Informed Investors and the Internet

Amir RUBIN and Eran RUBIN *

Abstract:

During the last decade the Internet has become an increasingly important source for gathering company related information. We employ Wikipedia editing frequency as an instrument that captures the degree in which the population is engaged with the processing of company-related information. We find that firms whose information is processed by the population more frequently are associated with lower analysts' forecast errors, smaller analysts' forecast dispersions, and significant changes in bid-ask spreads on analysts' recommendation days. These results indicate that information processing over the Internet is related to the degree to which investors and analysts are informed about companies.

JEL Classification: G14 G24

Keywords: Analysts, Dow Jones, information asymmetry, self efficacy, Internet, Wikipedia

Date: December 13, 2009

* Amir Rubin is from the Finance Area, Faculty of Business Administration, Simon Fraser University, Canada. Email: arubin@sfu.ca. Eran Rubin is from the Management Information System Division, Sauder Business School, University of British Columbia, Canada: Email Eran.Rubin@sauder.ubc.ca.

1. Introduction

It was almost two decades ago when the Internet was recognized as "the information superhighway". While there may be debates as to the source of the term, its adequacy is unequivocal. Since its introduction, the Internet has proven to revolutionize the way people learn, interact, and communicate. In the Internet, space is unlimited, censorship is almost inexistent, and individuals from all different backgrounds can have a voice. Hence, with the advent of the Internet, volumes of information can be accessed from any place at any time, raw data can be explored, and heterogeneous perspectives can be communicated.

Possessing such attributes, the Internet has proven to provide individuals with more and better information in many domains. In retail consumption, Internet technology was found to lower information asymmetries between customers and retailers (Bakos, 2001). The cost of acquiring price information, as well as information about the reputations of suppliers and retailers has reduced immensely compared to the pre-Internet era. In politics, the Internet has emerged as a tool that reduces the cost of acquiring and distributing information. Information that was not typically accessible prior to the Internet age is now available to the average person. For example, voting records and campaign donations, which were once the province of the elite, are now readily available online (DiMaggio, 2001). Finally, in the health care industry, patients today are much more informed about the details of their illness because of the Internet. Therefore, doctors too have to explore a much wider variety of options, provide a deeper layer of analysis, and read the most recent research in order to provide extra information that would help patients and doctors reach a decision about treatment (Anderson, 2004).

Similarly to other domains, the impact of the Internet on information processes in financial markets cannot be denied. In today's world, investors are free to browse the Securities and Exchange Commission website, Yahoo! Finance, and many other financial news and discussion sources to obtain as much information about companies as they choose; usually for free. Thus, with the abundance of information sources available to individuals, it would only seem natural that many investors become more informed.

In this paper we design empirical tests that allow us to investigate whether *processing* of information over the Internet is correlated with the degree to which different market participants are informed. In order to engage in such a task, it is essential to quantify the extent to which investors effectively use the Internet to gain information pertaining to the companies in which they invest – a task which we refer to as *Internet information processing*.

Many factors can influence one's ability to retrieve and comprehend company related information from the Internet. These include the availability of information on the web, the extent to which the information is easy to locate, and the extent to which individuals find it easy to engage with the information. According to a large body of Information Systems literature (e.g. Compeau and Higgins, 1995; Eastin and LaRose 2000; Hsu and Chiu, 2004), when one perceives his or her ability to gain company related information from the Internet to be low, one's use of the Internet for such purposes would be low as well.[1] I.e., deficiencies in Internet information processing would be reflected in low use of the Internet.

To proxy differences in Internet information processing across firms we analyze differences in the use of the Internet to gain company related information. Specifically we

[1] Formally, the information systems literature has defined the perception of one's ability to use the Internet effectively for a task as "Internet self efficacy".

analyze differences in the frequency with which Dow Jones Industrial (DJI) firms' entries on Wikipedia[2] are edited. Wikipedia is a web based encyclopedia, providing information on almost any topic, as well as links to related Internet information sources from which the relevant information was obtained. An important aspect of Wikipedia is its permissive editing policy: Any individual with access to the Internet can edit the information on a Wikipedia entry, contributing their knowledge on specific topics. By analyzing the frequency with which the DJI firms' entries are edited, we are able to quantify the degree to which company-related information is processed by the population. More frequent editing on Wikipedia would suggest more individuals engage and feel confident with company related information provided on the Internet.[3]

It is important to emphasize that we do not perceive the information provided on Wikipedia's firm-entries as important for pricing of securities. For our purposes, Wikipedia edit frequency serves as a proxy to quantify the extent to which the public is engaged with, and informed about, company related information.

We develop three hypotheses to analyze the correlation between Internet information processing and the degree to which investors are informed. Our first test uses analysts' EPS forecasts errors as the proxy for the information gap that exists between informed individuals and an individual with perfect foresight. Thus, we measure the deviation of analysts' forecasts from the reality that prevails based on earning announcements. Regressions explaining

---

[2] In a recent poll, Wikipedia came 9th most popular website on the Web amongst US households (see http://www.slipperybrick.com/2007/02/wikipedia-cracks-top-ten-most-popular-sites-in-us/).

[3] The advantage of Wikipedia editing frequency as opposed to other Internet usage measures is that editing is associated with the more sophisticated/informed individuals, while one does not really need to be an expert in order to view a webpage. Recently, Google has started an interface called Google Trends that allows the analysis of trends in search queries (Choi and Varian, 2009). The limitation of Google Trends with regard to the analysis of this paper is that it does not allow for distinguishing between queries on the company and queries on the company's products.

analysts' forecasts errors reveal that higher edit frequencies are associated with smaller forecast errors, suggesting that increased Internet information processing is associated with more informed analysts.

In a second test we analyze whether Internet information processing is correlated with the degree of information asymmetry amongst informed individuals. In this test we use dispersion between the different analysts’ forecasts as an asymmetry measure. Results reveal that higher frequency edits are correlated with smaller analysts' forecast dispersion, suggesting that increased Internet information processing is associated with a decrease in this type of information asymmetry.

Finally, we analyze whether increased Internet information processing is correlated with an increase in the portion of informed investors. Kim and Verrecchia (1994) show that the information gap between informed and liquidity traders tends to increase on days of public announcements. On these days, informed traders, who continuously process information pertaining to companies, are comparatively better than market makers and liquidity traders in appropriating on the new information. Therefore, the market maker tends to increase bid-ask spreads during these days. Further, as the market maker is aware of the process that determines the number of informed traders in the market, spreads should be larger when the market maker perceives that a large portion of the investor-base is informed. Our empirical findings show that changes in spreads during analysts’ recommendation days are positively correlated with edit frequency. This suggests that enhanced Internet information processing is associated with a relatively larger portion of informed investors.

Our study finds a relation between Internet information processing and outcomes in financial markets. However, we stress the preliminary nature of our analysis as Wikipedia entries provide various types of information, most of which are not associated with the pricing of securities. Thus, we do not claim for a causal relation between Wikipedia edit frequency and our information asymmetry measures because the former is merely a proxy that quantifies information processing over the Internet. Rather, we perceive Wikipedia edit frequency to be a novel way to measure the degree in which the public processes information pertaining to companies.

In terms of our analysis, because of the perceived lack of casual relation between Wikipedia edit frequency and outcomes in financial markets, we must be careful that our proxy (i.e., Wikipedia edits) and dependents are not spuriously related due to their relation to a third variable. A situation that makes intuitive sense is that higher levels of analysts' coverage may affect Wikipedia editing and at the same time lead to lower forecast errors, lower forecast dispersion, and changes in spreads on recommendation dates. We address this possibility in a two-stage least squares (2SLS) structural model where analyst coverage and information asymmetry are determined jointly. Our results remain unchanged, suggesting that Internet information processing is indeed related to information asymmetry in financial markets.

The main contribution of the paper is the linkage it provides between Internet usage and the degree of informed individuals following specific companies. There are two papers that are related to this study. Bogan (2008) shows that the increasing presence of computers and Internet usage has substantial effects on financial markets by increasing stock market participation levels in U.S. households. A key difference in our paper is the quantification of cross-sectional (firm-related) differences. Grullon et al. (2004) show that a firm's advertising expenditure is correlated with its liquidity. They claim that advertising attracts uninformed investors to buy the stock, and that this, in turn, increases liquidity. As opposed to advertising, the Internet has the capability of making individuals more informed. Hence, in some respect our analysis complements their work as we focus on informed individuals rather than on liquidity traders.

The remainder of the paper proceeds as follows. Section 2 reviews Wikipedia and its relevance for quantifying Internet information processing pertaining to specific companies. Section 3 presents our hypotheses. Section 4 describes the data. Section 5 presents the econometric analysis which includes the simultaneous modeling of analysts' coverage and the dependent variables of the paper. Section 6 concludes.

2. Wikipedia

Firms can differ in the extent to which information related to them is accessible on the Internet. For instance, firms are likely to provide different levels of transparency about their operations. Moreover, firms that are more attractive to the public are likely to be cited more often in the electronic media and more closely followed by investors and other stakeholders. Naturally, as company related information is easier to access, engage with, and process, the more likely it will be internalized and used. Internalization and use of company related information on the Internet is well captured on the Wikipedia website.

Wikipedia is an Internet-based, volunteer-contributed encyclopedia that has become a popular online reference since its launch in 2001. Wikipedia has thousands of international contributors and is presently the largest example of an open content wiki.[4] According to Lih (2004), the goal of Wikipedia is to create an encyclopedia that can be shared and copied freely while encouraging people to easily change and improve the content. Each and every entry has

[4] Wiki is the technology enabling easy editing of web content. The Hawaiian word for "quick" WikiWiki, is the basis for the name.

an *edit* button, allowing anyone, even anonymous passersby, to add or delete content on any page. What would surely seem to have created chaos has actually produced increasingly respected content that has been evaluated and revised over time by the site's thousands of visitors.

Wikipedia is guided by an ideology that promotes knowledge sharing. The content-based policies of Wikipedia require the contributor to provide information that is neutral and verifiable and prohibits original research (Hansen et al., 2007). These policies state that Wikipedia articles should present all significant facets or competing positions on a given subject in a way that is unbiased. Contributing editors should make efforts not to betray their personal preferences or opinions in presenting a topic.

The wiki concept is somewhat counterintuitive because the technical implementation itself provides no gatekeeping function to ensure the quality of the material being contributed. For example, no proof of identity or qualifications is needed to participate and a reputation tracking system is not used within the community. What allows this open editing system to function effectively is the ability to track the status of entries and review individual changes. Wikipedia tracks and stores every version of an entry, so no operation is ever permanently destructive. As a foil to malicious contributors, it takes much more effort to vandalize a page than to revert an entry back to a prior version; indeed, an inappropriate entry can be undone with just one click of a button. This crucial asymmetry tips the balance in favor of productive members of the wiki community and allows quality content to emerge and evolve.

An important attribute of a Wikipedia web page is the number of edits made on the page over time, or the frequency of the page's edits. In broad terms, more edits correspond to a higher level of public participation. In this respect, two well-cited determinants for high edit frequency on Wikipedia are higher levels of knowledge and higher confidence levels in using the Internet (e.g., Hansen et al., 2007; Kuznetsov, 2006; Ulrik and Jurgen, 2007). The reasons are quite intuitive – high confidence in the knowledge level is required to initiate a contribution. Similarly, confidence in using the Internet is required to be able to gain knowledge about the topics at hand. These two factors are especially important in the context of Wikipedia since contributions are constantly monitored- an inaccurate contribution would be quickly removed, rendering the contribution efforts as wasted.

Hence, Wikepdia appears to effectively capture Internet self efficacy (Compeau and Higgins, 1995; Hsu and Chiu, 2004). Positive experience in using the Internet to gain information pertaining to a specific company would result in high edit frequency on the Wikepdia entry and would suggest similar use of the Internet for information related to the company. Similarly, ineffective experience in using the Internet to gain information pertaining to a company would result in low edits on the Wikepdia entry and a reluctance to use the Internet to gain information pertaining to the company.

3. Hypotheses

3.1 Analysts' forecasts and analysts' forecasts dispersion

Our first hypothesis is that higher edit frequency is associated with more informed analysts. We test this hypothesis by relating edit frequencies to analysts' EPS forecast errors. As analysts can be considered relatively informed individuals, we are essentially conducting a test to see whether Internet information processing is related to information asymmetry between informed individuals and the reality that exists. Formally, our hypothesis is

H1: *Edit frequencies on firms' Wikipedia entries are negatively correlated with analysts' forecast errors.*

Our second hypothesis is based on a related aspect of analysts' forecast errors. We expect that with higher a Wikipedia edit frequency, different analysts will posses rather similar information sets, either because they have access to the same sources of information (the Internet), or because reputation effects and competition force them to invest more effort in the production of forecasts associated with companies the investors are more informed about. Analysts' forecast dispersion measures the information asymmetry that exists between the different analysts. Hence, our second hypothesis is that analysts' forecast dispersion is likely to decrease with the editing frequency of a firm's Wikipedia entry.

H2: *Edit frequencies on firms' Wikipedia entries are negatively correlated with analysts' forecast dispersions.*

3.2 Bid-Ask Spread on analysts' recommendation days

The typical information asymmetry model (e.g., Golsten and Milgrom, 1985; Kyle, 1985) assumes two types of traders: liquidity traders and potential information processors (or informed traders). When trading with informed traders, market makers sustain losses, which they recover through increased bid-ask spreads. In these models, information asymmetry increases as the proportion of informed traders rises. Consequently, if the market maker anticipates a greater probability of facing an informed trader, the spread widens.

In their theoretical model, Kim and Verrecchia (1994) show that informed traders have a comparative advantage when trading on public announcement days. They show that in the absence of announcements there are no opportunities for traders capable of informed judgments to exploit their ability to process public information. Contrary to that, during a period of an announcement information asymmetry increases. On these days informed traders have superior ability, compared to market makers and liquidity traders, to assess firms' performance on the basis of the new information.

Following this idea, we analyze the change in spreads on the day of analysts' recommendation. As information gaps between informed and liquidity traders tend to increase on announcement days, we expect to be able to quantify the portion of informed traders in the market by observing the change in spreads on an announcement day. If higher edit frequency is associated with reduced information gathering costs, increases in spreads during announcement days should be larger for high edit frequency firms (as a larger portion of the investor-base is expected to be informed).

We choose to focus on analysts' recommendation days (and not other announcements such as earning announcements, 10K filings reports, etc.) because according to regulation fair-disclosure, analysts' recommendations should be based only on publicly available information.[5] Other announcements could at least partially include "inside information".[6] Kim and Verrecchia (1994) noted that a caveat for the interpretation that spreads should increase on announcement days is that in some instances public disclosure could reduce information asymmetry problems by revealing to market makers information known only to corporate insiders. Thus, other announcements may include a component that reduces information asymmetry and therefore may have conflicting effects on spreads.

---

[5] Cornett et al. (2007) provide evidence that following fair disclosure the information content in affiliated analysts' recommendations is similar to that of non-affiliated analysts, suggesting that following the new regulation affiliated analysts do not have access to inside information not available to other analysts.

[6] For example, insiders may know (or have a better idea about) the company's quarterly earnings prior to the announcement itself.

Thus, we expect that higher edit frequency is associated with reduced information processing costs. These reduced costs should lead to a larger portion of investors who are informed about the company, which should in turn be associated with larger spread increases on analysts' recommendation days. Formally our hypothesis is

H3: *Edit frequencies on firms' Wikipedia entries are positively correlated with changes in bid-ask spreads on analysts' recommendation days.*

4. Data

4.1 Description of Major Variables

Our Wikipedia editing frequency measure is a monthly measure, based on the number of edits conducted on each DJI firm's Wikipedia entry, between July 2005 and December 2006. We use this sample period because some of the DJI firms' entries were not available in the first half of 2004 and firms' editing frequencies are incomparable during their early stage on Wikipedia.[7] Wikipedia data is kept in its original form for all years and can be obtained at any time by accessing the revisions page of the companies. Wikipedia makes this data available both for online browsing, and for application based analysis by downloading the data. Further, Wikipedia makes special efforts that editing data will not be tampered. It can be explored, monitored, looked at, and analyzed by anyone at any time. These properties of

[7] Upon initialization on Wikipedia, a firm's historical information is incorporated into its entry for the first time, links with other relevant entries are established, and new sub-entries are built. Thus, differences in editing frequencies during the initial stages of a firm's entry may be due to differences in, e.g., the volume of historical data available on the company, the number of links to their subsidiaries, and the number of links to their products. Indeed, Wilkinson and Huberman (2007), who analyze over 500 million Wikipedia edits, show that during the first 50 weeks of an entry's creation, there is abnormal variation in the mean and standard deviation of the entry's edits.

Wikipedia make it an appropriate source for academic research, not commonly provided by other sites.[8]

During our sample period, we observe a growing trend in the editing frequency of firms' entries. This is to be expected given the increased use of the Internet in general and of Wikipedia in particular (Voss, 2005). Thus, to eliminate any non-stationary problems, our primary dependent variable, ***Edits***, is defined as the number of Wikipedia updates for a DJI firm entry, adjusted to October 2006 levels based on the general growth of edits in Wikipedia (i.e., multiplied by the frequency of edits on Wikipedia during October 2006 and divided by the frequency of edits on Wikipedia during the respective month).[9]

Our initial sample of analyst forecast and actual earnings data come from I/B/E/S. As a proxy for the consensus forecast we use the median one-quarter-ahead earnings per share (EPS) forecast, which allows us to use the maximum number of forecast observations during the sample period. We use two measures of the accuracy of consensus forecasts and two measures of the dispersion among forecasts, employing only one observation per quarter to avoid the problem of stale forecasts. Specifically, our ***Closest forecast error*** measure is based on the consensus forecast closest to, but at least 20 days preceding, the date on which an actual earnings announcement is released. This measure is the absolute difference between the median earnings per share (EPS) forecast and the actual earnings, deflated by the stock price

[8] It should be noted that during our sample period the editing policy was permissive, under which anonymous users could edit the entries freely. This policy has changed over time, locking pages periodically, allowing only special administrators to approve edits.

[9] Both the non-trended and trended series are available from the authors upon request. Statistics about Wikipedia growth can be found at http://stats.wikimedia.org/EN/TablesWikipediaEN.htm. We reran all the specifications of our paper with the trended series of edits and all our results were qualitatively unchanged. We also employed other measures of edits, including unique edits (i.e., multiple edits by the same contributor are discarded) and a measure that discards edits that are flawed (reverted or deleted edits). All measures yield similar results and are highly correlated.

five days before the earnings announcement date.[10] ***Average forecast error*** is calculated similarly, but is based on all the median consensus forecasts during the quarter before, but at least 20 days preceding, the date on which actual earnings is released. Turning to the dispersion proxies, ***Closest dispersion*** is the standard deviation of analysts' EPS forecasts reported closest to, but at least 20 days preceding, the date on which actual earnings is released, and ***Average dispersion*** is the average standard deviation of analysts' forecasts during the period since the last EPS announcement, but at least 20 days preceding the date on which actual earnings is released. Both measures are deflated by the stock price five days before the earnings announcement date.[11] The short forecasting horizon is used for each of these four measures to minimize the optimism bias that appears to exist in forecasts made at the beginning of a fiscal year (e.g., O'Brien, 1988; Easterwood and Nutt, 1999), and to focus on firm-specific information rather than economy- or industry-wide information (Elton et al., 1984).

In order to study the information gathering costs, we use two types of variables. ***Revision*** is the difference between two consecutive analyst's recommendations (a number between -4 and 4), whereas ***Spread*** is the closing bid-ask relative spread (i.e., [closing ask – closing bid]/closing price), taken from the daily file of the CRSP data set.

4.2 Control Variables

[10] Brown and Pfeiffer (2008) highlight important biases if one deflates forecast errors by share price. However, the potential misleading effects of using share price as a deflator are associated with tests on abnormal returns, which we do not concern ourselves with.
[11] To preclude the possibility of flawed data, prior to the calculations of the closest and average variables, we eliminate the top 1% of forecast error estimates.

In order to examine our hypotheses, we control for several firm characteristics that have been previously associated with forecast errors and dispersion. We also control for firm recognition, which seems to be particularly important in our analyses.

A possible concern related to edit frequency may be that it is measuring a firm's recognition by the public rather than the degree to which individuals process the firm's information over the Internet. As was discussed in the introduction, availability of information on the web is likely to affect Internet information processing, so a reasonable claim is that the Internet media publishes more information on some firms compared to others because the public is more interested in those firms. This interpretation cannot be ruled out and it would probably be impossible to completely disentangle a firm's recognition from Internet information processing associated with it. The reasons are obvious – one would expect that with increased Internet information processing firm recognition would increase and vice-versa. However, the two constructs are still distinguishable. For example, individual ability of processing information over the Internet may decrease with the abundance of information available by causing increased confusion (commonly referred to as "information overload"), while recognition would tend to increase with more information. To rule out the possibility that our results are solely a recognition issue, we follow Baker et al. (2002) and we control for recognition using the ***Wall Street Journal*** variable, which is the number of times a firm is cited during the three months prior to the forecast or recommendation.[12]

A possible concern with a news based variable such as the *Wall Street Journal* that we utilize is that it may under represent firms that lack problems or success stories (as news tends to typically be associated with problems or successes). We therefore follow Grullon (2004) et

[12] Baker et al. (2002) also use the number of analysts as a measure for investors' recognition. As all our dependent variables are based on analysts' forecasts and recommendations, we would be concerned with endogeneity issues when using the number of analysts as a control. This is further addressed in Section 5.4.

al. and also include as a control the variable ***Advertising***, defined as the total advertising expenditure divided by total sales. The intuition behind this variable is that firms with larger advertising expenses tend to be more recognized by the public.

An obviously important determinant of forecast errors is volatility. As volatility increases, the amount of relevant information that analysts must process increases, and analysts' ability to forecast earnings declines. Thus, firms with higher volatility are expected to have larger forecast errors and less agreement across analysts. Similar to Alford and Berger (1999) and Thomas (2002), we define ***Excess volatility*** as the standard deviation of the excess return (equity's return minus value-weighted return) over the period from 210 to 11 days before an announcement day.[13]

Next, even though our sample includes only DJI firms, the size of the firm may have an impact on forecasting ability, and may increase forecast accuracy and reduce forecast dispersion (e.g., Atiase, 1985). Accordingly, we control for ***Size***, the market value of the firm's equity (share price multiplied by the number of shares outstanding) five days before the announcement date.

We also control for the possibility that analysts face more difficulty when forecasting earnings for firms with a high market to book ratio or a high intangible to total assets ratio. The latter of the two variables follows from Barth et al. (2001), who conjecture that the level of analysts' efforts and the quality of analysts' forecasts vary with the degree to which firm value is comprised of intangible assets. ***Market to book*** is the market value of equity divided by the book value of equity, and ***Intangible*** is the ratio of intangible to total assets.

[13] Using the standard deviation of the market model residuals and using total volatility yield similar results.

Another possible determinant of forecasting error is leverage, as it adds to the volatility of earnings. We define ***Leverage*** as the ratio of long-term debt and debt in current liabilities to total assets.

Finally, Thomas (2002) shows that an important determinant of forecasting error is the degree of corporate diversification. He shows that analysts who follow firms that operate in multiple segments have smaller forecasting errors.[14] Similar to his measure, we use the ***Herfindahl Index*** as the measure of corporate diversification. It is defined as the sum of squares of each reported segment's assets as a proportion of the firm's total assets.

Note that all control variables that are based on Compustat (i.e., Advertising, Market to book, Intangible, Herfindahl Index, Leverage*)* are calculated based on the 2004 annual statements for 2005 announcements, and on the 2005 annual statements for 2006 announcements.

5. Empirical analysis

5.1 Univariate

Table I provides details on the sample characteristics. Our sample comprises the 30 DJI firms. Based upon an editing frequency variable that is quantified monthly, we have 540 firm-month edit observations. Based upon quarterly earnings per share announcements and revisions that are issued on ad hoc basis, our sample includes 160 forecast error observations and 289 revisions.[15] We present the distribution of most firm characteristic variables based on the 160 observations that correspond to the forecast error variables. As spreads are used only in the

[14] Thomas (2002) claims that forecast errors are not perfectly positively correlated across different operating segments. Thus, the combined forecast error of conglomerates is smaller than that for focused firms.
[15] Most firms will have five forecasts in the sample, but depending on the annual statements' dates, some firms may have six forecasts in the sample period.

revision analysis, we present the spread distribution based on the 289 observations that correspond to the revision observations.

[Insert Table I about here]

Table I shows that the median editing frequency is roughly one edit per day (27 in a month). However, there is much variation in this measure: The top-5% frequency is approximately 11 edits per day, while the bottom-5% frequency is approximately 1 edit per week. Thus, even though our sample consists of the large DJI firms, the variation of the *Edits* variable is high.

With regard to size, the DJI firms are large with a median market cap of close to $100 billion over the sample period. The DJI firms are highly visible, as can be seen by the relatively high number of Wall Street Journal articles that discuss them. On average, each firm appears 108 (median 79) times in a three month period, with a high variation in appearance between firms. The 90% confidence interval of the *Wall Street Journal* variable is between 17-317 appearances in a given quarter.

Excess volatility has a median value of 0.93%, but for 5% of the observations excess volatility is more than 1.6%. *Herfindahl Index* indicates that more than 50% of the observations correspond to firms that are diversified in a few operating segments, at least to some extent.

Since our sample consists of the large DJI firms, large forecast errors are not expected. Indeed, we find that forecast errors are relatively small, with a mean of 0.23% for the closest forecast error and a mean of 0.24% for the average forecast error. Further, revisions tend to be moderate. Analyst recommendations are in the range of 1-5 (1 - strong buy, 2 - buy, 3 - hold, 4 - sell, 5- strong sell), and revisions (current recommendation minus last recommendation) are predominantly within the range of 2 to -2. Larger differences between consecutive recommendations are rarely observed.

In Table II we present cross-sectional correlation of the major independent variables used in the study. The cross-section correlations are based on the observations that correspond to the analysts' consensus forecast file; however, the cross-sectional correlation based on the observations from the analysts recommendation file are rather similar. Apart from the control variables of the study, we also include correlations with the number of analysts following the company in the three months prior to the earnings announcement date.

In general, the correlations in our sample are not high. Edits has a correlation of 0.46 with the Wall Street Journal, so it does seem that firms that are highly covered by the Wall Street Journal also tend to be highly edited on Wikipedia. However, advertising and the number of analysts, which could be considered also as a proxy for a firm's recognition, have a rather small correlation with edits. Also, there is no strong correlation between the size of the firm and edits. Taken together, we conclude that there is a weak correlation between Internet information processing on Wikipedia and what one may associate with a firm's visibility or recognition.

[Insert Table II about here]

In Table III we present information on how Wikipedia edits vary across our sample of firms. To eliminate time-series variation and the effect of size, we construct a normalized edit variable that equals the mean of the monthly edit variable divided by the mean market value of the company during the sample period. We then sort our normalized edit variable from high to low. This allows us to see cross-sectional differences after size is controlled for.

[Insert Table III about here]

We find that McDonalds' edit frequency is the highest, followed by General Motors. The least edited firms are American International Group and Johnson & Johnson. Certainly both McDonalds and General Motors are followed quite extensively by the media, which may contribute to the high number of edits these firms observe.

5.2 Multivariate

Our *Edits* variable for the multivariate analysis is the average monthly edit frequency in the quarter in which forecasts or recommendations are being made.

Our first hypothesis is that *Edits* is negatively correlated with forecast errors. Table IV presents regression results for different specifications where the dependent variable is either the closest forecast error or the average forecast error. Because forecast errors are skewed, we use ln(1+forecast error) as our dependent variable. The most important result to be taken from the table is that *Edits* is negatively correlated with forecast errors. In seven out of the eight specifications the *Edits'* coefficient is negative and significant at the 1% level.

Another interesting result is that a firm's recognition, as measured by Wall Street Journal citation count and advertising costs, is positively correlated with forecast errors. Table IV also confirms some of the results of previous studies. Specifically, we find that volatility is positively correlated with forecast errors, and that firms with more intangible assets appear to have more accurate analysts' forecasts, as evidenced by the negative and significant coefficient of *Intangible*. This latter result is similar to the findings of Thomas (2002). Similar to intangibles, the market to book ratio is also negatively correlated with forecast errors.[16] The other control variables do not seem to be significantly correlated with forecast errors. This is

[16] The result for Intangible and the Market to book ratio may be rationalized if one considers that analysts tend to exert more effort when following firms with high growth opportunities (see Barth et al., 2001).

not surprising given that our sample comprises the DJI firms. All these firms are large in size and have many analysts covering them. Further, it is clear that analysts understand the importance of leverage and figure out how earnings per share are affected by debt levels.

[Insert Table IV about here]

Table V shows the results of regression specifications in which the dependent variable corresponds to the various definitions of dispersion. Similar to the case of forecast errors, because the standard deviation of forecasts is skewed, we take the log of the standard deviation as our dependent variable.[17] In columns (1)-(3) the dependent variable is the closest dispersion, in columns (4)-(6) the dependent variable is the respective quarter's average dispersion, and in columns (7)-(8) we use all the one-quarter-ahead forecast dispersions. The latter specification can be tested because dispersion tends to change somewhat compared to the stale consensus forecast. However, because most of the dispersion value will depend on the previous dispersion value, we add a lagged dependent variable in this specification. Also, because of the possibility of correlated errors, we test the specifications of column (7)-(8) using Newey-West standard errors (with three lags).

The results show that the coefficient of *Edits* is negative and significant at the 1% level in all regression specifications. We also find that volatility is positively correlated with dispersion, while the Herfindahl Index, market to book, and the ratio of intangible assets to total assets are negatively correlated with dispersion. The firm's recognition variables (Wall Street Journal and advertising), do not seem to have a consistent and robust correlation with dispersion. Finally, columns (7)-(8) show that analyst dispersion is highly persistent.

[Insert Table V about here]

[17] We discard the observations with a zero standard deviation. A zero standard deviation may be a flawed observation as one does not expect all analysts to provide the exact same forecast.

Our next test (Table VI) concerns changes in spreads on analysts' recommendation days. Specifically, we examine percentage change between the average daily spread for the firm's share during the 1.5 years of our sample, and the observed spread on the analyst recommendation day.[18] Recall that on recommendation days spreads would in general increase as informed traders are able to better appropriate and use the new information to their advantage (Kim and Verrechia, 1994). Based on this argument, we can quantify the portion of informed investors pertaining to a company; the higher the change in spread on recommendation days, the larger the informed base and the smaller the portion of liquidity traders. In turn, if information processing costs are indeed lower for firms with high edit frequency (compared to firms with low edit frequency), then their stock spread increase should be higher. On top of the controls used in the previous tests, we also control for turnover levels in some of the specifications, as changes in spread tend to correlate with market activity. To control for the possibility that analyst revisions are due to non-information factors, e.g., institutional biases such as herding (Hong and Kubik, 2000), we cluster standard errors at the firm level.

[Insert Table VI about here]

The results in Table VI confirm that edit frequency is positively correlated with changes in spreads on recommendation days. In all eight specifications edit frequency is positively correlated with changes in spread at the 1% significance level. We also find a weak negative correlation between changes in spreads and the Herfindahl Index. None of the other variables is significantly correlated with changes in spreads, suggesting that edit frequency is an important characteristic of firms. Further, the positive correlation between *Edits* and

[18] We repeat the analysis by examining percentage change between the average spread for the firm's share during the 1.5 years of our sample, and the observed spread in the two days following the recommendation (the recommendation day and the following trading day) – the results are qualitatively similar.

changes in spreads is consistent with the idea that information gathering costs are smaller for firms with a high edit frequency.

5.3 *Robustness*

As we have emphasized throughout the paper, the advantage of Wikipedia edit frequency is that it measures the degree of informed individuals. Other Internet based measures may at least partially be associated with uninformed noise traders, leading to ambiguous results. However, to see how Wikipedia fairs out with other measures, we follow Choi and Varian (April 2009) and extract data from Google Trends on the DJI companies for the same sample period. Google Trends analyzes Google web searches to compute how many search queries have been done for a particular term, relative to a certain base. Thus, it is rather easy to obtain cross-sectional variation concerning queries on DJI firms. However, a major limitation is that it is difficult if not impossible to distinguish between queries on the product and queries on the company, as both typically have the same name.

To conserve space we do not tabulate the results; however, we report our main findings.[19] As one may expect, the most queried companies are Microsoft, Intel, and IBM – all computer related companies whose end users tend to query about troubleshooting with their respective products. While we find that the cross-section correlation between the Google Trends measure and our Wikipedia variable is 0.5-0.6; the multivariate regressions are not robust with the Google Trend variable. Further, when we include both variables in the regression jointly, all the results with regard to Wikipedia edit frequency remain basically the same, while the Google Trend coefficient becomes insignificant, and in some cases flips signs.

[19] Results are available from authors upon request.

5.4 Endogeneity

A concern that remains is that of endogeneity as causality has not been established. We first note that our results do not support a reverse causality story. A low forecast error (and low dispersion) should be associated with less asymmetry of information and more agreement between individuals. This in turn should lead to a smaller edit frequency on Wikipedia, so we would expect to find a positive correlation between forecast errors and Wikipedia edits. The fact that our results are opposite, suggests that reverse causality does not provide a viable alternative to our analysis.

However, the analysis in this paper may support a situation in which informed analysts' recommendations provide information that leads to increased Wikipedia edits. In other words, it is plausible that analyst coverage is positively correlated with Wikipedia editing on the one hand; and is positively correlated with our information asymmetry variable (lower forecast errors, lower forecast dispersion, and changes in spreads on recommendation dates) on the other. Thus, according to this latter view point, it is not Internet information processing that leads to a more informed investor base, but rather the higher level of analysts' coverage. We address this possibility next.

The commonly used measure for analysts' coverage is the number of analysts. However, one cannot simply include it as an extra control because the number of analysts is endogenously related to the degree of information asymmetry. We therefore follow Doukas et al. (2005) and estimate a two-stage least squares (2SLS) regression by employing the following structural model:

$$Dependent = f(Independent, Edit\ frequency, Number\ of\ analysts) \quad (1)$$

$$Number\ of\ analysts = f(Dependent, Size, Excess\ volatility, 1/Price) \quad (2)$$

where *Dependent* represents either forecast error, forecast dispersion, or change in spreads on recommendation day, and *Independent* represents the full set of control variables used in Table IV, V, and VI. The dependent variables enter as independent variables in equation (2) jointly with size, excess volatility, and the reciprocal of the share price. The reciprocal of the share price has been utilized by Doukas et al. (2005) and Brennan and Hughes (1991).

The above specification allows us to model our dependent variables and the number of analysts jointly, and by doing so to mitigate the concern that the correlations we find between the dependents and edit frequency are contaminated by analysts' coverage.

[Insert Table VII about here]

The results of Table VII are consistent with the previous tables. The correlation between edit frequency and forecast error is negative and significant at the 1% level, the correlation between edit frequency and forecast dispersion is negative and significant at the 1% level, and the correlation between edit frequency and change in spreads is positive and significant at the 10% level. Thus, while the latter result is less significant than those reported in Table VI, it is consistent with it. Interestingly, we find that the number of analysts does not strongly correlate with our dependents (except for forecast error). Thus, it seems that this type of endogeneity concern can be ruled out.

5.5 Different sample

Since our analysis is based on data of the large 30 Dow Jones Industrial firms, we now wish to also examine whether there is evidence supporting our findings for other firms. Thus in this section we extend our analysis using a new sample. The new sample is also of largesized firms since the degree of Wikipedia entries editing frequency drops tremendously for smaller firms.[20]

The sample was constructed by including all S&P 500 firms that are followed by at least 20 analysts (empirically, the number of analysts can take on values between 0 and 40). Considerable analysts' coverage assured that the firms are sufficiently followed by analysts (and investors), and thus their Wikipedia measure would not suffer from a low edit frequency. Second, we reduce the endogeneity concern that both Wikipedia edit frequency and the information asymmetry proxies are spuriously associated with the degree of analyst coverage (as all these firms are widely followed by analysts).

[Insert Table VIII about here]

Table VIII shows the regression results associated with this new sample. The sample includes 132 firms and 511 observations. For each announcement we tabulated regression results associated with the closest forecast error and closest dispersion (results for the mean forecast error and mean dispersion are similar). We used the same control variables and same specifications used in Table IV and V. The results are similar in nature to those reported for the DJI firms. In all regression specifications the coefficient of edit frequency is negative, and it is significant in 7 out of the 8 specifications.

6. Conclusion

Many characteristics of the Internet distinguish it from traditional information sources. Traditional information sources present information that has passed through the filter of the

[20] Note that even for smaller S&P 500 firms edit frequency is less than once a day. With such infrequent editing, the proxy becomes incapable of differentiating between the degrees of Internet information processing amongst the smaller firms. We wish however to note that because all our firms are large, are followed by many analysts, and are well know; it is less probable that our results are driven by some hidden size effect.

reporter or the anchor and fit within a limited reporting time and space. In contrast, the Internet provides a theoretically limitless news hole of up-to-date, mostly raw information that is available whenever the user wants it. Furthermore, while traditional media decide on what constitutes the news of the day, the Internet allows people to search for the information that interests them (Western, 2000). Indeed, the Internet has proven to have a significant impact in many different domains, including politics, health, and retail consumption. In this paper we analyze the effect of the Internet on information processes taken by analysts and investors.

Previous work in the field of Information Systems provides compelling evidence that the extent to which individuals use computers for a specific task is dependent on the extent to which they feel they can gain desirable attainments. The less confident one is in his or her ability to obtain company related information from the Internet, the less likely one is to use the Internet to learn about the company and vise-versa. In this study we capture this aspect by analyzing the extent to which individuals engage with company related Internet information processing. The unique attribute of Wikipedia that allows individuals to actively participate in the information gathering process allows us to quantify cross-sectional variation in Internet information processing.

We hypothesize that more Internet information processing pertaining to a company would result in analysts and investors that are more informed about that company. To test our hypothesis we conduct empirical experiments that relate the frequency of edits of Dow Jones Industrial firms' entries on Wikipedia to analysts' forecasts and recommendations.

We find that with higher Wikipedia edit frequencies, analysts' forecast errors and forecast dispersions are reduced. Further, we find that higher Wikipedia edit frequencies are correlated with increases in bid-ask spread following analysts' recommendation. These results are consistent with the idea that Internet information processing is correlated with the degree to which investors and analysts are informed about companies.

We believe that this study raises some interesting questions and directions for future research. For example, an interesting question to ask is what drives Internet information processing of company related information. Intuitively, variations in this aspect originate from the interaction between firm-specific attributes and individual attributes. The former being such issues as the way a company disseminates information about its operation; the latter being such issues as personal preferences in analyzing information pertaining to specific companies. Understanding what drives Internet processing can potentially help reduce information asymmetry between investors, which in turn should reduce the firm's cost of capital.

Wikipedia offers a wealth of accurate data that can be analyzed in future studies. For example, detailed data on edits of entries is accurately logged. This type of data includes the exact date, time, location, and revisions' content. The global nature of the Internet together with such data may allow researchers to better capture the degree in which company related information is processed across the world. Intuitively, the degree in which foreign investors own US companies' shares may be associated with the degree in which these investors feel informed about the company. Thus, with Wikipedia data on editors' location it should be possible to capture information asymmetry between local and foreign investors, and perhaps better understand such phenomena as the "home bias".

Another possible venue for future research would be to exploit Wikipedia to establish a better understanding of how informational events (i.e., conference calls, earnings announcements) affect financial markets. An alternative possibility would be to identify firm informational events by measuring the time-series variation of Internet information processing measures. This venue may provide evidence that links financial market volatility to information processing over the Internet.

To conclude, many directions for future research can build on the work presented in this paper, and explore ways by which the Internet affects financial markets. We believe that with the advent nature of Internet usage it is important to understand its effects on investors' behavior. This paper advances our understanding of such relations by providing evidence for a correlation between Internet information processing and reduced information asymmetry.

References

Anderson, J.G., 2004, Consumers of e-health, Social Science Computer Review 22, 242-248.

Alford, A., and P. Berger, 1999, A simultaneous equations analysis of forecast accuracy, analysts following, and trading volume, Journal of Accounting, Auditing, and Finance 14, 219-240.

Atiase, R., 1985, Predisclosure information, firm capitalization, and security price behavior around earnings announcements, Journal of Accounting Research 23, 21-36.

Baker, K., Nofsinger, J., and D. Weaver, 2002, International cross-listing and visibility, Journal of Financial and Quantitative Analysis 37, 495-521.

Bakos, Y., 2001, The emerging landscape of retail e-commerce, Journal of Economic Perspectives 15, 69-80.

Barth, M., Kasznik, R., and M. McNichols, 2001, Analyst coverage and intangible assets, Journal of Accounting Research 39, 1-34.

Bogan, V., 2008, Stock market participation and the Internet, Journal of Financial and Quantitative Analysis 43, 191–212.

Brennan, M., Hughes, P., 1991, Stock prices and the supply of information, Journal of Finance 46, 1665-1691.

Brown, W., Pfeiffer, J., 2008, Do investors under-react to information in analysts' forecasts? Journal of Business Finance and Accounting 35, 889-911.

Choi. H., Varian, H., 2009, Predicting the present with Google Trends, Google Inc.

Compeau, D.R., and C.A. Higgins, 1995, Computer self-efficacy: development of a measure and initial test, MIS Quarterly 19, 189–211.

Cornett, M., Tehranian, H., and A. Yalcin, 2007, Regulation Fair Disclosure and the market's reaction to analyst investment recommendation changes, Journal of Banking and Finance 31, 567-588.

DiMaggio, P., Hargittai, E., Neuman W.R, and J.P. Robinson, 2001, Social Implications of the Internet, Annual Review of Sociology 27, 307-336.

Doukas, J., Kim, C., and C. Pantzalis, 2005, The two faces of analyst coverage, Financial Management 34, 99-125.

Easterwood, J.C., and S. Nutt, 1999, Inefficiency in analysts' earnings forecasts: systematic misreaction or systematic optimism? Journal of Finance 54, 1777-1779.

Eastin, M.S., and R. LaRose, 2000, Internet self-efficacy and the psychology of the digital divide, Journal of Computer-Mediated Communication 6.

Elton, E., Gruber, M., and M. Gultekin, 1984, Professional expectations: accuracy and diagnosis of errors, Journal of Financial and Quantitative Analysis 19, 351-363.

Glosten, L.R., and P.R. Milgrom, 1985, Bid, ask and transaction prices in a specialist market with heterogeneously informed traders, Journal of Financial Economics 14, 71-100.

Grullon, G., Kanatas, G., and J. Weston, 2004, Advertising, breadth of ownership, and liquidity, Review of Financial Studies 17, 439-461.

Hansen, S., Berente, N., and K. Lyytinen., 2007, Wikipedia as rational discourse: An illustration of the emancipatory potential of information systems, 40th Annual Hawaii International Conference on System Sciences 2007.

Hsu, M., and C. Chiu, 2004, Internet self-efficacy and electronic service acceptance, Decision Support Systems 38, 369-381.

Hong, H., and J. Kubik, 2000, Security analysts' career concerns and herding of earnings forecasts, RAND Journal of Economics 31, 121-44.

Kim, O., and R. Verrecchia, 1994, Market liquidity and volume around earnings announcements, Journal of Accounting and Economics 17, 41-67.

Kuznetsov, S., 2006, Motivations of contributors to Wikipedia, SIGCAS Computers and Society, Vol. 36, No. 2.

Kyle, A.S., 1985, Continuous auctions and insider trading, Econometrica 53, 1315-1335.

Lih, A., 2004, Wikipedia as participatory journalism: Reliable sources? Metrics for evaluating collaborative media as a new resource, 5th International Symposium on Online Journalism, University of Texas at Austin.

O'Brien, P., 1988, Analysts' forecasts and earnings expectations, Journal of Accounting and Economics 10, 53-83.

Thomas, S., 2002, Firm diversification and asymmetric information: evidence from analysts' forecasts and earnings announcements, Journal of Financial Economics 64, 373-396.

Ulrik, B., and L. Jürgen, 2007, Revision and co-revision in Wikipedia, Pro Learn International Workshop Bridging the Gap Between Semantic Web and Web 2.0, 4th European Semantic Web Conference 2007.

Voss, J., 2005, Measuring Wikipedia, In Proceedings of the 10th International Conference of the International Society for Scientometrics and Informetrics.

Westen, T., 2000, E-Democracy: Ready or Not, Here It Comes, National Civic Review 89, 217–28.

Wilkinson, D.M. and B.A, Huberman, 2007, Assessing the value of cooperation in Wikipedia, First Monday 12 (4).

Table I
Distribution of Variables

The sample comprises the 30 Dow Jones Industrial firms during the period July 2005 - December 2006. *Edits* is the number of Wikipedia updates during the month (trend-adjusted to October 2006 levels). *Size* is the market value of the firm's equity (in $billion) five days before the announcement date. *Excess volatility* is the standard deviation of the excess return (equity's return minus value-weighted return) during the 210 calendar days prior to the announcement date. *Wall Street Journal* is the number of Wall Street Journal articles that quote the company name during the previous three months. *Advertising* is the percentage of advertising expenses out of total sales. *Market to book* ratio is market value of equity divided by book value of equity. *Intangible* is the ratio of intangible to total assets. *Herfindahl Index* is the sum of squares of each reported segment's assets as a proportion of the firm's total assets. *Leverage* is the ratio of long-term debt and debt in current liabilities to total assets. *Market to book*, *Intangible*, *Herfindahl Index*, and *Leverage* are calculated based on the 2004 annual statements for 2005 announcements, and based on the 2005 annual statements for 2006 announcements. *Spread* is the (closing ask – closing bid)/closing price. *Closest forecast error* is calculated based on the median consensus forecast for earnings per share (EPS) reported closest to, but at least 20 days preceding, the date on which actual earnings is released. *Average forecast error* is calculated based on the average of the median consensus forecasts for EPS reported during the period since the last EPS announcement, but at least 20 days preceding the date on which actual earnings is released. *Closest dispersion* is calculated based on the standard deviation of analysts' EPS forecasts as reported closest to, but at least 20 days preceding, the date on which actual earnings is released. *Average dispersion* is calculated based on the average standard deviation of analysts' forecasts during the period since the last EPS announcement, but at least 20 days preceding the date on which actual earnings is released. All of the analysts' forecasts variables are deflated by the stock price five days before the announcement date. *Revision* equals the difference between an analyst's recommendations in two consecutives reports.

				Percentile				
	No.	Mean	Standard deviation	5th	25th	50th	75th	95th
A1. Firm Characteristics								
Edits	540	64.71	108.07	4.28	13.03	26.97	59.06	323.14
Size	160	123.98	89.43	27.61	56.79	98.26	173.65	349.12
Excess volatility (%)	160	1.03	0.37	0.67	0.78	0.93	1.15	1.60
Wall Street Journal	160	108.37	88.40	17	46	79	146	317
Advertising (%)	160	2.25	3.12	0.00	0.00	1.27	2.59	10.43
Market to book	160	3.53	1.58	1.29	2.17	3.35	4.51	6.05
Intangible (%)	160	15.11	13.49	1.01	4.04	9.46	24.63	41.06
Herfindahl Index (%)	160	17.90	24.50	0.00	0.00	10.56	22.55	75.00
Leverage (%)	160	35.46	15.66	9.57	24.90	36.30	43.05	62.23
Spread (%)	289	0.048	0.020	0.008	0.043	0.052	0.065	0.071
A2: Analysts' Forecasts								
Closest forecast error (%)	160	0.23	0.63	0.00	0.03	0.09	0.20	0.76
Average forecast error (%)	160	0.24	0.68	0.00	0.03	0.08	0.20	0.80
Closest dispersion (%)	155	4.03	8.33	1.00	1.00	2.00	3.50	12.00
Average dispersion (%)	155	4.10	8.53	1.00	1.00	2.00	3.33	11.67
Revision	289	0.12	1.34	-2	-1	0	1	2

Table II
Cross-Section Correlation

Cross-section correlation of edits and control variables used in the study. The correlations are calculated based on the 160 observations that correspond to the analysts forecast file. The sample comprises the 30 Dow Jones Industrial firms during the period July 2005 - December 2006. *Edits* is the number of Wikipedia updates during the month (trend-adjusted to October 2006 levels). *Size* is the market value of the firm's equity (in $billion) five days before the announcement date. *Excess volatility* is the standard deviation of the excess return (equity's return minus value-weighted return) during the 210 calendar days prior to the announcement date. *Wall Street Journal* is the number of Wall Street Journal articles that quote the company name during the previous three months. *Advertising* is the percentage of advertising expenses out of total sales. *Market to book* ratio is market value of equity divided by book value of equity. *Intangible* is the ratio of intangible to total assets. *Herfindahl Index* is the sum of squares of each reported segment's assets as a proportion of the firm's total assets. *Leverage* is the ratio of long-term debt and debt in current liabilities to total assets. *Market to book*, *Intangible*, *Herfindahl Index,* and *Leverage* are calculated based on the 2004 annual statements for 2005 announcements, and based on the 2005 annual statements for 2006 announcements. *Analysts* is the number of analysts following the firm in the three months prior to the EPS announcement date.

	Edits	Size	Excess Volatility	Wall Street Journal	Advertise	Market to book	Intangible	Herfindahl Index	Leverage
Size	0.2008								
Excess volatility	0.0636	-0.3000							
Wall Street journal	0.4637	0.4260	0.2011						
Advertising	0.0794	0.0017	0.0029	0.0334					
Market to book	0.0823	0.0699	-0.2690	-0.1901	0.3068				
Intangible	-0.2056	-0.1185	-0.0591	-0.2424	0.4744	0.0654			
Herfindahl Index	0.0521	-0.1685	-0.1149	-0.2403	0.4016	0.1052	0.0773		
Leverage	0.118	-0.3785	0.2530	-0.0937	0.0759	0.2816	0.1247	0.0800	
Analysts	0.1457	-0.1241	0.3420	-0.1901	0.2611	0.1878	-0.0667	0.0192	0.0696

Table III
Edits Normalized by Size

The table provides a list (in descending order) of the relative updating frequency for the different Dow Jones Industrial firms' websites on Wikipedia. *Edits normalized* is the number of edits (adjusted to October 2006 levels) divided by the market value of the firm's equity ($ billions). *Size* is the market value of the firm's equity ($ billions). The values are calculated at the end of each month (July 2005 - December 2006), and then averaged across the sample period.

Company name	Ticker	Edits normalized	Size
McDonalds	MCD	6.9442	44.342
General Motors	GM	4.8058	15.870
Wal Mart Stores	WMT	1.6961	194.750
Walt Disney	DIS	1.5900	57.575
Microsoft	MSFT	1.5212	270.546
Boeing	BA	0.9420	60.586
International Business Machines (IBM)	IBM	0.6132	129.587
Du Pont	DD	0.5831	39.654
Caterpillar	CAT	0.5034	43.016
Hewlett-Packard	HPQ	0.4879	91.317
Intel	INTC	0.4735	128.209
AT&T	T	0.4642	105.641
Alcoa	AA	0.4480	25.434
Coca Cola Company	KO	0.4250	103.323
Honeywell International	HON	0.3931	33.207
Verizon Communications	VZ	0.3928	96.554
American Express	AXP	0.3359	66.797
Home Depot	HD	0.3170	82.457
Procter Gamble	PG	0.2107	184.573
3M Company	MMM	0.2002	57.704
Merck	MRK	0.1397	77.753
JP Morgan Chase	JPM	0.1357	145.234
Exxon Mobil	XOM	0.1156	387.511
United Technologies	UTX	0.1072	59.596
Altria Group	MO	0.1043	157.529
Citigroup	C	0.1031	240.934
Pfizer	PFE	0.0942	185.101
General Electric	GE	0.0839	359.003
American International Group	AIG	0.0730	168.003
Johnson & Johnson	JNJ	0.0727	185.731

Table IV
Forecast Error and Wikipedia Edits

The sample includes consensus forecasts for the 30 Dow Jones Industrial firms during the period July 2005 - December 2006. *Edits* is the monthly average number of Wikipedia updates (trend-adjusted to October 2006 levels) in the three months prior to the announcement date. The dependent variables are transformed to log(1+forecast error). Definitions of the other variables are provided in Table I. *N* is the number of forecasts; *t*-statistics are provided in parentheses and calculated using White-corrected standard errors. Significance at the 1% (***), 5% (**), and 10% (*) is tabulated.

Dependent variable	(1) Closest forecast error	(2) Closest forecast error	(3) Closest forecast error	(4) Closest forecast error	(5) Average forecast error	(6) Average forecast error	(7) Average forecast error	(8) Average forecast error
Edits	-2.79E-4	-4.35E-4	-4.85E-4	5.16E-4	-3.10E-4	-4.31E-4	5.00E-4	5.15E-4
	(-2.53**)	(-2.84***)	(-3.13***)	(-3.30***)	(-2.66***)	(-2.74***)	(-3.10***)	(-3.17***)
Size	-1.69E-5	-3.06E-4	-3.31E-4	-2.42E-4	-6.68E-6	-2.61E-4	-2.88E-4	-2.45E-4
	(-0.16)	(-1.76*)	(-2.22**)	(-1.70*)	(-0.06)	(-1.44)	(-1.72*)	(-1.54)
Excess volatility	48.13	40.44	39.05	37.61	47.28	39.64	38.22	37.52
	(4.81***)	(5.02***)	(5.19***)	(4.80***)	(4.51***)	(4.61***)	(4.78***)	(4.54***)
Wall Street Journal		6.26E-4	4.25E-4	3.98E-4		5.49E-4	3.34E-4	3.21E-4
		(2.28**)	(1.72*)	(1.64)		(1.91*)	(1.28)	(1.24)
Advertising		0.26127	1.5347	1.67893		0.10905	1.51490	1.58535
		(0.73)	(2.24**)	(2.57***)		(0.28)	(2.23**)	(2.38**)
Market to book		-0.02527	-0.03183	-0.03752		-0.02931	-0.03652	-0.03929
		(-2.91***)	(-3.45***)	(-3.94***)		(-2.96***)	(-3.49***)	(-3.84***)
Intangible			-0.47108	-0.50721			-0.54809	-0.56574
			(-3.48***)	(-3.76***)			(-3.81***)	(-4.00***)
Herfindahl Index			-0.05862	-0.06563			-0.04824	-0.05166
			(-1.02)	(-1.17)			(-0.81)	(-0.87)
Leverage				0.14641				0.07152
				(1.53)				(0.69)
N	160	160	160	160	160	160	160	160
R-squared	0.494	0.552	0.587	0.592	0.447	0.504	0.548	0.549

Table V
Forecast Dispersion and Wikipedia Edits

The sample includes consensus forecasts dispersion for the 30 Dow Jones Industrial firms during the period July 2005 - December 2006. The dependent variables are transformed to log(dispersion). *Edits* is the monthly average number of Wikipedia updates (trend-adjusted to October 2006 levels) in the three months prior to the announcement date. *Dispersion* is the standard deviation of analysts' EPS forecasts. Definitions of other variables are provided in Table I. Specifications (7)-(8) include all dispersion forecasts in the period (one-, two-, three-, and four-quarter forward-looking forecasts) and regressions include the lagged dispersion forecast. *N* is the number of observations; *t*-statistics are provided in parentheses and calculated using White-corrected standard errors for specifications (1)-(6) and the Newey-West procedure with three lags for specifications (7)-(8). Significance at the 1% (***), 5% (**), and 10% (*) is tabulated. Significance at the 1% (***), 5% (**), and 10% (*) is tabulated.

Dependent variable	(1) Closest dispersion	(2) Closest dispersion	(3) Closest dispersion	(4) Average dispersion	(5) Average dispersion	(6) Average dispersion	(7) Dispersion	(8) Dispersion
Lag (dep. var.)							0.86800	0.81096
							(56.60***)	(42.47***)
Edits	-0.00132	-0.00148	-0.00153	-0.00223	-0.00212	-0.00239	-3.40E-4	-3.09E-4
	(-2.91***)	(-2.97***)	(-2.89***)	(-5.33***)	(-4.94***)	(-4.94***)	(-3.81***)	(-3.09***)
Size	9.80E-4	1.58E-4	4.87E-4	0.00149	0.00121	0.00148	-8.22E-5	-8.14E-5
	(1.18)	(0.15)	(0.16)	(1.68*)	(1.20)	(1.99**)	(-0.66)	(-0.64)
Excess volatility	144.87	98.81	75.87	121.77	99.04	76.82	19.51	21.13
	(6.13***)	(4.84***)	(4.79***)	(5.40***)	(4.82***)	(4.62***)	(4.25***)	(4.48***)
Wall Street Journal		0.00176	-3.18E-4		0.00063	-0.00109	4.70E-4	2.34E-4
		(1.94*)	(-0.42)		(0.72)	(-1.49)	(3.68***)	(1.82*)
Advertising		1.33354	12.72		-4.05298	5.71406		
		(0.91)	(6.85***)		(-2.53**)	(2.64***)		
Market to book		-0.25675	-0.344407		-0.13516	-0.22377		-0.04083
		(-7.90***)	(-10.5***)		(-3.99***)	(-7.14***)		(-5.91***)
Intangible			-3.19639			-2.97473		-0.32440
			(-7.86***)			(-6.51***)		(-4.47***)
Herfindahl Index			-1.02758			-0.72782		-0.07881
			(-6.02***)			(-3.87***)		(-2.57***)
Leverage			0.72508			1.04032		-0.00483
			(2.31**)			(2.87***)		(-0.07)
N	155	155	155	155	155	155	1605	1605
R-squared	0.310	0.515	0.652	0.293	0.402	0.537	0.815	0.821

Table VI
Percentage Change in Spread on Analysts' Recommendation Days and Wikipedia Edits

The sample includes analysts' recommendations revisions for the 30 Dow Jones Industrial firms during the period July 2005-December 2006. The dependent variable is *Percentage change in spread*, which equals the percentage difference between the average spread of the company's stock during the period and the spread on the analyst recommendation date. *Absolute revision* is the difference in analysts recommendation between two consecutives recommendations of a particular analyst. *Turnover* is the number of shares traded during the month of the revision divided by the number of shares outstanding. Definitions of other variables are provided in Table I. *N* is the number of observations; *t*-statistics of coefficient are provided in parenthesis and calculated using clustered standard errors at the firm level.

Dependent variable	(1) Percentage change in spread	(2) Percentage change in spread	(3) Percentage change in spread	(4) Percentage change in spread	(5) Percentage change in spread	(6) Percentage change in spread	(7) Percentage change in spread	(8) Percentage change in spread
Turnover					-26.17	-35.80	-36.26	-43.60
					(-1.05)	(-1.58)	(-1.62)	(-1.39)
Absolute revision	0.03970	0.03816	0.04012	0.02925	0.03825	0.03928	0.04223	0.02605
	(0.61)	(0.62)	(0.67)	(0.49)	(0.57)	(0.63)	(0.69)	(0.44)
Edits	0.00259	0.00246	0.00250	0.00271	0.00285	0.00243	0.00249	0.00272
	(2.88***)	(2.66***)	(2.56***)	(3.48***)	(3.04***)	(2.74***)	(2.66***)	(3.51***)
Size	2.93E-4	2.40E-4	2.69E-4	-4.24E-5	-2.75E-5	-3.09E-4	-2.73E-4	-7.74E-4
	(0.53)	(0.48)	(0.55)	(-0.08)	(-0.04)	(-0.54)	(-0.50)	(-1.22)
Excess volatility	-6.45	-7.81	-8.32	-7.71	10.34	9.84	9.31	13.04
	(-0.86)	(-0.95)	(-1.00)	(-0.92)	(1.01)	(1.07)	(1.04)	(1.15)
Wall Street Journal		5.47E-4	4.50E-4	7.04E-4		0.00114	0.00128	0.00115
		(0.63)	(0.43)	(0.65)		(1.24)	(1.04)	(0.91)
Advertising		-2.69079	-2.53142	-3.89652		-1.98537	-1.73789	-1.27241
		(-1.46)	(-1.25)	(-1.65*)		(-1.49)	(-1.09)	(-0.62)
Market to book			-0.01100	0.00648			-0.01645	-0.00972
			(-0.24)	(0.14)			(-0.37)	(-0.19)
Intangible				0.92432				0.21793
				(1.70*)				(0.37)
Herfindahl Index				-0.33397				-0.45066
				(-1.61)				(-1.99**)
Leverage				-0.31564				-0.41641
				(-0.58)				(-0.75)
N	289	289	289	289	289	289	289	289
R-squared	0.0186	0.0207	0.0207	0.0248	0.0212	0.0244	0.0245	0.0283

Table VII
Two-Stage Least Squares Model for Analyst Coverage and
(1) Forecast Error (2) Forecast Dispersion and (3) Changes in Spread

The sample includes analysts' recommendations revisions for the 30 Dow Jones Industrial firms during the period July 2005-December 2006. *Number of analysts* is the number of analysts who provided forecasts and/or recommendations in the 3 months prior. *1/Price* is one divided by the stock price five days before the announcement date. Definitions of other variables are provided in Table I, IV, V, and VI. *N* is the number of observations; *t*-statistics of coefficient are provided in parenthesis and calculated using White-corrected standard errors. Significance at the 1% (***), 5% (**), and 10% (*) is tabulated.

	(1)		(2)		(3)	
Endogenous variables	Closest forecast error	Number of Analysts	Closest dispersion	Number of Analysts	Percentage change in spread	Number of Analysts
Closest forecast error		-6.39769 (-2.44**)				
Closest dispersion				-1.4008 (-3.12***)		
Percentage change in spread						-1.14416 (-1.36)
Number of Analysts	-0.02672 (-2.19**)		0.03790 (0.85)		0.08602 (1.06)	
Other variables						
Edits	-0.00037 (-2.63***)		-0.00159 (-2.83***)		0.00264 (1.91*)	
1/Price		101.08 (4.03***)		135.86 (4.83***)		244.39 (11.50***)
Size	-3.04E-4 (-1.56)	-0.00148 (-0.55)	3.90E-4 (0.51)	7.53E-5 (0.03)	0.00220 (1.18)	-0.00231 (-0.84)
Excess volatility	45.21 (7.85***)	506.04 (3.64***)	63.05 (3.07***)	363.36 (4.17***)	-11.60 (-0.54)	-26.63876 (-0.64)
Wall Street Journal	0.00038 (2.02**)		-3.12E-4 (-0.41)		-5.37E-4 (-0.29)	
Advertising	1.72564 (2.67***)		10.77 (4.19***)		-16.25 (-1.98**)	
Market to book	-0.03963 (-3.35***)		-0.37508 (-8.68***)		-0.19733 (-2.20**)	
Intangible	-0.40383 (-3.07***)		-2.809 (-5.73***)		3.76478 (2.49**)	
Herfindahl Index	-0.070116 (-1.34)		-0.94344 (-4.42***)		0.27301 (0.49)	
Leverage	0.09580 (0.94)		0.89979 (2.17**)		0.27711 (0.26)	
Absolute revision					0.05417 (0.40)	0.50022 (1.78*)
N	160	160	155	155	289	289
R-squared - within	0.575	0.219	0.653	0.218	0.0246	0.405

Table VIII
Robustness: Forecast Error, Dispersion and Wikipedia Edits

The sample includes all S&P 500 firms who are followed by at least 20 analysts during the years 2005-2006 (132 firms and 511 observations). *Edits* is the monthly average number of Wikipedia updates in the three months prior to the announcement date. Definitions of all variables are provided in Table I, III and V. *t*-statistics are provided in parentheses and calculated using White-corrected standard errors. Significance at the 1% (***), 5% (**), and 10% (*) is tabulated.

Dependent variable	(1) Closest forecast error	(2) Closest dispersion	(3) Closest forecast error	(4) Closest dispersion	(5) Closest forecast error	(6) Closest dispersion	(7) Closest forecast error	(8) Closest dispersion
Edits	-2.1E-6	-0.0013	-2.94E-6	-0.0016	-3.17E-6	-0.0026	-1.50E-6	-0.0026
	(-1.88*)	(-1.71*)	(-1.98**)	(-3.62***)	(-2.11**)	(-3.60***)	(-1.06)	(-3.51***)
Size	-5.18E-4	-0.1063	-7.05E-4	-0.3248	-7.23E-4	-0.3129	-8.99E-4	-0.3191
	(-3.00***)	(-3.49***)	(-3.30***)	(-7.02***)	(-3.30***)	(-6.75***)	(-3.33***)	(-5.70***)
Excess volatility	-1.58E-4	-5.4967	-0.0015	2.7456	6.74E-4	2.9797	-7.59E-4	3.1843
	(-0.02)	(-2.44**)	(-0.20)	(1.13)	(0.08)	(1.20)	(-0.09)	(1.22)
Wall Street Journal			9.15E-6	0.0047	9.3E-6	0.0044	9.81E-6	0.0039
			(3.54***)	(5.25***)	(3.53***)	(5.09***)	(3.39***)	(4.19**)
Advertising			4.80E-4	0.5356	-1.56E-4	0.7194	-0.0023	0.4726
			(0.22)	(0.65)	(-0.07)	(0.87)	(-0.85)	(0.52)
Market to book			-2.56E-5	-0.0191	-2.44E-5	-0.0200	-2.03E-5	-0.0187
			(-1.08)	(-1.75*)	(-1.04)	(-1.87*)	(-0.94)	(-1.75*)
Intangible					-9.33E-4	-0.4163	-5.01E-4	0.4216
					(-1.97**)	(-2.22**)	(-0.89)	(-1.95*)
Herfindahl Index							-5.87E-4	-0.0027
							(-1.49)	(-0.02)
Leverage							0.0023	0.1011
							(2.48**)	(0.39)
R-squared	0.0486	0.0339	0.0732	0.1564	0.0820	0.1613	0.1171	0.1630