**Checkley, Matthew, Anóñ Higón, D and Alles, C (2017) The Hasty Wisdom of the Mob: How Market Sentiment Predicts Stock Market Behavior. Expert Systems with Applications, 77. pp. 256-263. ISSN 09574174**

Official URL: http://dx.doi.org/10.1016/j.eswa.2017.01.029
DOI: http://dx.doi.org/10.1016/j.eswa.2017.01.029
EPrint URI: http://eprints.glos.ac.uk/id/eprint/4294

The Hasty Wisdom of the Mob: How Market Sentiment Predicts Stock Market Behavior

M. S. Checkleya, D. Añón Higónb, H. Allesc

ABSTRACT

We explore the ability of sentiment metrics, extracted from micro-blogging sites, to predict stock markets. We also address sentiments' predictive time-horizons. The data concern bloggers' feelings about five major stocks. Taking independent bullish and bearish sentiment metrics, granular to two minute intervals, we model their ability to forecast stock price direction, volatility, and traded volume. We find evidence of a causal link from sentiments to stock price returns, volatility and volume. The predictive time-horizon is minutes, rather than hours or days. We argue that diverse and high volume sentiment is more predictive of price volatility and traded volume than near-consensus is predictive of price direction. Causality is ephemeral. In this sense, the crowd is more a hasty mob than a source of wisdom.

Keywords: sentiment, stock market, social media, forecasting, micro-blogging, analytics.

1. Introduction

Sentiment metrics derived from social media are claimed to help predict financial market behavior (c.f. Bollen et al., 2011, Mittermayer & Knolmayer, 2006). By matching the wording employed, in a micro-blog, with a dictionary of mood-related words, millions of financial market-related messages can be sampled for the emotions shown by their authors. Mood can be extracted real-time from bloggers and, in a fashion, measured. These metrics are aggregated over assets and time to provide a dynamic sense of collective market mood. The resulting time series of sentiment metrics can be tested for its ability to predict market prices and returns, price volatility, and trading volume. In other words, sentiment is argued to predict market behavior.

While the research literature on this topic burgeons, there is divergence in scholars' views about the underlying predictive ability of social media sentiment metrics. Research findings range from sentiment analysis being consistently prophetic, to having modest or selective value, to having no forecasting value whatsoever, particularly for price direction (c.f. Nassirtoussi et al., 2014, for a review of the research literature).

To help resolve the growing debate over the supposed value of financial market forecasting with sentiment metrics, we explore intraday sentiment and price data for each of five high-profile US stocks. Focusing on more granular data than prior authors, we employ intraday market data; sampled-every-two-minutes data on price direction, price volatility and trading volume. We match the market data to equally-granular sentiment metrics. These data facilitate the study of shorter time-horizons – relative to prior research - in the predictive ability of sentiments for prices. We find evidence of sentiments causing market behavior, albeit with selective ability to reduce forecast errors. Uniquely, we find strongest results over time horizons of minutes, rather than hours or days. In sum, sentiment has some short-term predictive value, but more-so when emotions are divergent across the market. Forecasts are better in the cases of predicting volatility and trading volume, than price direction. Bloggers collectively evince uncertainty more than clarity.

This study contributes to literature by adding to arguments in favour of emotion being salient to investment decision making, by finding unique evidence on the rate at which most valuable information diffuses between market participants, and by specifying related classes of trading strategy most apt to building on such a predictive foundation. The findings are further embedded in the work of Surowiecki, (2004); we argue that the strongest causal links between sentiment and market behaviour are found in times of strident and discordant sentiment. Traders collectively form a mob more than a wise crowd, but elements of shared sentiment can be weakly-linked to predictable price direction.

This paper is organized as follows: the next section contains a literature review, followed by discussion of the study's data and methods. Then we address the results of the statistical tests. Finally, conclusions are drawn, along with discussion of the paper's limitations and opportunities for additional research.

## 2. Literature Review

This section selectively addresses prior literature on two interrelated themes. First, how theory frameworks – Efficient Markets, (Fama, 1965), Behavioural Finance (c.f. Lowenstein and Lerner, 2003), Information Diffusion (c.f. Hong & Stein, 1999), and The Wisdom of the Crowds (Surowiecki, 2004) – attempt to account for the predictability of markets. Second, we discuss major findings in the literature specific to social media sentiment-based prediction of financial markets.

### 2.1 Theory of Market Prediction with Sentiment Metrics

To what extent stock market prices are predictable is a long-standing and high profile debate in the finance and economics literature. The Efficient Market Hypothesis (Fama, 1965) states that prices fluctuate randomly and hence the very act of forecasting is based on misapprehension. Put crudely, long-term supernormal trading profits (adjusted for risk) are implausible in efficient financial markets.

However, responding to critiques of the Efficient Market Hypothesis (c.f. Cutler et al., 1989), recent studies in Behavioural Finance suggest that emotion has a substantial role in investment decision making (Lowenstein and Lerner, 2003). With the speedy expansion of the Web and social media, the influence of investors' and web users' emotions has become increasingly noteworthy. Traditional news media have evolved into diverse forms of blogging and micro-blogging (e.g. Twitter and StockTwits), chat rooms and discussion boards. Information can be authored with ease, searched, an opinion formed and thereafter rapidly and widely shared (Oh et al. 2013). Emotion about financial markets is somewhat contagious and can be diffused and pooled (Oliveira et al., 2013). Hence, most of the studies reviewed in this section share a supposition; en masse, social media displays measurable emotion and the derived sentiment metrics can plausibly predict buying and selling behaviour. This, in turn, affects market prices.

Consistent with the Gradual-Information-Diffusion model (c.f. Hong, Lim & Stein, 2000, Hong & Stein, 1999, Matsubara et al. 2012), investors track market-salient news and opinion flow, and gradually and incompletely incorporate such information into their trading decisions. The rate of information diffusion to and between market participants influences how market prices shift over time. Traded assets avidly-covered by mass and social media would be expected to respond to news flow faster than low-profile assets, for which mass media reportage would be sparse or non-existent (Matsubara et al. 2012).

A central issue, related to the diffusion of information, is whether such diverse and decentralized opinions on social media, when aggregated, reflect "the wisdom of crowds" (Surowiecki, 2004). Do the opinions of bloggers collectively form valuable predictions about markets? Perhaps financial markets are susceptible to "group think", wherein a trader's opinions are not wholly decentralized and independent from those of other traders. If so, this would violate the conditions necessary for aggregated opinion to outperform experts (Surowiecki, 2004). It has been argued, for example, that rather than being diverse and decentralized, opinions are diffused along lines strongly mediated by the social networks of influential bloggers (Armentano et al. 2013).

The study of information diffusion within markets leads to the question of the rate at which information spreads. This, in turn, might affect the time horizon for a forecast based on such information. Prior literature, whether or not it pinpoints a price-causal role for social media sentiments, has tended to focus on forecast horizons of a day or more, with a few engaging with data granular to the hour (c.f. Nassirtoussi et al. 2014). Yet we know that market participants typically respond to salient market data within one hour (Chordia et al. 2005), and plausibly within seconds or thousandths of a second (c.f. Lewis, 2014).

Meanwhile, social media tends to elicit responses to high-interest blogs within a timeframe of minutes, albeit with the potential for an ongoing stream of responses and re-tweets over hours or days (PsychSignal, 2014). Putting these insights together, and knowing that social media is already in trade-prompting use by a number of market participants (PsychSignal, 2014), it appears worthwhile to investigate causal relationships of shorter duration than those prevalent in the extant research literature. One might suppose that, for high-profile assets, the duration of any price-causal relationship - between social media sentiment metrics and market behaviour - would more typically last minutes, rather than hours or days.

2.2 Predicting Markets with Social Media Sentiment Metrics

Having reviewed some central theoretical frameworks in the prior sub-section, we now consider research specific to predicting market behaviour (particularly asset price movements) with social media sentiment metrics.

The number of Twitter users has grown to several hundreds of millions, resulting in the composition of many hundreds or thousands of tweets in a typical trading day for any major stock (PsychSignal, 2014). Many of these messages respond to unfolding market events. They can be considered a real-time and near-continual evocation of mood throughout the trading day (Oliveira et al., 2013).

The idea of extracting sentiment from microblogs, in order to forecast stock markets, inspires a burgeoning research stream. Prior research argues that sentiment metrics are materially predictive of stock price returns (c.f. Mittermayer, 2004), volatility (c.f. Antweiler and Frank, 2004), and traded volume (c.f. Oliveira et al., 2013). In contrast, several researchers find that social media sentiment metrics proffer little predictive advantage, particularly for stock returns (for a literature review, see Nassirtoussi et al., 2014).

Bollen et al., (2011), sampled sentiments in six dimensions (Calm, Alert, Sure, Vital, Kind, and Happy) and composed, employing artificial neural networks, a model of improved forecasts of the Dow Jones Industrial Average. In contrast, much recent research (including this paper) takes emotion to have two dimensions: valence (positive or negative); and arousal or “level” (low or high) (Nassirtoussi et al. 2014).

With a novel method, Makrehchi, Shah & Liao (2013), test the predictive ability of sentiments by retrospectively assessing what sentiments plausibly could have been able to predict just before large market movements. Using daily analysis, they show that sentiments have sufficient predictive value to enable supernormal trading profits. Similarly, Ruiz et al. (2012), explore links between the volume of microblog messages and social network properties of those blogging, and market price movement and traded volume. They find stronger correlations with traded volume than price movement, but nonetheless suggest their results offer promise as the basis of a trading signal.

While this stream of research provides novel findings, there remain reasons for doubt.

Many studies use only short periods of sampled data (c.f. Yu et al., 2013). Moreover, most test daily data, with very few granular to the half or quarter hour (c.f. Antweiler and Frank, 2004, for data tested at 15 minute intervals). Along with contradictory findings on sentiments’ forecasting value, most prior studies have employed daily sentiments and daily market closing prices. Yet, assuming that intraday sentiment metrics are predictive of market behavior, there is reason - as argued above - to believe any valuable signals would be used speedily by market participants (c.f. Chordia et al., 2005).

In this paper, we test five major stocks, with untypically granular data, and over a two year period. In this way we lose some of the generalizability of testing, say, hundreds of stocks over months of daily data (an approach prevailing in extant literature). But, with granular price and sentiment data, we gain deeper insight to the short-term dynamics of sentiments and prices (played out over two years of analysis) in the face of widely-exposed mass and social media commentary. In essence, we focus on short-term depth and rigor, at the cost of breadth in terms of assets-tested. Any relationships found can be tested and perhaps generalized, to additional assets, in subsequent studies.

Having addressed some of the main findings and limitations of the research literature linking social media sentiment to financial market prediction, we progress, in the following section, to outline the data and methods employed in modelling sentiments' putatively-causal relationships with market behavior. We assess if sentiments offer better market predictions than naïve forecasting models. There is emphasis on the hypothesis that granular data offers insight to shorter forecast horizons.

## 3. Data and Methods

We aim to test if social media sentiment metrics are able to predict financial market behavior, and the forecast horizon of any such predictive relationship. In addressing the limitations of prior research, we consider more granular data over longer sampled periods, and apply stringent tests of the forecasting power of sentiment-based models. In contrast with most prior studies employing daily data, and granular – at best – to every quarter hour, we use sentiments measured every two minutes and match those to contemporaneous and equally granular price data for five high-profile US stocks. We sample the intraday data over a two year period, from both Twitter and StockTwits, which compares favourably with prior studies' use of a few months of data from less voluminous sources of sentiment. In common with Oliveira et al., (2013), we test sentiments' abilities to diminish forecasting error for three prime market variables: price direction (also referred to as "returns"); volatility; and trading volume. For each forecast so produced we calculate forecasting errors with both Mean Absolute Percentage Error ("MAPE") and Root Mean Squared Error ("RMSE") metrics.

### 3.1 The Extraction and Nature of Sentiment Metrics

In order to extract sentiment, the content of each financial market-relevant tweet is matched to a dictionary of words salient to the author's emotional state. For example, the tweet, "Excited about IBM share: will buy this month", would be captured automatically as a positive statement about the IBM stock price. The emotive words used in nearly all stock relevant tweets can be so denoted as either positive (bullish) or negative (bearish). More strongly emotional content, used in more tweets by more writers, would result in higher metrics for either bullishness or bearishness. The process of sampling and analysing sentiments is described extensively in, for example, Li and Li (2013), or Kontopoulos et al. (2013).

The sentiment metrics used in this study are generated by the commercial firm PsychSignal (www.psychsignal.com). These data are distinctive from prior studies in three ways. First, the metrics for either bullish or bearish emotions are scaled from zero (no emotion) to four (strongest emotion). Second, the bull and bear metrics are measured independently. Historically, other sentiment metrics might have been taken as a continuous variable summing to 100% of extracted emotion at any given time. In contrast, the PsychSignal metrics for bullish and bearish sentiment form two independent variables, each taking any value from zero to four. Third, PsychSignal extracts sentiment from both Twitter and StockTwits.

Micro-blogging activity on either of the two sampled sources did not commence until 2009. It can be argued that a statistically notable volume of relevant user content was not generated until 2010 or 2011.

The software employed by PsychSignal is based on the Linguistic Inquiry and Word Count ("LIWC") framework, which is available to the public along with a description of the methodology (LIWC, 2016). The LIWC software reads the salient text and calculates the percentage of words in the text reflecting different emotions, thinking styles, social concerns, and parts of speech. The process rests on the software comparing the text analysed with a dictionary of mood-related words that have been assessed by panels of experts for relevance and the “strength” of emotion displayed by word-use.

We study 5 well-known stocks: Amazon, Apple, Goldman Sachs, Google, and IBM.

These were selected because they are well-followed, well-tweeted and large capitalization stocks i.e. they can claim to both influence and be influenced by popular sentiment.

Moreover, they are, uniquely for such equities, each linked to their own volatility markets.

This invites future research should there be a proven link from the relevant sentiments to

stock price volatility.

3.2 Data Corrections

We consider the alignment of tweeting activity with trading hours. Because we use financial data from the NASDAQ and NYSE, we align messages with US trading hours (9:30 am to 4:00 pm) by assigning messages posted after 4:00 pm to the next trading day. This is consistent with Antweiler and Frank (2004). Thus, sentiments posted after the markets close are bundled to assist in the prediction of the market opening price for the following day. This is because these mood metrics could only affect the market behavior of a subsequent trading period.

The financial market price data was sampled from the commercial trading platform TradeStation (www.tradestation.com), which is comparable to other such sources, with the typical retrospective corrections for mergers, stock splits, etc., normalizing the share price.

Prices and share trading volume were sampled every two minutes and time-matched to the sentiment data for the duration of each trading day.

3.3 Derived Metrics and Statistical Methods

In this section, we summarise the data preparation and statistical methods. Two derived measures of sentiment are used, following Antweiler and Frank (2004). The first defines an index or metric of bullishness (Bt) for each time window as:

$$B_t = \ln\left(\frac{1+bull_t}{1+bear_t}\right)$$

Here bull represents an index of positive sentiment tweeted within a particular 2-minute period t for a specific stock, while bear represents an index of negative sentiment in the same period for the same stock.

The second measure, also consistent with Antweiler and Frank (2004), is the index of agreement (At) between positive and negative sentiments. It is given by:

$$A_t = 1-\sqrt{1-\left(\frac{bull_t - bear_t}{bull_t + bear_t}\right)^2}$$

If all tweeted messages about a particular company are all either bullish or bearish (but not both), agreement would, in that case, be 1 at time t. If sentiment is equally bullish and bearish, then agreement would be 0. In the absence of tweeted messages for a particular time period, we define the bullishness (B) and agreement (A) index for these silent periods as zero, in line with previous studies.

### 3.3.1 Adjustments to Financial Market Data

We are interested in three aspects of stock behavior; returns, volatility and traded volume. In common with established practice, we compute returns as the difference of the natural logarithm of the closing value of the stock price of a particular time period and its lagged value:

$$return_t = \ln(close_t) - \ln(close_{t-1})$$

We provide two measures of volatility. The first measures realized volatility and follows from Andersen et al. (2007):

$$RV_t = \sqrt{\sum_{i=1}^{n} r_{it}^2}$$

where RVt is the realized volatility at time t for an interval of 10 minutes and is obtained as the squared root of the sum of the squared return, r, during a time window interval. We also compute the volatility of stock returns, at t-minutes frequency, as a moving average filter of the stock variance at a time window of 10 minutes. Traded volume is a simple metric of the number of shares traded in a period. We use the natural logarithm of the traded volume in our analysis.

### 3.3.2 Granger Causality Analysis

To determine what relationships might exist between stock outcomes (returns, volatility and traded volume) and tweet sentiment features (bullishness and agreement), we use Granger Causality Analysis (Granger, 1969). A variable X is said to "Granger-cause Y" if Y can be better predicted using the histories of both X and Y than by using the history of Y alone. Hence, if when controlling for the information contained in past values of Y, past values of X add significantly to the explanation of current Y, then X is said to "Granger-cause" Y (c.f. Datta, et al. 2006 for an example of Granger analysis in the case of information systems).

Formally, the possible Granger causal links between stock outcomes (variable "S", defined in three different ways, as, returns, volatility and volume), and sentiments (variable "T") can be expressed using the parameters of Equation (1):

$$S_t = \sum_{j=1}^{m} \beta_j S_{t-j} + \sum_{j=1}^{m} \delta_j T_{t-j} + \eta_t + \varepsilon_t$$

Therefore, there is Granger causality from T to S if the lagged values of T have a statistically significant correlation with S $(\delta_j \neq 0, \forall j)$. In all specifications we control for month, day and hour fixed effects.

We test six hypothesized relationships:

(1) Bullishness Granger causes stock returns, volatility, and traded volume.

(2) Agreement Granger causes stock returns, volatility, and traded volume.

The "direct Granger method" is used to test for Granger causality between sentiments and stock behavior. Such models offer indications of both the size and timing of causal effects. An advantage of this single-stage method is the estimates of the Autoregression Distributed Lag ("ADL") model can remain unbiased in the presence of autocorrelated time series data. In so far as the number of lags used in the model is enough to account for time series autocorrelation, no pre-whitening is required (Freeman, 1983). However, insufficient lags can yield autocorrelated errors (and therefore misleading test statistics); while too many lags reduce the power of the test.

The final test of our model is to assess if its forecasting errors are materially smaller than those generated by the naïve model of forecasting price as identical to the prior period. We assess both MAPE and RMSE metrics of forecasting error and apply them in a fashion consistent with Oliveira et al. (2013).

## 4 Research Findings

This section presents the results of the prescribed statistical tests. It begins by describing the data tested.

### 4.1 Descriptive Statistics

Descriptive statistics of the main variables of interests, per company stock, are presented in Table 1. We additionally compute the Spearman's rank pair-wise correlation coefficient between the sentiment data and the financial variables. The results are summarized in Table 2. We find significant correlation in most cases, particularly for Goldman Sachs, IBM and Google. For all the stocks, the Bullishness index is positively correlated with realized volatility; while Agreement is correlated with traded volume, variance and realized volatility. Except for Amazon, the Bullishness index is also positively correlated with returns.

Table 1.

Table 2.

### 4.2 Granger Causality Test Results

The outcomes of the tests of Granger causation are provided in this section. A sample of the full results of the statistical tests can be provided on request. Table 3 shows the p-values resulting from the Granger causation tests. The p-value is the probability of the null hypothesis being correct. Hence, a small p-value means we cannot reject the presence of a Granger-causal tie from sentiments to market behavior. Results are presented for 10 and 20 time lags. The AIC, SBIC and HQIC minimal criterion tests suggest that, except for returns, the lag 20 is more appropriate, while for returns it is lag 10.

Table 3.

Of the 80 p-values considered in Table 3, 44 are significant at the 10%, 5% or 1% levels. Of these significant p-values, realized volatility and volume are best represented amongst the market behavior variables. Bullishness is more material to returns, while Agreement has more explanatory power for volume and volatility.

Every one of the five tested stocks supports realized volatility and volume being Granger caused by either Agreement or Bullishness. Every stock but IBM shows evidence of Bullishness Granger-causing returns. Goldman Sachs shows most Granger-causal relationships.

In summary, the test results are consistent with the Agreement metric predicting realized volatility and trading volume. The Bullishness metric similarly predicts, in most cases, returns. The shift from 10 lags to 20 lags makes modest difference to most salient results. All three market behaviors – returns, volatility, and volume – show selective evidence of being Granger-caused by sentiment metrics. In this moderated sense, all six hypothesized relationships are supported by the test results.

4.3 Forecast Error Analysis

In this subsection, we focus on the forecasting power of sentiment indices. We conduct a one step ahead prediction over a three-month period based on a naïve model (the naïve model is concerned with forecasting price as simply-generated from a price from the period just prior to the forecast period), denoted M0, and an augmented model, M1 (the augmented model uses both prior prices and sentiment metrics). The models are represented as follows:

$$M_0: \quad Y_t = \alpha + \sum_{i=1}^{j} \beta_i Y_{t-i} + \varepsilon_t$$
$$M_1: \quad Y_t = \alpha + \sum_{i=1}^{j} \beta_i Y_{t-i} + \sum_{i=1}^{j} \gamma_i X_{t-i} + \varepsilon_t$$

where Y represents the particular financial indicator (return, trading volume, volatility) and X is the sentiment indicator. We estimate M1 for both the Bullishness index and the Agreement index, respectively.

To analyse the forecasting accuracy we compute the Mean Absolute Percentage Error (MAPE) and also the Root Mean Squared Error (RMSE) as follows:

$$MAPE = \frac{100}{h} \sum_{t=T}^{T+h} \left| y_{t+h} - \hat{y}_{t+h} \right| / y_{t+h}$$

$$RMSE = \frac{1}{h} \sqrt{\sum_{t=T}^{T+h} \left( \hat{y}_{t+h} - y_{t+h} \right)^2}$$

where the “capped” y is the predicted value, y is the actual value, and h is the number of time periods over which forecasting is performed. Our dataset goes from the 17th February 2012 to 17th October 2014; and we use the last three months, i.e. 17 July 2014 to 17 October 2014, as the forecasting period.

In the regressions, the lag i is chosen to be 20 according to results in Table 3, equivalent to 40 minutes, and in all models we include controls for hour, day and month fixed effects. The only exception is for returns, for which a lag 10 is chosen instead, due to being a better fit to the model.

Table 4 shows the forecasting errors expressed as MAPE and RMSE measures. Adding the Agreement metric reduces both the MAPE and the RMSE for the trading volume of Amazon, Apple, Google and IBM. It also reduces the MAPE for the returns of Apple and Goldman Sachs and RMSE for Goldman Sachs and IBM. Additionally, it reduces the MAPE for the realized variance of Apple and Google; and the RMSE for Amazon, Apple, Goldman Sachs and IBM. Nevertheless, these reductions are not large and, in some cases, the forecasting error increases with the addition of sentiment metrics. The forecasting power of the Bullishness metric is less impressive, overall, than the Agreement metric.

Note that in Table 4, estimation is from 17/10/2012 to 16/07/2014, and forecasting is from 17/07/2014 to 17/10/2014. Underlined data indicates that the sentiment-based model produces smaller errors than the base model (“Model M0”).

Table 4.

In summary, the data show modest reductions in forecast errors upon the introduction of sentiment metrics to the forecasting model for market behaviors. Reductions are evident in 29 out of 80 tests. This means that in 51 cases out of 80, forecast errors increase with the addition of sentiment metrics. For Apple and Goldman Sachs, a sentiment-based model reduces forecast errors for most estimates. The Agreement metric performs better than Bullishness, offering more cases of reduced forecast error for stock returns, volatility, and volume.

## 5 Discussion

Social media sentiment metrics, extracted from salient micro-blogging activity, are selectively Granger-causal of stock price returns, volatility and stock trading volume. However, we find limited support for material improvements in predicting those market behaviors. At best, sentiment metrics, used very selectively, can claim to help improve financial market forecasting. Importantly, the time frame for such forecasts is narrow. The typical prediction window is less than 30 minutes, albeit with some small variation between the five stocks analysed.

In common with researchers arguing that sentiment metrics are materially predictive of stock price returns (c.f. Mittermayer, 2004; Bollen et al., 2011), we too find such relationships; albeit in limited fashion. Our tests show that sentiment metrics can have material bearing on the forecasting of share price direction (resonating with, for example, Makrehchi, et al. 2013). However, the predictability of volatility (c.f. Antweiler and Frank, 2004), and traded volume (c.f. Oliveira et al., 2013, Ruiz et al. 2012, Sprenger and Welpe, 2010), are both better-supported by our tests. We partially-contradict those researchers finding that social media sentiment metrics proffer no predictive advantage for any of the three market behaviors we study (c.f. Nassirtoussi et al., 2014 for a range of such views).

We find, in our analysis of the time window of relevant effects, broad support for the view that markets respond to new information sources within one hour (Chordia et al., 2005). Most of the identified Granger-causal ties from sentiments to market outcomes are operating over a time window of less than 30 minutes. And hence, as proposed, high-profile stocks can be associated with quite-rapidly diffused sentiment. Our findings are consistent with such sentiment being spread via social media. It is also suggestive of prior research and related trading models (largely based on daily data) having somewhat missed the richer pickings offered by more granular analyses.

The model developed in the paper has implications for trading practice. Specifically, it means, if one trades on such a signal, that the holding time of a position – the elapsed time between buying and selling a financial asset (long or short) - would be typically in the region of 2 to 30 minutes. This time frame would be too long for high frequency trading (Lewis, 2014), and too short for most forms of options trading (c.f. Lehar et al. 2002). It is apt for short-term trading of assets such as equities, indices or forex. Finally, the fact that the model is more-suited to predicting volatility or trading volume has implications for risk and trade management (c.f. Dimson, 1979). In particular, our model helps predict liquidity and the stability of prices, and hence is an aid to trade management and execution, particularly for very large, high-value trades.

In support of critiques of the Efficient Market Hypothesis (c.f. Cutler, et al., 1989), and recent studies in behavioral finance, we too argue that emotion has a material – albeit limited - role in investment decision making (Lowenstein and Lerner, 2003). Furthermore, we acknowledge that social media creates the prospect of quite-swiftly diffused and shared emotion about events in financial markets (Oliveira et al., 2013); our evidence for time-constrained causal relationships from sentiment metrics to market behavior supports that view. Investors, on this understanding, are forming trade-prompting views progressively and in response to others' views, as evinced on social media (consistent with the model of Hong & Stein, 1999). Furthermore, our analysis shows sentiment metrics better predicting price volatility and accelerated share buying and selling. We are less-able, with this data, to predict the valuation of assets.

We argue, consistent with Oh et al. (2013), it is largely collective uncertainty, and perhaps some blending of fear and greed, being widely and quite-rapidly transmitted by social media. These socially volatile effects appear more significant than a communal expectation of share price returns. Hence, the very absence of “consensus” can itself produce the (meta-level) wisdom of knowing there is little directional wisdom. Dissonance of sentiment predicts unpredictability of price direction, along with elevated volatility and trading. The market participants we study indirectly and quite-rapidly furnish – rather than embody – wisdom. They are more a hasty mob than a wise crowd.

## 5.1 Limitations of the Study and Ideas for Further Research

In common with other such studies, this paper contains a number of limitations, many of which invite further study. First, while our study data is, to the best of our knowledge, the most granular and long-term to have been tested and published to date, there remains a strong case for studying tick price data matched to real time sentiments. It is widely recognized that financial markets respond very quickly to unfolding events and so it would be a natural extension of prior studies to consider causal effects over ever-shorter intervals. Moreover, our model finds mixed evidence for directional prediction from sentiments. Given the value of directional prediction to market participants, there is benefit in adding a binary test of directional predictive ability.

Amongst mixed findings for improved forecast error, we find strongest evidence of predictability for Apple and Goldman Sachs. Noting that Apple is the most-tweeted stock of all, this is suggestive of predictability being linked to the volume of relevant micro-blogging activity. With only five stocks in this study, we cannot yet definitively say if message volume is statistically important in comparing the predictability of the market behaviour of many traded assets.

In analysing five well-known stocks we might have missed interesting phenomena in other areas. Assets other than shares, such as commodities or currencies, plausibly could respond quite differently to sentiments. In addition, the two years of studied data do not contain a large and sustained market crash (as per 2008-2009, for example, during which there was no great volume of tweeting about stocks). We can suppose the relationship between sentiment metrics and market behavior will change in the course of such profound and sustained turmoil.

The use of Artificial Neural Networks (“ANNs”) – amongst other AI-based techniques - is, arguably, well-suited to exploring the relationships we study (c.f. Wong, 1991). However, as a first step in revealing notable causal ties between variables, the regression-based method we use (founded on the approach of Granger, 1969) has advantages. For example, Tu, (1996) argues that regression reduces the risk of over-fitting, and is more appropriate to exploring theory-based (rather than purely-empirical) relationships. Nevertheless, future studies would benefit from exploring ANNs applied to granular data for many market assets and related sentiment.

We report causal windows of several minutes for volatility. While this might be of high interest to some market participants, such as market makers, we remain unclear as to the general usefulness of this finding in market trading. The trading value for volatility indices, such as the VIX, remains untested.

Finally, this study has focused on informational advantages conferred by sentiment metrics. It does not address the issue of whether, how much, and how quickly any such advantages are eroded or changed by their use in trading.

## 5.2 Conclusions

We claim empirical contributions to the research literature in two areas. First, we find limited evidence of Granger-causality from social media sentiment metrics to all three stock market behaviours considered: returns; volatility; and trading volume. These causal effects are salient over a time window of a minutes, whereas prior studies focus on daily data, with a few studying data granular to the half or quarter hour. Second, we find some modest and selective evidence of improved forecast errors, upon using sentiment metrics, for all of the five stocks considered. In the cases of Goldman Sachs and Apple, most forecast errors were improved by a sentiment-based prediction model. Across all five stocks, price volatility and trading volume appear more predictable than price direction. We specify classes of trading strategy most apt to building on such a predictive foundation.

This study contributes to theory by adding to arguments in support of emotion being salient to investment decision making, and by finding unique evidence on the rate and manner in which valuable information diffuses between market participants (c.f. Hong, Lim & Stein, 2000). Our findings are consistent with traders' actions being selectively influenced by the steady accumulation of other traders' opinions. The findings are further embedded in the work of Surowiecki, (2004); the strongest causal links between sentiment and price behaviour are found in times of strident and discordant market mood. The aggregated views of market participants do not, therefore, wholly embody "the wisdom of the crowds", and yet they still have predictive value. Sometimes traders agree. Sometimes, and more notably, traders take sides. Either way, accumulated sentiment – concordant or discordant – has modestly predictable consequences for markets.

References

Andersen, T., Bollerslev, T., & Diebold, F. (2007). Roughing it up: Including Jump Components in the Measurement, Modeling and Forecasting of Return Volatility, Review of Economics and Statistics. 89, pp. 701-720.

Antweiler, W., & Frank, M.Z. (2004). Is all that talk just noise? The information content of internet stock message boards. The Journal of Finance. 59(3), pp. 1259–1294.

Armentano, M.G., Godoy, D., & Amandi, A.A. (2013). Followee recommendation based on text analysis of micro-blogging activity. Information Systems. 38. Pp. 1116-1127.

Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. Journal of Computational Science. 2(1), pp. 1–8.

Chordia, T., Roll, R., & Subrahmanyam, A. (2005). Evidence on the speed of convergence to market efficiency. Journal of Financial Economics. 76, pp. 271–292.

Cutler, D. M., Poterba, J. M. & Summers, L. H. (1989). What moves stock prices?, The Journal of Portfolio Management. 15 (3), pp. 4–12.

Datta, P., & Mbarika, V. W. (2006). A Global Investigation of Granger Causality between Information Infrastructure Investment and Service-Sector Growth. Information Society. Jul-Aug 2006, Vol. 22 issue 3, p149-163.

Dimson, E. (1979). Risk measurement when shares are subject to infrequent trading. Journal of Financial Economics. 7(2), 197-226.

Fama, E. (1965). The Behavior of Stock Market Prices. Journal of Business. 38: 34

Freeman, J.R. (1983). Granger Causality and the Times Series Analysis of Political Relationships. American Journal of Political Science. Vol. 27, No. 2 (May, 1983), pp. 327-358.

Granger, C. W. J. (1969). Investigating Causal Relations by Econometric Models and Cross-spectral Methods. Econometrica 37 (3). pp. 424–438.

Hong, H., Lim, T. & Stein, J. C. (2000), Bad News Travels Slowly: Size, Analyst Coverage, and the Profitability of Momentum Strategies. The Journal of Finance, 55: 265–295.

Hong, H., & Stein, J.C. (1999). "A Unified Theory of Underreaction, Momentum Trading and Overreaction in Asset Markets." Journal of Finance (6): 2143-2184.

Kontopoulos, E., Berberidis C., Dergiades, T. & Bassiliades, N. (2013). Ontology based sentiment analysis of twitter posts. Expert Systems with Applications. 40, pp. 4065–4074.

Lehar, A., Scheicher, M., & Schittenkopf, C. (2002). GARCH vs. stochastic volatility: Option pricing and risk management. Journal of Banking & Finance, 26(2), 323-345.

Lewis, M. (2014). Flash Boys: A Wall Street Revolt. New York : W.W. Norton & Company.

Li, Y.M., & Li, T.Y. (2013). Deriving Intelligence from Microblogs. Decision Support Systems. 55. pp. 206–217.

LIWC Linguistic Inquiry and World Count. (2016). http://liwc.wpengine.com/how-it-works/ Accessed 05.10.16

Lowenstein, G., & Lerner, J.S. (2003). The Role of Affect in Decision Making. In R. Davidson, K. Scherer, & H. Goldsmith (Eds.), Handbook of Affective Science, pp. 619-642. New York: Oxford University Press.

Makrehchi, M., Shah, S., & Liao, W. (2013). Stock prediction using event-based sentiment analysis. In Proceedings of IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT), November.

Matsubara, Y., Yasushi, S., Prakash, B.A., Li, L., & Faloutsos. C. (2012). Rise and fall patterns of information diffusion: model and implications. In Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining (KDD '12). ACM, New York, NY, USA, 6-14.

Mittermayer, M. A., & G.F. Knolmayer, (2006). Newscats: A News Categorization and Trading System, Proceedings of the 6th International Conference on Data Mining. (ICDM), IEEE, pp. 1002–1007.
Psychsignal. (2014). www.psychsignal.com Accessed July 14th 2014.

Nassirtoussi, A.K., Aghabozorgi, S., Ying Wah, T., & Ling Ngo, D.C. (2014). Text Mining for Market Prediction: A Systematic Review. Expert Systems with Applications. Volume 41, Issue 16, 15 November 2014, pp. 7653–7670.

Oh, O., Agrawal, M. & Raghav Rao, H. (2013). Community Intelligence and Social Media Services: A Rumor Theoretic Analysis of Tweets During Social Crises. Management Information Systems Quarterly. Volume: 37. Issue: 2 pp. 407-426.

Oliveira, N., Cortez, P., & Areal, N. (2013). On the Predictability of Stock Market Behavior Using StockTwits Sentiment and Posting Volume. Progress in Artificial Intelligence: Lecture Notes in Computer Science. Vol. 8154, 2013, pp. 355-365.

Ruiz, E.J., Hristidis, V., Castillo, C., Gionis, A., & Jaimes, A. (2012). Correlating financial time series with micro-blogging activity. In Proceeding of ACM International Conference on Web Search and Data Mining (WSDM), Seattle, WA, February.

Sprenger, T., & Welpe, I. (2010). Tweets and Trades: The Information Content of Stock Microblogs. Social Science Research Network Working Paper Series, pp. 1–89.

Surowiecki, J. (2004). The Wisdom of Crowds: Why the Many Are Smarter Than the Few and How Collective Wisdom Shapes Business, Economies, Societies and Nations. Little, Brown.

Tu, J.V. (1996). Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes. Journal of Clinical Epidemiology, Volume 49, Issue 11, Pages 1225-1231.

Twitter. (2014). https://about.twitter.com/company. Accessed August 18th 2014.

Wong, F.S. (1991). Time series forecasting using backpropagation neural networks, Neurocomputing, Volume 2, Issue 4, July 1991, Pages 147-159.

Yu, Y., Duan, W., & Cao, Q. (2013). The Impact of Social and Conventional Media on Firm Equity Value: A Sentiment Analysis Approach. Decision Support Systems. 55 (4), pp. 919–926.

Table 1. Descriptive Statistics (Feb 17th 2012 to Oct 17th 2014)

| | **Apple** Mean (s.d.) | **Amazon** Mean (s.d.) | **Goldman Sachs** Mean (s.d.) | **Google** Mean (s.d.) | **IBM** Mean (s.d.) |
|---|---|---|---|---|---|
| *B* | 0.103 (0.598) | 0.013 (0.262) | 0.007 (0.157) | 0.022 (0.303) | 0.005 (0.145) |
| *A* | 0.353 (0.470) | 0.067 (0.248) | 0.023 (0.149) | 0.089 (0.283) | 0.020 (0.139) |
| Return | 0.055 (0.122) | 0.225 (0.133) | 0.364 (0.098) | 0.350 (0.099) | -0.103 (0.079) |
| Volume | 12.596 (0.842) | 9.231 (0.925) | 9.290 (0.833) | 9.229 (1.011) | 9.519 (0.759) |
| Variance | 0.015 (0.231) | 0.018 (0.287) | 0.010 (0.052) | 0.010 (0.154) | 0.006 (0.127) |
| Realized Variance | 0.178 (0.206) | 0.209 (0.212) | 0.161 (0.148) | 0.154 (0.158) | 0.116 (0.133) |

Notes: Values in the table are the mean and standard deviation (in parenthesis) of the sentiment indices and financial indicators of the different stock indices. The mean of returns is in thousands. B represents the Bullishness index while *A* stands for the Agreement index.

Table 2. Spearman Correlation coefficients between sentiment data indices and financial variables of a sample of stock indices

| | **Apple** | | **Amazon** | | **Goldman Sachs** | | **Google** | | **IBM** | |
|---|---|---|---|---|---|---|---|---|---|---|
| | *B* | *A* | *B* | *A* | *B* | *A* | *B* | *A* | *B* | *A* |
| | | | | | | | | | | |
| **Return** | 0.009* | -0.003 | 0.003 | 0.004 | 0.006* | 0.008* | 0.013* | 0.002 | 0.008* | 0.008* |
| | | | | | | | | | | |
| **Volume** | 0.005 | 0.145* | 0.022* | 0.013* | 0.033* | 0.080* | 0.027* | 0.134* | 0.017* | 0.083* |
| | | | | | | | | | | |
| **Variance** | 0.006 | 0.154* | 0.015* | 0.106* | 0.027* | 0.078* | 0.023* | 0.111* | 0.017* | 0.076* |
| | | | | | | | | | | |
| **Realized Variance** | 0.007* | 0.162* | 0.015* | 0.112* | 0.028* | 0.082* | 0.023* | 0.116* | 0.018* | 0.077* |

Notes: (*) significant at 5%. Values in the table show the Spearman's rank correlation coefficients between the Bullishness index (B) and the financial variable; and between the Agreement index (A) and the financial variable for each stock.

Table 3. P-Values of Granger Causation Tests between sentiment indices and financial indicators of a sample of stock indices.

| | **Volatility** | | **Realized Volatility** | | **Volume** | | **Returns** | |
|---|---|---|---|---|---|---|---|---|
| Stock Inde | **B** | **A** | **B** | **A** | **B** | **A** | **B** | **A** |
| **AMAZON** | | | | | | | | |
| 10 LAGS | 0.2412 | 0.5402 | 0.4171 | 0.0028*** | 0.0487** | 0.0000*** | 0.0629* | 0.6951 |
| 20 LAGS | 0.4370 | 0.1018 | 0.5046 | 0.0000*** | 0.0525* | 0.0000*** | 0.0260** | 0.9556 |
| **APPLE** | | | | | | | | |
| 10 LAGS | 0.8513 | 0.1599 | 0.0386** | 0.0002*** | 0.0034*** | 0.1564 | 0.1127 | 0.1282 |
| 20 LAGS | 0.7048 | 0.0046*** | 0.0232* | 0.0000*** | 0.0155** | 0.0031*** | 0.0046*** | 0.3037 |
| **GS** | | | | | | | | |
| 10 LAGS | 0.0000*** | 0.0000*** | 0.0000*** | 0.0000*** | 0.8764 | 0.0002*** | 0.0084*** | 0.8748 |
| 20 LAGS | 0.0000*** | 0.0000*** | 0.0000*** | 0.0000*** | 0.4798 | 0.0000*** | 0.0089*** | 0.9890 |
| **GOOGLE** | | | | | | | | |
| 10 LAGS | 0.1392 | 0.2163 | 0.0480** | 0.0478** | 0.9014 | 0.0000*** | 0.0216** | 0.7922 |
| 20 LAGS | 0.0160** | 0.2874 | 0.0037*** | 0.0276** | 0.5029 | 0.0000*** | 0.0369** | 0.5542 |
| **IBM** | | | | | | | | |
| 10 LAGS | 0.1763 | 0.4411 | 0.0001*** | 0.0000*** | 0.3088 | 0.0000*** | 0.2768 | 0.4265 |
| 20 LAGS | 0.1384 | 0.2759 | 0.0001*** | 0.0001*** | 0.5999 | 0.0000*** | 0.2485 | 0.2452 |

Notes: Values in the table are the p-values of the Granger causality test as stated in Equation (1), where *, **, and *** indicate significance at the 10%, 5% and 1% levels, respectively. In columns B we test whether Bullishness Granger causes the financial indicator (volatility, realized volatility, volume and returns) for each stock. In columns A we test whether Agreement Granger causes the financial indicator (volatility, realized volatility, volume and returns) for each stock.

Table 4. Forecasting Results. MAPE and RMSE Estimates

**AMAZON**

***Model 0: Only includes lags (n=20) for dependent variable, except for Return where n=10. Controls for time effects***

(1) Return, (2) Volume, (3) Volatility, (4) Realized Variance.

| | | |
|---|---|---|
| (1) | MAPE=102.62892 | RMSE=0.15789524 |
| (2) | MAPE=4.771451 | RMSE=0.56122968 |
| (3) | MAPE=120.68136 | RMSE=0.30720081 |
| (4) | MAPE=16.69075 | RMSE=0.13058024 |

***Model 1A: (Adds the Bullishness metric)***

| | | |
|---|---|---|
| (1) | MAPE=103.06181 | RMSE=0.15793618 |
| (2) | MAPE=4.7729716 | RMSE=0.56132156 |
| (3) | MAPE=149.85669 | RMSE=0.30710101 |
| (4) | MAPE=16.719097 | RMSE=0.13060939 |

***Model 1B: (Adds the Agreement metric)***

| | | |
|---|---|---|
| (1) | MAPE=102.92471 | RMSE=0.15790816 |
| (2) | MAPE=4.7703562 | RMSE=0.56100849 |
| (3) | MAPE=143.09904 | RMSE=0.30720381 |
| (4) | MAPE=16.768402 | RMSE=0.13054429 |

**APPLE**

***Model 0: Only includes lags (n=20) for dependent variable, except for Return where n=10. Controls for time effects***

(1) Return, (2) Volume, (3) Volatility, (4) Realized Variance.

| | | |
|---|---|---|
| (1) | MAPE=99.768799 | RMSE=0.10024473 |
| (2) | MAPE=2.587934 | RMSE=0.41124581 |
| (3) | MAPE=181.23047 | RMSE=0.01917382 |
| (4) | MAPE=19.041611 | RMSE=0.05393209 |

***Model 1A: (Adds the Bullishness metric)***

| | | |
|---|---|---|
| (1) | MAPE=99.871727 | RMSE=0.10025626 |
| (2) | MAPE=2.5875969 | RMSE=0.41117315 |
| (3) | MAPE=212.96416 | RMSE=0.01924627 |
| (4) | MAPE=19.058155 | RMSE=0.0539343 |

***Model 1B: (Adds the Agreement metric)***

| | | |
|---|---|---|
| (1) | MAPE=99.72673 | RMSE=0.10024468 |
| (2) | MAPE=2.5873389 | RMSE=0.41118618 |
| (3) | MAPE=237.25314 | RMSE=0.01924928 |
| (4) | MAPE=18.902777 | RMSE=0.05388988 |

**GOOGLE**

***Model 0: Only includes lags (n=20) for dependent variable, except for Return where n=10. Controls for time effects***

(1) Return, (2) Volume, (3) Volatility, (4) Realized Variance.

| | | |
|---|---|---|
| (1) | MAPE=102.41462 | RMSE=0.09263055 |
| (2) | MAPE=6.0852337 | RMSE=0.64821957 |
| (3) | MAPE=129.16119 | RMSE=0.02726294 |
| (4) | MAPE=16.329178 | RMSE=0.05434002 |

***Model 1A: (Adds the Bullishness metric)***

| | | |
|---|---|---|
| (1) | MAPE=102.9474 | RMSE=0.09264848 |
| (2) | MAPE=6.0848203 | RMSE=0.64819014 |
| (3) | MAPE=166.01393 | RMSE=0.02731333 |
| (4) | MAPE=16.358107 | RMSE=0.05432975 |

***Model 1B: (Adds the Agreement metric)***

| | | |
|---|---|---|
| (1) | MAPE=102.53411 | RMSE=0.09263535 |
| (2) | MAPE=6.0791068 | RMSE=0.64780432 |
| (3) | MAPE=172.92926 | RMSE=0.02731346 |
| (4) | MAPE=16.275864 | RMSE=0.05436601 |

**Goldman Sachs**

***Model 0: Only includes lags (n=20) for dependent variable, except for Return where n=10. Controls for time effects***

(1) Return, (2) Volume, (3) Volatility, (4) Realized Variance.

| | | |
|---|---|---|
| (1) | MAPE=100.22709 | RMSE=0.09024394 |
| (2) | MAPE=4.9991579 | RMSE=0.57623335 |
| (3) | MAPE=142.74214 | RMSE=0.02559098 |
| (4) | MAPE=20.363062 | RMSE=0.05597522 |

***Model 1A: (Adds the Bullishness metric)***

| | | |
|---|---|---|
| (1) | MAPE=100.40341 | RMSE=0.0902298 |
| (2) | MAPE=4.999197 | RMSE=0.57623128 |
| (3) | MAPE=145.03264 | RMSE=0.02558823 |
| (4) | MAPE=20.337393 | RMSE=0.05592784 |

***Model 1B: (Adds the Agreement metric)***

| | | |
|---|---|---|
| (1) | MAPE=100.21034 | RMSE=0.09023957 |
| (2) | MAPE=4.998569 | RMSE=0.57628744 |
| (3) | MAPE=144.10599 | RMSE=0.02558749 |
| (4) | MAPE=20.367056 | RMSE=0.05596454 |

**IBM**

***Model 0: Only includes lags (n=20) for dependent variable, except for Return where n=10. Controls for time effects***

(1) Return, (2) Volume, (3) Volatility, (4) Realized Variance.

| | | |
|---|---|---|
| (1) | MAPE=99.819504 | RMSE=0.06583014 |
| (2) | MAPE=4.5258126 | RMSE=0.53248495 |
| (3) | MAPE=125.63831 | RMSE=0.00791805 |
| (4) | MAPE=18.606939 | RMSE=0.03603948 |

***Model 1A: (Adds the Bullishness metric)***

| | | |
|---|---|---|
| (1) | MAPE=99.940163 | RMSE=0.06582234 |
| (2) | MAPE=4.5261135 | RMSE=0.53255101 |
| (3) | MAPE=169.58861 | RMSE=0.00818384 |
| (4) | MAPE=18.648745 | RMSE=0.03606617 |

***Model 1B: (Adds the Agreement metric)***

| | | |
|---|---|---|
| (1) | MAPE=99.861809 | RMSE=0.06582611 |
| (2) | MAPE=4.5257297 | RMSE=0.53248565 |
| (3) | MAPE=164.94598 | RMSE=0.00817617 |
| (4) | MAPE=18.660328 | RMSE=0.03604659 |

Notes: Forecasting results in terms of MAPE (Mean absolute percentage error) and RMSE (Root mean squared error) for the financial variables studied, are presented. Underlined data indicates that the sentiment-based model produces smaller errors than the base model (“Model 0”).