Can Facebook Predict Stock Market Activity?*

Yigitcan Karabulut[†]

Goethe University Frankfurt

First Draft: August 29, 2011

This Draft: October 17, 2011

-Preliminary Draft-

Please do not quote without permission

Abstract

Using a novel and direct measure of investor sentiment, I find that Facebook's Gross National Happiness (GNH) has the ability to predict changes both in daily returns and trading volume in the US equity market. For instance, a one standard deviation increase in GNH predicts an increase in market returns equal to 11 basis points over the next day. Moreover, the impact of GNH appears to be stronger among small-cap stocks, and in the face of turmoil.

Keywords: Investor sentiment, social media, behavioral finance, Facebook.

JEL Codes: D81, G11, G12.

†Yigitcan Karabulut is with Goethe University Frankfurt, Retail Banking Competence Center, House of Finance, Grüneburgplatz 1, 60323, Frankfurt, Germany, E-mail: karabulut@finance.uni-frankfurt.de, Phone: +49 (0) 69 798 33859. The usual disclaimer applies.

1 Introduction

The question of whether sentiment affects stock market has long attracted a great deal of attention from academics. Following the influential work of De Long et al. (1990), a number of studies address this issue, and show that noise trader sentiment can persist in financial markets and influence asset prices (e.g. Lee et al., 1991, Barberis et al., 1998; Daniel et al., 2001; Tetlock, 2007).[1] Thus the relevant question is not any longer whether sentiment has an effect on stock prices, but rather how to measure investor sentiment more accurately and ascertain its effects on the stock market (Baker and Wurgler, 2006).

In this paper, I investigate the interactions between investor sentiment and stock market using a new and direct measure of sentiment. Specifically, I employ Gross National Happiness (hereafter referred to as, GNH) of Facebook, which captures daily sentiment using content from the individual status updates of almost 100 million Facebook users in the US.

I attempt to quantify the effects of Facebook sentiment on stock market by examining three questions, which are derived from the behavioral models of securities market (e.g. De Long et al., 1990; Campbell et al., 1993). First, I ask whether the sentiment measure compiled from Facebook displays any ability to predict both daily returns and trading volume in the stock market. Second, I analyze the question of how sentiment impacts on the cross-section of stock returns rather than its effect on the aggregate market returns. Finally, I study the interactions between Facebook sentiment and market returns during unusually volatile episodes of stock market. To answer these questions, I adopt a vector autoregressive (VAR) framework, using daily stock market and sentiment data over the four year period from September 10, 2007 through September 9, 2011.

Estimation results indicate that Facebook sentiment, as measured by GNH, displays ability to predict daily market returns. The results are statistically significant and eco-

---

[1]For instance, the model of De Long et al. (1990) infers that asset prices that are widely held by noise traders may deviate from their fundamental values for longer time periods if the noise trader sentiment is correlated.

nomically meaningful. For instance, a one standard deviation increase in GNH has an impact of 11 basis points increase in the next day's returns, which is higher than the mean market return during the sample period. Further, I find a stronger impact of GNH on daily returns among small-cap stocks, and in the face of market turmoil. The former finding is in line with the results of existing literature that small-cap stocks are disproportionately held by small investors, and hence, more strongly affected by sentiment (Kumar and Lee, 2006; Baker and Wurgler, 2006). Finally, unusually high levels of Facebook sentiment are also associated with higher future trading volume that provides direct support for the model predictions of Campbell et al. (1993).

The key contribution of this paper is to propose a novel and direct measure of investor sentiment, which has particularly attractive properties. First, the sentiment measure, GNH, is compiled from Facebook, which is the world's largest social network, with 750 million active users worldwide as of August 2011.[2] There are more than 150 million Facebook users in the US, covering almost 50 percent of the population (and almost 70 percent of the online population) from different age groups and genders. Hence, the observed pervasive use of Facebook makes GNH a reasonably representative sentiment measure for the entire US population.[3] Second, GNH is computed using content from individual status updates of Facebook users. Specifically, a status update is a short-format note as a response to the question of '*What's on your mind?*'.[4] Therefore, individual updates are generally self-descriptive, undirected, and hence, have affective content about the user. Keeping this in mind, it can be argued that Facebook's status update is a more appropriate and better choice for directly measuring sentiment as compared to other social media tools such as blog entries, messages in online forums or microblogging posts (Kramer, 2010). Third, GNH is measured on a daily basis from the beginning of September 2007, providing a high frequency sentiment measure for a relatively longer time period. Finally,

---

[2]Put differently, recall the fact that world population accounts for 6.96 billion as of September 2011, the number of Facebook users suggests that 1 in every 9 people on Earth is on Facebook.

[3]The Facebook user statistics are obtained from www.checkfacebook.com.

[4]This question and corresponding answer field show up in the homepage of Facebook whenever the user logs on the web site.

sample period includes one of the most striking episodes of the US economy and equity market that also allows me to study the relation between sentiment and stock market during an unusually volatile time period. Overall, all these characteristics motivate to use Facebook measure to capture sentiment among the US population.

To my knowledge, this paper is one of the first that utilizes information from online social networking sites in finance.[5] Therefore, apart from testing the theories of investor sentiment, this paper also highlights the usefulness of data from online social networks, which could possibly provide a rich source of information for other possible economics and finance applications.[6]

The remainder of the paper is organized as follows. Section 2 provides background information and some theory for studying the impact of sentiment on stock market activity. Section 3 describes the sentiment measure employed in the paper and further provides general information about Facebook. In Section 4, I introduce the dataset and provide some summary statistics. Section 5 describes the estimation strategy and reports the main findings. Section 6 concludes the paper.

2   Theory and Background

This section provides motivation for studying the impact of sentiment on stock market activity by briefly reviewing the existing theoretical and empirical literature on the role of investor sentiment in stock market activity.

The classic theory of securities market posits that market participants are rational, and hence, asset prices in equilibrium reflect rationally evaluated and discounted future cash flows and investment risks (e.g. Sharpe, 1964; Lintner, 1965). Nevertheless, highly

---

[5]For instance, in an earlier study, Bollen, Mao, and Zhang (2010) measure the collective mood of the US population using content from the microblogging posts (i.e. tweets) from Twitter; an online social network. Their results imply that changes in the public mood can be tracked from the Twitter mood. Further, among the 7 observed mood dimensions that they have constructed, only some are associated with shifts in the Dow Jones Industrial Average values.

[6]Data from online social networking sites can be especially insightful and interesting for applications, which focus on information transmission and social networks (e.g. Cohen et al., 2010).

speculative episodes in the stock markets such as the internet bubble in the late 1990s, or more recently the subprime credit crisis, create a significant hurdle, which challenges the premise of 'pure rationality' of the classic theory. To understand such wild movements in stock markets, recent theoretical models of securities markets relax the 'pure rationality' premise, and also give a role to investor sentiment in asset prices (e.g. De Long et al., 1990; Lee et al., 1991; Barberis et al., 1998).

In particular, behavioral models of securities markets consider two types of investors: Rational arbitrageurs who are immune to sentiment and form rational expectations about asset returns, and noise traders who are subject to exogenous sentiment and form either overly optimistic or pessimistic beliefs relative to rational expectations. These two types of investors compete in financial markets where the asset prices are determined based on their respective beliefs.

If asset prices deviate from their fundamental values due to a demand (supply) shock from noise traders, standard theory argues that rational arbitrageurs would force them to their fundamentals by taking positions against noise traders. However, arbitrageurs may not be willing to bet against mispricing as they are prone to several limitations such as having a short investment horizon or costs and risks of trading and short selling (De Long et al., 1990). Hence, noise trader sentiment can persist in the financial market and affect security prices in equilibrium when arbitrage is limited. In a nutshell, existing behavioral models establish the role of investor sentiment in asset price patterns.

Accordingly, a large body of empirical literature attempts to measure investor sentiment and to assess its effects on stock market activity (e.g. Neal and Wheatley, 1998; Lamont and Thaler, 2003). So far several different proxies have been employed to capture investor sentiment as there is no obvious sentiment measure.[7] For instance, Neal and Wheatley (1998) investigate the forecast power of three potential sentiment measures;

---

[7]The potential sentiment measures are either direct proxies that are derived from investor surveys or indirect measures such as indicators for fluctuations in investor mood, retail investor trades or mutual fund flows. For a detailed discussion of potential investor sentiment measures, please see Baker and Wurgler (2007).

discounts on closed-end funds, net mutual fund redemption, and the ratio of odd-lot sales to purchases. Their results indicate that two of these three proxies, namely fund discounts and net redemptions display some ability to predict the size premium and the difference between small and large firm returns. However, Neal and Wheatley (1998) only document a very weak evidence that odd-lot ratio predicts market returns. In another similar paper, Baker and Wurgler (2006) examine the question of how sentiment affects the cross-section of stock returns rather than its impact on the aggregate market returns. They measure investor sentiment by a composite index of six commonly used sentiment variables.[8] Baker and Wurgler (2006) document that smaller stocks, high volatility stocks, unprofitable stocks, non-dividend-paying stocks, extreme-growth stocks, and distressed stocks earn high returns following the periods of low sentiment, and they earn low returns when the sentiment is high.

Another line of the empirical literature attempts to link the stock market returns to fluctuations in human emotions that are creatively identified by employing exogenous mood indicators (Hirshleifer and Shumway, 2003; Kamstra et al., 2003). For instance, Edmans et al. (2007) construct a sport sentiment, using the results of international soccer games. The idea of using sport results as a mood indicator is motivated by evidence from the psychology literature. Specifically, it is documented that there is a significant change in the behavior of fans following wins and losses of their teams (Wann et al., 1994). Similarly, Edmans et al. (2007) also document a loss effect on stock market returns, which is even more pronounced among small-cap stocks, and after more important games.

In contrast to these studies that rely on indirect proxies of investor sentiment, this paper focuses on a direct and new measure of sentiment compiled from an online social network, and investigates the role of sentiment in daily market returns.[9]

---

[8]The sentiment index of Baker and Wurgler (2006) include widely used sentiment variables such as value-weighted dividend premium, the number of IPOs, the average first-day IPO return, the value-weighted CEFD, the equity share in new issues, and NYSE turnover.

[9]The measure of sentiment employed in this paper also has several advantages over other direct measures of sentiment such as those derived from investor surveys. For instance, Baker and Wurgler (2007) note that sentiment measures complied from investor surveys are subject to some criticism as investors may respond to a survey differently than they would actually behave, which is not the case for the sentiment measure employed in this paper.

Finally, the model of Campbell et al. (1993) infers that investor sentiment would also have an effect on trading volume in the markets. Specifically, when noise traders form overly optimistic (pessimistic) expectations about the future stock market outcomes, they would intensively buy (sell) stocks. Accordingly, the arising unusually high level of demand (supply) from noise traders will be offset by market makers to restore the market equilibrium, resulting in higher trading volume. In a recent paper, Tetlock (2007) addresses this issue and provides evidence, which is consistent with the predictions of the model of Campbell et al. (1993). In particular, he documents that unusually high or low levels of sentiment as measured by a self-constructed media index based on a popular Wall Street Journal column, appears to predict high trading volume in Dow Jones. Interestingly, Tetlock (2007) also finds a direct effect of media sentiment on trading volume, which he attributes to the trading costs argument (Antweiler and Frank, 2004).[10] The interactions between trading volume and stock market returns represent also another issue, which I examine later in the paper.

3 Facebook as a Measure of Investor Sentiment

One of the key contributions of this paper is to study a variable that captures investor sentiment in a direct and more timely manner. Specifically, I employ Gross National Happiness as the measure of investors sentiment that is provided by Facebook; a popular online social networking site.

Broadly, Facebook can be described as an online social networking tool which facilitates its members' communication with their friends and families in a more efficient way. In particular, the platform enables its users to present themselves in an online profile and to make friends who can post comments on each other's pages, and view each other's profiles.

---

[10]In particular, the measure of communication used by Antweiler and Frank (2004) is related to direct and indirect trading costs such as costs of liquidity and volume. Based on the findings of the existing literature, Tetlock (2007) argues that there should be a negative relationship between his media pessimism index and trading volume if pessimism proxies for trading costs.

In recent years, Facebook is becoming an increasingly important part of everyday life. To obtain an impression about its increased importance, I next highlight some statistics. First, it is estimated that Facebook has worldwide more than 750 million active users, of which 50 percent log on to the site on any given day. Further, the average Facebook user has 130 friends in her network and spends about 31.1 minutes a day on the site, which makes an aggregate total of 700 billion minutes per month.[11]

Table 1 provides some further statistics on the number of Facebook users and corresponding percentage shares in the online population for the ten largest Facebook nations. As can be seen in the Table, Facebook has more than 150 million members only in the US, covering almost 50 percent of the entire population and 70 percent of the online population in this country. Similarly, there are almost 30 million users (69.5 percent of the online population) in the UK and 20 million users (37.6 percent of the online population) in Germany. Taken as a whole, all these numbers emphasize the important role as well as worldwide pervasive use of Facebook among individuals.[12]

One possible concern associated with employing a sentiment measure that is compiled from Facebook is the representation of the entire population. Particularly, it is generally believed that Facebook is rather used by younger people, therefore, older population in this online social network is underrepresented. However, demographic characteristics of the US American Facebook users, as reported in Table 2, paint a picture that remove these possible concerns: Of the 154.5 million members in the US, only 10.4 percent are under 18 years whereas the share of users age 25 to 34, and 35 to 44 account for 23.6 and 16.6 percent, respectively. Finally, the share of Facebook users who age 55 and older accounts for 12.4 percent. Moreover, 54.7 percent of the Facebook users are female and the remaining 45.3 percent are male, suggesting that both genders in Facebook are almost equally represented. Overall, statistics on the demographic characteristics of

---

[11]The user statistics are obtained from the Facebook web site. For further information, please see Facebook factsheet that is available on the web site.

[12]One other possible indicator for the Facebook's increased importance is its estimated market value. In particular, Facebook's market value is estimated to be at 50 billion US Dollars as of January 2011 according to reports from *the New York Times*. For further details, please see *the New York Times* article, *Goldman Offering Clients a Chance to Invest in Facebook* from January 2, 2011.

the Facebook users underpin the common use of Facebook by different age groups and genders, indicating that the sentiment measure compiled from Facebook would reasonably represent the aggregate sentiment among the US population.

The sentiment variable employed in the analysis is constructed on a daily basis using content from the individual status updates of each active Facebook user in the US, of which there are more than 40 million posts on a given day (Kramer, 2010). Particularly, status updates in Facebook are short-format notes that contain text provided by the user as a response to the question of '*What's on your mind?*'. The question and corresponding answer field show up in the homepage of Facebook whenever the user logs on to the website. Figure 1 illustrates examples of Facebook's status update.

As noted by Kramer (2010), a status update is a self-descriptive text modality that is designed to share personal updates. Therefore, they generally include more emotional or affective content as compared to wall posts in Facebook or messages in other online social media tools (Kramer, 2010). Moreover, status updates are generally not directed to a specific target like wall posts or Twitter posts, both of which do not necessarily include any information or feelings about the user itself. Keeping this in mind, it can be argued that Facebook's status updates seem to be the most appropriate text modalities from online social network sites to measure the sentiment among the population.

The sentiment measure, GNH, is introduced and developed by Adam D.I. Kramer, Lisa Zhang and Ravi Grover from the Facebook Data Team.[13] Particularly, GNH is calculated using the 'word-count' methodology as explained in Pennebaker et al. (2007). In this procedure different sets of words are defined to have different psychological meanings, which are in my case positive and negative emotions.[14] Each individual update is assigned both a positivity and a negativity score by counting the positive and negative emotion words in every post. For instance, a status update of 'It was a good day' has a positivity score of 0.2, and a negativity score of 0 since the only positive emotion word in this

---

[13]I would like to thank the developers of *Gross National Happiness* and the *Facebook Data Team* for making sentiment data available for the analysis.

[14]For the full list of negative and positive words, please see Pennebaker et al. (2007).

example is 'good', and the rest of the words are neutral.[15] This procedure has been repeated on a daily basis for the entire status updates posted by almost 100 million active Facebook members in the US. Consequently, two different affective scores are computed for every day; positivity, and negativity factors. Finally, Facebook calculates the GNH as the standardized difference between positivity and negativity factors, i.e.:

$$GNH_t = \frac{\mu_t^p - \mu^p}{\sigma^p} - \frac{\mu_t^n - \mu^n}{\sigma^n} \tag{1}$$

where GNH_t is the GNH on day t, and μ_t^p and μ_t^n represent the average share of positive and negative words used in the individual updates by Facebook users on a given day, respectively. Finally, μ^p (μ^n) and σ^p (σ^n) represent the mean and the standard deviation of daily share of positive (negative) words across the sample period.

Since GNH consists of two equally-weighted components, I also employ each of the affective dimensions as possible measures for sentiment and investigate separately their impact on the stock market outcomes.[16]

Throughout the paper, a higher score of GNH would imply a higher sentiment whereas a lower GNH would suggest a lower sentiment among the US population. Moreover, each of the positivity and negativity scores can be interpreted as optimism and pessimism factors, respectively. Figure 2 and Figure 3 depict the three sentiment measures over the observation period. As can be seen in Figure 2, there is a sharp decline in the GNH in the beginning of 2008. Interestingly, GNH falls back to its lowest value (-0.058) on 16 September 2008, one of the most dramatic days in Wall Street's history, and jumps to its highest value on December 31, 2009. Finally, it is worth mentioning that positivity factor also follows a similar trend as GNH across the observation period whereas negativity score does not appear to display much variation during the sample period.

---

[15]Following the example of Kramer (2010), a status update of 'Today was kinda good, kinda bad' would have a positivity score of 0.17 because of the word 'good', and a negativity score of 0.17 because of the word 'bad'.

[16]The rationale for the normalization of the two affective components of the Facebook sentiment index is that the number of positive and negative emotion words differ in amounts. In order to compare these scores directly, normalization has been carried out. For further details, please see Kramer (2010).

4 Data and Variable Definitions

Since the primary objective of this paper is investigating the interactions between investor sentiment and stock market activity, both stock market and sentiment data are needed. For the analysis, I downloaded daily GNH, positivity and negativity scores for the US from the Facebook website for the period between September 10, 2007 and September, 9 2011. The observation period comprises 1,044 trading days after excluding the weekends. Since the stock market is idle on national holidays such as on Christmas or Thanksgiving, I conclude the sample selection by excluding national holidays that leaves a final sample of 1,009 trading days.

Stock market data that contain daily returns and volume originate from Thomson Reuters Datastream.[17] I compute the daily returns using total return index, assuming that dividends are reinvested. Moreover, following Campbell et al. (1993) and Tetlock (2007), detrended daily volume in logs is employed as the volume measure since the level of log volume is not stationary. Specifically, I use the detrending methodology proposed by Campbell et al. (1980) where volume trend is computed as the rolling average of the past 60 trading days of log volume and subtracted from the daily volume observation.[18]

Table 3 presents summary statistics for the final sample. In Panel A of Table 3, I first report descriptive statistics on variables of particular interest. The mean (median) value of the Facebook sentiment as measured by GNH is -0.0167 (-0.0155), suggesting that investor sentiment during the observation period was negative. Similarly, the mean daily return in the sample period accounts for -2 basis points. Considering the fact that the sample period contains one of the worst episodes of the US economy, these observations are not surprising. As noted earlier, the lowest value for GNH is observed on September 16, 2008; on the day when the securities firm, Lehman Brothers filed for bankruptcy protection and another big investment bank Merrill Lynch agreed to sell itself to Bank of

---

[17]I use the Thomson Reuters Datastream's mnemonics 'TOTMKUS' to obtain the time-series for return index and trading volume. Further, I note that I employ the stock market return index as measured in US Dollars.

[18]The estimation results are also robust to using longer and shorter rolling windows such as 30 days, 90 days and 120 days in detrending the daily volume.

America to avert a possible bankruptcy filing.[19] Keeping this in mind, it can be argued that there seems to be a link between the Facebook sentiment and stock market. Finally, Table 3 also reports descriptive statistics on other variables employed in the analysis, i.e. positivity and negativity dimensions, detrended volume and volatility.

I also include in the empirical analysis several environmental measures, which have been employed in the existing literature as mood proxies. First, I use the average daily temperature (as measured in Fahrenheits), precipitation (in mm) and wind speed to proxy for the weather-induced mood. The choice of these variables as mood indicator is motivated by the strong evidence from the psychology literature, which shows that almost 40 percent of the variation in mood can be explained by weather (e.g. Persinger and Levesque, 1983). For instance, the literature indicates that higher temperature and more hours of sunshine are associated with higher levels of optimism and lower levels of depression and skepticism (Cunnigham, 1979; Howarth and Hoffman, 1984). Following Saunders (1993) and Hirshleifer and Shumway (2003) who link the weather-induced mood on stock market returns, I also collect weather data for New York City where the stock market is located from the database of the National Climatic Data Center (NCDC). Further, I deseasonalize the weather variables based on the methodology of Hirshleifer and Shumway (2003) since these variables are seasonal. Particularly, I calculate the average value for temperature and precipitation for each calendar week and deduct the mean value from the daily observation to remove pure seasonal variation.

To isolate the possible effects of fluctuations in biorhythm of individuals from the Facebook sentiment, I next calculate the Seasonal Affective Disorder (SAD) variable as described in Kamstra et al. (2003). In their paper, Kamstra et al. (2003) employ seasonal variations in daylight as a mood indicator and investigate its effects on aggregate market returns. Their hypothesis is based on the strong evidence of a relationship between seasonal variation in daylight and depression from the psychology literature (Cohen et al.,

---

[19]See, for instance, New York Times article, *'Lehman in Bankruptcy; Merrill to Be Sold; A.I.G. Struggles'*, from September 15, 2008 for a brief overview of the events.

1992). Specifically, daily darkness duration in New York City is used to calculate the SAD measure. Further, following Kamstra et al. (2003), a dummy variable for fall is also included in the model in order to allow for an asymmetric effect of seasonal affective disorder in the fall relative to winter.

Finally, I collect data for lunar phases for the observation period to capture the possible effects of lunar cycles on the Facebook sentiment.[20] In fact, psychology literature fails to find a direct relationship between investor mood and lunar cycles. However, as noted by Dichev and Janes (2001) who document a significant effect of moon phases on stock returns, the observed effect of lunary phase on mood may be related to the tradition of individuals' beliefs about lunar effects on human behavior. Both data on daylight duration and lunar phases are obtained from the United States Naval Meteorology and Oceanography Command (NMOC) database. Panel B of Table 3 presents summary statistics on the environmental variables.

5   Results

In this section, I first introduce the econometric model employed in the empirical analysis and address some technical issues. The second subsection presents the main findings of the paper.

5.1   Econometric Issues

To study the impact of Facebook sentiment on stock market activity, I use a vector autoregressive (VAR) approach, which simultaneously estimates the bidirectional causality between stock market outcomes and Facebook sentiment. The model has the following form:

---

[20]The full moon dummy takes the value 1 up to three days before and after each full moon date and 0 otherwise.

$$z_t = \alpha + \sum_{j=1}^{n=5} \gamma_j \cdot z_{t-j} + \beta \cdot x_t + u_t \qquad (2)$$

where z_t is a three-variable vector; $Sentiment_t$, $Rets_t$, and Vol_t. $Sentiment_t$ is the index score for the Facebook sentiment as measured by GNH, $Positivity$ or $Negativity$ on day t, $Rets_t$ and Vol_t represent the daily return and detrended log daily volume in the stock market on day t, respectively. Finally, x_t is the vector of parameters for control variables.

All lags up to 5 days prior to market activity are included.[21] Three variables in the vector z_t represent the endogenous variables in the system whereas the variables in vector x_t are the exogenous control variables. As control variables, following Tetlock (2007), I first include the five lags of detrended squared residuals to proxy for past volatility in the market.[22] Further, various calendar controls are also included to account for the possible return anomalies. For instance, dummy variables for day of the week, and a dummy variable for the trading day after a national holiday when the stock market is idle are included. Moreover, I construct another variable, Tax_t, that takes the value 1 if day t is in the last trading day or first five trading days of the tax year, and equals to zero otherwise.[23] Monthly fixed effects are also controlled by including dummy variables for each month of the year.

It is also important to note that I control for the days when the GNH is unusually high to ensure that reported results are not driven by any outliers.[24] Finally, I control for

---

[21]In choosing the optimal length of lags, I rely on the Akaike Information Criterion (AIC) which is minimized at 5-lags in my VAR model. Nevertheless, I note that the Schwarz' Bayesian Information Criterion is minimized at 4-lags. I also estimate the VARs using 4-lags and obtain qualitatively similar results.

[22]Please see Tetlock (2007) for the detailed information for the calculation procedure of the proxy for past volatility. It is also important to note that using an alternative volatility measure, i.e. the Volatility Index (VIX) of CBOE does not affect the reported results.

[23]As noted by Kamstra et al. (2003), the tax year begins on January 1 in the United States.

[24]GNH is unusually high on December 31 and on other holidays probably because the Facebook users use widely positive emotions words in their status updates on these days as holiday salutations (e.g. 'Happy Holidays') that contribute to the spikes as shown in the Figure 2 and Figure 3. Therefore, I control for the high values by including dummy variables for each of these days. Nevertheless, Kramer (2010) notes that wishing someone a happy holiday is also a positive emotional act which is therefore emotionally not 'blank'.

environmentally induced mood fluctuations across population by using a set of environmental variables, i.e. weather proxies, SAD and dummy variable for full moon that are drawn from the existing behavioral finance literature.

The error terms in Equation (2) are assumed to be independent of lagged values of endogenous variables in the system, which enables me to estimate each equation separately by ordinary least squares (OLS) method. Finally, I correct the standard errors for any heteroskedasticity and autocorrelation in the residuals up to five lags by employing Newey and West (1987) robust standard errors.

5.2 Predicting Stock Market Activity using Facebook Sentiment

This section presents the results of different tests of whether Facebook sentiment has the ability to predict future stock market activity.

First I investigate the interaction between daily returns and sentiment measures compiled from Facebook. As noted by Kumar and Lee (2006), retail investors have high levels of direct stock ownership in the US equity market. Further, extant studies also indicate that individuals are sentiment-prone investors (e.g. Frazzinia and Lamont, 2008) as they have limited financial capability (Lusardi and Mitchell, 2007) and engage in more attention-based trading (Kumar and Lee, 2006; Barber and Odean, 2008). Keeping this in mind, I estimate the following equation to test the ability of Facebook sentiment to predict daily aggregate market returns:

$$Rets_t = \alpha + \sum_{j=1}^{n=5} \gamma_j \cdot Rets_{t-j} + \sum_{j=1}^{n=5} \theta_j \cdot Sentiment_{t-j} + \sum_{j=1}^{n=5} \eta_j \cdot Vol_{t-j} + \beta \cdot x_t + u_t \quad (3)$$

Since the underlying objective of this paper is to find out whether Facebook sentiment can predict future stock market activity, throughout the paper I focus on the estimates of coefficients on the sentiment variable, θ_j, that describe the dependence of various stock market measures on sentiment factors.

14

Table 4 presents the estimates of coefficients on Facebook sentiment. Each reported coefficient measures the impact of a one standard deviation increase in sentiment factors on daily returns. As reported in the last row of Table 4, the joint significance test for 5-lags of sentiment measure imply that past values of GNH and $Positivity$ factors have statistically significant forecasting power for daily market returns whereas I find no significant effect for $Negativity$ factor. Particularly, the effect of a one standard deviation increase in GNH on next day's return is 11 basis points and the effect of $Positivity$ factor is even slightly more pronounced, which has an impact of 12 basis points on the next day's returns.

Indeed, the magnitudes of these estimates are also economically meaningful. To obtain an impression about the economic importance of these results, I next compare Facebook sentiment's impact with other daily returns. For instance, daily average market return during the observation period accounts for -2 basis points that would be completely offset by a one standard deviation increase in GNH. Similarly, Tetlock (2007) reports that the effect of a one standard deviation increase in his media pessimism index predicts a decrease in Dow Jones returns equal to 8.1 basis points over the next day, which is also in absolute values smaller than the impact of Facebook sentiment. Overall, comparisons with other daily returns suggest that Facebook sentiment seems to have some reasonable forecasting power for future market returns.

Furthermore, as noted earlier, the sample period covers one of the most striking episodes of the US stock markets, namely the subprime crisis period. Apparently, one of the most dramatic events in the recent financial crisis was the failure of the prominent US securities firm, Lehman Brothers which made its bankruptcy filings on September 15, 2008. The collapse of Lehman Brothers triggered an abrupt turmoil in the financial markets, which probably represents one of the worst episodes in Wall Street's history since the Great Depression (Shiller, 2008). Particularly, the failure of Lehman Brothers ignited great uncertainty and anxiety among investors, bringing the solvency of many financial institutions into question. Figure 4 illustrates the daily option implied volatility in the markets as measured by the Volatility Index (VIX) of the Chicago Board Options Ex-

change over the period September 10, 2007 through September 9, 2011.[25] As illustrated in the Figure 4, volatility in the markets jumped to unusually high levels in the final months of 2008 and early 2009, suggesting that financial crisis reached its peak in the period after the collapse of Lehman Brothers.

There is ample evidence that asset prices would significantly deviate from their fundamental values during crisis periods. Particularly, Hu et al. (2010) argue that reluctance of arbitrageurs to bet against mispricing is especially severe in crisis periods since arbitrage capital is scarce and associated risks are high, which altogether may end up in 'more noise' in asset prices. Based on this consideration, it is a-priori to conjecture that Facebook sentiment would display stronger effect on stock market returns in the post-Lehman period, which refers to the 3-year time period from September, 16 2008 through September, 9 2011. Consistent with this hypothesis, as reported in Column 2 of Table 4, I observe that the forecasting power of Facebook sentiment for future market returns indeed improves significantly in the post-Lehman period. Specifically, the impact of a one standard deviation rise in the Facebook sentiment equals to 17 basis points increase in market returns over the next day. There is also a similar pattern in the predictive power of positivity factor whereas negativity factor does not show any significant effect even in the post-Lehman period.

Next I study the impact of Facebook sentiment on the cross-section of stock returns instead of its effect on aggregate market returns. Recall the sentiment theory that deviations from fundamental prices occur as a consequence of both a demand shock from noise traders and constraints to arbitrage. Therefore, the effect of sentiment on the prices of different stocks would vary based on stocks' sensitivity to demand shocks and/or their complexity for arbitrage. For instance, Baker and Wurgler (2006) suggest that both extreme growth and distressed firms are more prone to speculation, therefore, these stocks should be most affected by investor sentiment.[26] Additionally, it is in the literature well documented that

---

[25]VIX measures the implied volatility of options on the Standard and Poor's 500 stock index.

[26]The authors argue that young, small and growth stocks are more prone to speculative demand from the noise traders, probably because of the lack of long earnings history as well as due to the extreme

arbitrage restrictions vary across stocks. Specifically, arbitrage is especially risky, costly and sometimes even impossible for younger, smaller and extreme-growth stocks that impose a limitation to rational investors to completely offset the demand from noise traders, which may end up in more noise in prices (Amihud and Mendelson, 1986; Lamont and Thaler, 2003; Baker and Wurgler, 2006). Considered jointly, it can be conjectured that the impact of investor sentiment would be more pronounced for small cap stocks than its effects on the prices of large-cap stocks.

To address this hypothesis, I analyze the forecasting power of Facebook sentiment for the daily returns of different stock portfolios, which are formed on size.[27] I obtain a time series of daily value-weighted returns for different portfolios from Professor Kenneth French's web site. To test whether sentiment has distinct effects on different stock portfolios, I employ the model as expressed in (3), including the returns of particular stock portfolio into the system.[28]

Table 5 presents the estimation results for portfolio returns formed on size. For the sake of brevity, I report only the impact of Facebook sentiment as measured by GNH and *Positivity*.[29] Each coefficient in the Table measures the impact of a one standard deviation change in sentiment factor on daily returns of the portfolios. As can be seen in the Table, GNH has an incremental ability to predict returns among small-cap stocks, i.e. market value smaller than the median market equity of NYSE, whereas it has both statistically and economically weaker effect on the returns of large-cap stocks. Further, Facebook sentiment as measured by the *Positivity* factor does not show neither positive nor negative impact on the returns of large-cap stocks. Specifically, I cannot reject the null hypothesis

growth potentials of these stocks that allow uniformed traders to form random expectations in a wide range about the future cash flows and investment risks (Baker and Wurgler, 2006).

[27]The portfolios employed in the analysis are constructed as follows: Stocks are classified as small-cap and large-cap where the median value of market equity in NYSE at the end of June in each year constitutes the threshold value. For the detailed information about the construction of these portfolios, please see the data library of Professor Kenneth French at http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/.

[28]It is worth mentioning that I also estimate the VARs without including the overall stock market returns in the system. Excluding the market returns from the system produces qualitatively similar results.

[29]Similar to the previous regressions I also find no significant impact of *Negativity* factor on the daily returns.

that five lags of *Positivity* do not forecast returns of large-cap stocks ($\chi^2(5)$: 8.2). Taken as a whole, this evidence appears to be in line with the idea that small firm stocks are held widely by sentiment-prone individuals (Lee, Shleifer, and Thaler, 1991). Moreover, these results also provide additional support for the findings of the existing literature. For instance, Kumar and Lee (2006) document a relation between retail investor sentiment and returns of small-cap stocks, but no significant relation between sentiment and returns of larger size stock portfolios.

Analogous to the aggregate market return regressions, forecasting power of Facebook sentiment for the returns of small cap-stocks also increases in the post-Lehman period, underpinning the hypothesis that there is more noise in stock returns in crisis periods. Particularly, a one standard deviation increase in GNH has an impact of 19 basis points rise in the next day's returns of the small-cap portfolio. In short, when Facebook sentiment is high, stocks, in particular those which are predominantly held by sentiment-prone investors, display higher future returns as compared to other stocks.

Finally, I turn to the relationship between Facebook sentiment and trading volume. As noted earlier, the underlying rationale behind linking investor sentiment to trading volume is based on the model predictions of Campbell et al. (1993). Particularly, when there is an exogenous positive or negative shock to investor sentiment, noise traders would react to them by buying or selling securities. Accordingly, market makers would take positions against noise traders to restore market equilibrium, which results in higher market volume (Tetlock, 2007). In a nutshell, the hypothesis is that high values of negative or positive sentiment would predict higher future trading volume. To test this hypothesis, the following model is estimated where the sentiment factor in absolute values is also included in the equation (Tetlock, 2007):

$$Vol_t = \gamma_0 + \sum_{j=1}^{n=5} \gamma_j \cdot Rets_{t-j} + \sum_{j=1}^{n=5} \theta_j \cdot Sentiment_{t-j} + \sum_{j=1}^{n=5} \kappa_j \cdot |Sentiment_{t-j}| + \sum_{j=1}^{n=5} \eta_j \cdot Vol_{t-j} + \beta \cdot x_t + u_t$$

(4)

The coefficient estimates on sentiment measures are presented in Table 5. Consistent with the hypothesis derived from the model of Campbell et al. (1993), I observe that high absolute values of Facebook sentiment as measured by GNH and $Positivity$ both display the ability to predict higher trading volume whereas I find no significant effect of $Negativity$ factor. Particularly, the joint significance test statistics (reported in the last row) for lagged values of GNH and $Positivity$ are 28.00 (p-value<0.001) and 25.95 (p-value<0.001), which strongly indicate that Facebook sentiment is associated with higher future volume.

Interestingly, I also observe that Facebook sentiment displays some (weak) direct effect on trading volume in the full sample period whereas the effect completely disappears in the post-Lehman period. In his paper, Tetlock (2007) documents a direct role of the media sentiment measure in predicting the future trading volume. He argues that if the sentiment factor, pessimistic communication in the media, proxies for trading costs, a decline in negative sentiment predicts higher volume in the market.

Taken as a whole, estimation results indicate that Facebook sentiment appears to display ability to predict future stock market returns. Particularly, the impact of Facebook sentiment on daily returns is more pronounced in the face of market turmoil, and for those stocks, which are mostly held by noise traders. Finally, I provide evidence that high levels of Facebook sentiment, either positive or negative, have also forecasting power for future trading volume.

6 Conclusions

In this paper, I investigate the relationship between investor sentiment and stock market activity by using a new, direct and timely-fashioned measure of sentiment. The sentiment measure used in this study is constructed on a daily basis using content from the individual status updates of almost 100 million US American Facebook users for the time period between September 10, 2007 and September 9, 2011.

I perform a myriad tests of whether the sentiment measure compiled from Facebook correlates with various measures of stock market activity. The findings are in line with the predictions of investor sentiment theory. First, I show that Facebook sentiment displays ability to predict statistically significant and economically meaningful changes both in daily returns and trading volume in the US equity market. For instance, a one standard deviation increase in the Facebook sentiment, as measured by GNH, predicts an increase in returns equal to 11 basis points over the next day. Comparisons with other daily returns such as the mean daily return over the sample period or with other sentiment measures (e.g. Tetlock, 2007) imply that the effect of Facebook sentiment is economically highly strong as well. Moreover, I also document that the impact of Facebook sentiment is particularly stronger among small-cap stocks, and in the face of market turmoil. The latter finding provides direct evidence for the hypothesis that scarce arbitrage capital and higher risks associated with arbitrage in highly volatile periods seem to end up in more noise in stock prices. Finally, unusually positive or negative values of Facebook sentiment are also associated with higher future trading volume in the market, confirming the model predictions of Campbell et al. (1993). In short, the results presented in this paper support the role of investor sentiment as an important factor affecting the stock market activity.

In addition to studying the relation between investor sentiment and stock market activity, this paper also points out the usefulness of data from online social media in possible finance and economics applications. For instance, expanding the analysis to other countries, and exploring the interactions between Facebook sentiment and stock market activity seems to be promising that I am planning to explore in future research.

References

Amihud, Y. and H. Mendelson (1986). Asset pricing and the bid-ask spread. *Journal of Financial Economics 17*, 223–249.

Antweiler, W. and M. Z. Frank (2004). Is all that talk just noise? the information content of internet stock message boards. *Journal of Finance 59*, 1259–1293.

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

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

Barber, B. M. and T. Odean (2008). All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors. *Review of Financial Studies 21*, 785–818.

Barberis, N., A. Shleifer, and R. Vishny (1998). A model of investor sentiment. *Journal of Financial Economics 49*, 307–343.

Campbell, J., S. J. Grossman, and J. Wang (1993). Trading volume and serial correlation in stock returns. *Quarterly Journal of Economics 108(4)*, 905–939.

Cohen, L., A. Frazzini, and C. Malloy (2010). Sell side school ties. *Journal of Finance 65*, 1409–1437.

Cohen, R. M., M. Gross, T. E. Nordahl, W. E. Semple, D. A. Oren, and N. E. Rosenthal (1992). Preliminary data on the metabolic brain pattern of patients with winter seasonal affective disorder. *Archives of General Psychiatry 49*, 545–552.

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

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

Edmans, A., D. Garcia, and . Norli (2007). Sports sentiment and stock returns. *Journal of Finance 62*, 1967–1998.

Frazzinia, A. and O. A. Lamont (2008). Dumb money: Mutual fund flows and the cross-section of stock returns. *Journal of Financial Economics 88*, 299–322.

Hirshleifer, D. and T. Shumway (2003). Good day sunshine: stock returns and the weather. *Journal of Finance 58*, 1009–1032.

Howarth, E. and M. S. Hoffman (1984). A multidimensional approach to the relationship between mood and weather. *British Journal of Psychology 75*, 15–23.

Hu, X., J. Pan, and J. Wang (2010). Noise as information for illiquidity. Nber working paper, National Bureau of Economic Research.

Kamstra, M. J., L. A. Kramer, and M. D. Levi (2003). Winter blues: a sad stock market cycle. *American Economic Review 93*, 324–341.

Kramer, D. A. (2010). An unobtrusive behavioral model of gross national happiness. *In Proceedings of the ACM Conference on Human Factors in Computing Systems*, 287–290.

Kumar, A. and C. M. Lee (2006). Retail investor sentiment and return comovements. *Journal of Finance 61*, 2451–2486.

Lamont, O. A. and R. H. Thaler (2003). Can the market add and subtract? mispricing and tech stock carve-outs. *Journal of Political Economy 111*, 227–268.

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

Lintner, J. (1965). Security prices, risk and maximal gains from diversification. *Journal of Finance 20*, 587–616.

Lusardi, A. and O. Mitchell (2007). Financial literacy and retirement planning: New evidence from the rand american life panel. Working papers, University of Michigan, Michigan Retirement Research Center.

Neal, R. and S. M. Wheatley (1998). Do measures of investor sentiment predict returns? *Journal of Financial and Quantitative Analysis 33*, 523–545.

Pennebaker, J. W., C. K. Chung, M. Ireland, A. Gonzales, and R. J. Booth (2007). *The Development and Psychometric Properties of LIWC2007*. University of Texas at Austin and University of Auckland.

Saunders, Edward M., J. (1993). Stock prices and wall street weather. *American Economic Review 83*, 1337–1345.

Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *Journal of Finance 19*, 425–442.

Shiller, R. (2008). *The Subprime Solution. How Today's Global Financial Crisis Happened, and What to Do about It*. Princeton University Press.

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

Wann, D., T. Dolan, K. Mcgeorge, and J. Allison (1994). Relationships between spectator identification and spectators perceptions of influence, spectators emotions, and competition outcome. *Journal of Sport and Exercise Psychology 16*, 347–364.

Table 1: Facebook User Statistics

	Facebook Users	Share of Online Population (%)
United States	154,573,340	70.22%
Indonesia	40,139,480	100%
India	34,609,480	42.73%
Turkey	30,280,580	100%
United Kingdom	30,040,200	69.50%
Mexico	28,150,240	100%
Philippines	25,508,800	100%
Brazil	24,921,480	36.92%
France	22,599,080	55.31%
Germany	20,741,880	37,56%

Note: The table presents the number of Facebook users and their share in the online population in the top ten largest Facebook nations. The data come from www.checkfacebook.com, as of August 27, 2011.

Table 2: Demographics of the Facebook Users in the US

	Facebook Users	Facebook Users (in %)	US Population (in %)
Panel A: Age			
Age<14	835,480	0.50%	7.70%
13<Age≤24	38,007,200	24.70%	16.60%
23<Age≤34	36,319,060	23.60%	16.02%
33<Age≤44	25,518,060	16.60%	16.01%
43<Age≤54	18,847,640	12.20%	17.19%
53<Age≤64	11,967,940	7.70%	13.41%
Age≥65	7,253,200	4.70%	13.08%
Panel B: Gender			
Male	69,169,820	45.30%	49.40%
Female	83,452,100	54.70%	50.60%

Note: The table presents the number of Facebook users in the US. Panel A reports the age distribution of US American Facebook users and the entire US American population between age 9 and 85. Panel B reports the share and total number of Facebook users and the entire US American population grouped by gender. The data come from www.checkfacebook.com, as of August 27, 2011; and U.S. Census Bureau, Current Population Reports, as of December 15, 2010.

Table 3: Summary Statistics for the Sample

	Mean	25th pctile	Median	75th pctile	Standard Deviation	No of Obs.
Panel A:						
Gross National Happiness	-0.0167	-0.0300	-0.0155	-0.0072	0.0262	1,009
Positivity score	-0.0184	-0.0427	-0.0118	-0.0043	0.0269	1,009
Negativity score	-0.0017	-0.0091	-0.0032	0.0058	0.0103	1,009
Daily returns in logs	-0.0002	-0.0073	0.0008	0.0074	0.0159	1,009
Volatility (detrended)	0.00001	-0.0001	-0.0001	0.0000	0.0004	1,009
Volume (detrended)	0.0076	-0.1187	-0.0031	0.1443	0.2237	1,009
Panel B:						
Seasonal Affective Disorder	0.8694	0.0000	0.0000	1.8200	1.0962	1,009
Full Moon Dummy	0.2359	0.0000	0.0000	0.0000	0.4248	1,009
Preciption (deseasonalized)	-0.0051	-0.1315	-0.0745	-0.0225	0.2809	1,009
Temperature (deseasonalized)	-0.0028	-3.9800	-0.1900	3.6579	5.6374	1,009
Wind speed (deseasonalized)	-0.0103	-2.3300	-0.4300	2.0750	3.3631	1,009
Fall Dummy	0.2478	0.0000	0.0000	0.0000	0.4319	1,009
Tax Dummy	0.0178	0.0000	0.0000	0.0000	0.1324	1,009
Holiday Dummy	0.0347	0.0000	0.0000	0.0000	0.1831	1,009

Note: The table presents descriptive statistics for the variables employed in the analysis. The data come from Facebook, Datastream, National Climatic Data Center (NCDC), United States Naval Meteorology and Oceanography Command (NMOC).

Table 4: Predicting Daily Returns using Facebook Sentiment

Sentiment :	GNH		Positivity Factor		Negativity Factor	
	Full sample	Post-Lehman period	Full sample	Post-Lehman period	Full sample	Post-Lehman period
$Sentiment_{t-1}$	0.0011**	0.0017***	0.0012**	0.0016**	0.0001	-0.001
$Sentiment_{t-2}$	0.0001	0.0004	0.0001	0.0003	0.0015	0.003
$Sentiment_{t-3}$	0.0004	0.0011**	0.0003	0.0008	-0.0002	-0.0013
$Sentiment_{t-4}$	0.0003	0.0003	0.00	-0.0002	-0.0009	-0.0008
$Sentiment_{t-5}$	0.0004	0.0007*	0.0002	0.0005	0.00	0.0002
Environmental controls	Yes	Yes	Yes	Yes	Yes	Yes
Monthly fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Calendar controls	Yes	Yes	Yes	Yes	Yes	Yes
No of Obs	1,004	753	1,004	753	1,004	753
$\chi^2(5)[Sentiment]$	13.00**	12.25**	10.7*	9.3*	2.55	3.7

Note: The table reports the estimates of coefficients on Facebook sentiment. Each reported coefficient measures the impact of a one-standard deviation increase in the sentiment measures on daily returns. The regression is based on 1,004 daily observations from September 10, 2007 to September, 9 2011. The post-Lehman period refers to the 3-year time period over September, 16 2008 through September, 9 2011. Newey and West (1987) standard errors are used that are robust to heteroskedasticity and autocorrelation up to 5 lags. The data come from Facebook, Datastream, NCDC and NMOC. Three stars denote significance at 1 percent or less; two stars denote significance at 5 percent or less; one star denotes significance at 10 percent or less.

Table 5: How does Facebook sentiment affect the cross-section of stock returns? (Portfolios Formed on Size)

Sentiment :	GNH				Positivity factor			
	Full sample		Post-Lehman period		Full sample		Post-Lehman period	
	Small	*Big*	*Small*	*Big*	*Small*	*Big*	*Small*	*Big*
$Sentiment_{t-1}$	0.0012**	0.0011*	0.0019***	0.0017**	0.0014**	0.0013*	0.0018***	0.0016**
$Sentiment_{t-2}$	-0.0003	0.0001	0.0001	0.0004	-0.0004	0.0001	0.0000	0.0003
$Sentiment_{t-3}$	0.0009	0.0005	0.0016**	0.0011*	0.0008	0.0003	0.0014*	0.0008
$Sentiment_{t-4}$	0.0004	0.0002	0.0003	0.0002	0.0001	-0.0001	-0.0002	-0.0004
$Sentiment_{t-5}$	0.0003	0.0003	0.0007	0.0006	0.0001	0.0001	0.0003	0.0003
Environmental controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Monthly fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Calendar controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
No of Obs	998	998	747	747	998	998	747	747
$\chi^2(5)[Sentiment]$	12.9**	10.2*	14.7**	9.5*	10.8*	8.2	12.35**	6.45

Note: The table reports the estimates of coefficients on Facebook sentiment. Each reported coefficient measures the impact of a one-standard deviation increase in the sentiment measures on daily returns in each cross-section of stocks. The regression is based on 1,004 daily observations from September 10, 2007 to August, 31 2011. The post-Lehman period refers to the 3-year time period over September, 16 2008 through June, 30 2011. Newey and West (1987) standard errors are used that are robust to heteroskedasticity and autocorrelation up to 5 lags. The data come from Professor Kenneth French's website, Facebook, Datastream, NCDC and NMOC. Three stars denote significance at 1 percent or less; two stars denote significance at 5 percent or less; one star denotes significance at 10 percent or less.

Table 6: Predicting Daily Volume Using Facebook Sentiment

Sentiment :	GNH		Positivity factor		Negativity factor	
	Full sample	Post-Lehman period	Full sample	Post-Lehman period	Full sample	Post-Lehman period
$Sentiment_{t-1}$	-0.0212	-0.006	-0.019	-0.0089	-0.0007	-0.0093
$Sentiment_{t-2}$	0.02	0.0164	0.0294	0.0272	0.0031	-0.01
$Sentiment_{t-3}$	-0.0144	-0.0294**	-0.0031	-0.0159	0.0207	0.0317
$Sentiment_{t-4}$	0.0145	0.0125	0.0253	0.026	0.0158	0.0174
$Sentiment_{t-5}$	0.0236*	0.0176	0.0001	-0.0048	-0.0381	-0.0303
$|Sentiment_{t-1}|$	0.0477***	0.0335**	0.0542***	0.0388**	0.0011	0.0019
$|Sentiment_{t-2}|$	-0.0006	0.0066	-0.0073	-0.0031	0.0027	0.0004
$|Sentiment_{t-3}|$	0.0194*	0.0317***	0.0092	0.0179	-0.0027	0.001
$|Sentiment_{t-4}|$	-0.0125	-0.0103	-0.0215	-0.0227	0.0058	0.0045
$|Sentiment_{t-5}|$	-0.0224**	-0.0153	0.001	0.0063	0.0012	-0.0023
Environmental controls	Yes	Yes	Yes	Yes	Yes	Yes
Monthly fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Calendar controls	Yes	Yes	Yes	Yes	Yes	Yes
No of Obs	1,004	753	1,004	753	1,004	753
$\chi^2(5)[Sentiment]$	13.85**	7.7	20.05***	12.2**	8.65	10.35*
$\chi^2(5)[|Sentiment|]$	28.00***	18.65***	25.95***	16.4***	2.00	0.80

Note: The table reports the estimates of coefficients on Facebook sentiment. Each reported coefficient measures the impact of a one-standard deviation increase in the sentiment measures on daily volume. The regression is based on 1,004 daily observations from September 10, 2007 to September, 9 2011. The post-Lehman period refers to the 3-year time period over September, 16 2008 through September, 9 2011. Newey and West (1987) standard errors are used that are robust to heteroskedasticity and autocorrelation up to 5 lags. The data come from Facebook, Datastream, NCDC and NMOC. Three stars denote significance at 1 percent or less; two stars denote significance at 5 percent or less; one star denotes significance at 10 percent or less.

Figure 1: **Illustration of Individual Status Updates in Facebook**

Note: This figure illustrates an example of the homepage in Facebook. The question 'What's on your mind?' shows up in this page whenever the user logs on to Facebook. Further, the status updates of friends in the network as well as their other recent activities will be shown in this page.
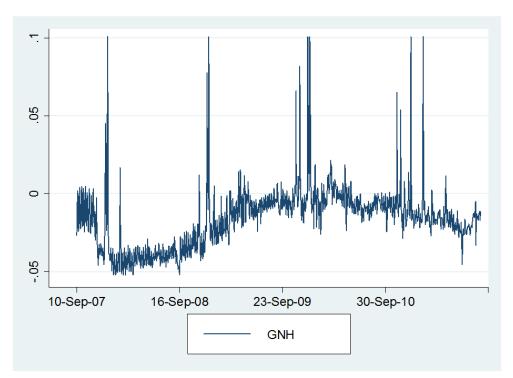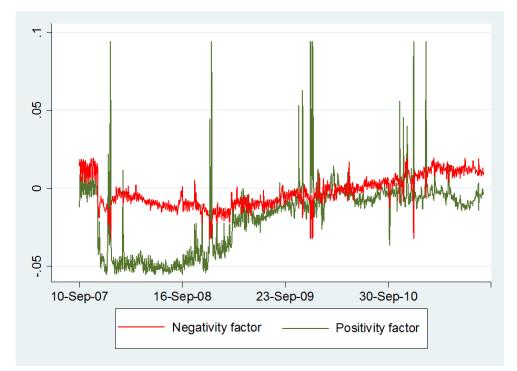
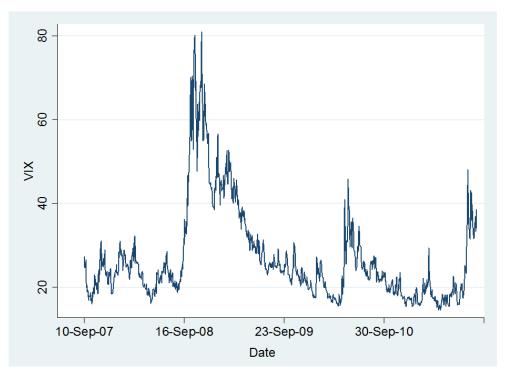Figure 2: **Facebook Sentiment Over the Sample Period (I)**

Note: This figure depicts the main Facebook sentiment measure over the observation period. Facebook sentiment is captured by Gross National Happiness (GNH). The data comes from Facebook.

29

Figure 3: **Facebook Sentiment Over the Sample Period (II)**

Note: This figure depicts the alternative Facebook sentiment measures over the observation period. The observation period is from September 10, 2007 to September, 9 2011. Facebook sentiment is captured by two different measures: Positivity, and Negativity factors. The data comes from Facebook.

Figure 4: **The Volatility Index (VIX) Over the Sample Period**

Note: This figure depicts the Volatility Index (VIX) of the CBOE over the observation period. The observation period is from September 10, 2007 to September, 9 2011. The data comes from the website of Chicago Board Options Exchange (CBOE).