A Review of High-Frequency Sentiment Analysis

Khurshid Ahmad,
Trinity College, Dublin, Ireland
August 2011

Introduction
The work on the impact of news, especially affect and sentiment laden phrases within news items, on financial markets is critical for the well-being of economies large and small. The irrational behaviour (Kahneman and Tversky 1979, Kahneman 2002), contrarian behaviour (Shiller 2000), herd behaviour (Haiss 2010), and often irresponsible behaviour (Mackenzie 2009) of the various stakeholders in the financial markets, has demonstrated the limits of efficient market hypothesis (EMH) (Mandelbrot and Hudson 2004). It was the central tenet of EMH that news can be discounted by rational stakeholders. The 2008 global financial crisis was predicted by practitioner/academics (e.g.Taleb 2007), by serious financial polemicians (Roubini 2004) as well by key economic thinkers: Vernon Smith has noted that ‘historically, a recurrent theme in economics is that the values to which people respond are not confined to those one would expect based on the narrowly defined canons of rationality.’ (Smith 2008). It appears that prices and volumes cannot be determined by discounting news and rumours. The inclusion of (reports about and of) the behavior of the stakeholders in the market in price/volume models appears intuitively to be a ‘good thing’.

The advent of sophisticated computer systems, facilitating high frequency (HF) trading (Goodhart and O’Hara 1997) on the one hand, and providing access to and analysis of news, blogs and edicts from a variety of sources (Tetlock 2007, Ahmad 2011) on the other, suggests that financial traders will have access to enough information to make informed decisions very quickly – in tens of seconds (Iyengar and Ma 2010, have presented a theoretical model for incorporating news impact in HF trading). This advantage of HF trading is not limited to the trading floors only, indeed this trading will help to investigate ‘the effects of market structure on the availability and interpretation of the data, methodological issues such as the treatment of time, the effects of intra-day seasonals, and the effects of time-varying volatility, and the information content of various market data’ (Goodhart and O’Hara 1997). However, HF trading may also lead to instances of stocks dropping from $40 to one cent in a matter of seconds and in the same trading another stock rising from $6 to $99,999 (Mackenzie 2011 commenting on a curious price spike/fall on 6 May 2010; the author also makes the point about algo(rithmic) sniffing where software systems look at the behavior of a competitors’ high-frequency trading algorithms in the market.

The key point for me at least is that HF trading patterns can be discerned at different scales from tick-by-tick trades to daily, weekly, monthly or yearly scales. HF trading and other innovations in financial/economic systems ‘contain variables that operate on a variety of time-scales simultaneously so that the relationships between variables may well differ across time-scales’ (Ramsey 1999:2594): Using the wavelet approach Ramsey and Zhang (1997) have successfully analysed foreign exchange tick trade data.

The impact of news on financial markets can be substantial and, according to the Nobel Laureate Robert Engle (2002), asymmetric: the arrival of ‘bad news’ has a longer lasting effect on prices, and particularly on volumes of shares traded, when compared to the arrival of ‘good news’. Economists typically use proxies for the news content – change in the values
of currencies, bonds, or aggregated indices like the Dow Jones, NASDAQ.

News was traditionally interpreted as a causal variable in financial market models: readily quantifiable aspects of news as a proxy for the news itself. The proxy includes the timing of news arrival, the volume of news items and the type of news (Antweiler and Frank 2004). News proxies were used with some effect to show that ‘negative’ news has much longer lasting impact than the positive news (Engle 2002, Engle & Ng 1991). This kind of sentiment analysis is almost always conducted post-hoc (Chang & Taylor 2003, Bauwens, Omrane and Giot 2005). Sentiment indices have been constructed using news proxies – in the context of equities it has been observed that ‘wave of investor sentiment has larger effects on securities whose valuations are highly subjective and difficult to arbitrage’ (Baker and Wurgler 2006). More recently, news proxies have been used in the analysis of the high frequency trading behaviour of German Bund futures: ‘news announcements have a highly positive impact on both efficient and noise volatility’ (Hautsch, Hess, and Veredas 2011).

News content analysis, rather than the use of proxies, is becoming more important recently. Continual records of market analysts’ opinions on commercial news channels shows that positive news has short term (1 minute) impact on prices but the negative news impacts for 15 minutes (Busse and Clifton 2002). It has been argued that the use of optimistic language in press releases of a firm appears to increase the firm’s future earnings, whilst pessimistic news has the opposite effect (Davies, Piger & Sedor 2006). Tetlock has shown that the negative affective component of news reports does have a longer lasting impact on the volatility of equities by analyzing opinion columns in financial news papers (Tetlock 2007). Groß-Klußmann and Hautsch (2011) have reported the impact of sentiment extracted from news streams on the patterns observed in high frequency trading using a VAR model for equity trading on the London Stock Exchange. The authors obtained ‘distinct responses in returns, volatility, trading volumes and bid-ask spreads due to news arrivals’ (2011:321). The results obtained by Groß-Klußmann and Hautsch depend critically on linguistic processing especially the pragmatic questions about the relevance of the news and the topics discussed therein.

The key to successful incorporation of sentiment in econometric models of financial time series, that are sometimes non-stationary, show GARCH effects, and have a scalar behaviour, requires an understanding of news and blogs at different levels of linguistic (Wilson, Wiebe& Hoffmann 2005) and ontological description (Valitutti, Strapparava & Stock 2004) . News agencies supply a vast quantity of undifferentiated general news and it is not clear what is the relationship between such news and a specific financial instrument or commodity. In order to use news specific to an instrument/commodity, it is important that a language processing system has access to the right terminology which is organised systematically under the rubric of an applications ontology. The next level of linguistic description is that of grammar and morphology which are essential for disambiguation – natural language is inherently vague and ambiguous and grammatical and morphological analysis can help in identifying and eliminating ambiguity (Rentoumi et al 2009). The higher level of linguistic description include semantic and pragmatics: sentiment bearing phrases comprise affect including evaluation (positive/negative), attitude (active/passive) and orientation, and are sometimes expressed metaphorically (Glucksberg 2001, Goatley 1997). Different types of texts – reportage, editorial, comments, blogs – require different kinds of analysis as these texts are structurally different and use different linguistic devices for communication.
One major problem in sentiment analysis is that of aggregation: sentiment in a text or blog is articulated within a phrase – one can assign a polarity evaluation to the phrase in a nominal sense. The question is how to aggregate the sentiment score of a set of phrases that comprise a text or blog? One can aggregate nominal values but what kind of a arithmetic will be used. Typically, this aggregation is carried out using multivariate analysis including factor analysis – but this analysis has its own limitations.

Similarly, the topic of a news story has to be determined and there are algorithms to do automatically perform such an assignment – but these algorithms have their own limitations and had to be used with some care as demonstrated by Groß-Klußmann and Hautsch (2011).

Goodheart and O’Hara (1997) do talk about the notion of time in trading: the cointegration of a news time series with prices/volumes time series is not easy to perform in that the dating of the news and that of prices are carried out using different conceptualisation of time. News affect can lag the prices (Tetclock 2007), but equally one can take the contrary view. In any case, the information in the news is absorbed in the prices once all stakeholders have access to the news – the question is this: how will this absorption be incorporated in a combined news/price models?

References:


