

**DO INTERFACES MATTER? A REEXAMINATION OF XBRL USING FINANCIAL
STATEMENT ACQUISITION AND MARKET ACTIVITY**

By

James J. Anderson

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Business Administration – Accounting – Doctor of Philosophy

2022

ABSTRACT

DO INTERFACES MATTER? A REEXAMINATION OF XBRL USING FINANCIAL STATEMENT ACQUISITION AND MARKET ACTIVITY

By

James J. Anderson

Starting in 2009 the eXtensible Business Reporting Language (XBRL) standard was mandated for financial statements by the SEC. The XBRL standard was intended to encourage less-sophisticated trader disclosure processing; however, previous literature has conjectured that the standard primarily aided more-sophisticated traders' disclosure processing. I reexamine the effect of XBRL on more- and less-sophisticated trader disclosure processing by testing whether XBRL influenced their information acquisition and testing whether the proportional relationship between information acquisition and market activity is different for more- and less-sophisticated traders. I find the staggered implementation of XBRL is associated with a 49% (26%) increase in less (more) sophisticated trader information acquisition. Next, I find the proportional relationship between information acquisition and market activity is greater for less-sophisticated traders when compared to more-sophisticated traders. Specifically, I find information acquisition for less-sophisticated traders has a greater proportional relationship with abnormal price movement, abnormal trading volume, and abnormal bid-ask spreads. Together these findings suggest that XBRL did not provide a disproportionate information advantage to more-sophisticated traders, but rather benefited less-sophisticated traders by decreasing their information acquisition costs.

ACKNOWLEDGEMENTS

I am grateful for the support, guidance, and insightful comments of my dissertation committee chair, Chris Hogan and committee members Kenneth Bills, Marilyn Johnson, and K. Ramesh. I appreciate the helpful comments and suggestions from Alon Kalay, Aaron Fritz, Hari Ramasubramanian, Shuting Wu, Sue Yang, Jing Kong, Sangmok Lee, Jennifer Madden, Camille Chippewa, Chris Ittner, and Musaib Ashraf, the doctoral students at Michigan State University and the 2020 AAA Doctoral Consortium. Lastly, I am grateful for the financial support generously provided by the Department of Accounting and Information Systems at Michigan State.

TABLE OF CONTENTS

LIST OF TABLES.....	vi
LIST OF FIGURES	vii
KEY TO ABBREVIATIONS	viii
I. INTRODUCTION.....	1
II. BACKGROUND	8
Disclosure Processing	8
Disparities Between More- and Less-Sophisticated Traders	9
The Reaction of Traders to Information Acquisition Costs	10
XBRL.....	10
Traditional Filings.....	11
Machine Readable Files.....	12
Interactive Viewer.....	13
XBRL and Information Acquisition Costs.....	15
Variations in Trader Sophistication and Size	16
Information Disparities among More-Sophisticated, Less-Sophisticated, and Uninformed Traders	17
Information Integration Costs.....	18
Consequences of the XBRL Standard.....	19
III. HYPOTHESIS DEVELOPMENT	20
10-K Filings and Interactive Viewer.....	20
Trader Sophistication, Market Liquidity, and Information Processing.....	23
IV. RESEARCH DESIGN	26
10-K Filing Acquisition and XBRL.....	26
Acquisition of 10-K Reports and Market Activity.....	28
IV. RESULTS.....	32
Descriptive Statistics.....	32
Tests of H1a and H1b: 10-K Filing Acquisition and XBRL	38
Tests of H2: Acquisition of 10-K Reports and Market Activity	41
Additional Analyses.....	45
Type of Information Acquisition and Market Activity	45
Channels of Information Acquisition.....	48
Interactive Viewer and Machine-Readable Acquisition effect on Market Activity	49
Robustness Test.....	53
Earnings Announcements, Information Acquisition, and Market Activity.....	55
Robustness Tests	58
Internet Service Provider Information Acquisition.....	58
Sensitivity Tests	61

Window Length	61
Short Window Length	62
Long Window Length	66
Conditional Estimation	70
V. CONCLUSION	75
APPENDICES	77
APPENDIX A: VARIABLE DEFINITIONS	78
APPENDIX B: EDGAR LOG	80
BIBLIOGRAPHY	88

LIST OF TABLES

TABLE 1: Summary Statistics.....	34
TABLE 2: Poisson Estimation of Acquisition.....	40
TABLE 3: Market Activity and Information Acquisition	42
TABLE 4: Channel Analysis	46
TABLE 5: Machine-Readable files, Interactive Viewer, and Market Activity	51
TABLE 6: Machine-Readable files and Market Activity	54
TABLE 7: OLS Estimation of Market Activity on Information Acquisition Subsampled by Earnings Announcement Status	56
TABLE 8: OLS Estimation of Market Activity on Information Acquisition for Internet Service Providers	60
TABLE 9: (0,1) Acquisition and Trading Window.....	63
TABLE 10: (0,7) Acquisition and Trading Window.....	67
TABLE 11: Population Limited to Firm-Years with Downloads.....	72

LIST OF FIGURES

FIGURE 1: Trader Disclosure Processing Steps.....	8
FIGURE 2: Example of Traditional 10-K.....	12
FIGURE 3: Example of a Machine Readable File.....	13
FIGURE 4: Example of Interactive Viewer.....	14
FIGURE 5: Information Acquisition Over Time	33
FIGURE 6: Information Acquisition by Type of Trader.....	37
FIGURE A: Variable Definitions.....	78
FIGURE B1: Information Acquisition Manual vs. Automated Definitions	87

KEY TO ABBREVIATIONS

XBRL	extensible Business Reporting Language
SEC	Securities and Exchange Commission
EDGAR	Electronic Data Gathering, Analysis, and Retrieval system
XML	extensible Markup Language
HTML	Hypertext Markup Language
ISP	Internet Service Provider
CIK	Central Index Key
IP	Internet Protocol
IPv4	Internet Protocol Version 4

I. INTRODUCTION

A primary objective of the SEC is to “maintain fair and orderly markets” (SEC 2020) and an associated goal has been to “level the playing field” for all traders (SEC 1998). In 2009, the SEC mandated financial statement reporting under the eXtensible Business Reporting Language (XBRL) standard for the largest firms. The XBRL standard required firms to provide XBRL labeled financial statements which made both Machine-Readable files as well as Interactive Data format (hereafter Interactive Viewer) available for traders. Both Machine-Readable files and Interactive Viewer (hereafter collectively referred to as XBRL components) were expected, in part, to continue the SEC’s efforts to protect smaller less-sophisticated traders (hereafter less-sophisticated traders).¹

XBRL components were expected to lower information acquisition costs, increase financial statement information acquisition, and diminish information barriers that separate less-sophisticated traders from larger more-sophisticated traders (hereafter more-sophisticated traders) with greater financial resources (SEC, 2009).² Prior literature that examines the requirement to provide XBRL components has provided mixed support for the SEC’s conjectures (Blankespoor, Miller, and White 2014; Dong, Li, Lin, and Ni 2016; Kim Li, and Liu 2019; Liu, Wang, and Yao 2014). In this study, I seek to contribute to the literature by examining how

¹ The XBRL standard was mandated in waves based on market capitalization. The firms required to publish XBRL filings in 2009 were firms with the highest market capitalization. Each year after 2009, firms with lower market capitalization were required to submit XBRL filings. According to the XBRL standard §232.405 (SEC 2009) all firms were expected to comply with the standard by 2013, however empirically this paper identifies firms making their first XBRL filing in 2016.

² The SEC provided no clear definition of “small investors” in §232.405 (SEC 2009) thus I follow the conceptual definition in Kalay (2015). I conceptually define less-sophisticated traders as those who devote less time and attention to their investments and are less proficient in analyzing investment-related information. My empirical measure of less-sophisticated traders follows Drake, Johnson, Roulstone, and Thornock (2020), who consider large institutional traders, such as Bank of America, as more-sophisticated traders and less-sophisticated traders as those who are not classified as more-sophisticated. Less-sophisticated traders may include institutional traders who are associated with smaller institutions. Refer to Appendix B for a discussion of the measurement approach.

XBRL components influence information acquisition behaviors of more- and less-sophisticated traders as well as examining the relationship between information acquisition and market activity for more- and less-sophisticated traders.

Blankespoor et al. (2014), in contrast to the SEC's conjectures, find the availability of XBRL components is associated with increased information asymmetry in the market. They conjecture that more-sophisticated traders utilize Machine-Readable files to gain an information advantage which results in additional market level information asymmetry. Empirically they document an association between the availability of XBRL components and an increase in information asymmetry. However, due to data limitations they do not directly observe the acquisition of Machine-Readable files. Thus, it remains an open question whether more-sophisticated traders acquire Machine-Readable files and whether more-sophisticated traders gain an information advantage following the availability of XBRL components by utilizing Machine-Readable files.

An unexplored consequence of the XBRL standard is the introduction of Interactive Viewer, a point and click tool on EDGAR, which likely disproportionately aids less-sophisticated traders' disclosure processing. XBRL components include both Machine-Readable files as well as Interactive Viewer. Interactive Viewer provides all traders access to a point and click interface that can be easily navigated by anyone with rudimentary computer processing skills. In addition, Interactive Viewer allows traders to identify information more quickly within the 10-K filing as compared to utilizing the traditional 10-K format.³ I expect Interactive Viewer makes

³ The traditional 10-K provides all the information required within the 10-K in a linear format with a table of contents, page numbers, etc. This traditional format is provided to traders in a PDF or web accessible format (HTML) providing essentially the same experience of a linear document. In contrast, Interactive Viewer provides an experience more akin to an online textbook where the trader can jump directly to the note of their interest via the user interface.

information acquisition disproportionately easier for less-sophisticated traders due to their lower average level of technical competency. In addition, I expect that XBRL components aided less-sophisticated traders', rather than more-sophisticated traders', disclosure processing due to less-sophisticated traders' utilization of Interactive Viewer.

My empirical analysis includes two parts and utilizes EDGAR log data from 2003 to 2017 to capture information acquisition.⁴ First, I investigate whether the availability of XBRL components increased the number of traders who acquire form 10-K and whether less-sophisticated traders have a greater proportional increase in information acquisition. I predict that XBRL components decrease information acquisition costs for traders resulting in increased information acquisition on EDGAR. Furthermore, I expect that Interactive Viewer decreased information acquisition costs disproportionately for less-sophisticated traders. Following these observations, I predict that less-sophisticated traders have a greater proportional increase in information acquisition following the availability of XBRL components when compared to their more-sophisticated counterparts.

Consistent with this hypothesis, I find the availability of XBRL components is associated with greater trader acquisition of 10-K filings and a disproportionate increase in less-sophisticated trader 10-K acquisition. I take advantage of the staggered implementation of XBRL to allow for a staggered difference-in-difference research design allowing for an examination of how XBRL components influence 10-K acquisition. I find that the number of less (more) sophisticated traders who acquire 10-K filings increases significantly by 49% (26%) following the availability of XBRL components. In addition, I find the increase in less-sophisticated trader information acquisition is proportionally larger than that of more-sophisticated traders. These

⁴ Specifically EDGAR log data extends from January 1st 2003 to June 30th 2017

results suggest XBRL disproportionately decreased information acquisition costs for less-sophisticated traders as compared to more-sophisticated traders likely due to the introduction of Interactive Viewer.

Second, I expect and find that the proportional relationship between information acquisition and market activity is greater for less-sophisticated traders. Prior to the release of a 10-K filing, less-sophisticated traders have a comparative information disadvantage in the market when compared to more-sophisticated traders. When less-sophisticated traders acquire the 10-K filing they have a greater improvement in their information set. Less-sophisticated traders seek to earn information rents from uninformed traders with their information advantage (Miller 2010; Li and Ramesh 2009; Drake, Roulstone, and Thornock 2015; Kim and Verrecchia 1994). Following these observations, I expect and find a greater proportional relationship between less-sophisticated trader information acquisition and the three measures of market activity: the cumulative absolute value of abnormal returns (*CAR ABS*), abnormal trading volume (*Volume*), and abnormal bid-ask spread ($\Delta Spread$).

Third, in additional analysis, I explicitly explore whether more-sophisticated traders gain an information advantage by utilizing Machine-Readable files and seek to better understand how traders utilize different channels of the 10-K.⁵ If more-sophisticated traders gain an information advantage from their acquisition of Machine-Readable files, then I expect them to acquire the Machine-Readable files immediately following the implementation of the XBRL standard. In addition, I expect a positive association between more-sophisticated trader acquisition of Machine-Readable files and market activity as traders capitalize on their information advantage.

⁵ Channels refers to the different file types/versions of the 10-K that traders can utilize. The Machine-Readable files, Interactive Viewer, and other forms of the 10-K all contain the same disclosure but are provided to the trader in separately identifiable files. Within this study I can measure different trader types' acquisition of the 10-K via these different channels.

However, I find very few downloads of Machine-Readable files in the first year of XBRL implementation. Similarly in multivariate analysis, I do not find a positive association between more-sophisticated trader acquisition of Machine-Readable files and measures of market activity: cumulative absolute value of abnormal returns (*CAR ABS*), abnormal trading volume (*Volume*), and abnormal bid-ask spread ($\Delta Spread$). Together these results do not suggest that more-sophisticated traders gain an information advantage from utilizing Machine-Readable files. In addition, I separately test the proportional relationship between more- and less-sophisticated trader acquisition of Machine-Readable files and Interactive Viewer files and market activity. I find that only less-sophisticated trader acquisition of Interactive Viewer has a significant relationship with market activity. Overall, the results do not suggest that more-sophisticated traders gain an information advantage by utilizing Machine-Readable files but rather suggest that less-sophisticated traders gain an information advantage by using Interactive Viewer.

This study makes several contributions to the literature related to XBRL and information acquisition more generally. First this paper finds that the XBRL standard disproportionately encouraged less-sophisticated trader acquisition of 10-K filings. Previous studies have conjectured that less-sophisticated traders could not utilize the data from the XBRL standard (Blankespoor et al. 2014; Bhattacharya et al. 2018), however this study shows that less-sophisticated traders had a larger proportional increase in information acquisition following the implementation of XBRL. This suggests that Interactive Viewer disproportionately decreased information acquisition costs for less-sophisticated traders resulting in a larger proportional increase in less-sophisticated trader information acquisition. Overall, these findings suggest that less-sophisticated traders may have benefited from the XBRL standard by the introduction of Interactive Viewer.

Second this study contributes to the literature by exploring whether more-sophisticated traders gained an information advantage by utilizing the Machine-Readable files made available with the XBRL standard. The descriptive statistics show that download rates of Machine-Readable files immediately following the implementation of the XBRL standard are low when compared to Interactive Viewer downloads. In addition, the multivariate relationship does not show a positive association between Machine-Readable file downloads and the measures of market activity. Together these results do not provide evidence that more-sophisticated traders gain an information advantage from the implementation of XBRL by their exclusive use of Machine-Readable files.

Next, this study contributes to the literature on the informativeness of 10-K filings. Previous literature suggests the informativeness of SEC filings has decreased over time (Easton and Zmijewski 1993). The findings in this study suggest that information processing costs may contribute to the lack of market activity surrounding the 10-K filings. 10-K filings may contain information that can be informative to the market; however, information acquisition costs could have hindered the utilization of the filings in the past.

Finally, this study contributes to the literature by providing an archival test of the results found in experimental studies. Previous studies such as Rennkamp (2012), Nelson and Rupa (2014) and Blankespoor et al. (2019) documented the effects of financial information formatting on decision making processes with experimental research methods. They find how information is provided to traders influences their ability and willingness to utilize the information. The archival evidence of this study supports the generalizability of their findings with large scale empirical evidence concerning how traders' disclosure processing is influenced by the format in which information is presented to traders.

There are limitations to the findings of this study due to the nature of the data utilized. First, the mechanism to differentiate between more- and less-sophisticated traders is imperfect. Second, information acquisition via downloads, specifically Machine-Readable downloads, may not necessarily indicate the information is being processed and utilized at the time of download. Similarly, a more-sophisticated trader download may populate a database that is latter utilized by a less-sophisticated trader. Next, counting the number of traders accessing 10-Ks is imperfect due to ISP addresses and the way EDGAR stores data. Fourth, when a trader accesses the 10-K via multiple channels, it becomes very difficult to attribute the aggregate market effect to a particular channel. In addition, there exists only one market reaction to the release of a 10-K filing with many different individuals acquiring the 10-K via multiple different channels outside of EDGAR. Overall, these limitations require assumptions and certain design choices that may limit the ability to attribute a market effect to Interactive Viewer, Machine-Readable files, or to a specific group of traders.

The paper proceeds as follows. Section II discusses background and hypothesis development, Section III discusses the research design, Section IV presents the sample and results, and Section V concludes.

II. BACKGROUND

Disclosure Processing

Traders experience disclosure processing costs when accessing and utilizing information to inform their trading decisions. Disclosure processing can be broken into three distinct sequential processes as shown in Figure 1: awareness, information acquisition, and price integration (Blankespoor et al. 2019; Blankespoor, deHaan, and Marinovic 2020). Traders sequentially progress through these three steps experiencing implicit and explicit costs at each step to impound information into price. For example, an explicit cost could be fees traders pay to data aggregators while an example of an implicit disclosure processing cost is the time traders spend analyzing a disclosure.

FIGURE 1: Trader Disclosure Processing Steps



Rational traders will not choose to utilize information if disclosure processing costs exceed the expected disclosure processing returns. Grossman and Stiglitz (1980) model the relationship between information acquisition costs and whether information is impounded into price. They find that as information acquisition costs increase, fewer traders are willing to acquire information and drift increases. Similar in outcome, Diamond and Verrecchia (1981) find that diverse interpretations of public information or unique interpretations of public information

can result in drift because not all information and viewpoints are immediately and perfectly impounded into price.

Disparities Between More- and Less-Sophisticated Traders

There exist persistent disparities between different types of traders within the market, however the level of sophistication and information endowments of traders operate on a continuum. To simplify the discussion of these traders, numerous theoretical papers categorize traders into different groups to discuss their incentives and behaviors. For instance, Kyle (1985) categorizes traders into three separate groups: informed traders, market makers, and noise traders. Informed traders process information in disclosures to earn information rents, market makers are entities such as banks who provide liquidity to the market, and noise traders are entities that trade for unspecified reasons such as liquidity.

Different types of traders will strategically choose on which companies to trade and when to trade depending upon their level of sophistication and their information endowments (Kyle 1985). Uninformed traders such as market makers or noise traders avoid paying information rents to informed traders by trading where there is little information asymmetry and high liquidity. By strategically trading, uninformed traders can limit the amount of information rent they pay to informed traders. For instance, if an uninformed trader executes trades immediately following a public disclosure, they would pay information rents to the informed traders because they have not processed the information within the disclosure. Rather, uninformed traders can either trade leading up to an information event or wait until after the information has disseminated into the market so they are less likely to pay information rents.

Predicting informed traders' behavior is less intuitive because they experience conflicting incentives. On one hand, informed traders will earn the highest information rent when they have

the greatest information advantage. On the other hand, informed traders' information advantage generates information asymmetry within the market. Uninformed traders will avoid trading when there is high information asymmetry which reduces liquidity within the market. Reduced market liquidity makes it more difficult for informed traders to exploit their information advantage. If an informed trader seeks information rents in an illiquid market, the rest of the market will extrapolate their information from their trades.

The Reaction of Traders to Information Acquisition Costs

The SEC's implementation of EDGAR was intended to decrease information acquisition costs by making financial statements available online (Asthana and Balsam 2001). Before 1994, companies provided physical filings to the SEC that were made publicly available by physical mailings, at SEC locations, or via third parties. After the SEC-mandated electronic filings on EDGAR, from 1994 to 1997, traders could acquire public filings over the internet. Obtaining filings from the internet is less costly because traders are not required to provide direct payments or experience other indirect costs. Consistent with the expectation that EDGAR decreased information acquisition costs, the literature has found that EDGAR filings are associated with greater short window market reactions to the filings (Asthana and Balsam 2001) and fewer discrepancies between more- and less-sophisticated traders (Asthana et al. 2004).

XBRL

XBRL became a mandated reporting requirement after the acceptance of SEC regulation §232.405 which required firms to submit data for SEC's Next Generation EDGAR system (Interactive Viewer) and XML files for machine processing (Machine-Readable files) in addition

to the traditional files that were required for 10-K filings.⁶ Interactive Viewer and Machine-Readable files require firms to label individual pieces of information within financial statements with a descriptive, unique, and computer readable label. For instance, the fixed assets line item on the balance sheet receives its own unique XBRL label which should be used by all companies that have a fixed assets line item. The introduction of the XBRL requirement only mandates the labeling of information and the submission of additional file formats but does not change the content requirements of the disclosures. Next, I briefly describe each channel of information acquisition available on EDGAR.

Traditional Filings

Traditional files are the incumbent form of the 10-K filing that was available before XBRL and is still available after XBRL. This form of the 10-K filing mimics the paper versions of the 10-K and are created with the intention of a human user. Computers can extract information from the filings using this form of the 10-K filing, however 10-K filings in this form can pose processing difficulties for computers based on their format (Allee and DeAngelis 2015). The amount of information and the reliability of the information extraction from this form of the 10-K filing will vary based on the format that the filer chooses to implement. See Figure 2 for an excerpt from a traditional 10-K filing form.⁷

⁶ Interactive Data files are legally defined in section 232.11. These files are colloquially described as files created under eXtensible Business Reporting Language (XBRL) reporting requirements. Unique to this regulation, all items included within the financial statements including footnotes need to be individually labeled with XBRL tags.

⁷ The inline XBRL standard introduced XBRL tags into this form of the 10-K, however the inline XBRL requirements were not implemented until after the time period explored within this study (SEC 2018).

FIGURE 2: Example of Traditional 10-K

UNITED STATES
SECURITIES AND EXCHANGE COMMISSION
Washington, DC 20549

FORM 10-K

☒ Annual report pursuant to Section 13 or 15(d) of the Securities Exchange Act of 1934
For the fiscal year ended December 31, 2021

or

☐ Transition report pursuant to Section 13 or 15(d) of the Securities Exchange Act of 1934
For the transition period from _____ to _____
Commission file number 1-3950

Ford Motor Company
(Exact name of Registrant as specified in its charter)

Figure 2 provides an excerpt of the Traditional 10-K for Ford Motor Company’s December 31, 2020 fiscal year end.

Machine Readable Files

Machine-Readable files are electronic documents that are created with the expectation that traders or other financial statement users will utilize a tool or application to access the information within the file. For purposes of this study Machine-Readable files includes all files associated with a 10-K filing that end with a “.xml” file extension that are not an Interactive Viewer file. These Machine-Readable files have information labeled within them utilizing the XBRL standard which makes it dramatically easier for a tool or application to successfully locate and acquire information from the filing. Figure 3 contains a brief excerpt from Ford Motor company’s 10-K filing in its Machine-Readable format. This format is difficult for a human user to utilize. However, freely available software packages such as Beautiful Soup in python can quickly parse the information within these files making it easy for a programmer to build

applications. A notable advantage of Machine-Readable files for applications is that all the financial figures and disclosures are contained within one file.

FIGURE 3: Example of a Machine Readable File

```
<xbrl xml:lang="en-US">
<link:schemaRef xlink:href="f-20211231.xsd" xlink:type="simple"/>
<context id="i7e5fe2e4007e4b6ea4035f6fbf03dbe9_D20210101-20211231">
<entity>
<identifier scheme="http://www.sec.gov/CIK">0000037996</identifier>
</entity>
<period>
<startDate>2021-01-01</startDate>
<endDate>2021-12-31</endDate>
</period>
</context>
<context id="i9c205e39a38347029797aec79fcc7713_D20210101-20211231">
<entity>
<identifier scheme="http://www.sec.gov/CIK">0000037996</identifier>
<segment>
<xbrldi:explicitMember
dimension="f:FixedRateDebtSecurityAxis">f:FPRBMember</xbrldi:explicitMember>
</segment>
</entity>
<period>
<startDate>2021-01-01</startDate>
<endDate>2021-12-31</endDate>
</period>
```

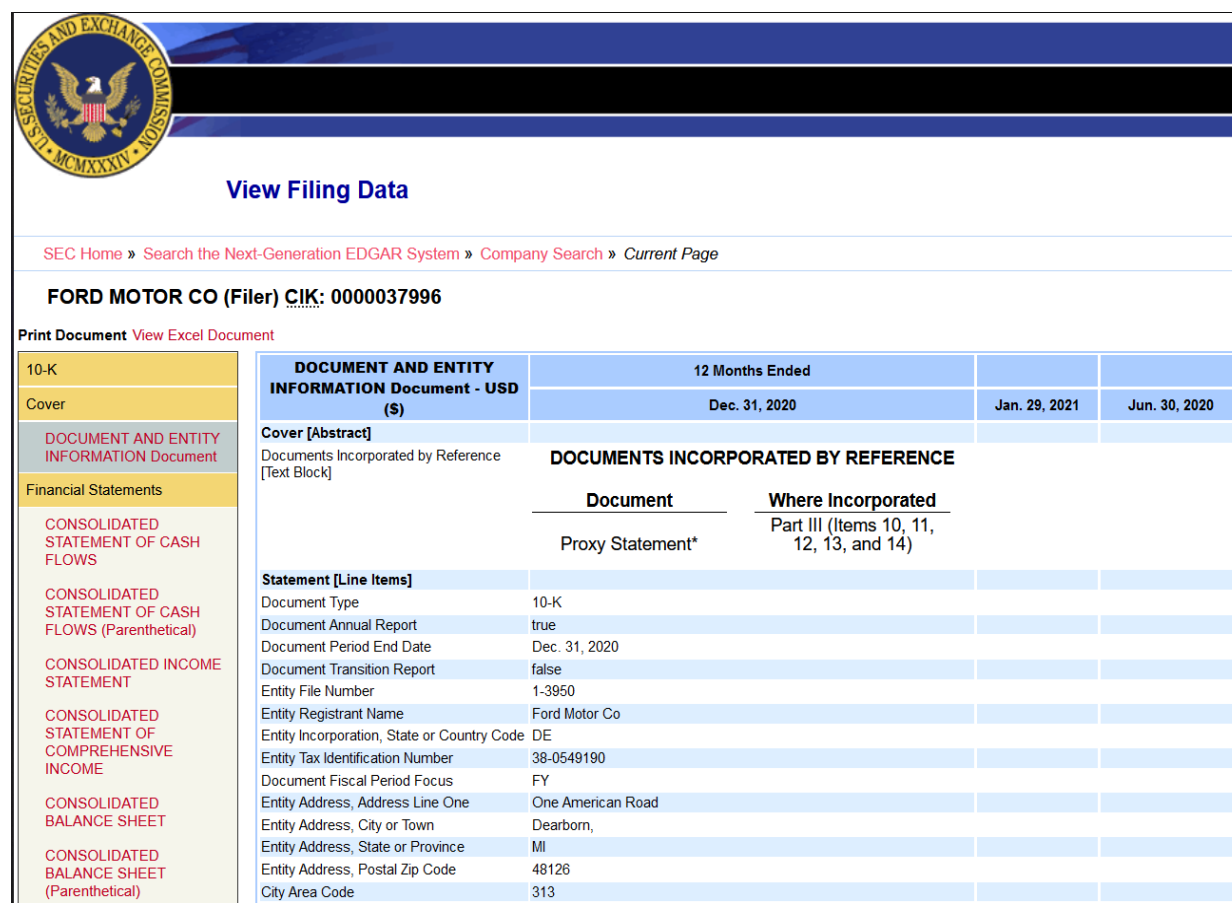
Figure 3 provides an excerpt of the Machine-Readable files for Ford Motor Company's December 31, 2020 fiscal year end 10-K filing.

Interactive Viewer

Interactive Viewer is a point and click tool on EDGAR that allows traders, regardless of technical competencies, to interact with XBRL data. Interactive Viewer allows the trader to quickly navigate to the relevant part of the 10-K filing by using the navigation pane on the left-hand side of the interface as seen in Figure 4. All of the XBRL labels required with the standard are available to the users in the pop-up boxes associated with all of the line items. Given the nature of Interactive Viewer, it is unlikely that traders are using Interactive Viewer to download

information via a computer application. All the information contained within the Interactive Viewer form of the 10-K filing is also available within the Machine-Readable format; however, the Interactive Viewer format will require the application to download hundreds of files to gain the same information that can be acquired via the download of one Machine-Readable file. In addition, Interactive Viewer files contain irrelevant data, such as html code to format the text, which is not useful for application programmers and will increase the resources to download the information.

FIGURE 4: Example of Interactive Viewer



SEC Home » Search the Next-Generation EDGAR System » Company Search » Current Page

FORD MOTOR CO (Filer) CIK: 0000037996

[Print Document](#) [View Excel Document](#)

10-K	DOCUMENT AND ENTITY INFORMATION Document - USD (\$)	12 Months Ended		
		Dec. 31, 2020	Jan. 29, 2021	Jun. 30, 2020
Cover	Cover [Abstract]			
DOCUMENT AND ENTITY INFORMATION Document	Documents Incorporated by Reference [Text Block]	DOCUMENTS INCORPORATED BY REFERENCE		
Financial Statements		Document	Where Incorporated	
CONSOLIDATED STATEMENT OF CASH FLOWS		Proxy Statement*	Part III (Items 10, 11, 12, 13, and 14)	
CONSOLIDATED STATEMENT OF CASH FLOWS (Parenthetical)	Statement [Line Items]			
CONSOLIDATED INCOME STATEMENT	Document Type	10-K		
CONSOLIDATED STATEMENT OF COMPREHENSIVE INCOME	Document Annual Report	true		
CONSOLIDATED BALANCE SHEET	Document Period End Date	Dec. 31, 2020		
CONSOLIDATED BALANCE SHEET (Parenthetical)	Document Transition Report	false		
	Entity File Number	1-3950		
	Entity Registrant Name	Ford Motor Co		
	Entity Incorporation, State or Country Code	DE		
	Entity Tax Identification Number	38-0549190		
	Document Fiscal Period Focus	FY		
	Entity Address, Address Line One	One American Road		
	Entity Address, City or Town	Dearborn,		
	Entity Address, State or Province	MI		
	Entity Address, Postal Zip Code	48126		
	City Area Code	313		

Figure 4 provides a screenshot of Interactive Viewer for Ford Motor Company's December 31, 2020 fiscal year end 10-K filing.

XBRL and Information Acquisition Costs

XBRL reduced information acquisition costs for traders by increasing data standardization, easing data aggregation costs, and improving trader data quality. Before XBRL, data aggregators or traders utilized manual collection and textual analysis techniques to aggregate and standardize information from the traditional form of the 10-K filing. Due to the lack of user interface standardization, data aggregation and standardization was laborious, expensive, slow, and error prone. Typically, data aggregation and standardization were only undertaken by large institutional traders or large data aggregators. After the introduction of XBRL labeling, the costs of standardizing and aggregating information in annual financial statements was dramatically reduced and data quality was improved (Blankespoor et al. 2014). XBRL filings have also led to new data aggregators, such as CalcBench, providing access to aggregated and standardized financial information at a lower cost for users. In addition, XBRL allowed smaller institutional traders to bypass data aggregators and directly acquire information from EDGAR in a timelier and cost-effective manner.

In congruence with previous analytical literature, the SEC expected XBRL would increase the number of traders acquiring annual financial statements. Grossman and Stiglitz (1980) describe a model where information acquisition costs and processing costs deter traders from acquiring and impounding information into price. A basic implementation of this model predicts that as information acquisition costs decrease the number of traders choosing to acquire information increases. In alignment with analytical literature, the SEC conjectured that XBRL would reduce information acquisition costs which would result in more small traders acquiring information (SEC 2009). Given smaller trader's greater sensitivity to information acquisition costs, the SEC anticipated the reduction of information acquisition costs would

disproportionately encourage smaller traders to acquire 10-K filings. In addition, the SEC conjectured the disparity between larger more-sophisticated traders and smaller less-sophisticated traders would diminish as the smaller traders acquired financial information.

Variations in Trader Sophistication and Size

The SEC did not provide a clear definition of “smaller less-sophisticated” traders within §232.405 that mandates XBRL labeling. Arguably, not all traders who could meet the broad definition of smaller less-sophisticated traders will acquire information from 10-K filings. For instance, liquidity traders as described in Kyle (1985) enter the market for their liquidity needs. These liquidity traders are unlikely to acquire information regardless of information acquisition costs and are more likely to strategically time their trades to avoid paying information rents. Next, smaller less-sophisticated traders could encompass small hedge funds, small institutional traders, as well as retail traders. When compared to large financial institutions such as Bank of America or Morgan Stanley, well established pension funds or boutique hedge funds can be considered relatively small. It is unclear if the SEC aimed with the XBRL standard to reduce the information disparity between the largest institutions and slightly smaller institutions or if they aimed to reduce the information disparity between institutions and the relatively unsophisticated traders such as retail traders and liquidity traders.

For ease of discussion, I follow the conceptual definition in Kalay (2015) to define less- and more-sophisticated traders. I conceptually define less-sophisticated traders as those who acquire 10-K filings but devote less time and attention to their investments and are less proficient in analyzing investment-related information. Next, I conceptually define more-sophisticated traders as those who acquire the 10-K filing and have greater resources which allows them to devote more time and attention to their investments. Finally, I define uninformed traders as

traders who do not acquire information from the 10-K filings. Uninformed traders includes liquidity traders and market makers from Kyle (1985), the uninformed traders from Grossman and Stiglitz (1980) or any other trader that enters the market without first acquiring relevant public disclosures.

Information Disparities among More-Sophisticated, Less-Sophisticated, and Uninformed Traders

Information disparities exist between more-sophisticated, less-sophisticated, and uninformed traders within the market. Traders must expend resources to process information within public disclosures. For traders without sufficient size, disclosure processing costs may be greater than the information rents paid to informed traders. With costly disclosure processing, information asymmetry between different groups of traders will persist in the market. When the SEC reduced information acquisition costs via the XBRL standard they reduced information acquisition costs and thus changed the equilibrium between different types of traders.

Decreasing information acquisition costs for all traders may have increased information asymmetry between traders. Decreasing information acquisition costs encourages the marginal trader to acquire information because their expected returns to disclosure processing have become positive. However, uninformed trader's expected returns to disclosure processing are unlikely to become positive given a reduction in disclosure processing costs due to their relative inability to capitalize on an information advantage. Encouraging more traders to become informed introduces more traders into the market looking to earn information rents from the uninformed traders. Reducing information acquisition costs may reduce the information asymmetry between more- and less-sophisticated traders, but at the same time put uninformed

traders at a greater information disadvantage, resulting in net greater information asymmetry between traders in the market.

Information Integration Costs

The information disparity between more- and less-sophisticated traders could be caused by information integration costs rather than information acquisition costs. Information acquisition costs are resources traders expend to gain access to information while information integration costs are the resources traders devote to process the information that they have acquired (Blankespoor et al. 2019). If less-sophisticated traders do not impound information due to information integration costs, then reducing information acquisition costs will not improve the information sets of less-sophisticated traders. If less-sophisticated traders are incapable of integrating information from public disclosures, then reducing information acquisition costs will prove ineffective in reducing the information disparity between more- and less-sophisticated traders.

Previous literature has found evidence consistent with the expectation that information integration costs delay and prohibit information from reflecting in price. You and Zhang (2009) find disclosures with higher information integration costs, as proxied by report length, are associated with higher post earnings announcement drift (PEAD). Similarly, Lee (2012) finds quarterly filings with higher information integration costs, as proxied by length and higher FOG index scores, are associated with higher PEAD. Miller (2010) finds that higher information integration costs, as proxied by report length, decreases abnormal trading volume. Additional cross sectional tests show that decreased abnormal trading volume is driven by fewer smaller traders actively trading on disclosures with higher information integration costs. Blankespoor et al. (2019) experimentally demonstrates that less-sophisticated traders do not impound

information into price when information awareness and acquisition costs are removed. They conclude that information integration costs prevent less-sophisticated traders from impounding information into price. In aggregate, these studies suggest that traders, especially less-sophisticated traders, experience considerable information integration costs which can deter them from impounding information into price.

Consequences of the XBRL Standard

Early empirical literature which directly investigates the consequences of XBRL has not supported all of the SEC's conjectured outcomes of the standard. Blankespoor et al. (2014) finds the implementation of XBRL is associated with higher abnormal bid ask spreads which suggests that XBRL increased information disparity between traders. They conclude that more-sophisticated traders benefited more from the XBRL mandate which caused information disparity between traders. Similarly, Liu et al. (2014) finds that analyst's forecast accuracy and analyst following increase following XBRL implementation. This suggests that more-sophisticated traders benefited from XBRL because more-sophisticated traders and analysts are expected to have similar backgrounds and expertise. Similarly, Bhattacharya et al. (2018) finds that XBRL reduced disparities between large and small institutional traders. Finally, Dong et al. (2016) find that XBRL filings are associated with decreased return synchronicity which suggests that more firm specific information was impounded into price after the release of XBRL filings. In aggregate, these studies suggest that XBRL filings are associated with more information impounded into price, but also that XBRL filings disproportionately assisted institutional traders, increasing the discrepancy between more- and less-sophisticated traders.

III. HYPOTHESIS DEVELOPMENT

10-K Filings and Interactive Viewer

Decreasing information acquisition costs can only have the intended market consequences if the 10-K contains valuable information that is not being utilized due to information acquisition costs. The information in a 10-K filing may not be decision useful for traders or the information contained in 10-K filings may have been previously disseminated to the market by other disclosure channels such as an earnings announcement.⁸ 10-K filings may serve a confirmatory role used to constrain managerial misreporting rather than providing information to the market (Ball 2008). Moreover, information within the 10-K filing can be costly for traders to acquire and integrate into their trading strategies. Acquiring and integrating detailed information from the 10-K filing into a trading decision may be difficult for traders due to the lack of standardized names for financial statement line items and a lack of standardized formatting in the financial statements.⁹ In addition, 10-K filings are long and repetitive (Brown and Tucker 2010), purposefully obfuscate information (Li 2008), and can use specific terminology that requires expertise to interpret (Loughran and McDonald 2011). Traders may not utilize information within the 10-K filing either due to the lack of information content or the difficulty associated with accessing the information.

However, form 10-K is a unique source of consolidated and audited financial information that provides an unrivaled level of detailed and specific disclosures that are not provided to the

⁸ Many companies provide a preliminary earnings announcement before the completion of the 10-K filing (Bronson, Hogan, Johnson, and Ramesh 2011; Li and Ramesh 2009; Schroeder 2016) to provide more timely financial information to the market, but the extent of information included in the earnings announcement varies (Marshall, Schroeder, and Yohn 2019). Easton and Zmijewski (1993) suggest that preliminary earnings announcements communicate financial information and 10-K filings provide little informational value to the market.

⁹ Financial statement line items are any line item that would appear on the balance sheet, income statement, or statement of shareholder's equity. For example, revenue, assets, and contributed capital are all financial statement line items.

market via other channels. In contrast to earnings announcements, 10-K filings are mandated by the SEC and must include full financial statements as well as additional financial information such as the notes to the financial statements.¹⁰ 10-K filings are the only periodic public disclosure that includes both financial statements and disaggregate details related to the information presented within the annual financial statements. For instance, many 10-K filings provide a financial statement note for debt and other financial obligations on the balance sheet. These notes typically contain descriptions of the types of debt the company has and its maturity dates. Previous literature has found that the 10-K filings provide incremental information to the market beyond what is provided in an earnings announcement or by other disclosure channels (Li and Ramesh 2009).

Traders, especially less-sophisticated traders, incur information acquisition costs because accessing information within 10-K filings requires effort to acquire and utilize for comparison between companies and across time. Information processing costs pose a significant hurdle for traders to utilize information in 10-K filings and creates a market for data aggregators, such as Compustat, to compile and standardize the information in 10-K filings. For traders to utilize the information in 10-K filings they must either subscribe to a costly data aggregator or devote time, effort, and expertise to gathering and standardizing data from 10-K filings.

As stated earlier, XBRL components introduced both Machine-Readable files as well as Interactive Viewer. Interactive Viewer is a point and click tool on EDGAR that allows traders to interact with the information within the 10-K filing in a more approachable format. Interactive Viewer allows for a side-by-side comparison of financial statements which makes information

¹⁰ Disclosures provided within the earnings announcement range from a full set of disaggregated financial statements (i.e., income statement, balance sheet, and cash flow statement) to an aggregated income statement (Francis et al. 2002; Chen, Defond and Park 2002; Collins, Li, and Xie 2009; D'Souza, Ramesh, and Shen 2010; Schroeder 2016).

acquisition more feasible for less-sophisticated traders. Notably, XBRL components do not require a change in the content of the disclosure but rather mandate how the information is provided to the SEC (Blankespoor et al., 2014) and by extension the market.

I predict XBRL components reduced information acquisition costs for traders and resulted in more traders choosing to acquire 10-K information (Grossman and Stiglitz 1980). For traders on the margin, decreased information acquisition costs convert the negative expected return of disclosure processing efforts to positive, increasing the number of traders who choose to acquire 10-K filing related information. This leads to the following directional hypothesis.

H1a: Acquisition of 10-K filings is greater following the availability of XBRL components.

Prior literature suggests that less-sophisticated traders are more sensitive to disclosure processing costs (Blankespoor et al. 2020; Allee , Bhattacharya , Black, and Christensen 2007). Consistently, the SEC anticipated the reduction of information acquisition costs from the availability of XBRL components would disproportionately encourage less-sophisticated traders to acquire 10-K filings (SEC 2009) likely due to the introduction of Interactive Viewer. Less-sophisticated traders' information acquisition costs may have decreased disproportionately due to their access to Interactive Viewer. Less-sophisticated traders can benefit disproportionately from Interactive Viewer because unlike their more-sophisticated counterparts they are less likely to have access to private sources of information such as data aggregators. I expect that the availability of XBRL components, specifically Interactive Viewer, disproportionately reduced processing costs for less-sophisticated traders leading to a greater increase in the information

acquisition of less-sophisticated traders relative to more-sophisticated traders. This leads to the following directional hypothesis.

H1b: Acquisition of 10-K filings increases more for less-sophisticated traders, relative to more-sophisticated traders, following the availability of XBRL components.

Trader Sophistication, Market Liquidity, and Information Processing

Blankespoor et al. (2014) find that availability of XBRL components increased information asymmetry in the market. They conjecture that more-sophisticated traders are better able to utilize the Machine-Readable files, providing them with an information advantage. Blankespoor et al. (2014) empirically document an increase in market level information asymmetry following the availability of XBRL components. Their results are consistent with XBRL benefiting the more-sophisticated traders in the market and increasing their relative information advantage resulting in higher levels of information asymmetry in the market (e.g., Kyle, 1985; Kim and Verrecchia, 1994). Similarly, Bhattacharya et al. (2018) investigate the trading activity of only relatively sophisticated traders and suggest that Machine-Readable files helped level the playing field between large and small institutional traders. They argue that non-institutional traders are incapable of capitalizing on the XBRL components because they are unable to utilize Machine-Readable files.

However due to data limitations these studies were unable to determine whether traders gain an information advantage due to Machine-Readable files or due to Interactive Viewer. Since Interactive Viewer became available contemporaneously with Machine-Readable files, less-sophisticated traders may have capitalized on the XBRL standard via Interactive Viewer. The

point and click interface provided by Interactive Viewer allows less-sophisticated traders to utilize XBRL filings thus reducing their information acquisition costs and providing them an information advantage over uninformed traders in the market. In addition, since less-sophisticated traders have an inferior information set prior to the 10-K filing, the XBRL standard may have disproportionately benefited less-sophisticated traders because information acquisition disproportionately improves their information sets.

I expect that less-sophisticated trader information acquisition will have a larger proportional relationship with market activity than information acquisition of more-sophisticated traders. Prior to the 10-K filing, less-sophisticated traders have an inferior information set as compared to that of their more-sophisticated counterparts. When a less-sophisticated trader acquires a 10-K filing, they experience a larger proportional gain in their information set. Given this change in information set, I expect that less-sophisticated traders will trade more aggressively after acquiring information from a 10-K filing as compared to their more-sophisticated counterparts. Following these observations, I expect that less-sophisticated trader information acquisition will have a larger proportional association with market activity when compared to the same proportional relationship for more-sophisticated traders. This leads to the following directional hypothesis.

H2: The association between information acquisition and market activity at the time of the 10-K release is greater for less-sophisticated trader filing acquisition relative to more-sophisticated trader filing acquisition.

In addition to testing H1 and H2, I provide descriptive information and exploratory analyses related to how the market is impacted by information acquisition via more- and less-

sophisticated traders via Interactive Viewer and Machine-Readable files. I do not hypothesize on these relationships due to challenges caused by information acquisition via multiple channels as discussed more fully in the results section.

IV. RESEARCH DESIGN

10-K Filing Acquisition and XBRL

To examine H1a, I utilize the implementation of XBRL and by proxy the availability of Machine-Readable files and Interactive Viewer as an exogenous event which reduced information acquisition costs for traders. The requirement for companies to submit XBRL financial statements was implemented in a phased approach between 2009 and 2014 based on a firm's market capitalization (Blankespoor et al. 2014). When firms were required to comply with the XBRL standard, both Machine-Readable files and Interactive Viewer became available for traders simultaneously. To capture the full effect of XBRL, I define *XBRL* as 1 for all years after the XBRL components became available to the public and 0 for all prior years. If the introduction of Machine-Readable files and/or Interactive Viewer decreased information acquisition costs for traders, then I expect *XBRL* will have a positive and significant association with information acquisition.

$$\begin{aligned} Acquisition_{i,t} = & \gamma_0 + \gamma_1 XBRL_{i,t} + \sum_{k=2}^{12} \gamma_k Controls_{i,t} + Filing\ Day\ Fixed\ Effects \\ & + Firm\ Fixed\ Effects + \epsilon_{i,t} \end{aligned} \quad (1)$$

In the Poisson regression model (1), *Acquisition* is measured separately as the total 10-K acquisition of all traders ($Total^{ACQ}$), as the 10-K acquisition by less-sophisticated traders ($LSop^{ACQ}$), and as the 10-K acquisition by more-sophisticated traders ($MSop^{ACQ}$).¹¹ To develop these

¹¹ Poisson regression models provide consistent estimates for models with non-negative dependent variables. I use a Poisson regression model because acquisition variables are by nature always non-negative. I utilize the non-logged version of the acquisition variables, those with an “ACQ” superscript, because Poisson estimation appropriately

measures of information acquisition, I use the publicly available SEC EDGAR log and specific IP addresses present within the log files to capture how many unique traders access 10-K filings. Total financial statement acquisition ($Total^{ACQ}$) is a proxy for the total number of users accessing information on EDGAR.¹² Following Drake et al. (2020), sophisticated trader financial statement acquisition ($MSop^{ACQ}$) is a count of unique requests for 10-K filings on EDGAR from IP addresses owned by a large financial institution. Less-sophisticated trader financial statement acquisition ($LSop^{ACQ}$) is a count of unique requests for 10-K filings on EDGAR from IP addresses that are not from large financial institutions or non-trading institutions such as universities or public accounting firms. $Total^{ACQ}$, $MSop^{ACQ}$, and $LSop^{ACQ}$ proxy for the number of traders (distinct IP addresses) accessing form 10-K information from the date of the 10-K release to two trading days following the release. Refer to Appendix A for variable definitions and Appendix B for a detailed explanation of how acquisition activity is categorized and measured.

The acquisition variables $Total^{ACQ}$, $MSop^{ACQ}$, and $LSop^{ACQ}$ are based on acquisition of form 10-K through any channel available on EDGAR, including Machine-Readable files, Interactive Viewer, and more traditional means of acquiring the 10-K filing. H1a predicts a positive and significant coefficient on $XBRL$ if a reduction in information acquisition costs encourages traders to acquire information. Similarly, H1b predicts that when $LSop^{ACQ}$ is regressed on $XBRL$ the coefficient on $XBRL$ will be significantly larger than the corresponding

estimates non-negative count variables without transformation. In this setting, a Poisson regression model is preferable to a negative binomial model because negative binomial regression models can provide incorrect statistical inference and may not converge with the inclusion of fixed effects. In contrast, Poisson estimates are fully robust, have no variance assumption, and allow for general serial correlation (Wooldridge 1999). For a broader discussion concerning the specification of Poisson regression refer to Wooldridge (2013, pages: 608-610).

¹² $Total^{ACQ}$ contains financial statement usage which includes non-trading institutions such as universities or public accounting firms. $MSop^{ACQ}$ and $LSop^{ACQ}$ do not include acquisition by academic institutions, auditing firms, law firms, and other entities that are identifiably not securities traders.

coefficient when the model is estimated with $MSop^{ACQ}$ as the dependent variable. The subscripts i, t represent the firm and the 10-K filing date respectively.

A vector of control variables is included to control for other determinants of financial statement acquisition. Report length and tone data from Loughran and McDonald (2011) proxy for the information content of the filings, to help separate the effect of XBRL on information acquisition from the potential effects of changes in information content (Blankespoor 2019). To proxy for information content, I include 10-K filing word count (*Words*), unique word count (*Unique Words*), negative word count (*Negative Words*), positive word count (*Positive Words*), and file size (*File Size*). In addition, I include control variables that proxy for why traders may choose to access 10-K filings. The natural logarithm of assets (*Size*), return on assets (*ROA*), and leverage (*Lev*) control for a trader's incentives to view a company's 10-K filing. Since an earnings announcement release before the 10-K filing date can impact the corresponding market activity (Li and Ramesh 2009), I include indicator variables capturing the status of the earnings announcement at the time of 10-K filing. The variables *FD>EA* and *No EA*, are each binary indicator variables for 10-K filings that are released after the earnings announcement and 10-K filings without an earnings announcement, respectively. Finally, I include firm fixed effects to control for firm-specific incentives for traders to view 10-K filings. I also include fixed effects for the 10-K filing date to control for other market effects on the date of the 10-K filing including information transfers between firms within an industry.¹³

Acquisition of 10-K Reports and Market Activity

To test H2, I investigate the proportional relationship between traders' acquisition of 10-K filings and market activity. I estimate the following OLS regression model.

¹³ Drake et al. (2015), Drake et al. (2016), and Chen and Zhou (2018) study the determinants of trader's acquisition of financial disclosures and Li and Ramesh (2009) document intra-industry information spillovers.

$$\begin{aligned}
\text{Market Activity} = & \beta_0 + \beta_1 MSop_{i,t}^{LN\ ACQ} + \beta_2 LSop_{i,t}^{LN\ ACQ} + \beta_3 XBRL_{i,t} + \sum_{k=4}^{16} \beta_k Controls_{i,t} \\
& + \text{Filing Date Fixed Effects} + \text{Firm Fixed Effects} + \epsilon_{i,t} \quad (2)
\end{aligned}$$

I test the difference between the coefficients on more- and less-sophisticated trader information acquisition ($MSop^{LN\ ACQ}$ and $LSop^{LN\ ACQ}$) to understand which has a greater proportional impact on market activity. I log trader information acquisition ($MSop^{LN\ ACQ}$ and $LSop^{LN\ ACQ}$) so that β_1 and β_2 capture the effect of a percentage increase in information acquisition on market activity, or in other words the proportional relationship between traders' information acquisition and market activity. Comparing β_1 and β_2 captures the difference in the proportional relationship between more- and less-sophisticated trader information acquisition and *Market Activity*.¹⁴ In H2, I predict that $\beta_1 < \beta_2$ which can be interpreted as a firm with a 10% additional $MSop^{ACQ}$ having less market activity than a similar firm with 10% additional $LSop^{ACQ}$, holding all else equal.

I utilize three measures of *Market Activity* to capture whether less-sophisticated traders gain an information advantage from acquiring 10-K filings.¹⁵ The first and second market activity measures are 1) the cumulative absolute value of excess returns based on the value-

¹⁴ Using the logged count is more appropriate than the unlogged count as the independent variable of interest because the logged count can be interpreted as the marginal effect which is ideal for comparing the coefficients. Theoretically we would not expect one download by a less-sophisticated trader to have a comparable effect to one download by a more-sophisticated trader. More-sophisticated traders have a significantly greater capacity to capitalize on their information acquisition due to their greater access to funds. However, I do expect that a greater level of proportional information acquisition by more- and less-sophisticated traders will have a similar effect on market activity if the 10-K is equally informative for each group. For instance, if more- and less-sophisticated traders gain a similar improvement to their information sets following the acquisition of a 10-K filing then a 10% increase in either more- or less-sophisticated traders' information acquisition should result in a similar effect on market activity.

¹⁵ All Market Activity variables are computed from the day of the 10-K filing (day zero) to two trading days after filing (day two).

weighted CRSP market return (*CAR ABS*) and 2) abnormal trading volume (*Volume*). I expect that if less-sophisticated traders benefit more from acquiring information in 10-K filings, then their information acquisition should result in greater proportional price movements (*CAR ABS*) and greater proportional trading volume (*Volume*) relative to that of more-sophisticated traders as evidenced by $\beta_1 < \beta_2$. The third measure of market activity is abnormal bid-ask spread ($\Delta Spread$). When less-sophisticated traders become informed they gain an information set closer to that of more-sophisticated traders. The decrease in information disparity between more- and less-sophisticated traders may reduce total information asymmetry within the market. Following these observations, I would expect that $\beta_2 < 0$ and $\beta_1 > \beta_2$.

On the other hand, if $\beta_2 > \beta_1$ is observed (and $\beta_2 > 0$), this suggests less-sophisticated traders' information acquisition creates aggregate information asymmetry in the market. Less-sophisticated traders within this study includes all traders who are not identifiable as large institutional traders. These less-sophisticated traders are sophisticated enough to download a 10-K filing. After acquiring the 10-K filing these less-sophisticated traders may gain an information advantage over traders who trade without becoming informed regardless of disclosure processing costs (hereafter uninformed traders.)¹⁶ If less-sophisticated traders acquire the 10-K filing to gain an information advantage over uninformed traders, then we would simultaneously expect greater price movement, trading volume, and information asymmetry as a result of their information rent-seeking trades (Kim and Verrecchia 1994).

I utilize a variety of variables to control for the information content present within 10-K filings as well as information available for the firm before the release of the annual financial

¹⁶ By definition uninformed traders do not download information from EDGAR and thus are not easily empirically identifiable. An example of an uninformed trader is a liquidity trader that executes a trade shortly after the 10-K filing. This liquidity trader executes the trade at an information disadvantage due to either their immediate cash needs or a lack of awareness of their information disadvantage.

statements. Similar to prior literature examining the availability of XBRL filings, I control for *XBRL* to identify whether the implementation of XBRL elicits greater market activity (Blankespoor et al. 2014; Liu et al. 2014). In addition, I control for information content within the 10-K filing (*ROA*, *Leverage*, and *Size*) that is made public and may influence trader's willingness to view the annual filings. Similarly, I control for the length of the 10-K (*Word Count*) to control for traders' expectation of information content within the 10-K filing.¹⁷ Next, I include the indicator variables *FD>EA* and *No EA* that capture the status of the earnings announcement at the time of 10-K filing to control for information already released to the market. Finally, this test includes 10-K filing date fixed effects as well as firm fixed effects to control for intra-industry information transfers and firm specific disclosures within the filing.

¹⁷ I use the contemporaneous measure of information content because financial disclosures are very persistent over time (Dyer, Lang, and Stice-Lawrence 2017) and the XBRL reporting standard is associated with greater information content within the annual financial statement (Blankespoor 2019)

IV. RESULTS

Descriptive Statistics

Table 1 Panel A displays an overview of the sample selection process. My sample is comprised of companies that file form 10-K for fiscal years ending after January 1, 2004 and on or before December 31, 2016. I begin the sample in 2004 to avoid data quality issues associated with the first year of EDGAR log files (Loughran and McDonald 2017). The sample ends for firms with year ends on or before December 31, 2016 because the SEC logs were only available through June 30, 2017 at the time of data collection. To generate the sample, I require that a firm has a 10-K filing on SEC EDGAR, is present in Compustat, and has trading activity on CRSP. I begin with 46,673 firm-year observations that are present in all three data sets. I then remove observations where the 10-K filing date occurs before the earnings announcement date, which removes 1,916 firm-years. Then I remove observations without all required control variables which removes 936 additional firm-years. Finally, I remove 1,326 firm-year observations from the analysis that have insufficient cluster size to avoid incorrect statistical inferences (Correia 2015). This leaves a sample of 42,495 firm-year observations from 5,959 unique firms. All continuous variables are winsorized at the 1 percent and 99 percent levels.

Descriptive statistics are provided in Table 1, Panel B. The mean firm has 311 traders acquiring information from the 10-K within 2 trading days after the release of the filing.¹⁸ The average firm has approximately 837 million dollars in assets (*Size* = 6.730). The average (median) return on assets (*ROA*) is 0.015 (0.023) and the average (median) book-to-market ratio (*BTM*) is 0.784 (0.502). The average 10-K has approximately forty-seven thousand words (*Words*

¹⁸ Acquisition variables are measured over a window 0 to 2 days with regards to the release of the 10-K filing. Refer to Appendix B for a detailed discussion of how the acquisition measures are computed.

= 10.761). Table 1 Panel C provides a correlation matrix with Pearson (Spearman) correlations displayed above (below) the diagonal.

Table 1 Panel D provides a descriptive analysis of financial statement acquisition during the year before (year = -1) and the first year XBRL was available (year = 0). The mean (median) number of traders and other users accessing documents ($Total^{ACQ}$) changes from 253 (218) to 317 (295). Figure 5 provides a graphical representation of the total acquisition ($Total^{ACQ}$) for firms that implemented XBRL in 2009 (Treatment) and firms that did not implement XBRL in 2009 (Control) for the year of implementation and the surrounding years.¹⁹ The value of the line represents the average number of file requests on EDGAR for the firm ($Total^{ACQ}$). The divergence of the lines in the Treatment year suggests that information acquisition increased following the implementation of XBRL.

FIGURE 5: Information Acquisition Over Time

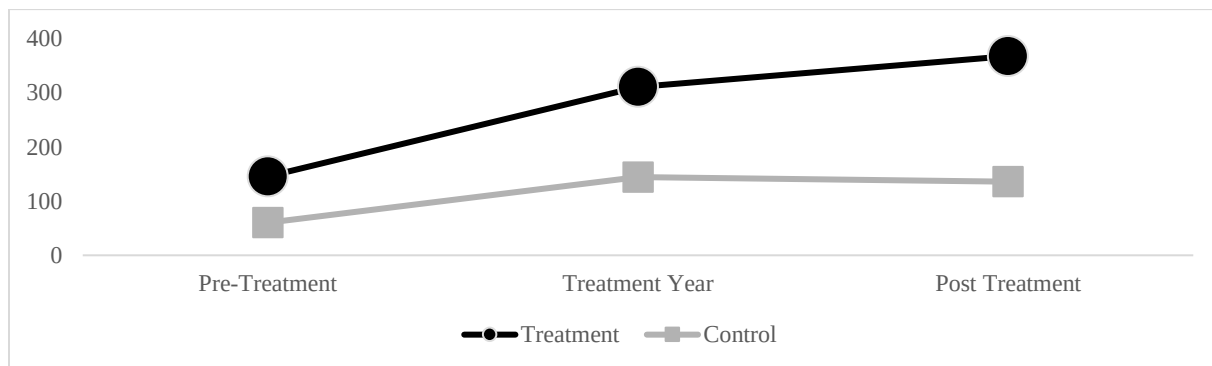


Figure 5 compares a treated group of firms to a control group of firms for acquisition following the availability of Interactive Viewer. The treatment is group is the first group of firms for which Interactive Viewer became available in 2009. The control firms are firms for which Interactive Viewer becomes available in 2011. The vertical axis represents the total acquisition of 10-K filings ($Total^{ACQ}$) for the average firm in for the treatment or control group. Pre-Treatment, Treatment, and Post Treatment years are 2008, 2009, and 2010.

¹⁹ The control group in Figure 2 is firms who have XBRL first available in 2011. The implementation year of 2011 to ensure that the control group is not treated within the graphic.

TABLE 1: Summary Statistics**Panel A: Sample Size**

Observations between 2004-2016	46,673
FD<EA	(1,916)
Missing Variables	(936)
Insufficient Cluster Size	(1,326)
Sample Size	42,495

Panel B: Descriptive Statistics

	Mean	Median	SD	P25	P75
<i>CAR ABS</i>	0.065	0.041	0.088	0.022	0.077
<i>Volume</i>	0.740	0.054	5.063	-0.352	0.867
<i>Δ Spread</i>	0.494	0.018	2.700	-0.685	1.056
<i>Total</i> ^{ACQ}	311.021	209.000	298.940	90.000	475.000
<i>Total</i> ^{LN ACQ}	5.041	5.347	1.655	4.511	6.165
<i>LSop</i> ^{ACQ}	289.557	187.000	285.628	83.000	440.000
<i>LSop</i> ^{LN ACQ}	4.965	5.236	1.640	4.431	6.089
<i>MSop</i> ^{ACQ}	13.962	11.000	14.727	3.000	20.000
<i>MSop</i> ^{LN ACQ}	2.149	2.485	1.193	1.386	3.045
<i>Mach MSop</i> ^{ACQ}	0.560	0.000	1.294	0.000	0.000
<i>Mach MSop</i> ^{LN ACQ}	0.257	0.000	0.525	0.000	0.000
<i>Non Mach MSop</i> ^{ACQ}	20.451	16.000	21.475	4.000	30.000
<i>Non Mach MSop</i> ^{LN ACQ}	2.433	2.833	1.318	1.609	3.434
<i>XBRL</i>	0.582	1.000	0.493	0.000	1.000
<i>Words</i>	10.761	10.726	0.496	10.441	11.042
<i>Unique Words</i>	8.018	8.016	0.207	7.887	8.149
<i>Negative Words</i>	6.672	6.668	0.605	6.293	7.041
<i>Positive Words</i>	5.793	5.787	0.536	5.464	6.116
<i>File Size</i>	15.312	15.158	1.269	14.271	16.419
<i>Lev</i>	0.186	0.110	0.230	0.003	0.295
<i>ROA</i>	0.015	0.023	2.285	-0.007	0.066
<i>BTM</i>	0.784	0.502	35.000	0.278	0.820
<i>Proir Spread</i>	-4.076	-3.301	2.811	-4.885	-2.279
<i>Turn</i>	-0.018	-0.649	6.577	-0.889	-0.229
<i>Volatility</i>	0.471	0.376	0.345	0.255	0.565
<i>Size</i>	6.730	6.715	2.017	5.355	8.033

TABLE 1 (cont'd)

Panel C: Correlations

	(1)	(2)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)	(21)	(22)
(1) <i>CARABS</i>		0.23	0.58	0.06	0.06	0.01	-0.06	-0.03	-0.02	-0.01	0.00	-0.02	-0.09	0.03	-0.03	0.01	-0.43	0.11	0.41	-0.21
(2) <i>Volume</i>	0.32		0.24	0.03	0.03	0.04	0.02	0.02	0.00	0.00	0.00	0.00	0.01	0.01	0.00	0.00	-0.02	0.00	0.01	-0.04
(5) <i>Δ Spread</i>	0.42	0.52		0.05	0.05	0.03	0.00	-0.01	0.00	0.00	0.01	0.00	-0.03	0.05	-0.02	0.00	-0.17	0.03	0.17	-0.09
(6) <i>Total</i> ^{LNACQ}	0.00	0.07	0.04		1.00	0.81	0.32	0.39	0.18	0.21	0.21	0.18	0.58	0.06	0.00	0.00	-0.04	0.00	0.04	0.16
(7) <i>LSop</i> ^{LNACQ}	0.00	0.07	0.04	1.00		0.80	0.32	0.40	0.17	0.21	0.21	0.18	0.58	0.06	0.00	0.00	-0.04	0.00	0.04	0.15
(8) <i>MSop</i> ^{LNACQ}	-0.04	0.09	0.06	0.83	0.81		0.47	0.36	0.24	0.26	0.26	0.23	0.58	0.13	0.01	0.00	0.06	-0.01	-0.06	0.29
(9) <i>Mach MSop</i> ^{LNACQ}	-0.10	0.04	0.04	0.43	0.42	0.54		0.41	0.07	0.09	0.08	0.05	0.45	0.01	0.00	0.00	0.15	-0.01	-0.14	0.12
(10) <i>XBRL</i>	-0.08	0.02	-0.01	0.53	0.54	0.38	0.45		0.11	0.15	0.15	0.11	0.62	0.04	-0.01	0.00	0.12	0.02	-0.11	0.05
(11) <i>Words</i>	-0.06	0.05	0.04	0.24	0.23	0.26	0.09	0.12		0.92	0.93	0.91	0.41	0.22	-0.03	0.00	0.00	0.02	-0.01	0.46
(12) <i>Unique Words</i>	-0.03	0.04	0.02	0.27	0.26	0.28	0.11	0.15	0.92		0.90	0.89	0.40	0.16	-0.03	0.01	-0.03	0.03	0.03	0.37
(13) <i>Negative Words</i>	-0.02	0.04	0.03	0.27	0.26	0.27	0.10	0.16	0.92	0.88		0.87	0.41	0.15	-0.04	0.00	-0.04	0.03	0.04	0.40
(14) <i>Positive Words</i>	-0.05	0.04	0.03	0.23	0.23	0.25	0.07	0.12	0.89	0.88	0.84		0.36	0.16	-0.05	0.00	-0.03	0.02	0.02	0.39
(15) <i>File Size</i>	-0.14	0.04	0.01	0.72	0.72	0.59	0.50	0.61	0.42	0.40	0.41	0.36		0.14	-0.02	0.01	0.20	-0.02	-0.18	0.40
(16) <i>Lev</i>	-0.05	0.06	0.05	0.10	0.10	0.17	0.03	0.04	0.28	0.20	0.18	0.19	0.20		-0.02	0.00	-0.02	0.02	0.01	0.22
(17) <i>ROA</i>	-0.21	0.01	-0.01	0.00	0.00	0.06	0.04	-0.06	-0.14	-0.15	-0.21	-0.14	-0.03	-0.03		0.00	0.05	-0.03	-0.05	-0.01
(18) <i>BTM</i>	0.07	0.01	0.00	-0.02	-0.02	-0.04	0.00	-0.01	0.01	-0.05	0.05	-0.08	0.05	-0.06	-0.23		-0.01	0.00	0.01	0.00
(19) <i>Prior Spread</i>	-0.52	-0.04	-0.03	0.06	0.05	0.12	0.17	0.12	0.05	-0.02	-0.02	0.01	0.24	0.13	0.34	-0.05		-0.22	-0.92	0.35
(20) <i>Turn</i>	0.20	-0.01	0.02	-0.26	-0.26	-0.20	0.01	-0.05	0.01	0.04	0.05	0.05	-0.20	-0.03	-0.26	-0.04	-0.33		0.24	-0.09
(21) <i>Volatility</i>	0.50	0.03	0.02	-0.06	-0.06	-0.12	-0.17	-0.11	-0.06	0.01	0.01	-0.02	-0.22	-0.13	-0.31	0.06	-0.93	0.31		-0.31
(22) <i>Size</i>	-0.31	0.00	-0.02	0.22	0.21	0.31	0.13	0.05	0.45	0.34	0.37	0.35	0.40	0.41	0.19	0.10	0.48	-0.25	-0.43	

TABLE 1 (cont'd)

Panel D: Before and After Interactive Viewer Acquisition

	Before XBRL (year = -1)				After XBRL (year = 0)				Change			
	mean	median	p1	p99	mean	median	p1	p99	mean	median	p1	p99
<i>Total</i> ^{ACQ}	251.852	217.000	103.000	849.000	346.162	331.000	117.000	1100.000	94.310 ***	114.000 ***	14.000	251.000
<i>LSop</i> ^{ACQ}	220.564	188.500	92.000	763.000	313.256	287.000	111.000	1004.000	92.692 ***	98.500 ***	19.000	241.000
<i>MSop</i> ^{ACQ}	19.711	17.000	3.000	62.000	20.488	18.000	1.000	75.000	0.777 ***	1.000 ***	-2.000	13.000
<i>Int LSop</i> ^{ACQ}	0.000	0.000	0.000	0.000	10.809	9.000	0.000	52.000	10.809 ***	9.000 ***	0.000	52.000
<i>Int MSop</i> ^{ACQ}	0.000	0.000	0.000	0.000	0.316	0.000	0.000	4.000	0.316 ***	0.000 ***	0.000	4.000
<i>Mach LSop</i> ^{ACQ}	0.000	0.000	0.000	0.000	35.165	34.000	1.000	67.000	35.165 ***	34.000 ***	1.000	67.000
<i>Mach MSop</i> ^{ACQ}	0.000	0.000	0.000	0.000	1.543	1.000	0.000	6.000	1.543 ***	1.000 ***	0.000	6.000
<i>Total</i> ^{LN ACQ}	5.440	5.384	4.644	6.745	5.698	5.805	4.771	7.004	0.258 ***	0.421 ***	0.126	0.259
<i>LSop</i> ^{LN ACQ}	5.303	5.244	4.533	6.639	5.606	5.663	4.718	6.913	0.303 ***	0.419 ***	0.186	0.274
<i>MSop</i> ^{LN ACQ}	2.880	2.890	1.386	4.143	2.636	2.944	0.693	4.174	-0.245 ***	0.054 ***	-0.693	0.031
<i>Int LSop</i> ^{LN ACQ}	0.000	0.000	0.000	0.000	2.202	2.303	0.000	3.970	2.202 ***	2.303 ***	0.000	3.970
<i>Int MSop</i> ^{LN ACQ}	0.000	0.000	0.000	0.000	0.151	0.000	0.000	1.099	0.151 ***	0.000 ***	0.000	1.099
<i>Mach LSop</i> ^{LN ACQ}	0.000	0.000	0.000	0.000	3.446	3.555	0.693	4.220	3.446 ***	3.555 ***	0.693	4.220
<i>Mach MSop</i> ^{LN ACQ}	0.000	0.000	0.000	0.000	0.680	0.693	0.000	1.792	0.680 ***	0.693 ***	0.000	1.792

Table 1 provides an overview of the sample selection. The sample is based on form 10-K for fiscal years ended after January 1st, 2004 – before December 31st, 2016 for firms with relevant variables in both CRSP and Compustat. Panel A provides sample selection details, Panel B provides summary statistics, Panel C provides a correlation matrix, and Panel D provides summary statistics for firms the year before and the year after the implementation of XBRL. In Panel C, Spearman (Pearson) correlations are above (below) the diagonal. Variables with an “ACQ” superscript are acquisition variables and variables with an “LN ACQ” superscript are acquisition variables that have been logged according to Appendix B. Correlations that are significant at the 5 percent level are in **bold**. All continuous variables are winsorized at the 1 percent and 99 percent levels. Significance at the 10 percent, 5 percent, and 1 percent levels is denoted as *, **, and ***. Variable definitions appear in Appendix A and a detailed discussion of acquisition variable measurement is provided in Appendix B.

Figure 6 provides a graphical representation of the average acquisition of files by trader group. Notably most of the increase in total file acquisition ($Total^{ACQ}$) appears to be driven by less-sophisticated traders acquiring files on EDGAR in the period where XBRL was implemented for more firms. Figure 4 Panels A and B provide the aggregate count of traders who acquire filings relative to the earnings announcement date and 10-K filing date, respectively. Notably, on the release of the earnings announcement in Panel A, a variety of new and old files are acquired via EDGAR. In contrast, in Panel B when a 10-K filing is released, the majority of the traders appear to utilize the new 10-K filing on EDGAR. This figure suggests that when a 10-K is released traders' information acquisition is focused on the new 10-K. In comparison, when earnings announcements are released, information acquisition includes both the new and previous disclosures. This may be due to the limited information presented in an earnings announcement encouraging traders to utilize old 10-K filings for additional information.

FIGURE 6: Information Acquisition by Type of Trader

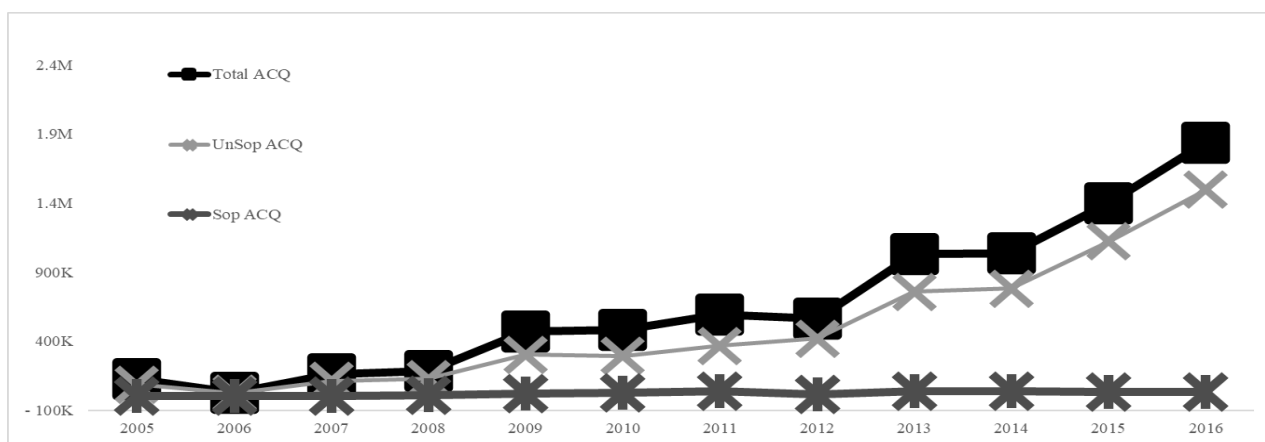


Figure 6 provides the aggregate acquisition of files by the type of trader. The x-axis is the filing year, and the y-axis is the count of traders in the population

Tests of H1a and H1b: 10-K Filing Acquisition and XBRL

Table 2 presents the Poisson regression results for equation (1) examining H1a. Columns (1), (2), and (3) report the results of estimating the relationship between *XBRL* and *Total*^{ACQ}, *LSop*^{ACQ}, and *MSop*^{ACQ} respectively. The positive and significant (p-value<0.01) coefficient on *XBRL* in column (1) means that more traders acquired form 10-K filings from EDGAR after XBRL was mandated.²⁰ The positive and significant coefficient (p-value<0.01) on *XBRL* (0.396), in column (1), translates into a 49% increase in the number of traders acquiring 10-K filings following the availability of XBRL.²¹ Similarly the positive and significant (p-value<0.01) coefficient on *XBRL* in columns (2) and (3) represent the change in less-sophisticated and more-sophisticated trader information acquisition and translate into a 49% and 26% increase in trader information acquisition. Overall, in support of H1a, total, less-sophisticated, and more-sophisticated trader acquisition of 10-K filings is higher following the implementation of XBRL.

To test H1b, whether acquisition of 10-K filings increases more for less-sophisticated traders, I compare the magnitude of the coefficient on *XBRL* between columns (2) and (3) (Clogg, Petkova, and Haritou 1995). In support of H1b, I find that the increase in information acquisition is greater for less-sophisticated traders (*LSop*^{ACQ}) as compared to more-sophisticated traders (*MSop*^{ACQ}) following the implementation of XBRL (p-value<0.01). The greater comparative increase in information acquisition for less-sophisticated traders following the implementation of XBRL suggests that their information acquisition costs decreased more than those of more-sophisticated traders.

²⁰ The significance of all p-values are based on two-tailed tests throughout this study.

²¹ The percentage change in the dependent variable due to a change in a binary indicator variable in a Poisson regression model is calculated as $[\exp(\text{Coefficient}) - 1] * 100$ (Wooldridge, 1999 page 608).

Size and *ROA* are positively associated with the *Acquisition* measures. I also find that $FD > EA$ in columns (1) and (2) each have a negative and significant coefficient (p-value<0.01) which indicates that 10-K filings with a contemporaneous earnings announcement are associated with more 10-K filing downloads than those with an earnings announcement before the 10-K filing date ($FD > EA$). This may be due to greater salience of the 10-K filings to traders when they are released contemporaneously with earnings announcements consistent with literature on market attention (deHaan Shevlin, and Thornock 2015; Blankespoor, deHaan, and Zhu 2018). Interestingly in column (3) the coefficient on $FD > EA$ is positive and significant (p-value<0.01) which suggests that more-sophisticated traders have a greater rate of 10-K acquisition for firms with a preceding earnings announcement as compared to firms with a contemporaneous release.

TABLE 2: Poisson Estimation of Acquisition

	<i>Total</i> ^{ACQ} (1)	<i>LSop</i> ^{ACQ} (2)	<i>MSop</i> ^{ACQ} (3)
	Coeff. (t-stat)	Coeff. (t-stat)	Coeff. (t-stat)
<i>XBRL</i>	0.396 *** (38.15)	0.398 *** (36.12)	0.232 *** (16.43)
<i>Words</i>	-0.099 *** (-3.24)	-0.105 *** (-3.35)	-0.074 * (-1.76)
<i>Unique Words</i>	0.036 (0.84)	0.039 (0.90)	-0.038 (-0.56)
<i>Negative Words</i>	0.112 *** (6.87)	0.114 *** (6.85)	0.128 *** (4.83)
<i>Positive Words</i>	0.006 (0.31)	0.007 (0.35)	0.007 (0.27)
<i>File Size</i>	0.029 *** (6.23)	0.030 *** (6.37)	0.000 (0.02)
<i>Lev</i>	-0.018 (-1.15)	-0.019 (-1.19)	0.028 (0.99)
<i>ROA</i>	-0.002 (-1.23)	-0.002 (-1.27)	0.000 (-0.21)
<i>Size</i>	0.076 *** (11.61)	0.073 *** (10.83)	0.147 *** (15.37)
<i>FD>EA</i>	-0.071 *** (-8.91)	-0.070 *** (-8.61)	0.012 *** (-9.38)
<i>No EA</i>	-0.235 (-0.94)	-0.233 (-0.98)	-0.507 (-0.81)
<i>Intercept</i>	4.838 *** (22.51)	4.780 *** (21.62)	0.357 *** (5.83)
<u>Fixed Effects</u>			
Filing Date	Included	Included	Included
Firm	Included	Included	Included
Number of Observations	42,495	42,495	42,495
Number of Clusters	5,959	5,959	5,959
Pseudo R2	0.946	0.945	0.664
<i>XBRL</i> difference between (2) and (3)		0.166 *** 12.552	

Table 2 reports the Poisson regression model with *Total*^{ACQ}, *LSop*^{ACQ}, and *MSop*^{ACQ} as the dependent variable. The t-statistics are clustered by firms. Significance at the 10 percent, 5 percent, and 1 percent levels is denoted as *, **, and ***. All continuous variables are winsorized at the 1 percent and 99 percent levels. Variable definitions appear in Appendix A and a detailed discussion of the *Acquisition* variable measurement (*LSop*^{ACQ}, and *MSop*^{ACQ}) is provided in Appendix B.

Tests of H2: Acquisition of 10-K Reports and Market Activity

Table 3 Panel A presents the results of OLS estimation of regression equation (2) utilizing $LSop^{LN\ ACQ}$ and $MSop^{LN\ ACQ}$ as the acquisition variables of interest. Both $LSop^{LN\ ACQ}$ and $MSop^{LN\ ACQ}$ are logged count variables which means that the coefficients can be interpreted as the proportional relationship between information acquisition and market activity. Consistent with the findings in prior literature, Table 3 Panel A shows that increased information acquisition results in increased market activity (Drake et al. 2017; Drake et al. 2015). In both Columns (1) and (2) the coefficients on both $LSop^{LN\ ACQ}$ and $MSop^{LN\ ACQ}$ are positive and significant (p-value<0.01) consistent with an increase in information acquisition being associated with an increase in abnormal returns and trading volume. Moreover, consistent with H2, the coefficients on $LSop^{LN\ ACQ}$ are significantly larger than the coefficients on $MSop^{LN\ ACQ}$ (p-value<0.01) which means that less-sophisticated trader information acquisition has a larger proportional relationship with abnormal returns and trading volume than that of their more-sophisticated counterparts.

Column (3) employs $\Delta Spread$ as the dependent variable. The coefficients on both $LSop^{LN\ ACQ}$ and $MSop^{LN\ ACQ}$ are positive and significant (p-value<0.01). These findings are generally consistent with the findings in Blankespoor et al. (2014), which document an increase in information asymmetry following the introduction of XBRL. More interestingly, the coefficient on $LSop^{LN\ ACQ}$ is significantly greater than the coefficient on $MSop^{LN\ ACQ}$ (p-value<0.01). This result suggests that less-sophisticated trader information acquisition generates proportionally greater information asymmetry within the market than that of their more-sophisticated counterparts.

TABLE 3: Market Activity and Information Acquisition

Panel A: OLS Estimation of Market Activity on Information Acquisition

	<i>CAR ABS</i> (1)	<i>Volume</i> (2)	<i>Δ Spread</i> (3)
	Coeff.	Coeff.	Coeff.
	(t-stat)	(t-stat)	(t-stat)
<i>LSop</i> ^{LNACQ}	0.020 *** (8.61)	0.714 *** (6.22)	0.476 *** (7.85)
<i>MSop</i> ^{LNACQ}	0.004 *** (3.14)	0.204 *** (5.66)	0.140 *** (3.83)
<i>XBRL</i>	-0.005 * (-1.95)	-0.358 *** (-3.84)	-0.099 (-1.48)
<i>Words</i>	0.000 (-0.25)	0.035 (0.54)	-0.026 (-0.62)
<i>Lev</i>	0.019 ** (2.48)	0.374 ** (2.01)	0.597 *** (2.92)
<i>ROA</i>	-0.001 (-1.15)	-0.002 (-1.07)	-0.021 (-1.14)
<i>BTM</i>	0.000 (0.33)	0.000 *** (5.41)	0.000 (-0.68)
<i>Spread</i>	-0.009 *** (-8.21)	-0.031 (-1.33)	-0.019 (-0.43)
<i>Turn</i>	0.000 ** (2.13)	-0.002 (-0.29)	-0.003 (-0.58)
<i>Volatility</i>	0.014 ** (2.13)	-0.394 ** (-2.33)	0.377 (1.31)
<i>Size</i>	-0.004 *** (-2.69)	-0.219 ** (-2.30)	-0.155 *** (-2.90)
<i>Intercept</i>	-0.021 (-0.97)	-1.039 (-1.45)	-0.090 (-0.16)
<i>FD > EA</i>	-0.041 *** (-16.24)	-1.149 *** (-10.77)	-1.328 *** (-17.67)
<i>No EA</i>	0.164 *** (2.88)	-0.193 (-0.17)	2.507 * (1.94)
<u>Fixed Effects</u>			
Filing Date	Included	Included	Included
Firm	Included	Included	Included
<u>F-Test</u>			
<i>LSop</i> ^{LNACQ} - <i>MSop</i> ^{LNACQ}	0.016 *** 6.379	0.511 *** 3.926	0.337 *** 4.449
Number of Observations	42,495	42,495	42,495
Number of Clusters	5,959	5,959	5,959
R-Square	0.441	0.315	0.367

TABLE 3 (cont'd)

Panel B: OLS Estimation of Market Activity on Information Acquisition for Post XBRL

	Pre-Period CAR ABS (1) Coeff. (t-stat)	Post-Period CAR ABS (2) Coeff. (t-stat)	Pre-Period Volume (3) Coeff. (t-stat)	Post-Period Volume (4) Coeff. (t-stat)	Pre-Period Δ Spread (5) Coeff. (t-stat)	Post-Period Δ Spread (6) Coeff. (t-stat)
<i>LSop</i> ^{LN ACQ}	0.010 *** (6.15)	0.020 *** (7.37)	0.329 *** (6.48)	0.488 *** (5.88)	0.288 *** (4.59)	0.421 *** (4.34)
<i>MSop</i> ^{LN ACQ}	0.001 (1.27)	0.004 *** (3.47)	0.085 *** (3.24)	0.274 *** (7.25)	0.075 ** (2.13)	0.156 *** (3.22)
<i>Words</i>	0.002 (1.53)	-0.001 (-0.66)	-0.002 (-0.05)	-0.051 (-1.06)	-0.007 (-0.13)	-0.005 (-0.09)
<i>Lev</i>	0.028 *** (4.35)	0.027 *** (4.15)	0.136 (0.80)	0.222 (1.22)	0.132 (0.56)	0.299 (1.18)
<i>ROA</i>	-0.009 (-1.46)	-0.022 *** (-3.92)	-0.258 (-1.60)	-0.107 (-0.74)	-0.394 (-1.57)	-0.366 * (-1.69)
<i>BTM</i>	0.010 *** (5.39)	0.010 *** (4.75)	-0.057 (-1.42)	-0.112 ** (-2.22)	0.026 (0.34)	-0.128 (-1.57)
<i>Prior Spread</i>	-0.006 *** (-6.24)	-0.007 *** (-8.23)	0.028 (1.47)	-0.036 * (-1.71)	-0.013 (-0.38)	-0.028 (-0.80)
<i>Turn</i>	0.002 ** (2.57)	0.002 *** (3.75)	-0.017 (-1.20)	-0.007 (-0.45)	0.020 (0.73)	0.040 (1.58)
<i>Volatility</i>	0.012 ** (1.98)	-0.002 (-0.33)	0.016 (0.11)	-0.359 *** (-2.61)	0.340 (1.36)	-0.387 * (-1.91)
<i>Size</i>	-0.001 (-0.80)	-0.002 (-1.19)	0.025 (0.45)	0.014 (0.24)	0.074 (1.06)	-0.096 (-1.39)
<i>Intercept</i>	-0.001 (-0.04)	-0.047 ** (-2.02)	-0.190 (-0.32)	-1.621 ** (-2.27)	-0.372 (-0.49)	-0.635 (-0.75)
<i>FD>EA</i>	-0.037 *** (-12.56)	-0.035 *** (-14.50)	-0.964 *** (-11.82)	-0.810 *** (-11.14)	-1.383 *** (-13.19)	-1.132 *** (-13.84)
<i>No EA</i>		0.146 * (1.82)		1.629 (1.10)		0.800 (1.36)
<u>Fixed Effects</u>						
Filing Date	Included	Included	Included	Included	Included	Included
Firm	Included	Included	Included	Included	Included	Included
<u>F-Test</u>						
<i>LSop</i> ^{LN ACQ} - <i>MSop</i> ^{LN ACQ}	0.009 *** 4.403	0.016 *** 4.982	0.244 *** 3.991	0.214 ** 2.285	0.213 *** 2.691	0.265 ** 2.312
Number of Observations	17,596	23,740	17,596	23,740	17,596	23,740
Number of Clusters	3,336	5,543	3,336	5,543	3,336	5,543
R-Square	0.573	0.608	0.422	0.489	0.428	0.484
<i>LSop</i> ^{LN ACQ} difference Pre- and Post-Period		-0.010 *** (-3.16)		-0.159 (-1.64)		-0.133 (-1.15)
<i>MSop</i> ^{LN ACQ} difference Pre- and Post-Period		-0.003 ** (-2.11)		-0.189 *** (-4.10)		-0.081 (-1.35)

TABLE 3 (cont'd)

Table 3 Panel A and B reports the ordinary least squares regression with *CAR ABS*, *Volume*, and *Spread* as the dependent variable. The t-statistics are clustered by firms. Significance at the 10 percent, 5 percent, and 1 percent levels is denoted as *, **, and ***. All continuous variables are winsorized at the 1 percent and 99 percent levels. Variable definitions appear in Appendix A and a detailed discussion of the *Acquisition* variable measurement ($LSop^{LN\ ACQ}$ and $MSop^{LN\ ACQ}$) is provided in Appendix B. Panel A utilizes the full sample and Panel B limits the sample to firm-years with XBRL filings ($XBRL = 1$) with sufficient cluster size which leaves 23,740 firm-years.

Noticeably, within the control variables the binary indicator variable that captures the presence of XBRL filings ($XBRL$) has a negative coefficient. This variable is included within the specification to control for the presence of XBRL filings however this coefficient captures only the partial direct effect. Interpreting the partial direct effect after controlling for information acquisition is not needed for testing the hypothesis in this study. Future research may consider investigating the direct and indirect effects of XBRL filings on market activity.

Next, I re-estimate equation (2) after limiting the sample to only firm-year observations where XBRL components are available ($XBRL = 1$) in Table 3 Panel B to ensure that the results found in Table 3 Panel A are identifiable post-XBRL implementation. Overall, the results tabulated in Table 3 Panel B are consistent in sign and relative magnitude with those presented in Table 3 Panel A. Similar to the results shown in Table 3 Panel A, the coefficients on more- and less-sophisticated trader information acquisition ($MSop^{LN\ ACQ}$ and $LSop^{LN\ ACQ}$) are positive and significant in columns (1), (2), and (3). In addition, similar to Table 3 Panel A, the coefficients on less-sophisticated trader information acquisition ($LSop^{LN\ ACQ}$) are larger than the corresponding coefficients on more-sophisticated trader information acquisition ($MSop^{LN\ ACQ}$) (p-value<0.05).

Additional Analyses

Type of Information Acquisition and Market Activity

Results of testing H1a, H1b, and H2 are consistent with (1) an increase in information acquisition following the implementation of XBRL, (2) the increase in information acquisition is greater for less-sophisticated traders, and (3) the information acquisition by less-sophisticated traders has a greater proportional impact on market activity relative to that of more-sophisticated traders. These analyses do not however address the type, or format, of the 10-K files acquired by traders (hereafter referred to as channels). Blankespoor et al. (2014) posit that the increase in spread they document following the implementation of XBRL may be the result of more-sophisticated traders being able to take advantage of the Machine-Readable files, whereas the less-sophisticated traders are not. To shed light on this explanation, in additional analyses I explore information acquisition segregated by the channel of information acquisition and explore their association with market activity.

These analyses present challenges as a result of the overlap in file types accessed by traders. For example, traders can and often do access both Machine-Readable and Interactive Viewer file types of the same 10-K filing, making it impossible to determine which file type the trader used to become informed. To conduct exploratory analyses, I first document the overlap between the use of different channels for the same 10-K filings when they acquire information.²² I define instances when a trader utilizes multiple channels of the 10-K as Multiple Channel information acquisition. In contrast, downloads by traders who only utilize one type of 10-K file (Interactive Viewer, Machine-Readable, or Other), are defined as a Single Channel information

²² Different forms of the same 10-K refers to the fact that the same disclosure included in the 10-K filing is provided to traders in Interactive Viewer, Machine-Readable files, and other forms of the 10-K filing. Other forms of the 10-K include text files, html files, and various other file types.

acquisition. Table 4 Panel A presents descriptive statistics for Multiple and Single Channel information acquisition at the trader level. Less-sophisticated traders consistently download information from multiple channels even when accessing Machine-Readable files. 62% (46%) of less-sophisticated traders who access Interactive Viewer (Machine-Readable files) also access another form of the same 10-K filing. In contrast, more-sophisticated traders who utilize Machine-Readable files are very unlikely to utilize any other channel (Multiple Channel $MSop^{ACQ} = 8\%$).

TABLE 4: Channel Analysis

Panel A: Information Acquisition – Single and Multiple Channel Information Acquisition

Table 5 Panel A Variables		Multiple Channel	Single Channel	Multiple and Single Channel
L Sop	$LSop^{ACQ}$	21%	79%	100%
	$Mach MSop^{ACQ}$	7%	93%	100%
	$Non Mach MSop^{ACQ}$	7%	93%	100%
L Sop	$Int LSop^{ACQ}$	59%	41%	100%
	$Mach LSop^{ACQ}$	37%	63%	100%
	$Other LSop^{ACQ}$	14%	86%	100%
MSop	$Int MSop^{ACQ}$	69%	31%	100%
	$Mach MSop^{ACQ}$	7%	93%	100%
	$Other MSop^{ACQ}$	5%	95%	100%

TABLE 4 (cont'd)**Panel B: Sample Selection – Channel Analysis**

Description	Variable	Firm-Year Observations	Insufficient Cluster Size	Remaining Sample
XBRL Components Available	$XBRL = 1$	22,190	(682)	21,508
At least one less-sophisticated trader accessing a Machine-Readable file	$Mach LSop^{ACQ} > 0$	18,665	(55)	18,610
At least one more-sophisticated trader accessing a Machine-Readable file	$Mach MSop^{ACQ} > 0$	9,363	(718)	8,645
At least one less-sophisticated trader accessing a Interactive Viewer	$Int LSop^{ACQ} > 0$	18,534	(75)	18,459
At least one more-sophisticated trader accessing Interactive Viewer	$Int MSop^{ACQ} > 0$	3,583	(1,742)	1,841

Table 4 Panel A reports the multiple channel and single channel information acquisition separated by the channel of information acquisition by both more and less-sophisticated traders. Panel B details the number of firm-year observations have different types of information acquisition as well as the number of firm-years removed in the regression due to insufficient cluster size.

Channels of Information Acquisition

To investigate which channel of information acquisition is driving the market response, ideally I would separate information acquisition into Machine-Readable, Interactive Viewer, and Other types of acquisition and then include all three channels in the same regression. Engaging in this strategy for both less- and more-sophisticated traders would result in six mutually exclusive counts of downloads on EDGAR. However, as documented in Table 4 Panel A, traders do not exclusively utilize one channel of information acquisition to inform their trading decisions which makes this research design infeasible. In addition, within this study there exists only one aggregate market response for each 10-K filing for which many traders access the 10-K via multiple channels. To investigate how the different channels of information acquisition affect market activity I estimate separate regressions with one channel of 10-K acquisition at a time for each trader type as described in more detail below.

In these analyses examining information acquisition separated by the channel of acquisition, I require that all firm-years have at least one download via the acquisition channel of interest. For instance, if a specification includes more-sophisticated traders' acquisition of Machine-Readable files ($MSop^{ACQ}$) I require that every firm-year included in the sample has at least one download of a Machine-Readable file by a more-sophisticated trader ($MSop^{ACQ} > 0$). The research design utilized in this study implements the logged count of downloads on EDGAR to allow for a comparison between more- and less-sophisticated traders' proportional relationship with the level of market activity. However, the coefficient loses this interpretation when a large percentage of firm-years have zero downloads. Empirically there are many firm-years with zero downloads for channel specific downloads. These firm-years are removed from the sample to maintain the interpretation of the coefficients.

Table 4 Panel B tabulates the number of firm-year observations to be included in regressions that investigate different channels of information acquisition within the post-XBRL period. There are 22,190 firm-year observations that have XBRL components available ($XBRL = 1$). Of those observations, 18,665 firm-year observations have at least one download of a Machine-Readable file by a less-sophisticated trader ($Mach LSop^{ACQ} > 0$) and 18,534 firm-year observations have at least one download of an Interactive Viewer file by a less-sophisticated traders ($Int LSop^{ACQ} > 0$). Interestingly, the number of firm-years that have at least one download of a Machine-Readable file or Interactive Viewer file by a more-sophisticated trader is 9,363 ($Mach MSop^{ACQ} > 0$) and 3,583 ($Int MSop^{ACQ} > 0$) firm-years which is noticeably less coverage than less-sophisticated traders.²³

Interactive Viewer and Machine-Readable Acquisition effect on Market Activity

I begin these analyses by separately examining how information acquisition via Machine-Readable files and Interactive Viewer affects market activity. First in Table 5 Panel A, I re-estimate regression equation (2) separately for less- and more-sophisticated trader acquisition of Machine-Readable files ($Mach LSop^{LN ACQ}$ and $Mach MSop^{LN ACQ}$). If traders gain an information advantage from the acquisition of Machine-Readable files then I expect that the coefficient on more- and less-sophisticated trader acquisition of Machine-Readable files ($Mach LSop^{LN ACQ}$ and $Mach MSop^{LN ACQ}$) will be positive and significant. The odd columns estimate the association between less-sophisticated trader acquisition of Machine-Readable files ($Mach$

²³ In un-tabulated analysis I re-perform Table 3 Panels A and B after limiting the sample to only firm-years with at least one download by both a more- and less-sophisticated trader ($MSop^{ACQ} > 0$ and $LSop^{ACQ} > 0$). This analysis provides similar estimations of the coefficients and t-statistics to those presented in Table 3 Panels A and B. In the re-performed version of Table 3 Panel A requiring that each firm-year observation has at least one download removes 1,985 firm-year observations due to zero downloads by less-sophisticated traders ($LSop^{ACQ}=0$) and removes an additional 3,837 firm-year observations due to zero downloads by more-sophisticated traders ($MSop^{ACQ}=0$). Finally, there is an additional 216 firm-years that are removed due to insufficient cluster size (Correia 2015). For the re-performed version of Table 3 Panel B the sample size is unchanged by the requirement that every firm-year has at least one download by both a more- and less-sophisticated trader ($MSop^{ACQ} > 0$ and $LSop^{ACQ} > 0$).

$LSop^{LN\ ACQ}$) and market activity ($CAR\ ABS$, $Volume$, and $\Delta Spread$).²⁴ Similarly, the even columns test the association between more-sophisticated trader acquisition of Machine-Readable files ($Mach\ MSop^{LN\ ACQ}$) and market activity ($CAR\ ABS$, $Volume$, and $\Delta Spread$). Second, Table 5 Panel B presents the results following the approach in Table 5 Panel A after replacing more- and less-sophisticated trader acquisition of Machine-Readable files with their acquisition of Interactive Viewer ($Int\ LSop^{LN\ ACQ}$ and $Int\ MSop^{LN\ ACQ}$).²⁵ The coefficients for less-sophisticated trader acquisition of Interactive Viewer files ($Int\ LSop^{LN\ ACQ}$) are presented in odd columns and the coefficients for more-sophisticated trader acquisition of Interactive Viewer files ($Int\ MSop^{LN\ ACQ}$) are presented in the even columns.

Overall, the results presented Table 5 Panels A and B suggest that less-sophisticated traders utilize Interactive Viewer to inform their trades but does not provide evidence that more-sophisticated traders utilize Machine-Readable files to inform their trades. In Table 5 Panel B the coefficients on less-sophisticated trader acquisition of Interactive Viewer ($Int\ LSop^{LN\ ACQ}$) are positive and significant (p-value<0.01). In contrast, more- and less-sophisticated trader acquisition of Machine-Readable files ($Mach\ LSop^{LN\ ACQ}$ and $Mach\ MSop^{LN\ ACQ}$) in Table 5 Panel A and more-sophisticated trader acquisition of Interactive Viewer in Table 5 Panel B even columns do not have a significant relationship (p-value>0.10) with market activity.

²⁴ In Table 5 Panel A there are 18,610 firm-year observations in the odd columns because I require that each firm year has at least one less-sophisticated trader downloading a Machine-Readable file ($Mach\ LSop^{ACQ} > 0$). Similarly, there are 8,645 firm-year observations in the even columns because I require that each firm year has at least one more-sophisticated trader downloading a Machine-Readable file ($Mach\ MSop^{ACQ} > 0$). Refer to Table 4 Panel B for more details regarding the sample selection criteria.

²⁵ There are 18,459 firm-year observations in the odd columns and 1,841 firm-year observations in the even columns because I require that each firm-year has at least one download of an Interactive Viewer file by a less- or more-sophisticated trader ($Int\ LSop^{ACQ} > 0$ and $Int\ MSop^{ACQ} > 0$). Refer to Table 4 Panel B for more details regarding the sample selection criteria.

TABLE 5: Machine-Readable files, Interactive Viewer, and Market Activity

Panel A: OLS Estimation of Sophisticated Trader Acquisition of Machine-Readable files and Market Activity

	<i>CAR ABS</i> (1)	<i>CAR ABS</i> (2)	<i>Volume</i> (3)	<i>Volume</i> (4)	<i>ΔSpread</i> (5)	<i>ΔSpread</i> (6)
	Coeff. (t-stat)	Coeff. (t-stat)	Coeff. (t-stat)	Coeff. (t-stat)	Coeff. (t-stat)	Coeff. (t-stat)
<i>Mach LSop</i> ^{<i>LNACQ</i>}	0.002 (1.09)		0.057 (0.53)		-0.006 (-0.11)	
<i>Mach MSop</i> ^{<i>LNACQ</i>}		0.004 (1.15)		0.120 (0.80)		0.008 (0.07)
Controls	Included	Included	Included	Included	Included	Included
<u>Fixed Effects</u>						
Filing Date	Included	Included	Included	Included	Included	Included
Firm	Included	Included	Included	Included	Included	Included
Number of Observations	18,610	8,645	18,610	8,645	18,610	8,645
Number of Clusters	4,150	3,184	4,150	3,184	4,150	3,184
R-Square	0.554	0.685	0.395	0.582	0.472	0.626
<i>Mach LSop</i> ^{<i>LNACQ</i>} - <i>Mach MSop</i> ^{<i>LNACQ</i>}		-0.002 (-0.52)		-0.063 (-0.34)		-0.014 (-0.11)

TABLE 5 (cont'd)

Panel B: OLS Estimation of Sophisticated Trader Acquisition of Interactive Viewer and Market Activity

	<i>CAR ABS</i> (1)	<i>CAR ABS</i> (2)	<i>Volume</i> (3)	<i>Volume</i> (4)	<i>ΔSpread</i> (5)	<i>ΔSpread</i> (6)
	Coeff. (t-stat)	Coeff. (t-stat)	Coeff. (t-stat)	Coeff. (t-stat)	Coeff. (t-stat)	Coeff. (t-stat)
<i>Int LSop</i> ^{LNACQ}	0.009 *** (5.83)		0.339 *** (2.77)		0.174 *** (3.75)	
<i>Int MSop</i> ^{LNACQ}		-0.002 (-0.53)		-0.096 (-0.63)		0.092 (0.95)
Controls	Included	Included	Included	Included	Included	Included
<u>Fixed Effects</u>						
Filing Date	Included	Included	Included	Included	Included	Included
Firm	Included	Included	Included	Included	Included	Included
Number of Observations	18,459	1,841	18,459	1,841	18,459	1,841
Number of Clusters	4,129	708	4,129	708	4,129	708
R-Square	0.554	0.738	0.398	0.771	0.467	0.704
<i>Int LSop</i> ^{LNACQ} - <i>Int MSop</i> ^{LNACQ}		0.011 ** (2.81)		0.435 * (2.23)		0.082 (0.76)

Table 5 Panel A and B report the ordinary least squares regression with *CAR ABS*, *Volume*, *Spread* as the dependent variable. The sample in the Table 5 Panel A odd columns are limited to firm-years with at least one download by a less-sophisticated traders using a Machine-Readable file (*Mach LSop*^{ACQ} > 0) and the even columns are limited for firm-years with at least one download by a more-sophisticated trader using a Machine-Readable file (*Mach MSop*^{ACQ} > 0). Similarly, the sample for Table 5 Panel B odd columns is limited to firm-years with at least one download by a less-sophisticated trader (*Int LSop*^{ACQ} > 0) and the sample in the even columns is limited to firm-years with at least one download by a more-sophisticated trader (*Int MSop*^{ACQ} > 0). Refer to Table 4 Panel A for more details regarding sample selection. The t-statistics are clustered by firm. Significance at the 10 percent, 5 percent, and 1 percent levels is denoted as *, **, and ***. All continuous variables are winsorized at the 1 percent and 99 percent levels. Variable definitions appear in Appendix A and a detailed discussion of the *Acquisition* variable measurement is provided in Appendix B.

Robustness Test

A valid criticism of Table 5 is that not including non-Machine-Readable or non-Interactive Viewer downloads of the 10-K filing introduces an omitted correlated variable bias. To address this criticism in Table 6 I re-estimate equation (2) after bifurcating more-sophisticated trader information acquisition ($MSop^{LNACQ}$) into more-sophisticated trader acquisition of Machine-Readable files ($Mach MSop^{LNACQ}$) and more-sophisticated trader acquisition of Non-Machine-Readable files ($Non Mach MSop^{LNACQ}$). I only separate Machine-Readable acquisition of more-sophisticated traders because it uniquely has a low rate of Multiple Channel information acquisition in Table 4 Panel B. This analysis allows for a more robust test of whether more-sophisticated traders gain an information advantage consistent with the expectations of Blankespoor et al. (2014) and Bhattacharya et al. (2018). I do not find evidence that more-sophisticated traders gain an information advantage from acquiring Machine-Readable files. In columns (1), (2), and (3) the coefficient on more-sophisticated trader acquisition of Machine-Readable files ($Mach MSop^{LNACQ}$) is not significant (p-value > 0.10). Consistent with the findings in Table 3, I find a positive and significant relationship (p-value < 0.01) between less-sophisticated trader information acquisition ($LSop^{LNACQ}$) and market activity. Interestingly, I find a positive and significant coefficient on more-sophisticated trader acquisition of Non-Machine-Readable files ($Non Mach MSop^{LNACQ}$). In aggregate, the results suggests that more-sophisticated traders gain an information advantage from acquiring 10-K filings, but they appear to gain this advantage from utilizing Non-Machine-Readable versions of the 10-K filing.

TABLE 6: Machine-Readable files and Market Activity

Robustness Test: OLS Estimation of Sophisticated Trader Acquisition of Machine-Readable files and Market Activity

	<i>CAR ABS</i> (1)	<i>Volume</i> (2)	<i>Δ Spread</i> (3)
	Coeff. (t-stat)	Coeff. (t-stat)	Coeff. (t-stat)
<i>LSop</i> ^{<i>LNACQ</i>}	0.040 *** (3.64)	1.033 *** (2.58)	0.698 *** (2.99)
<i>Mach MSop</i> ^{<i>LNACQ</i>}	-0.001 (-0.30)	-0.119 (-0.80)	-0.094 (-0.85)
<i>Non Mach MSop</i> ^{<i>LNACQ</i>}	0.013 *** (3.23)	0.974 *** (3.79)	0.292 *** (3.17)
<u>Fixed Effects</u>			
Filing Date	Included	Included	Included
Firm	Included	Included	Included
<u>F-Test</u>			
<i>Mach MSop</i> ^{<i>LNACQ</i>} - <i>Non Mach MSop</i> ^{<i>LNACQ</i>}	-0.014 ** (-2.41)	-1.093 *** (-3.55)	-0.386 ** (-2.50)
Number of Observations	8,645	8,645	8,645
Number of Clusters	3,184	3,184	3,184
R-Square	0.690	0.589	0.629

Table 6 reports the ordinary least squares regression with *CAR ABS*, *Volume*, *Spread* as the dependent variable. The sample is limited to firm-years that have at least one acquisition of all acquisition variables (*Non Mach MSop*^{*ACQ*} > 0, *Mach MSop*^{*LNACQ*} > 0, and *LSop*^{*LNACQ*} > 0). Refer to Table 4 Panel A for more details regarding sample size selection. The t-statistics are clustered by firm.

Significance at the 10 percent, 5 percent, and 1 percent levels is denoted as *, **, and ***. All continuous variables are winsorized at the 1 percent and 99 percent levels. Variable definitions appear in Appendix A and a detailed discussion of the *Acquisition* variable measurement is provided in Appendix B.

Overall, the evidence in this paper is consistent with the empirical findings of Blankespoor et al. (2014) but suggests a different causal mechanism. Blankespoor et al. (2014) attribute the association between XBRL availability and increased information asymmetry to more-sophisticated traders gaining an information advantage from processing Machine-Readable files. In this study, I find that very few more- and less-sophisticated traders download the Machine -Readable files immediately following the implementation of XBRL. In addition, I do not find a significant association between more- or less-sophisticated traders' acquisition of Machine-Readable files and information asymmetry. I do however find a strong positive association between less-sophisticated traders' acquisition of Interactive Viewer and market activity. Overall this suggests that less-sophisticated traders gain an information advantage via Interactive Viewer rather than more- or less- sophisticated traders gaining an information advantage via Machine-Readable files.

Earnings Announcements, Information Acquisition, and Market Activity

In the main analyses presented in Table 3, I include indicator variables for earnings announcement timing to control for possible differential effects on market activity conditional on the timing of the earnings announcement relative to the availability of form 10-K. If an earnings announcement is made public before the release of the 10-K filing, the release of form 10-K may confirm information for the market rather than providing new information, lowering the incentives to acquire new information. However, if there is no earnings announcement preceding the 10-K, there are greater incentives to acquire information, particularly for less-sophisticated traders who may have fewer alternative sources of information. The binary indicator variables *FD>EA* and *No EA* equal one when an earnings announcement is released prior to the 10-K filing or if there is no identified earnings announcement associated with the 10-K filing. In Table

3 Panel A, the coefficient on $FD > EA$ is negative and significant (p-value < 0.01) for $CAR\ ABS$, $Volume$, and $\Delta Spread$ in columns (1), (2) and (3). Therefore, 10-K filings that occur after an earnings announcement are associated with smaller cumulative absolute abnormal returns ($CAR\ ABS$), smaller abnormal trading volume ($Volume$) and smaller bid-ask spread ($\Delta Spread$) when compared to 10-K filings with a contemporaneous release of an earnings announcement and 10-K filing.

TABLE 7: OLS Estimation of Market Activity on Information Acquisition Subsampled by Earnings Announcement Status

	<i>CAR ABS</i> (1) <i>FD > EA = 1</i> Coeff. (t-stat)	<i>CAR ABS</i> (2) <i>FD > EA = 0</i> Coeff. (t-stat)	<i>Volume</i> (3) <i>FD > EA = 1</i> Coeff. (t-stat)	<i>Volume</i> (4) <i>FD > EA = 0</i> Coeff. (t-stat)	<i>Δ Spread</i> (5) <i>FD > EA = 1</i> Coeff. (t-stat)	<i>Δ Spread</i> (6) <i>FD > EA = 0</i> Coeff. (t-stat)
<i>LSop^{LN ACQ}</i>	0.029 *** (6.86)	0.052 *** (6.07)	0.843 *** (6.76)	2.152 *** (4.30)	0.637 *** (7.11)	1.354 *** (4.69)
<i>MSop^{LN ACQ}</i>	0.005 *** (2.92)	0.012 *** (2.66)	0.251 *** (6.81)	0.388 ** (2.00)	0.185 *** (4.34)	0.493 *** (2.77)
Controls	Included	Included	Included	Included	Included	Included
<u>Fixed Effects</u>						
Filing Date	Included	Included	Included	Included	Included	Included
Firm	Included	Included	Included	Included	Included	Included
<u>F-Test</u>						
<i>LSop^{LN ACQ} - MSop^{LN ACQ}</i>	0.024 *** 5.976	0.040 *** 3.848	0.592 *** 4.762	1.765 *** 2.914	0.453 *** 4.516	0.861 ** 2.289
<u>Difference between <i>FD > EA = 1</i> and <i>FD > EA = 0</i></u>						
<i>LSop^{LN ACQ}</i>		-0.023 ** (-2.43)		-1.310 ** (-2.54)		-0.717 ** (-2.37)
<i>MSop^{LN ACQ}</i>		-0.007 (-1.38)		-0.137 (-0.69)		-0.308 * (-1.68)
Number of Observations	28,333	6,253	28,333	6,253	28,333	6,253
Number of Clusters	4,491	1,464	4,491	1,464	4,491	1,464
R-Square	0.444	0.579	0.429	0.649	0.375	0.521

TABLE 7 (cont'd)

Table 7 reports the ordinary least squares regression with *CAR ABS*, *Volume*, *Spread* as the dependent variable. Odd columns estimate equation (2) on the subsample of firms with a preceding earnings announcement and the even columns estimate equation (2) on the subsample of firms without a preceding earnings announcement. The t-statistics are clustered by firm. Significance at the 10 percent, 5 percent, and 1 percent levels is denoted as *, **, and ***. All continuous variables are winsorized at the 1 percent and 99 percent levels. Variable definitions appear in Appendix A and a detailed discussion of the *Acquisition* variable measurement is provided in Appendix B.

To further investigate this relationship, in Table 7 I perform sub sample analyses based on the status of the earnings announcement. First, I partition the sample into two groups including firms with an earnings announcement before the 10-K filing date ($FD > EA = 1$) and firms that do not have an earnings announcement before the 10-K filing date ($FD > EA = 0$) and then estimate equation (2) separately for each partition.²⁶ Results of these analyses are presented in Table 7 with results for $FD > EA = 1$ presented in columns 1, 3 and 5, and results for $FD > EA = 0$ in columns 2, 4, and 6. Overall, the results in Table 7 suggest that more- and less-sophisticated trader 10-K filing acquisition has a significant association with market activity regardless of the status of the earnings announcement. The F-test for differences between the coefficients on less- and more-sophisticated information acquisition ($LSop^{LNACQ}$ and $MSop^{LNACQ}$) are all positive and significant (p-value<0.05). This suggests that less-sophisticated traders gain a larger proportional benefit from 10-K acquisition than their more-sophisticated counterparts regardless of the status of the earnings announcement. Next, I test the difference between coefficients on the acquisition variables ($LSop^{LNACQ}$ and $MSop^{LNACQ}$) between filings with a preceding earnings announcement ($FD > EA = 1$) in the odd columns and 10-K filings without a preceding earnings announcement ($FD > EA = 0$) in the even columns. The difference between the even and odd

²⁶ In addition, I remove the binary variables $FD > EA$ and $No\ EA$ from the modified equation 2 within the subsample analysis because there is no variation in these variables after subsampling.

columns is negative and significant (p-value<.05) for less-sophisticated trader information acquisition ($LSop^{LNACQ}$) and is negative for more-sophisticated trader information acquisition but is not statistically significant in all columns (p-value > 0.10). This suggests that less-sophisticated traders gain more information from 10-K filings when there is not a preceding earnings announcement, however the comparable relationship for more-sophisticated traders is not as robust.

Robustness Tests

Internet Service Provider Information Acquisition

A valid concern related to the results presented in Table 3 is that the less-sophisticated trader group contains more-sophisticated traders due to the way the groups are determined as explained in Appendix B. The following test is performed to ensure that the results presented in this study are not a product of the empirical definition used for less-sophisticated traders. As noted in Appendix B, less-sophisticated trader information acquisition ($LSop^{ACQ}$) includes all the information acquisition by traders that cannot be classified as more-sophisticated traders. Potentially this definition leaves more-sophisticated traders inappropriately classified as less-sophisticated traders due to an inability to identify their IP addresses. To address this concern, I re-estimate equation (2) after replacing less-sophisticated trader information acquisition ($LSop^{LNACQ}$) with two variables capturing ISP information acquisition (ISP^{LNACQ}) and non-ISP information acquisition ($Not\ ISP^{LNACQ}$). ISP^{ACQ} only includes the count of the number of file requests made from IP addresses that are known to be an ISP such as AT&T or Charter Internet. There is a lower likelihood that traders utilizing the services of an ISP are a more-sophisticated trader. $Not\ ISP^{ACQ}$ counts the number of requests from IP addresses that are classified as less-

sophisticated but are not specifically identified as an ISP.²⁷ If a more-sophisticated trader owns their own IP address but does not download enough information from EDGAR to be classified as a more-sophisticated trader their downloads will be counted in *Not ISP^{ACQ}*.

²⁷*ISP^{ACQ}* and *Not ISP^{ACQ}* are a bifurcation of *LSop^{ACQ}*. Thus: $ISP^{ACQ} + Not\ ISP^{ACQ} = LSop^{ACQ}$.

TABLE 8: OLS Estimation of Market Activity on Information Acquisition for Internet Service Providers

	<i>CAR ABS</i> (1)	<i>Volume</i> (2)	<i>Δ Spread</i> (3)
	Coeff. (t-stat)	Coeff. (t-stat)	Coeff. (t-stat)
<i>ISP</i> ^{<i>LN ACQ</i>}	0.006 *** (4.86)	0.264 *** (7.22)	0.076 * (1.80)
<i>Not ISP</i> ^{<i>LN ACQ</i>}	0.010 *** (8.18)	0.217 *** (5.65)	0.330 *** (7.10)
<i>MSop</i> ^{<i>LN ACQ</i>}	0.002 *** (3.12)	0.138 *** (6.70)	0.106 *** (3.93)
<i>XBRL</i>	-0.004 *** (-2.88)	-0.207 *** (-3.75)	-0.153 *** (-2.79)
<i>Words</i>	-0.001 (-0.59)	-0.044 (-1.51)	-0.027 (-0.77)
<i>Lev</i>	0.027 *** (7.45)	0.184 * (1.73)	0.202 (1.42)
<i>ROA</i>	-0.014 *** (-3.79)	-0.143 (-1.48)	-0.302 ** (-2.03)
<i>BTM</i>	0.010 *** (8.14)	-0.069 ** (-2.33)	-0.047 (-0.96)
<i>Prior Spread</i>	-0.006 *** (-11.29)	-0.006 (-0.50)	-0.031 (-1.36)
<i>Turn</i>	0.002 *** (5.16)	-0.010 (-1.01)	0.025 (1.48)
<i>Volatility</i>	0.006 * (1.66)	-0.191 ** (-2.09)	-0.086 (-0.58)
<i>Size</i>	-0.003 *** (-3.36)	-0.007 (-0.24)	-0.061 * (-1.72)
<i>Intercept</i>	0.009 (0.73)	-0.235 (-0.64)	0.104 (0.23)
<i>FD > EA</i>	-0.035 *** (-21.19)	-0.896 *** (-17.91)	-1.210 *** (-20.93)
<i>No EA</i>	0.158 * (1.74)	0.757 (0.48)	1.682 *** (3.58)
<u>Fixed Effects</u>			
Filing Date	Included	Included	Included
Firm	Included	Included	Included
<u>F-Test</u>			
<i>ISP</i> ^{<i>LN ACQ</i>} - <i>Not ISP</i> ^{<i>LN ACQ</i>}	-0.005 ** -2.394	0.047 0.785	-0.254 *** -3.518
<i>ISP</i> ^{<i>LN ACQ</i>} - <i>MSop</i> ^{<i>LN ACQ</i>}	0.003 ** 2.408	0.126 *** 2.936	-0.029 -0.542
<i>Not ISP</i> ^{<i>LN ACQ</i>} - <i>MSop</i> ^{<i>LN ACQ</i>}	0.008 *** 5.393	0.079 * 1.734	0.225 *** 3.999
Number of Observations	42,495	42,495	42,495
Number of Clusters	5,959	5,959	5,959
R-Square	0.534	0.383	0.388

Table 8 reports the ordinary least squares regression with *CAR ABS*, *Volume*, *Spread* as the dependent variable. The t-statistics are clustered by firm. Significance at the 10 percent, 5 percent, and 1 percent levels is denoted as *, **, and ***. All continuous variables are winsorized at the 1 percent and 99 percent levels. Variable definitions appear in Appendix A and a detailed discussion of the *Acquisition* variable measurement is provided in Appendix B.

Results are reported in Table 8. The difference in the coefficients on ISP^{LNACQ} , $Not\ ISP^{LNACQ}$, and $MSop^{LNACQ}$ vary depending upon the dependent variable in Table 8. The coefficients on ISP^{LNACQ} and $Not\ ISP^{LNACQ}$ are positive and significant similar to Table 3. The coefficient on $Not\ ISP^{LNACQ}$ is larger than the coefficient on ISP^{LNACQ} in columns (1) and (3) (p-value < 0.05), but the coefficient on $Not\ ISP^{LNACQ}$ is smaller than the coefficient on ISP^{LNACQ} in column (2) (p-value > 0.10). The larger coefficient on $Not\ ISP^{LNACQ}$ in columns (1) and (3) suggests that the traders identified within $Not\ ISP^{ACQ}$ generate more price movement and information asymmetry. This is consistent with the expectation that some traders with higher levels of sophistication are captured in $Not\ ISP^{ACQ}$ and by proxy $LSop^{ACQ}$. The coefficient on ISP^{LNACQ} is larger than the coefficient on $MSop^{LNACQ}$ in columns (1) and (2) (p-value < 0.05), but the coefficient on ISP^{LNACQ} is smaller than the coefficient on $MSop^{LNACQ}$ in column (3) (p-value > 0.10). This suggests that the less-sophisticated traders who gain their internet access via an ISP are using the information to inform their trading decisions, but are not generating the market level information asymmetry noted in Table 3. The coefficients on $Not\ ISP^{LNACQ}$ are larger than the corresponding coefficients on $MSop^{LNACQ}$ in columns (1), (2), and (3). This suggests that the results presented in Table 3 are driven by the less-sophisticated traders who do not utilize the services of an ISP.

Sensitivity Tests

Window Length

A concern with the results presented in this study is that the length of the window over which the market activity and information acquisition variables are measured impacts the results. To address this concern, I re-estimate equations (1) and (2) after modifying from a (0,2) day window to both a (0, 1) window in Table 9- and a (0,7) day window in Table 10. All variable definitions that leverage a (0,2) day window are modified to reflect the new window

specification. The window modification applies to the computation of *CAR ABS*, *Volume*, *ΔSpread*, $LSop^{LNACQ}$, and $MSop^{LNACQ}$.

Short Window Length

Table 9 Panel A re-estimates equation (1) after changing the event window from (0, 2) to (0, 1). The coefficient on *XBRL* is positive and significant similar to the results presented in Table 2. In addition, the coefficient on *XBRL* in column (2) is significantly larger than the corresponding coefficient in column (3) which suggests that less-sophisticated traders continue to have a greater proportional increase in information acquisition in a shorter window (p-value < 0.01). Similar to Table 2, this suggests that less-sophisticated traders have a greater proportional increase in their information acquisition than their more-sophisticated counterparts.

Next, in Table 9 Panel B I re-estimate equation (2) for the entire sample period after changing the event window from (0, 2) to (0, 1). The coefficients on $LSop^{LNACQ}$ and $MSop^{LNACQ}$ continue to be positive and significant (p-value < 0.01) similar to Table 3 Panel A. In addition, $LSop^{LNACQ}$ continues to be significantly larger than $MSop^{LNACQ}$ (p-value < 0.01). Similarly in Table 9 Panel C which estimates the coefficients separately for the pre- and post XBRL time periods the coefficients on $LSop^{LNACQ}$ are all positive and significant (p-value < 0.01). Similar to Table 3 Panel B the coefficients on $MSop^{LNACQ}$ are all positive, however the coefficient in column (1) is insignificant (p-value > 0.10) and the coefficient in column (3) is weakly significant (p-value < .10).

TABLE 9: (0,1) Acquisition and Trading Window

Panel A: Poisson Estimation of Acquisition with (0,1) Window

	<i>Total</i> ^{ACQ} (1)	<i>LSop</i> ^{ACQ} (2)	<i>MSop</i> ^{ACQ} (3)
	Coeff. (t-stat)	Coeff. (t-stat)	Coeff. (t-stat)
<i>XBRL</i>	0.418 *** (39.59)	0.418 *** (37.42)	0.241 *** (16.64)
<i>Words</i>	-0.102 *** (-3.17)	-0.108 *** (-3.28)	-0.072 (-1.63)
<i>Unique Words</i>	0.042 (0.95)	0.047 (1.07)	-0.053 (-0.75)
<i>Negative Words</i>	0.106 *** (6.17)	0.107 *** (6.15)	0.123 *** (4.38)
<i>Positive Words</i>	0.011 (0.49)	0.012 (0.52)	0.009 (0.30)
<i>File Size</i>	0.027 *** (5.89)	0.029 *** (6.01)	-0.001 (-0.10)
<i>Lev</i>	-0.016 (-1.04)	-0.017 (-1.06)	0.022 (0.72)
<i>ROA</i>	-0.002 (-1.17)	-0.002 (-1.26)	0.000 (-0.25)
<i>Size</i>	0.068 *** (9.72)	0.065 *** (9.01)	0.137 *** (13.51)
<i>FD>EA</i>	-0.067 *** (-7.57)	-0.066 *** (-7.36)	-0.112 *** (-8.18)
<i>No EA</i>	-0.241 (-0.84)	-0.244 (-0.88)	-0.407 (-0.71)
<i>Intercept</i>	4.696 *** (21.71)	4.633 *** (20.84)	2.001 *** (5.49)
<u>Fixed Effects</u>			
Filing Date	Included	Included	Included
Firm	Included	Included	Included
Number of Observations	42,421	42,421	42,421
Number of Clusters	5,645	5,645	5,645
Pseudo R2	0.939	0.939	0.614
<i>XBRL</i> difference between (2) and (3)		0.177 *** 12.760	

TABLE 9 (cont'd)

Panel B: OLS Estimation of Market Activity on Information Acquisition with (0,1) Window

	CAR ABS (1)	Volume (2)	Δ Spread (3)
	Coeff.	Coeff.	Coeff.
	(t-stat)	(t-stat)	(t-stat)
<i>LSop</i> ^{LN ACQ}	0.015 *** (7.97)	0.734 *** (5.92)	0.480 *** (6.77)
<i>MSop</i> ^{LN ACQ}	0.004 *** (3.44)	0.236 *** (5.39)	0.204 *** (4.54)
<i>XBRL</i>	-0.005 ** (-2.33)	-0.385 *** (-3.24)	-0.145 * (-1.78)
<i>Words</i>	-0.001 (-0.70)	0.050 (0.52)	-0.055 (-0.99)
<i>Lev</i>	0.014 ** (2.45)	0.291 (1.24)	0.694 ** (2.48)
<i>ROA</i>	0.000 (-0.96)	-0.001 (-0.46)	-0.018 (-1.13)
<i>BTM</i>	0.000 * (1.77)	0.000 *** (5.93)	0.000 (-1.03)
<i>Prior Spread</i>	-0.006 *** (-7.74)	-0.021 (-0.75)	-0.027 (-0.64)
<i>Turn</i>	0.000 * (1.73)	-0.005 (-0.43)	-0.006 (-1.18)
<i>Volatility</i>	0.011 ** (2.22)	-0.223 (-1.03)	0.415 (1.60)
<i>Size</i>	-0.003 ** (-2.14)	-0.245 * (-1.94)	-0.188 *** (-3.11)
<i>Intercept</i>	-0.008 (-0.41)	-0.431 (-0.47)	0.844 (1.19)
<i>FD > EA</i>	-0.035 *** (-16.37)	-1.613 *** (-11.83)	-1.748 *** (-17.72)
<i>No EA</i>	0.159 *** (2.64)	3.960 *** (3.27)	4.084 * (1.65)
<u>Fixed Effects</u>			
Filing Date	Included	Included	Included
Firm	Included	Included	Included
<u>F-Test</u>			
<i>LSop</i> ^{LN ACQ} - <i>MSop</i> ^{LN ACQ}	0.011 *** 5.896	0.498 *** 3.901	0.276 *** 3.115
Number of Observations	42,483	42,483	42,483
Number of Clusters	5,958	5,958	5,958
R-Square	0.407	0.319	0.362

TABLE 9 (cont'd)

Panel C: OLS Estimation of Market Activity on Information Acquisition for Pre and Post

XBRL Period with (0,1) Window

	Pre-Period CAR ABS (1) Coeff. (t-stat)	Post-Period CAR ABS (2) Coeff. (t-stat)	Pre-Period Volume (3) Coeff. (t-stat)	Post-Period Volume (4) Coeff. (t-stat)	Pre-Period $\Delta Spread$ (5) Coeff. (t-stat)	Post-Period $\Delta Spread$ (6) Coeff. (t-stat)
<i>LSop</i> ^{LNACQ}	0.009 *** (6.19)	0.019 *** (6.13)	0.420 *** (4.34)	1.108 *** (3.88)	0.345 *** (4.00)	0.609 *** (4.29)
<i>MSop</i> ^{LNACQ}	0.001 (1.21)	0.006 *** (3.68)	0.093 * (1.87)	0.472 *** (5.49)	0.150 *** (2.66)	0.251 *** (3.11)
<i>Words</i>	0.000 (-0.03)	-0.001 (-0.95)	0.082 (0.60)	-0.002 (-0.02)	-0.039 (-0.44)	-0.020 (-0.25)
<i>Lev</i>	0.019 *** (2.73)	0.016 ** (2.35)	0.407 (1.34)	0.457 (0.99)	0.687 ** (2.03)	0.741 * (1.72)
<i>ROA</i>	0.000 (-1.33)	-0.004 (-1.05)	0.000 (0.04)	0.045 (1.26)	-0.004 (-1.15)	-0.126 (-0.85)
<i>BTM</i>	-0.001 (-0.77)	0.000 *** (3.12)	-0.020 (-0.65)	0.000 *** (6.22)	-0.038 (-0.42)	0.000 (-0.55)
<i>Prior Spread</i>	-0.005 *** (-4.65)	-0.007 *** (-5.96)	0.052 (1.45)	-0.025 (-0.35)	-0.024 (-0.43)	-0.055 (-0.92)
<i>Turn</i>	0.000 (1.26)	0.000 (1.10)	0.025 (0.93)	-0.008 (-0.73)	0.006 (0.24)	-0.008 ** (-2.03)
<i>Volatility</i>	0.006 (0.83)	0.010 (1.46)	0.246 (0.74)	-0.268 (-0.83)	0.512 (1.32)	0.092 (0.28)
<i>Size</i>	0.001 (0.72)	-0.002 (-0.70)	-0.112 (-0.81)	-0.104 (-0.52)	0.030 (0.29)	-0.226 ** (-1.99)
<i>Intercept</i>	-0.002 (-0.12)	-0.049 * (-1.94)	0.363 (0.28)	-4.044 ** (-2.16)	0.375 (0.32)	-0.525 (-0.41)
<i>FD > EA</i>	-0.034 *** (-8.65)	-0.037 *** (-12.45)	-1.809 *** (-8.38)	-1.595 *** (-6.48)	-1.965 *** (-10.80)	-1.653 *** (-11.85)
<i>No EA</i>		0.185 *** (2.89)		7.774 *** (3.65)		4.922 *** (2.61)
<u>Fixed Effects</u>						
Filing Date	Included	Included	Included	Included	Included	Included
Firm	Included	Included	Included	Included	Included	Included
<u>F-Test</u>						
<i>LSop</i> ^{LNACQ} - <i>MSop</i> ^{LNACQ}	0.008 *** 4.145	0.013 *** 3.597	0.327 *** 2.818	0.636 ** 2.243	0.195 * 1.764	0.358 ** 2.112
Number of Observations	17,623	23,680	17,623	23,680	17,623	23,680
Number of Clusters	3,335	5,539	3,335	5,539	3,335	5,539
R-Square	0.468	0.546	0.437	0.407	0.391	0.479
<i>LSop</i> ^{LNACQ} difference Pre- and Post-Period		-0.009 *** (-2.69)		-0.688 ** (-2.28)		-0.264 (-1.59)
<i>MSop</i> ^{LNACQ} difference Pre- and Post-Period		-0.005 ** (-2.38)		-0.379 *** (-3.82)		-0.101 (-1.03)

TABLE 9 (cont'd)

Table 9 Panel A reports the Poisson regression model with $Total^{ACQ}$, $LSop^{ACQ}$, and $MSop^{ACQ}$ as the dependent variable. Table 9 Panel B and C report the ordinary least squares regression with $CAR\ ABS$, $Volume$, $Spread$ as the dependent variable. The t-statistics are clustered by firm. Significance at the 10 percent, 5 percent, and 1 percent levels is denoted as *, **, and ***. All continuous variables are winsorized at the 1 percent and 99 percent levels. Variable definitions appear in Appendix A; however, the window used for the variables: $Total^{ACQ}$, $LSop^{ACQ}$, and $MSop^{ACQ}$, $CAR\ ABS$, $Volume$, and $Spread$ is (0,1) instead of (0,2). A detailed discussion of the *Acquisition* variable measurement is provided in Appendix B.

Long Window Length

Table 10 panels A, B, and C mirror the panels presented in Table 9. The results in Table 10 Panels A, B, and C continue to be similar to those presented in Table 3 and Table 9. Notable exceptions include in Table 10 Panel B column (1) the coefficient on $MSop^{LN\ ACQ}$ is weakly significant (p-value < 0.10) as compared to the corresponding coefficient on Table 3 Panel A which demonstrates a strong significance (p-value < 0.01). In addition, the coefficients on $MSop^{LN\ ACQ}$ in Table 10 Panel C are not significant (p-value > 0.10) except for column (4) which is strongly significant (p-value < 0.01). This is a notable deviation from Table 3 Panel B and Table 9 Panel C where the corresponding coefficients have greater levels of statistical significance. This suggests that more-sophisticated traders process information to inform their trading quickly after the release of the 10-K filing.

TABLE 10: (0,7) Acquisition and Trading Window

Panel A: Poisson Estimation of Acquisition with (0,7) Window

	<i>Total</i> ^{ACQ} (1)	<i>LSop</i> ^{ACQ} (2)	<i>MSop</i> ^{ACQ} (3)
	Coeff. (t-stat)	Coeff. (t-stat)	Coeff. (t-stat)
<i>XBRL</i>	0.329 *** (33.26)	0.331 *** (31.45)	0.196 *** (14.12)
<i>Words</i>	-0.106 *** (-3.23)	-0.113 *** (-3.35)	-0.049 (-1.35)
<i>Unique Words</i>	0.061 (1.34)	0.063 (1.36)	0.013 (0.22)
<i>Negative Words</i>	0.109 *** (6.68)	0.113 *** (6.77)	0.085 *** (3.52)
<i>Positive Words</i>	0.009 (0.47)	0.011 (0.53)	0.002 (0.10)
<i>File Size</i>	0.035 *** (7.63)	0.039 *** (7.99)	-0.003 (-0.56)
<i>Lev</i>	-0.007 (-0.43)	-0.006 (-0.40)	0.023 (0.88)
<i>ROA</i>	-0.001 (-0.68)	-0.001 (-0.72)	0.000 (-0.13)
<i>Size</i>	0.097 *** (14.75)	0.092 *** (13.66)	0.182 *** (20.03)
<i>FD>EA</i>	-0.064 *** (-9.09)	-0.064 *** (-8.85)	0.011 *** (-7.95)
<i>No EA</i>	-0.108 (-0.59)	-0.103 (-0.56)	-0.706 (-1.82)
<i>Intercept</i>	4.935 *** (22.89)	4.866 *** (21.92)	0.325 *** (6.45)
<u>Fixed Effects</u>			
Filing Date	Included	Included	Included
Firm	Included	Included	Included
Number of Observations	42,451	42,451	42,451
Number of Clusters	5,950	5,950	5,950
Pseudo R2	0.953	0.952	0.752
<i>XBRL</i> difference between (2) and (3)		0.135 *** 10.924	

TABLE 10 (cont'd)

Panel B: OLS Estimation of Market Activity on Information Acquisition with (0,7) Window

	CAR ABS (1)	Volume (2)	$\Delta Spread$ (3)
	Coeff.	Coeff.	Coeff.
	(t-stat)	(t-stat)	(t-stat)
<i>LSop</i> ^{LNACQ}	0.041 *** (9.68)	0.780 *** (9.45)	0.359 *** (7.39)
<i>MSop</i> ^{LNACQ}	0.004 * (1.72)	0.154 *** (4.75)	0.060 ** (2.18)
<i>XBRL</i>	-0.011 ** (-2.14)	-0.321 *** (-3.64)	-0.022 (-0.42)
<i>Words</i>	0.001 (0.30)	0.027 (0.54)	-0.053 (-1.61)
<i>Lev</i>	0.040 ** (2.36)	0.313 ** (2.29)	0.342 * (1.88)
<i>ROA</i>	-0.001 (-0.92)	-0.001 (-0.69)	-0.010 (-1.04)
<i>BTM</i>	0.000 (0.57)	0.000 *** (4.61)	0.000 (-0.54)
<i>Prior Spread</i>	-0.023 *** (-12.01)	-0.019 (-1.03)	-0.053 (-1.56)
<i>Turn</i>	0.000 * (1.72)	-0.009 (-1.51)	-0.011 ** (-2.43)
<i>Volatility</i>	0.036 *** (3.24)	-0.215 (-1.45)	0.225 (1.09)
<i>Size</i>	-0.007 *** (-2.95)	-0.253 *** (-3.64)	-0.099 ** (-2.54)
<i>Intercept</i>	-0.098 ** (-2.13)	-1.718 *** (-2.67)	-0.323 (-0.73)
<i>FD>EA</i>	-0.062 *** (-15.80)	-0.849 *** (-9.76)	-0.858 *** (-15.58)
<i>No EA</i>	0.101 (1.45)	0.283 (0.25)	3.364 (1.12)
<u>Fixed Effects</u>			
Filing Date	Included	Included	Included
Firm	Included	Included	Included
<u>F-Test</u>			
<i>LSop</i> ^{LNACQ} - <i>MSop</i> ^{LNACQ}	0.037 *** 8.652	0.626 *** 6.898	0.299 *** 4.988
Number of Observations	42,483	42,483	42,483
Number of Clusters	5,958	5,958	5,958
R-Square	0.503	0.308	0.374

TABLE 10 (cont'd)

Panel C: OLS Estimation of Market Activity on Information Acquisition for Pre and Post XBRL with (0,7) Window

	Pre-Period CAR ABS (1) Coeff. (t-stat)	Post-Period CAR ABS (2) Coeff. (t-stat)	Pre-Period Volume (3) Coeff. (t-stat)	Post-Period Volume (4) Coeff. (t-stat)	Pre-Period Δ Spread (5) Coeff. (t-stat)	Post-Period Δ Spread (6) Coeff. (t-stat)
<i>LSop</i> ^{LNACQ}	0.024 *** (6.37)	0.061 *** (8.36)	0.494 *** (6.37)	1.133 *** (6.79)	0.257 *** (4.39)	0.495 *** (5.37)
<i>MSop</i> ^{LNACQ}	0.001 (0.65)	0.005 (1.49)	0.058 (1.64)	0.355 *** (5.03)	0.047 (1.36)	0.062 (1.27)
<i>Words</i>	0.003 (0.97)	-0.001 (-0.28)	-0.023 (-0.39)	0.088 (1.04)	-0.037 (-0.70)	-0.037 (-0.81)
<i>Lev</i>	0.046 *** (3.53)	0.051 *** (3.17)	0.235 (0.98)	0.396 * (1.69)	-0.078 (-0.37)	0.569 ** (2.21)
<i>ROA</i>	0.000 ** (-2.14)	-0.009 * (-1.92)	-0.001 (-0.79)	0.030 * (1.66)	-0.003 (-1.29)	-0.075 (-1.03)
<i>BTM</i>	0.004 (0.79)	0.000 (1.09)	-0.020 (-0.89)	0.000 *** (6.59)	-0.100 (-1.54)	0.000 (-0.27)
<i>Prior Spread</i>	-0.022 *** (-9.75)	-0.025 *** (-8.56)	0.012 (0.51)	-0.005 (-0.13)	-0.122 *** (-3.44)	-0.024 (-0.63)
<i>Turn</i>	0.000 (0.07)	0.000 (1.44)	0.015 (0.85)	-0.011 ** (-2.40)	-0.008 (-0.62)	-0.012 *** (-2.84)
<i>Volatility</i>	0.036 ** (2.07)	0.024 * (1.78)	-0.093 (-0.43)	-0.294 (-1.35)	0.009 (0.04)	0.168 (0.77)
<i>Size</i>	-0.005 (-1.17)	-0.013 *** (-2.80)	-0.057 (-0.74)	-0.361 *** (-2.68)	0.106 (1.60)	-0.249 *** (-3.39)
<i>Intercept</i>	-0.018 (-0.41)	-0.198 *** (-3.55)	-0.423 (-0.52)	-4.905 *** (-3.46)	-0.963 (-1.30)	-0.466 (-0.58)
<i>FD>EA</i>	-0.064 *** (-8.44)	-0.066 *** (-11.81)	-0.839 *** (-8.22)	-0.901 *** (-5.53)	-0.974 *** (-9.62)	-0.873 *** (-11.63)
<i>No EA</i>		0.132 * (1.81)		1.267 (0.83)		3.101 (1.15)
<u>Fixed Effects</u>						
Filing Date	Included	Included	Included	Included	Included	Included
Firm	Included	Included	Included	Included	Included	Included
<u>F-Test</u>						
<i>LSop</i> ^{LNACQ} - <i>MSop</i> ^{LNACQ}	0.022 *** 4.797	0.056 *** 6.949	0.437 *** 4.895	0.778 *** 4.285	0.210 *** 2.800	0.433 *** 3.931
Number of Observations	17,520	23,831	17,520	23,831	17,520	23,831
Number of Clusters	3,339	5,544	3,339	5,544	3,339	5,544
R-Square	0.612	0.622	0.380	0.402	0.411	0.485
<i>LSop</i> ^{LNACQ} difference Pre- and Post-Period		-0.037 *** (-4.56)		-0.639 *** (-3.47)		-0.238 ** (-2.18)
<i>MSop</i> ^{LNACQ} difference Pre- and Post-Period		-0.003 (-0.92)		-0.297 *** (-3.77)		-0.015 (-0.25)

Table 10 Panel A reports the Poisson regression model with *Total*^{ACQ}, *LSop*^{ACQ}, and *MSop*^{ACQ} as the dependent variable. Table 10 Panel B and C report the ordinary least squares regression with *CAR ABS*, *Volume*, *Spread* as the dependent variable. The t-statistics are clustered by firm. Significance at the 10 percent, 5 percent, and 1 percent levels is denoted as *, **, and ***. All continuous variables are winsorized at the 1 percent and 99 percent levels. Variable definitions appear in Appendix A; however, the window used for the variables: *Total*^{ACQ}, *LSop*^{ACQ}, and *MSop*^{ACQ}, *CAR ABS*, *Volume*, and *Spread* is (0,7) instead of (0,2). A detailed discussion of the *Acquisition* variable measurement is provided in Appendix B.

Together the results in Table 9 and Table 10 suggest that more-sophisticated traders process information within the 10-K within short order of the release of the 10-K filing. This is consistent with the expectation that more-sophisticated traders have greater access to resources which allows them to process the 10-K filing shortly after the release. In addition, less-sophisticated trader 10-K acquisition has a larger association with market activity than more-sophisticated traders in Table 9 and Table 10. This suggests that the 10-K is a more important information event for less-sophisticated traders than for more-sophisticated traders

Conditional Estimation

A key research design choice in this study is the decision to keep all firm-years within Tables 2-3, but then limit the sample to only firm years with at least one download for every acquisition variable in Tables 5-8. To ensure that Tables 2-3 are not impacted by the inclusion of firm-years without any downloads by more- or less-sophisticated traders, I re-perform these tests after limiting the sample.

In Table 11 I require that each firm-year has at least one download by a more-sophisticated trader ($MSop^{ACQ} > 0$) and a less-sophisticated trader ($LSop^{ACQ} > 0$). Requiring that each firm-year has at least one download by a less-sophisticated trader ($LSop^{ACQ} > 0$) removes 1,985 firm-year observations. Similarly, requiring that each firm-year has a more-sophisticated trader download ($MSop^{ACQ} > 0$) removes an additional 3,837 firm-year observations. Finally, I remove 1,542 firm-year observations from the analysis that have insufficient cluster size to avoid incorrect statistical inferences (Correia 2015). This leaves a sample of 36,457 firm-year observations from 5,482 unique firms.

In Table 11 Panels A, B, and C, I re-estimate equation (1) and (2) with the new limited sample. Similar to the results in Table 2, the coefficient on *XBRL* is positive and significant in

Table 11 Panel A. In addition, the coefficient on XBRL in column (2) is statistically larger than the corresponding coefficient in column (3) which provides support for H1 which predicts that less-sophisticated traders have a disproportionate increase in information acquisition following the implementation of XBRL. Similarly, in Table 11 Panel B, the coefficients on $LSop^{LNACQ}$ and $MSop^{LNACQ}$ are positive and significant. The difference in the coefficients on $LSop^{LNACQ}$ and $MSop^{LNACQ}$ are positive and significant similar to the results in Table 3 Panel A. The coefficients on $LSop^{LNACQ}$ and $MSop^{LNACQ}$ in Table 11 Panel C are positive and significant similar to Table 3 Panel B. In addition, the coefficient on $LSop^{LNACQ}$ is larger than the coefficient on $MSop^{LNACQ}$ (p-value < 0.01) in all three columns similar to the results in Table 3 Panel B. Overall, these findings suggest the results in Tables 2 and 3 are not sensitive to the removal of firm-years without any downloads by more- or less-sophisticated traders.

TABLE 11: Population Limited to Firm-Years with Downloads

Panel A: Poisson Estimation of Acquisition: Population Limited to Firm-Years with Downloads

	<i>Total</i> ^{ACQ} (1)	<i>LSop</i> ^{ACQ} (2)	<i>MSop</i> ^{ACQ} (3)
	Coeff.	Coeff.	Coeff.
	(t-stat)	(t-stat)	(t-stat)
<i>XBRL</i>	0.396 *** (37.87)	0.398 *** (35.82)	0.236 *** (16.90)
<i>Words</i>	-0.102 *** (-3.29)	-0.107 *** (-3.39)	-0.075 * (-1.80)
<i>Unique Words</i>	0.038 (0.88)	0.041 (0.93)	-0.024 (-0.37)
<i>Negative Words</i>	0.114 *** (6.87)	0.116 *** (6.83)	0.126 *** (4.85)
<i>Positive Words</i>	0.006 (0.28)	0.007 (0.32)	0.003 (0.11)
<i>File Size</i>	0.030 *** (6.29)	0.031 *** (6.40)	0.002 (0.27)
<i>Lev</i>	-0.017 (-1.08)	-0.018 (-1.13)	0.036 (1.35)
<i>ROA</i>	-0.002 (-1.19)	-0.002 (-1.24)	0.000 (-0.18)
<i>Size</i>	0.077 *** (11.42)	0.074 *** (10.64)	0.144 *** (15.36)
<i>FD>EA</i>	-0.069 (-8.59)	-0.068 *** (-8.29)	0.012 *** (-9.28)
<i>No EA</i>	-0.234 (-0.92)	-0.229 (-0.95)	-0.548 (-0.93)
<i>Intercept</i>	4.844 *** (22.16)	4.788 *** (21.25)	0.352 *** (5.73)
<u>Fixed Effects</u>			
Filing Date	Included	Included	Included
Firm	Included	Included	Included
Number of Observations	36,457	36,457	36,457
Number of Clusters	5,482	5,482	5,482
Pseudo R2	0.946	0.945	0.664
<i>XBRL</i> difference between (2) and (3)		0.162 *** 12.824	

TABLE 11 (cont'd)

Panel B: OLS Estimation of Market Activity on Information Acquisition: Population Limited to Firm-Years with Downloads

	<i>CAR ABS (1)</i>	<i>Volume (2)</i>	<i>Δ Spread (3)</i>
	Coeff. (t-stat)	Coeff. (t-stat)	Coeff. (t-stat)
<i>LSop</i> ^{LNACQ}	0.034 *** (9.39)	1.208 *** (6.00)	0.786 *** (8.84)
<i>MSop</i> ^{LNACQ}	0.007 *** (4.17)	0.303 *** (6.00)	0.251 *** (5.39)
<i>XBRL</i>	-0.010 *** (-3.39)	-0.552 *** (-5.28)	-0.224 *** (-3.08)
<i>Words</i>	-0.002 (-0.97)	-0.003 (-0.04)	-0.069 (-1.47)
<i>Lev</i>	0.025 *** (2.62)	0.440 ** (2.11)	0.627 *** (3.45)
<i>ROA</i>	-0.001 (-1.19)	-0.002 (-0.86)	-0.021 (-1.11)
<i>BTM</i>	0.000 (0.39)	0.000 *** (5.77)	0.000 (-0.58)
<i>Prior Spread</i>	-0.009 *** (-7.40)	-0.026 (-0.99)	-0.009 (-0.18)
<i>Turn</i>	0.000 ** (2.04)	-0.007 (-0.71)	-0.005 (-1.14)
<i>Volatility</i>	0.014 ** (1.96)	-0.400 ** (-2.07)	0.487 (1.54)
<i>Size</i>	-0.006 *** (-3.68)	-0.330 *** (-2.94)	-0.211 *** (-3.54)
<i>Intercept</i>	-0.088 *** (-3.03)	-3.059 *** (-3.10)	-1.425 ** (-2.13)
<i>FD > EA</i>	-0.041 *** (-14.62)	-1.149 *** (-9.72)	-1.312 *** (-15.94)
<i>No EA</i>	0.152 *** (3.38)	-0.406 (-0.47)	2.246 (1.36)
<u>Fixed Effects</u>			
Filing Date	Included	Included	Included
Firm	Included	Included	Included
<u>F-Test</u>			
<i>LSop</i> ^{LNACQ} - <i>MSop</i> ^{LNACQ}	0.027 *** 7.293	0.905 *** 4.061	0.535 *** 5.074
Number of Observations	36,457	36,457	36,457
Number of Clusters	5,482	5,482	5,482
R-Square	0.454	0.327	0.394

TABLE 11 (cont'd)

Panel C: OLS Estimation of Sophisticated Trader Acquisition of Machine-Readable files, CAR ABS, Volume, and Spread: Population Limited to Firm-Years with Downloads

	<i>CAR ABS</i> (1)	<i>Volume</i> (2)	Δ <i>Spread</i> (3)
	Coeff. (t-stat)	Coeff. (t-stat)	Coeff. (t-stat)
$LSop^{LNACQ}$	0.043 *** (7.29)	1.676 *** (4.25)	0.850 *** (5.28)
$MSop^{LNACQ}$	0.010 *** (3.87)	0.542 *** (5.31)	0.340 *** (4.28)
Controls	Included	Included	Included
<u>Fixed Effects</u>			
Filing Date	Included	Included	Included
Firm	Included	Included	Included
<u>F-Test</u>			
$LSop^{LNACQ} - MSop^{LNACQ}$	0.033 *** 4.938	1.133 ** 2.546	0.511 *** 2.686
Number of Observations	21,508	21,508	21,508
Number of Clusters	5,056	5,056	5,056
R-Square	0.572	0.407	0.502

Table 11 Panel A reports the Poisson regression model with $Total^{ACQ}$, $LSop^{ACQ}$, and $MSop^{ACQ}$ as the dependent variable. Table 11 Panel B reports the ordinary least squares regression with $CAR\ ABS$, $Volume$, and $Spread$ as the dependent variable. Table 11 Panel C limits the sample presented in Table 11 Panel B to only the firm-years with XBRL data ($XBRL = 1$). The t-statistics are clustered by firms. Significance at the 10 percent, 5 percent, and 1 percent levels is denoted as *, **, and ***. All continuous variables are winsorized at the 1 percent and 99 percent levels. Variable definitions appear in Appendix A and a detailed discussion of the *Acquisition* variable measurement ($LSop^{LNACQ}$ and $MSop^{LNACQ}$) is provided in Appendix B. Sample is limited to only firm-years with both a more-sophisticated trader download ($MSop^{ACQ} > 0$) and a less-sophisticated trader download ($LSop^{ACQ} > 0$)

V. CONCLUSION

This study reexamines the SEC's conjectured benefits from the implementation of the XBRL standard by examining how the standard influenced the rate of information acquisition and the related market activity. SEC regulation §232.405 required firms to make available Machine-Readable files and Interactive Viewer, a point and click tool on EDGAR, for 10-K filings. This study finds the number of less-sophisticated traders who acquire 10-K filings within two trading days after the 10-K release increases by 49% following the implementation of XBRL. This study also finds the increase in less-sophisticated trader acquisition is statistically higher than that of sophisticated traders. In the second set of analyses, this study finds that greater information acquisition is associated with higher levels of market activity and that this effect is disproportionately driven by the information acquisition of less-sophisticated traders. The disproportionate increase in less-sophisticated information acquisition suggests that Interactive Viewer decreased information acquisition costs disproportionately for less-sophisticated traders as they are the group more likely to utilize the point and click interface provided by Interactive Viewer.

This study contributes to the literatures that focus on XBRL, disclosure informativeness, and information processing costs. This study adds to the XBRL literature such as Blankespoor et al. (2014) by empirically documenting a significant change in information acquisition. In addition, the findings of this study suggest that less-sophisticated traders benefited from the XBRL standard via their utilization of Interactive Viewer. Similarly, this study contributes to the literature on the informativeness of 10-K filings by documenting that the limited market activity of 10-K filings may be due in part to disclosure acquisition costs that less-sophisticated traders experience. Finally, this study empirically documents the effects of information presentation on

trader behavior that has been previously explored in experimental studies such as Rennkamp (2012) and Nelson and Rupa (2014).

APPENDICES

APPENDIX A: VARIABLE DEFINITIONS

Variable Definitions

FIGURE A: Variable Definitions

	Variable Name	Source	Description
Dependent Variables	<i>CAR ABS</i>	CRSP	The absolute value of the cumulative value-weighted abnormal return from the day of the release of the annual filing and two trading days after the filing. $ABS\ CAR = \sum_{k=0}^2 ar_k $
	<i>Volume</i>	CRSP	The mean daily trading volume during the release of the annual filing from the day of the release of the annual filing and two trading days after the filing (0,2) less the mean daily trading volume during the non-filing period (-45, -5) scaled by the standard deviation of the daily trading volume during the non-filing period (-45, -5). $Volume = \frac{\mu(vol(0,2)) - \mu(vol(-45,-5))}{\sigma(vol(-45,-5))}$
	$\Delta Spread$	CRSP	The mean spread during the release of the annual filing from the day of the release of the annual filing and two trading days after the filing (0,2) less the mean spread during the non-filing period (-45, -5). $Spread = Spread_{(0,2)} - Spread_{(-45,-5)}$ $Spread_{(0,2)} = \text{avg}(\text{Daily Spread}_0, \text{Daily Spread}_1, \text{Daily Spread}_2)$ $\text{Daily Spread}_t = \text{Ask High}_t - \text{Bid Low}_t / ((\text{Bid Low}_t + \text{Ask High}_t)/2)$
	<i>XBRL</i>	EDGAR	A binary indicator variable that is equal to 1 for the year when the firm has one or more acquisitions of a Machine-Readable file or an Interactive Viewer file on EDGAR and all subsequent years. All years before the first year of interactive file acquisition on EDGAR are equal to 0.
Acquisition	<i>Total</i> ^{LN ACQ}	EDGAR Log	A count of all unique ip/file request on EDGAR. Transformed according to steps defined in Appendix B.
	<i>LSop</i> ^{LN ACQ}	EDGAR Log	A count of unsophisticated trader unique ip/file request on EDGAR. Transformed according to steps defined in Appendix B.
	<i>MSop</i> ^{LN ACQ}	EDGAR Log	A count of sophisticated trader unique ip/file request on EDGAR. Transformed according to steps defined in Appendix B.
	<i>Int LSop</i> ^{LN ACQ}	EDGAR Log	A count of all unique ip/file request for unsophisticated traders accessing interactive viewer files on EDGAR. Transformed according to steps defined in Appendix B.
	<i>Int MSop</i> ^{LN ACQ}	EDGAR Log	A count of all unique ip/file request for sophisticated traders accessing interactive viewer files on EDGAR. Transformed according to steps defined in Appendix B.
	<i>Mach MSop</i> ^{LN ACQ}	EDGAR Log	A count of all unique ip/file request for sophisticated traders accessing machine readable files on EDGAR. Transformed according to steps defined in Appendix B.
	<i>Mach LSop</i> ^{LN ACQ}	EDGAR Log	A count of all unique ip/file request for unsophisticated traders accessing machine readable files on EDGAR. Transformed according to steps defined in Appendix B.
	<i>Non Mach MSop</i> ^{LN ACQ}	EDGAR Log	A count of all unique ip/file request for unsophisticated traders accessing any file other than Machine-Readable files on EDGAR. Transformed according to steps defined in Appendix B.

FIGURE A (cont'd)

	Variable Name	Source	Description
	<i>Words</i>	EDGAR	The log of the count of words in the 10-K from Loughran and McDonald (2011).
Control Variables	<i>Lev</i>	Compustat	Liabilities divided by assets (Compustat variables "dltt" / "at").
	<i>ROA</i>	Compustat	Sales divided by assets (compustat variables "ib" / "at").
	<i>Size</i>	Compustat	The log of assets from compustat (compustat variable "at").
	<i>BTM</i>	Compustat	Book Value/Market Value.
	<i>Inst. Hold</i>	Thompson	From Thomson Reuter's Institutional Holdings database.
	<i>Turn</i>	CRSP	The firm's turnover 90 days before the filing of the 10-K $VOL/(PRC * SHROUT)$.
	<i>Prior Spread</i>	CRSP	The average firm spread for the 90 days before the filing of the 10-K $(BIDLO-ASKHI)/((BIDLO + ASKHI)/2) * 100$.
	<i>Volatility</i>	CRSP	Standard deviation of Log(1+ Daily Return, annualized by multiplying by $\sqrt{252}$ measured over the 90 days before the filing of the 10-K.
EA Status	<i>Intercept</i>	Compustat	The intercept of the multivariable estimation. Captures the effect of contemporaneous release of 10-K filing and earnings announcement.
	<i>FD<EA</i>	Compustat	A binary indicator variable that equals 1 when 10-K filing occurs after the earnings announcement. Otherwise is equal to 0.
Acquisition variables count the unique IP Address, channel, and day user access requests. Refer to Appendix B for a more complete discussion of the EDGAR filing system and the measurement approach.			

APPENDIX B: EDGAR LOG

Overview of SEC EDGAR Log

SEC EDGAR tracks all users that download information within EDGAR and has made this information public in the form of the EDGAR Log from January 1st 2003 to June 30th 2017. Each individual download on SEC EDGAR results in the system recording the download in the EDGAR Log with the user's IP address, time of request, and information requested. A user's location, organization, as well as their aggregate activity can be inferred from their IP address in the EDGAR log, with some limitations. Specifically, EDGAR users who gain internet access via contracting with internet service providers (ISPs) are more difficult to isolate.²⁸ IP addresses are the primary means to differentiate between users on EDGAR. ISPs dynamically change their users' IP addresses to balance internet usage across their infrastructure as well as ensure efficient usage of their limited IP addresses. This load balancing makes individual identification impossible because the user's IP address will be changed by their ISP each time they visit EDGAR.²⁹

EDGAR Log Trader Classification

I look up information from "ipwhois" for each IP Block that has at least 100 downloads from EDGAR during my sample period to classify traders as more- or less-sophisticated traders.³⁰ All IP address have organizational text associated with the IP address that can be identified via a service such as "ipwhois." IP addresses that are owned by large institutions such as large financial institutions, universities, and audit firms contain the institution's name in the IP

²⁸ ISPs are companies that consumers and businesses contract with for internet connectivity. Examples of ISPs include Charter, AT&T, Time Warner Cable, and Comcast.

²⁹ ISPs change their users IP addresses; however, they rarely balance load across large geographic distances given the physical limitations of networks. Previous studies have leveraged this limitation to explore EDGAR usage by geographic area (Drake et al. 2017; Bernile, Kumar, and Sulaeman 2015)

³⁰ IP Blocks in this context are all the IP addresses with the same first three octets of the four octets included in an IPv4 IP address. Each IP Block has 255 IP addresses.

address' organizational text. I utilize this organizational text to classify IP addresses as more-sophisticated, less-sophisticated, and non-trader entities.

To classify the organizational text from IP blocks I use a systematic approach that utilizes both manual and automated classification steps. First, all IP addresses that have financial institution's name in their organizational text are classified as more-sophisticated traders. For example, if an IP address' organizational text includes: "Bank of America" then the IP address is classified as a more-sophisticated trader. Second, all IP addresses with a non-trader entities' name in the organizational text are classified as non-trader. For example, if an IP address' organizational text includes: "University" the IP address is classified as a non-trader entity.³¹ Finally, all IP addresses that are owned by an ISP, have conflicting classifications, or are not classified via any other criteria are classified as a less-sophisticated trader.

To apply my IP address classification, first I obtain the organizational text for all IP Blocks with more than 100 downloads on EDGAR from January 1st, 2003 to June 30th 2017 from ipwhois. I manually classify 360 IP blocks with the most downloads from EDGAR during the time period into more-sophisticated, less-sophisticated, and non-trader entities. In addition, I utilize a key word list to classify more-sophisticated and non-trader entities based on the word list from Drake et al. (2020).^{32,33} If the organizational text for two different IP addresses within an IP Block map to different classifications than the entire IP block is classified as less-sophisticated. For instance, if one IP address meets the classification of more-sophisticated trader and a second IP address in the same block meets the criteria for a less-sophisticated trader then

³¹ Non-trader entities refer to entities such as universities, public accounting firms, and law firms that have incentives to download information from EDGAR but are unlikely to use this information to execute trades.

³² IP Block refers to a set of IP address with the same first three octets in the IPv4 address.

³³ My key word list is inspired by Drake et al. (2020), but the organization names have been shortened to increase classification rate. Sophisticated trader key words include: 'goldman', 'sigma', 'deutsche', 'bank of america', 'boa', 'bny', 'macquarie', 'barclays', 'maverick', 'jpmorgan', 'chase'. ISP key words include: 'time warner', 'verizon', 'comcast', 'earthlink', 'at&t', 'qwest', 'charter', 'hurricane', 'isp'

the entire IP block is classified as a less-sophisticated trader. In addition, if the manual and automated classifications provide different classifications for an IP block than the entire IP block is classified as a less-sophisticated trader.

EDGAR File Formats

Annual 10-K filings are provided to end users via many different channels on EDGAR such as Interactive Viewer or Machine-Readable files. All channels of 10-K acquisition contain the same audited financial statements regardless of the channel; however, the way the information is stored on EDGAR varies drastically based on channel.³⁴ EDGAR records every individual file request from the server rather than what channel the 10-K was accessed. Interactive Viewer stores the 10-K in many different individual files on EDGAR which can easily exceed 100 uniquely identifiable files. Every time a user clicks on a button in Interactive Viewer this creates a recorded download on the EDGAR log. The numerous individual files associated with Interactive Viewer creates significantly more downloads per individual user than the other channels of accessing the 10-K such as Machine-Readable or the text file (Ryans 2017). For instance, if a user wanted to view the balance sheet and income statement of a company via the text file channel EDGAR would record one download. In contrast, Interactive Viewer stores the balance sheet and the income statement in different files. If the trader clicks on the income statement and then clicks on the balance sheet two different downloads will be recorded on the EDGAR log. Traders who access the 10-K via Interactive Viewer will generate significantly more requests on EDGAR than a comparable trader using a text file. Downloads of a 10-K filing

³⁴ Interactive Viewer and Machine-Readable files only contain XBRL labeled data. MD&A and items from the company are not XBRL labeled thus they are only included in the text file and HTML versions of the annual filing. All XBRL labeled information by regulation should contain the same disclosures.

via Interactive Viewer are unlikely to be automated because the files that support Interactive Viewer are not well suited for automated downloading or processing.³⁵

I classify files into 10-K channels by the file name and extension of the file. To identify the channel of a download I systematically categorize the file's name into a channel.³⁶ Because the channel type affects the number of requests needed to extract the same amount of information, my count of file acquisition ($LSop^{ACQ}$, $MSop^{ACQ}$, etc.) counts the unique channel, day, and cik combination. This definition removes the higher number of requests driven by the nature of Interactive Viewer. This results in the example from the paragraph above counting both users as having as one request. This ensures that users moving from the text file channel to the Interactive Viewer channel do not generate more requests as a product of the channel used.

Machine vs. Robot Definition

Papers such as Loughran and McDonald (2017), Drake et al. (2015), and Ryans (2017) have anticipated that traders may use programs/scripts to systematically download data from EDGAR. A small group of traders could programmatically download large amounts of data from EDGAR. This type of robotic acquisition is not considered to be an accurate reflection of information acquisition because the data may be downloaded, archived, and not utilized for decision making immediately after the download from EDGAR. To exclude the automated download of files from EDGAR from measures of information acquisition, these papers create a

³⁵ Notably, Interactive Viewer files are larger than Machine-Readable files because they contain additional code for formatting. This formatting code is identical for every 10-K filing on EDGAR and do not provide useful information for the computer. This needlessly increases the storage requirements for a 10-K filing. In addition, to obtain an entire filing via Interactive Viewer a program would need to download every individual file used by Interactive Viewer. This can result in hundreds of downloads for one 10-K filing. In comparison Machine-Readable files at most require five downloads. The size of the downloaded files and the number of files impacts the amount of computer time required to download a single 10-K filing which is impactful if a program is downloading thousands of 10-K filings.

³⁶ Specifically, I identify Interactive Viewer files by their naming convention. All Interactive Viewer files start with with a "R" followed by a number and ended with either ".htm" or ".xml". Interactive Viewer files before 2011 end with a ".xml" and 2011 and after end with a ".htm." Similarly, I identify Machine-Readable files by identifying all files that end with ".xml" that do not meet the naming criteria for Interactive Viewer.

definition for automated download activity and exclude these downloads from their variables of interests.

A key component to the automated download definitions used in Loughran and McDonald (2017), Drake et al. (2015), and Ryans (2017) is the classification of activity as automated if the corresponding IP address downloads an excessive number of files within a day (hereafter referred to as a high search filter). For instance, the definition of automated download activity implemented by Ryans (2017) categorizes an IP address as automated if the IP address downloads more than 500 files in a day or downloads files from more than 25 CIKs in a 3 minute period.³⁷

A large proportion of activity from ISPs is classified as automated download activity when papers utilize human/robotic download classifications with a high search filter. ISPs are companies such as Comcast, Version, Sprint, etc. that sell internet connectivity to residential consumers and small businesses. The vast majority of ISP IP addresses are categorized as automated via high search filter used in as Loughran and McDonald (2017), Drake et al. (2015), and Ryans (2017). Given that ISPs make up a considerable portion of total EDGAR downloads the high- search filter systematically classify most activity on EDGAR as automated.

One IP address from an ISP may represent the aggregate activity of many individual users rather than identifying robotic downloads. There exists a limited number of IPv4 address in the world. Given the dramatic increase in the number of connected individuals and internet enabled devices there has become a scarcity of IP addresses, specifically IPv4 addresses. In response to this scarcity, ISPs have implemented Carrier-Grade Network Address Translation to their

³⁷ Drake et al. (2015) uses less than 5 downloads per minute and less than 1,000 downloads per day as a definition of non-automated IP addresses. Loughran and McDonald (2017) treats any IP address with more than 50 daily downloads as automated.

networks to accommodate the growing number of internet users (Richter, Wohlfart, Vallina-Rodriguex, Allman, Bush, Feldmann, Kreibich, Weaver, and Paxson. 2016). This technology allows the ISP to conserve IP address by funneling activity from multiple users into a single IP address when interacting with websites such as EDGAR. Thus, if an ISP funnels multiple human users who are utilizing Interactive Viewer into the same IP address the high search filter is likely to categorize the entire IP address as automated due to requesting more than 500 files in a day. This study does not utilize an automated definition because the hypothesis predicts an increase in aggregate information acquisition. Additionally, my study does not use high search filter on EDGAR log because technological changes in the IT infrastructure along with how Interactive Viewer stores data has likely resulted in the high search filter categorizing real human users as bot during the timer period of interest.

Limitations

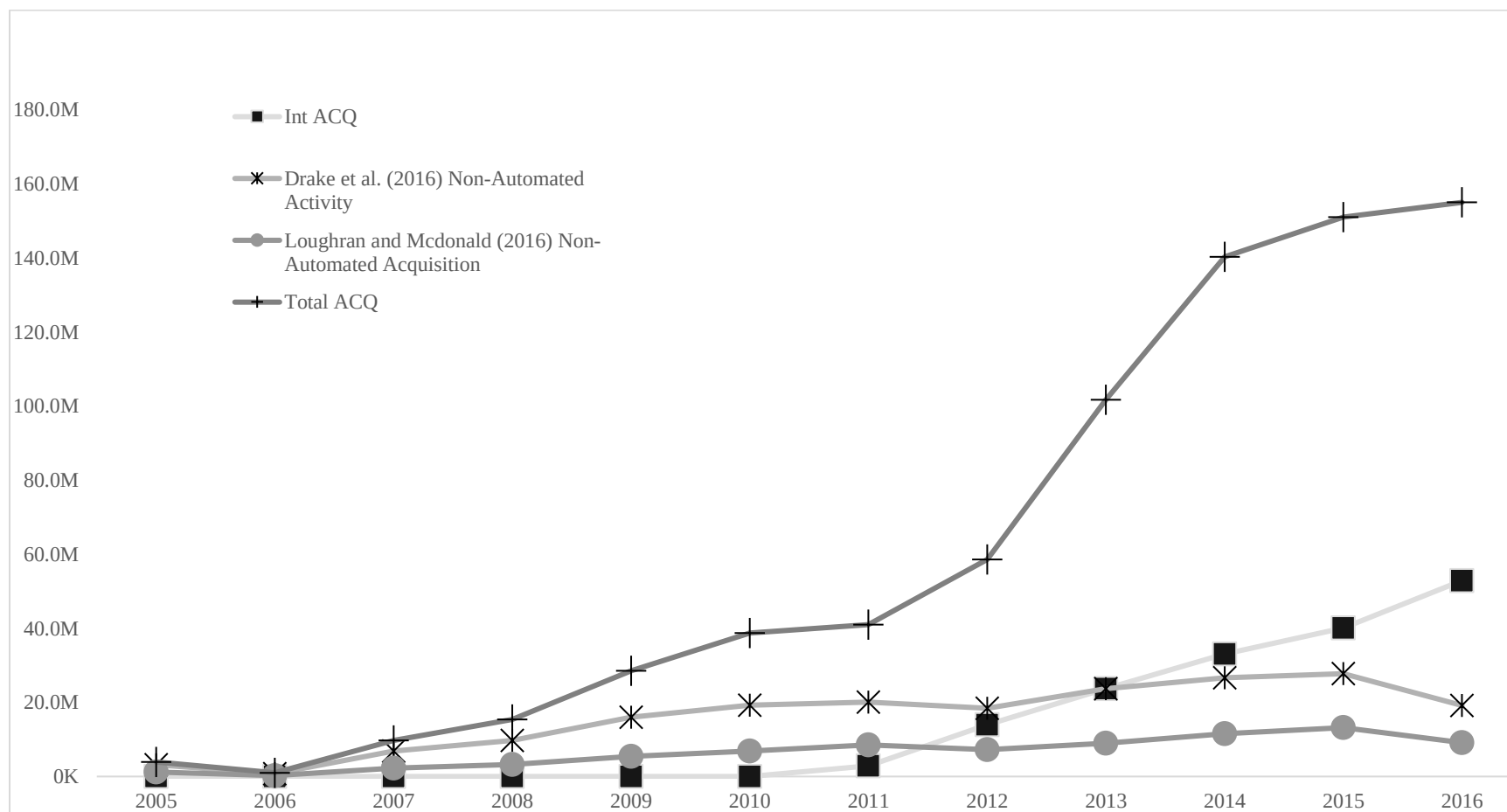
Given the nature of IP addresses there are inherent limitations to how trader activity on SEC EDGAR can be tracked. The large amount of internet users who utilize the service of an ISP makes the tracking of individual users on EDGAR over multiple days unreliable. Most individual traders will be dynamically assigned IP addresses either by their ISP or organization. Traders cannot be traced reliably day to day because every day the trader logs onto the internet they will be dynamically assigned a new IP address. In addition, given the implementation of Carrier-Grade Network Address Translation in ISP's networks one IP address from an ISP has a high likelihood to represent multiple users in one day. In this paper I count an access request of an IP address to a firm and channel on a day as one instance of information acquisition. This ensures that one user session on Interactive Viewer and other channels of the 10-K are only counted one acquisition event. However, if a trader accesses the same firm and channel on two different days

this study will count that acquisition twice. In addition, this measure will under count information acquisition if multiple users are assigned the same IP address and view the same 10-K via the same channel within a day. This means if two traders both view the same 10-K filing via Interactive Viewer on the same day and their ISP routes their activity into the same IP address then their will only be one recorded download for both users.

Descriptive Graphics

Figure B1 displays the measures used in this paper as compared to the commonly used measures in Drake et al. (2015) and Loughran and McDonald (2017). This figure combines all activity on EDGAR for the entire year for form 10-K. Notice that the Drake et al. (2015) and Loughran and McDonald (2017) do not trend upwards with the introduction of Interactive Viewer. This is because Interactive Viewer users are more likely to utilize the services of an ISP who's IP addresses get classified as automated via the high search filter utilized in Drake et al. (2015) and Loughran and McDonald (2017).

FIGURE B1: Information Acquisition Manual vs. Automated Definitions



BIBLIOGRAPHY

BIBLIOGRAPHY

- Admati, A. R., and P. Pfleiderer. 1988. A Theory of Intraday Patterns: Volume and Price Variability. *The Review of Financial Studies* 1 (1): 3–40.
- Allee K., N. Bhattacharya, E. Black, and T. Christensen. 2007. Pro forma disclosure and investor sophistication: External validation of experimental evidence using archival data. *Accounting, Organizations and Society*. 32 (3): 201-222
- Bhattacharya, N., Y. Cho, and J. Kim. 2018. Leveling the Playing Field between Large and Small Institutions: Evidence from the SEC’s XBRL Mandate. *The Accounting Review* 93 (5): 51–71.
- Bhushan, R. 1991. Trading Costs, Liquidity, and Asset Holdings. *The Review of Financial Studies* 4 (2): 343–360.
- . 1994. An informational efficiency perspective on the post-earnings announcement drift. *Journal of Accounting and Economics* 18 (1): 45–65.
- Blankespoor, E. 2019. The Impact of Information Processing Costs on Firm Disclosure Choice: Evidence from the XBRL Mandate. *Journal of Accounting Research* 57 (4): 919-967.
- Blankespoor, E., Ed deHaan, and I. Marinovic. 2020. Disclosure Processing Costs, Investors’ Information Choice, and Equity Market Outcomes: A Review. *Journal of Accounting and Economics*: 101344.
- Blankespoor, E., E. deHaan, and C. Zhu. 2018. Capital market effects of media synthesis and dissemination: evidence from robo-journalism. *Review of Accounting Studies* 23 (1): 1–36.
- Blankespoor, E., B. Miller, and H. White. 2014. Initial evidence on the market impact of the XBRL mandate. *Review of Accounting Studies* 19 (4): 1468–1503.
- Blankespoor, E., Ed deHaan, John Wertz, and Christina Zhu. 2019. Why Do Individual Investors Disregard Accounting Information? The Roles of Information Awareness and Acquisition Costs. *Journal of Accounting Research* 57 (1): 53–84.
- Blau, B., J. DeLisle, and S. M. Price. 2015. Do sophisticated investors interpret earnings conference call tone differently than investors at large? Evidence from short sales. *Journal of Corporate Finance* 31: 203–219.
- Bronson, S., C. Hogan, M. Johnson, & K. Ramesh (2011). The unintended consequences of PCAOB auditing standard nos. 2 and 3 on the reliability of preliminary earnings releases. *Journal of Accounting and Economics*, 51(1-2), 95-114.
- Brown, S. , and J. Tucker. 2010. Large-sample evidence on firms’ year-over-year MD&A modifications. *Journal of Accounting Research* 49 (2): 309–346.

- Chen, S., M. DeFond, and C. Park. 2002. Voluntary disclosure of balance sheet information in quarterly earnings announcements. *Journal of Accounting and Economics* 33 (2): 229–251.
- Chen, G., and J. Zhou. 2018. XBRL Adoption and Systematic Information Acquisition via EDGAR. *Journal of Information Systems* 33 (2): 23–43.
- Clogg C., E. Petkova, and A. Haritou. 1995. “Statistical Methods for Comparing Regression Coefficients Between Models” *American Journal of Sociology* 100 (5): 1261-1293
- Collins, D., O.Li, and H. Xie. 2009. What drives the increased informativeness of earnings announcements over time? *Review of Accounting Studies* 14 (1): 1–30.
- Correia, S. 2015. *Singletons, cluster-robust standard errors and fixed effects: A bad mix. Technical Note*, Working paper, Duke University.
- Debreceeny, R., S. Farewell, M. Piechocki, C. Felden, and A. Gräning. 2010. Does it add up? Early evidence on the data quality of XBRL filings to the SEC. *Journal of Accounting and Public Policy* 29 (3): 296–306.
- Dong, Y., O. Li, Y. Lin, and C. Ni. 2016. Does Information-Processing Cost Affect Firm-Specific Information Acquisition? Evidence from XBRL Adoption. *Journal of Financial and Quantitative Analysis* 51 (2): 435–462.
- Drake, M., B. Johnson, D. Roulstone, and J. Thornock. 2020. Is There Information Content in Information Acquisition? *The Accounting Review* 95 (2): 113–139.
- Drake, M., P. Quinn, and J. Thornock. 2017. Who Uses Financial Statements? A Demographic Analysis of Financial Statement Downloads from EDGAR. *Accounting Horizons* 31 (3): 55–68.
- Drake, M., D. Roulstone, and J. Thornock. 2015. The Determinants and Consequences of Information Acquisition via EDGAR. *Contemporary Accounting Research* 32 (3): 1128–1161.
- . 2016. The usefulness of historical accounting reports. *Journal of Accounting and Economics* 61 (2): 448–464.
- D’Souza, J., K. Ramesh, and M. Shen. 2010. Disclosure of GAAP line items in earnings announcements. *Review of Accounting Studies* 15 (1): 179–219.
- Dyer, T., M. Lang, and L. Stice-Lawrence. 2017. The evolution of 10-K textual disclosure: Evidence from Latent Dirichlet Allocation. *Journal of Accounting and Economics* 64 (2–3): 221–245.
- Easton, P., and M. Zmijewski. 1993. SEC Form 10K/10Q Reports and Annual Reports to Shareholders: Reporting Lags and Squared Market Model Prediction Errors. *Journal of Accounting Research* 31 (1): 113–129.

- Elliott, W., K. Rennekamp, and B. White. 2015. Does concrete language in disclosures increase willingness to invest? *Review of Accounting Studies* 20 (2): 839–865.
- Francis, J., K. Schipper, and L. Vincent. 2002. Expanded Disclosures and the Increased Usefulness of Earnings Announcements. *The Accounting Review* 77 (3): 515–546.
- Grinblatt, M., and M. Keloharju. 2000. The investment behavior and performance of various investor types: a study of Finland's unique data set. *Journal of Financial Economics* 55 (1): 43–67.
- Grossman, S., and J. Stiglitz. 1980. On the impossibility of informationally efficient markets. *The American economic review* 70 (3): 393–408.
- Jegadeesh, N., and S. Titman. 1993. Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency. *The Journal of Finance* 48 (1): 65–91.
- Kalay, A. 2015. Investor sophistication and disclosure clienteles. *Review of Accounting Studies* 20 (2): 976–1011.
- Kim, J., B. Li, and Z. Liu. 2019. Information-Processing Costs and Breadth of Ownership. *Contemporary Accounting Research* 36 (4): 2408–2436.
- Kim, O., and R. E. Verrecchia. 1994. Market liquidity and volume around earnings announcements. *Journal of Accounting and Economics* 17 (1): 41–67.
- Kyle, A. 1985 Continuous Auctions and Insider Trading. *Econometrica* 53, 6: 1315–335.
- Lee, C. 1992. Earnings news and small traders: An intraday analysis. *Journal of Accounting and Economics* 15 (2): 265–302.
- Lee, Y. 2012. The Effect of Quarterly Report Readability on Information Efficiency of Stock Prices*. *Contemporary Accounting Research* 29 (4): 1137–1170.
- Li, E. , and K. Ramesh. 2009. Market Reaction Surrounding the Filing of Periodic SEC Reports. *The Accounting Review* 84 (4): 1171–1208.
- Li, F. 2008. Annual report readability, current earnings, and earnings persistence. *Journal of Accounting and economics* 45 (2–3): 221–247.
- Li, R., (Wesley) Wang, Z. Yan, and Y. Zhao. 2019. Sophisticated Investor Attention and Market Reaction to Earnings Announcements: Evidence From the SEC's EDGAR Log Files. *Journal of Behavioral Finance* 20 (4): 490–503.
- Liu, C., T. Wang, and L. Yao. 2014. XBRL's impact on analyst forecast behavior: An empirical study. *Journal of Accounting and Public Policy* 33 (1): 69–82.
- Loughran, T., and B. McDonald. 2011. When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. *The Journal of Finance* 66 (1): 35–65.

- . 2017. The use of EDGAR filings by investors. *Journal of Behavioral Finance*: 1–18.
- Marshall, N., J. Schroeder, and T. Yohn. 2019. An Incomplete Audit at the Earnings Announcement: Implications for Financial Reporting Quality and the Market's Response to Earnings. *Contemporary Accounting Research* 36 (4): 2035–2068.
- Mahani, R., and A. Potesman. 2008. Overreaction to stock market news and misevaluation of stock prices by less-sophisticated investors: Evidence from the option market. *Journal of Empirical Finance* 15 (4): 635–655.
- Milian, J. 2015. Less-sophisticated Arbitrageurs and Market Efficiency: Overreacting to a History of Underreaction? *Journal of Accounting Research* 53 (1): 175–220.
- Miller, B. 2010. The effects of reporting complexity on small and large investor trading. *The Accounting Review* 85 (6): 2107–2143.
- Nelson, M., and K. Rupar. 2014. Numerical formats within risk disclosures and the moderating effect of investors' concerns about management discretion. *The Accounting Review* 90 (3): 1149–1168.
- Novy-Marx, R., and M. Velikov. 2015. A Taxonomy of Anomalies and Their Trading Costs. *The Review of Financial Studies* 29 (1): 104–147.
- Rennkamp, K. 2012. Processing Fluency and Investors' Reactions to Disclosure Readability. *Journal of Accounting Research* 50 (5): 1319–1354.
- Richter, P., F. Wohlfart, N. Vallina-Rodriguez, M. Allman, R. Bush, A. Feldmann, C. Kreibich, N. Weaver, and V. Paxson. 2016. *A multi-perspective analysis of carrier-grade NAT deployment*. (working) 215–229.
- Ryans, James, 2017. *Using the EDGAR Log File Data Set*. (last accessed September 15, 2020). Available at: <https://ssrn.com/abstract=2913612> or <http://dx.doi.org/10.2139/ssrn.2913612>
- Schroeder, J. H. 2016. The Impact of Audit Completeness and Quality on Earnings Announcement GAAP Disclosures. *The Accounting Review* 91 (2): 677–705.
- Securities and Exchange Commission. 2020. *What We Do* (last accessed November 19, 2020) Available at: <https://www.sec.gov/about/what-we-do>
- . 2018. *Inline XBRL Filing of Tagged Data*. (last accessed February 13, 2022) Available at: <https://www.sec.gov/rules/final/2018/33-10514.pdf>.
- . 2009. *Interactive data to improve financial reporting*. Vol. 232.405. (last accessed September 15, 2020) Available at: <http://www.sec.gov/rules/final/2009/33-9002.pdf>.

- . 1998. *A plain English handbook: How to create clear SEC disclosure*. SEC Office of Investor Education and Assistance. (last accessed November 19, 2020) Available at: <http://www.sec.gov/pdf/handbook.pdf>.
- Wooldridge, J. 1999. Distribution-free estimation of some nonlinear panel data models. *Journal of Econometrics* 90 (1): 77–97.
- Wooldridge, J. 2013. “Limited Dependant Variable Models and Sample Selection”. In *Introductory Econometrics: A Modern Approach*. 5th Edition. Boston, MA: Cengage Learning.