Can Facebook Predict Stock Market Activity?*

Yigitcan Karabulut†

Goethe University Frankfurt

JOB MARKET PAPER

This Draft: October 21, 2013

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 in both daily returns and trading volume in the US stock market. For instance, an increase of one standard deviation in GNH is associated with an increase of 11.23 basis points in market returns over the next day. Consistent with noise trader models, the influence of GNH on market returns is temporary and is reversed during the following trading weeks. I also verify the empirical validity of GNH by performing several tests in different natural settings.

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

JEL Codes: D81, G11, G12.

---

*I would like to thank Pooyan Amir Ahmadi, Brad Barber, Kelley Bergsma, Pengjie Gao, Michalis Haliassos, Cam Harvey, David Horn, Byoung-Hyoun Hwang, Danling Jiang, Bige Kahraman, Orcun Kaya, Gabriele Lepori, Simone Manganelli, Sebastian Mueller, Hersh Shefrin, Paul Tetlock, Erik von Schedvin, and seminar participants at the 2013 AFA Meetings in San Diego, NBER Behavioral Finance Conference in Stanford, 2nd. Miami Behavioral Finance Conference in Miami, European Summer Symposium in Financial Markets in Gerzensee (informal presentation), 2nd. European Retail Investment Conference in Stuttgart, Bundesbank-CFS-ECB Workshop on Macro and Finance in Frankfurt, and Sveriges Riksbank for very helpful comments and discussions. The background work for this paper was carried out in the Summer of 2011 while I was visiting as a dissertation intern the Research Division at the Sveriges Riksbank; kind hospitality is gratefully acknowledged.

†Yigitcan Karabulut is with Goethe University Frankfurt, Department of Money and Macroeconomics, House of Finance, Gruneburgplatz 1, 60323, Frankfurt, Germany, E-mail: karabulut@econ.uni-frankfurt.de, Phone: +49 (0) 69 798 33859. The usual disclaimer applies.

1 Introduction

The potential role of investor sentiment in financial markets has received considerable attention from economists since John Maynard Keynes (1936) referred to 'animal spirits' in explaining stock market anomalies. The behavioral models of asset pricing pioneered by De Long et al. (1990) introduced the concept of 'irrational noise traders' in financial markets, formally demonstrating the relationship between noise trader sentiment and asset prices. A vast body of empirical literature also examines the effects of sentiment on the stock market and indicates that investor sentiment may persist in financial markets and influence stock prices (e.g., Lee, Shleifer, and Thaler, 1991; Baker and Wurgler, 2006; Tetlock, 2007). As Baker and Wurgler (2006) argue, 'Now, the question is no longer, as it was a few decades ago, whether investor sentiment affects stock prices, but rather how to measure investor sentiment and quantify its effects.'

In this paper, I propose a novel and direct measure of investor sentiment that is based on credible direct reports about the subjective well-being of millions of people. More specifically, I measure investor sentiment using Facebook's Gross National Happiness (GNH) measure. GNH is calculated using the textual analysis of emotion words posted by more than 160 million users on Facebook. Using vector autoregressive models, I then examine the relationship between GNH and daily stock market activity. First, I find that GNH can predict future stock market returns, regardless of whether controlling for past stock market volatility, daily economic conditions or the turn-of-the-year effect. This effect is statistically significant and economically meaningful. For example, an increase of one standard deviation in the GNH measure is associated with an increase of 11.23 basis points in the following day's returns. This increase is greater than the mean market returns during the sample period. Consistent with noise trader models, my findings indicate that the positive influence of GNH on market returns is temporary and is fully reversed during the following two trading weeks. Moreover, the GNH values retain most of their predictive ability even when I substitute a more conservative return window into my analysis (for instance, using open-to-close returns in lieu of close-to-close returns). Second, I find that GNH significantly predicts increases in future trading volume. This finding is consistent with the theory that unusually high or low levels of investor sentiment are associated with a high market trading volume (De Long et al., 1990; Campbell, Grossman, and Wang, 1993). These findings support the interpretation of the GNH measure as a reflection of investor sentiment.

By contrast, the alternative hypothesis that GNH conveys information about macroeconomic fundamentals (rather than investor sentiment) does not appear to be supported by the data. For example, the relationship between GNH and daily stock market activity remains almost identical even after I control for daily economic and business conditions in the forecasting regressions. Likewise, I examine the predictive content of GNH for near-term macroeconomic conditions and find no predictive ability for GNH with respect to macroeconomic measures.

To alleviate any concern that my results are due to data mining, I run a series of out-sample-tests on my findings. In particular, I extend my analysis to international markets and test whether the predictive ability of the GNH measure also holds internationally. I show that the GNH measure from the UK and Germany also predicts future stock market returns and trading volume in the respective countries, which indicates that GNH is also predictive of future stock market activity in an international sample.

Lastly, I perform a validation analysis in a natural setting to reinforce the proposition that GNH serves as a proxy for investor sentiment. Following Baker, Wurgler, and Yuan (2011), I examine the effects of differential GNH on the relative price deviations of dual-listed companies. I show that the relative price deviations of the twin companies are positively associated with the relative GNH values of the respective markets. The results remain identical when I control for non-synchronous trading and exchange rate fluctuations. This evidence supports the hypothesis that country-specific sentiment may partly explain disparities in twin company pricing (Froot and Dabora, 1999) and further strengthens the argument that GNH reflects investor sentiment.

The key contribution of this paper is that it proposes a novel and direct measure of investor sentiment with particularly attractive properties. First and foremost, GNH is based on the linguistic tone of

Facebook status updates, which may be interpreted as credible direct reports of subjective well-being (Kahneman and Krueger, 2006). Specifically, a status update is a short, self-descriptive message provided by the user in response to the question *'What's on your mind?'*. Therefore, status updates often have informational content regarding what people actually think or feel (Kramer, 2010). Thus, GNH more directly reflects investor sentiment (i.e., the animal spirits) than do the existing market or survey-based sentiment measures. Second, GNH is compiled from Facebook, which is currently the largest social network worldwide and has more than one billion active users. In the US alone, there are over 160 million Facebook users across all age groups (i.e., almost 50 percent of the entire US population is 'on Facebook'). Thus, GNH is likely to capture the sentiment of the US population overall. Finally, Facebook computes GNH on a daily basis, which allows me to track investor sentiment with a high level of frequency. In general, GNH appears to be a reasonable proxy for investor sentiment.

To the best of my knowledge, this paper is one of the first to utilize information from online social networking sites in finance.[1] Therefore, apart from testing theories of investor sentiment, this paper also highlights the usefulness of data from online social networks, which may constitute a rich source of information with many economics and finance applications.[2]

This paper is organized as follows. Section 2 provides background information and the motivation for studying the effects of investor sentiment on the stock market. Section 3 describes the variables used and provides further information on the data sources. Section 4 first provides information about the empirical methodology implemented and then presents the results of the main analysis and a set of robustness analyses. I present additional out-of-sample evidence in Section 5. Section 6 reports the findings of a validation analysis conducted using a natural setting. Finally, Section 7 concludes the paper.

[1]For example, in contemporaneous work, Bollen, Mao, and Zeng (2011) measure the collective mood of the US population using content from microblogging posts on Twitter (i.e., tweets). The results imply that changes in the public mood can be tracked based on the mood on Twitter. Furthermore, of the seven observed mood dimensions that these researchers have identified, only certain dimensions are associated with shifts in the Dow Jones Industrial Average.

[2]Data from online social networking sites can be particularly interesting when used for research focused on information transmission and social networks (e.g., Cohen, Frazzini, and Malloy, 2010).

2 Theory and Background

The classic theory of securities markets posits that market participants are fully rational and therefore that asset prices in equilibrium reflect rationally discounted and evaluated future cash flows and investment risks (e.g., Sharpe, 1964; Lintner, 1965). However, economic agents such as noise traders can demand assets for reasons that are not related to fundamentals. This is what Keynes (1936) referred to as 'animal spirits' or what Hume (1748) much earlier called 'motivating passions'. Accordingly, highly speculative episodes in stock markets such as the stock market boom and crash of 1929 or the Internet bubble period seem not to be fully justified by fundamentals.

To understand such wild fluctuations in financial markets, behavioral models explore alternatives to the premise of 'full rationality'. For example, De Long et al. (1990) present a model that formalizes the role of investors who are not fully rational (i.e., noise traders) in financial markets. In their model, there are two types of investors: rational arbitrageurs and noise traders. Rational arbitrageurs hold rational expectations about future asset returns, whereas noise traders are subject to exogenous sentiment and form expectations that are either overly optimistic or overly pessimistic relative to rational expectations. The equilibrium price reflects the expectations of both noise traders and rational investors because both types of investors are assumed to be risk averse and because assets are risky. Essentially, the model of De Long et al. (1990) demonstrates that asset prices may deviate from fundamental values because of irrational waves of optimism and pessimism if the demand across noise traders is correlated and there are limits to arbitrage.[3,4] In summary, the theory formally shows that under costly arbitrage, noise trader sentiment may persist in financial markets and affect asset prices in equilibrium.[5]

Following the seminal work of De Long et al. (1990), a significant amount of empirical literature has attempted to measure investor sentiment and study its effects on securities markets (e.g., Lee, Shleifer,

[3]The demand across noise traders can be correlated when investors act on similar 'pseudo-signals', such as forecasts by Wall Street gurus or prior price and volume patterns. Pseudo-signals are signals that are non-informative with regard to a firm's fundamental value (Kumar and Lee, 2006).

[4]De Long et al. (1990) argue that rational arbitrageurs may not be willing to bet against noise traders to correct mispricing because of limitations such as short investment periods and the costs and risks of trading and short selling.

[5]A related strand of the theoretical literature on investor sentiment focuses on how investors form biased beliefs. See, for example, Barberis, Shleifer, and Vishny (1998); Daniel, Hirshleifer, and Subrahmanyan (2001).

and Thaler, 1991; Kumar and Lee, 2006; Baker and Wurgler, 2006; Lemmon and Portniaguina, 2006; Qiu and Welch, 2006; Tetlock, 2007; Garcia, 2013; Da, Engelberg, and Gao, 2013). Traditionally, the sentiment measures employed in the literature are either indirect (market-based) measures or direct (survey-based) measures. Lee, Shleifer, and Thaler (1991) provide an example of the former; they use closed-end fund discounts (CEFD) as a proxy for investor sentiment and show that the CEFD measure is an index of individual investors' optimism (pessimism) relative to the broader market sentiment. Lee, Shleifer, and Thaler (1991) demonstrate that decreases in CEFD are positively associated with the returns of stocks that are predominantly held by retail investors who are also assumed to be noise traders.[6] In another paper, Neal and Wheatley (1998) investigate the forecasting power of different market-based sentiment measures, including net mutual fund redemption, the ratio of odd-lot sales to purchases and CEFD. Their results indicate that two of these three indirect sentiment measures, CEFD and net redemptions, significantly predict the size premium. More recently, Baker and Wurgler (2006) examine how investor sentiment affects the cross-section of stock returns. They construct a measure of investor sentiment based on six commonly used market-based sentiment measures: value-weighted dividend premiums, the number of IPOs, the average first-day IPO returns, value-weighted CEFD, the equity share in new issues and NYSE turnover. Baker and Wurgler (2006) indicate that smaller stocks, high-volatility stocks, unprofitable stocks, non-dividend-paying stocks, extreme-growth stocks and distressed stocks earn higher returns following periods of low sentiment and display lower returns when sentiment is high.[7]

Unlike these studies, which rely on indirect proxies for sentiment, other papers focus on direct survey-based measures. For instance, Lemmon and Portniaguina (2006) and Qiu and Welch (2006) pro-

[6] In another paper, Qiu and Welch (2006) re-examine CEFD as a measure of investor sentiment over a longer time period. Specifically, they analyze the time-series relationship between CEFD and the UBS/Gallup Index for Investor Optimism, which is a direct survey-based sentiment measure. Qiu and Welch (2006) find that CEFD have no correlation with the UBS/Gallup measure. The authors conclude that CEFD may not be a good proxy for investor sentiment.

[7] Another strand of the empirical literature attempts to link stock market returns to fluctuations in human emotions, which are identified by exogenous mood indicators (Hirshleifer and Shumway, 2003; Kamstra, Kramer, and Levi, 2003). For example, Edmans, Garcia, and Norli (2007) construct a sports sentiment measure using the results of international soccer games. The authors find a loss effect on stock market returns that is especially pronounced among small-cap stocks and after important games.

vide evidence that the Michigan Index of Consumer Sentiment is strongly correlated with the returns of small-cap stocks and the returns of firms held predominantly by individual investors, which is consistent with the sentiment theory. In the same vein, using survey data from Investor's Intelligence, Brown and Cliff (2005) analyze the long-run effects of sentiment on stock prices and find that investor sentiment is related to market mispricing.

My paper takes an approach that is similar to the latter because the sentiment variable that I propose, GNH, directly captures investor sentiment. Relative to existing direct sentiment measures, GNH has several advantages. First, GNH is computed on a daily basis. Survey-based measures are generally available at a lower frequency (e.g., monthly or quarterly). Second, GNH is based on a textual analysis of emotion words posted by millions of people on Facebook. In other words, GNH is based on what people 'actually' think or feel. By contrast, survey-based measures are subject to some criticism because people may respond to surveys in a manner that belies what they actually think (Baker and Wurgler, 2007). Thus, GNH brings us closer to the source of investor sentiment (i.e., the animal spirits) than do the existing market or survey-based sentiment measures. However, the GNH measure also has potential pitfalls. For instance, survey-based measures such as the Michigan Index of Consumer Sentiment have longer time series than the GNH measure. Overall, it appears that GNH qualifies as a reasonable proxy for investor sentiment. In the following section, I provide a detailed description of the GNH measure.

3 Data and Variable Definitions

3.1 Gross National Happiness

In this paper, I propose a novel and direct measure of investor sentiment. The sentiment variable, GNH, is compiled from Facebook, a social networking site that allows its users to create online profiles and select a network of friends who can view and post on each other's profiles. Facebook is currently the largest online social network; it boasts more than one billion active users worldwide.[8] As shown in

[8]More than 50 percent of all users log on to Facebook on any given day. The average Facebook user has 130 friends in her network and spends approximately 31.1 minutes a day on Facebook, which yields an aggregate total of 700 billion minutes

Table 1, there are more than 160 million Facebook users in the US alone. This figure equals 50 percent of the entire population and 70 percent of the online population in the US. Similarly, during the period spanning from December 2004 to April 2012, the number of Facebook users correlates strongly with the Google's Search Volume Index (SVI) for 'Facebook', and both series start to radically increase at the beginning of 2008.[9] Figure 1 shows this relationship graphically.

Facebook determines the GNH measure via the textual analysis of content from status updates.[10] Figure 2 illustrates examples of status updates on Facebook. A status update is a self-descriptive text that contains information provided by the user in response to the question '*What's on your mind?*'. Therefore, status messages often contain informational content about what people think or feel (Kramer, 2010). To some extent, status updates on Facebook resemble the 'Experience Sampling Method' (ESM), which was developed to obtain information about people's feelings in natural settings (Csikzentmihalyi, 1990; Kahneman and Krueger, 2006). However, compared to existing ESMs, Facebook's status updates allow information to be collected about the subjective well-being of a larger sample of individuals over a relatively longer time period.[11] Moreover, status updates are often not directed to a specific target audience, unlike wall posts on Facebook or Twitter messages, neither of which necessarily have any informational content about the subjective well-being of individuals. Given this background, of all of the text used in online social networks, Facebook's status updates appear to be the most appropriate type of text to use in constructing a sentiment measure.[12]

GNH is computed using the 'word-count' methodology explained in Pennebaker et al. (2007). In this approach, different sets of words are defined as having different psychological meanings: in this case, as connoting positive and negative emotions.[13] Facebook measures a status update's positivity

per month. User statistics are obtained from the Facebook web site. For further information, please see the Facebook factsheet that is available on the website.

[9]For a detailed description of Google's SVI, see, for example, Da, Engelberg, and Gao (2011).

[10]There is a budding strand of literature that uses textual analysis for different finance applications. See, for example, Tetlock (2007); Tetlock, Saar-Tsechansky, and Mackassy (2008); Tetlock (2010); Garcia (2013).

[11]The existing studies (e.g., Diener and Suh, 1999) indicate that self-reported well-being measures are positively correlated with visible signs of happiness such as smiling, sleep quality and others that can be visualized using neuro-imaging.

[12]GNH was first introduced and developed by Adam D.I. Kramer, Lisa Zhang and Ravi Grover from the Facebook Data Team. I would like to thank the developers of the *Gross National Happiness* measure and the *Facebook Data Team* for making data available for analysis.

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

(negativity) according to the relative frequency with which positive (negative) emotion words are used in each individual update. For example, a status update that states, 'I had a good day' has a positivity score of 0.20 and a negativity score of 0 because the only emotion word in this post is 'good' (which is also positive) and the rest is neutral. Applying this procedure to all of its status updates, Facebook calculates aggregated positivity and negativity scores each day.[14] GNH is then computed as the normalized difference between these two affective factors as follows:

$$GNH_t = \frac{\mu_t^p - \mu^p}{\sigma^p} - \frac{\mu_t^n - \mu^n}{\sigma^n} \qquad (1)$$

where μ_t^p and μ_t^n represent the daily relative frequency of positive and negative word use in status updates by Facebook users, respectively. μ^p (μ^n) and σ^p (σ^n) are the mean and standard deviation of the daily frequency of positive (negative) word use across the sample period.

To alleviate any concerns about outliers and to address the issues of slow-moving time trends and seasonality, I make the following adjustments to the Facebook measures. First, to mitigate the concerns of any outliers, all Facebook variables (i.e., GNH, Positivity and Negativity) are winsorized at the 0.5% upper and lower tails of their distributions. Then, to eliminate possible seasonality, I regress these variables on weekday dummies and month dummies and calculate the residuals from the regression estimates. Finally, to address the slow-moving time trends, I subtract the past one-month moving average (i.e., 20 trading days) of the Facebook variables from the daily observations.[15]

One possible concern about the GNH measure is the degree to which it represents the US population. This factor is noteworthy because it is generally believed that Facebook is used predominantly by younger people and that the older population may therefore be underrepresented on Facebook. However,

[14] Positivity and negativity scores can be interpreted as optimism and pessimism factors, respectively. In the empirical analysis, I employ each of these affective dimensions as possible proxies for investor sentiment and separately quantify their effects on the stock market.

[15] In unreported analysis, I find that the happiest workday of the week is Friday and the happiest month of the year is December, suggesting that there is seasonality in the GNH series. Therefore, I seasonally adjust the GNH variable. The reason for detrending the GNH measure is the following. Because the user base of the Facebook social network grew over time, the composition of Facebook users may have changed over time. For example, 'happier' people with better outside opportunities could have increasingly joined Facebook over time, generating a spurious positive trend in happiness. To address the possibility of slow-moving time trends in GNH, I detrend this variable using a moving average detrending procedure.

this concern is not fully justified. Of the 164 million US Facebook users, only 8% are younger than 18; the shares of users who are 25 to 44 years old and older than 45 are 25% and 27%, respectively. As in the U.S. population, 54.7% of Facebook users are female and 45.3% are male. Table 2 provides detailed information on the demographics of US Facebook users.

I collect a time series of daily GNH measure for the time period beginning January 1, 2008 through April 27, 2012 from Facebook. The sample period includes $1,090$ trading days after weekends and national holidays are excluded.[16] During the sample period, the mean and median of GNH are negative, as reported in Table 3. Notably, the raw value of GNH falls to its lowest value on September 16, 2008, the day when Lehman Brothers filed for bankruptcy protection and Merrill Lynch agreed to sell to Bank of America to avert bankruptcy. This was one of the most dramatic days in Wall Street history. Thus, there is an apparent link between GNH and the stock market.

3.2 Other Data

I collect daily returns and trading volume data for the US from Thomson Reuters Datastream.[17] The observation period spans from January 1, 2008 to April 27, 2012 and includes 1,090 trading days after weekends and holidays are excluded.

As a proxy for stock market volatility, I use the Chicago Board Options Exchange (CBOE) Market Volatility Index (VIX), which measures the implied volatility of the options on the Standard and Poor's 100 stock index (Baker and Wurgler, 2007). The daily series of the VIX are obtained from the CBOE's online data library.[18]

I use the business and economic conditions index developed by Aruoba, Diebold, and Scotti (2009) (ADS) to control for general economic activity in my return regressions. The ADS index is designed

[16] Facebook has publicized the daily data on the GNH measure for the time period beginning from September 9, 2007. In the analysis, I consider the time period beginning from January 1, 2008 for two reasons. First, Facebook became the largest online social network in the US in terms of both unique monthly visitors and number of users in 2008. Second, starting the sample period in 2008 also makes it possible to eliminate the possible concern that Facebook users were 'younger' at this time (i.e., that most were college students).

[17] I use the Thomson Reuters Datastream's mnemonic 'TOTMKUS' to obtain the time-series data for the return index ('RI') and the trading volume ('VO') for the US stock market.

[18] The time-series data for VIX are available at http://www.cboe.com/micro/vix/historical.aspx.

to track macroeconomic conditions with high frequency in the US. More specifically, Aruoba, Diebold, and Scotti (2009) use a wide variety of macroeconomic indicators at different frequencies to construct their real-time economic conditions index. The underlying indicators are (seasonally adjusted) weekly initial jobless claims, monthly payroll employment, industrial production, personal income less transfer payments, manufacturing and trade sales and quarterly real GDP. The average value of the ADS index is zero, and values larger (smaller) than zero indicate better (worse) than average economic conditions. I obtain the time series of the ADS index data from the web site of the Federal Reserve Bank of Philadelphia.[19]

As additional controls, I employ several environmental measures that have been used in the literature as mood proxies. First, I include average daily temperature (in Fahrenheit), precipitation (in mm) and wind speed in my analysis to proxy for weather-induced mood. My inclusion of these variables is motivated by evidence from the psychology literature that nearly 40 percent of mood variation can be explained by the weather (Persinger and Levesque, 1983).[20] Following Saunders (1993), I collect weather data for New York City from the National Climatic Data Center database. To remove pure seasonal variation, I also adjust the environmental variables based on the methodology by Hirshleifer and Shumway (2003).[21] Second, following Kamstra, Kramer, and Levi (2003), I control for Seasonal Affective Disorder (SAD) in the return regressions. Kamstra, Kramer, and Levi (2003) use seasonal variations in daylight as mood indicators and investigate the effects of these variations on stock market returns. To compute the SAD variable, I collect data on the daily darkness duration in New York City from the database of the United States Naval Meteorology and Oceanography Command.

[19]The daily time-series data of the ADS index are available from the *Real-Time Data Research Center* of the Federal Reserve Bank of Philadelphia. See Aruoba, Diebold, and Scotti (2009) for a detailed description of the ADS index.

[20]For example, existing studies show that higher temperatures and more hours of sunshine are associated with higher levels of optimism and lower levels of depression and skepticism (Howarth and Hoffman, 1984).

[21]More specifically, I calculate the average values for daily temperature, precipitation and wind speed for each calendar week and then deduct the mean value from the daily observations.

4 Gross National Happiness and Stock Market Activity

I use a vector autoregressive (VAR) framework to examine the relationship between GNH and daily stock market activity. The VAR model enables me to simultaneously estimate the bidirectional causal relationship between stock market measures and the GNH measure. The model takes the following form:

$$z_t = \alpha + \sum_{j=1}^{5} \beta_j \cdot z_{t-j} + \theta \cdot x_t + \epsilon_t \tag{2}$$

where z_t denotes the following endogenous variables in the system: $Facebook_t$, $Rets_t$ and Vol_t. $Facebook_t$ represents the detrended and seasonally adjusted index score for GNH or one of its affective dimensions; $Rets_t$ and Vol_t are the log returns and detrended log trading volume on day t, respectively.[22] All VAR estimates include all lags of up to 5 days (i.e., one week of calendar time for the stock market) prior to market activity. The choice of lag length is based on the Akaike (1974) information criteria.[23] Finally, x_t represents the exogenous control variables in the model.

The exogenous variables include VIX up to five lags to control for past volatility;[24] the five lags of the ADS index proxy for general economic and business conditions, and several calendar controls are used, including monthly dummies (Thaler, 1987), a dummy variable for the trading day after a holiday (Ariel, 1990) and weekday dummies (Keim and Stambaugh, 1984). Following Ivashina and Scharfstein (2010), I also control for the bankruptcy of Lehman Brothers and the associated turbulence in the financial markets by including a dummy variable for the fourth quarter of 2008 (i.e., the peak period

[22] Because log trading volume in levels is not stationary, I detrend it using first-order differencing. It is also important to note that I obtain qualitatively similar results when I use alternative detrending procedures such moving average detrending as in Campbell, Grossman, and Wang (1993) and Tetlock (2007). I report the regression estimates for this alternative specification in Table A.1 in the Appendix.

[23] Based on the Schwarz (1978) criterion, the optimal lag length is three. In an unreported analysis, I re-estimate the return predictability regressions using all lags of the endogenous variables up to three days. I find that GNH is still a positive and significant predictor of future stock market returns. For example, an increase of one standard deviation in GNH is associated with an increase of 10.79 basis points in daily returns over the next day (*t-statistics=2.25; p-value<0.05*).

[24] Using alternative volatility measures such as the detrended squared residuals, as in Tetlock (2007), or the innovations in seasonal-adjusted VIX computed from the *ARFIMA(0,d, q)* model, as in Da, Engelberg, and Gao (2013), does not affect the results of the forecasting regressions. The regression estimates of these alternative specifications are reported in Table A.2 in the Appendix.

of the financial crisis) in the model. Furthermore, to account for the turn-of-the-year effect, I define an indicator variable that is set to one in the last week of December and in the first week of January and zero otherwise (Ritter, 1988; Bergsma and Jiang, 2013). In a recent paper, Bergsma and Jiang (2013) also show that the stock market returns are significantly higher in the time window spanning from five days before through five days after the New Year's Day in ten countries where the New Year holidays are not on January 1. The authors relate the price run up in the stock markets to the positive swings of investor mood during this time period.[25] Similarly, GNH is also high on December 31, most likely because Facebook users use widely positive emotion words in their status updates when using traditional phrases such as 'Happy New Year'. To ensure that the results are not driven by a limited number of extreme observations, I indicate these days using binary variables. I also account for environmentally induced mood fluctuations by using a set of variables that I introduced in the previous section. Finally, it is important to note that all stock market measures and control variables are winsorized at the 0.5% upper and lower tails of their distributions.

The error terms in Equation 2 are assumed to be independent of the lagged values of the endogenous variables in the system, which allows me to estimate each equation separately using ordinary least squares. I also 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.

I first examine the relationship between daily market returns and the GNH measure. To test whether GNH forecasts future daily returns, I estimate the following model:

$$Rets_t = \alpha + \sum_{i=1}^{5} \gamma_i \cdot Rets_{t-i} + \sum_{j=1}^{5} \beta_j \cdot Facebook_{t-j} + \sum_{k=1}^{5} \eta_k \cdot Vola_{t-k} + \theta \cdot x_t + u_t \qquad (3)$$

[25] In the days surrounding New Year's Day (i.e., five trading days before and after January 1), the mean log daily returns in my sample accounts for 9.77 basis points, whereas the average market return in the non-new-year-period is 0.22 basis points. Similarly, the mean value of the GNH measure is highly positive and equal to 0.009 around New Year's Day, whereas it is equal to −0.002 in the remaining period. In Table A.3, I examine the estimates of return predictability regressions after omitting the days surrounding New Year's Day. After excluding the turn-of-the-year effect, none of the Facebook variables has a statistically significant impact on future stock market returns at conventional levels. Still, the first lag of the GNH measure is positively associated with future trading volume, but the statistical significance is slightly smaller (*t-statistics=1.98*; *p-value<0.05*.)

The forecasting regressions are similar to those used by Tetlock (2007) and Garcia (2013) to predict future stock market activity using media news sentiment. Throughout the paper, I focus on the estimates of coefficients on the Facebook variable (i.e., β_j), which describe the dependence of the stock market measures on Facebook sentiment.[26]

Table 4 presents the estimates of the coefficients of the Facebook measure from the return predictability regressions. Each reported coefficient represents the effect of a one-standard-deviation increase in the GNH measure on daily returns in basis points. As reported in columns (i) to (iii) of Table 4, GNH has a statistically and economically significant effect on the next day's market returns, regardless of whether controlling for past volatility, past daily economic conditions or the turn-of-the-year effect. For instance, an increase of one standard deviation in GNH is associated with an increase of 11.23 basis points (*t-statistics=2.28*; *p-value<0.05*) in daily returns over the next day. Such increases are economically significant.[27]

To demonstrate the economic significance of the relationship between GNH and stock market returns, I compare the effect of GNH with other effects and standard daily returns. For example, the daily average market return during the sample period is 0.6 basis points, which would be completely offset by a one-standard-deviation change in GNH. Similarly, Tetlock (2007) reports that an increase of one standard deviation in his news media sentiment measure predicts a decrease in Dow Jones returns that is equal to 8.1 basis points, which is smaller on an absolute scale than the GNH effect.[28] In general, comparisons with other factors that influence daily returns suggest that GNH appears to have some economically meaningful forecasting power with regard to market returns.

Next, I separately investigate the forecasting power of each affective dimension underlying the GNH

[26] All estimates are available upon request.

[27] In Table A.4, I also examine the relationship between daily stock market activity and the raw GNH values, that is, I do not carry out any winsorization, seasonality or detrending on the GNH measure. Similar to the results reported in Table 4 and Table 7, the raw values of the GNH measure significantly predict future stock market returns and trading volume. For example, a one-standard-deviation change in the GNH measure is associated with an increase of 12.10 basis points in market returns over the next day (*t-statistics=2.54*; *p-value<0.05*). Similarly, the first lag of GNH is a significant positive predictor of future trading volume (*t-statistics=3.54*; *p-value<0.01*.)

[28] In a recent paper, Garcia (2013) constructs a news media sentiment measure by counting the positive and negative words from two financial columns from the New York Times. He reports that a one-standard-deviation change in his media news sentiment measure is associated with a change of 12 basis points in future daily returns during recessions, whereas the effect during expansion periods accounts for 3.5 basis points.

measure. I estimate the initial model as expressed in Equation (3); all independent variables in these regressions are identical except that GNH has been substituted for by either Positivity or Negativity. The first three columns of Table 5 show that Positivity has a statistically and economically significant effect, whereas I find no indication that Negativity can significantly predict future stock returns, as shown in columns (iv) to (vi) of Table 5. The effect of Positivity is slightly greater than (but not significantly different from) that of GNH, which has an effect of 11.46 basis points on the next day's returns (*t-statistics=2.38*; *p-value<0.05*).

The standard models of investor sentiment posit that sentiment-induced mispricing will be corrected over the longer horizon. In other words, asset prices exhibit return reversals. It is unclear, however, what the appropriate time horizon is for examining reversals (Han, 2008). In the reported return predictability regressions, I cannot reject the null hypothesis of no reversal within one week. Although there is some indication of return reversal (particularly, in the second and fourth lag of GNH), the sum of the coefficient estimates on all four lags (i.e., $t-2$ through $t-5$) is statistically not different from zero (χ^2-test*=0.78; p-value>0.1*).

To examine whether there is a reversion to fundamentals over the longer horizon, I next investigate the relationship between GNH and stock market returns for up to three weeks. The lag length in the weekly analysis is determined using the Akaike (1974) and Schwarz (1978) criteria. To reduce the 'noise' in the analysis, I use weekly data rather than daily data (Ben-Rephaela, Kandela, and Wohla, 2012).[29] The sample now includes 226 weekly observations.

I start with the contemporaneous relationship between GNH and weekly market returns. Consistent with sentiment theory, GNH is positively contemporaneously associated with weekly returns, with a correlation coefficient of 0.118, but the statistical power is low (*p-value=0.07*). In examining the relationship between GNH and subsequent market returns, I still observe that past GNH values exert a statistically and economically significant influence on future stock returns. In column (i) of Table 6, I

[29] In the weekly analysis, I use Wednesday-to-Wednesday changes in trading volume and returns to avoid potential seasonal effects and thin trading problems.

show that a change of one standard deviation in GNH corresponds to an increase of approximately 34.31 basis points in the market returns during the next week (*t-statistics=2.44; p-value<0.05*). However, this effect is only temporary; it reverses during the following two trading weeks. The sum of the coefficient estimates for lag 2 to lag 3 is -62.29 and is statistically significant ($\chi^2 - test$*=4.62; p-value<0.05*).[30] Therefore, I can reject the null hypothesis that the effect of GNH on market returns is permanent, which is consistent with noise trader models that show that short-term returns are reversed in the *long run*. In column (ii) and (iii) of Table 6, I further show that Positivity forecasts a temporary increase and reversal as predicted by the GNH measure, whereas the effect of the Negativity factor is marginally significant and the reversal in lags 2 through 3 is statistically not different from zero.

Next, I consider the effect of GNH on future trading volume, providing a measure of stock market activity different from market returns. The analysis of the relationship between trading volume and GNH is motivated both by recent empirical evidence provided by Tetlock (2007) and by the theoretical model of Campbell, Grossman, and Wang (1993).

In their model, Campbell, Grossman, and Wang (1993) demonstrate that unusually high or low values of investor sentiment will generate higher trading volume in financial markets. Specifically, when noise traders experience a negative (positive) belief shock, they will sell (buy) securities. To restore market equilibrium, market makers will absorb rising demand (supply) from noise traders, which will result in higher trading volume (Tetlock, 2007).

To examine whether GNH is associated with future trading volume, I estimate the following model, in which I include the absolute values of the Facebook measures for up to five lags in the regression:

$$Vol_t = \gamma_0+\sum_{i=1}^{5}\gamma_i\cdot Rets_{t-j}+\sum_{j=1}^{5}\beta_{1j}\cdot Facebook_{t-j}+\sum_{k=1}^{5}\beta_{2k}\cdot|Facebook_{t-k}|+\sum_{l=1}^{5}\eta_l\cdot Vol_{t-l}+\theta\cdot x_t+u_t \tag{4}$$

[30] It is important to note that the sum of the coefficients on the three lags is not significantly different from zero ($\chi^2 - test$*=0.88; p-value>0.1*). Thus, I cannot reject the hypothesis that the reversal in lags 2 through 3 exactly offsets the initial increase in market returns (Tetlock, 2007).

The coefficient estimates for both GNH and the absolute values of GNH are presented in Table 7. Each coefficient in the table represents how an increase of one standard deviation in the GNH measure affects daily trading volume. Again, I estimate the effect of Facebook using all three Facebook measures.[31]

Consistent with the model predictions of Campbell, Grossman, and Wang (1993), I find that GNH robustly forecasts increases in trading volume. The first lag of the absolute value of GNH is a significantly positive predictor of future trading volume (*t-statistics=3.62*; *p-value<0.01*). This effect is also robust to controlling for past daily economic conditions or the turn-of-the-year effect, as shown in columns (ii) and (iii) of Table 7.

Next, I examine the forecasting power of each affective dimension underlying the GNH measure on future trading volume. As reported in Table 8, unusually high or low values of Positivity also predict high market trading volume, whereas Negativity factor displays no significant effect on future trading volume.[32] Specifically, the first lag of Positivity significantly predicts increases in the trading volume on the next day (*t-statistics=3.16*; *p-value<0.01*). In summary, the results of the volume regressions provide direct support for the findings of Tetlock (2007), who indicates that the absolute values of sentiment correspond to higher trading volumes in the Dow Jones.

I also find some evidence for the hypothesis that GNH is directly associated with future trading volume. The first lag of GNH is negatively associated with the future trading volume (*t-statistics*=-*2.45*; *p-value<0.05*). In his paper, Tetlock (2007) also finds that his measure of media news sentiment has a direct negative effect on future trading volume, which he attributes to the trading costs argument (Antweiler and Frank, 2004).

Finally, in Table 9, I examine the relationship between Facebook measures and subsequent weekly

[31]To assess whether the results of the volume regressions are robust to alternative model specifications, I use the squared values of GNH (instead of absolute values of GNH) to capture the unusually high and low values of investor sentiment. The coefficient estimates are reported in Table A.5. As reported in the table, the first lag of the squared GNH is a significantly positive predictor of future trading volume. This suggests that the forecasting power of GNH for future trading volume is not driven by any model misspecification.

[32]In fact, the third lag of the Negativity factor is positively associated with the future trading volume and is statistically significant at the 5% level.

trading volume in the stock market. I observe that past GNH values exert a statistically and economically significant influence on future trading volume. In particular, the first lag of the absolute value of GNH is positively associated with the future trading volume (*t-statistics=2.60*; *p-value<0.01*). I also observe that the initial increase in trading volume is followed by a reversal in the following two trading weeks, although the sentiment theory makes no clear predictions about the relationship between sentiment and subsequent trading volume. The sum of the coefficient estimates for lag 2 to lag 3 is -0.083 and is statistically highly significant ($\chi^2 - test$*=9.14; p-value<0.01*).

In summary, the results presented imply that the GNH measure has the ability to predict changes in both daily and weekly returns and trading volume for the US stock market.

4.1 Timing Issues

Thus far, I have measured market returns using close-to-close prices and have investigated the ability of the GNH measure to forecast future market returns. However, one might argue that GNH may convey after-hours information that may not be fully incorporated into closing prices because Facebook computes the GNH measure every day at 12 p.m. ET.[33] For example, any news arriving after the close of the market will be incorporated into the contemporaneous GNH but will be reflected in the next day's market returns. This may, of course, cast doubts on the daily return predictability regressions presented here. In fact, the results of the weekly predictability regressions, as reported in Table 6, suggest that timing issues cannot fully explain the predictive power of GNH for future market returns, as the exact intraday timing of GNH matters less for weekly predictability. Still, an essential issue is to analyze whether the forecast power of the GNH measure is concentrated in after-hours trading (close-to-open) or is allocated equally throughout the trading day (Tetlock, 2007).

I use a more conservative time window to calculate daily returns and address after-hours information. More specifically, I use daily open-to-close returns as a dependent variable, which allows me to eliminate

[33] A recent study analyzed approximately 1.6 million posts and 7.5 million comments mode on Facebook during the period spanning from August 10, 2007 to October 10, 2010. This study shows that Facebook use in terms of number of status updates made and comments posted reaches its peak at 3 p.m. ET. The other largest spikes occur at 11 a.m. and 8 p.m. ET (Vitrue, 2011).

the close-to-open return component. To compute open-to-close returns, I use price information on three different Exchange Traded Funds (ETFs), which replicate the three major stock market indices in the US. I use the Dow Jones Total Market ETF (TMW) for the Dow Jones US Total Stock Market Index, the NYSE Composite Index ETF (NYC) for the NYSE Composite Index and the S&P 500 ETF (SPY) for the S&P 500 Index. The data on opening and closing prices come from Thomson Reuters Datastream.

As in the initial return predictability regression, all of the independent variables are identical except that close-to-close returns have been replaced by open-to-close returns. Table 10 summarizes the coefficient estimates for the GNH measure. Again, each coefficient in the table measures the effect of a one-standard-deviation change in GNH on open-to-close returns in basis points. Table 10 shows that the GNH values retain most of their predictive ability even when I substitute a more conservative return window in my analysis. The magnitudes of the next day's coefficients range from 6.28 (*TMW; t-statistics= 1.76; p-value<0.1*) to 5.69 (*NYC; t-statistics= 1.75; p-value<0.1*) basis points and retain most of their original economic significance.[34]

I also perform similar tests using each affective dimension underlying the GNH measure. For brevity, I only present the coefficient estimates for the Positivity factor because I find that the Negativity factor has no predictive power either for the close-to-close return or for the open-to-close returns.[35] I observe that Positivity has a statistically and economically significant influence on the next day's returns even when I use open-to-close returns as the dependent variable. For example, the magnitudes of the next day's coefficients range from 6.66 (*TMW; t-statistics= 1.98; p-value<0.5*) to 6.05 (*SPY; t-statistics= 1.71; p-value<0.1*) basis points.

In summary, despite the slight decline in the effect of GNH on market returns, the regression results imply that the predictive power of GNH is dispersed throughout the entire trading day rather than being concentrated in after-hours trading.[36]

[34]After substituting the open-to-close returns for close-to-close returns, I observe that GNH does not significantly predict one-day ahead returns of the S&P 500 ETF. The magnitude of the next day's coefficient diminishes from 10.40 to 5.27 basis points.

[35]For brevity, the results of these regressions are presented in Table A.6 in the Appendix.

[36]I also perform a similar test to investigate whether the results of the daily volume regressions are robust to after-hours information. In particular, I exclude the prior day's GNH (i.e., $t-1$) from the model and include GNH and the absolute

4.2 Interpreting the Results: Sentiment versus Information

Up to this point in the discussion, I have interpreted GNH as a measure of investor sentiment. The results provide evidence that is consistent with this interpretation. First, I find that GNH predicts returns for short horizons and that it predicts a reversion to fundamentals over the *long run*. Alternatively, one might argue that the GNH measure may reflect information about economic fundamentals that is not yet incorporated into stock prices, i.e., an information hypothesis. Under this hypothesis, GNH would still forecast short-term returns but no return reversal would occur. Thus, the market response to GNH would be permanent. The evidence that indicates the initial increase is followed by a reversal is therefore consistent with the sentiment hypothesis but inconsistent with the information hypothesis regarding GNH. I can also rule out the alternative hypothesis that GNH serves a proxy for stale information, which has been already incorporated into stock prices (Tetlock, 2010). The stale information theory posits that GNH should have no forecasting power with regard to future stock market activity (Tetlock, 2007), whereas I show that GNH forecasts future stock market returns irrespective of whether after-hours information is controlled for. Second, GNH significantly predicts high market trading volume. This finding is consistent with the predictions of the models presented in De Long et al. (1990) and Campbell, Grossman, and Wang (1993), which suggest that unusually high or low values for investor sentiment are associated with increases in future trading volume.

To further investigate whether GNH is a proxy for investor sentiment rather than for information about the economy, I examine the relationship between GNH and the stock market while controlling for general macroeconomic activity. More specifically, I use the ADS index to proxy for daily economic activity because it is designed to track the economic and business conditions in the US (Aruoba, Diebold, and Scotti, 2009). If the GNH measure truly conveys information about macroeconomic fundamentals, then it is reasonable to expect that its predictive power will decrease (or even disappear) when daily

value of GNH from lag 2 through lag 6 in the model. Table A.7 reports the results of this robustness exercise. I find that the past absolute GNH values still predict increases in future trading volume. The loading on the second lag of GNH is positive (0.019), however, the statistical power of this effect is low (*t-statistics=1.65; p-value<0.1*). This result implies that the predictive content of the GNH measure with regard to future trading volume is not driven by after-hours information.

economic conditions are controlled for in the regressions. I re-estimate the predictability regressions in (3) and (4) and include the ADS index up to five lags as additional controls in the model. As shown in columns (ii) and (iii) of Table 4, GNH has forecasting power for future returns even after I control for daily economic and business conditions. The magnitude of the next day's coefficient changes slightly relative to the initial estimates in column (i) of Table 4 but retains most of its original economic and statistical significance (10.88 versus 11.44). I perform a similar test to analyze the robustness of the effect of GNH on trading volume. As reported in columns (ii) and (iii) of Table 7, the absolute GNH values retain their ability to predict increases in future volume even after the ADS index is controlled for in the volume regressions. In particular, the first lag of the absolute value of GNH remains a significantly positive predictor of future trading volume, which strengthens the hypothesis that GNH serves as a proxy for investor sentiment.

In an additional test of the information hypothesis, I next investigate the predictive ability of the GNH measure with regard to future macroeconomic activity. If GNH is a proxy for new information about the economy, then it should forecast near-term macroeconomic conditions. For example, Tetlock, Saar-Tsechansky, and Mackassy (2008) show that the fraction of negative words in firm-specific news stories, which they interpret as an information proxy, can predict a firm's fundamentals, such as earnings. In a similar vein, I analyze the relationship between GNH and (seasonally adjusted) weekly initial jobless claims to test the predictive power of GNH with regard to future macroeconomic activity. I use initial jobless claims as a proxy for macroeconomic conditions for two reasons. First, weekly data on initial claims are available. By contrast, for other economic measures such as GDP and industrial production, only quarterly or monthly data are usually available. Second, as noted by Choi and Varian (2009), weekly initial claims have been documented to be good indicators of economic conditions in the US.[37] One potential challenge in this analysis is that of determining the appropriate time lag because the theory makes no clear predictions about it. Therefore, I use different lag structures (i.e., one and three lags)

[37]The data on seasonally adjusted weekly initial jobless claims come from the United States Department of Labor at http://workforcesecurity.doleta.gov/unemploy/claims.asp. The sample period extends from January 1, 2008 to April 7, 2012.

in the regressions. Table 12 presents the coefficients from the time-series regressions of the weekly initial jobless claims as dependent variables on the lagged values and the lagged GNH values.[38] As shown in Table 12, I find no evidence that past GNH values (or the affective dimensions underlying the GNH measure) are significantly associated with future initial jobless claims, which implies that GNH does not predict near-term macroeconomic conditions, which remains true when I use different lag structures in the estimation. In summary, the hypothesis that GNH serves as a proxy for information about macroeconomic fundamentals does not appear to be supported by the data.[39]

5 Out-of-Sample Tests: International Evidence

To further investigate the robustness of my results, I also conduct a series of out-of-sample tests. In particular, I extend my analysis to international markets (i.e., United Kingdom and Germany) and test whether the predictive ability of the GNH measure also holds internationally.

I obtain the daily series of the GNH measure for the UK and Germany from January 1, 2008 to April 27, 2012 from Facebook. As in the US regressions, I also use winsorized, seasonally adjusted and detrended values of the GNH measure in the out-of-sample tests. The daily returns and trading volume data for the UK and Germany come from Thomson Reuters Datastream.[40] To account for environmentally induced mood fluctuations, I also collect data on weather and daily darkness duration for London and Frankfurt, the cities where the stock exchange is located in the UK and Germany, respectively. The data on environmental variables come from the National Climatic Data Center database

[38] I use the weekly log changes in initial jobless claims as dependent variables in the regressions because the log weekly initial jobless claims in levels are not stationary.

[39] I also examine the relationship between GNH and the Michigan Index of Consumer Sentiment. The contemporaneous correlation between GNH and (log) consumer confidence in levels is negative at -0.059, but the relationship is statistically insignificant (*p-value=0.67*). For the relationship between GNH and log changes in consumer sentiment, the correlation increases to -0.29 and is statistically significant at the 5% level (*p-value=0.038*). When I study the forecasting ability of GNH with respect to changes in consumer confidence, I find no significant effect of GNH when examining future survey results for sentiment and consumer confidence. The regression results are reported in Table A.8. This finding remains the same when I use different lag structures (i.e., one, two or four lags). However, these results must be interpreted with caution because of the sample size and the limited statistical power of these tests.

[40] I use Thomson Reuters Datastream's mnemonic 'TOTMKUK' and 'TOTMKBD' to obtain the time-series data for the return index and the trading volume for the UK and German stock market, respectively.

and the database of the United States Naval Meteorology and Oceanography Command.[41]

First, I consider the effects of the GNH measure on stock market returns of the respective countries. I separately estimate the initial model as expressed in Equation (3) for the UK and Germany. Unlike in the initial model, I now measure the stock market volatility using the detrended squared residuals, as in Tetlock (2007). The results of the return predictability regressions are reported in columns (i) and (iii) of Table 13. Each coefficient in the table represents how an increase of one standard deviation in GNH affects daily market returns in basis points. As reported in the table, GNH has a statistically significant influence on the next day's market returns both in the UK and Germany. For example, an increase of one standard deviation in GNH is associated with an increase of 11.85 basis points in the UK (*t-statistics= 2.78; p-value<0.01*) and 13.96 basis points in Germany (*t-statistics= 2.21; p-value<0.05*). These effects are also economically significant, given that the unconditional mean log return in the UK and Germany over the sample period accounts for 0.13 basis points and -1.43 basis points, respectively.

In columns (ii) and (iv) of Table 13, I examine the relationship between the GNH measure and daily trading volume. Again, I use the model as presented in Equation (4), in which I include the absolute values of GNH for up to five lags in the model. I find that GNH robustly forecasts increases in trading volume both in the UK and Germany. Similarly to the US regressions, the first lag of the absolute value of the GNH measure is positively associated with future trading volume.

Overall, these out-of-sample tests confirm my key findings that the GNH measure has the ability to predict changes in daily returns and trading volume, suggesting that these results are not due to data mining.

[41] It should be noted that I do not control for daily precipitation in the international tests because the database of the United States Naval Meteorology and Oceanography Command does not provide any information on this weather variable for either London or Frankfurt.

6 Validation Exercise: Evidence from Dual-Listed Companies

In this section, I examine the effects of the GNH measure on the relative price deviations of dual-listed companies (DLCs). In fact, DLCs (often referred to as 'Siamese twins') provide a natural setting in which to test the validity of GNH as a sentiment measure. To begin this exercise, I will first provide a brief background on DLCs.

DLCs involve two companies that operate in different countries, where the shares of the twin companies are traded on the two countries' respective stock exchanges. The twin companies contractually act as a single entity and divide their cash flows among their shareholders using a fixed ratio (Baker, Wurgler, and Yuan, 2011). In theory, the stocks of these companies should be traded with reference to fixed price parity because the shares of the twin companies are claims to the same underlying cash flows. In practice, however, the relative prices of DLCs deviate considerably from the theoretical price ratio, which violates the law of one price (Froot and Dabora, 1999; DeJong, Rosenthal, and VanDijk, 2009).

Several hypotheses have been formulated to explain why twin price disparities may occur. For example, Froot and Dabora (1999) examine potential rational explanations such as discretionary use of dividend income by the parent company, currency fluctuations and tax-induced investor heterogeneity. These authors argue that tax-induced investor heterogeneity has the potential to explain some but not all of the facts. Accordingly, Froot and Dabora (1999) conclude that country-specific demand from noise traders may be one of the reasons for twin price disparities.

In a recent study, Baker, Wurgler, and Yuan (2011) empirically examine the relationship between differential investor sentiment and price gaps for DLCs. They find that the relative prices of Siamese twins (in this case, Shell/Royal Dutch) are positively associated with the relative investor sentiment of the respective markets, which supports the notion that differential noise trader sentiment may partly explain why these price gaps occur. Given this background, I analyze whether differences in country-specific sentiment (as measured using GNH) are associated with twin price disparities, which also allows me to validate GNH as a sentiment measure.

To address this issue, I obtain price data for four actively traded DLCs from Thomson Reuters Datastream. The twin companies in my sample are two Anglo-Australian twins, Rio Tinto and BHP Billiton, and two Anglo-Dutch twins, Unilever and Reed/Elsevier International. I collect daily GNH data for Australia, the UK and the Netherlands from Facebook.

The twin companies in my sample have different structures: Rio Tinto, BHP Billiton and Unilever are structured as separate entities, whereas Reed/Elsevier International is structured as a combined entity. In addition, the theoretical price ratio for Rio Tinto, BHP Billiton and Unilever equals 1, whereas the fixed-price parity for Reed/Elsevier International is 1.538.

Following DeJong, Rosenthal, and VanDijk (2009), I calculate the relative price deviations of the DLCs from their theoretical parity as follows:

$$Deviation_{i,t} = \frac{ln(Price_{i,A,t})}{ln(Price_{i,B,t})} - ln(\text{Theoretical Parity}_{i,A,B}) \tag{5}$$

where A and B represent the twin pairs, $Price_t$ is the price of the twin share denoted in a common currency (British Pounds) on day t and Theoretical Parity is the theoretical price parity of the twin pairs.

Figure 3 illustrates the log price deviations of the twin companies from the theoretical price parity over the sample period. The twin price disparities are large and fluctuate substantially over time. For example, the (pooled) mean relative price deviation in the sample amounts to 7.99$\%$ with a standard deviation of 10.1$\%$. Moreover, the Pearson correlation coefficients for the price gaps for the Anglo-Australian and Anglo-Dutch twins account for 0.78 (*p-value<0.01*) and 0.15 (*p-value<0.01*), respectively. Thus, there appears to be country-specific common factors (e.g., local investor sentiment) that determine the relative co-movement of twin companies' stock prices in the same country.

To test the hypothesis that twin companies' price gaps are related to time-varying differential investor sentiment, I separately estimate the following model for each twin pair:

$$Deviation_{i,t} = \alpha + \sum_{j=-1}^{1} \beta_j \cdot (GNH_{A,t+j} - GNH_{B,t+k}) + \sum_{k=-1}^{1} \theta_l \cdot ER_{l+k} + \epsilon_t \tag{6}$$

where $Deviation_{i,t}$ represents the log price deviation of twin pair i on day t and $GNH_{A,t}$ and $GNH_{B,t}$ are the winsorized, detrended and seasonally adjusted GNH for countries A and B, respectively. ER_t represents the daily log changes in the currencies of countries A and B. The daily exchange rate data come from Thomson Reuters Datastream.

All regressions include one lag, one contemporaneous and one lead coefficient for all independent variables to account for non-synchronous trading. I also control for possible heteroskedasticity and autocorrelation in the residuals by using Newey-West robust standard errors.

Panel A of Table 14 presents the estimation results. For the sake of brevity, I report only the sum of the coefficient estimates for each variable. The significance tests are the *Chi-squared tests* on the sum of the lead, current and lag coefficients for each variable. The test statistics are reported in parentheses.[42]

Table 14 shows that the GNH spread is positively related to the log price deviations of the Anglo-Dutch twins, which is consistent with the findings of Baker, Wurgler, and Yuan (2011). This effect is also statistically highly significant for the Anglo-Dutch twin pairs but not for the Anglo-Australian twins. For instance, when the Dutch GNH increases relative to the UK GNH, the price of Unilever NV also increases relative to that of Unilever PLC. This finding implies that differential investor sentiment as measured by GNH may partly explain the variations in twin pricing. The coefficient estimates are economically meaningful as well. For example, an increase of one standard deviation in the relative Facebook sentiment in the Netherlands and the UK is associated with an increase of 48 basis points in the price ratio of Reed/Elsevier International and 33 basis points in the price ratio of Unilever NV and Unilever PLC. In Panel B of Table 14, I separately examine the effects of each country's GNH on the log price deviations of the twin pairs. I find that GNH_A (GNH_B) is positively (negatively) associated with the relative prices of DLCs, which is consistent with the previous results.

[42] In the first two regressions, A and B denote the Netherlands and the UK, whereas the pair of countries is Australia and the UK in the latter two regressions.

In summary, the results of the exercise presented here provide additional support for the proposition that GNH is a proxy for investor sentiment.[43]

7 Conclusions

In this paper, I propose a new measure of investor sentiment that is based on the linguistic tone of the status updates posted by millions of people on Facebook. I examine the relationship between this measure of sentiment, GNH, and daily stock market activity using vector autoregressive models. First and foremost, I show that GNH has the ability to predict statistically significant and economically meaningful changes in aggregate market returns. Consistent with noise trader models, the positive influence of GNH on market returns is only temporary and completely reverses during the following trading weeks. The evidence of an initial increase and a subsequent return reversal supports the proposition that GNH serves as a proxy for investor sentiment. Comparisons with other daily returns also demonstrate the economic significance of the relationship between GNH and market returns. Second, I find that GNH robustly predicts increases in future trading volume. This result is consistent with the models presented in De Long et al. (1990) and Campbell, Grossman, and Wang (1993).

In addition, I perform several tests to scrutinize the hypothesis that GNH reflects investor sentiment. First, I show that the relationship between GNH and daily stock market activity remains identical even after I control for the daily macroeconomic conditions in the forecasting regressions. Second, I examine the predictive content of GNH for near-term macroeconomic conditions and find that GNH has no predictive ability with regard to macroeconomic measures. In summary, the alternative interpretation that GNH serves as a proxy for new information about macroeconomic fundamentals does not appear

[43] In a further robustness test, I consider the effects of differential GNH on the changes in the premiums of four country-closed end funds (CEFs). As Bodurtha, Kim, and Lee (1995) argue, 'under the investor sentiment hypothesis, the premium of a country fund captures the differential sentiment between the U.S. and foreign markets.' To address this issue, I use a multifactor pricing model in which weekly changes in premiums are regressed on a set of global risk factors and country-specific risk. The estimation results are reported in Table A.9 in the Appendix. The regression results imply that the weekly premium changes of CCEFs are positively associated with the US GNH but are negatively related to the GNH of the foreign market, although the relationship is not statistically significant in two of the four cases. In summary, the results of this exercise yet again support the empirical validity of GNH as a measure of investor sentiment.

to be supported by the data. I also present additional out-of-sample evidence that the predictive ability of GNH also holds internationally. Lastly, I conduct a validation exercise using data from dual-listed companies to reinforce the interpretation of GNH as a reflection of investor sentiment. More specifically, I examine the effects of differential GNH on the relative price deviations of twin companies from different countries and find that these price deviations are positively associated with the relative GNH of their respective markets. This result provides additional evidence that GNH is a valid measure of investor sentiment.

To the best of my knowledge, this paper is one of the first to utilize information from online social networking sites in finance. Therefore, in addition to testing the theories of investor sentiment, this paper also highlights the usefulness of data from online social media in finance and economics contexts. Additional applications of this nature should be considered in future research.

References

Akaike, H. (1974). A new look at the statistical model identification. *Automatic Control, IEEE Transactions on 19*(6), 716–723.

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.

Ariel, R. A. (1990). High stock returns before holidays: existence and evidence on possible causes. *Journal of Finance 45*(5), 1611–1626.

Aruoba, B., F. X. Diebold, and C. Scotti (2009). Real-time measurement of business conditions. *Journal of Business & Economic Statistics 27*, 417–427.

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.

Baker, M., J. Wurgler, and Y. Yuan (2011). Global, local, and contagious investor sentiment. *Journal of Financial Economics*.

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

Ben-Rephaela, A., S. Kandela, and A. Wohla (2012). Measuring investor sentiment with mutual fund flows. *Journal of Financial Economics 104*, 363–382.

Bergsma, K. and D. Jiang (2013). Let's celebrate! cultural new year and stock returns around the world. Working paper, Florida State University.

Bodurtha, J. N., D. S. Kim, and C. Lee (1995). Closed-end country funds and u.s. market sentiment. *Review of Financial Studies Vol. 8(3)*, pp. 879–918.

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

Brown, W. G. and M. T. Cliff (2005). Investor sentiment and asset valuation. *Journal of Business 78(2)*, 405–440.

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.

Choi, H. and H. R. Varian (2009). Predicting initial claims for unemployment benefits. Technical report, Google.

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

Csikzentmihalyi, M. (1990). *Flow: The Psychology of Optimal Experience*. Harper Collins (New York).

Da, Z., J. Engelberg, and P. Gao (2011). In search of attention. *Journal of Finance 66*(5), 1461–1499.

Da, Z., J. Engelberg, and P. Gao (2013). The sum of all fears: Investor sentiment and asset prices. Technical report, Working Paper.

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.

DeJong, A., L. Rosenthal, and M. A. VanDijk (2009). The risk and return of arbitrage in dual-listed companies. *Review of Finance 13*, 495–520.

Diener, E. and E. M. Suh (1999). National differences in subjective well-being. In D. Kahnemann, E.Diener, and N.Schwarz (Eds.), *Well-Being: The Foundations of Hedonic Psychology*, pp. 434–450. Russell-Sage (New York).

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

Froot, K. A. and E. M. Dabora (1999). How are stock prices affected by the location of trade? *Journal of Financial Economics 53*, 189–216.

Garcia, D. (2013). Sentiment during recessions. *Journal of Finance 68*(3), 1267–1300.

Han, B. (2008). Investor sentiment and option prices. *Review of Financial Studies 21*, 387–414.

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.

Hume, D. (1748). *An Enquiry concerning Human Understanding*. Oxford University Press.

Ivashina, V. and D. Scharfstein (2010). Bank lending during the financial crisis of 2008. *Journal of Financial Economics 97*(3), 319–338.

Kahneman, D. and A. B. Krueger (2006). Developments in the measurement of subjective well-being. *Journal of Economic Perspectives 20(1)*, 3–24.

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

Keim, D. B. and R. F. Stambaugh (1984). A further investigation of the weekend effect in stock returns. *Journal of Finance 39*(3), 819–835.

Keynes, J. M. (1936). *The General Theory of Employment, Interest, and Money*. MacMillan, London.

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.

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

Lemmon, M. and E. Portniaguina (2006). Consumer confidence and asset prices: Some empirical evidence. *Review of Financial Studies 19*, 1499–1529.

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

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

Newey, W. and K. West (1987). A simple, positive definite, heteroscedastic and autrocorrelation consistent covariance maxtrix. *Econometrica 55*, 703–708.

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.

Persinger, M. and B. F. Levesque (1983). Geophysical variables and behavior: Xii: The weather matrix accommodates large portions of variance of measured daily mood. *Perceptual and Motor Skills 57*, 868–870.

Qiu, L. and I. Welch (2006). Investor sentiment measures. Unpublished working paper, Brown University.

Ritter, J. R. (1988). The buying and selling behavior of individual investors at the turn of the year. *Journal of Finance 43*(3), 701–717.

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

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics 6*(2), 461–464.

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

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

Tetlock, P. (2010). All the news that's fit to reprint: Do investors react to stale information? *Review of Financial Studies 23*, 3250–3557.

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

Thaler, R. H. (1987). Amomalies: The january effect. *Journal of Economic Perspectives 1*(1), 197–201.

Table 1: Number of Facebook Users

	Total Number of Users	Share of Online Population (%)
United States	164,695,640	68.84%
Brazil	57,817,940	76.13%
India	54,755,360	67.60%
Indonesia	44,234,940	100.00%
Mexico	38,015,880	100.00%
United Kingdom	31,963,740	62.14%
Turkey	31,250,160	89.29%
Philippines	29,424,360	99.07%
France	24,914,940	55.83%
Germany	24,472,260	37.58%

Note: The table presents the number of Facebook users and their corresponding shares in the online population in the top ten largest Facebook nations, as of September 2012. The data come from www.checkfacebook.com, as of September 28, 2012.

Table 2: Demographics of the US Facebook Users

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

Note: The table presents information on the demographics of US Facebook users. Panel A reports the age distribution of both US Facebook users and the entire US population. Panel B presents the Facebook users and the entire US population grouped by gender. The data on age and gender distribution of Facebook users come from www.socialbakers.com, up to date as of September 28, 2012. The data on the age and gender decomposition of the US population are obtained from the US Census Bureau, Current Population Reports, up to date 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: Facebook & stock market measures						
GNH *(raw)*	-0.0016	-0.0135	-0.0006	0.0090	0.0203	1,090
Positivity *(raw)*	-0.0014	-0.0225	0.0051	0.0121	0.0226	1,090
Negativity *(raw)*	0.0001	-0.0083	-0.0016	0.0103	0.0103	1,090
GNH *(detrended, deseasonalized)*	-0.0020	-0.0070	-0.0019	0.0021	0.0170	1,090
Positivity *(detrended, deseasonalized)*	-0.0014	-0.0056	-0.0016	0.0021	0.0144	1,090
Negativity *(detrended, deseasonalized)*	0.0005	-0.0011	0.0004	0.0023	0.0040	1,090
Daily log returns	0.0001	-0.0070	0.0010	0.0075	0.0174	1,090
VIX	26.9899	19.3500	23.5450	30.6900	11.1521	1,090
Daily log volume *(detrended)*	0.0004	-0.1011	-0.0062	0.1022	0.1823	1,090
ADS Index	-0.7482	-1.2503	-0.2288	0.0435	1.2024	1,090
Panel B: Control variables						
SAD variable	0.8906	0.0000	0.0000	1.8400	1.0939	1,090
Daily temperature *(deseasonalized)*	0.0963	-4.0893	0.1118	3.9964	6.1553	1,090
Daily wind speed *(deseasonalized)*	0.0272	-2.3571	-0.2712	2.2371	3.4675	1,090
Daily precipitation *(deseasonalized)*	-0.0109	-0.1483	-0.0920	-0.0271	0.3009	1,090
New Year Dummy	0.0385	0.0000	0.0000	0.0000	0.1926	1,090
Crisis Dummy	0.0688	0.0000	0.0000	0.0000	0.2532	1,090
Holiday Dummy	0.0358	0.0000	0.0000	0.0000	0.1858	1,090
Monday Dummy	0.1862	0.0000	0.0000	0.0000	0.3895	1,090
Tuesday Dummy	0.2064	0.0000	0.0000	0.0000	0.4049	1,090
Wednesday Dummy	0.2073	0.0000	0.0000	0.0000	0.4056	1,090
Thursday Dummy	0.2018	0.0000	0.0000	0.0000	0.4016	1,090
Friday Dummy	0.1982	0.0000	0.0000	0.0000	0.3988	1,090

Note: The table presents descriptive statistics for the variables used in the econometric analysis. I adjust the Facebook variables (i.e., GNH, Positivity and Negativity) in the following manner. First, to mitigate concerns over any outliers, all Facebook variables are winsorized at the 0.5% upper and lower tails of their distributions. Then, to eliminate the possible seasonality, I regress these variables on weekday dummies and month dummies and calculate the residuals from the regression estimates. Finally, to address the slow-moving time trends, I subtract the past one-month moving average of the Facebook variables from the daily observations. Daily trading volume is detrended using first-order differencing. The weather variables are detrended using the methodology described in Hirshleifer and Shumway (2003). All stock market measures and control variables are also winsorized at the 0.5% upper and lower tails of their distributions. The table data come from Facebook, Thomson Reuters Datastream, FED Philadelphia, Chicago Board Options Exchange (CBOE), National Climatic Data Center (NCDC), and United States Naval Meteorology and Oceanography Command (NMOC).

Table 4: Predicting Daily Market Returns Using GNH (I)

	(i)		*(ii)*		*(iii)*	
	β	*t-stat*	β	*t-stat*	β	*t-stat*
GNH_{t-1}	10.8751**	2.12	11.4364**	2.24	11.2316**	2.28
GNH_{t-2}	-0.302	-0.06	-0.4177	-0.09	-0.4128	-0.09
GNH_{t-3}	7.383**	2.06	6.855*	1.94	6.4788*	1.7
GNH_{t-4}	-0.7513	-0.17	-0.4531	-0.11	-0.2476	-0.05
GNH_{t-5}	2.5981	0.76	2.2801	0.64	2.3146	0.64
Past volatility	*Yes*		*Yes*		*Yes*	
Environmental controls	*Yes*		*Yes*		*Yes*	
Calendar controls	*Yes*		*Yes*		*Yes*	
Daily economic activity	*No*		*Yes*		*Yes*	
Turn-of-the-year effect	*Yes*		*Yes*		*Yes*	
New Year's Day Dummies	*No*		*No*		*Yes*	
$R-squared$	0.0644		0.0753		0.0773	
No of Obs	1,090		1,090		1,090	
	χ^2-test	*p-value*	χ^2-test	*p-value*	χ^2-test	*p-value*
$\beta_{t-1}=0$	4.51	0.0339	5.00	0.0256	5.21	0.0226
$\sum_{j=1}^{5}\beta_{t-j}=0$	3.74	0.0535	3.5	0.0615	3.48	0.0623
$\sum_{j=2}^{5}\beta_{t-j}=0$	1.04	0.3082	0.81	0.3671	0.78	0.3783

Note: The table reports the estimates of the coefficients on the GNH measure. Each reported coefficient measures the impact of a one-standard-deviation increase in the GNH measure on daily returns in basis points. Daily economic activity is measured by the ADS index (Aruoba et al., 2009). I use VIX up to five lags to control for past volatility. The regressions are based on 1,090 daily observations from January 1, 2008 to April, 27 2012. Newey and West (1987) standard errors that are robust to heteroskedasticity and autocorrelation for up to five lags are used. All stock market measures and control variables are also winsorized at the 0.5% upper and lower tails of their distributions. The table data come from Facebook, Thomson Reuters Datastream, CBOE, FED Philadelphia, NCDC and NMOC. Three stars denote significance at 1% or less; two stars denote significance at 5% or less; one star denotes significance at 10% or less.

Table 5: Predicting Daily Market Returns Using GNH (II)

	Positivity						*Negativity*					
	(i)		*(ii)*		*(iii)*		*(iv)*		*(v)*		*(vi)*	
	β	*t-stat*	β	*t-stat*	β	*t-stat*	β	*t-stat*	β	*t-stat*	β	*t-stat*
$Facebook_{t-1}$	10.8912**	2.18	11.7056**	2.35	11.464**	2.38	-6.0186	-0.96	-5.6509	-0.89	-5.7267	-0.91
$Facebook_{t-2}$	0.9765	0.21	0.8824	0.2	0.9531	0.21	10.3472	1.33	9.9461	1.3	10.1205	1.32
$Facebook_{t-3}$	6.5993*	1.94	6.3376*	1.91	6.0472*	1.7	-6.7241	-1.21	-5.4873	-1.01	-4.8727	-0.85
$Facebook_{t-4}$	-2.9431	-0.71	-2.4238	-0.61	-2.4611	-0.53	-6.9586	-1.18	-6.5588	-1.12	-7.3653	-1.21
$Facebook_{t-5}$	2.512	0.76	2.2329	0.66	2.2663	0.67	1.4022	0.28	1.4759	0.29	1.3401	0.26
Past volatility	*Yes*		*Yes*		*Yes*		*Yes*		*Yes*		*Yes*	
Environmental controls	*Yes*		*Yes*		*Yes*		*Yes*		*Yes*		*Yes*	
Calendar controls	*Yes*		*Yes*		*Yes*		*Yes*		*Yes*		*Yes*	
Daily economic activity	*No*		*Yes*		*Yes*		*No*		*Yes*		*Yes*	
Turn-of-the-year effect	*Yes*		*Yes*		*Yes*		*Yes*		*Yes*		*Yes*	
New Year's Day Dummies	*No*		*No*		*Yes*		*No*		*No*		*Yes*	
$R-squared$	0.0648		0.0758		0.0779		0.0643		0.0744		0.0768	
No of Obs	1,090		1,090		1,090		1,090		1,090		1,090	
	χ^2-test	*p-value*	χ^2-test	*p-value*	χ^2-test	*p-value*	χ^2-test	*p-value*	χ^2-test	*p-value*	χ^2-test	*p-value*
$\beta_{t-1}=0$	4.76	0.0294	5.52	0.019	5.67	0.0174	0.92	0.3382	0.8	0.3719	0.82	0.3654
$\sum_{j=1}^{5}\beta_{t-j}=0$	3.67	0.0557	3.96	0.0469	3.84	0.0502	0.98	0.3217	0.49	0.485	0.53	0.4678
$\sum_{j=2}^{5}\beta_{t-j}=0$	0.79	0.3729	0.76	0.3821	0.70	0.4025	0.05	0.822	0.01	0.9488	0.01	0.9360

Note: The table reports the coefficient estimates for the Positivity (Negativity) measure. Each reported coefficient measures the impact of a one-standard-deviation increase in the Positivity (Negativity) measure on daily returns in basis points. Daily economic activity is measured by the ADS index (Aruoba et al., 2009). I use VIX up to 5 lags to control for past volatility. The regressions are based on 1,090 daily observations from January 1, 2008 to April, 27 2012. Newey and West (1987) standard errors that are robust to heteroskedasticity and autocorrelation for up to five lags are used. All stock market measures and control variables are also winsorized at the 0.5% upper and lower tails of their distributions. The table data come from Facebook, Thomson Reuters Datastream, CBOE, FED Philadelphia, NCDC and NMOC. Three stars denote significance at 1% or less; two stars denote significance at 5% or less; one star denotes significance at 10% or less.

Table 6: Predicting Weekly Market Returns Using GNH

	GNH		$Positivity$		$Negativity$	
	(i)		*(ii)*		*(iii)*	
	β	*t-stat*	β	*t-stat*	β	*t-stat*
$Facebook_{t-1}$	34.3124**	2.44	27.3565**	2.02	-38.0276*	-1.67
$Facebook_{t-2}$	-27.0258	-1.00	-35.2618	-1.47	-25.9713	-1.19
$Facebook_{t-3}$	-35.2691***	-2.74	-27.8283**	-2.31	40.2291*	1.68
Past volatility	*Yes*		*Yes*		*Yes*	
Environmental controls	*Yes*		*Yes*		*Yes*	
Calendar controls	*Yes*		*Yes*		*Yes*	
Weekly economic activity	*Yes*		*Yes*		*Yes*	
Turn-of-the-year effect	*Yes*		*Yes*		*Yes*	
$R-squared$	0.1733		0.1717		0.1943	
No of Obs	226		226		226	
	χ^2-test	*p-value*	χ^2-test	*p-value*	χ^2-test	*p-value*
$\beta_{t-1}=0$	5.93	0.0158	4.08	0.0449	2.8	0.096
$\sum_{j=1}^{3}\beta_{t-j}=0$	0.88	0.3496	1.65	0.2002	0.45	0.5053
$\sum_{j=2}^{3}\beta_{t-j}=0$	4.62	0.0329	5.46	0.0204	0.24	0.6236

Note: The table reports the estimates of the coefficients on the GNH measure or one of its affective dimensions. Each reported coefficient measures the impact of a one-standard-deviation increase in GNH (Positivity or Negativity) on weekly returns in basis points. Weekly economic activity is measured by the ADS index (Aruoba et al., 2009). I use VIX up to five lags to control for past volatility. The regressions are based on 226 weekly observations from January 1, 2008 to April, 27 2012. Newey and West (1987) standard errors that are robust to heteroskedasticity and autocorrelation for up to three lags are used. All stock market measures and control variables are winsorized at the 0.5% upper and lower tails of their distributions. The table data come from Facebook, Thomson Reuters Datastream, CBOE, FED Philadelphia, NCDC and NMOC. Three stars denote significance at 1% or less; two stars denote significance at 5% or less; one star denotes significance at 10% or less.

Table 7: Predicting Daily Trading Volume Using GNH (I)

	(i)		*(ii)*		*(iii)*	
	β	*t-stat*	β	*t-stat*	β	*t-stat*
GNH_{t-1}	-0.0311**	-2.52	-0.0306**	-2.45	-0.0307**	-2.45
GNH_{t-2}	0.0262**	2.32	0.0262**	2.34	0.0258**	2.29
GNH_{t-3}	-0.0079	-0.74	-0.0086	-0.79	-0.0089	-0.82
GNH_{t-4}	0.003	0.28	0.0024	0.23	0.0025	0.23
GNH_{t-5}	0.0148	1.48	0.0149	1.5	0.0152	1.51
$\|GNH_{t-1}\|$	0.0487***	3.62	0.0482***	3.56	0.0478***	3.47
$\|GNH_{t-2}\|$	0.0017	0.15	0.0016	0.14	0.0022	0.19
$\|GNH_{t-3}\|$	0.0174*	1.73	0.0174*	1.72	0.0181*	1.76
$\|GNH_{t-4}\|$	-0.002	-0.18	-0.0015	-0.13	-0.0026	-0.24
$\|GNH_{t-5}\|$	-0.0097	-0.96	-0.0097	-0.96	-0.0099	-0.97
Past volatility	*Yes*		*Yes*		*Yes*	
Past volume	*Yes*		*Yes*		*Yes*	
Environmental controls	*Yes*		*Yes*		*Yes*	
Calendar controls	*Yes*		*Yes*		*Yes*	
Daily economic activity	*No*		*Yes*		*Yes*	
Turn-of-the-year effect	*Yes*		*Yes*		*Yes*	
New Year's Day Dummies	*No*		*No*		*Yes*	
$R-squared$	0.296		0.2986		0.300	
No of Obs	1,090		1,090		1,090	
	$\chi^2 - test$	*p-value*	$\chi^2 - test$	*p-value*	$\chi^2 - test$	*p-value*
$\beta_{1t-1} = 0$	6.33	0.012	6.00	0.0144	5.99	0.0145
$\sum_{j=1}^{5} \beta_{1t-j} = 0$	0.12	0.734	0.09	0.7664	0.07	0.7969
$\beta_{2t-1} = 0$	13.13	0.0003	12.67	0.0004	12.02	0.0005
$\sum_{j=1}^{5} \beta_{2t-j} = 0$	10.64	0.0011	10.16	0.0015	9.68	0.0019

Note: The table reports the coefficient estimates for both the GNH measure (β_{1t}) and the absolute values of GNH (β_{2t}). Each reported coefficient measures the impact of a one-standard-deviation increase in the GNH measure (the absolute values of GNH) on daily trading volume. Daily trading volume is detrended using first-order differencing. Daily economic activity is measured by the ADS index (Aruoba et al., 2009). I use VIX up to five lags to control for past volatility. The regressions are based on 1,090 daily observations from January 1, 2008 to April, 27 2012. Newey and West (1987) standard errors that are robust to heteroskedasticity and autocorrelation for up to five lags are used. All stock market measures and control variables are also winsorized at the 0.5% upper and lower tails of their distributions. The table data come from Facebook, Thomson Reuters Datastream, CBOE, FED Philadelphia, NCDC and NMOC. Three stars denote significance at 1% or less; two stars denote significance at 5% or less; one star denotes significance at 10% or less.

Table 8: Predicting Daily Trading Volume Using GNH (II)

	Positivity						*Negativity*					
	(i)		*(ii)*		*(iii)*		*(iv)*		*(v)*		*(vi)*	
	β	*t-stat*	β	*t-stat*	β	*t-stat*	β	*t-stat*	β	*t-stat*	β	*t-stat*
$Facebook_{t-1}$	-0.0297**	-2.28	-0.0287**	-2.18	-0.0288**	-2.18	0.0112	0.98	0.0119	1.04	0.0121	1.04
$Facebook_{t-2}$	0.0315***	2.65	0.0308***	2.57	0.0304**	2.51	-0.0148**	-1.96	-0.0158**	-2.09	-0.0156**	-2.06
$Facebook_{t-3}$	0.0003	0.02	0.0002	0.02	0.0001	0.01	0.012*	1.71	0.0132*	1.9	0.0132*	1.87
$Facebook_{t-4}$	-0.0053	-0.45	-0.0057	-0.49	-0.0055	-0.47	-0.0042	-0.71	-0.0036	-0.6	-0.0033	-0.54
$Facebook_{t-5}$	0.0165	1.31	0.0166	1.32	0.0167	1.32	-0.0115*	-1.77	-0.0115*	-1.74	-0.012*	-1.8
$\|Facebook_{t-1}\|$	0.0471***	3.35	0.0462***	3.25	0.046***	3.16	0.0114	0.97	0.0117	0.99	0.0112	0.93
$\|Facebook_{t-2}\|$	-0.0045	-0.37	-0.0041	-0.33	-0.0034	-0.27	0.0062	0.85	0.0058	0.8	0.0055	0.77
$\|Facebook_{t-3}\|$	0.0104	0.88	0.0097	0.82	0.0102	0.84	0.0145**	2.27	0.0151**	2.37	0.0155**	2.38
$\|Facebook_{t-4}\|$	0.0064	0.56	0.0068	0.59	0.0057	0.5	-0.0014	-0.24	-0.0012	-0.2	-0.0018	-0.31
$\|Facebook_{t-5}\|$	-0.012	-0.94	-0.0119	-0.95	-0.012	-0.93	0.0085	1.33	0.0088	1.37	0.0091	1.4
Past volatility	*Yes*		*Yes*		*Yes*		*Yes*		*Yes*		*Yes*	
Past volume	*Yes*		*Yes*		*Yes*		*Yes*		*Yes*		*Yes*	
Environmental controls	*Yes*		*Yes*		*Yes*		*Yes*		*Yes*		*Yes*	
Calendar controls	*Yes*		*Yes*		*Yes*		*Yes*		*Yes*		*Yes*	
Daily economic activity	*No*		*Yes*		*Yes*		*No*		*Yes*		*Yes*	
Turn-of-the-year effect	*Yes*		*Yes*		*Yes*		*Yes*		*Yes*		*Yes*	
New Year's Day Dummies	*No*		*No*		*Yes*		*No*		*No*		*Yes*	
$R-squared$	0.29		0.2921		0.2936		0.2754		0.2801		0.2819	
No of Obs	1,090		1,090		1,090		1,090		1,090		1,090	
	χ^2-test	*p-value*	χ^2-test	*p-value*	χ^2-test	*p-value*	χ^2-test	*p-value*	χ^2-test	*p-value*	χ^2-test	*p-value*
$\beta_{1t-1}=0$	5.22	0.0225	4.74	0.0297	4.73	0.0298	0.96	0.328	1.08	0.2999	1.08	0.2993
$\sum_{j=1}^{5}\beta_{1t-j}=0$	1.06	0.3042	1.03	0.311	0.94	0.3326	0.38	0.5373	0.23	0.6292	0.21	0.6454
$\beta_{2t-1}=0$	11.21	0.0008	10.54	0.0012	9.99	0.0016	0.94	0.3325	0.98	0.3224	0.87	0.3515
$\sum_{j=1}^{5}\beta_{2t-j}=0$	8.02	0.0047	7.48	0.0063	7.13	0.0077	7.34	0.0068	7.52	0.0062	7.28	0.0071

Note: The table reports the coefficient estimates for the Positivity (Negativity) measure (β_{1t}) and the absolute values of Positivity (Negativity) (β_{2t}). Each reported coefficient measures the impact of a one-standard-deviation increase in the Positivity (Negativity) measure (the absolute values of Positivity (Negativity)) on daily trading volume. Daily trading volume is detrended using first-order differencing. Daily economic activity is measured by the ADS index (Aruoba et al., 2009). I use VIX up to five lags to control for past volatility. The regressions are based on 1,090 daily observations from January 1, 2008 to April, 27 2012. Newey and West (1987) standard errors that are robust to heteroskedasticity and autocorrelation for up to five lags are used. All stock market measures and control variables are also winsorized at the 0.5% upper and lower tails of their distributions. The table data come from Facebook, Thomson Reuters Datastream, CBOE, FED Philadelphia, NCDC and NMOC. Three stars denote significance at 1% or less; two stars denote significance at 5% or less; one star denotes significance at 10% or less.

Table 9: Predicting Weekly Trading Volume Using GNH

	GNH		$Positivity$		$Negativity$	
	(i)		*(ii)*		*(iii)*	
	β	*t-stat*	β	*t-stat*	β	*t-stat*
$Facebook_{t-1}$	-0.0294	-1.07	-0.0297	-1.27	-0.0201	-1.55
$Facebook_{t-2}$	0.0268	1.12	0.0365**	1.99	-0.0041	-0.38
$Facebook_{t-3}$	0.0848***	3.71	0.0768***	3.38	-0.0516***	-3.47
$\lvert Facebook_{t-1}\rvert$	0.078***	2.6	0.0812***	3.44	0.0315**	2.28
$\lvert Facebook_{t-2}\rvert$	-0.0225	-0.72	-0.0332	-1.43	0.0292**	2.21
$\lvert Facebook_{t-3}\rvert$	-0.0652**	-2.45	-0.0602***	-2.65	-0.0345**	-2.28
Past volatility	*Yes*		*Yes*		*Yes*	
Past volume	*Yes*		*Yes*		*Yes*	
Environmental controls	*Yes*		*Yes*		*Yes*	
Calendar controls	*Yes*		*Yes*		*Yes*	
Weekly economic activity	*Yes*		*Yes*		*Yes*	
Turn-of-the-year effect	*Yes*		*Yes*		*Yes*	
$R-squared$	0.4897		0.4828		0.4841	
No of Obs	226		226		226	
	χ^2-test	*p-value*	χ^2-test	*p-value*	χ^2-test	*p-value*
$\beta_{t-1}=0$	1.15	0.2854	1.61	0.2065	2.41	0.1222
$\sum_{j=1}^{3}\beta_{t-j}=0$	21.98	0.00	8.15	0.0048	10.62	0.0013
$\beta_{t-1}=0$	6.75	0.01	11.81	0.0007	5.18	0.0239
$\sum_{j=1}^{3}\beta_{t-j}=0$	9.14	0.002	0.15	0.6945	1.05	0.3063

Note: The table reports the estimates of the coefficients on the GNH measure or one of its affective dimensions (β_{1t}) and and the absolute values of the GNH measure or one of its affective dimensions (β_{2t}). Each reported coefficient measures the impact of a one-standard-deviation increase in GNH (Positivity or Negativity) on weekly trading volume. Weekly trading volume is detrended using first-order differencing. Weekly economic activity is measured by the ADS index (Aruoba et al., 2009). I use VIX up to five lags to control for past volatility. The regressions are based on 226 weekly observations from January 1, 2008 to April, 27 2012. Newey and West (1987) standard errors that are robust to heteroskedasticity and autocorrelation for up to three lags are used. All stock market measures and control variables are winsorized at the 0.5% upper and lower tails of their distributions. The table data come from Facebook, Thomson Reuters Datastream, CBOE, FED Philadelphia, NCDC and NMOC. Three stars denote significance at 1% or less; two stars denote significance at 5% or less; one star denotes significance at 10% or less.

Table 10: Predicting Daily Market Returns Using GNH: Timing Issues (I)

	SPDR Dow Jones Total Market ETF				iShares NYSE Composite Index ETF				SPDR SP 500 ETF			
	Close-to-close		*Open-to-close*		*Close-to-close*		*Open-to-close*		*Close-to-close*		*Open-to-close*	
	β	*t-stat*	β	*t-stat*	β	*t-stat*	β	*t-stat*	β	*t-stat*	β	*t-stat*
GNH_{t-1}	9.3834**	2.02	6.2817*	1.76	10.1337**	2.36	5.6964*	1.75	10.4028**	2.27	5.2796	1.45
GNH_{t-2}	-1.1895	-0.28	-3.4423	-0.93	0.6414	0.16	-2.367	-0.66	-0.5312	-0.13	-1.5841	-0.47
GNH_{t-3}	8.5481**	2.25	1.803	0.64	9.0813**	2.11	-1.7564	-0.52	7.754**	2.13	2.6886	0.99
GNH_{t-4}	-1.8127	-0.37	2.0193	0.55	-0.6795	-0.08	1.0793	0.33	-0.7137	-0.16	0.9093	0.25
GNH_{t-5}	2.4131	0.68	1.7865	0.61	2.2221	0.55	0.8688	0.34	2.2588	0.67	4.0927	1.39
Past volatility	*Yes*		*Yes*		*Yes*		*Yes*		*Yes*		*Yes*	
Environmental controls	*Yes*		*Yes*		*Yes*		*Yes*		*Yes*		*Yes*	
Calendar controls	*Yes*		*Yes*		*Yes*		*Yes*		*Yes*		*Yes*	
Economic activity	*Yes*		*Yes*		*Yes*		*Yes*		*Yes*		*Yes*	
Turn-of-the-year effect	*Yes*		*Yes*		*Yes*		*Yes*		*Yes*		*Yes*	
New Year's Day Dummies	*Yes*		*Yes*		*Yes*		*Yes*		*Yes*		*Yes*	
$R-squared$	0.0714		0.0751		0.0497		0.0634		0.0866		0.0795	
No of Obs	1,086		1,086		1,086		1,086		1,086		1,086	
	χ^2-test	*p-value*	χ^2-test	*p-value*	χ^2-test	*p-value*	χ^2-test	*p-value*	χ^2-test	*p-value*	χ^2-test	*p-value*
$\beta_{t-1}=0$	4.09	0.0433	3.11	0.0781	5.57	0.0185	3.05	0.0811	5.13	0.0237	2.12	0.146
$\sum_{j=1}^{5}\beta_{t-j}=0$	3.23	0.0727	1.08	0.2994	4.94	0.0265	0.23	0.6314	4.21	0.0404	2.02	0.1553
$\sum_{j=2}^{5}\beta_{t-j}=0$	0.9	0.3433	0.1	0.7569	1.6	0.2067	0.09	0.7614	1.13	0.288	0.82	0.3662

Note: The table reports the estimates of coefficients on the GNH measure. Each reported coefficient measures the impact of a one-standard deviation increase in the GNH measure on either close-to-close or open-to-close returns of different ETFs in basis points. I use the Dow Jones Total Market ETF (TMW) for the Dow Jones US Total Stock Market Index, the NYSE Composite Index ETF (NYC) for the NYSE Composite Index and the S&P 500 ETF (SPY) for the S&P 500 Index. Daily economic activity is measured by the ADS index (Aruoba et al., 2009). I use VIX up to five lags to control for past volatility. The regressions are based on 1,086 daily observations from January 1, 2008 to April, 27 2012. Newey and West (1987) standard errors that are robust to heteroskedasticity and autocorrelation for up to five lags are used. All stock market measures and control variables are also winsorized at the 0.5% upper and lower tails of their distributions. The table data come from Facebook, Thomson Reuters Datastream, CBOE, FED Philadelphia, NCDC and NMOC. Three stars denote significance at 1% or less; two stars denote significance at 5% or less; one star denotes significance at 10% or less.

Table 11: Predicting Daily Market Returns Using GNH: Timing Issues (II)

	Close-to-close		*Open-to-close*		*Close-to-close*		*Open-to-close*		*Close-to-close*		*Open-to-close*	
	β	*t-stat*	β	*t-stat*	β	*t-stat*	β	*t-stat*	β	*t-stat*	β	*t-stat*
$Positivity_{t-1}$	9.6139**	2.14	6.6565**	1.98	9.8309**	2.43	6.086*	1.89	10.3449**	2.3	6.0478*	1.71
$Positivity_{t-2}$	0.0252	0.01	-2.906	-0.89	2.2849	0.52	-1.6055	-0.47	0.6449	0.16	-0.7125	-0.23
$Positivity_{t-3}$	7.7424**	2.2	1.4292	0.54	8.0801**	2.07	-2.3593	-0.72	7.0682**	2.1	2.2846	0.91
$Positivity_{t-4}$	-3.4718	-0.73	0.7338	0.21	-2.8184	-0.55	-0.2065	-0.1	-3.2441	-0.73	-0.791	-0.23
$Positivity_{t-5}$	2.228	0.67	1.423	0.5	2.3058	0.59	0.4741	0.21	2.1264	0.66	3.7672	1.35
Past volatility	*Yes*		*Yes*		*Yes*		*Yes*		*Yes*		*Yes*	
Environmental controls	*Yes*		*Yes*		*Yes*		*Yes*		*Yes*		*Yes*	
Calendar controls	*Yes*		*Yes*		*Yes*		*Yes*		*Yes*		*Yes*	
Economic activity	*Yes*		*Yes*		*Yes*		*Yes*		*Yes*		*Yes*	
Turn-of-the-year effect	*Yes*		*Yes*		*Yes*		*Yes*		*Yes*		*Yes*	
New Year's Day Dummies	*Yes*		*Yes*		*Yes*		*Yes*		*Yes*		*Yes*	
$R-squared$	0.0716		0.0751		0.0499		0.0636		0.0869		0.0797	
No of Obs	1,086		1,086		1,086		1,086		1,086		1,086	
	χ^2-test	*p-value*	χ^2-test	*p-value*	χ^2-test	*p-value*	χ^2-test	*p-value*	χ^2-test	*p-value*	χ^2-test	*p-value*
$\beta_{t-1}=0$	4.59	0.0324	3.94	0.0475	5.89	0.0154	3.57	0.059	5.27	0.0219	2.93	0.0873
$\sum_{j=1}^{5}\beta_{t-j}=0$	3.35	0.0675	1.01	0.3142	5.49	0.0193	0.12	0.7312	3.95	0.0473	2.15	0.1432
$\sum_{j=2}^{5}\beta_{t-j}=0$	0.75	0.388	0.01	0.9122	1.64	0.2011	0.36	0.5469	0.76	0.384	0.58	0.4481

Note: The table reports the coefficient estimates for the Positivity (Negativity) measure. Each reported coefficient measures the impact of a one-standard deviation increase in the Positivity (Negativity) on either close-to-close or open-to-close returns of different ETFs in basis points. I use the Dow Jones Total Market ETF (TMW) for the Dow Jones US Total Stock Market Index, the NYSE Composite Index ETF (NYC) for the NYSE Composite Index and the S&P 500 ETF (SPY) for the S&P 500 Index. Daily economic activity is measured by the ADS index (Aruoba et al., 2009). I use VIX up to five lags to control for past volatility. The regressions are based on 1,086 daily observations from January 1, 2008 to April, 27 2012. Newey and West (1987) standard errors that are robust to heteroskedasticity and autocorrelation for up to five lags are used. All stock market measures and control variables are also winsorized at the 0.5% upper and lower tails of their distributions. The table data come from Facebook, Thomson Reuters Datastream, CBOE, FED Philadelphia, NCDC and NMOC. Three stars denote significance at 1% or less; two stars denote significance at 5% or less; one star denotes significance at 10% or less.

Table 12: Predicting Weekly Initial Claims Using GNH

	GNH				Positivity				Negativity			
	(i)		*(ii)*		*(iii)*		*(iv)*		*(v)*		*(vi)*	
	β	*t-stat*	β	*t-stat*	β	*t-stat*	β	*t-stat*	β	*t-stat*	β	*t-stat*
$Claims_{t-1}$	-0.0094***	-3.38	-0.0129***	-3.84	-0.0094***	-3.38	-0.0127***	-3.79	-0.0093***	-3.54	-0.0127***	-3.97
$Claims_{t-2}$	-	-	-0.0062*	-1.76	-	-	-0.0063*	-1.79	-	-	-0.0063*	-1.9
$Claims_{t-3}$	-	-	-0.0019	-0.56	-	-	-0.0021	-0.63	-	-	-0.0025	-0.78
$Facebook_{t-1}$	-0.0005	-0.16	-0.0033	-1.37	-0.0006	-0.17	-0.0033	-1.27	0.0002	0.07	0.0014	0.56
$Facebook_{t-2}$	-	-	0.0054**	2.06	-	-	0.0051*	1.86	-	-	-0.0039	-1.53
$Facebook_{t-3}$	-	-	0.0023	1.5	-	-	0.0014	0.9	-	-	-0.0048**	-2.32
$R-squared$	0.0619		0.1289		0.062		0.1242		0.0618		0.1314	
No of Obs	222		220		222		220		222		220	
	χ^2-test	*p-value*	χ^2-test	*p-value*	χ^2-test	*p-value*	χ^2-test	*p-value*	χ^2-test	*p-value*	χ^2-test	*p-value*
$\beta_{t-1}=0$	0.03	0.8704	1.88	0.1719	0.03	0.8676	1.6	0.2071	0.01	0.9388	0.32	0.5729
$\sum_{j=1}^{3}\beta_{j-1}=0$	-	-	1.04	0.3091	-	-	0.46	0.4967	-	-	5.01	0.0263

Note: The table reports the coefficient estimates on GNH, Positivity and Negativity. Each reported coefficient measures the impact of a one-standard deviation increase in the Facebook measures on weekly log changes in initial jobless claims. The regressions are based on 222 (220) weekly observations from January 1, 2008 to April, 7 2012. Newey and West (1987) standard errors that are robust to heteroskedasticity and autocorrelation for up to five lags are used. The table data come from Facebook and the web site of the US Department of Labor. Three stars denote significance at 1% or less; two stars denote significance at 5% or less; one star denotes significance at 10% or less.

Table 13: Out-of-Sample Tests: International Evidence

	UK				Germany			
	Daily log returns		*Daily trading volume*		*Daily log returns*		*Daily trading volume*	
	(i)		*(ii)*		*(iii)*		*(iv)*	
	β	*t-stat*	β	*t-stat*	β	*t-stat*	β	*t-stat*
GNH_{t-1}	11.8549***	2.78	-0.0261**	-2.1	13.9552**	2.21	-0.0231	-1.39
GNH_{t-2}	7.0593*	1.83	-0.0003	-0.02	0.758	0.09	-0.0126	-0.74
GNH_{t-3}	2.1679	0.54	-0.0001	-0.01	1.9808	0.28	0.0045	0.21
GNH_{t-4}	6.79	1.41	0.0204	1.29	3.8155	0.5	0.0086	0.52
GNH_{t-5}	-0.804	-0.22	0.0194	1.47	-4.3898	-0.71	0.0159	1.16
$\|GNH_{t-1}\|$	-		0.0389***	2.75	-		0.038**	2.49
$\|GNH_{t-2}\|$	-		0.0189	1.28	-		0.0063	0.41
$\|GNH_{t-3}\|$	-		-0.0109	-0.8	-		-0.0367*	-1.66
$\|GNH_{t-4}\|$	-		-0.0048	-0.32	-		0.0018	0.1
$\|GNH_{t-5}\|$	-		-0.0084	-0.61	-		-0.0106	-0.79
Past volatility	*Yes*		*Yes*		*Yes*		*Yes*	
Environmental controls	*Yes*		*Yes*		*Yes*		*Yes*	
Calendar controls	*Yes*		*Yes*		*Yes*		*Yes*	
Turn-of-the-year effect	*Yes*		*Yes*		*Yes*		*Yes*	
$R-squared$	0.0522		*0.4028*		0.0333		0.1641	
No of Obs	1,088		1,088		1,099		1,099	
	χ^2-test	*p-value*	χ^2-test	*p-value*	χ^2-test	*p-value*	χ^2-test	*p-value*
$\beta_{t-1}=0$	7.72	0.0055	4.41	0.0359	4.88	0.027	1.93	0.1651
$\sum_{j=1}^{5}\beta_{t-j}=0$	8.81	0.0031	0.52	0.4725	4.54	0.0333	0.18	0.6678
$\beta_{t-1}=0$	-	-	7.56	0.0061	-	-	6.18	0.013
$\sum_{j=1}^{5}\beta_{t-j}=0$	-	-	3.9	0.048	-	-	0.01	0.945

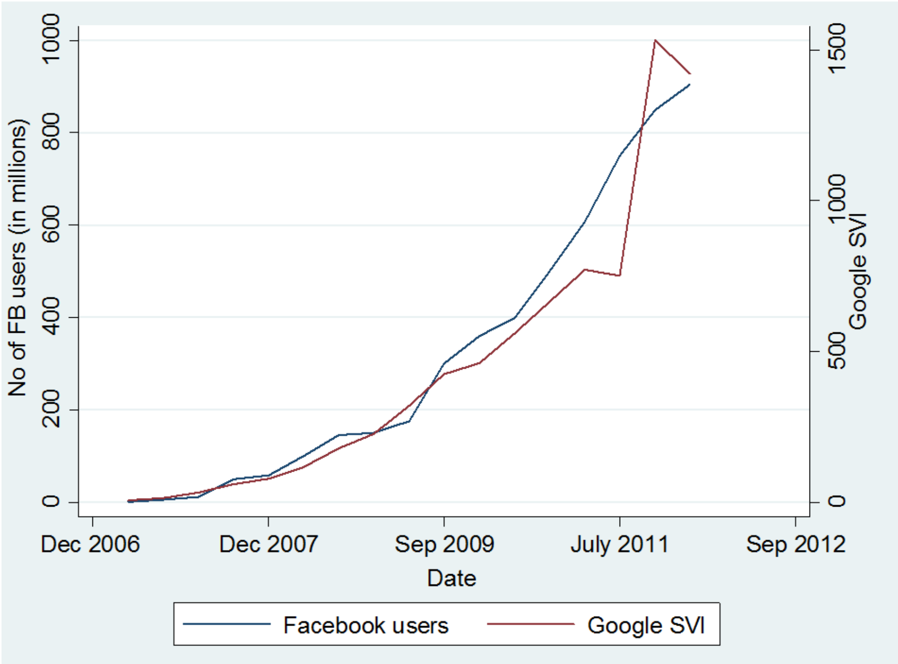
Note: The table reports the estimates of the coefficients on the GNH measure. Each reported coefficient measures the impact of a one-standard-deviation increase in the GNH measure on the daily returns in basis points and daily trading volume. The first two columns present the predictability regressions for the UK market, while the latter two present the results for German market. In columns (i) and (ii), I use the GNH measure from the UK. In columns (iii) and (iv), I use the GNH measure from Germany. The regressions are based on 1,088 (1,099) daily observations from January 1, 2008 to April, 27 2012 for the UK (Germany). Stock market volatility is measured using the detrended squared residuals, as in Tetlock (2007). Daily trading volume is detrended using first-order differencing. Newey and West (1987) standard errors that are robust to heteroskedasticity and autocorrelation for up to five lags are used. All stock market measures and control variables are also winsorized at the 0.5% upper and lower tails of their distributions. The table data come from Facebook, Thomson Reuters Datastream, CBOE, NCDC and NMOC. Three stars denote significance at 1% or less; two stars denote significance at 5% or less; one star denotes significance at 10% or less.

Table 14: Validation Exercise: Dual-Listed Companies and GNH

	Reed / Elsevier Int	Unilever	Rio Tinto	BHP Billiton
Panel A:				
$(GNH^{NL} - GNH^{UK})$	0.0048***	0.0033***	-	-
	(10.18)	(9.60)	-	-
$(GNH^{AUS} - GNH^{UK})$	-	-	-0.0035	0.0005
	-	-	(0.64)	(0.07)
Exchange rate changes	-0.0026	-0.0014	0.0045	0.0046
	(0.1463)	(0.90)	(0.74)	(0.1515)
Panel B:				
(GNH^{UK})	-0.009**	-0.0032*	0.0072	-0.0016
	(6.39)	(2.73)	(0.52)	(0.11)
(GNH^{NL})	0.01***	0.0058***	-	-
	(10.24)	(8.21)	-	-
(GNH^{AUS})	-	-	-0.0083	0.0011
	-	-	(0.70)	(0.04)
Exchange rate changes	-0.0026	-0.0013	0.0045	0.0045
	(2.21)	(0.94)	(0.68)	(2.04)
No of Obs	1,088	1,088	954	954

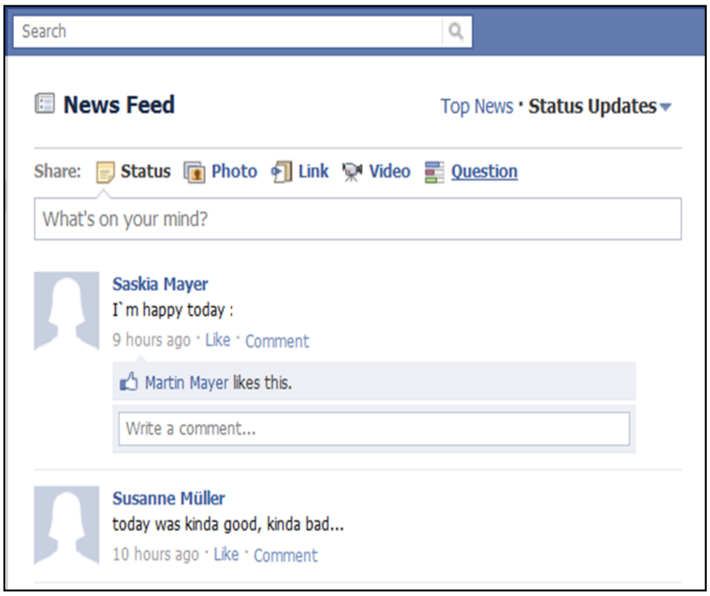
Note: The table reports the estimation results of time-series regressions of relative price deviations of dual-listed companies. Newey and West (1987) standard errors that are robust to heteroskedasticity and autocorrelation are used. The dependent variable is the company twins' log price deviation from their theoretical parity. GNH^{UK} represents GNH from the UK, and GNH^{NL} and GNH^{AUS} represent GNH from Australia and the Netherlands, respectively. I compute daily exchange rate changes by dividing the (log) lagged local currency British pound exchange rate by the current day's exchange rate. The significance tests are $\chi^2 - tests$ on the sum of the lead, current and lag coefficients for GNH and changes in exchange rates, and they are presented in parentheses. The table data come from Facebook and Thomson Reuters Datastream. Three stars denote significance at 1% or less; two stars denote significance at 5% or less; one star denotes significance at 10% or less.

Figure 1: Number of Facebook Users over Time



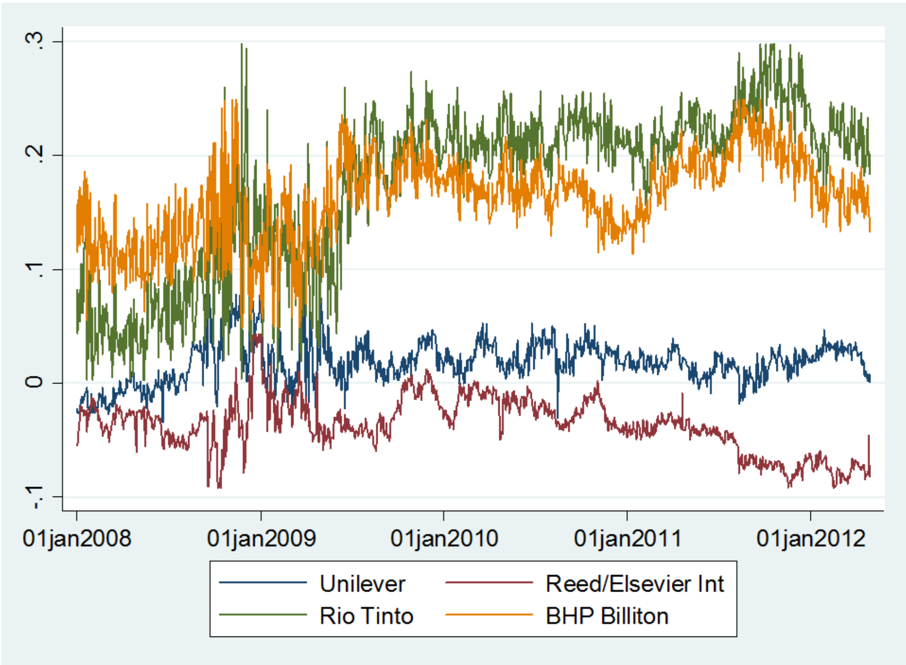
Note: This figure illustrates the total number of Facebook users and Google's Search Volume Index (SVI) for the keyword 'Facebook' over the period beginning from December, 2004 through April, 2012. The data come from Google and Facebook.

Figure 2: Status Updates in Facebook



Note: This figure illustrates an example of Facebook's start page. The question 'What's on your mind?' shows up in this page whenever the user logs on to Facebook. The status updates and the recent activity of friends in the network will be shown on this page as well.

Figure 3: Log Price Deviations of Dual-Listed Companies



Note: This figure illustrates the log price deviations of dual-listed companies (Rio Tinto, BHP Billiton, Unilever and Reed/Elsevier International) from their theoretical price parity. The observation period is from January 1, 2008 to April, 27 2012. The data come from Thomson Reuters Datastream.

A Appendix

Can Facebook Predict Stock Market Activity?

Table A.1: Predicting Daily Trading Volume Using GNH: Alternative Trading Volume Measure

	(i)		*(ii)*		*(iii)*	
	β	*t-stat*	β	*t-stat*	β	*t-stat*
GNH_{t-1}	-0.0353***	-3.02	-0.035***	-2.98	-0.0347***	-2.92
GNH_{t-2}	0.0231**	2.09	0.0231**	2.1	0.0226**	2.05
GNH_{t-3}	-0.0119	-1.13	-0.0125	-1.17	-0.0129	-1.2
GNH_{t-4}	0.0007	0.07	0.0003	0.03	0.0004	0.04
GNH_{t-5}	0.0076	0.8	0.0079	0.83	0.0083	0.86
$\lvert GNH_{t-1}\rvert$	0.0464***	3.73	0.0459***	3.69	0.0454***	3.62
$\lvert GNH_{t-2}\rvert$	-0.003	-0.28	-0.0033	-0.31	-0.0028	-0.27
$\lvert GNH_{t-3}\rvert$	0.018*	1.81	0.0178*	1.78	0.0182*	1.81
$\lvert GNH_{t-4}\rvert$	-0.0002	-0.02	-0.0001	-0.01	-0.0003	-0.03
$\lvert GNH_{t-5}\rvert$	-0.0036	-0.37	-0.0039	-0.4	-0.0042	-0.44
Past volatility	*Yes*		*Yes*		*Yes*	
Past volume	*Yes*		*Yes*		*Yes*	
Environmental controls	*Yes*		*Yes*		*Yes*	
Calendar controls	*Yes*		*Yes*		*Yes*	
Daily economic activity	*No*		*Yes*		*Yes*	
Turn-of-the-year effect	*Yes*		*Yes*		*Yes*	
New Year's Day Dummies	*Yes*		*No*		*No*	
$R-squared$	*0.5421*		*0.5435*		*0.5442*	
No of Obs	*1,090*		*1,090*		*1,090*	
	χ^2-test	*p-value*	χ^2-test	*p-value*	χ^2-test	*p-value*
$\beta_{1t-1}=0$	9.13	0.0026	8.89	0.0029	8.53	0.0036
$\sum_{j=1}^{5}\beta_{1t-j}=0$	1.31	0.2526	1.38	0.2411	1.37	0.2429
$\beta_{2t-1}=0$	13.93	0.0002	13.63	0.0002	13.08	0.0003
$\sum_{j=1}^{5}\beta_{2t-j}=0$	14.41	0.0002	13.52	0.0002	13.05	0.0003

Note: The table reports the coefficient estimates for both the GNH measure (β_{1t}) and the absolute values of GNH (β_{2t}). Each reported coefficient measures the impact of a one-standard-deviation increase in the GNH measure (the absolute values of GNH) on daily trading volume. Daily trading volume is detrended using the moving average detrending as in Campbell et al. (1993) and Tetlock (2007). Daily economic activity is measured by the ADS index (Aruoba et al., 2009). I use VIX up to 5 lags to control for past volatility. The regressions are based on 1,090 daily observations from January 1, 2008 to April, 27 2012. Newey and West (1987) standard errors that are robust to heteroskedasticity and autocorrelation for up to five lags are used. All stock market measures and control variables are also winsorized at the 0.5% upper and lower tails of their distributions. The table data come from Facebook, Thomson Reuters Datastream, CBOE, FED Philadelphia, NCDC and NMOC. Three stars denote significance at 1% or less; two stars denote significance at 5% or less; one star denotes significance at 10% or less.

Table A.2: Predicting Daily Market Returns Using GNH: Alternative Volatility Measures

	Detrended squared residuals				Innovations in seasonal-adjusted VIX			
	Daily log returns		*Daily trading volume*		*Daily log returns*		*Daily trading volume*	
	(i)		*(ii)*		*(iii)*		*(iv)*	
	β	*t-stat*	β	*t-stat*	β	*t-stat*	β	*t-stat*
GNH_{t-1}	11.2172**	2.14	-0.031**	-2.5	11.4935**	2.28	-0.0325***	-2.58
GNH_{t-2}	1.616	0.38	0.024**	2.12	0.7363	0.15	0.0269**	2.39
GNH_{t-3}	6.2813*	1.65	-0.0104	-1.01	6.7292*	1.7	-0.01	-0.94
GNH_{t-4}	-0.1277	-0.02	0.003	0.28	-0.7315	-0.15	0.0021	0.2
GNH_{t-5}	3.4686	0.91	0.0152	1.5	3.9929	1.11	0.0153	1.54
$\|GNH_{t-1}\|$	-		0.0474***	3.44	-		0.0493***	3.58
$\|GNH_{t-2}\|$	-		0.0035	0.31	-		0.0018	0.16
$\|GNH_{t-3}\|$	-		0.0196**	1.97	-		0.0196**	1.97
$\|GNH_{t-4}\|$	-		-0.0025	-0.24	-		-0.0026	-0.25
$\|GNH_{t-5}\|$	-		-0.0097	-0.93	-		-0.0101	-0.98
Past volatility	*Yes*		*Yes*		*Yes*		*Yes*	
Environmental controls	*Yes*		*Yes*		*Yes*		*Yes*	
Calendar controls	*Yes*		*Yes*		*Yes*		*Yes*	
Economic activity	*Yes*		*Yes*		*Yes*		*Yes*	
New Year's Day Dummies	*Yes*		*Yes*		*Yes*		*Yes*	
$R-squared$	0.0685		0.3063		0.0609		0.302	
No of Obs	1,090		1,090		1,090		1,090	
	χ^2-test	*p-value*	χ^2-test	*p-value*	χ^2-test	*p-value*	χ^2-test	*p-value*
$\beta_{1t-1}=0$	4.58	0.0327	6.26	0.0125	5.2	0.0228	6.64	0.0101
$\sum_{j=1}^{5}\beta_{1t-j}=0$	4.9	0.0271	0.01	0.9605	4.62	0.0318	0.01	0.9041
$\beta_{2t-1}=0$			11.83	0.0006			12.79	0.0004
$\sum_{j=1}^{5}\beta_{2t-j}=0$			10.31	0.0014			10.43	0.0013

Note: The table reports the estimates of the coefficients on the GNH measure. Each reported coefficient measures the impact of a one-standard-deviation increase in the GNH measure on daily returns (in basis points) and daily trading volume. Daily trading volume is detrended using the first order differencing. Daily economic activity is measured by the ADS index (Aruoba et al., 2009). In columns (i) and (ii), the volatility is measured using the detrended squared residuals, as in Tetlock (2007). In columns (iii) and (iv), I measure stock market volatility using the innovations in seasonal-adjusted VIX computed from the *ARFIMA(0,d,q)* model, as in Da, Engelberg, and Gao (2013). The regressions are based on 1,090 daily observations from January 1, 2008 to April, 27 2012. Newey and West (1987) standard errors that are robust to heteroskedasticity and autocorrelation for up to five lags are used. All stock market measures and control variables are also winsorized at the 0.5% upper and lower tails of their distributions. The table data come from Facebook, Thomson Retuers Datastream, CBOE, FED Philadelphia, NCDC and NMOC. Three stars denote significance at 1% or less; two stars denote significance at 5% or less; one star denotes significance at 10% or less.

Table A.3: Predicting Daily Stock Market Activity Using GNH: Turn-of-the-year effect

	GNH				Positivity				Negativity			
	Daily log returns		*Daily trading volume*		*Daily log returns*		*Daily trading volume*		*Daily log returns*		*Daily trading volume*	
	(i)		(ii)		(iii)		(iv)		(v)		(vi)	
	β	*t-stat*	β	*t-stat*	β	*t-stat*	β	*t-stat*	β	*t-stat*	β	*t-stat*
$Facebook_{t-1}$	1.9222	0.42	-0.0393***	-3.87	1.6552	0.36	-0.0355***	-3.41	-2.4713	-0.36	0.0296***	3.18
$Facebook_{t-2}$	-1.389	-0.21	0.0088	1.01	1.2354	0.18	0.0106	1.13	11.6461	1.26	-0.0093	-1.3
$Facebook_{t-3}$	4.311	1	-0.0113	-1.31	3.1389	0.73	-0.0004	-0.04	-6.6066	-1.03	0.0137**	2
$Facebook_{t-4}$	4.4977	1.19	-0.0047	-0.61	1.7713	0.46	-0.0109	-1.31	-12.5399**	-2.15	0.0013	0.22
$Facebook_{t-5}$	-3.3768	-0.83	0.0092	1.37	-2.7641	-0.68	0.0097	1.17	6.4508	1.23	-0.0129**	-2.36
$\|Facebook_{t-1}\|$	-		0.0162**	1.98	-		0.0138	1.6	-		-0.0113	-1.45
$\|Facebook_{t-2}\|$	-		-0.0013	-0.14	-		-0.0033	-0.33	-		-0.0004	-0.05
$\|Facebook_{t-3}\|$	-		0.0138*	1.78	-		0.0047	0.53	-		0.0155**	2.28
$\|Facebook_{t-4}\|$	-		-0.0061	-0.75	-		0.0017	0.21	-		-0.0056	-0.94
$\|Facebook_{t-5}\|$	-		-0.0046	-0.58	-		-0.0043	-0.45	-		0.0059	1
Past volatility	*Yes*		*Yes*		*Yes*		*Yes*		*Yes*		*Yes*	
Environmental controls	*Yes*		*Yes*		*Yes*		*Yes*		*Yes*		*Yes*	
Economic additivity	*Yes*		*Yes*		*Yes*		*Yes*		*Yes*		*Yes*	
Calendar controls	*Yes*		*Yes*		*Yes*		*Yes*		*Yes*		*Yes*	
$R-squared$	0.0776		0.3085		0.077		0.2979		0.0825		0.309	
No of Obs	1,043		1,043		1,043		1,043		1,043		1,043	
	χ^2-test	*p-value*	χ^2-test	*p-value*	χ^2-test	*p-value*	χ^2-test	*p-value*	χ^2-test	*p-value*	χ^2-test	*p-value*
$\beta_{1t-1}=0$	0.17	0.6776	14.96	0.0001	0.13	0.7197	11.63	0.0007	0.13	0.7179	10.11	0.0015
$\sum_{j=1}^{5}\beta_{1t-j}=0$	0.46	0.499	9.74	0.0019	0.39	0.8143	6.08	0.0138	0.2	0.656	6.24	0.0127
$\beta_{2t-1}=0$			3.93	0.0478			2.56	0.1			2.09	0.1486
$\sum_{j=1}^{5}\beta_{2t-j}=0$			2.29	0.1309			1.14	0.285			0.18	0.6705

Note: The table reports the estimates of the coefficients on GNH, Positivity, and Negativity. Each reported coefficient measures the impact of a one-standard-deviation increase in the corresponding Facebook measure on daily returns (in basis points) and daily trading volume. Daily trading volume is detrended using the first order differencing. Daily economic activity is measured by the ADS index (Aruoba et al., 2009). I use VIX up to 5 lags to control for past volatility. The regressions are based on 1,043 daily observations from January 1, 2008 to April, 27 2012 after omitting the trading days surrounding the New Year's Day. Newey and West (1987) standard errors that are robust to heteroskedasticity and autocorrelation for up to five lags are used. All stock market measures and control variables are also winsorized at the 0.5% upper and lower tails of their distributions. The table data come from Facebook, Thomson Reuters Datastream, CBOE, FED Philadelphia, NCDC and NMOC. Three stars denote significance at 1% or less; two stars denote significance at 5% or less; one star denotes significance at 10% or less.

Table A.4: Predicting Daily Stock Market Activity Using the Raw GNH Values

	Daily log returns		*Daily trading volume*	
	(i)		*(ii)*	
	β	*t-stat*	β	*t-stat*
GNH_{t-1}	12.1024**	2.54	-0.0186	-1.22
GNH_{t-2}	-0.9641	-0.27	0.0385***	2.64
GNH_{t-3}	5.9166*	1.76	-0.0174	-1.49
GNH_{t-4}	-1.6812	-0.36	0.0067	0.55
GNH_{t-5}	0.0877	0.03	0.0377***	3.03
$\lvert GNH_{t-1}\rvert$	-		0.0494***	3.54
$\lvert GNH_{t-2}\rvert$	-		-0.0023	-0.2
$\lvert GNH_{t-3}\rvert$	-		0.0288***	2.79
$\lvert GNH_{t-4}\rvert$	-		-0.0008	-0.08
$\lvert GNH_{t-5}\rvert$	-		-0.0288***	-2.66
Past volatility	*Yes*		*Yes*	
Economic activity	*Yes*		*Yes*	
Environmental controls	*Yes*		*Yes*	
Calendar controls	*Yes*		*Yes*	
$R-squared$	0.0776		*0.3092*	
No of Obs	1,090		1,090	
	$\chi^2 - test$	*p-value*	$\chi^2 - test$	*p-value*
$\beta_{t-1} = 0$	6.43	0.0114	1.49	0.223
$\sum_{j=1}^{5} \beta_{t-j} = 0$	3.97	0.0466	7.67	0.0057
$\beta_{t-1} = 0$	-	-	12.5	0.0004
$\sum_{j=1}^{5} \beta_{t-j} = 0$	-	-	8.34	0.004

Note: The table reports the estimates of the coefficients on the raw GNH measure, that is, I do not carry out any winsorization, seasonality and detreding on the GNH measure. Each reported coefficient measures the impact of a one-standard-deviation increase in the GNH measure on daily returns (in basis points) and daily trading volume. Daily trading volume is detrended using the first order differencing. Daily economic activity is measured by the ADS index (Aruoba et al., 2009). I use VIX up to 5 lags to control for past volatility. The regressions are based on 1,090 daily observations from January 1, 2008 to April, 27 2012 after omitting the trading days surrounding the New Year's Day. Newey and West (1987) standard errors that are robust to heteroskedasticity and autocorrelation for up to five lags are used. All stock market measures and control variables are also winsorized at the 0.5% upper and lower tails of their distributions. The table data come from Facebook, Thomson Reuters Datastream, CBOE, FED Philadelphia, NCDC and NMOC. Three stars denote significance at 1% or less; two stars denote significance at 5% or less; one star denotes significance at 10% or less.

Table A.5: Predicting Daily Trading Volume Using GNH: Alternative Model Specification

	(i)		*(ii)*		*(iii)*	
	β	*t-stat*	β	*t-stat*	β	*t-stat*
GNH_{t-1}	-0.0537***	-2.95	-0.0536***	-2.94	-0.0561***	-3.1
GNH_{t-2}	0.0201	1.39	0.0211	1.46	0.0218	1.5
GNH_{t-3}	0.0019	0.16	0.0011	0.09	0.001	0.09
GNH_{t-4}	-0.0046	-0.45	-0.0055	-0.54	-0.0086	-0.83
GNH_{t-5}	0.0141	1.46	0.0144	1.48	0.016*	1.66
GNH^2_{t-1}	0.067***	3.14	0.0668***	3.12	0.0688***	3.25
GNH^2_{t-2}	0.0093	0.78	0.0081	0.68	0.0083	0.68
GNH^2_{t-3}	0.0078	0.78	0.0077	0.77	0.0091	0.88
GNH^2_{t-4}	0.0075	0.71	0.0082	0.77	0.0094	0.87
GNH^2_{t-5}	-0.0091	-0.89	-0.0093	-0.91	-0.0103	-0.98
Past volatility	*Yes*		*Yes*		*Yes*	
Past volume	*Yes*		*Yes*		*Yes*	
Environmental controls	*Yes*		*Yes*		*Yes*	
Calendar controls	*Yes*		*Yes*		*Yes*	
Economic activity	*No*		*Yes*		*Yes*	
Turn-of-the-year effect	*Yes*		*Yes*		*Yes*	
New Year's Day Dummies	*No*		*No*		*Yes*	
$R-squared$	0.3075		0.3101		0.3137	
No of Obs	1,090		1,090		1,090	
	$\chi^2 - test$	*p-value*	$\chi^2 - test$	*p-value*	$\chi^2 - test$	*p-value*
$\beta_{1t-1} = 0$	8.71	0.0032	8.64	0.0034	9.61	0.002
$\sum_{j=1}^{5} \beta_{1t-j} = 0$	2.42	0.1203	2.49	0.1151	3.23	0.0724
$\beta_{2t-1} = 0$	9.86	0.0017	9.73	0.0019	10.54	0.0012
$\sum_{j=1}^{5} \beta_{2t-j} = 0$	6.99	0.0083	6.58	0.0104	6.4	0.0116

Note: The table reports the coefficient estimates for both the GNH measure (β_{1t}) and the squared values of GNH (β_{2t}). Each reported coefficient measures the impact of a one-standard-deviation increase in the GNH measure (the squared values of GNH) on daily trading volume. Daily trading volume is detrended using first-order differencing. Daily economic activity is measured by the ADS index (Aruoba et al., 2009). I use VIX up to 5 lags to control for past volatility. The regressions are based on 1,090 daily observations from January 1, 2008 to April, 27 2012. Newey and West (1987) standard errors that are robust to heteroskedasticity and autocorrelation for up to five lags are used. All stock market measures and control variables are also winsorized at the 0.5% upper and lower tails of their distributions. The table data come from Facebook, Thomson Reuters Datastream, CBOE, FED Philadelphia, NCDC and NMOC. Three stars denote significance at 1% or less; two stars denote significance at 5% or less; one star denotes significance at 10% or less.

Table A.6: Predicting Daily Market Returns Using Negativity: Timing Issues

	SPDR Dow Jones Total Market ETF				iShares NYSE Composite Index ETF				SPDR S&P 500 ETF			
	Close-to-close		*Open-to-close*		*Close-to-close*		*Open-to-close*		*Close-to-close*		*Open-to-close*	
	β	*t-stat*	β	*t-stat*	β	*t-stat*	β	*t-stat*	β	*t-stat*	β	*t-stat*
$Negativity_{t-1}$	-4.5062	-0.75	-2.5957	0.12	-6.6136	-1.01	-0.639	-0.13	-6.1038	-1.03	-0.2944	-0.3
$Negativity_{t-2}$	10.0975	1.47	6.9146	0.28	11.7555	1.5	-4.1364	-0.48	10.126	1.47	7.2	-1.01
$Negativity_{t-3}$	-8.747	-1.55	-2.043	-0.05	-9.4428	-1.32	3.699	0.43	-6.5904	-1.26	-2.6827	1.56
$Negativity_{t-4}$	-3.4615	-0.58	-5.6461	-0.28	-6.0594	-0.89	-3.4719	-0.51	-7.9334	-1.24	-6.5541	-0.38
$Negativity_{t-5}$	0.4207	0.08	-0.9147	-1.4	2.3557	0.41	-1.5608	-0.24	1.5256	0.3	-1.5619	-0.9
Past volatility	*Yes*		*Yes*		*Yes*		*Yes*		*Yes*		*Yes*	
Environmental controls	*Yes*		*Yes*		*Yes*		*Yes*		*Yes*		*Yes*	
Calendar controls	*Yes*		*Yes*		*Yes*		*Yes*		*Yes*		*Yes*	
Economic activity	*Yes*		*Yes*		*Yes*		*Yes*		*Yes*		*Yes*	
Turn-of-the-year effect	*Yes*		*Yes*		*Yes*		*Yes*		*Yes*		*Yes*	
New Year's Day Dummies	*Yes*		*Yes*		*Yes*		*Yes*		*Yes*		*Yes*	
$R-squared$	0.0709		0.0758		0.0502		0.0646		0.0869		0.0813	
No of Obs	1,086		1,086		1,086		1,086		1,086		1,086	
	χ^2-test	*p-value*	χ^2-test	*p-value*	χ^2-test	*p-value*	χ^2-test	*p-value*	χ^2-test	*p-value*	χ^2-test	*p-value*
$\beta_{t-1}=0$	0.57	0.4507	0.31	0.5759	1.01	0.3141	0.59	0.4416	1.07	0.301	0.001	0.9465
$\sum_{j=1}^{5}\beta_{t-j}=0$	0.53	0.4673	0.28	0.5996	0.57	0.4513	0.33	0.566	1.23	0.267	0.34	0.558
$\sum_{j=2}^{5}\beta_{t-j}=0$	0.04	0.844	0.04	0.8339	0.01	0.9702	0.01	0.9337	0.13	0.72	0.31	0.577

Note: The table reports the estimates of coefficients on the Negativity measure. Each reported coefficient measures the impact of a one-standard deviation increase in the Negativity measure on either close-to-close or open-to-close returns of different ETFs in basis points. I use the Dow Jones Total Market ETF (TMW) for the Dow Jones US Total Stock Market Index, the NYSE Composite Index ETF (NYC) for the NYSE Composite Index, and the S&P 500 ETF (SPY) for the S&P 500 Index. Daily economic activity is measured by the ADS index (Aruoba et al., 2009). I use VIX up to 5 lags to control for past volatility. The regressions are based on 1,086 daily observations from January 1, 2008 to April, 27 2012. Newey and West (1987) standard errors that are robust to heteroskedasticity and autocorrelation for up to five lags are used. All stock market measures and control variables are also winsorized at the 0.5% upper and lower tails of their distributions. The table data come from Facebook, Thomson Reuters Datastream, CBOE, FED Philadelphia, NCDC and NMOC. Three stars denote significance at 1% or less; two stars denote significance at 5% or less; one star denotes significance at 10% or less.

Table A.7: Predicting Daily Trading Volume Using GNH: Timing Issues

	GNH		*Positivity*		*Negativity*	
	(i)		*(ii)*		*(iii)*	
	β	*t-stat*	β	*t-stat*	β	*t-stat*
$Facebook_{t-2}$	0.0083	0.72	0.0129	1.02	-0.0073	-1.14
$Facebook_{t-3}$	-0.0095	-0.84	-0.0001	0	0.0108	1.52
$Facebook_{t-4}$	0	0	-0.0064	-0.52	-0.0029	-0.46
$Facebook_{t-5}$	0.0031	0.29	0.0028	0.2	-0.0077	-1.15
$Facebook_{t-6}$	0.0201**	1.98	0.0173	1.52	-0.0135**	-2.02
$\|Facebook_{t-2}\|$	0.0198*	1.65	0.0141	1.04	0.0105	1.61
$\|Facebook_{t-3}\|$	0.0215**	2.11	0.0127	1.04	0.0144**	2.17
$\|Facebook_{t-4}\|$	0.0028	0.25	0.0091	0.78	0.0007	0.11
$\|Facebook_{t-5}\|$	0.0054	0.53	0.0052	0.41	0.0064	1.05
$\|Facebook_{t-6}\|$	-0.0126	-1.15	-0.0105	-0.84	0.0117*	1.72
Past volatility	*Yes*		*Yes*		*Yes*	
Environmental controls	*Yes*		*Yes*		*Yes*	
Calendar controls	*Yes*		*Yes*		*Yes*	
Turn-of-the-year effect	*Yes*		*Yes*		*Yes*	
$R-squared$	0.2803		0.2757		0.2793	
No of Obs	1,090		1,090		1,090	
	χ^2-test	*p-value*	χ^2-test	*p-value*	χ^2-test	*p-value*
$\beta_{1t-2}=0$	0.52	0.4712	1.05	0.3059	1.31	0.2532
$\sum_{j=2}^{6}\beta_{1t-j}=0$	2.6	0.1072	4.72	0.03	3.38	0.0664
$\beta_{2t-2}=0$	2.74	0.09	1.09	0.2972	2.61	0.1067
$\sum_{j=2}^{6}\beta_{2t-j}=0$	5.34	0.021	3.69	0.05	12.63	0.0004

Note: The table reports the coefficient estimates for both the GNH measure (β_{1t}) and the absolute values of GNH (β_{2t}). Each reported coefficient measures the impact of a one-standard-deviation increase in the GNH measure (the absolute values of GNH) on daily trading volume. Daily trading volume is detrended using first-order differencing. Daily economic activity is measured by the ADS index (Aruoba et al., 2009). I use VIX up to 5 lags to control for past volatility. The regressions are based on 1,090 daily observations from January 1, 2008 to April, 27 2012. Newey and West (1987) standard errors that are robust to heteroskedasticity and autocorrelation for up to five lags are used. All stock market measures and control variables are also winsorized at the 0.5% upper and lower tails of their distributions. The table data come from Facebook, Thomson Reuters Datastream, CBOE, FED Philadelphia, NCDC and NMOC. Three stars denote significance at 1% or less; two stars denote significance at 5% or less; one star denotes significance at 10% or less.

Table A.8: Predicting Consumer Confidence Using GNH

	GNH				$Positivity$				$Negativity$			
	(i)		*(ii)*		*(iii)*		*(iv)*		*(v)*		*(vi)*	
	β	*t-stat*	β	*t-stat*	β	*t-stat*	β	*t-stat*	β	*t-stat*	β	*t-stat*
$CCONF_{t-1}$	-0.0003	-0.04	0.0011	0.1	-0.001	-0.01	0.0015	0.14	-0.0013	-0.17	-0.0015	-0.12
$CCONF_{t-2}$			-0.0177	-2.38			-0.0173**	-2.35			-0.0199**	-2.37
$CCONF_{t-3}$			-0.0029	-0.23			-0.0028	-0.22			-0.0055	-0.41
$CCONF_{t-4}$			0.0063	0.76			0.0063	0.76			0.0047	0.61
$Facebook_{t-1}$	-0.0110	-1.04	-0.0158	-1.51	-0.0096	-0.81	-0.0149	-1.31	0.0148*	1.68	0.0161*	1.68
$Facebook_{t-2}$			0.0009	0.2			0.0019	0.48			0.0037	0.41
$Facebook_{t-3}$			-0.0006	-0.1			0.0004	0.08			0.0057	0.46
$Facebook_{t-4}$			-0.007	-1.68			-0.0069*	-1.74			0.0069	0.74
R-squared	0.0242		0.1413		0.0184		0.1372		0.0431		0.1439	
No of Obs	51		48		51		48		51		48	
	$\chi^2 - test$	*p-value*	$\chi^2 - test$	*p-value*	$\chi^2 - test$	*p-value*	$\chi^2 - test$	*p-value*	$\chi^2 - test$	*p-value*	$\chi^2 - test$	*p-value*
$\beta_{t-1} = 0$	1.09	0.3026	2.27	0.1404	0.66	0.4202	1.71	0.1982	2.82	0.0994	2.84	0.1002
$\sum_{j=1}^{4} \beta_{t-j} = 0$			4.55	0.0393			3.7	0.0619			2.39	0.1305

Note: The table reports the coefficient estimates on GNH, Positivity and Negativity. Each reported coefficient measures the impact of a one-standard deviation increase in the Facebook measures on monthly log changes in the Michigan Index of Consumer Sentiment. The regressions are based on 51 (48) monthly observations from January 1, 2008 to April, 7 2012.Newey and West (1987) standard errors that are robust to heteroskedasticity and autocorrelation for up to five lags are used. The table data come from Facebook and Thomson Reuters Datastream. Three stars denote significance at 1% or less; two stars denote significance at 5% or less; one star denotes significance at 10% or less.

Table A.9: Validation Exercise: Closed-end Fund Discounts and GNH

	Australia	Chile	Germany	Mexico
$GNH^{Foreign}$	-1.4252	-1.1498**	0.2875	-0.7854**
	(1.08)	(3.84)	(0.04)	(6.54)
GNH^{US}	1.0369	1.1208**	-0.0733	0.5913**
	(2.14)	(4.87)	(0.53)	(6.27)
Exchange rate changes	-1.0739***	0.3666	-0.1173	-0.0199
	(5.38)	(0.47)	(0.69)	(0.01)
R-squared	0.1096	0.0931	0.1051	0.0591
No of Obs	209	235	235	235

Note: The table reports the estimation results of weekly time-series regressions of changes in weekly premiums of country closed-end funds (CCEFs) on GNH and exchange rate changes. The sample period covers 235 week period between January 1, 2008 and April 27, 2012 for Chilean, German, and Mexican CCEFs whereas it includes 209 week observations from January 1, 2008 to November 11, 2011 for the Australian CCEF. Newey and West (1987) standard errors that are robust to heteroskedasticity and autocorrelation are used. The dependent variable is the changes in weekly premium. GNH^{US} represents GNH from the US while $GNH^{Foreign}$ is GNH from the foreign market that are Australia, Chile, Germany, and Mexico, respectively. I compute weekly exchange rate changes by dividing the (log) lagged local currency U.S. dollar exchange rate by the current week exchange rate. All regressions include one lag, contemporaneous and one lead of all independent variables to account for non-synchronous trading. For brevity, I report only the sum of coefficient estimates for each variable. The significance tests are χ^2-*tests* on the sum of the lead, current, and lag coefficients for each variable. The data come from Thomson Reuters Datastream, and funds themselves. hree stars denote significance at 1% or less; two stars denote significance at 5% or less; one star denotes significance at 10% or less.