Impact of Wikipedia on Market Information Environment: Evidence on Management Disclosure and Investment Reaction

Sean Xin Xu
School of Economics and Management, Tsinghua University
Beijing, CHINA  {xuxin@sem.tsinghua.edu.cn}

Xiaoquan (Michael) Zhang
School of Business and Management, Hong Kong University of Science and Technology,
Clear Water Bay, HONG KONG  {zhang@ust.hk}

In this paper, we seek to determine whether a typical social edia platform, Wikipedia, improves the information environment for investors in the financial market. Our theoretical lens leads us to expect that information aggregation about public companies on Wikipedia may influence how management’s voluntary information disclosure reacts to market uncertainty with respect to investors’ information about these companies. Our empirical analysis is based on a unique data set collected from financial records, management disclosure records, news article coverage, and a Wikipedia modification history of public companies. On the supply side of information, we find that information aggregation on Wikipedia can moderate the timing of managers’ voluntary disclosure of companies’ earnings disappointments, or bad news. On the demand side of information, we find that Wikipedia’s information aggregation moderates investors’ negative reaction to bad news. Taken together, these findings support the view that Wikipedia improves the information environment in the financial market and underscore the value of information aggregation through the use of information technology.

Keywords: Social media, Wikipedia, information environment, financial market, management disclosure, information aggregation

Introduction

One primary function of markets is to aggregate information. When market participants cannot communicate with each other freely, they collect their information piecemeal and it “never exists in concentrated or integrated form” (Hayek 1945, p. 519). In this context, market price makes important contributions to information transmission, but such markets may still be inefficient, with price dispersion often observed in practice (Ba et al. 2012; Brynjolfsson and Smith 2000). Researchers thus ask whether market participants can use any explicit mechanism to directly exchange and aggregate information (e.g., Chen et al. 2011; Dellarocas and Wood 2008; Fan et al. 2000; Granados et al. 2010; Sun 2011; Sun and Tyagi 2012; Zhu and Zhang 2010).

A burgeoning literature examines whether social media can serve as such a mechanism (e.g., Gu et al. 2007; Tetlock et al. 2006).
Social media are Internet-based applications “that allow the creation and exchange of user-generated content” (Kaplan and Haenlein 2010, p. X). On social media, users are not only consumers, but also active contributors of content. Following this line of research, we focus on Wikipedia, a typical social media platform (Greenstein and Zhu 2012a, 2012b; Kane and Fichman 2009; Ransbotham and Kane 2011). Wikipedia is an online encyclopedia based on the “wiki” technology that allows individuals to make collaborative contributions to web pages through a web browser (Kane and Fichman 2009). This technology allows people who have access to the Internet to freely contribute and modify the content of its web pages. Volunteers contributed all of the entries in the encyclopedia. After its creation in January 2001, Wikipedia quickly developed to become one of the largest reference websites, with currently 400 million unique visitors monthly and more than 90,000 active contributors. The English version of Wikipedia, the most popular of the 270 different language versions, attracts hundreds of thousands of visitors who make tens of thousands of modifications every day. As of September 2011, Wikipedia is the seventh-most-popular website worldwide, receiving 2.7 billion page views every month.

On Wikipedia, several thousand articles are about traded companies. For these companies, Wikipedia keeps a good record of their significant events. Appendix A depicts a section of the Wikipedia entry for Dell. All sentences in this section follow the format “On [Month Day, Year], [Dell [did something]].” As such, Wikipedia effectively maintains a virtual memory where qualitative information exists in a concentrated and integrated form (Hayek 1945, Kankanhalli et al. 2005; Stein and Zwass 1995). Visitors thereby can obtain aggregated, timely information about the firm’s products, management structure, operations, marketing, sales, and other important firm events. With this rich information, Wikipedia has become an important source of knowledge about companies. In September 2011, for example, Wikipedia entries for Microsoft and Google received 400,000 and 2,000,000 visits, respectively. In particular, there is some evidence that Wikipedia has substantial impact on active seekers of information. According to HitWise, an Internet traffic-monitoring service, more than 70 percent of Wikipedia’s incoming traffic comes from search engines. Because of Wikipedia’s importance in search engines, Wikipedia pages for listed companies are often ranked among the highest in search results. As such, when investors conduct research online, Wikipedia entries of listed companies are highly visible and accessible to them. As noted by Financial Times, “the capital market is looking for information from companies in various social media channels and at a greater extent than we anticipated….Wikipedia is the most popular social media site for individuals looking for such information, used by more than three quarters of respondents.”

Our research examines the value of Wikipedia for aggregating firm information, and we choose the modern financial market as our research context. The market’s information environment consists of quantified information provided by financial analysts (mainly their forecasts of firm earnings) and qualitative descriptions of firms’ business environments, operations, and prospects published by media. These linguistic descriptions are a valuable complement to analyst forecasts, because summative earnings forecasts cannot entirely convey the limitless variety of firms’ financial standings (Tetlock et al. 2008) and because analyst forecasts are often incomplete or even biased measures of firm performance (Jacob et al. 1999; Karamanou and Vafeas 2005). The accounting and finance literatures thus emphasize a mosaic notion of investors’ information environment (Francis et al. 1997; Lang and Lundholm 1996; Loss and Seligman 1995). The central tenet of the mosaic perspective, that information is multifaceted and that both quantified and qualitative types of information are useful, implies that the advent of detailed, linguistic information on Wikipedia would profoundly change investors’ information environment. Data on qualitative information aggregation have been hard to come by, however, and the value of linguistic descriptions of a firm’s environments, operations, and prospects has been severely neglected in prior studies.

Fortunately, Wikipedia offers not only a practical means for information aggregation, but also a way for scholars to investigate the process of information aggregation. On Wikipedia, the history of user contributions is open to the public. This gives us a unique opportunity to develop a proxy for information aggregation on Wikipedia—that is, the number of user modifications of public firms’ Wikipedia entries (called “Wikipedia modifications” hereafter). Through collaborative


With this proxy for information aggregation on Wikipedia, our approach to identify its impact is based on a rich literature on management disclosure (for literature reviews, see Dye 2001; Healy and Palepu 2001). We examine how information aggregation on Wikipedia may influence managers’ and investors’ behaviors. From the disclosure literature, the timing of management disclosure about firms’ unfavorable news is related to investors’ information set about the firms (Healy and Palepu 2001). The intuition is that, driven by self-interest, managers decide whether to admit unfavorable news based on what investors already know (Dye 1985, 2001). So far, the literature widely recognizes the quantified information provided by analysts as the foremost source of information for investors, and, accordingly, relates management disclosure to characteristics of the quantified information. Our overall prediction, motivated by research on media impact, is that “linguistic media content captures otherwise hard-to-quantify aspects of firms’ fundamentals” (Tetlock et al. 2008, p. 1437). If Wikipedia’s qualitative information really informs investors, then it has a potential to alter how management discourse is conditioned on characteristics of the quantified information.

We address the following research questions: (1) Does Wikipedia’s aggregation of qualitative information affect how management disclosure is conditioned on quantified information in the market about firm earnings? (2) If Wikipedia can really inform investors in the market, how would that change investors’ reactions when managers disclose unfavorable news?

The first assumption is about information asymmetries in the financial market. That is, managers have an information advantage relative to investors; managers have more complete information about their firm’s economic reality (Healy and Palepu 2001).

The second assumption is that managers observe (at least partially) what investors know. There is evidence that managers announce news about cash flows “to meet investor demand for cash flow information” (Wasley and Wu 2006, p. 391). What is implied is that managers know investors’ information demand and what investors already know, so that managers can leverage disclosure to meet investors’ information demand. Prior studies also document evidence “consistent with managers being concerned with the risk for litigation and issuing preemptive earnings forecasts to adjust investor expectations” (Wasley and Wu 2006, p. 390). This, again, implies that managers are aware of investor expectations, so that they can use preemptive forecasts to influence investors. The literature on media coverage also assumes that consumers’ prior beliefs are not proprietary information; rather, firms know it (Gentzkow and Shapiro 2006).

In particular, it is arguable that managers are aware that investors use Wikipedia to get information about their firms. A article in Financial Times reported how Klaus Kleinfeld, the former CEO of Siemens, described his reaction to the growing influence of Wikipedia and argued that “Wikipedia, the peer-produced online encyclopedia, is a popular way for people to gain information about companies and business people.” Steve Goodman, CEO of PacketTrap, believed that “if our potential customers want to learn about something, they either go to…Wikipedia, or to Google. And Google search results often lead right back to Wikipedia” (Zetlin 2010). A recent survey, by Webranking, found that 81 percent of surveyed companies cared about their information released to investors via Wikipedia; the survey suggested that

---

6A large body of literature supports this notion (e.g., Abarbanell et al. 1995; Francis et al. 1997; Healy and Palepu 2001; Kasznik and Lev 1995; Lang and Lundholm 1996; Roulstone 2003; Skinner 1994, 1997).

7We thank an anonymous reviewer for pointing out this key assumption.

8http://blogs.ft.com/businessblog/2008/01/wikipedia-is-pohtml/#axzz1mPeEZxlW.
Management Disclosure

Based on the assumption that managers are aware of what investors know, the literature offers two perspectives to explain why managers choose to withhold or release information.

Withholding Perspective: When there is a high level of information asymmetry between managers and investors, managers tend to suppress or withhold unfavorable information (Dye 2001). A seminal model explains why managers’ withholding of bad news hinges on information asymmetry between managers and investors (Dye 1985). The underlying rationale is that withholding bad news may give managers the necessary time to wait for the arrival of more favorable news or to make adjustments to accounting measures. Managers are better placed to do this when firm information is opaque to investors. By contrast, when there is lower information asymmetry between managers and investors, there is less room for managers to manipulate information.  

Preemption Perspective: When the market holds an overly optimistic view about firm performance, managers face several risks; to alleviate these risks, managers may choose to release information promptly to adjust market expectation (Healy and Palepu 2001). First, if the market has an overly optimistic expectation about firm performance, firm disclosure on the official announcement date would trigger a downward movement of stock price (Bartov et al. 2002). Managers may be held accountable for poor stock performance.  

Information Aggregation Mechanisms

From each perspective above, management disclosure is conditioned on investors’ information. As mentioned in the “Introduction,” investors’ information environment consists of two major sources: One is quantified information provided by financial analysts, who collect information from various sources, evaluate firms’ current performance, and make quantitative forecasts about future firm profitability (Roulstone 2003). The other is qualitative information provided by media—that is, linguistic descriptions of firms’ current and future profit-generating activities. Qualitative information sources include traditional media, such as individual journalists and the business press (Frankel and Li 2004; Mitchell and Mulherin 1994), and, more recently, social media, which plays an increasingly important role in disseminating firm information on the Internet (Gu et al. 2007; Tetlock et al. 2008).

Prior studies find poor stock performance to be associated with management changes (Dahya et al. 2002; Warner et al. 1988; Weisbach 1988), as well as hostile takeovers, which in turn result in high CEO turnover (Franks and Mayer 1996; Morck et al. 1990).
With these various information sources, it is important to examine mechanisms that can help aggregate information for investors. Prior studies suggest several mechanisms.

First, given the quantitative nature of analyst forecasts about firm earnings, it is convenient to combine analyst forecasts and create an aggregate measure (e.g., average analyst forecasts; Francis et al. 1997; Karamanou and Vafeas 2005; Lang and Lundholm 1996; Roulstone 2003). Such aggregation, however, may still result in inaccurate and incomplete information for investors. Analysts’ forecasts are often biased in that they may overestimate a firm’s earnings, resulting in an upward analyst bias (Jacob et al. 1999; Karamanou and Vafeas 2005). Analysts have incentives to make optimistic forecasts because they may be rewarded for providing information that generates trading volume and investment banking fees for their brokerage houses (Dechow et al. 2000; Lin and McNichols 1998). Also, analysts’ summative earnings forecasts cannot entirely convey the limitless variety of firms’ financial standings (Tetlock et al. 2008). Investors, however, need detailed, qualitative information about various aspects of firm operations in order to assess firm value, according to the “mosaic” perspective. The disclosure literature documents evidence in support of this mosaic view (e.g., Francis et al. 1997; Lang and Lundholm 1996). We thus need to address mechanisms that can aggregate qualitative information for investors.

An approach to aggregate qualitative information is to store and present news articles and investor discussions in a unified depository. Mitchell and Mulherin (1994) find that the number of news articles reported by Dow Jones & Company is related to trading volume in the market, suggesting that investors become informed and thus buy and sell stocks. Antweiler and Frank (2004) find that stock-market discussions posted on Yahoo! Finance help predict market volatility. The huge number of articles, however, may give investors information overload. Also, a large proportion of online discussions may have repetitive content or even “noise”; as a result, investors bear high costs of information processing on online message boards (Gu et al. 2007).

One way to eliminate noise of online postings is through users’ collective reporting. Gu et al. (2007, p. 74) examined virtual investing-related communities (VICs):

One approach VIC providers use to improve posting quality is to actively monitor and filter low-quality postings… allowing users to report abusive postings, which are then investigated manually by VIC providers.

Another way to eliminate abusive postings is to have companies select and present relevant news articles in the “investor relations” section on corporate websites (Geerings et al. 2003). These approaches are controlled by VICs or companies, however, so information aggregation may be subject to their selection bias. For instance, companies behind some online communities may choose not to engage in noise filtering (Gu et al. 2007).

In Appendix B, we summarize these mechanisms as documented in previous research. In the next section, we explain how information aggregation on Wikipedia, the focus of our research, differs from these previous mechanisms.

Research Framework and Hypotheses

We develop a research framework for Wikipedia’s information aggregation and managers’ voluntary disclosure (Figure 1). The framework closely follows the premise of the disclosure literature that managers’ decisions of voluntary disclosure are based on investors’ information environment. This study moves one step forward by analyzing the additional value of Wikipedia’s information aggregation in investors’ information set.

Timing of Voluntary Disclosure

The dependent variable of our framework (Figure 1) is the timing of voluntary disclosure. Disclosure timing is determined by the tension between withholding information and releasing it in a timely manner. Following the literature, we...
restrict our attention to the disclosure of earnings performance (among others, Aniolowski et al. 2007; Baginski et al. 1994; Cohen et al. 2007; Kasznki and Lev 1995; Miller 2002; Skinner 1994, 1997). We focus on voluntary disclosure and exclude mandatory disclosure, because the latter type is required by the SEC and thus may not be attributable to features of the information environment. Following the literature, we examine disclosure timing of bad news, that is, when forthcoming earnings are below market expectations (e.g., Baginski et al. 1994; Dye 2001; Healy and Palepu 2001; Skinner 1994, 1997). Disclosure timing of bad news can be leveraged because of the following considerations. First, managers generally have an incentive to withhold unfavorable information (Dye 1985, 2001). Second, according to the preemption perspective, it is a firm’s poor rather than good earnings performance that may result in management turnover and damage the firm’s relationship with the professional investment community (Kasznki and Lev 1995; Skinner 1994, 1997). In particular, litigation risks are attributable to bad rather than good news. Based on a review of prior studies on litigation cases, Skinner (1994) summarizes that more than 95 percent of firms that are targets of earnings-related shareholder lawsuits can be classified as having bad news on official earnings-announcement dates. Therefore, managers face a crucial decision of whether to

---

14Skinner (1994, p. 42) explains that “The legal reasons for this asymmetry [i.e., the dominant fraction of lawsuits related to bad news] appear to relate to proof of damages and the need to show a sufficient causal connection between the plaintiff’s injury and the wrongful conduct. If an investor purchases a stock whose price subsequently declines, it is relatively easy to show both an out-of-pocket loss and (if the decline is accompanied by the release of information) causation. Conversely, a plaintiff who sells before good news is revealed suffers an opportunity loss and must show that he or she would not have sold had the information been available.”
release unfavorable news promptly prior to the mandatory earnings announcement.

Following this line of studies, the dependent variable in our framework (Figure 1) is disclosure lag, defined as the number of days between the end of a fiscal quarter and the date when managers voluntarily disclose bad news about earnings performance. The shorter the disclosure lag, the more timely is the voluntary disclosure of adverse information.

**Quantified Information**

Prior studies on investors’ information environment use two variables to characterize analysts’ quantified forecasts for firm earnings: dispersion and bias (Francis et al. 1997; Karamanou and Vafeas 2005; Lang and Lundholm 1996; Roulstone 2003). The literature uses analyst dispersion to proxy for information asymmetry in the information environment (Abarbanell et al. 1995; Roulstone 2003). Analyst dispersion refers to the degree to which analysts’ opinions are in disagreement. The withholding perspective suggests that when analysts cannot agree with each other, investors, who obtain information from analysts, are unsure about firm operations and performance.

This gives managers an information advantage; that is, managers can keep silent (about their firm’s bad earnings performance) for a longer time if the market is more uncertain about the information that they have. Prior studies in the disclosure literature suggest that when there is greater information asymmetry in the market, managers can delay their voluntary disclosure for a longer time, waiting for investors to acquire information and achieve a consensus about firm performance (Francis et al. 2008; Miller 2002; Wasley and Wu 2006). Greater information asymmetry generally gives managers more time to adjust accounting information to suppress bad news (Matsumoto 2002; Richardson et al. 2004).

In sum, the withholding perspective suggests that disclosure lag is positively related to analyst dispersion.

The literature uses analyst bias to measure the difference between true firm performance and the quantified information accessible to investors (Jacob et al. 1999; Karamanou and Vafeas 2005). According to the preemption perspective, the greater the analyst bias, the larger is the negative earnings surprise on the official earnings announcement date (Kasznik and Lev 1995). The negative earnings surprise exposes managers to risks of job changes, reputation damage, and stockholder lawsuits, as reviewed above. Early disclosure is probably the best strategy to reduce these risks for three reasons (Skinner 1994): (1) only investors who bought or sold the firm’s stock before management disclosure can sue, simply because investors who conduct transactions after the disclosure are already aware of the bad news; (2) early disclosure undercuts plaintiffs’ argument that managers failed to disclose promptly, because managers made a relatively timely disclosure; and (3) early disclosure helps spread out (negative) market reactions over a longer time, which helps prevent the stock price from plummeting and thus helps mitigate the risk of management changes because of poor stock performance (Kasznik and Lev 1995). In sum, the pre-emption perspective suggests that disclosure lag is negatively related to analyst bias.

**Qualitative Information**

Prior literature characterizes media information using the number of news articles and the amount of newsworthy content (Antweiler and Frank 2004; Atiase 1985; Foster 1987; Mitchell and Mulherin 1994; Tetlock et al. 2008). We follow the literature and address these factors in our research. Since the focus of this study is the role of Wikipedia’s information aggregation, they are posited as control variables in our model. We discuss their effects in detail later in the subsection “Empirical Model for Management Disclosure.”

As defined earlier, information aggregation on Wikipedia refers to a process by which individuals synthesize their information by contributing to Wikipedia’s company entries.

**How Does Wikipedia Aggregate Information?**

Earlier we summarized the various mechanisms of information aggregation as documented in prior research. Wikipedia differs from previous mechanisms of information aggregation in that it can better generate the “wisdom of crowds” (see Appendix B). Surowiecki (2004) suggests that a social media platform must satisfy four conditions for a crowd to be smart: diversity, independence, decentralization, and aggregation. Wikipedia meets these conditions.

Creating and revising any entry on Wikipedia is completely decentralized. It allows anyone to freely edit any details of all entries (Kane and Fichman 2009; Te’eni 2009). Wikipedia thus supports “democratization of contributions (e.g., leveling the playing field so anyone can contribute an idea)” (Majchrzak 2009, p. 19; Zwass 2010). This guarantees diversified sources of contributors, in contrast to corporate websites or some communities that control information content (Geerings et al. 2003). Wikipedia ensures independent expression of
users’ opinions and supports “deep profiling,” allowing users
to provide detailed information to define their own personal
and social identities (Zhang and Zhu 2011). This encourages
users to identify with Wikipedia and motivates them to con-
tribute (Ma and Agarwal 2007). Contributors use the wiki
technology to edit the same pages, thus effectively aggregating information (Kane and Fichman 2009) and reducing
readers’ information-processing costs (Gu et al. 2007). Wiki-
pedia also provides functions to easily search and display
other users’ inputs. This leads to collaborative efforts to
screen noise and eliminate abusive postings (Moon and
Sproull 2008). The difference between a wiki and a tradi-
tional discussion board is that good content is retained on a
wiki, and openness and transparency make a wiki naturally
resistant to spam (Wagner and Majchrzak 2006).

A relevant literature examines the editing behavior of Wiki-
pedia contributors. There is evidence that individuals’ con-
tributing patterns are often determined by their personalities
and thus do not change much over time (Panciera et al. 2009).

Some prominent Wikipedians argue that a group of prolific
users is the driving force behind the success of Wikipedia
(Wales 2005), because some contributors tend to express
ownership of entries in the collaborative authoring (Thom-
santelli et al. 2009). Recent research, however, suggests a
dramatic shift such that the collective contributions play an
increasingly bigger role (Kittur et al. 2008). Not only quality
evolves out of coordination (Kittur and Kraut 2008), but
 colaborative efforts (Swarts 2009), diversity (Chen et al.
2010), and conflicts (Arazy et al. 2011; Kittur and Kraut
2010) may also determine the success of Wikipedia. Collobor-
ative authoring thus enables the aggregation of diverse, and
even conflicting, opinions from individual contributors.

Nature of Company Information on Wikipedia

Wikipedia describes firms’ basic profiles. Additionally, and
importantly, Wikipedia aggregates comprehensive and timely
information about firms’ up-to-date activities.

Wikipedia aggregates a comprehensive array of details about
a firm’s activities, such as launching new products, changes
in top management, outsourcing, etc. This type of informa-
tion aggregation contrasts with the summative earnings fore-
casts issued by analysts (Carlson and Zmud 1999). Wikipedia
also covers detailed firm operations more broadly than any
single medium channel, because it aggregates information
from various media. On March 23, 2009, for example, a
modification on Dell’s Wikipedia page suggested that Dell
sold a call center to a French outsourcing firm, Teleperfor-
ance. Only a weekly magazine, BusinessWorld, reported
this deal. In this case, traditional information channels may
leave information holes between parties with asymmetric and
incomplete information (Lin et al. 2005; Schultze and Leidner
2002), while Wikipedia helps reduce information asymmetry by keeping a comprehensive record of facts.

Wikipedia aggregates information in a timely manner. Wiki-
pedia entries for Microsoft and Dell, for example, are edited
multiple times a day, which is more frequent than how often
analysts update their coverage of these companies. On
June 1, 2009, almost immediately after the Dow Jones’
announcement that Cisco was going to replace GM as a
component of the Dow Jones Index, the Wikipedia entry for
Cisco was modified to reflect it. The earliest newspaper
coverage, by the Washington Post, came as late as June 2.
These cases demonstrate that important information about
firm fundamentals becomes public on Wikipedia very quickly.

Next we proceed to analyzing the impact of Wikipedia’s
information aggregation.

Wikipedia and Management Disclosure

Much prior research considers the impact of information
aggregation on market price and returns. Classical asset-
pricing models assume perfect information aggregation in the
financial market (Banks 1985; Radner 1979). Diamond and
Verrecchia’s (1981) model shows that when the process of
information aggregation via market pricing is not perfect,
market outcomes can be different, and Axelson (2007) theo-
retically shows that managers should react (in the context of
securities design) to different levels of investor private infor-
mation. Forsythe and Lundholm (1990) and Sunder (1992)
examine financial information aggregation with experiments
and show evidence that various information aggregation
mechanisms affect stock prices. DellaVigna and Pollet (2009)
show that information aggregation eliminates the effect of
limited attention, thus affecting stock returns. Collectively,
these prior studies demonstrate the impact of information
aggregation on investors. An important implication for our
research is that information aggregation influences investors’
information set.

We carry this implication to the setting of information aggre-
gation on Wikipedia. We have offered some anecdotal evi-
dence (in the “Introduction”) suggesting that Wikipedia can
benefit investors. In the same vein, Gu et al. (2007) suggest
that investors value high-quality linguistic descriptions about
firms. Butler (2001) argues that the key driver of participa-
tion in collaborative communities is that participants can gain
We thus hypothesize: disclosure lag and analyst dispersion would be weakened. As such, the positive relationship between management information even though there is high analyst dispersion. If this is the case, managers would be less able to withhold unfavorable information even though there is high analyst dispersion. According to the withholding perspective, when analyst bias is high, managers’ disclosure lag will be relatively shorter (i.e., a negative relationship between disclosure lag and analyst bias). This is because managers are concerned about the market’s overly optimistic expectation of firm earnings. Wikipedia can help establish a more accurate expectation of firm earnings for investors in two ways. First, when investors are better informed through Wikipedia’s information aggregation, they possess more comprehensive and timely information about firms’ fundamentals. As a result, they can better evaluate the firm’s true performance. Second, Wikipedia supports a neutral point of view (NPOV) stance (Majchrzak 2009). The NPOV policy explicitly prohibits either avoiding or highlighting favorable or unfavorable facts. Wikipedia encourages all users to comply with the NPOV policy and allows anyone to modify a firm’s entry through rounds of addition, deletion, and reorganization. In such an information aggregation process, contributors are able to correct any details of an entry that violate the NPOV policy. Because this policy requires contributors to use neutral words and to only document facts, information aggregation on Wikipedia prevents one point of view from dominating. This is markedly different from analysts’ recommendations, which are often biased (Jacob et al. 1999; Karamanou and Vafeas 2005). Traditional media also have a more emotional touch than Wikipedia entries, be it for ideological (Mullainathan and Shleifer 2005) or financial (Reuter and Zitzewitz 2006) reasons. When the market is replete with an overly optimistic sentiment, the NPOV policy can help investors obtain a more objective assessment of firm performance. Overall, information aggregation on Wikipedia should help alleviate managers’ concerns about high analyst bias, thus weakening the negative relationship between management disclosure lag and analyst bias. Formally, we can write the following hypothesis:

**Hypothesis 1 (H1):** Information aggregation on Wikipedia weakens the relationship between analyst dispersion and management disclosure lag.

According to the preemption perspective, when analyst bias is high, managers’ disclosure lag will be relatively shorter (i.e., a negative relationship between disclosure lag and analyst bias). This is because managers are concerned about the market’s overly optimistic expectation of firm earnings. Wikipedia can help establish a more accurate expectation of firm earnings for investors in two ways. First, when investors are better informed through Wikipedia’s information aggregation, they possess more comprehensive and timely information about firms’ fundamentals. As a result, they can better evaluate the firm’s true performance. Second, Wikipedia supports a neutral point of view (NPOV) stance (Majchrzak 2009). The NPOV policy explicitly forbids either avoiding or highlighting favorable or unfavorable facts. Wikipedia encourages all users to comply with the NPOV policy and allows anyone to modify a firm’s entry through rounds of addition, deletion, and reorganization. In such an information aggregation process, contributors are able to correct any details of an entry that violate the NPOV policy. Because this policy requires contributors to use neutral words and to only document facts, information aggregation on Wikipedia prevents one point of view from dominating. This is markedly different from analysts’ recommendations, which are often biased (Jacob et al. 1999; Karamanou and Vafeas 2005). Traditional media also have a more emotional touch than Wikipedia entries, be it for ideological (Mullainathan and Shleifer 2005) or financial (Reuter and Zitzewitz 2006) reasons. When the market is replete with an overly optimistic sentiment, the NPOV policy can help investors obtain a more objective assessment of firm performance. Overall, information aggregation on Wikipedia should help alleviate managers’ concerns about high analyst bias, thus weakening the negative relationship between management disclosure lag and analyst bias. Formally, we can write the following hypothesis:

**Hypothesis 2 (H2):** Information aggregation on Wikipedia weakens the relationship between analyst bias and management disclosure lag.

**Investor Reaction to Disclosure**

Now we use the lens of investor reaction to see how Wikipedia benefits investors. The disclosure literature has long
observed investor reaction to management disclosure of firm earnings (Kasznik and Lev 1995; Miller 2002; Pownall and Waymire 1989; Waymire 1984). A literature review concludes that “disclosure is associated with stock price performance” (Healy and Palepu 2001, p. 431). In this line of research, Kasznik and Lev (1995) show a significant association between market reaction and unexpected earnings surprise to investors (because of analyst bias). The higher the bias, the more negative is the market reaction. Their study also suggests that if investors receive warnings before management disclosure of bad news, the disclosure is less of a surprise to investors. In the same vein, we propose that when investors obtain detailed and timely information about firms’ economic reality, they adjust their firm valuation accordingly. Later, when managers release bad news about firm earnings, it will trigger a smaller market reaction because of the earlier market adjustment. We, therefore, hypothesize that information aggregation on Wikipedia, which occurs before management disclosure (see Figure 1), would weaken how analyst bias triggers investor reaction.16

**Hypothesis 3 (H3):** Information aggregation on Wikipedia weakens the relationship between analyst bias and investor reaction to management disclosure.

### Method

### Measures

To develop our measures for variables in the research framework (Panel A of Figure 1), we follow the disclosure literature and present a time line of disclosures in Panel B of Figure 1 (Skinner 1997). In a fiscal quarter (between T1 and T2), analysts make forecasts about firm earnings-per-share (EPS), media publish news articles, and contributors modify firm entries on Wikipedia. The disclosure literature assumes that managers know their own firms’ EPS at fiscal quarter end T2, when they realize whether they are confronting forthcoming bad news about EPS (Skinner 1997). The SEC requires public companies to announce their EPS at T4 (the official announcement date). During the period between T2 and T4, managers may voluntarily make disclosures about their firms’ EPS (e.g., at T3).

Consistent with the literature (Kasznik and Lev 1995; Skinner 1994), we measure disclosure lag (LAG) as the number of calendar days between fiscal quarter end (T2) and the date when management voluntarily discloses bad news about EPS (T3).

Analyst bias (BIAS) is the difference between the mean of analyst forecasts of a firm’s EPS during a fiscal quarter (between T1 and T2) and the firm’s actual EPS (Francis et al. 1997; Karamanou and Vafeas 2005). Analyst dispersion (DISPERSION) is the standard deviation of analyst forecasts of a firm’s EPS during a fiscal quarter (between T1 and T2) (Abarbanell et al. 1995; Roulstone 2003).

We measure information aggregation on Wikipedia by Wikipedia modifications (WikiMOD)—specifically, the number of times modifications are made on Wikipedia about a firm in one quarter (between T1 and T2). WikiMOD is a proxy for information aggregation, because by modifying the same firm entry on Wikipedia, individuals add their information, combine different viewpoints about the firm’s operations and performance, and filter out noise and biased statements, thus generating synthesized information about the firm. In order for WikiMOD to play a role in our analysis, a premise is that managers can be informed about the magnitude of WikiMOD, that is, the extent of information aggregation about their firms. This premise is likely to hold. For each company’s Wikipedia entry, there is a “revision history” page on Wikipedia, listing information about each historical modification on that entry (including the time of each modification and the content modified). For example, if one wants to check the modification history of IBM’s Wikipedia entry, s/he can easily see the entire history (http://en.wikipedia.org/w/index.php?title=IBM&action=history) with no need for any special technical skills. In addition, managers can use some publicly available tools, like “Wikipedia Page History Statistics,” which provide the number of modifications for any single Wikipedia entry in the last day/week/month. As an illustration, Appendix C shows the summary statistics for Apple’s Wikipedia page modifications. Using such tools, it may not be too difficult for managers to track and thus realize how many changes were made on their company’s Wikipedia entry. Finally, although our estimation uses the number of modifications as a proxy for information arrivals on Wikipedia, managers do not really need to know the number of modifications of their Wikipedia entries in order to assess how information gets leaked through Wikipedia. A quick skimming of the whole page is likely to give a manager a very good idea about the transparency of the information environment.

It is possible that not all modifications are related to a company’s financial performance (e.g., modifications related to word choice, grammar mistakes, etc.). We refrain from removing these unrelated modifications because that would unavoidably involve subjective human judgment. These modifications are likely to be orthogonal to the variables of

---

16We appreciate the Associate Editor’s suggestion to formally putting forth this hypothesis.
interest. Hence, including these modifications would not affect the direction of our estimation results, while reducing the efficiency of the estimation. As such, the introduction of noise would bias against our evidence. Another potential concern is that WikiMOD may involve modifications in “edit wars” (i.e., back-and-forth changes due to disagreement between contributors). As reported later in our sensitive analysis in Appendix F, our results hold up after we remove possible back-and-forth changes.17 We also note that if there are other social-media platforms that can achieve information aggregation, focusing on Wikipedia is only likely to bias our results downward. Therefore, our conclusion regarding the effect of information aggregation would be conservative.

As for investor reaction to bad news, we follow the disclosure literature (Kasznik and Lev 1995) and develop a proxy measure by computing the sum of market-adjusted returns in two windows: a five-day window around voluntary disclosure and a five-day window around the official EPS announcement. Market-adjusted returns (RET) in the combined two windows represent market reaction to bad news. In addition, we estimate cumulative abnormal returns (CAR) in the combined two windows.18 These two methods yield highly consistent results, so the discussion below is based on one method (CAR).

**Empirical Model for Management Disclosure**

To test H1 and H2, we use a hazard regression model, developed in the statistics literature to assess the impact of explanatory variables on the timing of an event. In our research context, the event under investigation is management transition from withholding to disclosure, and the timing of the event refers to the disclosure lag (LAG). Because the distribution of event timing is often far from normal, a hazard model is usually superior to ordinary least squares regression (Kalbfleisch and Prentice 1980). The dependent variable of the hazard analysis is the hazard rate \( h(t) \), which is the probability of management transition from withholding to disclosure at time \( t \). Following Kauffman et al. (2000), we specify a Cox proportional-hazard model as follows:

\[
h(t) = h_0(t) \exp(-βX)
\]

where \( X \) is a vector of explanatory variables, \( β \) is a row vector of coefficients to be estimated, and \( h_0(t) \) is a hazard function with \( X = 0 \). It is worth noting that we specify \( \exp(-βX) \) rather than \( \exp(βX) \), because with this specification, the coefficient \( β \) tells how the explanatory variables \( X \) affect the disclosure lag as follows: if an explanatory variable has a positive coefficient, it reduces the hazard rate \( h(t) \) and thus increases LAG; conversely, a negative coefficient indicates an effect to decrease LAG. In the literature, Kauffman et al. use a similar approach for exposition reasons.

Because H1 and H2 are about whether WikiMOD moderates how LAG relates to analyst characteristics (DISPERSION and BIAS), the explanatory variables \( X \) in model (0) include DISPERSION, BIAS, WikiMOD, WikiMOD*DISPERSION, and WikiMOD*BIAS. We have the following hazard model for testing H1 and H2:

\[
h(t) = h_0(t) \exp(-θ_1DISPERSION - θ_2BIAS - θ_3WikiMOD - θ_4WikiMOD*DISPERSION + θ_5WikiMOD*BIAS + ζ)\]

We need two sets of control variables. First, we need to include controls identified by the disclosure literature as directly affecting management disclosure:

- We control for firm risk, measured as earnings variability (VAR)—that is, the standard deviation of earnings across the eight previous fiscal quarters (Kothari et al. 2002). To compensate for high risk resulting from high earnings variability, investors tend to demand high returns that increase the firm’s costs of capital. Voluntary disclosure may help reduce costs of capital by providing more information to investors (Francis et al. 2008). As such, high earnings variability may motivate firms to disclose earlier.

- The literature suggests the need to control for firm size (MV), measured as firm market value at the beginning of a fiscal quarter, because larger firms are more exposed to litigation risks (Kasznik and Lev 1995).

- The literature suggests the need to control for two specific industry sectors: high tech (HIGHTECH) and regulated industry (REG). High-tech firms are likely to be exposed to a larger-than-average risk of shareholder lawsuits and thus may be motivated to disclose to deter investors’ litigation (Kasznik and Lev 1995).19 Yet,
The recent literature on media impact suggests that among the reasons for shareholder lawsuits are high-tech firms’ aggressive accounting techniques (e.g., excessive capitalization of software development costs). Disclosing poor EPS may not help fend off such lawsuits (Kasznik and Lev 1995). In regulated industries, firms may be obliged to release information about their operations, in addition to quarterly financial reports. But firms in regulated industries may be perceived as having low risk (Dewan et al. 2007), which may reduce firms’ incentive to disclose. We thus include the two industry dummies (HIGHTECH and REG) but do not predict their signs.

• We also control for fiscal quarters. Managers are more likely to withhold bad news in the fourth quarter than in earlier quarters (Baginski et al. 1994), as releasing bad news in the last fiscal quarter affects firm performance for the entire fiscal year.

Second, we need to control for news coverage, because modifications on Wikipedia and news coverage may be triggered by common events. If news coverage also informs investors, we need to address their moderation effects in model (1) as well. This helps rule out alternative explanations for the moderating role of Wikipedia modifications.

• The recent literature on media impact suggests that readers are sensitive to newsworthy content, which can be measured with the amount of positive words (POSITIVE) and negative words (NEGATIVE) in news articles (Tetlock et al. 2008). Given the same level of dispersion and uncertainty, an increased level of positive news affords managers the possibility to take advantage of it, causing further delay in disclosure. We therefore expect a positive interaction effect between POSITIVE and DISPERSION on disclosure lag; conversely, we expect a negative interaction effect between NEGATIVE and DISPERSION. Regarding the moderation on BIAS, because a positive media bias means that EPS is already below expectation, negative news may further hasten managers to disclose bad news to minimize damage. We therefore expect a negative interaction effect between NEGATIVE and BIAS on disclosure lag; conversely, we expect a positive interaction effect between POSITIVE and BIAS.

• We count the number of news articles (NEWS) in the business press about a firm in one quarter (Antweiler and Frank 2004; Frankel and Li 2004). We expect the role of NEWS is similar to that of NEGATIVE, because prior research finds that readers are generally more affected by negative than positive words (Tetlock et al. 2008).

**Empirical Model for Investor Reaction**

To test H3, our analysis builds on the management disclosure literature (Kasznik and Lev 1995), which has established a regression model relating analyst bias (BIAS) to investor reaction (i.e., cumulative abnormal returns—CAR—as defined above). We extend the model by adding the moderating effect of WikiMOD, and have an ordinary least squares (OLS) model as follows:

\[ \text{CAR} = \alpha + \eta \text{BIAS} + \delta_1 \text{WikiMOD} + \delta_2 \text{WikiMOD*BIAS} + \zeta \text{controls} + \varepsilon \]  

Regarding controls in the above model, we follow prior research and include firm size, which moderates the effect of BIAS, in that investors are more tolerant of analyst bias for large firms than for small firms (Kasznik and Lev 1995). The reason is that poor performance in one quarter may have less impact on large firms’ long-term economic viability as compared with that of small firms (Kasznik and Lev 1995). We also include news articles and newsworthy content, for the abovementioned reasons. If negative news content can inform investors about forthcoming bad news, it can mitigate investor reaction to bad news, while positive news content may play an opposite role. Regarding the number of news articles, we expect that its role is similar to that of negative news (Tetlock et al. 2008).

**Data and Sample**

We collect data from several sources: (1) management disclosure data from the First Call Historical Database (FCHD), (2) editing history data from Wikipedia, (3) firm data from Compustat and Center for Research in Security Prices (CRSP), and (4) news-coverage data from Lexis-Nexis. Table 1 summarizes the measures and data sources of all variables, which are described in detail below.
<table>
<thead>
<tr>
<th>Variable</th>
<th>Computation</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Theoretical Variables</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>LAG</strong> (disclosure lag)</td>
<td>The number of calendar days between fiscal quarter end and the date when management voluntarily discloses bad news about EPS (i.e., between T2 and T3 in Figure 1)</td>
<td>First Call</td>
</tr>
<tr>
<td><strong>DISPERSION</strong> (analyst dispersion)</td>
<td>The standard deviation of analyst forecasts about a firm’s EPS in a fiscal quarter (i.e., between T1 and T2 in Figure 1), scaled by the mean EPS forecast</td>
<td>First Call</td>
</tr>
<tr>
<td><strong>BIAS</strong> (analyst bias)</td>
<td>The absolute value of the difference between firm EPS in a quarter and the mean EPS forecast made by analysts in the quarter (i.e., between T1 and T2 in Figure 1), scaled by the mean EPS forecast</td>
<td>First Call</td>
</tr>
<tr>
<td><strong>WikiMOD</strong> (Wikipedia modifications—a proxy for information aggregation on Wikipedia)</td>
<td>Natural logarithm of one plus the number of Wikipedia modifications for a company in a fiscal quarter (between T1 and T2 in Figure 1)</td>
<td>Wikipedia</td>
</tr>
<tr>
<td><strong>CAR</strong> (cumulative abnormal returns)</td>
<td>Sum of abnormal returns in a combined window including 5 days around voluntary disclosure (i.e., T3 in Figure 1) and 5 days around the official earnings announcement (i.e., T4 in Figure 1)</td>
<td>CRSP</td>
</tr>
<tr>
<td><strong>RET</strong> (market adjusted returns)</td>
<td>Sum of market adjusted returns (i.e., less market average returns) in a combined window including 5 days around voluntary disclosure (i.e., T3 in Figure 1) and 5 days around the official earnings announcement (i.e., T4 in Figure 1)</td>
<td>CRSP</td>
</tr>
<tr>
<td><strong>Control Variables</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>NEWS</strong> (number of news articles)</td>
<td>Natural logarithm of one plus the number of news articles about a company in a fiscal quarter (between T1 and T2 in Figure 1), as documented in Lexis-Nexis</td>
<td>Lexis-Nexis</td>
</tr>
<tr>
<td><strong>POSITIVE</strong> (positive news)</td>
<td>Standardized proportion of positive words. Proportion of positive words (POS%) is calculated as total number of positive words for a company in a fiscal quarter divided by total number of words for that company in that quarter. We calculate the mean (Mean_POS) and the standard deviation (Sd_POS) of the proportion of positive words for the prior calendar year, ( \text{POSITIVE} = (\text{POS}% - \text{Mean}_\text{POS}) / \text{Sd}_\text{POS} ).</td>
<td>Lexis-Nexis</td>
</tr>
<tr>
<td><strong>NEGATIVE</strong> (negative news)</td>
<td>NEGATIVE is calculated similar to POSITIVE, with the proportion of negative words.</td>
<td>Lexis-Nexis</td>
</tr>
<tr>
<td><strong>VAR</strong> (earnings variability)</td>
<td>The standard deviation of quarterly earnings across eight fiscal quarters before the quarter under examination (i.e., prior to T1 in Figure 1)</td>
<td>Compustat</td>
</tr>
<tr>
<td><strong>MV</strong> (market value)</td>
<td>Natural logarithm of firm market value at the beginning of the fiscal quarter (i.e., at T1 in Figure 1)</td>
<td>Compustat</td>
</tr>
<tr>
<td><strong>HIGHTECH</strong> (high-tech industries)</td>
<td>HIGHTECH equals 1 when the firm belongs to Drugs (SIC 2833-2836), R&amp;D Services (8731-8734), Programming (7371-7379), Computers (3570-3577), or Electronics (3600-3674), and 0 otherwise</td>
<td>Compustat</td>
</tr>
<tr>
<td><strong>REG</strong> (regulated industries)</td>
<td>REG equals 1 when the firm belongs to Telephone (SIC 4812-4813), TV (4833), Cable (4841), Communications (4811-4899), Gas (4922-4924), Electricity (4931), Water (4941), or Financial sectors (6021-6023, 6035-6036, 6141, 6311, 6321, 6331), and 0 otherwise</td>
<td>Compustat</td>
</tr>
</tbody>
</table>
We obtain access to FCHD from Wharton Research Data Services (WRDS). FCHD reports the history of analysts’ estimates of companies’ EPS, based on which we compute analyst bias (BIAS) and dispersion (DISPERSION). FCHD also records company-issued guidelines about forthcoming EPS and labels whether the guidelines are bad news. We use this information to identify bad news and calculate disclosure lag (LAG).

FCHD contains quarterly earnings information for 8,500 U.S. securities; we limit our attention to common stocks. For each stock, we obtain firm identity (provided by FCHD) and then search on Wikipedia for the corresponding entry. A PERL program is used to search our list of companies on Wikipedia. After obtaining the URLs of these entries, we manually go over all of these Wikipedia pages to ensure that the entries are correctly matched to the companies. This practice yields 375 Wikipedia entries of public companies. To obtain Wikipedia revision information, we use a software robot program to collect the complete “revision history” of all these entries.

After comparing each revision with the previous revision, we calculate the number of words added or deleted and record the number of modifications (WikiMOD) in each fiscal quarter for each firm. Overall, for the period between March 21, 2001, and May 19, 2006, 8,789 registered users and 5,450 unregistered users contributed a total of 77,921 modifications on these Wikipedia firm entries.

We manually search the Lexis-Nexis database for news articles about each of the 375 companies in our final sample. For Lexis-Nexis’ HTML output, we develop a program to parse the result pages and record the newspaper name, date, and content of each piece of news. These are our raw data for quantifying news coverage (NEWS) and analyzing news content. We follow prior research to conduct a content analysis to quantify the language used in news stories. We compute the proportions of positive words (POSITIVE) and negative words (NEGATIVE) in all news stories related to a firm in one quarter (between T1 and T2).

The algorithm of our computation comes directly from the literature on stock-market responses to news content (Das and Chen 2007; Tetlock 2007; Tetlock et al. 2008).

We obtain stock-return data from CRSP and use the data to calculate market-adjusted returns (RET) and cumulative abnormal returns (CAR), as defined above. We use firm data from Compustat for earnings variability (VAR), market value (MV), a high-tech industry dummy (HIGHTECH), and a regulated industry dummy (REG).

Our research context adds two necessary restrictions to our sample. First, a firm’s EPS in a quarter should be below the mean EPS forecast made by analysts during the same quarter. This implies a situation in which managers faced forthcoming bad news. Second, the firm should voluntarily disclose bad news about EPS (between T2 and T4 in Figure 1). These two restrictions reduce our sample to a set of 161 warnings (i.e., voluntary disclosures of bad news about EPS) released by 96 public companies during the observation period when the Wikipedia modification history is available to us. These 161 warnings form our final sample used in the subsequent analysis.

Summary Statistics

Table 2 reports summary statistics. In the final sample, the mean disclosure lag is 14.11 days, and the maximum is 43

\[ \text{NEGATIVE} = \frac{\mu_{\text{NEG}}}{\sigma_{\text{NEG}}} \]

When companies make voluntary earnings disclosures in press releases and interviews, the disclosures can be a number (e.g., $2.35), a range of numbers (e.g., from $2.35 to $2.68), or a simple text item (e.g., the earnings will be above/below analysts’ expectations). FCHD reports the form of each specific disclosure and uses a field to describe the disclosure and indicate whether it is a positive or negative surprise. Please refer to FCHD Technical Guide for further details.


Wikipedia attributes modifications to users’ registered names if contributors log into the system. Anonymous contributors who do not log in are identified only by their IP addresses. We exclude all modifications contributed by anonymous users in our subsequent analysis because, given the low cost of anonymous editing, spammers and abusers often contaminate entries and because knowledge contribution on social media is found to be strongly associated with identity verification (Ma and Agarwal 2007; Wasko and Faraj 2005; Zhang and Zhu 2011).

NEGATIVE is a standardized measure of the proportion of negative words in all news stories related to a firm (Tetlock et al. 2008). Specifically, we calculate the proportion of negative words (\( \text{NEG} \)), the total number of negative words for a firm in a fiscal quarter divided by the total number of words for that firm in the same quarter. We then calculate the mean (\( \mu_{\text{NEG}} \)) and the standard deviation (\( \sigma_{\text{NEG}} \)) of the proportion of negative words for the prior calendar year, and define \( \text{NEGATIVE} = (\text{NEG} - \mu_{\text{NEG}}) / \sigma_{\text{NEG}} \). Similarly, we compute \( \text{POSITIVE} = (\text{POS} - \mu_{\text{POS}}) / \sigma_{\text{POS}} \). To categorize the words as positive or negative, we follow Tetlock (2007) and Tetlock et al. (2008) and use the Harvard-IV-4 psychosocial dictionary.

Table 2. Summary Statistics (Variable Definitions Are in Table 1)

<table>
<thead>
<tr>
<th>Variable</th>
<th>LAG</th>
<th>DISPER-</th>
<th>BIAS</th>
<th>NEWS</th>
<th>Wiki-</th>
<th>POSITIVE</th>
<th>NEGATIVE</th>
<th>VAR</th>
<th>MV</th>
<th>HIGH</th>
<th>TECH</th>
<th>REG</th>
<th>CAR</th>
<th>RET</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>2.2667</td>
<td>0.3246</td>
<td>1.0083</td>
<td>2.0316</td>
<td>1.5947</td>
<td>−1.0366</td>
<td>0.0491</td>
<td>0.4418</td>
<td>18.1280</td>
<td>0.3625</td>
<td>0.0188</td>
<td>−0.0484</td>
<td>−0.0684</td>
<td></td>
</tr>
<tr>
<td>Std. Dev.</td>
<td>0.8765</td>
<td>0.8691</td>
<td>3.0376</td>
<td>1.3455</td>
<td>1.0907</td>
<td>1.6965</td>
<td>1.3778</td>
<td>1.4268</td>
<td>1.2513</td>
<td>0.4822</td>
<td>0.1361</td>
<td>0.1112</td>
<td>0.1126</td>
<td></td>
</tr>
</tbody>
</table>

Bivariate Correlations (p-values shown in parentheses)

<table>
<thead>
<tr>
<th>Variable</th>
<th>LAG</th>
<th>DISPERSION</th>
<th>BIAS</th>
<th>NEWS</th>
<th>WikiMOD</th>
<th>POSITIVE</th>
<th>NEGATIVE</th>
<th>VAR</th>
<th>MV</th>
<th>HIGHTECH</th>
<th>REG</th>
<th>CAR</th>
<th>RET</th>
</tr>
</thead>
<tbody>
<tr>
<td>LAG</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>BIAS</td>
<td>0.151</td>
<td>0.499</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>NEWS</td>
<td>−0.231</td>
<td>−0.155</td>
<td>−0.008</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>WikiMOD</td>
<td>−0.158</td>
<td>−0.132</td>
<td>−0.051</td>
<td>0.332</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>POSITIVE</td>
<td>−0.136</td>
<td>0.054</td>
<td>0.056</td>
<td>−0.202</td>
<td>−0.019</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>NEGATIVE</td>
<td>−0.018</td>
<td>0.086</td>
<td>−0.002</td>
<td>−0.021</td>
<td>−0.061</td>
<td>−0.108</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>VAR</td>
<td>−0.122</td>
<td>0.220</td>
<td>0.418</td>
<td>0.066</td>
<td>−0.140</td>
<td>0.091</td>
<td>0.030</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>MV</td>
<td>−0.260</td>
<td>−0.246</td>
<td>−0.303</td>
<td>0.445</td>
<td>0.354</td>
<td>−0.056</td>
<td>−0.128</td>
<td>−0.227</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>HIGHTECH</td>
<td>0.121</td>
<td>0.272</td>
<td>0.203</td>
<td>−0.267</td>
<td>−0.184</td>
<td>0.045</td>
<td>−0.082</td>
<td>0.224</td>
<td>−0.234</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>REG</td>
<td>0.026</td>
<td>−0.050</td>
<td>−0.046</td>
<td>0.009</td>
<td>−0.144</td>
<td>0.038</td>
<td>0.007</td>
<td>0.037</td>
<td>0.016</td>
<td>−0.104</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>CAR</td>
<td>−0.080</td>
<td>−0.226</td>
<td>−0.250</td>
<td>0.074</td>
<td>0.007</td>
<td>−0.054</td>
<td>−0.056</td>
<td>−0.195</td>
<td>0.215</td>
<td>−0.254</td>
<td>0.004</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>RET</td>
<td>−0.113</td>
<td>−0.221</td>
<td>−0.259</td>
<td>0.086</td>
<td>0.020</td>
<td>−0.071</td>
<td>−0.053</td>
<td>−0.198</td>
<td>0.225</td>
<td>−0.247</td>
<td>0.007</td>
<td>0.993</td>
<td>1</td>
</tr>
</tbody>
</table>

Both Wikipedia modifications (WikiMOD) and news coverage (NEWS) are positively correlated with firm size (MV), suggesting that larger firms receive more media coverage. Firm size is negatively correlated with earnings variability (VAR) and analyst dispersion (DISPERSION), suggesting that larger firms have less uncertainty in earnings and may release more information to analysts (Kothari et al. 2002; Lang and Lundholm 1996). These correlations together may explain why WikiMOD is negatively correlated with VAR and DISPERSION.

Disclosure lag is negatively correlated with WikiMOD and NEWS, suggesting that information arrivals may result in earlier disclosure. Given the positive correlation between WikiMOD and NEWS, however, this univariate analysis cannot tell us whether LAG is attributable to information aggregation or news arrivals per se, or both. LAG is also negatively correlated with firm size, because larger firms are more exposed to litigation risks than small firms (Kasznik and Lev 1995). Hence, large firms are more likely to disclose earlier to immunize themselves from litigation. LAG is positively correlated with analyst dispersion, suggesting that announcements occur later, in the presence of higher information uncertainty in the market.

Regarding analyst forecasts, both analyst bias and analyst dispersion are negatively correlated with firm size and positively correlated with earnings variability and the high-tech dummy. It could be the case that more information is available for larger firms, making analysts’ forecasting relatively easier (Lang and Lundholm 1996). In addition, high-tech firms feature greater earnings variability, which may make it more difficult for analysts to predict future EPS (Dewan et al. 2002).
2007). The correlation between BIAS and DISPERSION is positive, suggesting that analysts’ forecasts are more accurate when they agree more with each other. Overall, the correlations are consistent with observations in previous research.

Results

Results about Management Disclosure

To test H1 and H2, Table 3 reports regression results for model (1). The software package used is R and the function is coxph. From columns (1) through (5), Table 3 presents results when explanatory variables are added step by step. The purpose is to compare different model specifications to check the robustness of the results. We mean-center variables involved in interactions to ease the interpretation of results. We discuss the results in turn below.

Analyst Forecasts (DISPERSION and BIAS)

We first examine the impact of analyst dispersion and bias. In column (1) of Table 3, the coefficients on DISPERSION and BIAS represent their average (or constant) effects across all values of moderators (Aiken and West 1991). In columns (2) through (5), as we add moderators, the coefficients on DISPERSION and BIAS change in magnitude and significance. This is not surprising because in columns (2) through (5), the coefficients on DISPERSION and BIAS represent their impact conditioned on the mean values of the moderators (Aiken and West 1991).

As show in column (1) of Table 3, the positive sign of DISPERSION indicates that its average effect is to increase disclosure lag. This confirms the notion that when the external information environment regarding firm EPS is highly dispersed, managers tend to delay disclosure. The negative sign of BIAS indicates that its average effect is to decrease disclosure lag. This supports the notion that managers are more likely to disclose unfavorable information when analysts are far off in estimating firm earnings.

Information Aggregation on Wikipedia (WikiMOD)

Regression results in Table 3 support our hypotheses about information aggregation on Wikipedia (H1 and H2). As seen in Table 3, the results for WikiMOD are consistent across columns (2) through (5). Below we use column (2) to present the results.

H1, supported: The average effect of DISPERSION in column (1) suggests that managers tend to withhold adverse information if they enjoy an information advantage. The interaction with Wikipedia modification (−0.1321, p < 0.05, column (2) of Table 3), however, moderates the effect. The interaction’s negative sign suggests that, in the presence of more Wikipedia modifications, the extent to which managers leverage their information advantage is weakened. This supports our hypothesis that WikiMOD may mitigate the impact of DISPERSION on disclosure lag, such that the positive relationship between DISPERSION and disclosure lag is less conspicuous in the presence of higher WikiMOD.

H2, supported: The average effect of BIAS in column (1) suggests that when analysts are too optimistic in estimating earnings, managers typically shorten the lag to offer information more promptly. The significant interaction between WikiMOD and BIAS (0.0623, p < 0.05, column (2) of Table 3), however, moderates the effect. If the reason behind an earlier release of bad news is managers’ concern about large stock-price declines on official announcement dates, then Wikipedia modifications may play the role of releasing information to the public earlier, thereby alleviating that concern. As such, when there are more Wikipedia modifications, managers are not as worried as before. The positive sign on WikiMOD*BIA supports our hypothesis that WikiMOD may mitigate the impact of BIAS on disclosure lag, such that the negative relationship between BIAS and disclosure lag is less severe in the presence of higher WikiMOD.

Number of News Articles and News Content

As shown in Table 3, the coefficients on NEGATIVE are negative, suggesting that when there are more negative words about a company, managers tend to disclose earlier. This is consistent with our expectation and suggests a plausible “backfire” effect associated with NEGATIVE. Managers in our research context face forthcoming bad news given an upward market bias introduced by analysts. If investors follow analysts first and realize analyst bias later (on the official announcement date), then the inconsistency between market expectation and NEGATIVE would lead investors to suspect that managers had withheld unfavorable information. The inconsistency would backfire, exposing managers to risks resulting from negative earnings surprises.

The interaction between NEGATIVE and BIAS is negative as expected, although the significance level is weak. The interaction between NEGATIVE and DISPERSION is significant and positive, different than our expectation. The positive interaction suggests that NEGATIVE’s effect of shortening disclosure lag would be salient when DISPERSION is low.
Table 3. Wikipedia and Management Disclosure (H1 and H2) (Regression specification is based on model (1) in text. Variable definitions are in Table 1.)

<table>
<thead>
<tr>
<th>Theoretical Variables</th>
<th>Analysts</th>
<th>Aggregation via Wikipedia</th>
<th>Number of News Articles</th>
<th>Content of News (Positive &amp; Negative Words)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
</tr>
<tr>
<td>Wikipedia Modifications (WikiMOD)</td>
<td>-0.1350**</td>
<td>-0.1048*</td>
<td>-0.1255**</td>
<td>-0.1206*</td>
</tr>
<tr>
<td>WikiMOD*DISPERSION</td>
<td>-0.1321**</td>
<td>-0.1495***</td>
<td>-0.3480***</td>
<td>-0.3473***</td>
</tr>
<tr>
<td>WikiMOD*BIA S</td>
<td>0.0623**</td>
<td>0.0672***</td>
<td>0.1036***</td>
<td>0.0990***</td>
</tr>
<tr>
<td>Analyst Dispersion (DISPERSION)</td>
<td>0.1732**</td>
<td>0.1395**</td>
<td>0.1711**</td>
<td>0.0724</td>
</tr>
<tr>
<td>Analyst Bias (BIAS)</td>
<td>-0.0349*</td>
<td>-0.0378**</td>
<td>-0.0322**</td>
<td>0.0121</td>
</tr>
</tbody>
</table>

| Control Variables | | | | |
| Number of News Articles (NEWS) | -0.0995* | -0.0987* | -0.0947* |
| NEWS*DISPERSION | 0.0565 | 0.2613** | 0.2476 |
| NEWS*BIA S | -0.0113 | -0.0562** | -0.0531 |
| Negative Words (NEGATIVE) | -0.0804** | -0.0929* |
| NEGATIVE*DISPERSION | 0.2862*** | 0.2661* |
| NEGATIVE*BIA S | -0.0667* |
| Positive Words (POSITIVE) | -0.0726 |
| POSITIVE*DISPERSION | -0.0201 |
| POSITIVE*BIA S | -0.0052 |
| Earnings Variability (VAR) | -0.1413*** | -0.1240*** | -0.1099*** | -0.0920*** |
| Market Value (MV) | -0.2973*** | -0.2536*** | -0.2325*** | -0.2296*** |
| High-tech (HITECH) | 0.2423 | 0.2129 | 0.1387 | 0.1785 |
| Regulation (REG) | 0.0420 | -0.1332 | -0.0930 | -0.1137 |
| Quarter 1 Dummy (Q1) | -0.4517*** | -0.5662*** | -0.5431*** | -0.4925** |
| Quarter 2 Dummy (Q2) | -0.0826 | -0.1424 | -0.1405 | -0.0869 |
| Quarter 3 Dummy (Q3) | -0.4644** | -0.5340*** | -0.5413*** | -0.5587*** |
| Observations | 161 | 161 | 161 | 161 |
| Likelihood Ratio Test | p = 0.001 | p < 0.001 | p < 0.001 | p = 0.001 |
| Pseudo R-Squared | 0.158 | 0.203 | 0.215 | 0.230 |

Note: Robust estimates of standard errors are reported in parentheses. Significance levels are one-tailed for directional predictions and two-tailed otherwise. *p < 0.10; **p < 0.05; ***p < 0.01. The pseudo R² refers to Nagelkerke’s (1991) R².
This implies that managers may perceive the risk associated with negative media content (i.e., the threat of backfire as mentioned above) to increase as DISPERSION decreases. When analysts are consistent with each other (i.e., low DISPERSION), their forecasts may become more convincing to investors. This would expose managers to a greater threat of backfire; consequently, managers may disclose earlier to minimize damage.

The impacts of NEWS are similar to those of NEGATIVE. The impacts of POSITIVE are not significant. These findings are consistent with prior research suggesting that readers are more affected by negative than positive news content (Tetlock et al. 2008).

Overall, our results seem to indicate that, unlike the salient role played by WikiMOD, some variables capturing media coverage turn out to be not significant. Similar to our result, Frankel and Li (2004) find that media coverage does not seem to improve investors’ information environment. A possible explanation offered by the literature is media bias. The literature on media bias suggests that media are often biased for both demand- and supply-side reasons (Knight and Chiang 2008). On the demand side, consumers may prefer to consume information that confirms their prior beliefs; profit-maximizing publishers thus have incentives to bias their reports (Gentzkow and Shapiro 2006; Mullainathan and Shleifer 2005). On the supply side, individual journalists and financial analysts may also bias reports to reflect their own views or achieve their own financial objectives (Baron 2006; Reuter and Zitzewitz 2006). As a result, financial reports in the media often slant information through “selective omission, choice of words, and varying credibility ascribed to the primary source” (Gentzkow and Shapiro 2006, p. 281). Wikipedia, owing to its NPOV policy and because contributions on Wikipedia follow a democratic and collaborative model, is less susceptible to media bias. Prior studies establish that this kind of “wisdom-of-crowds” production model tends to generate facts (Kittur and Kraut 2008; Swarts 2009). If this is the case, one may expect that Wikipedia, instead of media, plays a salient role in improving the market’s information environment. Our results offer supportive evidence, although we cannot completely rule out the possibility of the existence of bias on Wikipedia.

Results about Investor Reaction

To test H3, Table 4 presents the regression results for model (2). Columns (1) and (2) report regressions using CAR over a five-day time window to measure investors’ reaction; columns (3) and (4) examine CAR over a three-day time window, for a robustness check; and columns (5) through (8) examine market-adjusted returns (RET) based on a five-day time window and a three-day time window. These columns show highly consistent results. Below, we use columns (1) and (2) to discuss our findings.

Column (1) of Table 4 establishes the baseline result. Since our sample includes only bad news, the market reaction should be negative. As expected, the more optimistic the analysts, the more disappointed are the investors (i.e., a negative coefficient on BIAS).

H3, supported: In column (2), WikiMOD*BIAS has a positive sign, suggesting that WikiMOD moderates investor reaction to bad news. Had there existed a medium that disclosed information to the public, this medium would have offset the surprise caused by BIAS. Our results suggest that WikiMOD seems to play such a role, in support of H3.

Regarding controls, firm size plays a positive moderating role as expected. NEWS and NEGATIVE play a negative moderating role. These results suggest that (1) more news coverage exacerbates the problem of overly optimistic analyst estimates, and (2) when there is an upward analyst bias, investors are more disappointed if traditional media used many negative words to describe that firm. The first effect seems to suggest that more news coverage may merely magnify the effect of analyst bias. Or, investors are generally affected by negative instead of positive words, so NEWS and NEGATIVE play similar roles. Judging from the sign, we can at least argue that more news coverage does not offer more information to reduce the impact of analyst bias or improve the information environment (Frankel and Li 2004). The second effect is consistent with the backfire story discussed above. Given the negative coefficient on BIAS, and the nonsignificant coefficient on NEGATIVE, it is plausible to argue that investors tend to “trust” analysts more than newspapers. That is, investors seem to follow analysts’ upward-biased recommendation, even when newspapers use many negative words to describe a company. When analysts are wrong, the inconsistency between news and analysts’ optimistic estimates would backfire. This may help explain why the interaction between NEGATIVE and BIAS is significantly negative.

Additional Analysis

Possible Endogeneity of WikiMOD: As with other empirical studies, we cannot control for all possible information sources to completely rule out the existence of alternative channels through which managers and investors get informed. Endogeneity concerns are legitimate in this context for two reasons.
Table 4. Investor Reaction to Disclosure (H3) (Regression specification is based on model (2) in text. Variable definitions are in Table 1.)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Analyst Bias (BIAS)</td>
<td>-0.0085*** (0.0019)</td>
<td>-0.0027*** (0.0321)</td>
<td>-0.0053*** (0.0222)</td>
<td>-0.0057*** (0.0332)</td>
</tr>
<tr>
<td>WikiMOD</td>
<td>-0.0084 (0.0099)</td>
<td>-0.0087 (0.0102)</td>
<td>-0.0074 (0.0099)</td>
<td>-0.0075 (0.0103)</td>
</tr>
<tr>
<td>WikiMOD*BIAS</td>
<td>0.0035*** (0.0015)</td>
<td>0.0033** (0.0018)</td>
<td>0.0035** (0.0015)</td>
<td>0.0033** (0.0019)</td>
</tr>
<tr>
<td>Control Variables</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Market Value (MV)</td>
<td>0.0057 (0.0096)</td>
<td>0.0133 (0.0099)</td>
<td>0.0061 (0.0099)</td>
<td>0.0134 (0.0100)</td>
</tr>
<tr>
<td>MV*BIAS</td>
<td>0.0046*** (0.0017)</td>
<td>0.0034** (0.0019)</td>
<td>0.0045*** (0.0019)</td>
<td>0.0035** (0.0020)</td>
</tr>
<tr>
<td>NEWS</td>
<td>0.0075 (0.0088)</td>
<td>0.0034 (0.0086)</td>
<td>0.0087 (0.0088)</td>
<td>0.0045 (0.0087)</td>
</tr>
<tr>
<td>NEWS*BIAS</td>
<td>-0.0039*** (0.0013)</td>
<td>-0.0056*** (0.0015)</td>
<td>-0.0042*** (0.0015)</td>
<td>-0.0059*** (0.0015)</td>
</tr>
<tr>
<td>NEGATIVE</td>
<td>0.0025 (0.0045)</td>
<td>0.0022 (0.0035)</td>
<td>0.0031 (0.0044)</td>
<td>0.0025 (0.0034)</td>
</tr>
<tr>
<td>NEGATIVE*BIAS</td>
<td>-0.0100*** (0.0042)</td>
<td>-0.0104** (0.0036)</td>
<td>-0.0109** (0.0047)</td>
<td>-0.0109*** (0.0038)</td>
</tr>
<tr>
<td>POSITIVE</td>
<td>-0.0016 (0.0036)</td>
<td>-0.0022 (0.0019)</td>
<td>-0.0029 (0.0036)</td>
<td>-0.0033* (0.0020)</td>
</tr>
<tr>
<td>POSITIVE*BIAS</td>
<td>0.0000 (0.0033)</td>
<td>-0.0015 (0.0034)</td>
<td>0.0008 (0.0037)</td>
<td>-0.0007 (0.0037)</td>
</tr>
<tr>
<td>Observations</td>
<td>161</td>
<td>161</td>
<td>161</td>
<td>161</td>
</tr>
<tr>
<td>R–Squared</td>
<td>0.0752</td>
<td>0.1498</td>
<td>0.0307</td>
<td>0.1323</td>
</tr>
</tbody>
</table>

Note: Robust estimates of standard errors are reported in parentheses. Significance levels are one-tailed for directional predictions and two-tailed otherwise. *p < 0.10; **p < 0.05; ***p < 0.01.

First, it is possible that managers may release information through Wikipedia themselves. If a manager decides to release information through Wikipedia, and if the decision to edit the entry correlates with the decision about disclosure lag (e.g., because of an incentive to release information to the public), then the manager’s self-editing may confound Wikipedia’s effect.

Second, if there exists an alternative information channel that influences both Wikipedia modifications and disclosure lags, then the identified effect of WikiMOD cannot be attributed to Wikipedia. More broadly, if there are some omitted factors that affect disclosure lags and are correlated with Wikipedia modifications, endogeneity is a concern. We address these endogeneity issues by conducting an instrumental variable (IV) estimation with two possible IVs. Our results suggest that these concerns are not serious. Appendix D reports the details of our IV estimation.

An Alternative Explanation: We need to control for the possible impact of firm visibility. It is an alternative explanation for the impact of WikiMOD, in that firm visibility is correlated with WikiMOD and may influence management disclosure. We reviewed prior studies on firm visibility and found that they used three sets of proxy variables for firm visibility. The first set refers to fundamental firm characteristics that are associated with firm visibility, including firm size (e.g., market value), firm profitability (e.g., ROA), firm age, and advertising expenditure. Firms with a larger size and greater profitability and older firms may be associated with higher levels of visibility (Bushee and Miller 2012, Grullon et al. 2004), and advertising may also help increase firm visibility (Grullon et al. 2004). The second set refers to media coverage in that the amount of news reports tends to covary with firm visibility. The third set refers to firm size and other characteristics, including industry affiliation, stock price, and other factors that may affect visibility.

26We thank an anonymous reviewer for suggesting this important test.
with firm visibility (Bushee and Miller 2012). The third set concerns firm visibility in the capital market (Baker et al. 1999; Bushee and Miller 2012; Grullon et al. 2004) and suggests that firm visibility may be positively related to NYSE listing (a dummy variable indicating firms listed at NYSE), institutional ownership (the percentage of a firm’s common stock held by institutions), and analyst following (the number of analysts following a firm).

In addition to the traditional measures of firm visibility, we seek to control for online firm visibility, given that the interest of our research is the effect of an online information channel. In a very recent paper, Da et al. (2011) use Google’s Search Volume Index (SVI) of company names as a proxy for investor attention. We collect Google SVI data from Google Insights.\(^{27}\) SVI for a search term is the number of searches for that term scaled by its time-series average. To make search volumes of different companies comparable, we obtain all firm SVIs relative to the generic word “cotton,” which remains stable throughout the observation period. Figure 2 gives an example of how the relative search volume for “Dell” is obtained. The blue (lower) line is the search volume for “Dell.” The red (upper) line is the search volume for “Dell.” Google reports relative measures of the two search volume indices. Throughout the observation period, the search volume for cotton is normalized to 1. Some company names, such as Palm and Gap, may have alternative meanings that are not directly related to the respective companies. We manually go through our list and create an indicator for each company. When reporting our results, we include all such companies to avoid a subjectivity bias. The results remain qualitatively unchanged after removing such companies from our sample.

Appendix E presents the results. For ease of comparison, Column (1) shows the base model (i.e., Column (5) in Table 3). Column (2) includes all of the traditional controls for firm visibility. Column (3) further includes the control for Google SVI. Appendix E conveys two key messages. First, firm visibility generally plays a role in leading firms to disclose earlier, as evident in the significant and negative coefficients on firm size, NYSE listing, analyst following, and firm profitability. Second, importantly, we see qualitatively unchanged effects of WikiMOD, after incorporating the various controls for firm visibility.

In Appendix F, we examine the economic significance of Wikipedia modifications. We show that information aggregation on Wikipedia has an economically meaningful impact on market reaction. Appendix G shows that our results are robust to alternative samples (e.g., excluding firms with very low and very high values of WikiMOD, removing modifications that are involved in edit wars), and to alternative measures (e.g., using the number of words added to proxy for information aggregation on Wikipedia).

**Concluding Remarks**

**Major Findings**

In this paper, we explore the question of whether Wikipedia improves the information environment for financial market investors, and if so, how public-firm managers change their voluntary disclosure behavior concerning bad news when there are variations in Wikipedia coverage. Our analysis is based on a unique data set collected from financial records of public firms, management-disclosure records, news articles related to these firms, and the modification history of firm entries on Wikipedia. The following new findings add to the existing literature.

First, we find that information aggregation on Wikipedia about public firms can moderate the timing of management disclosure of earnings disappointments. The timing of management disclosure is a function of the information environment, as described by the characteristics of analyst forecasts, a finding consistent with the disclosure literature. Although managers generally tend to delay disclosure of bad news when analyst forecasts are highly dispersed, the delay is shortened by Wikipedia’s information aggregation, suggesting that Wikipedia weakens information asymmetry between managers and investors.

Second, although the literature points out that when facing high analyst bias, managers tend to release bad news early to avoid various risks, this study finds that information aggregation on Wikipedia helps alleviate managers’ concerns, thus making management’s disclosure timing less dependent on analyst bias.

Finally, by examining how investors react to management disclosure of bad news, we find that investors’ negative reaction to bad news is moderated by Wikipedia modifications. Although Wikipedia may not be the only means by which investors can enjoy the benefits of social media, our results suggest a market adjustment associated with information aggregation on Wikipedia before a firm announces bad news.

\(^{27}\)http://www.google.com/insights/search/. For more information about how the indices were constructed, please refer to Google’s help documents: http://support.google.com/insights/. Similar to Da et al. (2011), we search for both the company names and their stock tickers. The results are highly consistent.
**Theoretical Implications**

This paper makes several theoretical contributions. Most prior studies of Wikipedia focus on users’ contribution behavior (e.g., Greenstein and Zhu 2012a; Kittur and Kraut 2008, 2010; Thom-Santelli et al. 2009; Zhang and Zhu 2011). In many ways, these studies have improved our understanding of how individuals may be motivated to contribute to Wikipedia. Without establishing the value of the outcomes of such collaborations, however, research on contribution incentives is limited. This study fills such a gap and examines how Wikipedia makes an impact to the real world. To this end, we document IT’s capability of aggregating information (Benbasat and Zmud 2003) and analyze the impact of an IT artifact (Wikipedia) directly (Orlikowski and Iacono 2001; Venkatesh et al. 2007).

Our finding suggests that the informational impact of social media such as Wikipedia can be derived from social interactions and collaborative efforts. This study supports a shift to conceptualizing users as social actors rather than individualistic entities (Lamb and Kling 2003). With such a shift, our study goes beyond the view of Wikipedia contributors as disjointed individuals, with each following his or her own agenda. Rather, we promote a theoretical anchor for identifying the value of information aggregation arising from collaborative efforts.

In building the theory, this paper taps into the accounting literature of management disclosure. This not only introduces a new lens to the information systems literature but also demonstrates the complementarity of information systems research with that of other related fields. The increasing reliance of business on IT opens a door for information systems research to contribute to these other business disciplines. As this study shows, the analysis of the impact of Wikipedia sheds light on how managers and investors obtain and use information in the financial market.

**Managerial Implications**

This study also has implications for investors, managers, and policy makers in the financial market. Importantly, different from news stories, which fade away quickly, Wikipedia’s aggregating and accumulating information keeps a good record of a company’s important events. This suggests that Wikipedia’s effects on investors, managers, and regulators can be long-lasting.

Our results suggest that investors can benefit greatly from the value of information aggregation offered by services such as Wikipedia. This type of information aggregation stands in stark contrast to traditional ways for investors to get informed. In the traditional model of information acquisition, individual investors suffer a significant information disadvantage. The channel to obtain information is very limited and both newspapers and analysts can introduce significant biases. Unlike these channels, the decentralized information aggregation on Wikipedia seems to be offering a new channel that is either free of these biases or resistant to them. Interestingly, some individual investors have already started to build an information aggregation portal that utilizes the wiki technology. This new service is called “Wikinvest”28. On their “About Us”

---

page, they write, “We’re regular, everyday investors who are sick of the level of innovation at the major finance portals.”

One major difference between Wikinvest and traditional financial information websites is that Wikinvest relies heavily on user-generated content, and it extracts the value of information aggregation.

For managers, our study suggests that their information advantage is increasingly compromised with the advent of information aggregation channels like Wikipedia. In the pre-Wikipedia era, they could utilize their private information and manipulate what and when the market could learn about their firms. With Wikipedia and other types of social-media channels, it becomes increasingly difficult to maintain the same level of control over information. It may be necessary for managers to change their mindset and embrace the benefits brought about by new media. One of our findings suggests that Wikipedia can alleviate their litigation concerns when analyst bias is high. Our view is that the information asymmetry between managers and investors is not necessarily a zero-sum game. If managers can recognize the unavoidable impact of IT, and make good use of it, they will be able to better capture opportunities and resolve challenges.

Regulators such as the SEC should welcome the improved information transparency introduced by Wikipedia and other social-media channels. To restore investors’ confidence in the market after a series of corporate scandals, the Sarbanes-Oxley Act (SOX) was enacted in 2002. SOX aims to reform public company accounting and protect investors, and one major objective is to enhance corporate transparency (Coates 2007). The information transparency achieved through information aggregation on Wikipedia offers a complementary mechanism to achieve the same goals as those of regulators. This paper examines information aggregation on Wikipedia for only public companies, and thus it is important to note that many private companies also have Wikipedia pages. If Wikipedia can improve the information transparency of these privately held companies, given that SOX does not apply to them, then our results suggest that technology-enabled information aggregation by individuals plays a broader role than the legal mechanism in enhancing companies’ information environment.

**Limitations and Future Research**

This paper has several limitations that may need future work to address. First, as explained above, we focus on only Wikipedia and do not capture the full range of social-media channels (e.g., blogs, Twitter, Facebook, etc.). Although other social-media channels can also aggregate information, the mechanisms through which these other channels affect management information disclosure may be different. Blogs, for example, are typically managed in a centralized manner, such that one blog can have only one or a few contributing bloggers (Sun and Zhu 2012). Only some frequent bloggers can attract the public’s attention, and their potential influence is limited to loyal followers of these blogs. Although it is easy to find some information about almost all public companies on Wikipedia, blogging information of these companies is scattered and may be difficult to find. Twitter is different from Wikipedia in that Twitter’s social influence follows a directional path. In addition, similar to traditional news, information about specific companies on Twitter has short-lived popularity. On Facebook, network ties are based on friends, and the tie strength is much stronger than the poster–reader relation on Wikipedia. As a result, the information-exchange mechanism is different. As a result of the structural differences between Wikipedia and other social media platforms, whether other social media channels affect management disclosure in a similar way needs careful examination before our results can be generalized. Our prediction of such effects is rather optimistic. If the unmeasured information aggregations were to be correlated with what we have incorporated in our analysis, then our estimates would partially reflect this broader range of information over social media. This is likely to be true, because information can diffuse across social-media channels (Watson-Manheim and Bélanger 2007). A future study could deepen our understanding by studying other types of social-media channels.

Second, Wikipedia may play other roles in the financial market. This study merely touches the tip of the iceberg. We employ a narrow lens to examine managers’ behavior of information disclosure. There are other significant issues, such as investor sentiments, insider trading, changes in liquidity, and stock-price volatility associated with social-media coverage. These would be fruitful avenues for future research.

Third, this study adopts a positivist perspective to infer the relationship between Wikipedia and the financial market’s information environment. Our results could be complemented by case studies, surveys, or ethnographic methods. For instance, it remains an assumption that managers know that information aggregation over Wikipedia provides information to investors. Future research can conduct a survey to ask managers whether they are aware of information aggregation over Wikipedia and how they perceive the extent to which investors get informed from that. While our results present evidence of Wikipedia’s role in the financial market, these alternative methods offer important insights about why and how such roles take place.
Acknowledgments

We thank the senior editor, the associate editor, and the two anonymous reviewers for their constructive feedbacks, which enabled us to improve the paper significantly. We appreciate the suggestions that we received from Kalok Chan, Eric Clemons, Sudipto Dasgupta, Alok Gupta, Chia-Chun Hsieh, Kirill Novoselov, Mike So, Feng (Alex) Tian, Bruce Weber, Hong Xu, Mingshan Zhang, Yong Zhang, Feng Zhu, and participants at the 2009 International Conference on Information Systems. Sean Xin Xu acknowledges financial support from the Program for New Century Excellent Talents in University (NCET-10-0508) and the National Natural Science Foundation of China (#70990080 and #71001056). Michael Zhang acknowledges generous financial support from Hong Kong Research Grant Council and HKUST (RPC07/08.BM16 and DAG_S09/10.BM03). An early version of this paper was presented and nominated for the Best Paper Award at the International Conference on Information Systems in 2009. Both authors contributed equally to this research.

References


Xu & Zhang/Impact of Wikipedia on Market Information Environment


About the Authors

Sean Xin Xu is an associate professor at School of Economics and Management, Tsinghua University. He received his Ph.D. from the University of California, Irvine. His current research interests include IT value under different corporate governance structures, post-adoptive behaviors of IT use, social media, enterprise systems and information risk, and IT use in supply chain contexts. His papers have been published in Management Science, MIS Quarterly, Information Systems Research, Journal of MIS, Strategic Management Journal, European Journal of Information Systems, and International Journal of Technology Management. He serves on the editorial board of Information Systems Research as an associate editor, and that of MIS Quarterly as a guest associate editor. He was nominated for the Best Paper Award at the International Conference on Information Systems (ICIS) in 2009. He won the Vernon Zimmerman Best Paper Award at the Asian-Pacific Conference on Information Accounting Issues in 2007, two Best Paper Awards at ICIS in 2002 and 2003, and a Best Paper Award (International Track) at the 2004 Americas’ Conference on Information Systems.

Xiaquan (Michael) Zhang is an assistant professor of Information Systems, Business Statistics, and Operations Management at the Hong Kong University of Science and Technology. He holds a Ph.D. in Management from MIT Sloan School of Management, an M.Sc. in Management, a BE in Computer Science, and a BA in English from Tsinghua University. Before joining academia, he worked as a consultant for an investment bank, and as an international marketing manager for a high-tech company. He holds a U.S. patent, and started a social-network company. Michael’s research interests are on issues related to creation, dissemination, and processing of information in business and management contexts. His work studies pricing of information goods, online word-of-mouth, online advertising, incentives of creation in open source and open content projects, and use of information in financial markets. His research has appeared in American Economic Review, Management Science, Journal of Marketing, MIS Quarterly, Journal of Management Information Systems, Decision Support Systems, and Journal of Interactive Marketing. He currently serves as an associate editor for Information Systems Research and as a guest associate editor for MIS Quarterly, and is also on the editorial boards of Production and Operations Management and Electronic Commerce Research and Applications.