component $K_t dq_t$ refers to pure jump counting process where $dq_t$is 1 if there is a jump at time *t* else 0. Boudt et al (2012) identified intra-day jumps $K_t dq_t$ in a high frequency price data holding two assumptions:

a) Noise process ε_t^X is independent of X; and
b) $\varepsilon_t^X \sim N(0, \sigma_{\varepsilon_x}^2)$

With these assumptions, it can be implied that with time interval being sufficiently small, high frequency return is normally distributed even in the presence of noise but with no jumps. This study identified intra-day jumps in two steps

1. Identification of jump days using non-parametric Barndorff-Nielsen and Shepherd (BNS) test (Barndorff-Nielsen and Shephard, 2006).
2. Intra-day jump time is identified using Bollerslev's Time of the Day (TOD) measure (Bollerslev, Todorov, & Li, 2013).

Jumps can be identified by detecting time stamps where variation due to jump component is significant. For this, separation of variations due to jump as well as the continuous part is important. The BNS test detects jump related variations using the difference between a measure of quadratic variation and a jump robust measure of variation such a bi-power variation (BPV).

For sampling frequency δ that divided a day into m equal time intervals, realized variance RV_t is taken as a measure of quadratic variation:

$$RV_t(\delta) = \sum_{j=1}^{m} r_{t-1+j\delta}^2 \rightarrow \int_{t-1}^{t} \sigma(s)^2 ds + \sum_{t-1<s\leq t} K^2(s) \qquad (3)$$

Where, r is the log price return series. Estimated quadratic variation RV consists of two components; one due to price variation from continuous flow while another due to discontinuous jumps.

Bi-power variation converges to the integrated variance alone.

$$BV_t(\delta) = \left(\frac{2}{\pi}\right)^{-1} \left(\frac{m}{m-1}\right) \sum_{j=2}^{m} \left|r_{t-1+j\delta}\right| . \left|r_{t-1+(j-1)\delta}\right| \rightarrow \int_{t-1}^{t} \sigma(s)^2 ds \qquad (4)$$

Hence, BNS relative jump measure RJ_t is calibrated as the contribution of jumps to the total variation as follows:

$$RJ_t = \frac{RV_t - BV_t}{RV_t} \qquad (5)$$

Where, $RJ_t > 0$ is only for the days with at least one jump. Huang and Tauchen (2005) further improvised the relative jump measure by *max-adjusting* the metric that is asymptotically standard normal.

$$z_t = \frac{RJ_t}{\sqrt{\left(\left(\frac{\pi}{2}\right)^2 + \pi - 5\right)\frac{1}{m} max\left\{1, \frac{TP_t}{BV_t^2}\right\}}} \qquad (6)$$

Where TP_t is the tri-power quarticity that converges to the integrated quarticity.

$$TP_t = \mu_{4/3}^3 m \left(\frac{m}{m-2}\right) \sum_{j=3}^{m} \left|r_{t-1+j\delta}\right|^{4/3} . \left|r_{t-1+(j-1)\delta}\right|^{4/3} . \left|r_{t-1+(j-2)\delta}\right|^{4/3} \rightarrow \int_{t-1}^{t} \sigma^4(s) ds \qquad (7)$$

Where, $\mu_{4/3} = 2^{2/3}\Gamma\left(\frac{7}{6}\right)\Gamma\left(\frac{1}{2}\right) \approx 0.8309$. The exact time of the jump is further identified using Time of Day (TOD) measure proposed by Bollerslev et al (2013). TOD_i is the ratio of the diffusive variation over different part of the day *j=1,2,…m* relative to its average value of the day for *t=1,2,…T* days, given as following:

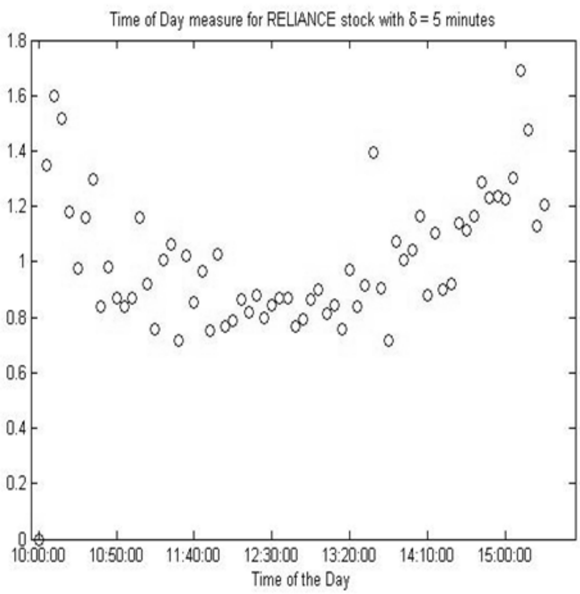
$$TOD_j = \frac{m \sum_{t=1}^{T} r_{t-1+j\delta}^2 . I\left(\left|r_{t-1+j\delta}\right| \leq \tau\sqrt{BV_t \Lambda\, RV_t} m^{-\bar{\omega}}\right)}{\sum_{t=1}^{T}\sum_{j=1}^{m} r_{t-1+j\delta}^2 . I\left(\left|r_{t-1+j\delta}\right| \leq \tau\sqrt{BV_t \Lambda\, RV_t} m^{-\bar{\omega}}\right)} \qquad (9)$$

Where I(.) is the indicator function that ensures that abnormally high returns are not included in the calculations. This study adopts empirically optimal value of $\tau = 2.5$ and $\bar{\omega} = 0.49$ observed in Bollerslev's simulation results (Bollerslev, et al., 2013). The resulting TOD measure generally exhibit a U-shaped pattern as a function of time interval j over a trading day as can be noticed in figure 1. The intra-day threshold that separates continuous price moves from jump process for day t and time interval j can then be given by $\alpha_{t,j}$ such that:

$$\alpha_{t,j} = \tau\sqrt{(BV_t \Lambda\, RV_t) * TOD_j} \tag{10}$$

Time stamp (t,j) for which $\left|r_{t-1+j\delta}\right| \geq \left(\alpha_{t,j} * m^{-\bar{\omega}}\right)$ correspond to intra-day jumps.

Figure 1: Time of the day measure for Reliance



Non-parametric models like asset price diffusion model proposed by Barndorff-Nielsen and Shepherd (2006) assume that observed high frequency data has true underlying returns. In BNS test, it is assumed that realized variation RV as the estimation of quadratic variation will have no micro-structure noise ε_t^X; and that bi-polar variation BV is a jump robust measure of variation. However, observing RV for different sampling periods as shown in figure 2 for RELIANCE stock shows that RV shoots up with shorter sampling periods. This suggests the presence of the micro-structure noise. Also, if the outliers in BV are removed, statistically the expected BV is greater than jump adjusted BV as can be observed in figure 3, implying presence of jumps. Hence bi-polar variation BV is not a jump robust measure of variation.

Figure 2: Realized Volatility of Reliance against sampling period

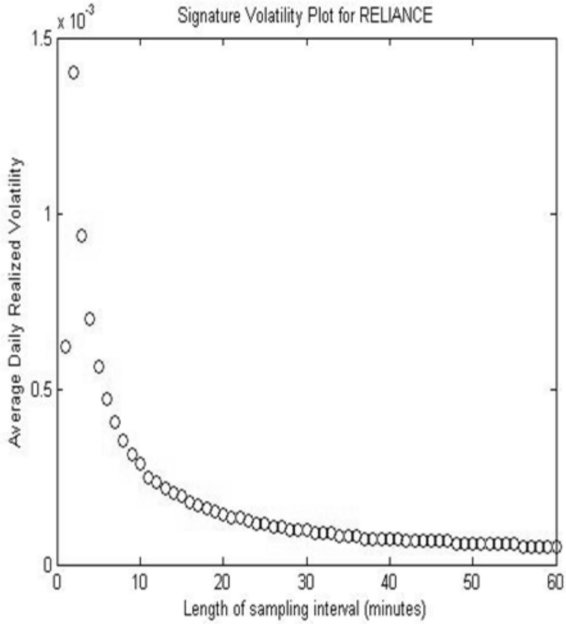
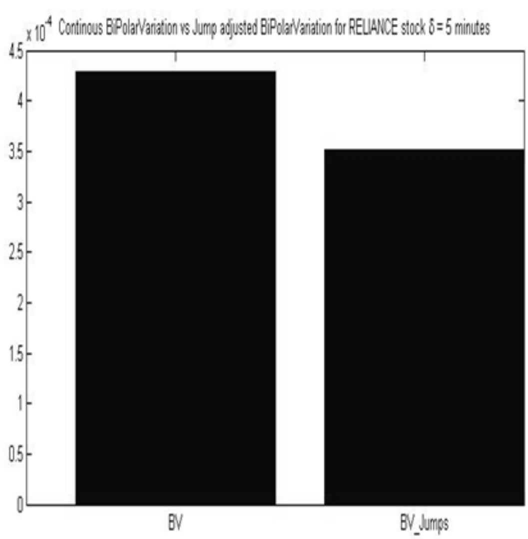


Figure 3: Bivariate Volatility   of Reliance with and without jumps



The jumps were identified using the TOD measure and return lags around TOD is considered for this study to align news with jumps. This is not the best method but is the only way to capture the jumps to isolate them from the time series.

Alternatively in order to identify price jumps, separating variations due to jump and continuous part is very important. To detect jump locations efficiently when prices follow a jump diffusion process, wavelet methods have been found to be a robust approach (Fan and Wang, 2007) (Sun and Meinl, 2012). This study adopts a modified wavelet based de-noising algorithm for detecting jumps, using local linear scaling approximation (LLSA) on linear maximal overlap discrete wavelet transform (MODWT) as proposed by Sun and Meinl (2012).

Wavelets are compact small orthonormal wave base that are capable of analysing signals in a localized time-frequency scale. This orthonormal wave base provides a multi-resolution analysis of a time series dilated and translated on different scales, usually dyadic. Discrete Wavelet Transform (DWT) is an orthogonal transform of a time series, in our case log-price process X of length nT, into L wavelet coefficient vector $W_j \in \mathbb{R}^{nT/2^l}$, $1 \leq l \leq L$ where $L \sim log2(nT)$ and one scaling coefficient vector $V_L \in \mathbb{R}^{nT/2^L}$ i.e.

$$\Psi X = [W_1, \dots W_L, V_L] \tag{11}$$

Fan and Wang (2007) used log-price process' DWT transform to detect the jumps. Applying DWT on price diffusion model, we obtain

$$X_{l,k} = \overline{X_{l,k}} + \varepsilon_{l,k}^{X} \tag{12}$$

Where, $l = 1, \dots, log2(nT); k = 1, .., 2^l$. Fan and Wang (2007) proposed that if a function contains a jump but is Holder-continuous, then the wavelet coefficients of the high pass-filter at some level l^{jump} close to jump-point decay at order$2^{\frac{1}{2}l^{jump}}$; while at no-jump point's decay at order $2^{\frac{3}{2}l^{jump}}$. Authors suggest an optimal level $2^{l^{jump}} \sim nT/log2(nT)$ where near-by jump points of log-price process will be significantly larger than others. This optimal scale for the jump identification will be referred as the "jump" scale in this study.

Because of DWT transformation, jump location could only be estimated with an information loss in time domain. Hence, this study used maximum overlap discrete wavelet transformation (MODWT) that generates an equal number of wavelet coefficients as the original time series at every level. Combined with zero phase correction, the locations of the wavelet coefficients naturally reveal information concerning the real data in original time domain.

This study used local linear scaling approximation (LLSA) on linear maximal overlap discrete wavelet transform (MODWT) as proposed by Sun and Meinl (2012).

4 NEWS SENTIMENT ANALYSIS

Affective content of a news can be calculated using a financial sentiment dictionary. Typically, a dictionary based sentiment is calculated using word counts belonging to different polarities and strengths used in the text (Das and Chen, 2007) (Schumaker and Chen, 2008). One of the most conventional methods of calculating positive or negative sentiment of a news story is to count the number of positive or negative sentiment words respectively in the given news text. However, a news article, as a standard piece of journalism, follows an inverted triangle layout (Dijk, 1988). Hence sentiment words featuring higher in a news story carry more value as compared to the words featuring in the last. Adapting an inverted triangle layout assumption for a standard news story reported by a financial source, this study proposes a new dictionary based *text* sentiment algorithm as follows.

Let a news story n_i contain $|P_i|$ paragraphs where P_i is set of paragraphs in news story n_i and p_{i0} is the heading

$$n_i = \{p_{i0}, p_{i1}, p_{i2}, p_{i3}, ...\} \tag{13}$$

For each paragraph p_{ij}, let d_{ij} is the set of dictionary based word counts for different polarities or sentiment categories; for instance, Loughran and McDonald (2009) features six sentiment categories – *positive, negative, weak, strong, uncertain,* and *litigious*. d_{ih} is the sentiment word count for heading. Then, sentiment metric d_i based on inverted triangle layout can be approximated as follows:

$$d_i = d_{i0} + d_{i1} + d_{i2}/2 + d_{i3}/2^2 + ...d_{i|P|}/2^{|P|-1} \tag{14}$$

While positive and, negative sentiment categories correspond to the news text sentiment, weak and strong categories of words correspond to the tone of sentiment. This study integrates the two in one final sentiment metric. Litigious words are ignored for the current work. Loughran and McDonald (2009) sentiment dictionary is supplemented with a list of following negation tokens – *no, not, none, neither, never, nobody, 'nt*, that can reverse the meaning or sentiment of a sentence.

Assuming an inverted triangle layout of a news story, the paper proposes the following algorithm for a dictionary based sentiment calibration of news. Each event tuple belonging to a news story inherits this sentiment measure. Let the dictionary based sentiment information for an i^{th} textual instance be denoted by positive - d_i^+, negative - d_i^-, uncertain - d_i^x, weak - d_i^{weak}, strong

- d_i^{strong} , and negation - $d_i^{negation}$ sentiments. And for a paragraph p_{ij} of news story n_i, let s_{ijk} be the k^{th} sentence of paragraph p_{ij}. For a given new story n_i, sentiment vector $[d_i^+, d_i^-, d_i^x]$ is calibrated as follows:

Input:
News Story n_i
Output:
Sentiment vector $[d_i^+, d_i^-, d_i^x]$
BEGIN
Step 1: For paragraph j = 0, let its weight be w_j (w_j=1 for j=0,1) as per inverted triangle layout

Step 1.1.: Get sentiment metric for positive d_{ijk}^+, negative d_{ijk}^-, uncertain d_{ijk}^x, weak d_{ijk}^{weak}, strong d_{ijk}^{strong} , and negation $d_{ijk}^{negation}$ sentiments for each sentence s_{ijk}.

Step 1.2: Scale up/down positive and negative sentiment with strong/weak sentiment tone respectively; and reverse the sentiment if negation is present in sentence s_{ijk}

$$d_{ijk}^+ = d_{ijk}^+ \times \left(2\middle|d_{ijk}^{strong} > 0\right) \times \left(0.5\middle|d_{ijk}^{weak} > 0\right) \times (-1|d_{ijk}^{negation} \sim = 0)$$

$$d_{ijk}^- = d_{ijk}^- \times \left(2\middle|d_{ijk}^{strong} > 0\right) \times \left(0.5\middle|d_{ijk}^{weak} > 0\right) \times (-1|d_{ijk}^{negation} \sim = 0)$$

Step 1.3: Sentiment for paragraph p_j (where $j=0$ refers to heading) is given by:

$d_{ij}^+ = \sum_k d_{ijk}^+$; $d_{ij}^- = \sum_k d_{ijk}^-$; $d_{ij}^x = \sum_k d_{ijk}^x$

Step 1.4: Downscale paragraph p_j sentiment with the location of paragraph in news story and add to the overall sentiment of the story.

$d_i^+ = d_i^+ + d_{ij}^+/w_j$; $d_i^- = d_i^- + d_{ij}^-/w_j$; $d_i^x = d_i^x + d_{ij}^x/w_j$

Step 2: Repeat Step for next paragraph of news story n_i with updated weight $w_j = w_j/2$ if $j>1$ else $w_j=1$.
END

Algorithm I: Dictionary based Sentiment Analysis algorithm based on Inverted Triangle Layout.

5 DATA ANALYSIS

This work is set in the Indian Futures Market of year 2009. Year 2009 was an eventful year for the Indian financial markets that had the famous 2008 credit crisis in the background to start with. In the year 2009, Indian financial markets went through troughs and crests with the Satyam bankruptcy crisis (NSE, 2009 accessed in 2014) weakening rupee, and euphoric election results (Economic, 2009 accessed in 2014) to name a few. Duration of the study is from January 1, 2009 to December 31, 2009.This study analyses top eight most traded companies of the National Stock Exchange, Mumbai (India), for year 2009. To detect jumps in the log-price process, underlying assumption of BNS test is that the microstructure noise is independently and identically distributed. However, the assumption's validity is conditional upon the choice of sampling interval (Rognlie, 2010). At small intervals, realized variance is very high due to inflated microstructure noise. This effect diminishes as the sampling interval is increased, but too large a sampling interval may obscure the jump detection process. An optimal sampling interval; with which the objectives of robustness to microstructure noise and preserving of asymptotic properties of our estimators is satisfied; can be heuristically chosen where the variance appears to stabilize (Rognlie, 2010). Hence, an average of daily realized volatility is plotted against varying sampling intervals to visually choose an optimal sampling interval where the variance seems to stabilize; as mentioned earlier and depicted in figure 1. For all the eight assets, a sampling interval of ten minutes was found to be an optimal choice.

Real-time news was collected from the *Bloomberg* real-time news service for the year 2009. There were 88821 news stories analysed from 1 January, 2009 to 31 December, 2009. Out of these, 5133 news were relevant to the eight companies under analysis as shown in table 1. The total trades that formed part of the sample are also shown for each company.

Table 1: Sampled companies, Trades and news events identified.

Asset	Trades in 2009	No. of news relevant in 2009
Unitech	45510704	288
DLF	39204274	422
Suzlon	37211025	494
ICICI Bank	32668776	777
Tata Steel	30341414	1001
Reliance	28723448	1833
Reliance Capital	28258500	85
JP Associates	28059111	233
Total		5133

6 RESULTS AND DISCUSSION

The results of the study show that the wavelet based jump identification seems to perform better than the BNS based jump identification methodology. Table 2 presents the results of jumps identified for the eight sampled companies, while table 3 present the same results using the BNS method. The total observations that were in the study were 7896 for each company which were sampled at 10 minute frequency. The total number of jumps were identified with both positive jumps and negative jumps. Positive jumps have an upward trend and negative jumps have a downward trend. The average duration of the jump for a company was found to be 21 minutes. The average positive jump duration was 21.72 minutes while an average negative jump

Table 2: No of jumps identified for each company using Wavelets

Company	No. of observations	No. of Jumps	Average Jump Duration in Minutes	No. of Positive Jumps	Average Positive Jump Duration in Minutes	No. of negative jumps	Average Negative Jump Duration in Minutes	Number of Jumps mapped with news	Percent News mapped with Jumps	Average News stories of Dominant News Cluster
DLF	7986	19	19.21674	9	23.22608	10	15.60833	15	0.789474	4.263158
ICICI Bank	7986	14	26.29167	7	32.00099	7	20.58234	13	0.928571	5.928571
JP Associates	7986	17	21.43137	8	27.23438	9	16.27315	11	0.647059	3.882353
Reliance Capital	7986	20	18.25625	10	19.62708	10	16.88542	7	0.35	1.5
Reliance	7986	16	22.81858	8	27.25781	8	18.37934	15	0.9375	13.125
Suzlon	7986	12	30.35822	6	30.08218	6	30.63426	11	0.916667	5
Tata Steel	7986	21	17.3502	11	19.56187	10	14.91736	19	0.904762	6.190476
Unitech	7986	20	18.21771	10	18.81944	10	17.61597	13	0.65	2.75
Average			21.74		24.72		18.86		0.765504	

Table 3: No of jumps identified for each company using BNS method

Company	No. of log-price observations	No. of Jumps	Optimal Lag (0-60 minutes)	Correlation	p-Value	No. of Positive Jumps	No. of negative jumps	Number of Jumps mapped with news	Percent jumps mapped with news	Average News stories of Dominant News Cluster
DLF	7986	330	60	0.022745	0.042177*	158	166	42	0.127273	0.239394
ICICI Bank	7986	263	0	0.015904	0.155291	139	120	11	0.041825	0.08365
JP Associates	7986	319	20	0.020031	0.073502	148	156	7	0.021944	0.112853
Reliance Capital	7986	270	60	-0.03938	0.000434*	119	146	1	0.003704	0.02963
Reliance	7986	264	0	0.019738	0.077776	123	138	18	0.068182	0.212121
Suzlon	7986	338	60	0.015845	0.156966	152	162	28	0.08284	0.263314
Tata Steel	7986	3758	30	0.028322	0.011386*	1880	1787	319	0.084886	0.223789
Unitech	7986	326	0	0.025868	0.020794*	153	152	4	0.01227	0.033742
Average									0.055365	

lasted for 18.86 minutes. According to behavioural finance literature, the markets react asymmetrically to news. It reacts sharply to the bad news as compared to the good news (Byrne and Brooks, 2008). Jump durations observed by the wavelet decomposition of trade instances corroborates with behavioural finance notion that following news the prices drift for some time with a lag.

One major impediment with the BNS method is the absence of the concept of jump duration. It only identifies the presence of jumps at particular time. In order to analyse the occurrence of jumps with respect to the market news and sentiments, it is imperative to assume an alignment lag or jump duration. To alleviate the problem, we observed a range of alignment lags from 0 minute to 60 minutes to find an optimal lag for each asset where the jumps identified are the most correlated with the market sentiment.

The BNS method identified far more jumps than the wavelet based method. Number of jumps identified by the BNS algorithm is around 9% of the trade instances observed. On the other hand, wavelets identify only 0.2% of the trade instances as significant jumps. The jumps identified in the proposed jump identification method using wavelets have noise removed at the optimal "jump" scale. The average percent of jumps that could be identified with news was only about 5% with BNS; while with the wavelets, 76% of the jumps could be identified with market sentiment. This suggests that the proposed wavelet based method clearly dominates the BNS methodology, as far as mapping the news sentiment is concerned.

The jump polarity is associated with the upward (positive) and downward (negative) direction. This study hypothesises that the jump polarity should essentially match the news polarity which is consistent with Bernard and Thomas (1990) in high frequency drift setup. News polarity is measured by the dictionary based methodology suggested earlier, through which news sentiment can be identified. Needless to say, the news sentiment will have a polarity because any cluster of news that enter the market should determine the positive sentiment or negative sentiment in the market at that moment. Table 4 shows the polarity match and mismatch between news sentiment and the jumps identified using wavelets and BNS methods. The polarity match on an average is far higher for the wavelet based method vis-à-vis the BNS method. The polarity match is 70% for the wavelet based method and 59% for the BNS method. This also suggests that the wavelet based jump identification can easily be associated with contemporaneous news sentiment.

7 SIMULATION TRADES IN THE OPTIONS MARKET

If trades based on trends are profitable one can prove that investors overreact to news and are deliberately swayed by the trend. Since jumps can be identified and their duration can also be assessed, we tried to simulate profitable trades using options market. The options market typically is a pref-

Table 4: Jumps Polarity against News Polarity

Jump Tool	Wavelets			BNS Bollersev		
	Jumps Polarity Match	Jumps Polarity Mismatch	Percent Match	Jumps Polarity Match	Jumps Polarity Mismatch	Percent Match
DLF	11	4	0.733333	23	23	0.5
ICICI Bank	11	2	0.846154	4	7	0.363636
JP Associates	10	1	0.909091	5	2	0.714286
Reliance Capital	4	3	0.571429	1	0	1
Reliance	12	3	0.8	5	11	0.3125
Suzlon	8	3	0.727273	14	11	0.56
Tata Steel	10	8	0.555556	158	129	0.550523
Unitech	6	6	0.5	3	1	0.75
Average			0.705354			0.593868

erred destination to execute speculative trades. The effect of news sentiment therefore would have larger impact on the profitability of trades initiated in the options market. Whenever there is a positive sentiment as indicated by the news sentiment and there is a corresponding jump identified, call options will be purchased and will be sold at the peak of the jump after a few minutes. The put options will be purchased. The results of the trades are reported in the table 5 which is for the wavelet based method. Significant profits can be noticed in both the calls and puts and all the profits are positive. This clearly indicates that the news sentiment can be used for profiting in the options market. Table 6 reports similar profits in the options market but the jump identification method in this case was the BNS method. The results indicate that the profits are largely negative.

8 CONCLUSION

The paper compares two jump identification methods; the BNS method and the wavelets based method. News sentiment is computed using a dictionary based method. High frequency news and high frequency stock market and option market data is used in the study. The study correlates news sentiment around the jumps and finds that the news sentiment polarity matches with the jump polarity and lead the price trends. Wavelet based methodology dominates the BNS method as far as the jump relationship with news sentiment is concerned. Wavelet based decomposition of trade instances not only removes noise while identifying jumps, but also introduces an estimate for the jump duration. Jump duration can be a useful parameter in trading and also brings out the behavioural patterns of the trading participants. Simulation trades in the options market also show significant profits in the wavelet based method as against the BNS method suggesting that investors overreact to news and h'ave an inclination to follow trends.

9 REFERENCES

Aït-Sahalia Y 'Disentangling diffusion from jumps' (2004) 74 *Journal of Financial Economics* 487–528

Aït-Sahalia, Y 'Maximum Likelihood Estimation of Discretely Sampled Diffusions: A Closed-form Approximation Approach' (2002) *70*(1) *Econometrica,* 223-262.

Aït-Sahalia, Y and Jacod, J 'Estimating the degree of activity of jumps in high frequency data' (2009b) *The Annals of Statistics,* 2202-2244.

Aït-Sahalia, Y and Jacod, J 'Testing for jumps in a discretely observed process' (2009a *) The Annals of Statistics*, 184-222.

Andersen T, Bollerslev T and Diebold F, 'Roughing it up, including jump components in the measurement, modelling and forecasting of return volatility' (2007) 89 *Review of Economics and Statistics*, 701-720.

Barker Richard 'The market for information: evidence from financial directors, analysts, and fund managers' (1998) 29 *Accounting and Business Research* 3-21

Barndorff-Nielsen, Ole E, and Neil Shephard 'Econometrics of testing for jumps in financial economics using bipower variation' (2006) 4 *Journal of financial Econometrics* 1-30.

Bernard V , Thomas J 'Evidence that stock prices do not fully reflect the implications of current earnings for future earnings', (1990) 13 *Journal of Accounting and Economics* 305-40

Bollerslev, Tim, Viktor Todorov, and Sophia Zhengzi Li 'Jump tails, extreme dependencies, and the distribution of stock returns' (2013) 172 (2) *Journal of Econometrics* 307-324.

Boudt, K., Croux, C. and Laurent, S 'Robust estimation of intraweek periodicity in volatility and jump detection' (2011) *18*(2) *Journal of Empirical Finance* 353-367.

Cont Rama and Tankov Peter, *Financial Modelling with Jump Processes*, (Chapman and Hall, CRC Press, New York 2009) Chapter 8

Cont, R. and Mancini, C 'Nonparametric tests for analyzing the fine structure of price fluctuations' (2007) *manuscript, Duke University*

Cox JC, Ross SA 'The Valuation of Options for Alternate Stochastic Processes' (1976) 3 *Journal of Financial Economics*.145-166.

Daniel K, Hirshleifer and A Subrahmanyam, 'Investor psychology and security market under and overreaction' (1998) 52 *Journal of Finance* 1 - 33

Das, S., & Chen, M. Y 'Yahoo! for Amazon: Sentiment Extraction from Small Talk on the Web' (2001) *EFA 2001 Barcelona Meetings*.

Devitt, A., & Ahmad, K. 'Sentiment Polarity Identification in Financial News: A Cohesion-based Approach' (2007) *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics. Prague, Czech Republic: Association for Computational Linguistics*, pp 984–991

Eraker B , Johnnes M S , Polson N 'The Impact of Jumps on Returns and Volatility' (2003) 58 *Journal of Finance*, 1269-1300

Fan, J. and Wang, Y 'Multi-scale jump and volatility analysis for high-frequency financial data' (2007) *102*(480) *Journal of the American Statistical Association,* , 1349-1362.

Huang X 'Macroeconomic news announcements, systematic risk, financial market volatility and jumps' (2015),Working paper, Federal Reserve Board, Washington DC.

Huang, X 'Macroeconomic news announcements, financial market volatility and jumps' (2007) Working Paper, Duke University,

Huang, Xin, and George Tauchen 'The relative contribution of jumps to total price variance' (2005) 3 (4) *Journal of financial econometrics* 456-499.

Jiang, G.J. and Oomen, R.C 'Testing for jumps when asset prices are observed with noise–a "swap variance" approach' (2008) 144(2) *Journal of Econometrics* .352-370.

Jorion P 'On Jump Processes in the Foreign Exchange and Stock Markets' (1989) 4, *Review of Financial Studies* 427- 445

Lakonishok J, A Shleifer 'Contrarian investment extrapolation and risk' (1994) 49(5) *Journal of Finance* 1541-78

Lee, S.S. and Mykland, P.A 'Jumps in financial markets: A new nonparametric test and jump dynamics' (2008) 21(6) *Review of Financial studies* 2535-2563.

Loughran T, and Mc Donald. B, 'When is a Liability not a Liability?, Textual Analysis, Dictionaries, and 10-Ks' (2009) 66(1) *The Journal of Finance* 35-65.

Maheu J , McCurdy T 'News Arrival, Jump Dynamics, and Volatility Components for Individual Stock Returns' (2004) 59 *Journal of Finance* 755-793.

Merton Robert C 'The Impact on Option Pricing of Specification Error in the Underlying Stock Price Returns' (1976) 31 *Journal of Finance* 333-350

Ritu Y, Ashwani K, Vinay K A 'Stock Market Sentiment Analysis: A Review of Dictionary Based Unsupervised Methods', (2013) IIM Lucknow Working Paper Series, 2012-13/ 40

Rognlie, Matthew 'Spurious Jump Detection and Intraday Changes in Volatility' (2010) Dissertation. Duke University Durham,

Schumaker, R. P., Zhang, Y., Huang, C.-N., & Chen, H. 'Evaluating sentiment in financial news articles' (2012) 53 *Decision Support Systems* 458–464.

Stone, P. J., Bales, R. F., Namenwirth, J. Z. and Ogilvie, D. M, 'The general inquirer: A computer system for content analysis and retrieval based on the sentence as a unit of information' (1962) 7 (4) *Behavioral Science* 484-498,.

Sun E W, Meinl T 'A New Wavelet based Denoising Algorithm for High Frequency Financial Data Mining' (2012) 217 *European Journal of Operation Research* 589-599

Tetlock, P. C 'Giving Content to Investor Sentiment: The Role of Media in the Stock Market' (2007) 62 *The Journal of Finance* 1139-1168.

Todorov, V. and Tauchen, G 'Activity signature plots and the generalized blumenthal-getoor index'. (2008) *Manuscript, Duke University and Northwestern University.*

Wei, S 'Simultaneous occurrence of price jumps and changes in diffusive price volatility*'* (2012) Doctoral dissertation, Duke University Durham.

Wysocki, P. 'Cheap talk on the web: The determinants of postings on stock message boards', (1998) University of Michigan Business School Working Paper 98025.

Yan Shu 'Jump Risk Stock Returns and Slope of Implied Volatility Smile' (2011) 99 *Journal of Financial Economics* .216-233