Geographic Dispersion of Investors and Price Discovery around Earnings Announcements

by

Jason Victor Chen

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy (Business Administration) In the University of Michigan 2016

Doctoral Committee:

Professor Reuven Lehavy, Chair
Associate Professor Feng Li
Assistant Professor David Maber
Associate Professor Yusufcan Masatlioglu
Associate Professor Paolo Pasquariello

Acknowledgements

I am grateful to my dissertation committee for their guidance and support: Reuven Lehavy (Chair), Feng Li, David Maber, Yusufcan Masatlioglu, and Paolo Pasquariello. I also appreciate the feedback and encouragement from fellow colleagues at the Ross business school: Lindsey, Greg, Venky, Lloyd, Ryan, Randy, Jed, and Christina. I am also grateful for the financial support provided by the Paton Accounting Fellowship, Auerbach Scholarship, and the Harry Jones endowment.

Table of Contents

Acknowledgements ii
List of Figures v
List of Tables vi
Abstract vii
Chapter 1. Introduction 1
Chapter 2. Prior Research and Predictions 7
  2.1 Prior Research of the Effects of Geographic Location on Information Acquisition 7
  2.2 Predictions 8
    2.2.1 Information Asymmetry 9
    2.2.2 Trading 10
    2.2.3 Informativeness of Price 10
    2.2.4 Geographic Distribution of Information 11
Chapter 3. Estimating Investors Locations and Measuring their Geographic Dispersion 13
  3.1 Estimating Investors Locations 13
    3.1.1 Investor Location Sample Characteristics 15
  3.2 Measuring Investor Geographic Dispersion 16
Chapter 4. Market Variables, Control Variables, and Sample Descriptive Statistics 19
  4.1 Market Variables 19
    4.1.1 Abnormal Bid-Ask Spread 19
    4.1.2 Abnormal Market Depth 20
    4.1.3 Abnormal Illiquidity 20
    4.1.4 Abnormal Trading Volume 21
    4.1.5 Abnormal Price Drift or Reversal 21
  4.2 Control Variables 23
  4.3 Sample Descriptive Statistics 24
Chapter 5. Research Design and Findings 26

5.1 Information Asymmetry around Earnings Announcements....................................26
5.2 Trading around Earnings Announcements 28
5.3 Informativeness of Price...30
5.4 Geographic Distribution of Information 32
Chapter 6. Robustness...35
6.1 Placebo Earnings Announcement Dates...35
6.2 Market Reaction around Earnings Announcements...36
6.3 Outsourcing Companies using EDGAR...37
6.4 American Depository Receipt (ADR) Listings 38
Chapter 7. Conclusion...39
Figures...41
Tables 46
Appendix...57
Bibliography 60

List of Figures

Figure 1. Locations of Requests for Firms' Filings World-Wide 2005 – 2012 42

Figure 2. Locations of Requests for Firms' Filings within the United States 2005 – 2012 43

Figure 3. Locations of Requests for Firms' Filings within New York City 2005 – 2012 .44

Figure 4. High versus Low Investor Geographic Dispersion 2011 45

List of Tables

Table 1. Sample of Requests to EDGAR 47

Table 2. Top 10 Countries, States, and Cities by Number of Requests to EDGAR 2005 – 2012..48

Table 3. Descriptive Statistics...49

Table 4. Pearson and Spearman Correlations 51

Table 5. Results of Regressing Information Asymmetry around Earnings Announcements on Investor Geographic Dispersion 52

Table 6. Results of Regressing Trading around Earnings Announcements on Investor Geographic Dispersion 53

Table 7. Results of Regressing Informativeness of Price around Earnings Announcements on Investor Geographic Dispersion 54

Table 8. Results of Regressing Information Asymmetry, Trading, and the Informativeness of Price on Investor Geographic Dispersion, Partitioned by Geographic Distribution of Information 55

Table 9. Results of Regressing Absolute Value of Abnormal Returns around Earnings Announcements on Investor Geographic Dispersion 56

Abstract

Geographic Dispersion of Investors and Price Discovery around Earnings Announcements

By

Jason Victor Chen

Chair: Reuven Lehavy

Past research has documented a geographic role in the information available to investors. Specifically, studies have shown that value relevant information available to investors differs across locations and that it is less costly for local investors to obtain local information. Drawing on this body of research, I examine the influence of investor geographic dispersion on the price discovery process around earnings announcements. In particular, I explore whether the geographic dispersion of investors leads to a greater diversity of information amongst investors and, by extension, greater information asymmetry, greater trading volume, and ultimately more informative price following earnings announcements. Consistent with my predictions, I find that firms whose investors are more geographically dispersed experience higher abnormal bid-ask spreads, lower abnormal market depth, higher abnormal illiquidity, higher abnormal trading volume, and lower price drift or reversal around quarterly earnings announcements. I further find that these results are more pronounced for firms with a greater geographic distribution of

information. These findings contribute to our understanding of the role of investor dispersion in the price discovery process.

CHAPTER 1

Introduction

When valuing firms, investors draw on a variety of sources in obtaining their information. For example, investors may use firm information regarding facilities, distribution hubs, or sales centers to gain insight into a company's operations. Or, they may use information about competitors, consumer preferences, political policies, and numerous other aspects of the business landscape to form a more complete base of knowledge about a firm. However, a key challenge in acquiring this information is that it does not reside in a central repository, but instead is spread across numerous locations (Hayek 1945). This geographic distribution imposes a significant cost on investors who attempt to aggregate the information (Addoum et al. 2013). Much of this information may only be acquired locally through direct experience or is prohibitively costly for non-local investors to obtain due to limitations on their attention (Grinblatt and Keloharju 2001; Coval and Moskowitz 2001). While prior research has documented a relation between an individual investor's location and his or her trading behavior, the role of the dispersion in investor's locations in the price discovery process remains an open question. In this study, I contribute to our understanding of the role of dispersion by providing novel empirical evidence that firms with geographically dispersed investors have more diversely informed investors facilitating their price discovery around earnings announcements.

Specifically, I posit that the impact of investor dispersion on information diversity affects three aspects of the price discovery process around earnings announcements: information asymmetry, trading volume, and the informativeness of price. To examine the impact of dispersion on these three elements of price discovery, I develop a series of predictions based upon prior research which suggests that investors utilize private information when evaluating public disclosures (Kim and Verrecchia 1994). First, I predict that firms with greater investor dispersion will exhibit greater information asymmetry between the market and market makers during the price discovery process around earnings announcements. Next, I predict that greater investor dispersion will lead to greater trading among investors, who draw on their individual knowledge base when interpreting earnings announcements. Third, I predict that greater dispersion will lead to more informative prices, as investors are able to incorporate their local information into the price discovery process. Finally, to validate my predictions, I hypothesize that the above effects will be stronger for firms with a greater distribution of information across locations.

In my tests of these predictions, I use a unique dataset which contains the latitudinal and longitudinal locations of market participants who use the Electronic Data Gathering, Analysis, and Retrieval (hereafter, EDGAR) filing system during each of the fiscal quarters between 2005Q1 and 2012Q1. This sample allows me to estimate the location of individual market participants who have actively sought information about a specific firm.

To measure information asymmetry, I use the following three proxies: abnormal bid-ask spreads, abnormal market depth, and abnormal illiquidity around quarterly earnings announcements. Applying these three measures of information asymmetry to the firms in my sample, I find that a greater geographic dispersion of investors is associated with higher abnormal bid-ask spreads, lower abnormal market depth, and higher abnormal illiquidity around quarterly earnings announcements. Higher abnormal bid-ask spreads suggest that market makers are less willing to transact at a low cost during price discovery (Lee and Ready 1991). Lower abnormal market depth and higher abnormal illiquidity suggest greater abnormal movements in price for each trade during the price discovery process (Goyenko et al. 2009). Together, these findings support the idea that firms with greater investor dispersion experience greater information asymmetry during price discovery around earnings announcements.

In addition to information asymmetry, I examine the effect of geographic dispersion on the extent of trading during the price discovery process. Here, I find that investor dispersion is positively associated with abnormal trading volume in the immediate period around quarterly earnings announcements. This finding is consistent with my prediction that greater investor geographic dispersion leads to greater diversity in individual investor knowledge, which in turn contributes to greater differences in investor interpretations of earnings announcements.

Next, I examine the impact of investor dispersion on the informativeness of price following earnings announcements, using the extent of price drift or reversal following quarterly earnings announcement (hereafter, DriftRev) as my proxy for the informativeness of price. Prior research suggests that under (over) reactions to new information are corrected over time as realizations of the expectations based upon the new information make their way into price or as other information is slowly incorporated by the market (Bernard and Thomas 1989; Rangan and Sloan 1998). Based on this, I measure the informativeness of price using the difference between the market's immediate and long term reaction to earnings news. Specifically, I measure DriftRev as the absolute value of the percentage difference between the abnormal return in the immediate period around the earnings announcement and the abnormal return over a long window period. Doing so, I find that firms with greater investor dispersion have lower DriftRev, supporting the prediction that greater dispersion is associated with a more informative price.

Finally, I predict that my observed relations will be stronger for firms whose information is more geographically distributed. To estimate the geographic distribution of information about a firm, I tally the number of unique states and countries mentioned in the firm's 10-K filing. I then partition my sample into firms with high and low distribution of information based on whether they are above or below the median of this tally. My findings show that the effect of investor dispersion on abnormal bid-ask spreads, abnormal depth, abnormal illiquidity, abnormal trading volume, and price drift or reversal, respectively, is more pronounced for firms with a high number of unique states and countries mentioned in their 10-K filings. Combined, this set of results provides further evidence for my hypotheses that investor dispersion leads to greater information diversity in the price discovery process.

My analysis of the effect of investor geographic dispersion on price discovery contributes to the literature related to the effect of investor location on trading behavior. Much of this research has focused on the effect of an individual investor's geographic location on his or her information acquisition and trading profits (Feng and Seasholes 2004; Grinblatt and Keloharju 2011; Coval and Moskowitz 2001). The findings in my study extend our understanding of the role of investor location by providing evidence that the dispersion of firm's investors plays a role in price discovery process around earnings announcements.[1] My study most closely relates to a concurrent working paper which examines the relation between the aggregate average distance of investors to the firm's headquarters and the markets response to earnings announcements (Chi and Shanthikumar 2014).

This study also contributes to the literature which explores the use of the EDGAR filing system. One criticism of prior studies which have explored this setting is that little is known about the characteristics of the individuals accessing the EDGAR filing system. By focusing on investor location, my study contributes to our understanding of EDGAR users, thereby enriching our understanding of the results from other studies based on the EDGAR filing system (Drake et al. 2012a, 2012b; Lee et al. 2013).

The rest of the paper is organized as follows. Chapter 2 further develops my hypothesis and formally outlines my predictions. Chapter 3 provides details regarding my estimation of the geographic dispersion of firm investors. Chapter 4 details the market and control variables used in my study and discusses sample descriptive statistics. Chapter 5

[1] The findings in this study also contribute to research which explores the conditions under which price is more likely to be informative (Lee 2001; Kothari 2001; Beyer et al. 2010).

discusses my research design and findings. Chapter 6 presents robustness, and Chapter 7 concludes the study.

CHAPTER 2

Prior Research and Predictions

2.1 Prior Research on the Effects of Geographic Location on Information Acquisition

Prior research has shown that an investor's distance to information affects whether that investor acquires and trades upon it. It suggests that investor attention and resource constraints make it more difficult for non-local investors to acquire local information. Consequently, local investors earn greater returns than non-local investors since they are better able to acquire value relevant information about a given firm. Ivkovic and Weisbenner (2005) examine individuals' trades placed through a discount brokerage and find that the average household earns greater returns from their holdings in local companies when compared to their non-local holdings. Massa and Simonov (2006) find similar evidence for a sample of individual level trades by Swedish investors.

Studies find that the advantage of a local presence extends to managers and sophisticated investors suggesting that this effect is not isolated to a specific class of investor. For instance, Giroud (2013) provides evidence that when managers are in closer proximity to a plant, their information asymmetry decreases and they are thus better able to monitor a plant's operations. In another study, Coval and Moskowitz (2001) find that the local investments of fund managers outperform their non-local picks. They argue that these findings are a result of fund managers having a local information advantage to non-local investors.

My study primarily draws from the prior literature on the relation between investor location and information diversity. Generally speaking, extant research argues that the costs imposed on acquiring information by investor's distance to information lead to geographically segmented pools of knowledge. Corroborating this, research finds that individual investors who are located near one another exhibit similar trading behavior. For example, Feng and Seasholes (2004) find evidence of correlated trading among investors who place trades from the same branch office of a local brokerage. In another study, Hong, Kubik, and Stein (2005) show that this effect extends to sophisticated investors, finding that individual mutual fund managers who are located in the same city exhibit similar buying and selling behavior.

2.2. Predictions

I build upon these findings in the extant literature and argue that when an investor base is more geographically dispersed, they are more likely to be located in a greater number of distinct local pools of knowledge. Ultimately, this leads to greater diversity of value relevant information among a firm's investors. Theory suggests that this effect of the geographic dispersion of investors on the diversity of information among investors has potential important implications on how earnings announcements are incorporated by the market during the price discovery process.

In classical theories of price discovery, information is incorporated into firm valuations easily and uniformly across investors (Grossman 1976). These theories assume that investors inherently know how new information is incorporated into price. Extending this assumption, one would expect that public disclosures of new information would be followed by immediate price adjustment and potentially no abnormal market activity, as this information would be instantly reflected in price (Milgrom and Stokey 1982). However, empirical studies contradict this prediction, and instead find evidence of abnormal market activity around public releases of information (Bamber 1987; Ball and Brown 1968; Carter and Soo 1999). A more recent stream of research proposes that investor's interpretation of new disclosures depends upon the information they already possess (Harris and Raviv 1993; Kim and Verrecchia 1994). Therefore, a public disclosure, such as an earnings announcement, isn't necessarily value relevant in and of itself but rather the market's reaction to the disclosure will depend upon the private information of each investor.

2.2.1 Information Asymmetry

If geographically dispersed investors draw from diverse local information sources then there is greater potential for more diversity in the private information that investors possess during the price discovery process around earnings announcements. This diversity of information suggests that the market as a whole is potentially more informed. If so, then this places market makers at a greater information disadvantage to the market when investors are geographically dispersed thus leading market makers to price protect themselves (Kyle 1985, Glosten and Milgrom 1985, Madureira and Underwood 2008). This leads to my first prediction:

Prediction 1: There is a positive association between the geographic dispersion of investors and information asymmetry during the price discovery process around a firm's earnings announcement.

2.2.2 Trading

Following the above argument, if geographically dispersed investors possess diverse private information, they may draw different conclusions regarding how new information pertains to price, which in turn would lead to trading among investors (Kim and Verrecchia 1994, Verrecchia 2001). Indeed, Kandel and Pearson (1995) provide empirical evidence supporting this idea, finding that the trading volume which takes place around the release of earnings can be partially explained by different interpretations of the announcement due to differences in investor private information. Therefore, firms with greater investor geographic dispersion may exhibit greater trading volume during the price discovery process.[2] This leads to my second prediction:

Prediction 2: There is a positive association between the geographic dispersion of investors and trading during the price discovery process around a firm's earnings announcement.

2.2.3 Informativeness of Price

Price is informative following price discovery in that it aggregates the information possessed by investors which is revealed through their trades. Therefore, if geographically

---

[2] Informed investors face a subtle tradeoff between executing a trade when liquidity is potentially lower and being informed. My argument assumes that informed investors, on average, rush to trade upon their private information.

dispersed investors draw from a greater diversity of unique local information sources, then they are potentially incorporating a greater amount of information into price during the price discovery process which takes places around earnings announcements (Beaver 1997; Hong and Page 2001; Watson et al. 1993). This leads to my third prediction:

Prediction 3: There is a positive association between the geographic dispersion of investors and the informativeness of price following the price discovery process around a firm's earnings announcement.

2.2.4 Geographic Distribution of Information

My final set of predictions extends the previous predictions. Predictions (1) through (3) are based on the idea that geographically dispersed investors acquire and incorporate unique local knowledge into price discovery. Cross-sectionally, this idea implies that firms with a greater geographic distribution of information should provide geographically dispersed investors with more opportunities to acquire and incorporate unique local knowledge. Therefore, I predict that the respective associations between the geographic dispersion of investors and a firm's price discovery (information asymmetry, trading, and informativeness of price) should be more pronounced for firms with a greater geographic distribution of information. This leads to my final set of predictions:

Prediction 4a: The positive association between the geographic dispersion of investors and information asymmetry during the price discovery process around a firm's earnings announcement is greater for firms with a greater geographic distribution of information.

Prediction 4b: The positive association between the geographic dispersion of investors and trading during the price discovery process around a firm's earnings announcement is greater for firms with a greater geographic distribution of information.

Prediction 4c: The positive association between the geographic dispersion of investors and the informativeness of price following the price discovery process around a firm's earnings announcement is greater for firms with a greater geographic distribution of information.

CHAPTER 3

Estimating Investors Locations and Measuring their Geographic Dispersion

3.1 Estimating Investors Locations

To estimate where firms' investors are geographically located, I use the locations of the requests for companies' filings to the EDGAR online filing system.[3] This system, created by the SEC in 1995, provides an electronic method for public companies to submit their filings. EDGAR also provides a comprehensive repository of all public companies' filings, which is freely accessible to the public via the Internet.

The web server log files of the EDGAR filing system where provided by the SEC through a freedom of information act request. The log files span from 2/4/2003 to 3/30/2012, providing approximately 9 years of requests to the EDGAR system. The log files provide the Internet address, filing accessed, and date of access for each request among other information.[4] [5]

---

[3] This measure of investors encapsulates individuals who have shown an active interest searching for information about the firm.

[4] The log files do not include requests made to the EDGAR FTP server.

[5] Internet address refers to the unique Internet Protocol Version 4 (IPV4) address of the request.

I implement heuristics to remove requests likely to have been generated from automated programs. First, I check if there have been more than 100 filings downloaded by a unique Internet address in a given day. I then check if at least 90 percent of those are text documents; if so, I conclude that the requests have been automatically generated and remove the requests from the associated Internet address from my sample for the day. The rationale behind this process is that the text documents contain meta-data and the raw code for attachments which are, for the most part, only machine readable. I next check if there have been more than 10 downloads of any type of filing in any given minute from a specific Internet address. If so, I remove all requests from that Internet address for the day.[6][7] Much like the first heuristic, the purpose of this check is to determine if an automated script is accessing the EDGAR website. These heuristics reduce the initial sample of 3,822,564,344 requests to 508,619,475, a reduction of approximately 87%. The magnitude of this reduction is consistent with that of prior studies using the EDGAR server logs after they have been subjected to a similar set of heuristics (Lee et al. 2013).

For the remaining sample of requests, I use Internet address to location lookup tables to find the latitudinal and longitudinal locations of the Internet addresses associated with each of the requests. Specifically, these tables map blocks of the Internet address space to specific latitude and longitude locations. The Internet address to location tables are indicated as being approximately 99% accurate at the country level. At the US level, most

---

[6] If users are behind a static network address translation (NAT) box then this heuristic may overly remove Internet addresses which are associated with multiple distinct individuals. Static NAT has 1 IP address and hence all users behind a static NAT box will appear in the EDGAR server logs as having the same IP address. Static NAT is typically used for small networks while dynamic NAT is used for more complex networks. I don't believe that this will systematically affect my sample selection.

[7] The limit of 10 was chosen based upon my estimates of the speed of automated requests which appear in the server logs

of the locations are accurate to approximately the center of the postal ZIP code of the request.[8] To verify the accuracy of the locations, I cross-check a sample of locations with those obtained using other Internet address location services. I use monthly historic mappings of Internet addresses to locations as there is approximately a 1% to 5% monthly turnover in the Internet address space each month. The location lookup tables provide location information from 1/1/2005 to 12/31/2012.

3.1.1 Investor Location Sample Characteristics

Table 1 provides a summary of the requests to the EDGAR filing system used in my study. Removing automated downloads as well as requests without location information or a CIK-to-GVKEY match yields a final sample size of 258,001,440 requests between 1/1/2005 and 3/30/2012. Interestingly, there is a clear growth trend in the number requests, from 33,609,976 distinct requests in 2005 to 111,475,245 in 2011.

I examine the locations of the requests to better understand where the variation in investor's locations comes from and to ensure that my sample is consistent with prior research and anecdotal evidence about the locations of investors in the US capital market. Table 2 shows that approximately 77% of all requests originate from within the United States. This finding is not surprising since the EDGAR filings system is primarily for US based companies. It is also consistent with estimates of the percentage of the US-based capital flows in the US capital market (OFII 2013).

[8] Further analysis suggests that certain areas of the US appear to have location information more accurate than ZIP code level. However, this was not officially confirmed by the provider.

The findings in Table 2 also show that the largest proportion of US-based requests comes from the New York City area. This geographic area comprises approximately 9.75% of all requests in my sample. This is consistent with my priors as well as anecdotal evidence supporting New York City as the financial capital of the US. The other cities which comprise the Top 10 cities for total number of requests in my sample include Chicago, Los Angeles, Houston, San Francisco, Washington, Dallas, Boston, Toronto, and Philadelphia. In general, these cities are also known for having active financial centers.

I conduct a borough-level analysis of the requests originating from the New York City area to provide further validation for the idea that requests tend to come from areas with greater financial activity. In particular, I expect most of these requests to come from Manhattan. According to the 2011 employment analysis of the city, approximately 12% of all those employed within Manhattan are associated with the financial services industry. This is in comparison to only 2.8% within the other four boroughs (NYCEDC 2013). If market participants are the primary drivers of the requests to the EDGAR filing system, then I would expect most New York City based requests to originate from within the Manhattan area. Figure 3 illustrates the locations of the requests for firms' filings within New York City. As illustrated, the majority of the requests indeed originate from the Manhattan area.

3.2 Measuring Investor Geographic Dispersion

After determining the locations of a firms' investors, I estimate the geographic dispersion of a firm's investor base using a measure of dispersion based upon the Herfindahl-Hirschman Index (hereafter, Herfindahl Index). The Herfindahl Index is calculated as the sum of squares of the characteristic of interest divided by the square of the sum of that characteristic (Berger and Ofek 1995). The Herfindahl Index was originally created to measure industry competition (Li et al. 2013; Giroud and Mueller 2011). However, it can also be used to measure the concentration of various characteristics. For example, Demsetz and Lehn (1985) adopt the Herfindahl Index to measure the concentration of firm ownership while Ahuja (2000) uses it to measure international patent concentration.

To calculate my Investor Geographic Dispersion Index, I divide the globe into segments of approximately 5 degree shifts in latitude and longitude, or square geographic segments of approximately 700km in diagonal length based on the Haversine formula.[9][10] I choose this distance because it is approximately equal to the size of the average US state. This process yields approximately 18,792 possible unique geographic segments, excluding areas covered by water.[11] Applying this to the requests in my sample, I find that there are 5,528 unique geographic segments from which filing requests originate. My Investor Geographic Dispersion Index is calculated as follows.

$$InvestorGeoDisp_{i,t} = -1 * \left[\frac{\sum_{r}^{GEOREGIONS} uniquerequests_{r,i,t}^{2}}{\left(\sum_{r}^{GEOREGIONS} uniquerequests_{r,i,t}\right)^{2}}\right] \qquad (1)$$

---

[9] 700km diagonal length is approximately equal to the size of Oregon.

[10] The Haversine formula tends to over/under estimate distances between two latitude points depending on their distance from the equator. Since the majority of the requests in my sample originate from within +45 degrees latitude of the equator differences in distances due to this are negligible.

[11] There are approximately 64,800 total possible unique geographic segments.

where *InvestorGeoDisp* is the geographic dispersion of investors measured over the quarter prior to a firm's earnings announcements, *GEOREGIONS* is each of the unique 5x5 degree segments of the globe and *uniquerequests* is the total number of unique requests that originate from the given geographic segment over the prior quarter. I multiply the ratio by -1 so that larger values of *InvestorGeoDisp* can be interpreted intuitively as reflecting greater investor geographic dispersion. *i* and *t* are firm and year-quarter, respectively.

Figure 4 illustrates differences in the dispersion of the locations of the requests for firms' filings between firms with high and low investor geographic dispersion. The figure is constructed by first sorting firms in my sample in 2011 by *InvestorGeoDisp.*[12] The locations of the request are then plotted for firms in the top 10% and in the bottom 10% of *InvestorGeoDisp,* respectively Figure 4a and Figure 4b. The figures show that the locations of the requests originate from a greater number of unique locations for the high investors geographic dispersion sample than the low sample.

[12] I restricted this to 2011 due to limitations on the mapping software. 2011 was chosen because it is the last full year of data in my sample.

CHAPTER 4

Market Variables, Control Variables, and Sample Descriptive Statistics

4.1 Market Variables

In this section, I define the market variables used in my study. I calculate abnormal measures of bid-ask spread, depth, illiquidity, and trading volume as the difference between the measures during the event window and a non-event window. I use abnormal measures for each of the variables to remove any potential firm specific cross-sectional variation. I follow prior research and use the trading days T-1 to T+1, where T is the day of the quarterly earnings announcement as the event period, (Bamber 1987; Asthana and Balsam 2001). The non-event window is defined as trading days T-54 to T-5. For my tests of abnormal drift or reversals in price, I measure abnormal price drift or reversal in relation to a value-weighted measure of market performance.

4.1.1 Abnormal Bid-Ask Spread

Prior studies suggest that the bid-ask spread is a reflection of adverse selection risk (Lee et al. 1993; Lee and Ready 1991). That is, specialists may protect themselves from trading at an information disadvantage through price, which is reflected in the size of the bid-ask spread. In particular, larger bid-ask spreads imply that market participants perceive greater information asymmetry and hence are less willing to transact at a low cost.

To calculate abnormal bid-ask spreads around quarterly earnings announcements for the firms in my sample, I use the bid and offer prices from the Trade and Quotes (TAQ) database. For each day in the event period, I calculate the average difference between the offer and bid prices for a firm's stock quotes, scaled by the average of the offer and bid price for the quote. I scale the spread because prior studies suggest that the size of the spread can be mechanically related to the magnitude of a firm's stock price (Garfinklel 2009). I then calculate abnormal bid-ask spread as the difference between the average bid-ask spread during the event period and the average over the non-event period, scaling by the standard deviation of the non-event bid-ask spread to normalize its distribution.

4.1.2 Abnormal Market Depth

Specialists may also attempt to protect themselves from risk by offering a lower quantity of shares at a given price (Lee et al. 1993), thereby impacting market depth. Since market depth is related to the order size needed to move market price, a deeper market suggests a lower perceived risk of transacting at an information disadvantage.

I use trading information obtained from the TAQ database to calculate abnormal market depth. I multiply the average bid size by the bid price plus the offer size multiplied by the offer price for all quotes of a firm's stock during a particular day. I calculate abnormal depth around earnings announcements by subtracting the average trading depth during the event period from the average trading depth during the non-event period, scaling by the standard deviation of the daily trading depth during the non-event period to normalize the distribution.

4.1.3 Abnormal Illiquidity

Illiquidity is used in prior studies to capture the relation between order flow and price movement (Amihud 2002). Kyle (1985) suggests that market makers use order flow to control for risk. Thus, greater illiquidity can be interpreted as an indication of greater perceived risk or greater information asymmetry.

I use the method described in Amihud (2002) to calculate illiquidity around quarterly earnings announcements. I first calculate the absolute price change per dollar of daily trading volume. The ratio is averaged over the event period around a firms' quarterly earnings announcement. I then estimate the average change per dollar in daily trading volume during the non-event windows and subtract this average from the average over the event window, scaling by the standard deviation of the ratio over the non-event window to normalize the distribution.

4.1.4 Abnormal Trading Volume

In addition to differences in perceptions of risk, investors may differ in their valuations of a firm thereby leading to increased trading. To measure abnormal trading volume, I follow a similar approach to that outlined in Asthana and Balsam (2001). Specifically, I calculate abnormal volume, *AVolume*, as the average daily trading volume over the event period minus the average daily trading volume for the firm over the non-event period, using daily trading volume information obtained from CRSP, scaling by the standard deviation of the volume over the non-event period. I truncate *AVolume* at the 99% level to mitigate the impact of a small subset of extreme outliers.

4.1.5 Abnormal Price Drift or Reversal

I estimate the informativeness of price around quarterly earnings announcements by calculating the difference between the event window price reaction and the future long window abnormal returns. To assess the initial price movement, I make an assumption about the correct price reaction to the earnings announcement. Prior studies suggest that under (over) reactions to news are corrected over time as realizations of the expectations based upon the news event are impounded into price or as the information itself is slowly incorporated by the market. However, since the timeframe for this correction process is open to question, I use the market price after several windows following the earnings announcement to proxy for what the correct price reaction should have been during the event window. Specifically, I use the firm market price at 5, 10, 15, 20, and 25 days after the announcement as my respective proxies for the correct price.

After identifying a correct price, I calculate the percentage difference between the reaction during the event window and the reaction over each of five long window returns. Specifically, I subtract the abnormal returns in the three day window around the earnings announcement from the abnormal returns starting one day before the earnings announcement and ending on each of the five long window return days. I then divide this number by the abnormal returns over the long window and take the absolute value of the result. I use value-weighted abnormal returns to account for changes in a firm's stock price caused by market-wide movements. Finally, I take the log of the percentage to reduce the size of the distribution:

$$DriftRev_{i,t} = log\left(\left|\frac{(AbnormalReturn_{i,t-1,t+T} - AbnormalReturn_{i,t-1,t+1})}{AbnormalReturn_{i,t-1,t+T}}\right|\right) \tag{2}$$

where *DriftRev* is the abnormal drift or reversal in price following an earnings announcement. *AbnormalReturn* is calculated as the compounded returns of a firm's stock minus the value-weighted compounded returns of the market over the same window. *T* is the long window time period and is equal to 5, 10, 15, 20, or 25 trading days following the earnings announcement. *i* and *t* are firm and year-quarter, respectively.

4.2 Control Variables

The controls that I include in my analysis capture firm and information environment characteristics that prior studies have found to be associated with the price discovery process. Below is a brief discussion of the motivation for each control; detailed definitions of each variable can be found in Appendix A.

The first characteristic that I control for is investor attention surrounding an earnings announcement. To measure investor attention, I use three distinct proxies (Dellavigna & Pollet 2008; Hirshleifer et al. 2009; Drake et al. 2012c): 1) a dummy for whether the earnings announcement was released on a Friday (*Friday)*, 2) the total number of earnings announcements released on the same day (*NumEA*), and 3) the total number of requests for a firm's filings around the earnings announcement (*Log_TotalRequests(-1,1)*). I include the number of requests for the firm's filings during the quarter (*Log_TotalRequests_Pre)* to proxy for overall investor attention on the firm.

In addition to investor attention, I include several firm-specific controls in my analysis. First, I include the log of both the market value of the firm's equity (*Log_MarketValue*) and the firm's book-to-market ratio (*Log_BooktoMarket)* to proxy for firm size and growth, respectively (Fama and French 1992). To proxy for firm performance, I use the absolute value of the abnormal returns of the firm in the pre-earnings announcement period (*Abs_Abn_Return_Pre*) (Solomon and Soltes 2011). I also include the absolute value of ROA (*Abs_ROA*) and the absolute value of the earnings surprise (*Abs_Median_Miss)* to control for any performance-related information disclosed during the announcement.

I include the percentage of institutional holdings (*Pcnt_InstitHoldings)* and the log of the number of earnings forecasts (*Log_NumAnalysts*) to control for the effect of shareholder sophistication and information intermediaries on the price discovery process (Drake et al. 2012). I include a dummy variable which is set to 1 if the total number of unique states and countries mentioned in the firm’s 10-K filing is above the median (*FirmGeoDist)* to proxy for the geographic distribution of information about the firms.

4.3 Sample Descriptive Statistics

I combine the investor geographic dispersion data with quarterly accounting information from Compustat Fundamentals Quarterly, information about analysts and their estimates from Institutional Brokers' Estimate System (I/B/E/S), market information from CRSP and Trade and Quote (TAQ), and information about the distribution of the firm’s business from firms’ annual 10-K filings. The sample, after requiring the necessary data, comprises approximately 74,200 firm quarter observations between 1/1/2005 and 3/31/2012 and includes 5,695 unique firms. Descriptive statistics for the sample are provided in Table 3. The table shows that the mean (median) firm in my sample has a market value of $694.36 million ($659.84 million), book-to-market value of 0.63 (0.71), percentage of institutional investment of 55.4% (62.2%), and analyst following of 4.1 (4.00).[13]

Table 4 presents the Pearson (below the diagonal) and Spearman (above the diagonal) correlations between the main variables in my study. The correlations show that investor geographic dispersion (*InvestorGeoDisp)* is highly and positively correlated with the amount of attention on the firm (*Log_TotalRequets_Pre*) and with the amount of attention around the earnings announcement (*Log_TotalRequets(-1,1)*). It is also positively correlated with abnormal bid-ask spread around the earnings announcement (*ABid-Ask(-1,1)*), the absolute value of ROA (*abs_ROA*), the absolute value of performance during the quarter (*Abs_Abn_Return_Pre*), the percentage of institutional investment in the firm (*Pcnt_InstitHoldings*), and analyst following (*Log_NumAnalysts*). It is negatively correlated with abnormal illiquidity (*AIlliquidity(-1,1)*), price drift or reversal (*DriftRev5*), whether the announcement occurred on a Friday (*Friday*), the number of concurrent earnings announcements released (*NumEA*), the absolute value of the earnings miss (*Abs_Median_Miss*), and the size of the firm (*Log_MarketValue*).

[13] The mean (median) analyst following in my sample is large since I require at least one analyst forecast to control for the markets expectation of earnings.

CHAPTER 5

Research Design and Findings

5.1 Information Asymmetry around Earnings Announcements

I use the following OLS regression model to examine whether investor geographic dispersion is associated with greater information asymmetry around a firms quarterly earnings announcement:

$$IA_{i,t} = \beta_0 + \beta_1 InvestorGeoDisp_{i,t} + \beta_i \Sigma Controls_{i,t} + IndFE_{i,t} + YearFE_t + MonthFE_t + \epsilon \quad (3)$$

where *IA* represents the three measures of information asymmetry: abnormal bid-ask spreads, abnormal trading depth, and abnormal liquidity around quarterly earnings announcements. Each abnormal measure is calculated as its average during the event period minus its average during the non-event period divided by its standard deviation during the non-event period. *Controls* include firm size, firm growth, institutional investment, absolute performance during the quarter, analyst following, overall investor attention, the geographic distribution of the firm, investor attention toward earnings announcements, the absolute value of the earnings miss, and the absolute value of ROA. *i* and *t* are firm and year-quarter, respectively. I include industry-fixed effects to control for industry-specific idiosyncratic differences in geographic dispersion. I include fixed effects for year and month to control for idiosyncratic time effects. Standard errors are clustered by firm and report date.

The results of model (3) for abnormal bid-ask spreads are shown in Table 5 Columns 2 and 3. As predicted, the coefficient estimate on *InvestorGeoDisp* is positive and significant at the 1% level. The magnitude of the estimated coefficient of 1.4318 suggests that a one standard deviation increase in *InvestorGeoDisp* of 0.036 leads to a change in abnormal *ABid-Ask(-1,1)* of 0.0515. The magnitude of this effect is 3.82% of the standard deviation of *ABid-Ask(-1,1)* (1.349) and 0.10 times the magnitude of the mean (0.499).

The findings in Table 5 Columns 4 through 7 suggest that the price impact of the average trade is higher when there is greater investor geographic dispersion and support the prediction that investor dispersion is associated with greater information asymmetry within the market during price discovery. Specifically, Table 5 Column 5 shows the results for the regression of abnormal depth around quarterly earnings announcements on investor geographic dispersion and controls. The coefficient estimate on *InvestorGeoDisp* is negative and statistically significant at the 1% level. The estimated coefficient of -0.3270 on *InvestorGeoDisp* suggests that a one standard deviation increase in *InvestorGeoDisp* leads to a -0.0118 difference in *ADepth(-1,1).* This is approximately 1.16% of the standard deviation of *ADepth(-1,1)* (1.011) and 5.89 times the magnitude of the mean (-0.002). Next, Table 5 Column 7 presents results for the regression of abnormal illiquidity around quarterly earnings announcements (*AIlliquidity(-1,1)*) on investor geographic dispersion (*InvestorGeoDisp)* and controls. The coefficient estimate on *InvestorGeoDisp* is 0.2374 and is statistically significant at the 5% level. A one standard deviation increase in *InvestorGeoDisp* leads to a 0.0085 increase in *AIlliquidity(-1,1)* which is approximately 1.20% of the standard deviation in *AIlliquidy(-1,1)* (0.713) and 1.42 times the magnitude of the mean of *AIlliquidity(-1,1)*(-0.003).

In general, the coefficient estimates on the control variables for investor attention on earnings announcements (*Log_TotalRequests*, *Friday, and NumEA*) suggest that greater attention is associated with lower information asymmetry. This is consistent with greater attention contributing to a more efficient price discovery (Drake et al. 2014). The coefficients on the proxies for absolute performance (*Abs_Median_miss and Abs_ROA*) suggest that absolute performance is positively associated with information asymmetry potentially due to its impact on investor's prior beliefs. Firm size (*Log_MarketValue*) and book-to-market (log_BooktoMarket) are both positively related to information asymmetry. Finally, analyst following (*Log_NumAnalysts*), institutional investment (*Pcnt_InstitHoldings*), and attention during the pre-period (*Log_TotalRequests_Pre*) provide mixed results as to their relation with information asymmetry around quarterly earnings announcements.

5.2 Trading around Earnings Announcements

To examine whether investor geographic dispersion is associated with greater trading around a quarterly earnings announcement, I use the following OLS regression model:

$$AVolume_{i,t} = \beta_0 + \beta_1 InvestorGeoDiv_{\ i,t} + \beta_i \Sigma Controls_{i,t} + IndFE_{i,t} + YearFE_t + MonthFE_t + \epsilon \quad (4)$$

where *AVolume* is abnormal trading volume around a quarterly earnings announcement, calculated as the average trading volume of the firm during the event period minus the average trading volume of the firm during the non-event period scaled by the standard deviation of the trading volume during the non-event period. *Controls* include the controls listed in Chapter 4.2, the absolute value of ROA, and the absolute value of the earnings surprise. *i* and *t* are firm and year-quarter, respectively. Industry, year, and month fixed effects are also included in the model. Standard errors are clustered by firm and report date.

The results of model (4) are shown in Table 6. The coefficient estimate on *InvestorGeoDisp* is positive and statistically significant at the 1% level suggesting that greater investor geographic dispersion is associated with higher trading volume. This finding is consistent with my prediction that greater investor geographic dispersion leads to greater trading during the price discovery process. The coefficient estimate of 0.5278 suggests that a one standard deviation increase in *InvestorGeoDisp* leads to an increase in *AVolume(-1,1)* of 0.0190. This is 2.08% of the standard deviation in *AVolume(-1,1)* (0.914) and 1.636 times the magnitude of the mean (-0.055).

Consistent with the findings in prior studies, the controls for greater attention on the earnings announcements (*Log_TotalRequests(-1,1), Friday, and NumEA*) are negatively associated with abnormal trading volume around the earnings announcements (Drake et al. 2012). Total attention on the firm during the pre-period (*Log_TotalRequests_Pre*) is positively associated with total trading volume around earnings announcements. I find a positive relation between the percentage of institutional holdings in the firm and trading volume (Knayazeva et al. 2013). The positive coefficient on firm size (*Log_MarketValue*) and the negative coefficient on book-to-market (*Log_BookToMarket*) suggest that larger firms and growth firms exhibit greater trading within the market around quarterly earnings announcements.

5.3 Informativeness of Price

I examine the relation between investor dispersion and the informativeness of price around a firm's quarterly earnings announcement using the following OLS regression:

$$DriftRev_{i,t} = \beta_0 + \beta_1 InvestorGeoDisp_{i,t} + \beta_i \Sigma Controls_{i,t} + IndFE_{i,t} + YearFE_t + MonthFE_t + \epsilon \quad (5)$$

where *DriftRev* is the drift or reversal in price following an earnings announcement calculated over the long window periods of 5, 10, 15, 20, and 25 trading days after the announcement. Details about the calculation of *DriftRev* can be found in Chapter 4.1.5. *Controls* include the controls listed in Chapter 4.2, the absolute value of ROA, and the absolute value of the earnings surprise. i and t are firm and year-quarter, respectively. Industry, year, and month fixed effects are also included. Standard errors are clustered by firm and report date.

Table 7 presents the result of model (5) for each of the long window periods of 5, 10, 15, 20, and 25 trading days following the earnings announcement. The findings suggest that the price reaction during the event period around quarterly earnings announcements is closer to the long window reaction for firms with a greater geographic dispersion of investors. For each of the long window periods of 5, 10, 15, and 25 trading days I find that the estimated coefficient on *InvestorGeoDisp* is negative and statistically associated with *DriftRev* at either the 1% or 5% level for all five windows. The test of the difference between the coefficient estimates for *InvestorGeoDisp* for the five day window (-0.4188) and the fifteen day window (-0.5927) yields a t-stat of -0.69 and a p-value of 0.487 therefore I cannot reject the null that the coefficient estimates are equal at conventional levels of significance. A test of the equivalence of the coefficient estimates on *InvestorGeoDisp* between the five day window and the twenty five window (-0.2815) yields a t-stat of 0.68 and a p-value of 0.495. Again, I cannot reject the null that the coefficient estimates are equal at conventional levels.

I measure the economic significance of a one standard deviation change *InvestorGeoDisp* using the average of the coefficient estimates across each of the five long windows (*DriftRev5 through DriftRev25*). The average coefficient estimate on *InvestorGeoDisp* across the five windows is -0.396 which suggests that a one standard deviation change in *InvestorGeoDisp* leads to a change in price drift or reversal of -1.415%.

Consistent with prior research, I find that the amount of attention on the earnings announcements (*Log_TotalRequests(-1,1), Friday, and NumEA*) is negatively related to drift or reversals in price (Drake et al. 2012c). The coefficient on the magnitude of the surprise is negatively related to the amount of drift. This is consistent with big surprises leading to greater attention and therefore more efficient incorporation of information. The coefficient estimate on the absolute value of ROA (*Abs_ROA*) is positive suggesting that larger magnitudes of performance are more difficult to interpret and incorporate into price. A greater abnormal change in price during the prior quarter (*Abs_Abn_Return_Pre*) is associated with greater drift or reversals in price, consistent with price momentum (Lee and Swaminathan 2000). Analyst following (*Log_NumAnalysts*) and the percentage of institutional holdings in the firm (*Pcnt_InstitHoldings*) is associated with lower drift or reversals in price. Larger firms (*Log_MarketValue*) and growth (*Log_BooktoMarket*) firms have greater drift or reversal in price suggesting that price formation may be more difficult for these firms.

5.4 Geographic Distribution of Information

Table 8 presents the results for the OLS regression of information asymmetry, trading, and the informativeness of price on investor geographic dispersion, partitioned by firms with high and low geographic distribution of information. A firm is categorizes as a high (low) geographic distribution of information firm if the number of unique states and countries mentioned in the 10-K is above (below) the median this tally.[14][15] Overall, the findings are consistent with my prediction that the effect on investor geographic dispersion on the price discovery process around quarterly earnings announcements is greater for firms with high geographic distribution of information.

Table 8 Panel A shows the results for the regression of information asymmetry on investor geographic dispersion and controls, partitioned by firms with high and low geographic distribution of information. As predicted, for high distribution firms I find that *ABid-Ask(-1,1)*, *ADepth(-1,1)*, and *AIlliquidity(-1,1)* are statistically associated with

---

[14] A firm's 10-K provides an overview of the company and thus provides information to estimate the geographic distribution of value relevant information (Addoum et al. 2013)

[15] An alternative measure of the geographic distribution of information about a firm could be industry classification. However, research suggests that industry classifications are at best noisy proxies of a firms business (Bhojraj et al. 2003)

investor geographic dispersion at the 1% level. The sign of the estimated coefficient on *InvestorGeoDisp* for all three regressions is consistent with a positive association between investor geographic dispersion and information asymmetry around quarterly earnings announcements. The differences between the coefficient estimates on *InvestorGeoDisp* between firms with high and low distributions of information for the regressions of *ABid-Ask(-1,1), ADepth(-1,1), and AIlliquidity(-1,1)* on *InvestorsGeoDisp* are 0.8105, -0.3852, and 0.3123 and are significant at the 1%, 10%, and 5% levels, respectively. The magnitudes of the estimated coefficients suggest that a one standard deviation increase in *InvestorGeoDisp* leads to a 0.02917, -0.1386, and 0.0112 change in *ABid-Ask(-1,1), ADepth(-1,1),* and *AIlliquidity(-1,1)* which is equivalent to 2.16%, 1.37%, and 1.57% of the standard deviations and 0.06, 69.3, and 1.87 times the magnitude of each respective measure.

Next, the results in Table 8 Panel B suggest that the effect of investor geographic dispersion on trading around quarterly earnings announcements is greater for firms with high geographic distribution of information. The estimated coefficients on *InvestorGeoDisp* are 0.8033 and 0.3000 and are statically significant at the 1% and 10% level for firms with high and low distributions of information, respectively. The difference between the estimated coefficients of 0.5033 is statically significant at the 1% level. The magnitude of this coefficient suggests that a one standard deviation increase in *InvestorGeoDisp* leads to a 0.0181 increase in *AVolume(-1,1)* which is 1.98% of its standard deviation and 9.15 times the magnitude of its mean.

Lastly, in Table 8 Panel C, I find evidence of a negative and statistically significant association between *InvestorGeoDisp* and *DriftRev* for the firms with a high distribution of information and less evidence of an association for firms with a low distribution of information. For high distribution firms, the estimated coefficient on *InvestorGeoDisp* is negative and statistically significant at either the 1% or 5% levels of significance across all five windows of *DriftRev*. On the other hand, for low distribution firms, I only find a negative and statistically significant association at the 5% level between *InvestorsGeoDisp* and *DriftRev* for the 15 day window. The differences in the estimated coefficients on *InvestorGeoDisp* between high and low distribution of information firms is significant at the 1% and 10% levels for the 5 day and the 10 day windows, respectively. The magnitudes of these differences suggest that a one standard deviation increase in *InvestorGeoDisp* leads to -2.96% and -1.72% changes in *DriftRev5* and *DriftRev10*, respectively.

CHAPTER 6

Robustness

6.1 Placebo Earnings Announcement Dates

One potential concern is that my empirical findings could be driven by overall differences in the market and not necessarily differences in the price discovery of quarterly earnings announcements for firms due to investor dispersion. I explore this possibility by examining the relation between investor dispersion and price discovery around a placebo quarterly earnings announcement date. If my empirical findings are driven by overall differences in the market then I expect to find similar relations between investor dispersion and my proxies for price discovery around placebo quarterly earnings announcement dates.

I specify a placebo quarterly earnings announcement date as approximately 1 month (25 trading days) after the actual quarterly earnings announcement date. I then estimate my measures of information asymmetry, trading, and the informativeness of price around these placebo earnings announcement dates and estimate models (3), (4), and (5) using these new measures. In untabulted results, I find that investor geographic dispersion is not associated with abnormal information asymmetry, trading, or the informativeness of price around placebo earnings announcement dates.[16][17] This is consistent with my hypothesis that the relation between investor dispersion and price discovery is driven by differing interpretations of earnings announcements and not by overall differences in the market among firms due to the geographic dispersion of investors.

6.2 Market Reaction around Earnings Announcements

A potential concern for my tests of the efficient of price formation is that less information may be incorporated into price around quarter earnings announcements for firms with a greater geographic dispersion of investors. This would result in a lower price reaction around earnings announcements and, potentially, lower price drift or reversal. While possible, this would not explain why I find greater information asymmetry and trading around quarterly earnings announcement for firms with a greater geographic dispersion of investors.

To address this concern, I explore the relation between investor geographic dispersion and the absolute value of abnormal returns around quarterly earnings announcements. The results of the regression of the absolute value of abnormal returns around quarterly earnings announcements (*Abs_Abn_Return(-1,1)*) on investor geographic dispersion (*InvestorGeoDisp*) and controls are presented in Table 9. I find that the estimated coefficient on *InvestorGeoDisp* is positive and statistically significant at the 1% level. The magnitude of the coefficient (0.0243) suggests that a one standard deviation

---

[16] I also use 5, 10, 15, and 20 weekdays following the quarter earnings announcement as placebo earnings announcement dates and find similar results.

[17] I still find a positive and statistically relation between investor's geographic dispersion and abnormal bid-ask spread around the placebo quarterly earnings announcement date. However, the magnitude of the relation is economically smaller by approximately 50%.

increase in *InvestorGeoDisp* leads to an increase in the absolute value of the abnormal returns around the earnings announcement of 9 basis points. This suggest that when a firms investors are geographically dispersed more information is incorporated into price around quarterly earnings announcement.

Next, to further address this concern, I examine the relation between the geographic dispersion of investors and the magnitude of new information within the disclosure released during quarterly earnings announcements. I proxy for the magnitude of new information using the absolute value of actual earnings minus the median analyst consensus forecast. I regress my proxy for new information on my measure of investor geographic dispersion and controls. In untabulated results, I find that the geographic dispersion of investors is not related to the magnitude of the earnings surprise.

6.3 Outsourcing Companies using EDGAR

A possible concern with my measure of the geographic dispersion of investors is that individuals who access the EDGAR system from certain countries abroad are not investors but rather are individuals from outsourcing firms. Specifically, these concerns are about outsourcing companies primarily located in India. These companies can be hired by large investors or data aggregators to hand code disclosures into machine readable format. Descriptive evidence provided in Figure 1 and Table 2 is consistent with this concern.[18]

I address this concern by removing requests which originating from India in my sample. I then recalculate my measure of investor geographic dispersion and rerun my main

[18] A representatives from a large US based financial data aggregator confirmed that some of their datasets are collected through operations located in India.

tests using this new measure. In untabulated results, I find that this does not affect my findings.

6.4 American Depository Receipt (ADR) Listings

American Depository Receipt (ADR) listings represent shares in a foreign stock that are traded within the US market. Since the operations of such firms are primarily non-US, this may impact investor dispersion. Moreover, the trading volume for the shares of these firms may be more greatly determined by their shares in their primary market, thereby potentially skewing my tests of trading which rely on trading volume within the US market.

To address this concern I identify 552 potential ADR listed firms within Compustat which exist during my sample period and remove them from my sample. Untabulated findings suggest that this does not materially change my findings.

CHAPTER 7

Conclusion

This study provides novel empirical evidence that the geographic dispersion of investors affects the diversity of information among investors and, by extension, the price discovery process around a firm's earnings announcements. Specifically, I find that investor geographic dispersion is positively associated with information asymmetry in the market and greater trading volume around quarterly earnings announcements. Moreover, I find that these firms are associated with lower price drift or reversal following these announcements. These results are more pronounced for firms with a greater geographic distribution of information.

The findings in this study make two main contributions to the extant literature. First, the findings contribute to the burgeoning literature which explores the impact of location on the information an investor possesses about a firm. Prior research in this area has primarily examined the effects of an investor's location on trading behavior and profits. I contribute to this area of research by providing empirical evidence that investor location dispersion has implications for the price discovery process of a firm. Second, my study contributes to the literature by providing a method for determining investor location from the EDGAR filing system. This insight into investor's locations enhances the descriptive evidence available about investors using the EDGAR system and thus contributes to a fuller understanding of the results in prior research based on investor requests for firm filings to the EDGAR system.

FIGURES

Figure 1 – Locations of Requests for Firms Filings World-Wide 2005 – 2012

Figure 1 presents a summary of the locations of the requests to the EDGAR filing system for the entire sample. The size of each dot corresponds to the relative number of request from that location. A latitude and longitude position is plotted only if there have been more than 10,000 requests from that location over the entire sample. This limitation was imposed due to constraints on the size of the dataset which could be use in the mapping software.

Figure 2 – Locations of Requests for Firms Filings within the United States 2005 – 2012

Figure 2 presents a summary of the locations of the requests to the EDGAR filing system within the lower 48 contiguous U.S. states for the entire sample. The size of each dot corresponds to the relative number of request from that location. A latitude and longitude position is plotted only if there have been more than 10,000 requests from that location over the entire sample. This limitation was imposed due to constraints on the size of the dataset which could be use in the mapping software.

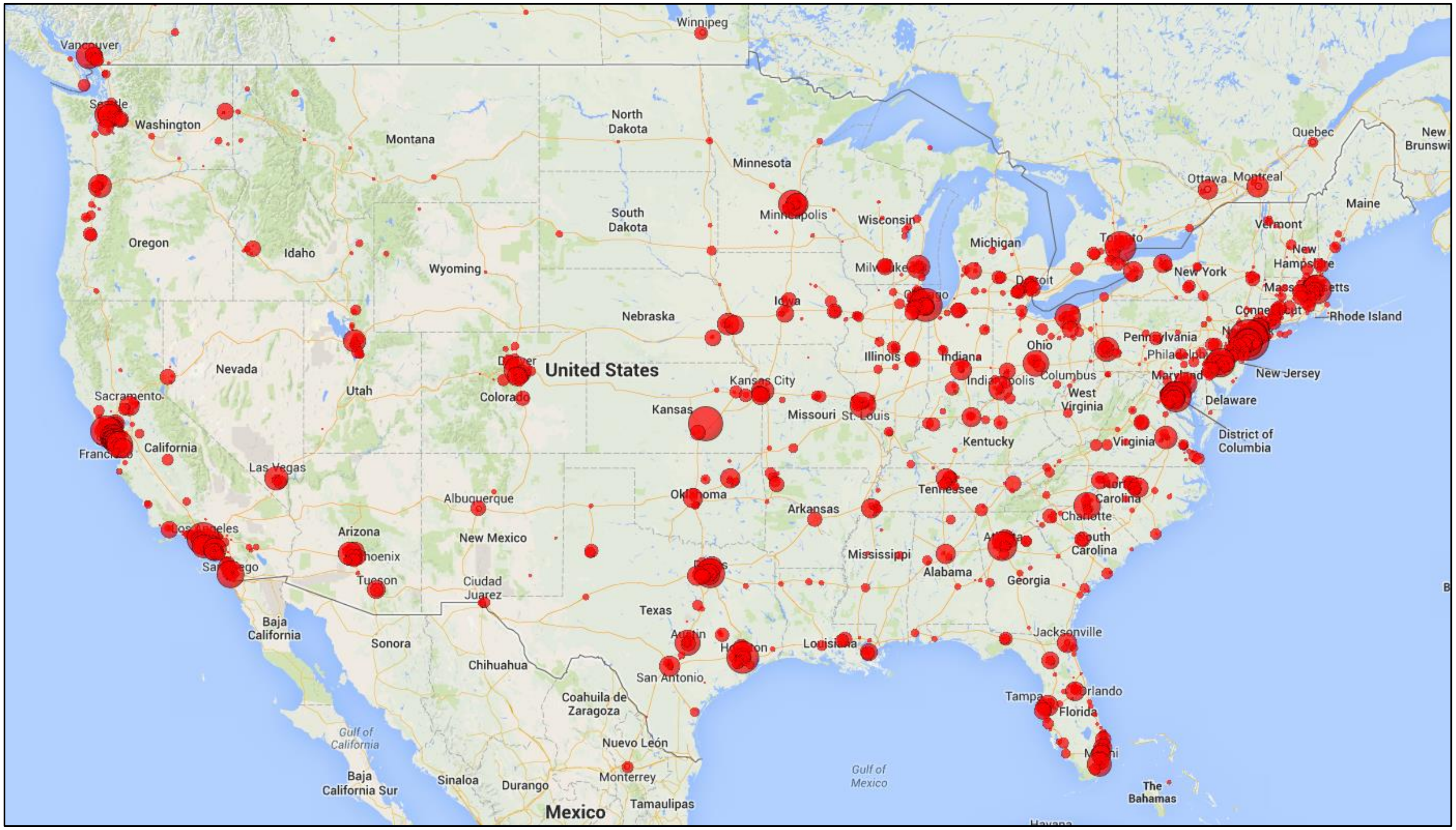
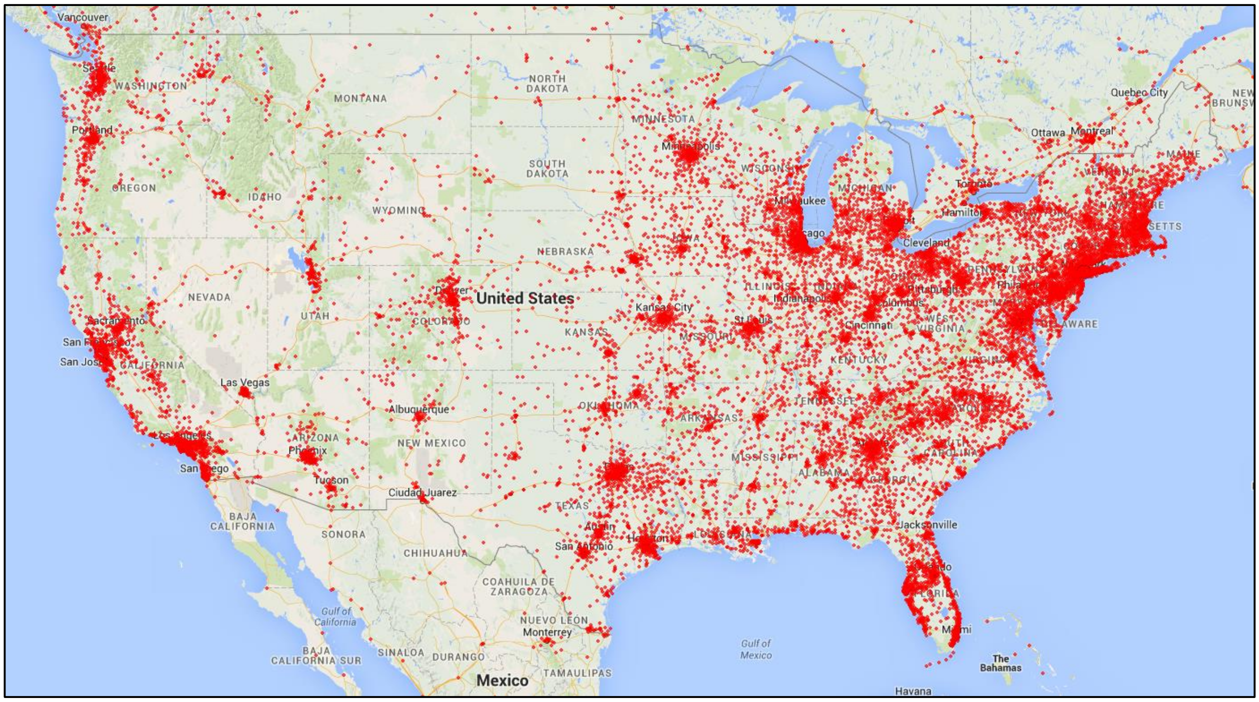


Figure 3 – Locations of Requests for Firms Filings within New York City 2005 - 2012

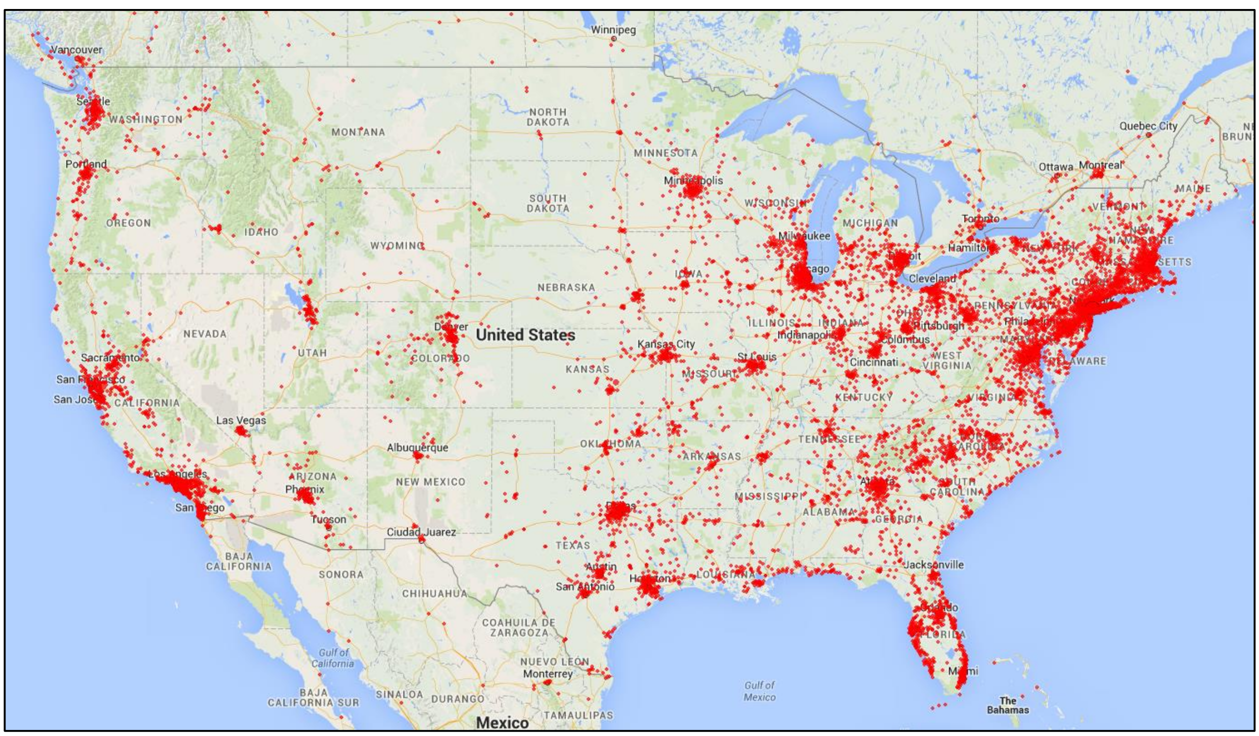
This figure presents the locations of the request for firms' filings within the New York City area. Each of the five distinct boroughs of New York City is marked on the map, Staten Island (lower left), The Bronx (upper right), Manhattan (center), Brooklyn (lower center), and Queens (center left). Due to the size of the data set and the constraints of the mapping software, a latitude and longitude position is plotted only if there have been more than 10,000 requests from that location of the entire sample period. The size of each dot represents the relative number of request which originated from that specific latitude and longitude location.

Figure 4 – High versus Low Investor Geographic Dispersion 2011

Figure 4 presents the locations of the requests for firms filings to the EDGAR filings system for firms with high and low investor geographic dispersion. Panel A presents the locations of the request for firms' filings for firms in the top 10% of investor geographic dispersion. Panel B presents the locations of the request for firms' filings for firms in the bottom 10%.



(a) High Investor Geographic Dispersion



(b) Low Investor Geographic Dispersion

TABLES

Table 1 – Sample of Requests to EDGAR

Table 1 provides a summary of the number of requests to the EDGAR web server used in this study. The main sample of requests spans from 1/1/2005 to 3/31/2012. Panel A provides a summary of the total number of requests and the number of requests removed during each stage of the sample selection processes. *Automated requests* are those which come from automated programs; details about how automated requests were identified can be found in Chapter 3. Panel B presents the number of requests and the number of request with location information and a GVKEY by year.

Panel A - Total Requests

	Requests	Less	% Loss
Total Requests	3,822,564,344		
Less: Automated Requests		3,313,944,869	87%
Requests	508,619,475		
Less: No Location Information		161,366,759	4%
Less: No GVKEY to CIK Match		89,251,276	2%
Requests with Location and GVKEY	258,001,440		

Panel B - Summary of Requests and Requests with Location and GVKEY by Year

Year	Months	Requests	Requests with Location and GVKEY
2005	12	33,609,976	21,484,570
2006	12	31,910,119	17,114,825
2007	12	51,535,524	28,547,522
2008	12	62,837,925	33,622,515
2009	12	93,912,186	45,399,859
2010	12	106,454,684	50,683,731
2011	12	111,475,245	52,518,456
2012	3	16,883,816	8,629,962
Total	87	508,619,475	258,001,440

Table 2 – Top 10 Countries, States, and Cities by Number of Requests to EDGAR 2005 – 2012

Table 2 presents the top 10 locations by number of request to the EDGAR filings system by country, U.S. state, and city. *Requests* is the total number of requests after removing automated downloads and requiring location and GVKEY information. *Percentage* is the total number of requests in the given location divided by the total number of requests word-wide in my sample (n = 258,001,440).

(a) Top 10 Countries

Country	Requests	Percentage
United States	199,136,738	77.18%
India	12,931,296	5.01%
Canada	7,507,438	2.91%
U.K.	5,220,715	2.02%
China	4,940,127	1.91%
Hong Kong	2,869,908	1.11%
Japan	2,449,060	0.95%
Germany	1,728,459	0.67%
Taiwan	1,280,957	0.50%
Singapore	1,216,275	0.47%
Israel	1,163,686	0.45%

(b) Top 10 U.S. States

State	Requests	Percentage
New York	34,053,350	13.20%
California	28,466,284	11.03%
Texas	13,688,249	5.31%
Illinois	10,414,978	4.04%
New Jersey	9,208,464	3.57%
Massachusetts	8,346,189	3.23%
Pennsylvania	7,593,219	2.94%
Florida	7,026,174	2.72%
Ohio	5,858,123	2.27%
Virginia	5,687,727	2.20%
Georgia	4,328,990	1.68%

(c) Top 10 Cities

City	Requests	Percentage
New York City	25,137,610	9.74%
Chicago	5,214,951	2.02%
Los Angeles	4,223,921	1.64%
Houston	3,902,142	1.51%
San Francisco	3,574,423	1.39%
Washington	3,261,591	1.26%
London	2,977,769	1.15%
Dallas	2,751,587	1.07%
Boston	2,495,090	0.97%
Toronto	2,475,453	0.96%
Philadelphia	2,324,689	0.90%

Table 3 - Descriptive Statistics

Table 3 shows the summary statistics for the sample used in this study. The event period is defined as T-1 to T+1 event days around the earnings announcements and the non-event period is the window T-54 to T-5. *InvestorGeoDisp* is the geographic dispersion of investors; details can be found in Chapter 3.2. *ABid-Ask(-1,1)* is abnormal bid-ask spread during the event period calculated as the average difference between the offer and the bid prices of the firm's stock scaled by the average of the offer and bid prices around the earnings announcement minus the average difference during the non-event period. The difference is scaled by the standard deviation over the non-event period. *ADepth(-1,1)* is the abnormal depth during the event period calculated as the bid price multiplied by the bid size plus the offer size multiplied by the offer price during the event period minus the non-event period scaled by the standard deviation over the non-event period. *AIlliquidity(-1,1)* is the Amihud illiquidity of the firm's stock calculated around the firm's earnings announcement; details can be found in Chapter 4. *AVolume(-1,1)* is the abnormal trading volume during the event period calculated as the difference between the average trading volume during the event period and the non-event period scaled by the standard deviation of the trading volume over the non-event period. *DriftRevX* is the log percentage of price drift or reversal following quarterly earnings announcements over the long window X trading days following the quarterly earnings announcement. It is calculated as the abnormal returns from T-1 to T+X minus the abnormal returns during the event period. The difference is then divided by the abnormal return over the long window then logged. *Log_TotalRequests_Pre* is the total number of requests for the firm's filings during the quarter prior to the quarterly earnings announcement. *Log_TotalRequests(-1,1)* is the total number of requests for the firms' filings during the event period. *Friday* is a flag which equals 1 if the earnings announcement takes place on a Friday. *NumEA* is number of earnings announcement on the given earnings announcement date. *Abs_Median_Miss* is the absolute value of the difference between analyst's consensus forecasted earnings and actual earnings. *Abs_ROA* is the absolute value of the firm's return-on-assets. *Abs_Abn_Return_Pre* is the absolute value of the firm's abnormal return during the quarter prior to the earnings announcement. *Pcnt_InstitHoldings* is the percentage of institutional holdings in the firm calculated as the total number of shares held by institutional investors divided by the total number of shares outstanding. *Log_NumAnalysts* is the log of the number of unique analyst forecasts for the firm during the quarter prior to the earnings announcement. *Log_MarketValue* is the log of the price of the firm's stock at the end of the quarter multiplied by the total number of shares outstanding. *Log_BooktoMarket* is the log of total assets divided by the sum of the market value of equity plus total liabilities.

Table 3 - Descriptive Statistics (continued)

Variable	N	Mean	P25	Med	P75	Std. Dev.
Investor Dispersion						
InvestorGeoDisp	74,200	-0.094	-0.110	-0.085	-0.069	0.036
Information Asymmetry						
ABid-Ask(-1,1)	73,990	0.499	-0.394	0.197	1.071	1.349
ADepth(-1,1)	73,990	-0.002	-0.646	-0.168	0.477	1.011
AIlliquidity(-1,1)	74,151	-0.006	-0.493	-0.158	0.325	0.713
Trading						
AVolume (-1,1)	74,184	-0.055	-0.597	-0.243	0.148	0.914
Informativeness of Price						
DriftRev5	74,197	-0.733	-1.524	-0.701	0.051	1.415
DriftRev10	74,197	-0.475	-1.158	-0.433	0.203	1.303
DriftRev15	74,197	-0.375	-1.003	-0.331	0.249	1.256
DriftRev20	74,197	-0.305	-0.884	-0.262	0.266	1.206
DriftRev25	74,197	-0.277	-0.819	-0.230	0.263	1.175
Investor Attention						
Log_TotalRequests_Pre	74,200	6.969	6.504	7.009	7.496	0.717
Log_TotalRequests(-1,1)	74,200	4.123	3.555	4.205	4.779	0.934
Friday	74,200	0.067	0.000	0.000	0.000	0.250
NumEA	74,200	289.750	163.000	272.000	402.000	166.898
Performance						
Abs_Median_Miss	74,200	0.110	0.015	0.040	0.100	0.229
Abs_ROA	74,200	0.027	0.005	0.014	0.029	0.044
Abs_Abn_Return_Pre	74,200	0.127	0.040	0.089	0.170	0.131
Other Characteristics						
Pcnt_InstitHoldings	74,200	0.554	0.267	0.622	0.854	0.339
Log_NumAnalysts	74,200	1.411	0.693	1.386	2.079	0.897
Log_MarketValue	74,200	6.543	5.440	6.492	7.599	1.556
Log_BooktoMarket	74,200	-0.465	-0.760	-0.344	-0.065	0.520

Table 4 - Pearson and Spearman Correlations

Table 4 provides Pearson (Spearman) correlations below (above) the diagonal for the approximately 74,200 observations in my main sample. Correlations at the 1% level are bolded. The event period is defined as T-1 to T+1 event days around the earnings announcements and the non-event period is the window T-54 to T-5. *InvestorGeoDisp* is the geographic dispersion of investors; details can be found in Chapter3.2. *ABid-Ask(-1,1)* is the abnormal bid-ask spread during the event period calculated as the average difference between the offer and the bid prices of the firm's stock scaled by the average of the offer and bid prices around the earnings announcement minus the average difference during the non-event period. The difference is scaled by the standard deviation over the non-event period. *ADepth(-1,1)* is the abnormal depth during the event period calculated as the bid price multiplied by the bid size plus the offer size multiplied by the offer price during the event period minus the non-event period scaled by the standard deviation over the non-event period. *AIlliquidity(-1,1)* is the Amihud illiquidity of the firm's stock calculated around the firm's earnings announcement; details can be found in Chapter 4. *AVolume(-1,1)* is the abnormal trading volume during the event period calculated as the difference between the average trading volume during the event period and the non-event period scaled by the standard deviation of the trading volume over the non-event period. *Log_TotalRequests_Pre* is the total number of requests for the firm's filings during the quarter prior to the quarterly earnings announcement. *DriftRev5* is the log percentage of price drift or reversal following quarterly earnings announcements over the long window X trading days following the quarterly earnings announcement. It is calculated as the abnormal returns from T-1 to T+X minus the abnormal returns during the event period. The difference is then divided by the abnormal return over the long window then logged. *DriftRev10,15,20,*and *25* were not included for brevity; untabulated results suggest that their correlations are similar to *DriftRev5*. *Log_TotalRequests(-1,1)* is the total number of requests for the firms' filings during the event period. *Friday* is a flag which equals 1 if the earnings announcement takes place on a Friday. *NumEA* is number of earnings announcement on the given earnings announcement date. *Abs_Median_Miss* is the absolute value of the difference between the median analyst's consensus forecasted earnings and actual earnings. *Abs_ROA* is the absolute value of the firm's return-on-assets. *Abs_Abn_Return_Pre* is the absolute value of the firm's abnormal return during the quarter prior to the earnings announcement. *Pcnt_InstitHoldings* is the percentage of institutional holdings in the firm calculated as the total number of shares held by institutional investors divided by the total number of shares outstanding. *Log_NumAnalysts* is the log of the number of unique analyst forecasts for the firm during the quarter prior to the earnings announcement. *Log_MarketValue* is the log of the price of the firm's stock at the end of the quarter multiplied by the total number of shares outstanding. *Log_BooktoMarket* is the log of total assets divided by the sum of the market value of equity plus total liabilities.

	Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	*InvestorGeoDisp*		**0.05**	**0.01**	**-0.02**	0.00	**-0.02**	**0.16**	**0.11**	**-0.01**	**-0.05**	**0.02**	**0.06**	0.01	**0.05**	**0.02**	**-0.04**	**0.02**
2	*ABid-Ask(-1,1)*	**0.05**		**-0.19**	**0.20**	**0.10**	**-0.01**	**0.05**	**0.04**	-0.01	**0.05**	**-0.02**	**-0.02**	**-0.06**	**0.11**	**0.10**	**0.12**	**-0.01**
3	*ADepth(-1,1)*	0.01	**-0.17**		**-0.25**	**-0.11**	**-0.05**	**-0.02**	**-0.01**	0.00	**-0.05**	**0.02**	**0.04**	**0.01**	**-0.02**	**-0.06**	**-0.09**	0.00
4	*AIlliquidity(-1,1)*	**-0.01**	**0.18**	**-0.23**		**0.08**	**-0.06**	**-0.04**	**-0.05**	**0.01**	**0.08**	0.00	**-0.03**	**-0.02**	**-0.03**	**-0.03**	0.00	**0.02**
5	*AVolume(-1,1)*	-0.01	**0.15**	**-0.10**	**0.08**		0.00	**-0.02**	**-0.02**	**-0.01**	**-0.07**	**-0.03**	**-0.01**	**-0.05**	**0.04**	**0.03**	**0.07**	**-0.04**
6	*DriftRev5*	**-0.02**	-0.01	**-0.04**	**-0.03**	0.01		**-0.04**	**-0.07**	**0.02**	**0.04**	**-0.02**	**-0.03**	**0.02**	**-0.07**	**-0.07**	**-0.06**	**0.05**
7	*Log_TotalRequests_Pre*	**0.19**	**0.06**	-0.01	**-0.04**	**-0.06**	**-0.03**		**0.72**	**-0.03**	0.00	**0.09**	**0.08**	**0.01**	**0.21**	**0.42**	**0.42**	**-0.04**
8	*Log_TotalRequests(-1,1)*	**0.11**	**0.05**	0.00	**-0.05**	**-0.05**	**-0.06**	**0.71**		**-0.09**	**-0.07**	**0.07**	**0.09**	0.01	**0.21**	**0.35**	**0.36**	**-0.06**
9	*Friday*	**-0.01**	-0.01	0.00	**0.02**	**-0.02**	**0.02**	**-0.04**	**-0.10**		**-0.21**	**0.04**	**-0.03**	0.00	**-0.08**	**-0.07**	**-0.03**	**0.06**
10	*NumEA*	**-0.04**	**0.04**	**-0.04**	**0.08**	**-0.03**	**0.04**	**0.01**	**-0.06**	**-0.19**		**0.01**	**0.03**	**0.05**	**-0.05**	-0.01	**-0.08**	**-0.02**
11	*Abs_Median_Miss*	**-0.02**	**-0.01**	-0.01	**0.03**	**-0.01**	0.00	**0.05**	**0.05**	**0.04**	**0.01**		**0.03**	**0.09**	**-0.05**	**-0.04**	**-0.03**	**0.20**
12	*Abs_ROA*	**0.02**	**-0.04**	**0.01**	**0.02**	-0.01	**0.01**	**0.01**	**0.02**	0.00	**0.07**	**0.13**		**0.10**	0.00	**0.05**	**0.02**	**-0.54**
13	*Abs_Abn_Return_Pre*	**0.02**	**-0.08**	**0.05**	**-0.02**	**-0.04**	**0.03**	**0.04**	**0.03**	0.01	**0.06**	**0.11**	**0.19**		**-0.09**	**-0.08**	**-0.25**	**0.01**
14	*Pcnt_InstitHoldings*	**0.06**	**0.10**	**-0.02**	**-0.05**	**0.02**	**-0.06**	**0.22**	**0.22**	**-0.08**	**-0.07**	**-0.11**	**-0.12**	**-0.12**		**0.36**	**0.34**	**-0.10**
15	*Log_NumAnalysts*	**0.04**	**0.10**	**-0.05**	**-0.05**	**0.02**	**-0.06**	**0.41**	**0.35**	**-0.07**	**-0.03**	**-0.06**	**-0.07**	**-0.09**	**0.35**		**0.59**	**-0.16**
16	*Log_MarketValue*	**-0.03**	**0.12**	**-0.09**	**-0.03**	**0.06**	**-0.05**	**0.40**	**0.35**	**-0.03**	**-0.10**	**-0.08**	**-0.17**	**-0.27**	**0.31**	**0.58**		**-0.27**
17	*Log_BooktoMarket*	0.00	0.00	0.00	**0.03**	**-0.06**	**0.04**	**-0.04**	**-0.06**	**0.05**	**-0.03**	**0.15**	**-0.38**	0.00	**-0.06**	**-0.14**	**-0.23**	

Table 5 – Results of Regressing Information Asymmetry around Earnings Announcements on Investor Geographic Dispersion

Table 5 presents results from the OLS regression of information asymmetry in period around quarterly earnings announcements (T-1 to T+1) on investor geographic dispersion. *InvestorGeoDisp* is the geographic dispersion of investors; details can be found in Section 3.2. Information asymmetry is measured using three proxies: abnormal bid ask spread (*ABid-Ask(-1,1))*, abnormal depth (*ADepth(-1,1))*, and abnormal illiquidity (*Ailliquidity(-1,1))*. Detailed descriptions for each of the variables can be found in Appendix A. P-values are reported in parenthesis below their respective coefficients. All continuous variables are winsorized at 1% and 99% of their respective sample distributions. ***, **, and * denote two-tailed statistical significance at 1%, 5%, and 10% significance levels respectively. Month, year, and industry fixed effects are included. Standard errors are clustered by firm and by report date.

	Abnormal Bid-Ask Spread		**Abnormal Depth**		**Abnormal Illiquidity**	
	ABid-Ask(-1,1) (*p-value*)		*ADepth(-1,1)* (*p-value*)		*AIlliquidity(-1,1)* (*p-value*)	
InvestorGeoDisp	1.3553***	1.4318***	-0.2892**	-0.3270***	0.2094**	0.2374**
	(0.000)	(0.000)	(0.024)	(0.009)	(0.043)	(0.019)
Log_TotalRequests(-1,1)		0.0292***		0.0170**		-0.0195***
		(0.002)		(0.035)		(0.001)
Friday		0.0541		0.0023		0.0513*
		(0.245)		(0.948)		(0.059)
NumEA		0.0003**		-0.0002**		0.0002***
		(0.033)		(0.046)		(0.004)
Abs_Median_Miss		0.0097		0.0106		0.0428***
		(0.684)		(0.552)		(0.003)
Abs_ROA		0.0153		-0.2944***		0.4287***
		(0.913)		(0.005)		(0.000)
Log_TotalRequests_Pre		-0.0356**		-0.0401***		0.0219**
		(0.015)		(0.004)		(0.021)
Abs_Abn_Return_Pre		-0.4250***		0.1844***		-0.0888***
		(0.000)		(0.000)		(0.002)
FirmGeoDist		0.0099		0.0005		-0.0123**
		(0.375)		(0.949)		(0.042)
Pcnt_InstitHoldings		0.2407***		-0.0168		-0.0258***
		(0.000)		(0.283)		(0.009)
Log_NumAnalysts		0.0575***		-0.0120*		-0.0231***
		(0.000)		(0.079)		(0.000)
Log_MarketValue		0.0709***		-0.0468***		0.0025
		(0.000)		(0.000)		(0.470)
Log_BooktoMarket		0.0990***		-0.0634***		0.0453***
		(0.000)		(0.000)		(0.000)
Constant & Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	73,984	73,984	73,984	73,984	74,145	74,145
Adjusted R-Squared	0.050	0.070	0.045	0.053	0.032	0.037

Table 6 – Results of Regressing Trading around Earnings Announcements on Investor Geographic Dispersion

Table 6 presents results from the OLS regression of trading in the period around quarterly earnings announcements (T-1 to T+1) on investor geographic dispersion. *InvestorGeoDisp* is the geographic dispersion of investors; details can be found in Chapter 3.2. *AVolume(-1,1)* is abnormal trading volume calculated as the difference between the average trading volume in the period around quarterly earnings announcement and a non-event period (T-54 to T-5) scaled by the standard deviation of the trading volume over the non-event period. Detailed descriptions for each of the variables can be found in Appendix A. All continuous variables are winsorized at 1% and 99% of their sample distributions. ***, **, and * denote two-tailed statistical significance at 1%, 5%, and 10% significance levels respectively. Month, year, and industry fixed effects are included. Standard errors are clustered by firm and by report date.

	Abnormal Trading Volume		
	AVolume(-1,1) *(p-value)*		
InvestorGeoDisp	0.5589***		0.5278***
	(0.001)		(0.002)
Log_TotalRequests(-1,1)		-0.0624***	-0.0619***
		(0.000)	(0.000)
Friday		-0.0827	-0.0812
		(0.199)	(0.206)
NumEA		0.0000	0.0000
		(0.886)	(0.873)
Abs_Median_Miss		0.0041	0.0039
		(0.833)	(0.841)
Abs_ROA		0.0126	0.0106
		(0.913)	(0.926)
Log_TotalRequests_Pre		0.0689***	0.0679***
		(0.004)	(0.004)
Abs_Abn_Return_Pre		-0.0184	-0.0162
		(0.606)	(0.653)
FirmGeoDist		-0.0116**	-0.0123**
		(0.049)	(0.039)
Pcnt_InstitHoldings		0.0383***	0.0377***
		(0.001)	(0.001)
Log_NumAnalysts		-0.0002	0.0003
		(0.983)	(0.974)
Log_MarketValue		0.0136***	0.0138***
		(0.007)	(0.006)
Log_BooktoMarket		-0.0339***	-0.0329***
		(0.005)	(0.007)
Constant & Fixed Effects	Yes	Yes	Yes
Observations	74,178	74,178	74,178
Adjusted R-Squared	0.164	0.167	0.167

Table 7 – Results of Regressing Informativeness of Price around Earnings Announcements on Investor Geographic Dispersion

Table 7 presents the OLS regression of price drift or reversal following quarterly earnings announcements on investor geographic dispersion. *InvestorGeoDisp* is the geographic dispersion of investors; details can be found in Chapter 3.2. *DriftRevX* is calculated as the absolute value of the difference between the long window abnormal return (T-1 to T+X) minus the abnormal returns in the period around quarterly earnings announcements (T-1 to T+1) divided by the abnormal long window returns. Detailed descriptions for each of the variables can be found in Appendix A. P-values are reported in parenthesis below their respective coefficients. All continuous variables are winsorized at 1% and 99% of their sample distributions. ***, **, and * denote two-tailed statistical significance at 1%, 5%, and 10% significance levels respectively. Month, year, and industry fixed effects are included. Standard errors are clustered by firm and by report date.

	Price Drift or Reversal				
	DriftRev5	*DriftRev10*	*DriftRev15*	*DriftRev20*	*DriftRev25*
	(p-value)	*(p-value)*	*(p-value)*	*(p-value)*	*(p-value)*
InvestorGeoDisp	-0.4095**	-0.4703***	-0.5437***	-0.2780**	-0.2785**
	(0.018)	(0.003)	(0.000)	(0.049)	(0.045)
Log_TotalRequests(-1,1)	-0.0804***	-0.0652***	-0.0581***	-0.0451***	-0.0524***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Friday	0.0353	0.0383*	0.0238	0.0141	0.0095
	(0.133)	(0.067)	(0.246)	(0.457)	(0.618)
NumEA	0.0003***	0.0001***	0.0001***	0.0001*	0.0000
	(0.000)	(0.001)	(0.006)	(0.061)	(0.815)
Abs_Median_Miss	-0.1246***	-0.0731***	-0.0826***	-0.0777***	-0.0899***
	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)
Abs_ROA	0.3492**	0.3566***	0.2794**	0.2487*	0.2810**
	(0.014)	(0.007)	(0.025)	(0.052)	(0.021)
Log_TotalRequests_Pre	0.0632***	0.0422***	0.0395***	0.0175	0.0396***
	(0.000)	(0.001)	(0.001)	(0.124)	(0.000)
Abs_Abn_Return_Pre	0.1696***	0.0995**	0.1130***	0.1262***	0.0689*
	(0.000)	(0.014)	(0.003)	(0.001)	(0.060)
FirmGeoDist	-0.0064	-0.0168	0.0053	-0.0099	0.0074
	(0.602)	(0.120)	(0.608)	(0.326)	(0.458)
Pcnt_InstitHoldings	-0.1072***	-0.0742***	-0.0462***	-0.0488***	-0.0526***
	(0.000)	(0.000)	(0.005)	(0.002)	(0.000)
Log_NumAnalysts	-0.0468***	-0.0345***	-0.0319***	-0.0147**	-0.0136**
	(0.000)	(0.000)	(0.000)	(0.034)	(0.042)
Log_MarketValue	-0.0138**	-0.0035	-0.0021	0.0014	-0.0034
	(0.013)	(0.507)	(0.679)	(0.780)	(0.447)
Log_BooktoMarket	0.0449***	0.0558***	0.0335***	0.0466***	0.0340***
	(0.003)	(0.000)	(0.008)	(0.000)	(0.004)
Constant and Fixed Effects	Yes	Yes	Yes	Yes	Yes
Observations	74,191	74,191	74,191	74,191	74,191
Adjusted R-Squared	0.016	0.011	0.008	0.005	0.005

Table 8 – Results of Regressing Information Asymmetry, Trading, and the Informativeness of Price on Investor Geographic Dispersion, Partitioned by Geographic Distribution of Information

Table 8 presents the effect of investor geographic dispersion (*InvestorGeoDisp)* on information asymmetry, trading, and the informativeness of price, partitioned by the geographic distribution of information about the firm. Firms are partitioned into high and low distribution of information using the number of unique states and countries mentioned in their 10-K filing. Panels A, B, and C present the estimated coefficients on *InvestorGeoDisp* from the regressions of information Asymmetry (*ABid-Ask(-1,1), ADepth(-1,1)*, and *AIlliquidity(-1,1)*), trading (*AVolume(-1,1)),* and the informativeness of price (*DriftRevX)* on investor geographic dispersion (*InvestorGeoDisp)* with controls for firms with high and low geographic distribution of information, respectively. The regressions are run on my main sample which contains approximately 74,200 observations. Detailed descriptions for each of the variables can be found in Appendix A. P-values are reported in parenthesis below their respective coefficients. All continuous variables are winsorized at 1% and 99% of their sample distributions. ***, **, and * denote two-tailed statistical significance at 1%, 5%, and 10% significance levels respectively. Month, year, and industry fixed effects are included. Standard errors are clustered by firm and by report date.

Panel A - Effect of Investor Geographic Dispersion on Information Asymmetry

Information Asymmetry Measure	**High Distribution** *(p-value)*	**Low Distribution** *(p-value)*	**Difference** *(p-value)*
ABid-Ask(-1,1)	1.8866***	1.0759***	0.8105***
	(0.000)	(0.000)	(0.005)
ADepth(-1,1)	-0.5283***	-0.1431	-0.3852*
	(0.004)	(0.351)	(0.082)
AIlliquidity(-1,1)	0.4086***	0.0963	0.3123**
	(0.002)	(0.440)	(0.044)

Panel B - Effect of Investor Geographic Dispersion on Trading

Trading Measure	**High Distribution** *(p-value)*	**Low Distribution** *(p-value)*	**Difference** *(p-value)*
AVolume(-1,1)	0.8033***	0.3000*	0.5033***
	(0.000)	(0.081)	(0.002)

Panel C - Effect of Investor Geographic Dispersion on the Informativeness of Price

Informativeness of Price Measure	**High Distribution** *(p-value)*	**Low Distribution** *(p-value)*	**Difference** *(p-value)*
DriftRev5	-0.8793***	-0.0433	-0.8360***
	(0.000)	(0.846)	(0.007)
DriftRev10	-0.7423***	-0.2612	-0.4811*
	(0.000)	(0.221)	(0.089)
DriftRev15	-0.6848***	-0.4317**	-0.2531
	(0.001)	(0.027)	(0.356)
DriftRev20	-0.4381**	-0.1605	-0.2776
	(0.025)	(0.366)	(0.251)
DriftRev25	-0.4451**	-0.1623	-0.2834
	(0.020)	(0.349)	(0.228)

Table 9 – Results of Regressing Absolute Value of Abnormal Returns around Earnings Announcements on Investor Geographic Dispersion

Table 9 presents results from the regression of the absolute value of returns in the period around quarterly earnings announcements (T-1 to T+1) on investor geographic dispersion. InvestorGeoDisp is the geographic dispersion of investors; details can be found in Chapter 3.2. Abs_Abn_Return(-1,1) is the absolute value of the firm's abnormal return in the period around quarterly earnings announcements. P-values are reported in parenthesis below their respective coefficients. All continuous variables are winsorized at 1% and 99% of their respective sample distributions. ***, **, and * denote two-tailed statistical significance at 1%, 5%, and 10% significance levels respectively. Standard errors are clustered by firm and by report date.

	|Abnormal Returns around Earnings Announcement|		
		Abs_Abn_Return(-1,1) *(p-value)*	
InvestorGeoDisp	0.0231**		0.0243***
	(0.015)		(0.004)
Log_TotalRequests(-1,1)		0.0093***	0.0093***
		(0.000)	(0.000)
Friday		0.0009	0.0010
		(0.522)	(0.492)
NumEA		-0.0000*	-0.0000*
		(0.064)	(0.074)
Abs_Median_Miss		0.0205***	0.0205***
		(0.000)	(0.000)
Abs_ROA		0.0029	0.0028
		(0.763)	(0.770)
Log_TotalRequests_Pre		-0.0003	-0.0004
		(0.585)	(0.539)
Abs_Abn_Return_Pre		0.0483***	0.0484***
		(0.000)	(0.000)
FirmGeoDist		0.0022***	0.0021***
		(0.000)	(0.001)
Pcnt_InstitHoldings		0.0045***	0.0045***
		(0.000)	(0.000)
Log_NumAnalysts		0.0044***	0.0044***
		(0.000)	(0.000)
Log_MarketValue		-0.0099***	-0.0099***
		(0.000)	(0.000)
Log_BooktoMarket		-0.0066***	-0.0065***
		(0.000)	(0.000)
Constant & Fixed Effects	Yes	Yes	Yes
Observations	74,191	74,191	74,191
Adjusted R-Squared	0.080	0.137	0.138

APPENDIX

Appendix A - Variable Definitions

Appendix A provides definitions for each of the main variables used in this study. The *event period* is defined as the three trading days around quarterly earnings announcements (T-1 to T+1). The *non-event period* is defined as trading days T-54 to T-5 prior to quarterly earnings announcements.

Variable	Definition
Geographic Dispersion	
InvestorGeoDisp	Investor geographic dispersion is calculated over the prior quarter as the sum of squares of the number of unique requests in each geographic segment divided by the square of the sum of the number of unique requests. I multiply the ratio by -1 so that larger values can be interpreted intuitively as reflecting greater dispersion. Details can be found in Chapter 3.2.
Information Asymmetry	
ABid-Ask(-1,1)	Abnormal bid-ask spread is calculated as the difference between the average daily bid-ask spread during the event period minus the non-event period. Average daily bid-ask spread is calculated as the average of the difference between the offer and bid prices divided by the average of the offer and bid prices for the quotes over the given period. The difference is normalized by the standard deviation of the measure during the non-event period.
ADepth(-1,1)	Abnormal depth is the average bid size multiplied by the bid price plus the offer size multiplied by the offer price over the quotes of the firm's stock during the event period minus the non-event period. The difference is normalized by the standard deviation of the measure during the non-event period.
AIlliquidity(-1,1)	Abnormal illiquidity is calculated as the absolute value of the daily return divided by the dollar value of trading volume for the day during the event period minus the non-event period. The difference is normalized by the standard deviation of the ratio during the non-event period.
Trading	
AVolume(-1,1)	Abnormal trading volume is the difference between the average trading volume during the event period and the average trading volume during the non-event period. The difference is normalized by the standard deviation of volume over the non-event window.
Informativeness of Price	
DriftRevX	Price drift or reversal is calculated as the absolute value of the percentage difference between the abnormal return during the event period and the abnormal return calculated over a long window period (T-1 to T+X, where X is 5, 10, 15, 20 and 25 trading days).
Investor Attention	
Log_TotalRequests_Pre	Total number of request to the EDGAR filing system for the firm's filings during the quarter prior to the earnings announcement.
Log_TotalRequests(-1,1)	Total number of request to the EDGAR filing system for the firm's filings during the event period around earnings announcements.
Friday	Flag set to 1 if the earnings announcement takes place on a Friday, 0 otherwise.
NumEA	Number of earnings announcements which take place on the same day as the given quarterly earnings announcement.

Appendix A - Variable Definitions (continued)

Variable	Definition
Performance	
Abs_Median_Miss	Absolute value of the quarterly earnings surprise is calculated as the absolute value of the difference between actual earnings minus the median analyst consensus forecast.
Abs_ROA	Absolute value of return-on-assets is calculated as the absolute value of income before extraordinary items divided by total assets.
Abs_Abn_Return	Absolute value of value weighted abnormal returns for the given period is calculated at the absolute value of the compounded returns of the firm minus the value weighted compounded returns of the market over the same window.
Other Characteristics	
Log_NumAnalysts	Quarterly number of analyst forecasts.
Log_MarketValue	Log of the market value of the firm. The market value of the firm is calculated as the total number of shares outstanding multiplied by the share price at the end of the quarter.
Log_BooktoMarket	Log of the firm's book-to-market ratio. The book-to-market ratio is calculated as total assets divided by the market value of equity plus total liabilities.
Pcnt_InstitHoldings	Percent of institutional holdings in the firm is calculated as the total number of shares held by institutional investors divided by the total number of shares outstanding from firms' 13F filings.
FirmGeoDist	Flag set to 1 if the number of unique states plus the number of unique countries mentioned in the company's 10-K filing is above the median for all firms. Countries were restricted to the top 50 countries in terms of economic development according to the United Nations 2012 GDP estimates.

BIBLIOGRAPHY

Addoum, J. M., Kumar, A., & Law, K. (2013). Geographic Diffusion of Information and Stock Returns.

Ahuja, G. (2000). Collaboration Networks, Structural Holes, and Innovation: Longitudinal Study. *Administrative Science Quarterly*, *45*(3), 425–455.

Amihud, Y. (2002). Illiquidity and Stock Returns: Cross-Section and Time-Series Effects. *Journal of Financial Markets*, *5*(1), 31–56.

Asthana, S., & Balsam, S. (2001). The Effect of EDGAR on the Market Reaction to 10-K Filings. *Journal of Accounting and Public Policy*, *20*.

Ball, R., & Brown, P. (1968). An Empirical Evaluation of Accounting Income Numbers. *Journal of Accounting Research*, *6*(2), 159–178.

Bamber, L. S. (1987). Unexpected Earnings, Volume around Firm Size, Quarterly and Trading Earnings Announcements. *The Accounting Review*, *62*(3), 510–532.

Beaver, W. H. (1997). Financial Reporting: An Accounting Revolution (3rd Edition). Prentice Hall

Berger, P. G., & Ofek, E. (1995). Diversification's Effect on Firm Value. *Journal of Financial Economics*, *37*, 39–65.

Bernard, V. L., & Thomas, J. K. (1989). Post-Earnings-Announcement Drift: Delayed Price Response or Risk Premium? *Journal of Accounting Research*, *27*, 1–36.

Beyer, A., Cohen, D. A., Lys, T. Z., & Walther, B. R. (2010). The financial reporting environment: Review of the recent literature. *Journal of Accounting and Economics*, *50*(2-3), 296–343.

Bhojraj, S., Lee, C. M. C., & Oler, D. K. (2003). What's My Line? A Comparison of Industry Classification Schemes for Capital Market Research. *Journal of Accounting Research*, *41*(5), 745–774.

Carter, M. E., & Soo, B. S. (1999). The Relevance of Form 8-K Reports. *Journal of Accounting Research*, *37*(1), 119–132.

Chi, S., & Shanthikumar, D. (2014). *The Geographic Dispersion of Google Search and the Market Reaction to Earnings Announcements*.

Coval, J. D., & Moskowitz, T. J. (2001). The Geography of Investment: Informed Trading and Asset Prices. *Journal of Political Economy*, *109*(4), 811–841.

Demsetz, H., & Lehn, K. (1985). The Structure of Corporate Ownership: Causes and Consequences. *Journal of Political Economy*, *93*(6), 1155–1177.

Dellavigna, S., & Pollet, J. M. (2009). Investor Inattention and Friday Earnings Announcements. *The Journal of Finance*, *LXIV*(2).

Drake, M. S., Roulstone, D. T., & Thornock, J. R. (2012a). The Informativeness of Stale Financial Disclosures.

Drake, M. S., Roulstone, D. T., & Thornock, J. R. (2012b). What Investors Want: Evidence from Investors' Use of the EDGAR Database. *SSRN Electronic Journal*.

Drake, M. S., Roulstone, D. T., & Thornock, J. R. (2012c). Investor Information Demand: Evidence from Google Searches Around Earnings Announcements. *Journal of Accounting Research*, *50*(4), 1001–1040.

Fama, E. F., & French, K. R. (1992). The Cross-Section of Expected Stock Returns. *The Journal of Finance*, *XLVII*(2).

Feng, L., & Seasholes, M. S. (2004). Correlated Trading and Location. *The Journal of Finance*, *LIX*(5), 2117–2144.

Garfinkel, J. A. (2009). Measuring Investors' Opinion Divergence. *Journal of Accounting Research*, *47*(5), 1317–1348.

Giroud, X. (2013). Proximity and Investment: Evidence from Plant-Level Data. *The Quarterly Journal of Economics*, 861–915.

Grinblatt, M., & Keloharju, M. (2001). How Distance, Language, and Culture Influence Stockholdings and Trades. *The Journal of Finance*, *LVI*(3).

Giroud, X., & Mueller, H. M. (2011). Corporate Governance, Product Market Competition, and Equity Prices. *The Journal of Finance*, *LXVI*(2), 563–600.

Goyenko, R. Y., Holden, C. W., & Trzcinka, C. a. (2009). Do liquidity measures measure liquidity? *Journal of Financial Economics*, *92*(2), 153–181.

Grossman, S. (1976). On the Efficiency of Competitive Stock Markets Where Trades Have Diverse Information. *The Journal of Finance*, *31*(2).

Harris, M., & Raviv, A. (1993). Differences of Opinion Make a Horse Race. *Review of Financial Studies*, *6*, 473–506.

Hayek, F. A. (1945). The Use of Knowledge in Society. *American Economic Review*, *35*(4), 519–530.

Hirshleifer, D., Lim, S. S., & Teoh, S. H. (2009). Driven to Distraction: Extraneous Events and Underreaction to Earnings News. *The Journal of Finance*, *LXIV*(5), 2289–2325.

Hong, H., Kubik, J. D., & Stein, J. C. (2005). Thy Neighbor's Portfolio: Word-of-Mouth Effects in the Holdings and Trades of Money Managers. *The Journal of Finance*, *LX*(6).

Hong, L., & Page, S. E. (2001). Problem Solving by Heterogeneous Agents. *Journal of Economic Theory*, *97*(1), 123–163.

Ivkovic, Z., & Weisbenner, S. (2005). Local Does as Local Is : Information Content of the Geography of Individual Investors' Common Stock Investments. *The Journal of Finance*, *LX*(1).

Kandel, E., & Pearson, N. D. (1995). Differential Interpretation of Public Signals and Trade in Speculative Markets, *103*(4), 831–872.

Kim, O., & Verrecchia, R. E. (1994). Market liquidity and volume around earnings announcements. *Journal of Accounting and Economics*, *17*(1-2), 41–67.

Knyazeva, A., Knyazeva, D., & Kostovetsky, L. (2013). *Investor Heterogeneity and Trading Around Earnings Announcements*.

Kothari, S. P. (2001). Capital Market Research in Accounting. *Journal of Accounting and Economics*, *31*, 105–231.

Kyle, A. S. (1985). Continuous Auctions and Insider Trading. *Econometrica*, *53*(6), 1315–1336.

Lee, C. M. C., Mucklow, B., & Ready, M. J. (1993). Spreads, Depths, and the Impact of Earnings Information: An Intraday Analysis. *The Review of Financial Studies*, *6*(2), 345–374.

Lee, C. M. C., & Ready, M. J. (1991). Inferring Trade Direction from Intraday Data. *The Journal of Finance*, *46*(2), 733–746.

Lee, C. M. C., & Swaminathan, B. (2000). Price Momentum and Trading Volume. *The Journal of Finance*, *LV*(5).

Lee, C. M. C., Wang, C. C. Y., & Ma, P. (2013). Crowdsourcing Peer Firms: Evidence from EDGAR Search Traffic.

Li, F., Lundholm, R., & Minnis, M. (2013). A Measure of Competition Based on 10-K Filings. *Journal of Accounting Research*, *51*(2)

Madureira, L., & Underwood, S. (2008). Information, Sell-side Research, and Market Making. *Journal of Financial Economics*, *90*(1), 105–126.

Massa, M., & Simonov, A. (2006). Hedging, Familiarity and Portfolio Choice. *Review of Financial Studies*, *19*(2), 633–685.

Milgrom, P. R., & Stokey, N. (1982). Information, Trade and Common Knowledge. *Journal of Economic Theory*, *26*, 11–21.

New York City Economic Development Corporation (NYCEDC) (2013), Borough Trends & Insights: Analyzing New York City's Local Economies

Organization for International Investment (OFII) (2013). Foreign Direct Investment in the United States 2013 Report

Sloan, R. G., & Rangan, S. (1998). Implications of the Integral Approach to Quarterly Reporting for the Post-Earnings-Announcement Drift. *The Accounting Review*, *73*(3), 353–371.

Solomon, D., & Soltes, E. (2011). *The Determinants of Coverage in the Business Press*.

Watson, W. E., Kumar, K., & Michaelsen, L. K. (1993). Cultural Diversity's Impact on Interaction Process and performance: Comparing Homogeneous and Diverse Task Groups. *Academy of Management Journal*, *36*(3), 590–602.

Verrecchia, R. E. (2001). Essays on Disclosure. *Journal of Accounting and Economics, 32*, 97–180.