8-2016

Social Media Information and Analyst Forecasts

Mahfuja Malik
Sacred Heart University, malikm@sacredheart.edu

Rajib Hasan
University of Houston, Clear Lake

Abu S. Amin
Central Michigan University

Recommended Citation
Malik, M., Hasan, R., & Amin, A.S. (2016). Social media information and analyst forecasts. *Financial Management Association conference paper*.

Social Media Information and Analyst Forecasts

Mahfuja Malik
Assistant Professor of Accounting
John F. Welch College of Business
Sacred Heart University
malikm@sacredheart.edu

Rajib Hasan
Assistant Professor of Accounting
College of Business
University of Houston Clear Lake
hasan@uhcl.edu

Abu S Amin
Assistant Professor of Finance
College of Business Administration
Central Michigan University
amin1a@cmich.edu

August 2016

Social Media Information and Analyst Forecasts

Abstract

In the past decade, social networking has changed the landscape of information dissemination. The rapid diffusion of social media services such as Facebook and Twitter is unprecedented and offers immense possibilities for corporations to communicate with, and engage core stakeholders in, various business decisions. In this study, we investigate whether social media play any role as a source of information for financial analysts. We specifically focus on information revealed on the official Facebook pages of S&P 500 firms. We define information content on a Facebook page as the total number of posts by the corporations and the comments, likes and shares (CLS) by Facebook users. By using the data of 4,929 quarterly forecasts from 2008 to 2012, we find that analyst forecast errors decrease significantly with the amount of information content on Facebook. This finding is robust, using the information content on Facebook pages for various time windows before the forecast dates. We further find that the information that helps analysts with forecasting is generated from public reaction, i.e., the CLS provided by the public and subscribers, but not from the number of posts provided by the corporations. Our findings confirm the increasing role of social media as a means of information dissemination, and the evidence of the efficient use of that information by sophisticated users such as financial analysts.

1. Introduction

Facebook has gradually become a popular media for corporate disclosures. Firms use Facebook for dissemination of information that encompass a wide variety of news that can include performance related news, product promotions, employee related news etc. Such disclosures may apparently seem unrelated to performance, however, to a discerning stakeholder such disclosures may provide ample source for deducing future performance.

In this study, we examine Facebook posts of firms on S&P 500 index that maintained Facebook pages. We analyze whether the information disclosed on corporate Facebook pages have any impact on financial analysts' forecast accuracy. Our results show analyst forecast error decrease significantly with the amount of information revealed on Facebook. Our results also highlight that posts made by companies that generate high reactions tend to have higher impact on improving analysts' forecast performance. Studies have shown that most of the reactions on corporate Facebook pages come from retail investors. Our findings suggest that the information retail investors find more useful, as indicated by their reactions, is the one that is instrumental in helping analysts making more accurate forecasts.

One key aspect of social media disclosure is that the information itself may not be novel to the market. Companies generally disclose publicly available information through their social media pages. The dynamic connectedness of social media platform makes the information disclosed more attractive to the subscribers. The interactive nature of social media highlights information that others have considered important, and thus contribute to faster dissemination and quicker impounding of information in market response. Moreover, disclosures in traditional media are aimed at a broader audience. Investors need to sift through volumes of information to make informed decision. But, social media has made it easier for investors to follow certain companies and be on the forefront of receiving company relevant information at real time and at practically no cost. Companies are increasingly adopting social media to reach relevant audience, who actively seek information about the company. Disclosure frequency and the content of the message both can play integral roles in shaping market sentiment. Existing literature provides evidence of social media disclosures influencing market reactions in the form of price movements and narrowing bid-ask spread. However, there is no evidence of how firm initiated information disseminated through social media affects the behavior or performance of expert information intermediaries. Thus, an investigation of how information disclosed through social media affect the performance of information intermediaries is imperative. Our study makes direct contribution to the nascent literature of social media disclosures and also contributes to the disclosure literature is general. Our results add to the literature by showing that dissemination of information through social media helps expert information intermediaries to make better forecasts. This finding is important because it identifies the type of information that makes the market more information efficient.

The rest of the paper is organized as follows: section 2 discusses background and literature, section 3 discusses the hypotheses, section 4 elaborates the methodology, section 5 explains the results, and lastly, section 6 provides conclusion.

2. Background and literature

Corporate disclosures on Facebook are different from disclosures through traditional media. Traditional media outlets, such as newspaper, websites, blogs, RSS feed etc., do not have the dynamic platform offered by Facebook. People can interact with the company and other people on the Facebook page. A simple ‘like’, ’share’, or ‘comments’ make it available to friends and sometimes friends of friends. This dynamic connectivity makes disclosures on Facebook unique.

Posts on corporate Facebook pages are generally open to public, however, new posts are delivered instantly to someone’s newsfeed only if that person subscribes or follows that page. People who are interested in following certain companies will follow the Facebook pages of those companies and as soon as new information is made available to those pages, they will be able to see new posts on their newsfeed. Interactions (likes, shares, or comments) on Facebook pages by the followers make the post available to their circle of friends. These interactions can change overtime, people can unlike, delete comments, unfollow the company, or even disable their Facebook profiles. Hasan & Wang (2016) explored this issue and their findings show that shares on average tend to go down over time, but likes and comments tend to remain fairly stable over time. We anticipate such biases in interaction variables are not likely to pose big issues for our analyses, since any subsequent decrease in shares will go against our finding of association between corporate Facebook activities and related analyst forecast accuracy.

Generally, a significant portion of Facebook post interactions tend to happen fairly quickly. A study by Optimal Social (Brand Networks as of October 2013) showed that 75% of Facebook engagement or interactions happen within the first three hours[1]. This has two implications for corporate Facebook pages; first, the ability of the post to go viral or create buzz and to move the market may happen fairly quickly, and second, the subsequent changes in these interactions may not be too important from the market reaction perspective.

[1] http://venturebeat.com/2013/03/28/75-of-facebook-engagement-is-in-the-first-180-minutes-says-facebook-competition-winning-tool/

Research on social media's impact on capital market and/or its' participant is limited. Studies have shown that social media-based activities affect stock market behavior. Luo, Zhang, and Duan (2013) use a sample of 9 firms in computer hardware and software industry and find that social media-based activities such as web-blogs and consumer ratings are leading indicators of firm equity value and have stronger predictive value than conventional online consumer behavioral metrics. Luo and Zhang (2013) use a similar sample of these 9 tech firms and indicate that consumer buzz and traffic in social media explain a substantial portion of the total variance of firms' value. Lee, Huttun, and Shu (2015) find that social media disclosure related to recall announcements attenuate the negative price reactions. Changes in daily posting volume on internet message boards are associated with investors trading behaviors (Wysocki 1998; Antweiler & Frank, 2004). Blankespoor, Miller, and White (2014) show that social media dissemination is positively associated with liquidity for low visibility technology firms. They find that dissemination of firm-initiated news via Twitter reduces bid-ask spreads and increases depth, consistent with reduction in information asymmetry. However, there is limited evidence of how dissemination on social media affects the beliefs of sophisticated information intermediaries, such as analysts. Our study aims to examine the influence of information on corporate social media on analysts' beliefs and forecasts.

3. Hypotheses

Analysts are considered expert information intermediaries with abilities to make more precise inferences from a set of information. Subscribers on a corporate Facebook page may like, comment, or share the news of a "sale" and spread it across the network. But analysts are believed to be capable of making more precise predictions about the earnings impact of such posts. Analysts are also capable of incorporating the relative interest generated by such posts, from subscribers' reactions to the post, into their predictions. It is very much possible for the analysts to have come to similar conclusions regarding the information content of the posts from other public disclosures, however, Facebook uniquely provides information about post related interests shown by general public or subscribers. This unique aspect of social media disclosure offers the analysts valuable additional insights about future impacts of the posts.

We conjecture that if Facebook posts are able to revise the beliefs of analysts, it will help them get a better understanding of the company's future performance. Therefore, they will be able to make more accurate forecasts. Thus, analysts following companies that provide more posts on social media, will be able to make more accurate forecasts. Our first hypothesis stated in alternative form is,

Hypothesis 1: Corporate Facebook posts have negative association with forecast error.

Next, we address the aspect of subscriber reactions to such social media posts. We further assume that analysts will be able to extract more insights from this unique characteristic of social media information on post related reactions, as revealed through "likes", "comments", and "shares". Our second hypothesis stated in alternative form is,

Hypothesis 2: Reactions to corporate Facebook posts have negative association with forecast error.

Evidence in support of the above hypotheses can help us understand the type of information that can make the market more information efficient.

4. Methodology

Sample Selection

We first collect the analyst quarterly forecast data from the I/B/E/S database. The initial sample consists of all quarterly forecasts, 51,618,039 observations, from the I/B/E/S database from 2008 to 2012. We restrict our sample to the analysts' latest forecasts. After excluding the non-latest forecasts, 51,362,206 observations, the sample reduces to 255,733. We also exclude 78,101 forecasts that came after the actual announcement dates. The number of latest forecasts that are made before the announcement dates is 177,632. The Facebook information used in this study is hand-collected. We find the official Facebook links for the S&P 500 firms from corporate websites, and then collect the number of posts, comments, likes and shares information on each day for the period 2008 to 2012. After excluding the non-S&P 500 firms, and merging the I/B/E/S data with the Facebook dataset, the number of observations becomes 5,164. We collect the financial information of the firms from the COMPUSTAT quarterly database. Because of the unavailability of the financial information and the missing values of other forecast related variables, we had to delete 235 observations. Thus, our final sample consists of 4,929 quarterly forecasts for 436 firms for which we are able to collect all required information for our analysis.

[Insert Table 1 Here]

Model Specification

To test the research question of this study, we employ the following multivariate regression:

$$\begin{aligned} Forecast\_Error_{i\,t} &= \alpha + \beta_1\, Post\_\&\_CSL_{i\,t} + \beta_2\, Size_{i\,t} + \beta_3\, Leverage_{\,i\,t} + \beta_4\, ROE_{i\,t} \\ &+ \beta_5\, Book\_to\_Mkt_{i\,t} + \beta_6\, Loss\_Firm_{i\,t} + \beta_7\, Change\_in\_EPS_{i\,t} \\ &+ \beta_8\, No\_Analyst_{i\,t} + \beta_9\, Forecast\_Horizon_{i\,t} + \sum \gamma_j\, Quarter\_Dummy + \varepsilon_t \end{aligned}$$

The dependent variable of this study is $Forecast\_Error_{i\,t}$. We estimate the forecast errors by using two measures, forecast errors based on mean forecasts ($Forecasts\_Error\_Mean_{\,i\,t}$) and forecast errors based on median forecasts ($Forecasts\_Error\_Median_{\,i\,t}$). By following the literature (Brown, 1993; Schipper, 1991), we define the forecast errors as the absolute value of the percentage of errors, where subscripts *i* and *t* denote firm *i* and quarter *t*:

$$Forecasts\_Error_{i\,t} = \left|\frac{Actual_{i\,t} - Forecasted_{\,i\,t}}{Actual_{i\,t}}\right|$$

In this study, the variable of interest is $Post\_\&\_CSL_{i\,t}$, which captures the volume of information content on corporate Facebook pages. $Post\_\&\_CSL_{i\,t}$ is calculated as the average number of posts, comments, likes and shares on an official Facebook page for a specific firm during 15 days before the I/B/E/S reported quarterly forecast dates. We also disaggregate the information content into two parts, $Posts_{i\,t}$ and $CSL_{i\,t}$, and estimate the regression separately.

By following previous literature, we control for firm-specific economic factors and other determinants for analyst forecast errors. We control also for the size of the firm because firm size is closely related to the information environment and to the disclosures that influence analyst forecasts (Lang and Lundholm, 1996). Analyst forecasts are more accurate for larger firms (Bushan, 1990; Lys and Soo, 1995; Wiedman, 1996; Brown, 1997; Hope, 2003; and Lang et al., 2003). $Size_{i\,t}$ is measured as the natural logarithm of the firm's market capitalization.

In our model we also include leverage, profitability and company growth prospects, since these relate to the nature and complexity of company operations and to analyst incentives for gathering information about them. We estimate $Leverage_{\,i\,t}$ as the total long-term debt divided by total stockholder equity. The profitability variable is $ROE_{i\,t}$, which is calculated as net income divided by total stockholder equity. $Book\_to\_Mkt_{i\,t}$, the ratio of book value per share to market value per share, captures the growth prospects of the firm. Prior studies find that analyst forecasts are less accurate for loss firms (Brown, 2001; Ciccone 2001; Hwang et al., 1996). We control the effect of loss firms on forecast errors by including an indicator variable, $Loss\_Firm_{i\,t}$, which equals 1 if the firm reports negative earnings and 0 otherwise. Lang and Lundholm (1996) find that larger changes in earnings are associated with less accurate forecasts. To control the effect of earnings surprises on forecast errors, we include another variable, $Change\_in\_EPS_{i\,t}$, the absolute value of the changes in earnings per share from the previous quarter to the current quarter. All the financial control variables are measured based on quarterly reporting.

Lys and Soo (1995) suggest that the greater the number of analysts following a company, the more intense their competition and the higher the incentive to reduce forecast errors in the ongoing competition for forecast reputation. By following the literature, we include the variable, $No\_Analyst_{i\,t}$, which is the natural logarithm of the number of analysts following the firm throughout the quarter (Dhaliwal et al., 2012). Many researchers believe that the degree of forecast errors depends on the time horizon between the forecast dates and the actual announcement dates (Brown et al., 1987; Lys and Soo, 1995; Jaggi and Jain, 1998). Earlier forecasts are subject to greater uncertainty and yield over or under optimism. Forecasts made closer to the time of announcement are more likely to be accurate (Das and Saudaraga, 1998; Jacob et al., 1999; Duru and Reeb, 2002). We control for the effect of the forecast time horizon by including the variable $Forecast\_Horizon_{i\,t}$ in the model. $Forecast\_Horizon_{i\,t}$ is the number of days between announcement dates and forecast dates. In addition to these control variables, we include dummy variables for each quarter to control the time-fixed effect. We estimate the model by clustering the standard errors by firms. Appendix A summarizes the definitions of all variables and the data bases used to collect those variables.

5. Results

Sample Distributions

Table 2 reports the distribution of the sample by year. Panel A shows that in 2011 the number of observations was 1,202 (24.30%) of the sample. 2012, with 543 (11.02%) of the sample, had the least number of observations. Panel B presents the volume of information content on Facebook pages per day. Although Facebook was launched in 2004, its use was limited to students at Harvard University. In September 2006, it became available to anyone aged 13 and over with a valid email address. The culture of the corporate Facebook page develops in 2007, although few companies shared information on their Facebook pages at that time. Panel B in Table 2 shows that the average number of posts and CLS per day in 2008 was 0. However, the average number of posts and CLS per day grew quickly in the later years of our sample period. For example, for 2011 and 2012 the average number of posts and CLS per day is 214.62 and 377.54, respectively, although the average number of posts per day throughout the sample years is less than one. In terms of volume, the information content comes primarily from the CLS provided by the public and stakeholders.

Panel C reports the sample distribution based on the SIC industry classifications. In our sample, manufacturing is the largest sector and contains 1,747 observations (35.44%) of the entire sample. The second largest sector is the wholesale and retail trade with 1,022 observations (20.73%), followed by transportation, communication and utilities with 820 observations (16.64%). The smallest sectors are agriculture, forestry and fishing, and the public administration sector, which have 17 and 19 observations, respectively. Panel D shows the data for the average Facebook content across industries. Transportation, communication and utilities have the highest number of average posts and CLS per day, 193.81, followed by the wholesale and retail trade sector with 178.97 posts and CLS per day.

[Insert Table 2 Here]

Descriptive Statistics

Table 3 presents the descriptive statistics of the key variables used in this study. The mean value of posts and CLS is 112.139 per day with a maximum value of 35,580. The standard deviation of posts and CLS is 951.625. The mean and standard deviation values of posts per day are 0.343 and 0.771. With a median value of 0, the distributions of all Facebook variables are highly skewed. The mean value of the analyst mean-estimates based forecast errors is 11.3%, and median-based forecast errors are 11.2%. The standard deviations of *Forecast_Error_Mean* and *Forecast_Error_Median* are 0.252 and 0.248, respectively. While the average number of analysts following a firm is 7.479, the minimum and maximum numbers are 1 and 41. The average forecast horizon is 283.19 days with a standard deviation value of 49.669. The minimum and maximum values of forecast horizon are 0 and 1,581 days.

The average size (natural logarithm of the market capitalization) of the firms in the sample is 9.09 and the standard deviation is 1.333. The minimum and maximum values of firm size are 3.415 and 12.509. The mean and standard deviation of the leverage values are 0.708 and 11.867. The mean and median values of *ROE* are 0.007 and 0.034, and the standard deviation is 1.01. The variable *Book_to_Mkt* has a mean of 0.60 and standard deviation of 0.892. The average value of the indicator variable for the loss firms is 0.12 and the median value for *Loss_Firm* is 0. The mean and standard deviation values of the variable *Change_in_EPS* are 0.625 and 2.308. The minimum and maximum values for *Change_in_EPS* are 0 and 111.01. All financial variables are winsorized at the 1% and 99% levels.

[Insert Table 3 Here]

Correlation Matrix

Table 5 reports the Pearson correlation values among the variables. All Facebook variables, *Post_&_CLS*, *Post* and *CLS* are highly correlated with each other. The correlation between *Post* and *CLS* is 0.809. The Facebook variables that capture the information volume on corporate Facebook pages are significantly negatively related to analyst forecast errors. These negative correlations provide the univariate supports for our predictions. The correlation value between *Post_&_CLS* and *Forecast_Error_Mean* is -0.111. *Forecast_Error_Mean* and *Post* is significantly correlated with a correlation value of -0.079. *Forecast_Error_Median* is significantly correlated with *Post_&_CLS*, *Post* and *CLS* with correlation values of -0.113, -0.084 and -0.113, respectively. This negative correlation implies that forecast errors decrease with the volume of information content on Facebook. The Facebook variables are significantly and positively related to firm *Size*, and negatively correlated with *Book_to_Mkt*. However, there is no correlation between Facebook variables and *Leverage*, *ROE*, *Loss_Firm* and *Change_in_EPS*. Facebook information content is positively correlated with *No_Analyst* but unrelated to *Forecast_Horizon*.

Analyst forecast errors are positively correlated with *Book_to_Mkt*, *Loss_Firm* and *Change_in_EPS*. For loss firms and firms with more earnings surprises, forecast errors are higher. For large firms, analysts forecast errors are lower, as they are for firms followed by a higher number of analysts. The correlation between *Forecast_Error_Mean* and *Size* is -0.052, and the correlation between *Forecast_Error_Mean* and *No_Analyst* is -0.144. Firm size is negatively and significantly correlated with *Book_to_Mkt*, *Loss_Firm* and *Change_in_EPS*, and positively correlated with *ROE*, *No_Analyst* and *Forecast_Horizon*. *ROE* is negatively correlated with *Book_to_Mkt*, *Loss_Firm* and *Change_in_EPS*. *Book_to_Mkt* is positively correlated with *Loss_Firm*, *Change_in_EPS* and *Forecast_Horizon*, and positively related to *No_Analyst*.

Loss_Firm is negatively related to *No_Analyst* and *Forecast_Horizon. No_Analyst* is positively related to *Forecast_Horizon.*

[Insert Table 4 Here]

Regression Results

Table 5 presents the multivariate regression results. The explanatory variable in Panel A is *Post_&_CLS.* The first column in Panel A shows the regression results where the forecast errors are calculated based on the mean forecasts. The variable *Post_&_CLS* is significantly and negatively associated with *Forecast_Error_Mean*. The value of the coefficient is -0.0062 with a *t*-stat value -3.02. This finding supports our hypothesis that analyst forecast errors decrease with the volume of information content on corporate Facebook pages. Analysts are using Facebook information and interpreting such information to provide more accurate forecasting. *Forecast_Error_Mean* is positively correlated with firm size and *Loss_Firm*, and negatively associated with *No_Analyst*. The model's adjusted R^2 value is 11.26%. Table 5, Panel A, also reports the regression model where the dependent variable is *Forecast_Error_Median.* The results are consistent with column one, and provide additional support for our analysis that the forecast errors based on the median forecast estimates decline with the contents on Facebook pages. The coefficient value on *Forecast_Error_Median* is -0.0058 with a *t*-stat value -3.12. The adjusted R^2 value of the second model is 8.04%.

We then disaggregate our variable of interest, *Post_&_CLS,* into two parts, *Post* and *CLS*, and re-estimate the regression models separately. Panel B shows the regression results for the Facebook variable *Post*. *Post* variable identifies the corporation provide information on Facebook. The regression results show that there is no significant association between the dependent variables, *Forecast_Error_Mean* and *Forecast_Error_Median*, and *Post*. This finding implies that financial analysts may not gather incremental information from Facebook posts to help improve their forecast accuracy. One possible explanation for this finding is that Facebook post information is available already to financial analysts. The analysts could collect the information from other sources before the companies post their information on social media. It is also possible that, because of the Regulation FD, corporations would not provide material information to the users of the Facebook. Therefore, the information in the posts does not add incremental value to the information environment for the analysts.

Panel C in Table 5 reports the regression results where the key explanatory variable is *CLS*. The variable *CLS* captures the Facebook user reactions to corporate posts. The information content in *CLS* is generated by the public and key stakeholders such as investors and customers. The results for both regression models in Panel C show that there are significant negative associations between *CLS* and forecast errors. This finding implies that financial analysts use the information from public and stakeholder reaction and incorporate those into their forecasting decisions. The information in Facebook posts varies from promotional sales to acquisition and merger-related key decisions, and it appears that Facebook, a vast and interactive platform, engages the public and stakeholders in key corporate decisions. Stakeholder reactions to various corporate decisions are often critical, adding value to the information environment for financial analysts. The more CLS, the richer the information content to help financial analysts reduce forecasting errors.

[Insert Table 5 Here]

We do additional analysis to confirm that our results are not biased by the selection of the time-window for Facebook information before the analyst forecast dates. In our main analysis, we consider the Facebook information content for 15 days before the forecast dates. In Table 6, we change the time-window to a 30 day period before the announcement dates. The results are consistent with the original findings. The first column in Table 6 shows that *Post_&_CLS* is significantly and negatively associated with *Forecast_Error_Mean*. The value of the coefficient is -0.0063 with a *t*-stat value -2.92. This finding supports our hypothesis. We find the similar results using *Forecast_Error_Median* as the dependent variable. The findings for the control variables are also consistent with the findings in Table 5. We do the same analysis altering the time-window for Facebook information to 45 days and 7 days before the forecast dates. All regression results (un-tabulated) provide supports for our hypothesis. These consistent findings confirm that the results are not biased by the selection of the time-window for Facebook information content, and financial analysts look continuously for information on social media and incorporate them into their forecast decisions.

[Insert Table 6 Here]

6. Conclusion

This study provides the first evidence on whether financial analysts use information on social media for company forecasting. We focused primarily on the information content in corporate Facebook pages, since Facebook is the largest social network with more than 1.7 billion active users worldwide. We hand-collected information on posts, comments, likes and shares for companies in the S&P 500, and measured the information content in their Facebook pages. By using the data of 4,929 quarterly forecasts from 2008 to 2012, we find that there is a significant negative association between analyst forecast errors and the volume of information on corporate official Facebook pages. We also separate the Facebook contents into two parts – posts and CLS – and find that financial analysts use the information in CLS provided by the public and company stakeholders. On the other hand, we found no evidence that corporate posts provide analysts with information that aids more accurate forecasting.

Our study has some limitations. First, we focus on the S&P 500 firms, all of whose companies are large and already in a rich information environment. The findings may differ for small or mid-sized firms. Second, the period of our sample, 2008 to 2012, coincides with the early period of the corporate Facebook culture, and many firms in our sample did not have a Facebook presence in early period of the sample. Although the number of Facebook users grew rapidly from 2008, frequent posts on company Facebook pages did not appear until later. A more recent dataset on Facebook content may provide better results for our analysis. Third, a further limitation of our study is that we did not classify the type of information in the posts. Future studies could do so, paying particular attention to the comments and reactions that stakeholders have about corporate postings, and examining whether they influence the revision of analyst forecasts.

Overall, our study provides evidence of the increasing importance of social media as an effective and interactive channel for providing and circulating corporate information for sophisticated users such as financial analysts.

References:

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

Blankespoor, E. Miller, G., and White, H. 2014. The role of dissemination in market liquidity: evidence from firms' use of Twitter. *The Accounting Review*, 89(1), 79-112.

Brown, L. 1993. Earnings forecasting research: Its implications for capital market research. *International Journal of Forecasting*, 9, 295-320.

Brown, L. 1997. Analyst forecasting errors: Additional evidence. *Financial Analysts Journal*, November-December, 81-88.

Brown, L. 2001. A temporal analysis of earnings surprises: Profits vs. losses. *Journal of Accounting Research,* 39, 221-241.

Brown, L. D., G. Richardson, and S. Schwager. 1987. An information interpretation of financial analyst superiority in forecasting earnings. *Journal of Accounting Research*, 25, 49-67.

Bushan, R. 1990. Firm characteristics and analysts following. *Journal of Accounting and Economics*, 13, 2, 255-274.

Ciccone, S. 2001. Analyst forecast properties, financial distress, and business risk. Working paper, University of New Hampshire.

Das, S., and S.M. Saudaragan. 1998. Accuracy, bias, and dispersion in analysts' earnings forecasts: The case of cross-listed foreign firms. *Journal of International Financial Management and Accounting*, 9, 1, 16-33.

Dhaliwal, D. S., Radhakrishnan, S., Tsang, A., and Yang, Y. G. 2012. Nonfinancial disclosure and analyst forecast accuracy: International evidence on corporate social responsibility disclosure. *Accounting Review*, 87 (3), 723-759.

Duru, A. and D.M. Reeb. 2002. International diversification and analysts' forecast accuracy and bias. *The Accounting Review*, 77, 2, 415-433.

Hasan, R., Wang, W. 2016. Social network activity, financial performance, and equity value. *Working paper*.

Hope, O. 2003. Disclosure practices, enforcement of accounting standards, and analysts' forecast accuracy: An international study. *Journal of Accounting Research* 41 (May), 235–272.

Hwang, L., C. Jan, and S. Basu. 1996. Loss firms and analysts' earnings forecast errors. *Journal of Financial Statement Analysis*. 1 (Winter), 18-31.

Jacob, J, T. Z. Lys, and M. Neale. 1999. Expertise in forecasting performance of security analysts. *Journal of Accounting Economics*, 28, 51-82.

Jaggi, J, and R. Jain. 1998. An evaluation of financial analysts' earnings forecasts for Hong Kong firms. *Journal of International Financial Management and Accounting*, 9, 3, 177-200.

Lang, M., and R. Lundholm. 1996. Corporate disclosure policy and analyst behavior. *The Accounting Review*, 71 (4), 467–493

Lang, M., K. Lins, and D. Miller. 2003. ADRS, analysts, and accuracy: Does cross-listing in the United States improve a firm's information environment and increase market value? *Journal of Accounting Research*, 41 (2), 317–345.

Lee, F., Hutton, A., and Shu, S. 2015. The role of social media in the capital market: evidence from consumer product recalls. *Journal of Accounting Research*, 53(2), 367-404.

Luo, X., & Zhang, J. (2013). How do consumer buzz and traffic in social media marketing predict the value of the firm? *Journal of Management Information Systems*, *30*(2), 213-238.

Luo, X., Zhang, J., and Duan, W. 2013. Social media and firm equity value. *Information Systems Research*, 24(1), 146–163.

Lys, T., and L. Soo. 1995. Analysts' forecast precision as a response to competition. *Journal of Accounting, Auditing and Finance*, 10 (4), 751–765.

Schipper, K. 1991. Commentary on analysts' forecasts. *Accounting Horizons* 5 (December), 105–121.

Wiedman, C. 1996. The relevance of characteristics of the information environment in the selection of a proxy for the market's expectations for earnings: An extension of Brown, Richardson & Schwager, *Journal of Accounting Research,* 34, 2, 313- 324.

Wysocki, P. 1998. Cheap talk on the Web: The determinants of postings on stock message boards. *Working paper. University of Michigan.*

Appendix A

Variable	Definition	Data Source
Dependent Variables		
Forecast_Error_Mean	Percentage of forecast errors calculated as the difference between actual values and mean forecasts, divided by the actual values	I/B/E/S*
Forecast_Error_Median	Percentage of forecast errors calculated as the difference between actual values and median forecasts, divided by the actual values	I/B/E/S
Explanatory Variables		
Post_&_CLS	Natural logarithm of average number of posts, comments, likes and shares in an official Facebook page for a specific firm during the 15 days before the quarterly forecast dates	Hand-collect
Post	Natural logarithm of average number of posts in an official Facebook page for a specific firm during the 15 days before the quarterly forecast dates	Hand-collect
CLS	Natural logarithm of average number of comments, likes and shares in an official Facebook page on corporate posting for a specific firm during the 15 days before the quarterly forecast dates	Hand-collect
Control Variables		
Size	Natural logarithm of the market capitalization of the firm at the end of the quarter	CRSP and Compustat**
Leverage	Total long-term debt divided by total stockholders' equity	Compustat
ROE	the net income divided by total stockholders' equity	Compustat
Book_to_Mkt	Book value per share and market value per share	CRSP and Compustat
Loss_Firm	An indicator variable equals 1 if the firm reports negative earnings, and 0 otherwise	Compustat
Change_in_EPS	Absolute value of the changes in earnings per share from the previous quarter to current quarter	Compustat
No_Analyst	Natural logarithm of the number of analysts that follow the firm throughout the quarter	I/B/E/S
Forecast_Horizon	Number of days between the actual announcement dates and forecast dates	I/B/E/S

* We collect the forecast data from I/B/E/S Summary Statistics.
** Financial variable information has been collected from Compustat Quarterly.

Table 1

Sample Selection Procedure

This Table reports the data collection and sample selection procedures. Analysts' forecasts data have been collected from the IBES database. All financial information are from the Compustat. Facebook information is hand-collected. The sample period is from 2008 to 2012. Final sample is restricted to S&P 500 firms.

Criteria	**Number of Observation**
Initial sample consists of all the quarterly forecasts from the IBES database for the years from 2008 to 2012	51,618,039
Less, Number of observations deleted for the exclusion of non-latest forecasts	(51,362,306)
Number of observations after deleting the non-latest forecasts	255,733
Less, Number of observations deleted for the exclusion of the forecasts that were made after the actual announcements	(78,101)
Number of observations after deleting the forecasts made after the actual announcement dates	177,632
Less, Number of observations deleted for non-S&P 500 firms and for the data unavailability of the Facebook information	(172,468)
Number of observations after merging the IBES and Facebook dataset	5,164
Less, Number of observations deleted for the unavailability of the financial information in the Compustat data base and for other forecast related missing values in IBES database	(235)
Final Sample [a]	**4,929**

[a.] Final Sample consists of 4,929 quarterly forecasts for 436 firms from S&P 500 for which all required information for the forecast related, financial and Facebook variables for the period from 2008 to 2012 are available.

Table 2

Panel A Year-Wise Sample Distribution

This table presents the distribution of the final sample used in this study. The sample consists of 436 firms' 4,929 quarterly forecasts for the period 2008 to 2012. The sample is restricted to the S&P 500 firms. The sample excludes the non-latest forecasts and forecasts made after the announcement date. Panel A reports the distribution of the sample by year. Panel B presents the average number of Facebook information content per day over the years. Panel C represents sample distribution by one-digit SIC industry classifications. Panel D shows the average number of Facebook information content per day across the industries.

Year	Frequency	Percentage	Cumulative Percentage
2008	885	17.95%	17.95%
2009	1124	22.80%	40.76%
2010	1175	23.84%	64.60%
2011	1202	24.39%	88.98%
2012	543	11.02%	100.00%
Total	**4929**	**100.00%**	

Panel B Year-Wise Average Facebook Information Content Per Day

Year	Frequency	Average Number of Post & CLS Per Day	Average Number of Post Per Day	Average Number of CLS Per Day
2008	885	0.00	0.00	0.00
2009	1,124	8.89	0.06	8.82
2010	1,175	67.88	0.34	67.53
2011	1,202	214.62	0.64	213.98
2012	543	377.54	0.80	376.74
Total	**4,929**			

Table 2

Panel C Industry-Wise Sample Distribution

Industry	SIC Codes	Frequency	Percentage	Cumulative Percentage
Agriculture, Forestry & Fishing	0100 - 0999	17	0.34%	0.34%
Mining and Constructions	1000 - 1799	172	3.49%	3.83%
Manufacturing	2000 - 3999	1,747	35.44%	39.27%
Transportation, Communication and Utility	4000 - 4999	820	16.64%	55.91%
Wholesale and Retail Trade	5000 - 5900	1,022	20.73%	76.64%
Finance, Insurance and Real Estate	6000 - 6700	579	11.75%	88.39%
Services	7000 - 8999	553	11.22%	99.61%
Public Administrations	9100 - 9999	19	0.39%	100.00%
Total		**4,929**	**100.00%**	

Panel D Industry- Wise Average Facebook Information Content Per Day

Industry	SIC Codes	Frequency	Average Number of Post & CLS Per Day	Average Number of Post Per Day	Average Number of CLS Per Day
Agriculture, Forestry & Fishing	0100 - 0999	17	4.74	0.20	4.55
Mining and Constructions	1000 - 1799	172	3.29	0.16	3.13
Manufacturing	2000 - 3999	1,747	100.14	0.26	99.88
Transportation, Communication and Utility	4000 - 4999	820	193.81	0.48	193.33
Wholesale and Retail Trade	5000 - 5900	1,022	178.97	0.47	178.49
Finance, Insurance and Real Estate	6000 - 6700	579	9.11	0.20	8.91
Services	7000 - 8999	553	51.90	0.37	51.53
Public Administrations	9100 - 9999	19	70.29	0.31	69.98
Total		**4,929**			

Table 3

Descriptive Statistics

This table presents the descriptive statistics of the key variables used in this study. The sample consists of 436 firms' 4,929 quarterly forecasts for the period 2008 to 2012. The sample is restricted to the S&P 500 firms. The sample excludes the non-latest forecasts and forecasts made after the announcement date.

Variable	N	Mean	Std. Dev	Minimum	Lower Quartile	Median	Upper Quartile	Maximum
Facebook Variables								
Post_&_CLS	4929	112.139	951.629	0.000	0.000	0.000	5.733	35580.130
Post	4929	0.343	0.771	0.000	0.000	0.000	0.400	11.867
CLS	4929	111.796	951.464	0.000	0.000	0.000	5.267	35579.600
(Log) Post_&_CLS	4929	1.225	2.003	0.000	0.000	0.000	1.907	10.480
(Log) Post	4929	0.214	0.352	0.000	0.000	0.000	0.336	2.555
(Log) CLS	4929	1.195	2.001	0.000	0.000	0.000	1.835	10.480
Forecast Variables								
Forecast_Error_Mean	4929	0.113	0.252	0.000	0.024	0.059	0.127	11.336
Forecast_Error_Median	4929	0.112	0.248	0.000	0.025	0.059	0.126	11.336
Number_of _Analysts	4929	7.479	6.168	1.000	3.000	6.000	10.000	41.000
(Log) No_Analyst	4929	1.906	0.684	0.693	1.386	1.946	2.398	3.738
Forecast_Horizon	4929	283.194	49.669	0.000	279.000	287.000	294.000	1581.000
Financial Variables								
Size	4929	9.0944302	1.3335208	3.4156445	8.287347	9.094788	9.8726592	12.50966
Leverage	4929	0.7086107	11.867096	-316.1235	0.232466	0.526171	1.062383	329.42225
ROE	4929	0.0074477	1.0102535	-52.73511	0.017084	0.034478	0.055769	14.589744
Book_to_Mkt	4929	0.6007091	0.892143	-31.15204	0.290598	0.496379	0.7991454	12.626033
Loss_Firm	4929	0.1203084	0.3253548	0	0	0	0	1
Change_in_EPS	4929	0.6257963	2.3083629	0	0.08	0.2	0.48	111.01

Table 4

Correlation Matrix

This table presents the Pearson correlation values among the key variables used in this study. The sample consists of 436 firms' 4,929 quarterly forecasts for the period 2008 to 2012. The sample is restricted to the S&P 500 firms. The sample excludes the non-latest forecasts and forecasts made after the announcement date.

	Post_&_CLS	***Post***	***CLS***	***Forecast_Error_Mean***	***Forecast_Error_Median***	***Size***	***Leverage***	***ROE***	***Book_to_Mkt***	***Loss_Firm***	***Change_in_EPS***	***No_Analyst***
Post_&_CLS												
Post	0.822***											
CLS	0.999***	0.809***										
Forecast_Error_Mean	-0.111***	-0.079***	-0.111***									
Forecast_Error_Median	-0.113***	-0.084***	-0.113***	0.989***								
Size	0.153***	0.039***	0.157***	-0.052***	-0.054***							
Leverage	0.013	0.007	0.015	0.004	0.003	-0.001						
ROE	0.007	-0.010	0.009	-0.010	-0.011	0.034**	0.650***					
Book_to_Mkt	-0.082***	-0.054***	-0.082***	0.102***	0.103***	-0.124***	0.019	-0.003				
Loss_Firm	-0.005	0.0175	-0.004	0.141***	0.143***	-0.243***	0.024*	-0.039***	0.103***			
Change_in_EPS	0.003	0.041***	0.003	0.081***	0.082***	-0.093***	0.001	-0.050***	0.037***	0.198***		
No_Analyst	0.225***	0.165***	0.225***	-0.144***	-0.145***	0.357***	0.002	0.019	-0.095***	-0.098***	-0.069***	
Forecast_Horizon	0.011	-0.003	0.011	0.006	0.006	0.061***	-0.005	-0.005	0.046***	-0.036**	-0.055***	0.115***

***, **, * Indicates the statistical significance of the correlations among the variables at the 0.01, 0.05, and 0.10 levels, respectively, based on a two-tailed test.
All financial variables are winsorized at the 1% and 99% levels.

Table 5

Panel A Multiple Regression of Analysts' Forecast Errors on Facebook Information Content (Post & CLS)

This table reports the results of the multiple regressions of analysts' forecast errors on the average of Facebook information content for 15 days before the forecast dates. The interest variable in this table is the average number of total Facebook posts and CLS (Comments, Likes and Shares). The sample consists of 436 firms' 4,929 quarterly forecasts for the period 2008 to 2012. The sample is restricted to the S&P 500 firms. The sample excludes the non-latest forecasts and forecasts made after the announcement date. Statistically significant variables are highlighted as bold.

	Dependent Variable = *Forecast_Error_Mean*			**Dependent Variable = *Forecast_Error_Median***		
Variable	**Parameter**	**t Value**	**Pr > |t|**	**Parameter**	**t Value**	**Pr > |t|**
Intercept	0.01039	0.19	0.852	0.01169	0.21	0.833
Post_&_CLS	**-0.00622**	**-3.02**	**0.003**	**-0.00588**	**-3.17**	**0.002**
Size	**0.00929**	**1.65**	**0.100**	0.00900	1.58	0.114
Leverage	0.00004	0.18	0.858	0.00005	0.25	0.804
ROE	-0.00059	-0.23	0.821	-0.00081	-0.30	0.768
Book_to_Mkt	0.01814	1.52	0.130	0.01808	1.52	0.131
Loss_Firm	**0.08247**	**3.46**	**0.001**	**0.08257**	**3.43**	**0.001**
Change_in_EPS	0.00517	1.50	0.134	0.00511	1.49	0.136
No_Analyst	**-0.04025**	**-2.49**	**0.013**	**-0.03925**	**-2.43**	**0.016**
Forecast_Horizon	0.00009	1.50	0.135	0.00009	1.51	0.132
Firm Fixed-Effect	Yes			Yes		
Quarter Fixed Effect	Yes			Yes		
Adjusted R^2	11.26%			8.04%		
N	4,929			4,929		

Panel B Multiple Regression of Analysts' Forecast Errors on Facebook Information Content (Post)

This table reports the results of the multiple regressions of analysts' forecast errors on the average of Facebook information content for 15 days before the forecast dates. The interest variable in this table is the average number of Facebook posts. The sample consists of 436 firms' 4,929 quarterly forecasts for the period 2008 to 2012. The sample is restricted to the S&P 500 firms. The sample excludes the non-latest forecasts and forecasts made after the announcement date. Statistically significant variables are highlighted as bold.

	Dependent Variable = *Forecast_Error_Mean*			**Dependent Variable = *Forecast_Error_Median***		
Variable	**Parameter**	**t Value**	**Pr > |t|**	**Parameter**	**t Value**	**Pr > |t|**
Intercept	0.01849	0.33	0.740	0.01940	0.35	0.727
Post	-0.01117	-0.72	0.472	-0.01275	-0.81	0.418
Size	0.00847	1.48	0.139	0.00821	1.44	0.152
Leverage	0.00003	0.16	0.872	0.00005	0.24	0.814
ROE	-0.00058	-0.22	0.824	-0.00082	-0.30	0.765
Book_to_Mkt	0.01858	1.56	0.119	0.01847	1.56	0.121
Loss_Firm	**0.08076**	**3.38**	**0.001**	**0.08108**	**3.36**	**0.001**
Change_in_EPS	0.00515	1.50	0.134	0.00510	1.50	0.135
No_Analyst	**-0.04150**	**-2.53**	**0.012**	**-0.04036**	**-2.48**	**0.014**
Forecast_Horizon	0.00010	1.63	0.104	0.00010	1.62	0.106
Firm Fixed-Effect	Yes			Yes		
Quarter Fixed Effect	Yes			Yes		
Adjusted R^2	7.65%			7.20%		
N	4,929			4,929		

Panel C Multiple Regression of Analysts' Forecast Errors on Facebook Information Content (Comments, Likes and Shares)

This table reports the results of the multiple regressions of analysts' forecast errors on the average of Facebook information content for 15 days before the forecast dates. The interest variable in this table is the average number of CLS (Comments, Likes and Shares) on the Facebook posts. The sample consists of 436 firms' 4,929 quarterly forecasts for the period 2008 to 2012. The sample is restricted to the S&P 500 firms. The sample excludes the non-latest forecasts and forecasts made after the announcement date. Statistically significant variables are highlighted as bold.

	Dependent Variable = *Forecast_Error_Mean*			**Dependent Variable = *Forecast_Error_Median***		
Variable	**Parameter**	**t Value**	**Pr > |t|**	**Parameter**	**t Value**	**Pr > |t|**
Intercept	0.00966	0.17	0.862	0.01109	0.20	0.842
CLS	**-0.00643**	**-3.04**	**0.003**	**-0.00601**	**-3.25**	**0.001**
Size	**0.00937**	**1.66**	**0.091**	0.00907	1.59	0.112
Leverage	0.00004	0.19	0.853	0.00005	0.25	0.800
ROE	-0.00058	-0.22	0.824	-0.00080	-0.29	0.771
Book_to_Mkt	0.01811	1.52	0.130	0.01806	1.51	0.131
Loss_Firm	**0.08260**	**3.46**	**0.001**	**0.08267**	**3.44**	**0.001**
Change_in_EPS	0.00518	1.50	0.133	0.00511	1.49	0.136
No_Analyst	**-0.04018**	**-2.49**	**0.013**	**-0.03920**	**-2.43**	**0.016**
Forecast_Horizon	0.00009	1.50	0.136	0.00009	1.51	0.132
	0.00966	0.17	0.862	0.01109	0.20	0.842
Firm Fixed-Effect	Yes			Yes		
Quarter Fixed Effect	Yes			Yes		
Adjusted R^2	11.29%			8.05%		
N	4,929			4,929		

Table 6

Multiple Regression of Analysts' Forecast Errors on Facebook Information Content (Post & CLS) for 30 Days Before the Forecast Dates

This table reports the results of the multiple regressions of analysts' forecast errors on the average of Facebook information content for 30 days before the forecast dates. The interest variable in this table is the average number of total Facebook posts and CLS (Comments, Likes and Shares). The sample consists of 436 firms' 4,929 quarterly forecasts for the period 2008 to 2012. The sample is restricted to the S&P 500 firms. The sample excludes the non-latest forecasts and forecasts made after the announcement date. Statistically significant variables are highlighted as bold.

	Dependent Variable = *Forecast_Error_Mean*			**Dependent Variable = *Forecast_Error_Median***		
Variable	**Parameter**	**t Value**	**Pr > |t|**	**Parameter**	**t Value**	**Pr > |t|**
Intercept	0.01022	0.18	0.854	0.01158	0.21	0.835
Post_&_CLS	**-0.00630**	**-2.92**	**0.004**	**-0.00591**	**-3.1**	**0.002**
Size	**0.00931**	**1.65**	**0.100**	0.00902	1.59	0.114
Leverage	0.00004	0.18	0.855	0.00005	0.25	0.801
ROE	-0.00061	-0.23	0.816	-0.00083	-0.3	0.763
Book_to_Mkt	0.01813	1.52	0.130	0.01807	1.51	0.131
Loss_Firm	**0.08253**	**3.46**	**0.001**	**0.08261**	**3.44**	**0.001**
Change_in_EPS	0.00518	1.5	0.134	0.00511	1.49	0.136
No_Analyst	**-0.04026**	**-2.49**	**0.013**	**-0.03926**	**-2.43**	**0.016**
Forecast_Horizon	0.00009	1.49	0.137	0.00009	1.51	0.133
Firm Fixed-Effect	Yes			Yes		
Quarter Fixed Effect	Yes			Yes		
Adjusted R^2	11.26%			8.04%		
N	4,929			4,929		