FROM DATA TO INFORMATION

By

JOSEPH H. STICE

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

2016

To Amber and the Babies

ACKNOWLEDGMENTS

I thank my dissertation committee members: Michael Drake, Zhifeng Gao, Marcus Kirk, David Reppenhagen, and Jenny Tucker (chair). I also thank Michael Mayberry, Derrald Stice, Roger White, Devin Williams, workshop participants at the University of Florida, Singapore Management University, The National University of Singapore, Nanyang Technical Institute, Chinese University of Hong Kong, University of Hong Kong, Southern Methodist University, and University of Arkansas for helpful comments and discussions.

TABLE OF CONTENTS

page

LIST OF TABLES

LIST OF FIGURES

Abstract of Dissertation Presented to the Graduate School
of the University of Florida in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy

FROM DATA TO INFORMATION

By

Joseph H. Stice

August 2016

Chair: Jennifer Tucker
Major: Business Administration

The increased accessibility of firm-specific data may speed price discovery and decrease information asymmetry in capital markets through decreased search and discovery costs. Too much data, however, may slow price discovery and increase information asymmetry because of increased friction arising from market participants being distracted or overwhelmed by an overabundance of data. In this I study examine how the amount of publicly available data influences price discovery and information asymmetry. I create a measure of firm data availability. I examine the relation between this measure and investor search for information on the Internet, price formation, and information asymmetry. I find that abnormally high data availability is negatively associated with the speed at which earnings information is impounded into price and positively associated with information asymmetry measures around earnings announcements and that these findings are driven by large firms. I find some evidence that the relation between data availability and price discovery is concave. My findings suggest that an overabundance of publicly available data "muddies the water" and impedes the impounding of information into stock prices.

CHAPTER 1
INTRODUCTION

The Internet has grown exponentially since the launch of the first website in August 1991.[1] By 2000, the number of websites was over 17 million and the number of users had surpassed 400 million. By the end of 2014 there were more than one billion websites and three billion Internet users.[2] This abundance of data may facilitate market participants' search-and-discovery activities by making data readily accessible. The Internet also increases investor awareness that such data exist, inducing them to more aggressively search for information. The abundance of data, however, may hamper search-and-discovery activities by distracting market participants. The negative impact of data may arise from the limited processing ability of investors (Libby et al. (2002)) or the increased difficulty of extracting useful statistics from publicly available data (Bloomfield (2002)). This study examines the effect of publicly available data on (1) the speed at which information is impounded into stock price and (2) information asymmetry around firm earnings announcements.

Prior studies have established the importance of information dissemination and information search via online sources. These studies find that information dissemination online is associated with more efficient capital markets. For example, tweets containing links to earnings information (Blankespoor et al. (2014)), discussions on internet message boards (Chen et al. (2014)), and dissemination of corporate news on the web (Drake et al. (2015)) decrease information asymmetry, substitute for more traditional information sources, and increase the

[1]This first web page consisted of a simple white background with black text describing the World Wide Web (W3 as it was referred to in the early days). See http://info.cern.ch/hypertext/WWW/TheProject.html for more information.

[2]The number of websites has slightly decreased since mid-2014 to below one billion, but is slowly climbing back up toward this threshold. Seehttp://www.internetworldstats.com/emarketing.htm; http://www.internetlivestats.com/total-number-of-websites/ for more information.

speed at which earnings news is impounded into stock price.[3] These studies also highlight the importance of retail investors in the market (the likely users of online data sources). While these studies have examined a subset of information available on the Internet, my study investigates the extent to which data availability on the Internet influences price discovery and information asymmetry in capital markets.

I make an important distinction between data and information. Data is a collection of facts and statistics and becomes useful only when it has been processed (i.e. turned into information). Information, on the other hand, is anything that will lead to a "change in expectations about [the] outcome of an event" and induce individuals to act in response to these updated expectations (Beaver (1968)). Data becomes information, and therefore useful, after it has been processed and presented in a form that can be used by market participants to change their assessments of future performance. Prior studies have examined information. My study examines how the increasing availability of data affects market participants' ability to turn data into information.[4] My study complements prior studies by examining how the availability of firm-specific data on the internet (a firm's "web presence") is used by investors to price stocks and expand their information set about a firm.

Too much data availability may introduce additional friction (barriers to trading activity) into capital markets. If investors become overwhelmed by an overabundance of data they will be unable to process these data in a timely manner. If investors are unable to process data, they will not be able to execute trades based on the information contained therein. As a result, stock prices

[3] Drake et al. (2015) find that certain types of web dissemination (blogs) decrease the speed at which earnings information is impounded into stock price.

[4] At the theoretical level, data and information are distinct constructs. However, at the operational level, this distinction is blurry. I acknowledge a significant amount of overlap between data and information in my measure.

will not fully reflect the information available in public data (because this information has not been fully extracted). This increased friction at the individual level will not necessarily result in inefficient markets; rather, the aggregate market will take longer to fully incorporate publically available data into stock prices and reach an efficient price.

I introduce a measure of firm data availability. My measure is the number of search results returned when performing an Internet search for a company ticker symbol for every week between 1/1/2007 and 12/31/2014.[5] Thus, for each company-week I have an approximation of the amount of data an investor would have access to if s/he had performed an internet search for the firm.[6] My measure can be seen as an estimate of the amount of publicly available data that exists for a firm at a given point in time. My measure allows me to examine changes in the availability of firm-specific public data around corporate events and whether the amount of data available aids in price discovery and effects information asymmetry.[7]

My measure of data availability is distinct from measures of information availability and investor search for information. My measure is distinct from the former because, as discussed above, data and information are fundamentally different constructs with information being inherently useful and data containing facts and statistics that can be turned into information.[8] My

[5] All searches are performed in Google.

[6] I recognize that not all information is equally weighted. Internet search engines filter data in an effort to present the most relevant or useful data higher up in a set of search results. To the extent that search engines are effective in presenting more useful data to searchers, I expect data availability to increase information efficiency and bias against finding results consistent with data availability slowing price discovery and increasing information asymmetry.

[7] I am indifferent to whether these changes in data availability are brought about through creation of new data or broader dissemination of existing data.

[8] There is overlap between my measure of data availability and previous measures of information availability. However, any similarities should be biased toward finding results consistent with prior literature – more information availability speeds price discovery and decreases information asymmetry.

measure is distinct from the latter because investor search for information involves examining existing information or creating new information by processing existing data. Thus, understanding the impact of data availability adds value above and beyond understanding the impact of information availability and investor search activities.

I validate this measure by examining the relation between my measure and measures of attention around corporate events. My measure varies predictably with the known release of firm-specific data and is associated with firm characteristics that should affect the prevalence of publicly available data. I find that in the week of and the weeks before and after an earnings announcement the level of data is abnormally high. This is consistent with earnings announcements being data-creating events for the market. This pattern is not true for 10K or 10Q filings. Data availability in the week of a 10K or 10Q filing is abnormally low while data availability in the week following a 10K or 10Q filing is abnormally high. Thus, it appears that an abnormally high amount of data is created/disseminated around earnings announcements and in the week following required filings, but less data is created/disseminated in the week of a required filing.[9]

In my formal test of hypotheses, I examine the relation between my measure of data availability and the speed at which earnings news is impounded into price and information asymmetry. I find that price discovery (i.e. the speed at which information is impounded into price) of earnings news is hindered as the abnormal availability of data increases. This finding is consistent with retail investors (those who would search for firm information using the Internet)

[9] One possible reason for this pattern is that earnings announcements precede 10K and 10Q filings. This could shift the baseline level of data up, resulting in lower abnormal data in the weeks of 10K and 10Q filings simply as an artifact of how "normal" levels of data are measured. To account for this possible issue, I estimate my baseline level of data using several alternate time windows – 2 weeks, 4 weeks, and 52 weeks. The results are robust to these alternate estimation windows. The tables presented use a 10-week estimation period following prior literature (Drake et al. (2012) and Da et al. (2011)).

being either overloaded by an overabundance of information or relying on stale/recycled (Tetlock (2011)) or inaccurate information obtained from searching activities. This result is only true for large firms (top tercile with respect to firm size). Small and medium-sized firms do not appear to be affected by public information availability. I also find that more data availability is associated with greater information asymmetry in the week of an earnings announcement. Larger firms, again, drive this effect. My results support the notion that a stronger "web presence" does not always lead to more efficient pricing. Furthermore, the fact that large firms drive my results provides evidence that it is the incremental availability of data and not the broadness of dissemination that distracts investors.

I further examine the possibility that the relation between data availability and price discovery and information asymmetry is concave. That is, that data availability aids price discovery and decreases information asymmetry up to a point, but then becomes detrimental to the firm's information environment. I find that firms with the weakest information environments benefit from increasing data availability, and this benefit decreases as the strength of the information environment increases.[10]

In additional tests I examine the effect of investor information search on the speed at which information is impounded into price and information asymmetry around earnings announcements. Effort expended by market participants in the form of search-and-discovery activities (information search) is expected to speed price discovery and decrease information asymmetry. However, data availability may either help or hurt. Thus, finding different effects of data availability and information search on the speed at which information is impounded into

[10] The weakest information environment firms are defined as the smallest firms (by size) with the least amount of Internet data availability.

price and information asymmetry would provide insight into the usefulness of data accessibly on the web.

I use the measure of investor attention/demand used in Da et al. (2011) and Drake et al. (2012) – Google SVI – to measure investor search activity.[11] Google SVI measures how much searching investors do for a particular keyword. By contrast, my measure represents what an investor would find upon performing a search for a particular keyword.[12] Consistent with my expectations, I find that investor search activities are associated with faster price discovery of earnings information for firms of any size. However, investor search activities around earnings announcements do not appear to affect information asymmetry in the week of the earnings announcement. These results suggest a differential effect of information availability and investor search activities in capital markets – investor search activities increase the speed at which information is impounded into stock price but do not influence information asymmetry, whereas data availability can decrease the speed at which information is impounded into stock price and increases information asymmetry.

My study differs from Drake et al. (2015) in two ways. First, my measure of data availability differs from their measure because I do not limit my results to websites that publish corporate information. Rather, I measure what investors would find if they used the Internet to search for financial information about a firm. Drake et al.'s measure captures mostly information, which is, by definition, useful while my measure captures all data available to an

[11] Google SVI represents the relative frequency (on a 100-point scale) at which a term was searched for on the internet at a given point in time benchmarked against the week or day on which that search term experienced the most searching activity.

[12] Not every investor needs to see every piece of data available on the Internet. A greater number of search results indicates that the available set of data is larger (i.e. more data exists for firms with a greater number of returned search results).

investor at a point in time. As an example, the mean (median) number of websites covering earnings events in Drake et al. (2015) is 8.5 (4). In my sample the mean (median) number of search results (individual web pages) returned from a search of company tickers in the week of an earnings announcement is 1,254,081 (274,000). Thus, the measure used in Drake et al. (2015) can be considered a subset of the information contained in my data. Second, in addition to examining how data availability influences the speed at which information is impounded into stock price, I examine its effect on information asymmetry and what firm events and characteristics influence information creation (at least on the Internet).

My paper contributes to the existing literature in four ways. First, I create a useful and innovative measure of firm data availability. Previous studies have examined aspects of a firm's information environment. My measure captures the total amount of public data available to investors. Second, I contribute to the literature on price formation by showing that, at least on the Internet, greater information availability can hinder price formation and result in greater information asymmetry. Third, I find evidence that retail investors, the likely users of information from Internet searches, appear to influence overall price discovery and information asymmetry. Finally, I identify an unlikely source of market friction – data availability brought about by advances in technology – that appears to slow price discovery and increase information asymmetry.

The rest of the paper is organized as follows. Section 2 reviews the extant literature and develops my hypotheses. Section 3 discusses the data. Section 4 develops my models. Section 5 reports findings. Section 6 reports additional analysis. Section 7 concludes.

CHAPTER 2
BACKGROUND AND HYPOTHESES

Efficient Market Hypothesis vs Information Overload

The Internet has become an increasingly vast repository of data. For investors, this means that now, more than ever, data about potential investments is just a few mouse-clicks away. The availability of this data may be a blessing or a curse. On the one hand, more publicly available data could decrease information asymmetry among market participants by providing less sophisticated investors with useful information about a firm. On the other hand, this influx of publicly available data could overwhelm or distract less sophisticated investors and therefore increase information asymmetry.

Strong-form and semi-strong-form efficient market theories predict that the market instantaneously and completely impounds all public information into stock price (Fama (1970)). In practice, however, this process takes time and the speed at which information is impounded into price depends on the amount of friction (anything that prevents a trade from being executed) in the market (Merton (1987)). In my setting an overabundance of data may act as a market friction if market participants are rational only with respect to their limited information set and have limited cognitive processing ability and time (i.e. bounded rationality Gigerenzer and Selten (2002)). Multiple studies have documented market frictions ranging from localized power outages (Shive (2012), to extreme weather (Loughran and Schultz (2004), to the enforcement of distracted driving laws (Brown et al. (2015)). In each case, exogenous shocks increase friction, which limits information transfer and results in fewer trades.

Bushee et al. (2010) find that earnings information disseminated in the business press (facilitating information accessibility and decreasing friction) results in less information asymmetry around announced earnings (i.e. improved price discovery consistent with Grossman

and Stiglitz (1980) and Hong and Stein (1999)). The studies mentioned above highlight the fact that market participants are 1) worse off when information exchange between market participants is hampered and 2) better off when the dissemination of value-relevant information is broader – both consistent with the conjecture that market participants benefit from increased levels of information. Consistent with D'Avolio, Gildor, and Shleifer (2001)), who argue that new technology "democratizes the market," The Internet, one of the greatest technological innovations of the 20th Century – with its vast array of websites and data – should facilitate information diffusion and result in less information asymmetry and more speedy price discovery.

Data is useful only if 1) it is relevant and 2) it is processed and turned into information by market participants. As the amount of data increases, investors must choose whether to use new data (i.e. expend energy to convert data into information) to update their beliefs about firm performance. The tradeoff arises because market participants have limited data processing capacity (Kahneman (1973)) and an overabundance of data may result in an "overload" in which investors cannot use all relevant and available data.

Market participants may ignore relevant information because of limited information processing ability (Hirshleifer and Teoh (2003)). Additionally, the Incomplete Revelation Hypothesis (Bloomfield (2002)) posits that investors will not respond to some relevant data because the extraction costs associated with obtaining useful statistics from those data (i.e. turn the data into information) are too high – an overabundance of data has the potential to increase extraction costs for the retail investor. Similarly, Rubenstein (2001) argues that markets are at least minimally rational – rational market participants cannot earn abnormal profits, in part, because of transaction costs – and in minimally rational markets investors are more likely to believe they can "beat the market."

Dellavigna and Pollet (2009) show that earnings news announced on a Friday, when investors are less attentive, results in a smaller immediate market response and larger delayed response than earnings news announced on other days of the week. Furthermore, Libby et al. (2002) review evidence that market participants (both sophisticated and unsophisticated) "[overrely] on unreliable information." One particular example, Bloomfield et al. (2003), is that experimental participants over-rely on past performance when predicting future performance. Similarly, Tetlock (2011) finds evidence that retail investors overreact to stale news (news already factored into price). Thus, news recycled on the Internet may result in what Drake et al. (2015) describe as an "echo chamber" and hinder price formation. These studies suggest that an excess of information (either useful or superfluous) may lead to difficulties in information processing.

Information Search

Internet search activity has been used in a number of studies to examine demand and attention allocation (Choi and Varian (2009) and Mondria et al. (2010)). Importantly, Barber et al. (2009) find evidence that trades of retail investors can materially affect stock prices. Thus, understanding where retail investors focus their attention is valuable. Initial work (Da et al. 2011) in finance uses the frequency of Google searches (SVI) as a proxy for retail investor attention and showed that abnormal SVI predicts higher prices (and subsequent price reversals).[1] Drake et al. (2012) use SVI as a measure for the demand of firm-specific information and find that abnormal SVI spikes at important corporate events (i.e. earnings announcements, management forecasts, analyst forecast, etc.) and persists for several weeks thereafter. In

[1] Institutions and professional investors are likely not relying on results from an Internet search to inform them about a potential investment.

addition, they find evidence of abnormally high SVI before earnings announcements but not before other important accounting events. These findings suggest that market participants use the Internet to search for information relevant to a stock. As a complement, my measure allows me to examine how much data are available via these search activities and test if investors do indeed succumb to "information overload" resulting from limited attention (Kahneman (1973)).

Information Availability

There is a difference between data availability and information availability. Data is a collection of facts and statistics that have not yet been processed. When data is processed it becomes information and can be used to update expectations of future performance. Thus, greater data availability is only beneficial if it can be turned into information. By contrast, information availability is always beneficial because information is, by definition, useful and can be used to update expectations. Prior literature, discussed below, has focused on information availability while my study is concerned with data availability.

Several studies have examined the effects of firm-specific information dissemination by the press. Miller (2006) finds that by reprinting information from other sources and by performing independent investigation and analysis the press serves as a "watchdog" for accounting misconduct. Bushee et al. (2010) investigate the relation between a firm's information asymmetry and business press coverage and find that, in general, information asymmetry decreases in coverage. Blankespoor et al. (2014) have similar findings with respect to firm-initiated releases of financial information on Twitter™ – tweets with links to financial information result in less information asymmetry for the tweeting firms. Similarly, Solomon (2012) finds that greater news coverage of good-news corporate events results in short-term price increases and smaller returns around subsequent earnings announcements. These studies illustrate the role media plays in disseminating firm-specific information.

The media may also focus investor attention on certain firms (Barber and Odean (2008) and Frederickson and Zolotoy (2015)). Li et al. (2011) find that newswire alerts, which serve to focus investor attention on particular firms, are more prevalent for firms that face certain risk factors or fail to release preliminary earnings. Upon examination of intraday trading patterns, Li et al. (2011) find that the market responds almost immediately to these alerts. Drake et al. (2014) find that the press facilitates the broad dissemination of earnings information mitigating cash flow mispricing and improving the market's ability to efficiently price securities. Twedt (2015) reports that management earnings guidance disseminated via newswires is more quickly impounded into price than other management guidance that is not disseminated by newswires. These studies highlight the effects of news outlets on price discovery.

Chen et al. (2014) examine opinions posted on an investor website.[2] They find that these opinions are associated with future returns and future earnings surprises, indicating that "internet chatter" may hold information relevant for updating expectations of future firm performance. Furthermore, they find that these relations are strongest for firms with less analyst coverage, suggesting that web coverage is a substitute for more traditional avenues of firm-specific information. Drake et al. (2015) take the initial steps to discovering the role the internet plays as an information intermediary in capital markets. They examine earnings announcement web coverage across a host of websites and show that this coverage increases the speed at which announcement period information is impounded into stock price. However, when they split web coverage into different categories they find that Blog and other non-financial website coverage appear to slow price discovery. In addition, they find that these effects depend on the visibility of the covered firm - only firms with high visibility benefit from more web coverage.

[2] http://seekingalpha.com/

Finally, Fang and Peress (2009) find that, in the cross section, stocks with no media coverage experience higher returns than do stocks with high media coverage. They conclude that this result arises because firms with media coverage are more recognizable and investors demand less of a return premium for these firms than for firms without media coverage. Fang and Peress' sample ends in 2003 to avoid the confounding effects of the Internet becoming a large portion of the mainstream media.

The studies mentioned in this section suggest that increased media coverage of earnings information aids price discovery leading up to earning announcements. Furthermore, these studies suggest that the media is able to focus investor attention on salient news events and enable investors to update their expectations about future firm performance. To the extent that data available on the Internet can be turned into information, I expect data availability to have similar positive effects on both the speed at which information is impounded into price and on information asymmetry. However, to the extent that data availability overwhelms or distracts investors and results in increased market frictions, I expect data availability to decrease the speed at which information is impounded into price and increase information asymmetry.

My first two hypotheses address the effect of information availability on the speed of price discovery and information asymmetry. Given the conflicting theoretical and empirical evidence on this subject (discussed above), I state my first two hypotheses in the null form:

- H1: Higher levels of public data availability have no effect on the speed at which earnings information is impounded into stock price.
- H2: Higher levels of public data availability have no effect on information asymmetry around earnings announcements.

The usefulness of publicly available data also depends on the information environment of the firm. For example, firms with weak information environments will likely benefit from more publicly available data because investors will fare better with some data than with no data. The

same may not be true for firms with strong information environments. For these, more data may provide less incremental benefit to investors and simply confuse or overload them. I expect different effects of data availability on information poor and on information rich firms. I state my third hypothesis in the alternative form:

- H3: The availability of data has a differential effect on the price discovery and information asymmetry of firms with strong versus weak information environments.

CHAPTER 3
DATA

I perform historical Google searches for each firm currently traded on the New York Stock Exchange, American Stock Exchange, or National Association of Securities Dealers Automated Quotation System. I require that each firm also have coverage in COMPUSTAT and CRSP. This yields an initial sample of 5,283 firms. Similar to Drake et al. (2012) and Da et al. (2011) I perform my searches using company ticker symbols and assume, as stated in Drake et al. (2012), "that a Google user who enters a ticker symbol into a search engine is looking for some piece of financial information about the firm." This assumption is reasonable. For example, a search for "MSFT" returns a graph of daily stock performance, articles about earnings, and analysts' recommendations, whereas a search for "Microsoft" returns a link to the company's homepage, Wikipedia, and other relevant but non-financial news. In essence, search results obtained from a search of a company's ticker can be seen as the estimate of the amount of publicly available financial data existing for the firm.

I perform a search every week between 1/1/2007 and 12/31/2014 by the company ticker symbol.[1,2] This results in a total of 2,239,992 individual company searches across the 8-year sample period. For each search I collect the number of web results returned. To be included in my sample, a firm must search results for the entire search period. A firm must also have an average of at least 1,000 search results or more across the search period. It is important to note

[1] I perform a search on the first day of every year (1/1/20XX) and every seven days thereafter – 1/8/20XX, 1/15/20XX, 1/22/20XX, etc. I exclude searches for Dec 30 and 31 and instead use the search for Jan 1 of the following year. This results in 52 searches per year per firm.

[2] Searching for firm data using the company ticker symbol is consistent with both Da et al. (2011) and Drake et al. (2012).

that search result counts are not a measure of search frequency (Google SVI).[3] Rather, this number provides a rough approximation of the total number of results that would have been returned if a person had searched by a company ticker in the given week. I use the total number of results returned as a measure of the total amount of financial-specific data publicly available about a firm.

To account for changes in the total amount of data available on the Internet over time (size of the Internet), I create a proxy for the size of the Internet. In theory, this could be done by leaving the search field blank (i.e. searching for everything) and counting the number of results returned for this general search. In practice, Google does not allow a user to perform this search. I therefore perform the most general search possible, a search for the integer "1", and use the number of results returned as a measure of the size of the internet in the search period.[4,5] I scale each weekly firm search result by the size of the internet for the same period. This yields a measure that is equal to the proportion of company-specific search results relative to the total number of search results available in Google.

Figure D-1 and Figure D-2 present a time-series of scaled data availability for two firms – Apple and Microsoft. While data availability is changes little from day to day, there is significant variation over time. Additionally, note the trend in data availability for both Microsoft and Apple. Over my sample period the amount of data (relative to the size of the Internet)

[3] Google SVI is a measure of "…users' propensity to search for a certain topic in Google." My measure is an estimation of how much the user would find upon performing a search of a certain topic.

[4] A Google search for any integer, as well as several individual English letters, results in a similarly "general" search.

[5] It becomes apparent when performing these searches that Google places a limit on the total number of search results returned for any single query. For example, on 07/17/2015 the cap was 25,270,000,000 and was returned for all integers and several individual English letters. This cap changes significantly over time, giving confidence that the number of results returned for these broad searches is a reasonable measure of total publicly available data (at least on Google).

available for Apple increases while the amount of data (relative to the size of the Internet) available for Microsoft decreases. This trend is consistent with the growth and activities of these two firms during my sample period.

I obtain firm financial data from COMPUSTAT and daily return data from CRSP. In addition, I collect data on institutional holdings and analyst coverage/forecasts from Thomson 13F and IBES, respectively. I construct a dataset using the 605,338 firm-weeks in my sample between Jan 1 2007 and Dec 31 to validate my measure. I then select the week around firm earnings announcements to test my hypotheses, resulting in a final sample 42,616 firm-week observations.[6]

[6] My variable of interest (Abnormal Information Availability) is measured once a week. I match each earnings announcement to the closest date on which information availability was measured.

CHAPTER 4
EMPIRICAL MODELS

Validation

I validate my measure by examining abnormal changes in data around corporate events and the relation between my measure of data availability and Google SVI. I expect data availability to vary around different types of corporate events. Following the method laid out in Drake et al. (2012), I regress weekly abnormal data availability on a set of corporate events and control variables. I find that my measure of data availability is associated with corporate events and varies with measures of attention and firm characteristics. I discuss the results of these validation tests in detail in Section 5 and outline the formal tests in Appendix B.

Hypothesis Testing

I examine how the availability of data influences the speed at which earnings information is impounded into stock price. Following Drake et al. (2015) I use the intraperiod timeliness measure (IPT) as the dependent variable in these tests. IPT is calculated using the following equation:

$$IPT\ [N] = \tfrac{1}{2} \sum^{N}_{t=0} (AbRet_{t-1} + AbRet_{t})/AbRet_{N} \quad (4\text{-}1)$$

IPT measures the portion of abnormal returns for a period realized up to and including each day in that period. I calculate the cumulative buy-and-hold abnormal return for the 10 days following the announcement of firm earnings (including the announcement date) and scale it by the buy-and-hold return for the entire period. This gives me the proportion of abnormal returns earned up to and including each day in my measurement window.

Higher values of IPT indicate that information is impounded more quickly into a firm's stock price Following Drake et al. (2015), I calculate IPT for the window, [0, +10] and use the IPT decile rank as the dependent variable in the following ordered logistic regression:

$$\begin{aligned} \text{IPT [N]} = {} & \beta_0 + \beta_1 \textit{Abnormal Data Availability} + \beta_2 \textit{Abnormal SVI} + \beta_3 \textit{Ln (1 + Analyst Reports)} + \beta_4 \textit{Absolute Earnings Surprise} + \beta_5 \textit{Negative Surprise} + \beta_6 \textit{Book to Market} + \beta_7 \textit{Institutional Ownership} + \beta_8 \textit{Size} + \beta_9 \textit{Quarterly Turnover} + \beta_{10} \textit{Return Volatility} + \beta_{11} \textit{Quarterly CAR} + \varepsilon \end{aligned} \quad (4\text{-}2)$$

In my second test I examine the relation between the abnormal availability of data and information asymmetry in the week of an earnings announcement. For this test I regress weekly abnormal data availability on the weekly abnormal bid-ask-spread for each firm, following Blankespoor et al. (2014). I estimate the following regression:

$$\begin{aligned} \textit{AbnormalSpread}_{it} = {} & \beta_0 + \beta_1 \textit{Abnormal Data Availability} + \beta_2 \textit{Abnormal SVI} + \beta_3 \textit{Abnormal Turnover} + \beta_4 \textit{Institutional Ownership} + \beta_5 |\textit{Weekly CAR}| + \beta_6 \textit{Ln (1 + Analyst Estimates)} + \beta_7 \textit{Size} + \beta_8 \textit{Book to Market} + \beta_9 \textit{Ln (Price)} + \beta_{10} |\textit{Earnings Surprise}| + \beta_{11} \textit{Negative Surprise} + \beta_{12} \textit{Quarterly Turnover} + \beta_{13} \textit{Ln (Quarterly Volatility)} + \beta_{14} \textit{Quarterly CAR} + \varepsilon \end{aligned} \quad (4\text{-}3)$$

Larger firms tend to have more institutional owners and a greater amount of analyst coverage, which results in these firms having stronger information environments. Therefore, in my estimation of Models (1) and (2) I split my sample into terciles to represent small, medium,

and large firms. I do this to test the differential effects of data availability across firms with differing types of information environments (i.e. weak vs strong).

Theoretically, the relation between data availability and price discovery and information asymmetry could be concave. Investors will benefit when more data become available for firms with weak information environments, but investors may become overloaded when more data become available for firms with strong information environments. I test for this relation by double sorting firms – first on size (keeping the smallest tercile) and then by the level of abnormal data availability (again by tercile). The smallest firms with the least data availability represent firms with the weakest information environments. I estimate Models (1) and (2) for this subsample.

My variable of interest is *Abnormal Data Availability*, which is defined as the difference between data availability in week t minus the median data availability for the 10 weeks preceding week t. I include a measure of investor demand/search for information for firm *i* in week *t, Abnormal SVI.* I expect investor search for information to speed price discovery and decrease information asymmetry. I also include several controls for the strength of the firm information environment (*Size, Institutional Ownership, and Analyst Coverage*). I further include measures of liquidity (*Abnormal Turnover* and *Quarterly Turnover*) control for abnormal returns and volatility leading up to the earnings announcement (*Quarterly Turnover* and *Quarterly Volatility).* Finally, I control for the news contained in the earnings announcement (*Earnings Surprise* and *Negative Surprise*). In all models I cluster standard errors by firm and include year and industry fixed effects.

CHAPTER 5
RESULTS

Descriptive Statistics

Table D-1, Table D-2, Table D-3, and Table D-4 present descriptive statistics used in my tests. As discussed above, I have two distinct samples. Descriptive statistics for my first sample, used to validate my measure, are reported in Table D-1 and Table D-3. Notably, *Abnormal Data Availability* and *Abnormal SVI* are 0.000 in any given week. In Table D-2 and Table D-4 are descriptive statistics for my sample around earnings announcements. Not surprisingly, I see that *Abnormal Data Availability* and *Abnormal SVI* are greater than zero for these events. The results for my first sample highlight the persistence of firm-specific data availability and investor search activity on the Internet and the results from my second sample provide preliminary evidence that, around earnings announcements, there is an increase in investor search activities and in data availability. Furthermore, in both Table D-3 and Table D-4, *Abnormal Data Availability* and *Abnormal SVI* are positively and significantly correlated, indicating that data availability and information search are at least somewhat related.

Data Availability around Corporate Events and Measures of Attention

In my first set of analyses, I validate my measure of firm data availability. Table D-5 presents results of my examination of data availability at corporate events and the relation between data availability and measures of attention and firm characteristics. Consistent with my univariate results, I find that more data is available on the Internet around corporate earnings announcements. This is consistent with earnings announcements supplying more new data to market participants or promoting the dissemination of existing information than occurs in an average week. Interestingly, I find that in the week of 10K and 10Q filing dates there is less information creation or dissemination than would be expected. One potential reason for this is

that the announcement of earnings happens in the weeks leading up to 10K and 10Q filings. This could skew "expected" data availability upward and result in abnormally less data availability. [1,2]

Similar to the univariate results, I find a positive association between weekly *Abnormal Data Availability* and *Abnormal SVI*. Furthermore, firm size appears to positively influence abnormal data availability. Interestingly, both analyst coverage and institutional ownership are associated with less abnormal data availability. This result could be evidence that firms with weaker information environments (i.e. less analyst coverage and smaller institutional holdings) experience a greater demand for publicly available data, and this demand is met by relatively more information being available on the Internet.

Abs (Weekly CAR) is negatively associated with abnormal data availability. This result is consistent with less recognizable stocks (i.e. firms with lower data availability) offering higher returns to their shareholders because these stocks are less liquid and therefore more difficult to trade (Merton 1987). Abnormal data availability is also positively associated with weekly abnormal returns. This suggests that at least some of the information available from Google searches is associated with value-relevant firm events. Finally, turnover is positively associated with abnormal data availability.

Table D-6 examines the pattern of data creation *around* corporate events. For earnings announcements there is an increase in data in the week leading up to the earnings announcement, a spike in the earnings announcement week, and increased data in the week following the

[1] *Expected data availability* is defined the as median data availability in the 10 weeks leading up to the target week.

[2] I perform a falsification test in which I randomly assign earnings announcement, 10-K filing, and 10-Q filing dates to the firms in my sample. Running this same test on the random event dates results in no significant loading for any of these randomly assigned corporate events. This test is described in detail in Appendix C.

announcement. For 10K and 10Q Filings there is a negative dip in data availability in the week of the filing and a seeming reversal of this dip in the week following the filing.

Data Availability and Price Discovery

The results in the previous section provide evidence that my measure of data availability is associated with corporate events and firm characteristics. I next turn to an examination of the influence of data availability and information search on price discovery of earnings information. Table D-7 presents results of this examination. I find that abnormal data availability is negatively associated with the speed at which earnings information is impounded into price. This leads to the conclusion that greater abnormal data availability on the web seems to hamper price discovery. This is consistent with an overabundance of data acting to overload or confuse investors. This effect only exists for large firms.[3] This is consistent with smaller firms (i.e. firms with weaker information environments) benefiting from increased web coverage and larger firms (i.e. firms with stronger information environments) being disadvantaged by increased coverage on the web.

I also find that *Abnormal SVI* is positively associated with the speed at which earnings information is impounded into stock price. This is consistent with increased investor search and discovery activities being associated with faster price discovery of value-relevant information. Furthermore, the result is true for firms of all sizes, suggesting that all firms benefit from increased search-and-discovery activities. One important caveat to consider with these investor search-and-discovery results is that investor search activity, as measured by Google SVI, might not lead to speedier price discovery. Rather, Google SVI could proxy for general interest in a

[3] While the coefficient on *Abnormal Data Availability* is significant for large firms and insignificant for small firms, these coefficients are not significantly different from each other.

firm. Thus, it is not information gathered from the Internet, per se, that leads to speedier price discovery but other search and discovery activities correlated with Google searches.

I further test for a concave relation between data availability and price discovery. I report results of this test in Table D-9. In the first column of Table D-9, the coefficient on *Abnormal Data Availability* is positive and significant, indicating that for the smallest firms with the lowest level of Internet data availability more data availability speeds price discovery. In the third column I see that for the smallest firms with the highest level of Internet data availability more data is negatively associated (although insignificantly) with the speed of price discovery. The difference between the coefficients on *Abnormal Data Availability* in the first and third columns is statistically significant ($p < 0.01$). This pattern of results suggests that, for these small firms, data availability initially speeds price discovery for firms with the least available data, but the benefit goes away and more data appears to slow price discovery for small firms with the most available data.

Data Availability and Information Asymmetry

I next examine the relation between abnormal data availability and information asymmetry around earnings announcements. Table D-8 presents results of these tests. Overall, abnormal data availability is positively associated with information asymmetry as measured by abnormal spread. This suggests, again, that as the amount of data available on the Internet increases, the ability of market participants (specifically retail investors) to process this data is hampered to the point that more data leads to more information asymmetry. Again, as in Table D-7, this result is only true for larger firms, which suggests that firms with weaker information

environments are not worse off when more data is publicly available on the Internet.[4] There is no significant relation between investor search activities and information asymmetry around earnings announcements.

I also examine the shape of the relation between data availability and information asymmetry and report the results of this test in Table D-10. The concave shape seen in the relation between data availability and the speed of price discovery is not present in the relation between data availability and information asymmetry.

[4] The coefficients on *Abnormal Data Availability* are significantly different between large and small firms ($p < 0.01$).

CHAPTER 6
ADDITIONAL ANALYSIS

Firms with different quality information environments may be differentially affected by data availability. To further explore this possibility, I examine cross-sectional differences in the relation between data availability and price discovery as well as data availability and information asymmetry for firms with different levels of institutional ownership. I report the results of my first test, the relation between data availability and price discovery, in Table D-11. Similar to the results for firm size reported in Table D-7, firms with the highest level of institutional ownership experience slower price discovery of earnings information in the presence of greater data availability, while firms with lower levels of institutional ownership do not exhibit any relation between data availability and price discovery. This finding is, again, consistent with my results being driven by data availability and not by the broader dissemination of firm-specific data. I report the result of my second test in Table D-12. There is no evident variation across institutional ownership levels in the relation between data availability and information asymmetry in the week of an earnings announcement. Information asymmetry in the week of an earnings announcement is positivity associated with data availability for firms with all levels of institutional ownership.

CHAPTER 7
CONCLUSION

This paper offers a new measure of firm data availability and validates this measure by examining how the availability of data week to week is associated with important corporate events. I examine whether data availability is associated with price discovery and information asymmetry around corporate earnings announcements.

I find that abnormal data availability is negatively associated with the speed at which earnings information is impounded into stock price and positively associated with information asymmetry around earnings announcements. These results are driven by large firms. I find that investor search activities (measured using Google SVI) are positively associated with the speed at which earnings information is impounded into stock price but not associated with information asymmetry around corporate earnings events.

My results suggest that increases in data availability lead to slower price discovery and more information asymmetry. This finding is consistent with the literature that suggests that market participants have limited data-processing capacity and can become overloaded by an overabundance of data. Excess data may make it more difficult for market participants to glean information from publicly available data. My results are also consistent with market participants' relying on stale information (Tetlock 2011) and more data creating an "echo chamber" of useless information (Drake et al. (2015)).

APPENDIX A
VARIABLES DEFINITONS

Table A-1. Variable Definitions

Variable	Description
Abnormal Spread	estimated as average spread in week t minus the average spread in the previous 10 weeks divided by the prior 10-week average
Abnormal Data Availability	weekly information availability less the average weekly information availability for the past 10 weeks scaled by the prior 10-week average
Earnings Announcement	indicator variable set equal to 1 if an earnings announcement happened within three days of a ticker search date for firm i.
10K Filing Date	indicator variable set equal to 1 if a 10K filing happened within three days of a ticker search date for firm i.
10Q Filing Date	indicator variable set equal to 1 if a 10Q filing happened within three days of a ticker search date for firm i.
Weekly Return	the average weekly stock return for firm I in the week around a ticker search date.
Mean Turnover	average weekly turnover for firm I in week t. Calculated as total dollar value of shares traded divided by the market value of shares outstanding.
Abnormal Turnover	estimated as average turnover in week t minus the average turnover in the previous 10 weeks divided by the prior 10-week average
Weekly CAR	the raw return minus the market return for firm i in week t
# Announcements	the number of earnings announcements that occurred in week t.
Abnormal SVI	weekly relative search volume less the average weekly search volume for the past 10 weeks scaled by the prior 10-week average
Analyst Following	the average number of analysts included in each consesus forecast for the period temporally closest to week t for firm i.
Book-to-Market	book value of equity (COMPUSTAT ceqq) as of most recent fiscal quarter divided by the market value of equity (calculated each week)
Institutional Ownership	the percentage of shares outstanding owned by institutional investors as of the most recent fiscal quarter;
Fourth Quarter	indicator variable set equal to one if week t is in the fourth fiscal quarter.
Mean Spread	average difference between closing ask and closing bid for firm i on each day in week t.
Size	market value of shares outstanding calculated for each firm week
Earnings Surprise	the difference quarterly earnings and median consensus forecast in IBES

Table A-1. Continued

Variable	Description
Negative Surprise	an indicator variable equal to 1 if earnings surprise < 0 and 0 otherwise.
Quarterly Turnover	the average weekly turnover for each firm in the 90 days leading up to an earnings announcement
IPT [0,10]	the intraperiod timeliness measure calculated over the time window beginning on the earnings announcement date and continuing for 10 trading days. Calculated using the following formula: IPT [N] = ½ ∑N t=0 (AbRett-1 + AbRett)/AbRetN
Quarterly Volatility	the standard deviation of stock returns for each firm in the 90 days leading up to an earnings announcement
Quarterly CAR	the cumulative buy-and-hold return for each firm in the 90 days leading up to an earnings announcement

APPENDIX B
MEASURE VALIDATION

I validate my measure by examining data availability around three corporate events – earnings announcements, 10K filings, and 10Q filings. My first validation test examines the creation of information at corporate events. I follow the method laid out in Drake et al. (2012). I estimate the following regression:

$$\begin{aligned} \textit{Abnormal Data Availability}_{it} = {} & \beta_0 + \beta_1 \textit{ Earnings Announcement} + \beta_2 \textit{ 10K Filing Date} \\ & + \beta_3 \textit{ 10Q Filing Date} + \beta_4 |\textit{Weekly Return}| + \beta_5 \textit{ Rank SVI} + \beta_6 \\ & \textit{Turnover} + \beta_7 |\textit{Weekly CAR}| + \beta_8 \textit{Bid} - \textit{Ask Spread} + \beta_9 \textit{Rank of \# Announcements} + \beta_{10} \textit{ Rank of Size} + \beta_{11} \textit{ Ln (1 + Analyst Following)} + \beta_{12} \textit{ Rank of Book-to-Market} + \beta_{13} \textit{ Institutional Ownership} + \beta_{14} \textit{Fourth Quarter} + \beta_{15} \textit{Abnormal Turnover} + \varepsilon \end{aligned} \quad \text{(B-1)}$$

My dependent variable in this regression, $\textit{Abnormal Data Availability}_{it}$, is calculated as the difference between scaled search results for firm *i* in week *t* and the median weekly search results for firm *i* in the previous 10 weeks. *Earnings Announcement* is 1 if an earnings announcement occurred during week *t* and 0 otherwise. I construct similar indicators for *10K Filings 10Q Filings*. I include the absolute weekly buy-and-hold stock return, $|\textit{Weekly Return}|_{it}$, for firm *i* in week *t* because both theory and prior empirical evidence suggest that investors demand smaller returns of stocks for which more information is available (Merton (1987) and Feng and Peress (2009)).

I include several measures of attention because the amount of publicly available data should be related to the amount of investor attention that a particular firm experiences. I include the decile ranking of SVI, *Rank SVI,* because it measures investor attention and investor demand for information. I also include measures of liquidity (*Turnover and Bid-Ask-Spread*) and a measure of distraction, *Rank of # Announcements.* Finally, I include a set of firm specific

controls – *Rank of Size, Rank of Book-to-Market, Ln (1 + Analyst Following), Institutional Ownership, and Fourth Quarter* – to control for changes in firm specific characteristics.[1]

In a separate regression I examine the timing of the change in data availability *around* these corporate events. For each event I create two additional indicator variables (*Event (Week Prior) and Event (Week Post)). Event (Week Prior) is* equal to 1 for each week before a corporate event (Earnings Announcement, 10K Filing, or 10Q Filing) and 0 otherwise. *Event (Week After)* is equal to 1 for each week after a corporate event (Earnings Announcement, 10K Filing, or 10Q Filing) and 0 otherwise. I estimate the following regression:

$$\begin{aligned} \textit{Abnormal Data Availability}_{it} = {} & \beta_0 + \beta_1 \textit{Earnings Announcement} + \beta_2 \textit{10k Filing Date} + \\ & \beta_3 \textit{10Q Filing Date} + \beta_4 \textit{Earnings Announcement (Week Prior)} + \\ & \beta_5 \textit{10k Filing Date(Week Prior)} + \beta_6 \textit{10Q Filing Date(Week Prior)} + \beta_7 \textit{Earnings Announcement(Week After)} + \beta_8 \textit{10k Filing Date(Week After)} + \beta_9 \textit{10Q Filing Date(Week After)} + \beta \textit{Controls} \\ & + \varepsilon \end{aligned} \quad \text{(B-2)}$$

I present the results of these validation tests in Table D-5 and Table D-6 and discuss the results in detail in Section 5. The results suggest that my measure captures changes in the public data availability for publicly traded firms.

[1] See Appendix A for variable definitions.

APPENDIX C
RANDOM EVENT DATES

In Appendix B I validate my measure and show that my measure of data availability is associated with important corporate events. However, this association could arise randomly or as the result of some artifact of the data. To mitigate this concern, I construct a set of random corporate event dates during my sample period and run the first stage of the validation test described in Appendix B above. Table B-1 below reports results of this falsification test. There is no statistically significant relation between my measure of data availability and these randomly generated corporate event dates. This gives added confidence that my measure of data availability captures real fluctuations in the data environment of the firm.

Table B -1. Falsification Test

	DV = Abnormal Data Availability
Variable	Model 1
Earnings Announcement (Random)	0.001
	(1.460)
10-K Filing (Random)	0.002
	(1.030)
10-Q Filing (Random)	0.000
	(0.420)
Year-Week Fixed Effects	Yes
N	605338
R-square	0.160

Table B-1 reports results of Model (1). The notable difference here is that a random date has been assigned for each firm as an earnings announcement, 10K filing, or 10Q filing. Variable definitions are available in Appendix A. All variables are winsorized at 1 and 99 percent. Standard errors are clustered by firm. T-values are shown in parentheses below the coefficient estimates. This sample consists of all firm weeks between Jan 2007 and Dec 2014. *, **, *** indicate statistical significance at the p < 0.10, 0.05, and 0.01, respectively. All tests are two-tailed.

APPENDIX D
TABLES

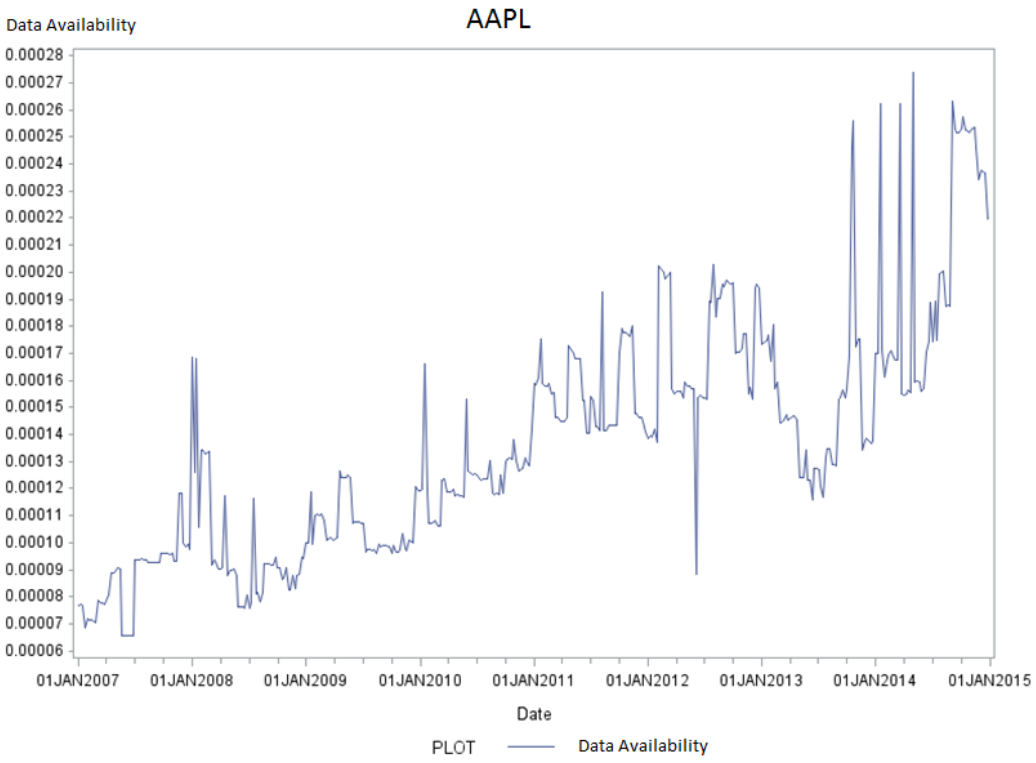

Figure D-1. Apple Data Trend

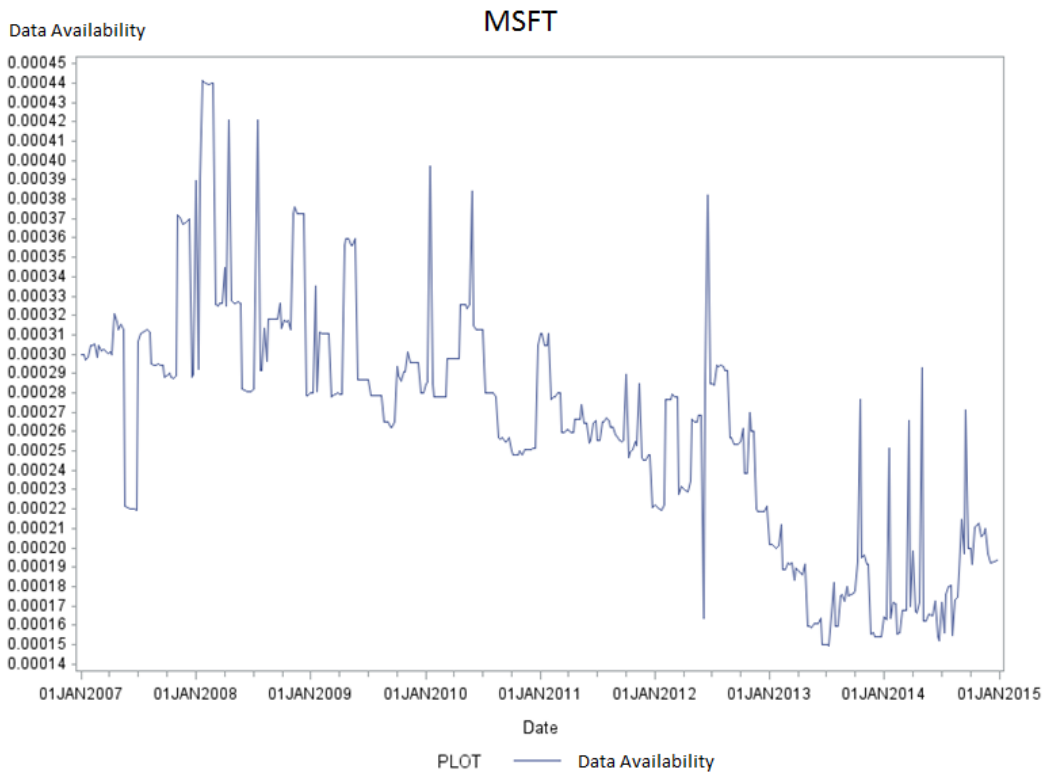

Figure D-2. Microsoft Data Trend

Table D-1. Descriptive Statistics Weekly Sample

Variable	N	Mean	Standard Deviation	25th Percentile	Median	75th Percentile
Mean Spread	605,338	0.002	0.005	0.000	0.001	0.002
Mean Turnover	605,338	0.010	0.010	0.004	0.007	0.013
Abs (Weekly Ret)	605,338	0.043	0.046	0.013	0.029	0.057
Size	605,338	21.130	1.790	19.891	21.127	22.303
Book to Market	605,338	1.044	2.070	0.356	0.597	0.949
Ln (Analyst Estimates)	605,338	1.726	0.932	1.099	1.825	2.468
Institutional Ownership	605,338	0.643	0.289	0.453	0.711	0.863
Abnormal Turnover	605,338	0.000	0.005	-0.002	0.000	0.001
Abs (Weekly CAR)	605,338	0.036	0.039	0.010	0.024	0.047
Abnormal SVI	605,338	0.014	0.171	-0.068	0.000	0.071
Abnormal Data Availability	605,338	0.016	0.166	-0.063	0.000	0.066

Table D-1 reports descriptive statistics for my sample of weekly observations between January 2007 and Dec 2014. All variables are winzorized at the 1 and 99 percent levels. See Appendix A for variable definitions.

Table D-2. Descriptive Statistics

Variable	N	Mean	Standard Deviation	25th Percentile	Median	75th Percentile
Abnormal Data Availability	42,616	0.024	0.161	-0.056	0.002	0.068
Abnormal Spread	42,616	0.092	0.476	-0.173	-0.003	0.233
Abnormal Turnover	42,616	0.004	0.008	0.000	0.002	0.006
Institutional Ownership	42,616	0.643	0.288	0.454	0.711	0.864
Abs (Weekly CAR)	42,616	0.059	0.060	0.018	0.040	0.079
Book to Market	42,616	1.037	2.054	0.356	0.596	0.944
Ln (Price)	42,616	3.024	1.001	2.449	3.180	3.742
Quarterly Volatility	42,616	0.027	0.016	0.016	0.023	0.033
Quarterly CAR	42,616	0.013	0.182	-0.087	0.002	0.098
Abnormal SVI	42,616	0.034	0.191	-0.056	0.006	0.082
Abs (Earnings Surprise)	42,616	0.010	0.034	0.001	0.002	0.006
Negative Surprise	42,616	0.435	0.496	0.000	0.000	1.000
Quarterly Turnover	42,616	0.010	0.009	0.005	0.008	0.013
Ln (Analyst Estimates)	42,616	1.945	0.747	1.386	1.969	2.546
Size	42,616	21.132	1.786	19.894	21.133	22.300
IPT [0,10]	42,225	3.678	1.784	2.456	3.562	4.660

Table D-2 reports descriptive statistics for my sample of earnings announcements between January 2007 and Dec 2014. All variables are winzorized at the 1 and 99 percent levels. See Appendix A for variable definitions.

Table D-3. Correlation Matrix 1

Variables	1	2	3	4	5	6	7	8	9	10	11
Mean Spread		-0.204	0.245	-0.744	0.258	-0.556	-0.336	0.009	0.295	0.002	0.015
Mean Turnover	-0.164		0.243	0.243	-0.028	0.366	0.342	0.179	0.217	0.019	0.019
Abs (Weekly Ret)	0.203	0.315		-0.181	0.080	-0.074	-0.006	0.153	0.677	0.012	0.014
Size	-0.559	0.105	-0.210		-0.316	0.658	0.264	0.016	-0.238	0.002	0.006
Book to Market	0.106	0.126	0.059	-0.184		-0.278	-0.168	0.000	0.084	-0.002	-0.005
Ln (Analyst Estimates)	-0.422	0.234	-0.087	0.648	-0.236		0.388	-0.030	-0.120	0.003	-0.002
Institutional Ownership	-0.352	0.173	-0.029	0.299	-0.325	0.435		-0.032	-0.050	0.005	-0.004
Abnormal Turnover	0.006	0.409	0.246	0.006	0.001	-0.007	-0.001		0.182	0.032	-0.003
Abs (Weekly CAR)	0.243	0.314	0.854	-0.255	0.059	-0.125	-0.069	0.288		0.016	0.019
Abnormal SVI	0.001	0.034	0.023	0.005	-0.003	0.007	0.001	0.056	0.027		0.036
Abnormal Data Availability	-0.001	0.020	0.017	0.011	-0.002	0.002	-0.002	0.004	0.017	0.029	

Spearman correlations are reported above the diagonal, and Pearson correlations are reported below the diagonal. This sample is all weekly observations between January 2007 and Dec 2014. See Appendix A for variable definitions.

Table D-4. Correlation Matrix 2

Variables	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1. Abnormal Data Availability		0.012	-0.004	0.001	0.024	-0.018	0.019	-0.012	-0.008	0.040	-0.017	0.002	0.005	-0.007	0.012	-0.012
2. Abnormal Spread	0.044		-0.030	0.010	0.007	-0.020	0.036	-0.078	-0.075	0.010	-0.029	0.011	-0.035	0.005	-0.024	0.005
3. Abnormal Turnover	0.007	0.000		0.258	0.350	-0.124	0.107	-0.054	-0.037	0.076	-0.015	-0.031	0.324	0.289	0.132	0.149
4. Institutional Ownership	0.006	0.020	0.185		0.045	-0.173	0.281	-0.054	-0.037	0.019	-0.174	0.009	0.368	-0.385	-0.261	0.067
5. Abs (Weekly CAR)	0.011	0.040	0.427	0.021		0.031	-0.208	0.345	-0.032	0.022	-0.189	0.046	0.146	-0.038	-0.202	0.256
6. Book to Market	0.001	-0.016	-0.042	0.328	-0.009		-0.342	0.170	0.130	-0.038	0.330	-0.115	-0.026	0.278	0.315	-0.041
7. Ln (Price)	-0.033	-0.051	-0.065	0.329	-0.243	-0.129		-0.505	-0.149	-0.039	0.442	-0.145	-0.100	-0.413	-0.692	-0.032
8. Quarterly Volatility	-0.027	-0.041	-0.084	0.084	-0.372	-0.106	-0.514		-0.020	0.000	-0.347	-0.125	-0.301	-0.209	-0.451	-0.005
9. Quarterly CAR	-0.011	-0.063	-0.038	-0.015	-0.021	-0.069	-0.121	-0.026		-0.014	-0.086	-0.100	-0.011	-0.017	-0.089	-0.016
10. Abnormal SVI	-0.031	-0.003	-0.106	-0.020	-0.035	-0.028	-0.050	-0.003	-0.011		-0.017	-0.007	-0.044	-0.073	-0.067	-0.022
11. Abs (Earnings Surprise)	-0.004	-0.002	-0.023	-0.129	-0.182	-0.172	-0.379	-0.362	-0.096	-0.010		-0.115	-0.003	-0.316	-0.381	-0.000

Table D-4. Continued

Variables	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
12. Negative Surprise	0.001	-0.007	-0.018	-0.009	0.048	0.043	-0.138	-0.135	-0.087	-0.015	-0.092		0.036	0.036	-0.089	-0.007
13. Quarterly Turnover	-0.002	-0.021	0.365	0.199	0.154	0.135	0.046	0.341	0.019	0.049	0.067	0.037		0.390	0.244	0.030
14. Ln (Analyst Estimates)	-0.008	0.000	0.218	0.423	-0.063	-0.228	-0.421	-0.199	0.009	-0.099	-0.179	0.036	0.277		0.659	0.062
15. Size	-0.031	-0.009	0.050	0.298	-0.231	-0.182	-0.696	-0.415	0.057	-0.099	0.263	-0.087	0.120	0.658		0.028
16. IPT [0,10]	-0.014	0.009	0.107	0.060	0.197	0.032	0.032	0.013	0.011	0.028	0.016	0.005	0.010	0.055	0.027	

Spearman correlations are reported above the diagonal, and Pearson correlations are reported below the diagonal. This sample is of earnings announcements between January 2007 and Dec 2014. See Appendix A for variable definitions.

Table D-5. Abnormal Data Availability and Corporate Events

	DV = Abnormal Data Availability			
Variable	Full	Small	Medium	Large
Earnings Announcement	0.003***	0.003**	0.002	0.002*
	(3.450)	(2.130)	(1.330)	(1.900)
10-K Filing	-0.006***	-0.004	-0.006**	-0.007***
	(-4.010)	(-1.460)	(-2.410)	(-2.890)
10-Q Filing	-0.007***	-0.008***	-0.008***	-0.004***
	(-7.570)	(-4.880)	(-5.950)	(-2.690)
# Announcements	0.022***	0.020***	0.022***	0.023***
	(95.060)	(44.900)	(55.950)	(57.540)
Rank_SVI	0.001***	0.001***	0.001***	0.001***
	(10.320)	(6.710)	(7.500)	(4.580)
Mean Spread	-0.021	-0.094	0.378	3.271***
	(-0.280)	(-1.080)	(0.900)	(5.570)
Mean Turnover	0.241***	0.221***	0.236***	0.288***
	(6.470)	(3.850)	(3.990)	(3.680)
Abs (Weekly Ret)	-0.041***	-0.051***	-0.073***	-0.011
	(-4.450)	(-3.800)	(-4.570)	(-0.570)
Size_Decile	0.001***	0.001*	0.001	0.002**
	(5.010)	(1.690)	(1.310)	(2.570)
BTM_Decile	0.000***	0.000	-0.001**	0.000*
	(-2.720)	(-1.290)	(-2.430)	(-1.650)
Q4	-0.006***	-0.006**	-0.006**	-0.006**
	(-4.420)	(-2.440)	(-2.420)	(-2.280)
Ln (Analyst Estimates)	-0.001*	-0.002	-0.002	0.000
	(-1.670)	(-1.580)	(-1.550)	(-0.260)
Institutional Ownership	-0.003**	-0.002	-0.004*	-0.003
	(-2.560)	(-1.140)	(-1.810)	(-1.100)
Abs (Weekly CAR)	0.062***	0.065***	0.089***	0.048**
	(5.730)	(4.150)	(4.640)	(2.050)
Year-Week Fixed Effects	Yes	Yes	Yes	Yes
N	605,338	201,780	201,779	201,779
R-square	0.160	0.155	0.161	0.165

Table D-5 reports results of Model (A1). Table D-5 examines the relation between my measure of data availability and corporate events and firm characteristics. Variable definitions are available in Appendix A. All variables are winsorized at 1 and 99 percent. Standard errors are clustered by firm. T-values are shown in parentheses below the coefficient estimates. This sample consists of all firm weeks between Jan 2007 and Dec 2014. *, **, *** indicate statistical significance at the p < 0.10, 0.05, and 0.01, respectively. All tests are two-tailed.

Table D-6. Abnormal Data around Corporate Events

	DV = Abnormal Data Availability			
Variable	Full	Small	Medium	Large
Earnings Announcement (Week Prior)	0.002**	0.001	0.001	0.002*
	(2.120)	(1.060)	(0.690)	(1.740)
Earnings Announcement	0.003***	0.004**	0.002	0.003**
	(3.350)	(2.360)	(1.010)	(1.960)
Earnings Announcement (Week After)	0.003***	0.004***	0.000	0.003**
	(3.400)	(3.170)	(0.220)	(2.220)
10-K Filing (Week Prior)	-0.002	-0.007***	0.006**	-0.003
	(-1.130)	(-2.720)	(2.130)	(-0.930)
10-K Filing	-0.005***	-0.004	-0.004	-0.006**
	(-3.400)	(-1.590)	(-1.350)	(-2.190)
10-K Filing (Week After)	0.003**	-0.003	0.010***	0.006**
	(2.050)	(-1.020)	(3.560)	(2.400)
10-Q Filing (Week Prior)	-0.001	0.000	-0.004**	0.000
	(-0.780)	(-0.110)	(-2.330)	(0.160)
10-Q Filing	-0.006***	-0.008***	-0.009***	-0.003
	(-5.380)	(-3.730)	(-4.490)	(-1.590)
10-Q Filing (Week After)	0.003**	0.000	0.002	0.005***
	(2.510)	(-0.070)	(1.080)	(2.740)
Year-Week Fixed Effects	Yes	Yes	Yes	Yes
N	605,338	201,780	201,779	201,779
R-square	0.160	0.155	0.161	0.165

Table D-6 reports results of Model (A2). Table D-6 examines date creation around corporate events.Variable definitions are available in Appendix A. Control variables in this table are as follows: # Announcements, Rank_SVI, Mean Spread, Mean Turnover, Abs Weekly Ret, Size_Decile, BTM_Decile, Q4, Ln (Analysts), Institutional Ownership, Abs Weekly CAR, Abnormal Turnover. All variables are winsorized at 1 and 99 percent. Standard errors are clustered by firm. T-values are shown in parentheses below the coefficient estimates. This sample consists of all firm weeks between Jan 2007 and Dec 2014. *, **, *** indicate statistical significance at the $p < 0.10$, 0.05, and 0.01, respectively. All tests are two-tailed.

Table D-7. Abnormal Data Availability and Speed by Firm Size

	DV = IPT [0,10]			
Variable	Full	Small	Medium	Large
Abnormal Data Availability	-0.141**	-0.096	-0.086	-0.231**
	(-2.430)	(-0.920)	(-0.890)	(-2.400)
Abnormal SVI	0.219***	0.174**	0.271***	0.249***
	(4.860)	(2.070)	(3.290)	(3.590)
Abs (Earnings Surprise)	-0.132	0.138	-0.034	0.328
	(-0.610)	(0.550)	(-0.050)	(0.590)
Negative Surprise	-0.004	-0.010	0.008	-0.014
	(-0.200)	(-0.330)	(0.270)	(-0.430)
Book to Market	-0.005	-0.010	-0.006	0.007
	(-1.040)	(-1.570)	(-0.760)	(0.520)
Quarterly Turnover	-2.771**	1.790	-5.226**	-10.433***
	(-2.090)	(0.760)	(-2.390)	(-3.840)
Ln (Analysts Estimates)	0.089***	0.048	0.095***	0.166***
	(4.880)	(1.430)	(3.000)	(5.390)
Ln (Quarterly Volatility)	0.023	-0.013	0.012	0.129***
	(0.910)	(-0.320)	(0.290)	(2.630)
Institutional Ownership	0.308***	0.356***	0.300***	0.098
	(8.350)	(5.790)	(4.890)	(1.200)
Size	-0.003	0.028	-0.009	-0.049**
	(-0.350)	(1.190)	(-0.240)	(-2.570)
Quarterly CAR	0.106**	0.118*	0.051	0.102
	(2.190)	(1.770)	(0.550)	(0.900)
Year-Industry Fixed Effects	Yes	Yes	Yes	Yes
N	42045	13904	14035	14106
Pseudo R-square	0.003	0.003	0.004	0.003

Table D-7 reports results of Model (1). Table D-7 examines the relation between the speed of price discovery and my measure of data availability. Variable definitions are available in Appendix A. All variables are winsorized at 1 and 99 percent. Standard errors are clustered by firm. Z-values are shown in parentheses below the coefficient estimates. This sample consists of earnings announcement weeks between Jan 2007 and Dec 2014. *, **, *** indicate statistical significance at the $p < 0.10$, 0.05, and 0.01, respectively. All tests are two-tailed.

Table D-8. Abnormal Data Availability and Information Asymmetry by Firm Size

	DV = Abnormal Spread			
Variable	Full	Small	Medium	Large
Abnormal Data Availability	0.075***	0.014	0.075***	0.116***
	(4.770)	(0.520)	(2.620)	(4.390)
Abnormal SVI	0.014	0.001	-0.001	0.014
	(1.090)	(0.040)	(-0.060)	(0.670)
Abnormal Turnover	-1.859***	-4.581***	-0.719	0.534
	(-4.930)	(-7.220)	(-1.230)	(0.680)
Institutional Ownership	0.005	0.047***	-0.024	0.019
	(0.540)	(2.680)	(-1.470)	(0.830)
Abs (Weekly CAR)	0.482***	0.512***	0.436***	0.537***
	(9.440)	(7.700)	(4.580)	(3.710)
Ln (Analyst Estimates)	0.019***	0.006***	0.001	0.002**
	(3.940)	(3.320)	(1.100)	(2.190)
Size	-0.028***	-0.043***	-0.040***	-0.010*
	-(12.680)	(-5.810)	(-4.070)	(-1.820)
Book to Market	-0.003**	-0.003*	-0.005**	-0.002
	(-2.260)	(-1.820)	(-2.300)	(-0.350)
Ln (Price)	0.052***	0.041***	0.070***	0.056***
	(14.310)	(6.520)	(10.620)	(7.400)
Abs (Earnings Surprise)	0.183***	0.096	0.337*	0.656
	(2.710)	(1.310)	(1.690)	(1.250)
Negative Surprise	0.023***	0.031***	0.004	-0.005
	(4.690)	(4.230)	(0.440)	(-0.650)
Quarterly Turnover	-0.339	1.196**	0.517	-1.767***
	(-1.030)	(2.010)	(0.910)	(-2.710)
Ln (Quarterly Volatility)	-0.037***	-0.075***	-0.054***	-0.001
	(-5.200)	(-6.290)	(-4.270)	(-0.070)
Quarterly CAR	-0.150***	-0.197***	-0.139***	0.003
	(-11.260)	(-11.160)	(-5.560)	(0.100)
Year-Industry Fixed Effects	Yes	Yes	Yes	Yes
N	42433	14097	14163	14173
R-square	0.035	0.045	0.043	0.039

Table D-8 reports results of Model (2). Table D-8 examines the relation between my measure of data availability and information asymmetry around an earnings announcement. Variable definitions are available in Appendix A. All variables are winsorized at 1 and 99 percent. Standard errors are clustered by firm. T-values are shown in parentheses below the coefficient estimates. This sample consists of earnings announcement weeks between Jan 2007 and Dec 2014. *, **, *** indicate statistical significance at the $p < 0.10$, 0.05, and 0.01, respectively. All tests are two-tailed.

Table D-9. Concave Relation between Data Availability and Price Discovery

	DV = IPT [0,10]		
Variable	Low	Medium	High
Abnormal Data Availability	0.862**	0.059	-0.261
	(2.450)	(0.050)	(-1.380)
Abnormal SVI	-0.050	0.174	0.379***
	(-0.350)	(1.140)	(2.690)
Abs (Earnings Surprise)	-0.235	-0.064	0.643*
	(-0.600)	(-0.140)	(1.650)
Negative Surprise	0.014	-0.032	-0.020
	(0.250)	(-0.610)	(-0.380)
Book to Market	-0.016	-0.013	-0.002
	(-1.500)	(-1.160)	(-0.190)
Quarterly Turnover	1.038	2.881	1.785
	(0.280)	(0.670)	(0.480)
Ln (Analysts Estimates)	0.050	0.083	0.019
	(0.890)	(1.380)	(0.310)
Ln (Quarterly Volatility)	-0.024	0.045	-0.063
	(-0.340)	(0.630)	(-0.860)
Institutional Ownership	0.383***	0.297	0.385***
	(3.630)	(2.610)	(3.900)
Size	0.039	0.068*	-0.026
	(0.900)	(1.690)	(-0.650)
Quarterly CAR	0.167	-0.020	0.176
	(1.490)	(-0.170)	(1.460)
Year-Industry Fixed Effects	Yes	Yes	Yes
N	4686	4591	4617
Pseudo R-square	0.0045	0.0045	0.0039

Table D-9 reports results of Model (1). Variable definitions are available in Appendix A. All variables are winsorized at 1 and 99 percent. Standard errors are clustered by firm. Z-values are shown in parentheses below the coefficient estimates. This sample consists of earnings announcement weeks between Jan 2007 and Dec 2014 for the smallest tercile of firms. Firms are further sorted into terciles based on the amount of data available in the week of the earnings announcement. *, **, *** indicate statistical significance at the $p < 0.10$, 0.05, and 0.01, respectively. All tests are two-tailed.

Table D-10. Concave Relation between Data Availability and Information Asymmetry

	DV = Abnormal Spread		
Variable	Low	Medium	High
Abnormal Data Availability	-0.047	-0.085	0.055
	(-0.530)	(-0.270)	(1.210)
Abnormal SVI	-0.014	-0.013	0.031
	(-0.360)	(-0.350)	(0.800)
Abnormal Turnover	-4.245***	-5.043***	-4.473***
	(-4.070)	(-5.180)	(-4.180)
Institutional Ownership	0.039	0.035	0.070**
	(1.340)	(1.310)	(2.280)
Abs (Weekly CAR)	0.332***	0.592***	0.597***
	(2.850)	(4.990)	(5.070)
Ln (Analyst Estimates)	0.007**	0.005**	0.005*
	(2.240)	(2.120)	(1.800)
Size	-0.036***	-0.042***	-0.053***
	(-2.920)	(-3.840)	(-4.470)
Book to Market	-0.002	-0.005*	-0.002
	(-0.900)	(-1.950)	(-0.820)
Ln (Price)	0.045***	0.039***	0.042***
	(4.420)	(3.980)	(3.990)
Abs (Earnings Surprise)	0.317**	-0.066	0.058
	(2.460)	(-0.580)	(0.480)
Negative Surprise	0.021*	0.042***	0.034**
	(1.720)	(3.140)	(2.490)
Quarterly Turnover	-1.103	3.022***	1.585
	(-1.210)	(3.090)	(1.520)
Ln (Quarterly Volatility)	-0.039*	-0.071***	-0.109***
	(-1.860)	(-3.770)	(-5.120)
Quarterly CAR	-0.224***	-0.208***	-0.163
	(-7.600)	(-6.800)	(-5.240)
Year-Industry Fixed Effects	Yes	Yes	Yes
N	4749	4653	4586
R-square	0.0426	0.0548	0.0532

Table D-10 reports results of Model (2). Variable definitions are available in Appendix A. All variables are winsorized at 1 and 99 percent. Standard errors are clustered by firm. T-values are shown in parentheses below the coefficient estimates. This sample consists of earnings announcement weeks between Jan 2007 and Dec 2014 for the smallest tercile of firms. Firms are further sorted into terciles based on the amount of data available in the week of the earnings announcement. *, **, *** indicate statistical significance at the p < 0.10, 0.05, and 0.01, respectively. All tests are two-tailed.

Table D-11. Abnormal Data Availability and Speed by Institutional Ownership

	DV = IPT [0,10]		
Variable	Fewest	Medium	Most
Abnormal Data Availability	-0.010	-0.096	-0.290***
	(-0.100)	(-0.960)	(-3.010)
Abnormal SVI	0.239***	0.142*	0.292***
	(2.940)	(1.930)	(3.830)
Abs (Earnings Surprise)	-0.155	0.341	0.149
	(-0.560)	(0.720)	(0.260)
Negative Surprise	0.011	0.032	-0.055*
	(0.360)	(1.010)	(-1.750)
Book to Market	-0.004	-0.053**	-0.019
	(-0.770)	(-2.470)	(-0.840)
Quarterly Turnover	-1.085	-3.531	-4.541*
	(-0.510)	(-1.500)	(-1.840)
Ln (Analysts Estimates)	0.029	0.140***	0.114
	(1.020)	(4.260)	(3.310)
Ln (Quarterly Volatility)	0.012	0.014	0.057
	(0.310)	(0.310)	(1.240)
Institutional Ownership	0.372***	0.081	0.113
	(3.780)	(0.360)	(0.610)
Size	0.016	-0.024*	-0.025
	(1.470)	(-1.780)	(-1.380)
Quarterly CAR	0.111	0.119	0.075
	(1.550)	(1.330)	(0.820)
Year-Industry Fixed Effects	Yes	Yes	Yes
N	13813	14119	14084
Pseudo R-square	0.003	0.0026	0.003

Table D-11 reports results of Model (1). The columns in Table D-11 represent terciles of institutional ownership. Variable definitions are available in Appendix A. All variables are winsorized at 1 and 99 percent. Standard errors are clustered by firm. Z-values are shown in parentheses below the coefficient estimates. This sample consists of earnings announcement weeks between Jan 2007 and Dec 2014. *, **, *** indicate statistical significance at the $p < 0.10$, 0.05, and 0.01, respectively. All tests are two-tailed.

Table D-12. Abnormal Data Availability and Information Asymmetry – Institutional Ownership

	DV = Abnormal Spread		
Variable	Fewest	Medium	Most
Abnormal Data Availability	0.071***	0.080***	0.062**
	(2.630)	(3.030)	(2.230)
Abnormal SVI	0.005	0.023	0.000
	(0.240)	(1.100)	(0.010)
Abnormal Turnover	-4.952***	-1.044	-0.049
	(-8.330)	(-1.510)	(-0.080)
Institutional Ownership	0.020	-0.065	-0.088
	(0.760)	(-1.200)	(-1.640)
Abs (Weekly CAR)	0.496***	0.527***	0.406***
	(6.440)	(5.510)	(4.320)
Ln (Analyst Estimates)	0.004***	0.001*	0.002**
	(3.440)	(1.670)	(2.370)
Size	-0.028***	-0.025***	-0.038***
	(-8.290)	(-6.250)	(-6.490)
Book to Market	-0.004***	0.022***	0.016**
	(-3.400)	(2.630)	(2.430)
Ln (Price)	0.041***	0.058***	0.081***
	(7.180)	(9.010)	(10.710)
Abs (Earnings Surprise)	0.068	0.092	0.398**
	(0.850)	(0.670)	(2.310)
Negative Surprise	0.011	0.010	0.009
	(1.400)	(1.290)	(1.020)
Quarterly Turnover	0.331	-0.528	-0.597
	(0.750)	(-0.840)	(-0.900)
Ln (Quarterly Volatility)	-0.051***	-0.030**	-0.027*
	(-4.450)	(-2.420)	(-1.940)
Quarterly CAR	-0.212***	-0.134***	-0.067**
	(-10.660)	(-5.630)	(-2.500)
Year-Industry Fixed Effects	Yes	Yes	Yes
N	14088	14177	14138
R-square	0.036	0.0395	0.0418

Table D-12 reports results of Model (2). The columns in Table D-12 represent terciles of institutional ownership. Variable definitions are available in Appendix A. All variables are winsorized at 1 and 99 percent. Standard errors are clustered by firm. T-values are shown in parentheses below the coefficient estimates. This sample consists of earnings announcement weeks between Jan 2007 and Dec 2014. *, **, *** indicate statistical significance at the $p < 0.10$, 0.05, and 0.01, respectively. All tests are two-tailed.

REFERENCES

Barber, B. M., & Odean, T. (2008). All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors. *Review of Financial Studies, 21*(2), 785-818.

Barber, B. M., Odean, T., & Zhu, N. (2009). Do retail trades move markets? *Review of Financial Studies, 22*(1), 151-186.

Beaver, W. H. (1968). The information content of annual earnings announcements. *Journal of Accounting Research, 6*(Empirical Research in Accounting: Selected Studies 1968), 67-92.

Blankespoor, E., Miller, G. S., & White, H. D. (2014). The role of dissemination in market liquidity: Evidence from firms' use of twitter™. *The Accounting Review, 89*(1), 79-112.

Bloomfield, R. J. (2002). The "Incomplete revelation hypothesis" and financial reporting. *Accounting Horizons, 16*(3), 233-243.

Bloomfield, R. J., Libby, R., & Nelson, M. W. (2003). Do investors overrely on old elements of the earnings time series?*. *Contemporary Accounting Research, 20*(1), 1-31.

Brown, N. C., Stice, H., & White, R. M. (2015). Mobile communication and local information flow: Evidence from distracted driving laws. *Journal of Accounting Research, 53*(2), 275-329.

Bushee, B. J., Core, J. E., Guay, W., & Hamm, S. J. W. (2010). The role of the business press as an information intermediary. *Journal of Accounting Research, 48*(1), 1-19.

Chen, H., De, P., Hu, Y. (., & Hwang, B. (2014). Wisdom of crowds: The value of stock opinions transmitted through social media. *Review of Financial Studies, 27*, 1367-1403.

Choi, H., & Varian, H. (2012). Predicting the present with google trends. *Economic Record, 88*, 2-9.

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

D'Avolio, G., Gildor, E., & Shleifer, A. (2001). Technology, information production, and market efficiency. *Economic Symposium Conference Proceedings: Economic Policy for the Information Economy*, , 125-160.

Dellavigna, S., & Pollet, J. M. (2009). Investor inattention and friday earnings announcements. *The Journal of Finance, 64*(2), 709-749.

Drake, M., Thornock, J., & Twedt, B. (2015). The internet as an information intermediary. *Working Paper, Brigham Young University*,

Drake, M. S., Guest, N. M., & Twedt, B. J. (2014). The media and mispricing: The role of the business press in the pricing of accounting information. *The Accounting Review, 89*(5), 1673-1701.

Drake, M. S., Roulstone, D. T., & Thornock, J. R. (2012). Investor information demand: Evidence from google searches around earnings announcements. *Journal of Accounting Research, 50*(4), 1001-1040.

Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *The Journal of Finance, 25*(2, Papers and Proceedings of the Twenty-Eighth Annual Meeting of the American Finance Association New York, N.Y. December, 28-30, 1969), 383-417.

Fama, E. F. (1998). Market efficiency, long-term returns, and behavioral finance1. *Journal of Financial Economics, 49*(3), 283-306.

Fang, L., & Peress, J. (2009). Media coverage and the cross-section of stock returns. *The Journal of Finance, 64*(5), 2023-2052.

Frederickson, J. R., & Zolotoy, L. (2015). Competing earnings announcements: Which announcement do investors process first? *The Accounting Review*

Gigerenzer, G., & Selten, R. (2002). *Bounded rationality: The adaptive toolbox* Mit Press.

Grossman, S. J., & Stiglitz, J. E. (1980). On the impossibility of informationally efficient markets. *The American Economic Review, 70*(3), 393-408.

Hirshleifer, D., & Teoh, S. H. (2003). Limited attention, information disclosure, and financial reporting. *Journal of Accounting and Economics, 36*(1–3), 337-386.

Hong, H., & Stein, J. C. (1999). A unified theory of underreaction, momentum trading, and overreaction in asset markets. *The Journal of Finance, 54*(6), 2143-2184.

Kahneman, D. (1973). *Attention and effort* Prentice-Hall.

Li, E. X., Ramesh, K., & Shen, M. (2011). The role of newswires in screening and disseminating value-relevant information in periodic SEC reports. *The Accounting Review, 86*(2), 669-701.

Libby, R., Bloomfield, R., & Nelson, M. W. (2002). Experimental research in financial accounting. *Accounting, Organizations and Society, 27*(8), 775-810.

Loughran, T., & Schultz, P. (2004). Weather, stock returns, and the impact of localized trading behavior. *Journal of Financial and Quantitative Analysis, 39*(02), 343-364.

Merton, R. C. (1987). A simple model of capital market equilibrium with incomplete information. *The Journal of Finance, 42*(3), 483-510.

Miller, G. S. (2006). The press as a watchdog for accounting fraud. *Journal of Accounting Research, 44*(5), 1001-1033.

Mondria, J., Wu, T., & Zhang, Y. (2010). The determinants of international investment and attention allocation: Using internet search query data. *Journal of International Economics, 82*(1), 85-95.

Rubinstein, M. (2001). Rational markets: Yes or no? the affirmative case. *Financial Analysts Journal, 57*(3), 15-29.

Shive, S. (2012). Local investors, price discovery, and market efficiency. *Journal of Financial Economics, 104*(1), 145-161.

Soloman, D. H. (2012). Selective publicity and stock prices. *The Journal of Finance, 67*(2), 599-638.

Tetlock, P. C. (2011). All the news that's fit to reprint: Do investors react to stale information? *Review of Financial Studies, 24*, 1481-1512.

Twedt, B. (2015). Spreading the word: Price discovery and newswire dissemination of management earnings guidance. *The Accounting Review*

BIOGRAPHICAL SKETCH

Han Stice earned both a bachelor's and master's degree in Accounting from Brigham Young University and his Ph.D. from the University of Florida. Han comes from a family of accounting academics. His father is an accounting professor, his uncle is an accounting professor, his brother is an accounting professor and his sister is currently a PhD student in Accounting at UNC – Chapel Hill.