The Effect of XBRL on Insider Trading Profitability

by

Paulien Zhu
#368399

A THESIS SUBMITTED FOR THE AWARD OF
MASTER OF SCIENCE
IN ACCOUNTING & CONTROL
OF ERASMUS SCHOOL OF ECONOMICS

Supervisor: Drs. R. Van der Wal RA
Co-reader: Dr. P.Y.E. Leung

July 26, 2018

Abstract

In 2009 the SEC required that financial statements are filed in eXtensible Business Reporting Language (XBRL). The SEC argued that XBRL holds many benefits some of which are reducing information asymmetry and democratizing the information environment. However, prior research did not reach consensus on the effect of XBRL on the information environment. Using a difference-in-difference test this thesis contributes to the current discussion by providing evidence using a proxy that has not been used before: insider trading profitability. This thesis found that XBRL is associated with a decrease in insider trading profitability. Furthermore, the information gap between the best informed investors (insiders), the medium informed investors (institutional investors) and the least informed investors (individual investors) was found to decrease. This indicates that XBRL is associated with democratizing of the information environment.

Keywords: *XBRL, Disclosure, Information asymmetry, Insider trading profitability*

Foreword and Acknowledgements

For the past several months I have dedicated myself to the best extent possible to write this thesis. This thesis represents the ending of my master Accounting, Auditing and Control. I finally understand the utmost dedication and efforts that my predecessors put forward to obtain their master degree and why the master thesis is the embodiment of one's academic skills. To say that writing this thesis was hard is an understatement. However, looking back at the whole process, there was no better feeling than the rewarding feeling of overcoming each obstacle and to finally present you my master thesis.

I could have never accomplished this without the guidance, support and feedback of my supervisor Drs. R. Van der Wal RA. I want to thank him for his unconditional efforts, inspirational ideas and time that have helped me to continuously improve my thesis.

I also want to thank Dr. P.Y.E. Leung for her valuable input and ideas regarding my thesis proposal and the welcomed feedback on my thesis.

Furthermore, I would like to thank my family and friends for their support. I especially want to thank my parents for never stop believing in me and for their loving words of encouragement, my boyfriend for always coping with me during my stressful moments and my brother for being always by my side as my mental support.

Paulien B. Y. Zhu

Table of Contents

List of Abbreviations

AR	Abnormal return
CAR	Cumulative abnormal return
EBR	Enhanced business reporting
EFFH	Extended functional fixation hypothesis
H	Hypothesis
IAS	International Accounting Standards
M	Model
OLS	Ordinary least squares
RQ	Research question
SBR	Standard business reporting
SEC	Securities and Exchange Comission
U.S.	United States
U.S. GAAP	United States general accepted accounting principles
XBRL	eXtensible Business Reporting Language
XBRL-GL	eXtensible Business Reporting Language-General Ledger

1 Introduction

Electronic formats of financial statements have been mandatory in the U.S. for registered companies since 1996. Electronic financial statements led to wider distribution of financial statements at a lower cost. However, the distribution of financial statements have created problems regarding information exchange for companies. To counter this problem several electronic languages have been developed to increase the compatibility of financial statements, with the most recent development being eXtensible Business Reporting Language (XBRL). XBRL is designed to improve accuracy and reliability for electronic transmission of business and financial data. The SEC adopted rules in 2009 requiring firms to file their financial statements in XBRL. Companies with fiscal periods ending on or after June 15, 2009, were required to comply with the mandate if these companies use U.S. GAAP and have a worldwide public float that is greater than 5 billion dollars. By adopting XBRL the required disclosures are made more accessible and usable, helping the SEC to increase investors' protection, obtain more fair, orderly and efficient markets, and to better facilitate capital information. Other benefits of structured data are: better access to data, improved ways to manipulate data, more comparable cross-sectional disclosures, and better comparable inter-time disclosures, which results in better and easier analysis of financial reports (SEC, 2016). If XBRL is effective, then the abovementioned benefits will lead to increased quality of financial information (Baldwin et al. 2006). However, adopting XBRL was very costly for firms (Hannon, 2006). In addition, prior literature has shown that the initial XBRL filings contain many errors (Du et al., 2013). Furthermore, debates have been raised whether the early mandatory adoption of XBRL provides informational value (Bartley et al., 2011; Gunn, 2007). This tension creates interesting ground for research. This thesis examines whether the adoption of XBRL has improved the information environment. More specifically, this thesis aims to examine whether XBRL has reduced insider profits through increased quality of financial reporting information by providing an answer to the following research question:

RQ: Does XBRL reduces information asymmetry with respect to insider trading?

Particularly, this thesis contributes to the discussion of the effectiveness of XBRL as it is the first research that examines the effect of XBRL on information asymmetry with respect to insider profits. The results of this research question are not only useful for regulators but also for investors and other stakeholders. The results on this research question can be used as motivation for future regulation, not only regarding XBRL regulation but also regarding regulation on insider trading. Regulators can use the conclusion of the study as an argument to enforce the use of XBRL, especially when adopting XBRL leads to a decrease of (unfavorable) insider trading. Investors and other stakeholders can benefit from these results because on the effectiveness of XBRL they can decide whether and how to allocate resources, such as, human capital and other investments into XBRL. Investors and other stakeholders have more incentive to use XBRL Web services if it is proven that this will convey more value-relevant information. Furthermore, since structured data is perceived as the future of financial reporting, auditing and other financial services, the results of this research question can be valuable for the development of XBRL and other types of structured data in the future.

This thesis relates to two streams of literature. First, this thesis contributes to the literature regarding the effects of XBRL on information asymmetry. This relation is important for regulators because it provides insight in the effectiveness of XBRL. Prior researches have examined the effect of XBRL on information asymmetry using other proxies. Blankespoor et al. (2014) examined the effect of XBRL on bid-ask spread, price impact of trade and trading volume. Kim et al. (2012) investigated the relationship between XBRL on event return volatility, information efficiency, standard deviations of daily stock returns and analysts' forecast errors. These proxies above measure the willingness of the investors to trade at a lower cost, the liquidity and the information uncertainty. Whilst these proxies provide interesting insight into the effectiveness of XBRL on various aspects, they do not measure the value-relevant information that ought to be disclosed through XBRL filings. Efendi et al. (2016) examined the value relevant information as disclosed by XBRL by measuring the abnormal return. Similar to Efendi et al. (2016) this research examines the value relevant information of XBRL. However, it is the first research that provides insight into the discrepancy of value-relevant information between insider managers and outsider investors and whether this is reduced through XBRL filings.

This thesis hypothesize that XBRL is associated with decreased information asymmetry. In other words, this thesis deems that XBRL improves the information environment by closing the information gap between insiders and investors. Furthermore, this thesis examines whether XBRL holds greater benefits for institutional investors or individual investors. This thesis hypothesize that XBRL democratizes the information environment, resulting in less information asymmetry between institutional investors and individual investors. To examine this research question and test the proposed hypotheses, XBRL 10-K filings for both the pre-and post-mandatory XBRL periods will be used. The sample consists of companies which are classified as large accelerated filers by the SEC of which the initial XBRL 10-K filing was filed in fiscal year 2009. Each XBRL-filer is then matched with a non-XBRL filer. The sample size consists of 382 companies. Consistent with Blankespoor et al. (2014) the sample period consists of a pre- and post-XBRL period. The pre-XBRL period starts from 15 June, 2008 until 14 June, 2009 and the post-XBRL starts from 15 June, 2009 until 14 June, 2010. This thesis explores the effect of XBRL on information asymmetry by using a difference-in-difference OLS regression. Furthermore, this thesis aims to deepen the understanding of the effect of XBRL on information asymmetry by analyzing the effect of XBRL on investor sophistication. An OLS regression with an indicator variable on investor sophistication is utilized to explore this relation.

This thesis found that XBRL is negatively associated with insider trading profitability as proxy for information asymmetry. Furthermore, empirical evidence suggest that XBRL indeed leads to a smaller information gap between institutional investors and individual investors. These findings contributes to practice as well as to theory. Since investors benefit from XBRL, regulators and firms have increased incentive to adopt XBRL. For theory, these findings provide a better understanding of the dynamics of the information environment and suggest that heterogeneous investors indeed exists.

The remaining part of this research is divided in the following sections: *Background & Motivation* – this section discusses the development of XBRL and the motivation for this research; *Literature Review and Hypotheses Development* – this section reviews the existing literature on XBRL, information asymmetry and investor sophistication, and provides a comprehensive theoretical framework for the hypotheses; *Methodology & Measurement* – this section explains the research design and measurement of the variables; *Sample Selection and Data* – this section explains the sample selection process, descriptive statistics and testing of the OLS assumptions; *Results* – this section interprets and explains the results following the analyses; *Conclusion* – this section provides an overall summary of the thesis, the conclusion, limitations and suggestions for future research.

2 Background & Motivation

2.1 Introduction

This section examines the history, development and adoption of XBRL. Furthermore, this section explains how XBRL works and what the related benefits and risks are. Consequently, this section explains how to overcome these risks. Finally, this section discusses XBRL developments in The Netherlands and Europe.

2.2 History of XBRL

Since 1996 electronic formats of financial statements have been mandatory in the U.S. for registered companies. The availability of electronic versions made financial statements more accessible to the public at a lower cost. However, the distribution of electronic formats via the Internet have created difficulties for cross-company and cross-geographical information exchange. These difficulties arose from several challenges including inconsistencies of content, navigation and electronic filing formats. Furthermore, the systems that were used to display the financial statements were incompatible. Prior to the development of XBRL several other languages have been developed, despite the development of these languages a standard language specifically for financial reporting did not yet existed in the 90s. This has prompted Charles Hoffman to propose an open standard for digitizing financial reports, which resulted in the development of XBRL (Perdana et al. 2015).

2.3 XBRL Explained

XBRL is designed to improve accuracy and reliability for electronic transmission of business and financial data. XBRL is freely available and a global standard for structured data. Hence, an XBRL document is an electronic file that contains data on business information. All the values of the data are defined by using tags. These tags are essential to XBRL because the tags add structure to the documents and convey information to computer applications about what each piece of data means and how this relates to other data.[1] The XBRL taxonomy collects all the tags and gives a precise definition for each tag. The tagging allows for structured and interactive data. This type of data can be analyzed by widely available software without manual

[1] See Appendix I for an example of XBRL tagging.

processing. This results in access and manipulation of data by investors, analysts and regulators in one disclosure as well as allowing for cross-sectional comparison and longitudinal comparison of disclosures (Gunn, 2007). For instance, XBRL can be used to address tasks ranging from auto-populating analytical spreadsheets to sort through vast amounts of data and identify anomalies. In addition, software can also be used to enhance readability of structured data.[2]

2.4 XBRL Adoption in the U.S.

The SEC adopted mandatory XBRL filing in 2009. To make the adoption to XBRL filings easier for firms, the SEC decided to create a phased-in approach with three phases. During phase one companies with fiscal periods ending on or after June 15, 2009, were required to comply with the mandate if these companies use U.S. GAAP and have a worldwide public float that is greater than 5 billion dollars. In phase two with fiscal periods ending on or after June 15, 2010, the remaining large accelerated filers using U.S. GAAP are required to comply with the mandate. Finally, for fiscal periods ending on or after June 15, 2011, all remaining smaller domestic filers as well as foreign private issuers using IFRS are required to comply with the mandate. In addition, a phased-in approach for the tagging requirements for the companies that comply with the mandate is used. In the first year that a company files financial information in XBRL, the firm must provide detail tagging for each item in its face financial statements. The company must also provide block tagging for footnote disclosures and schedules. In the second year of XBRL filings, the firm must comply with the first-year requirements but also needs to include detail tagging of footnotes and schedules[3].

2.5 XBRL's Potential Benefits and Risks

Traditional financial statements summarize the financial position of the firm and its performance with some explanation and details on the disaggregated underlying data of the financial statements. The current technology and data-management processes allows internal managers to analyze high-level aggregated performance measures to the source elements of the underlying transactions and events. For external users a trade-off between "uniformity" and "flexibility" of financial statements has always existed. Uniformity of financial statements

[2] According to SEC https://www.sec.gov/structureddata/what-is-structured-data
[3] See Appendix I for examples on block tagging and detail tagging

maximizes perceived comparability after aggregation, whilst flexibility allows for more relevant and meaningful firm-specific information. The XBRL taxonomies allow for major applications of the tagged data for external users and its potential is enormous. Tagging the underlying accounting meta-data will structure the data more efficiently and more effectively resulting in timely and more detailed information. The XBRL data can be formatted to the user's own presentational structure because the data will be tagged at the elementary level. Thus, the need for uniformity and the presentational format will not be relevant (Harris & Morsfield, 2012). According to Alles et al. (2004) XBRL sublanguages can extend the purpose of XBRL financial reporting as an output of the reporting process to an input of the reporting process. For instance, XBRL-GL is the language for general ledger entries at the start of the reporting process. Direct links to these transactions allow for a broad spectrum of specialized and customized accounting queries to be answered real-time. Hence, XBRL enables real-time financial reporting, that is, with zero latency and information that is available at any time. One major benefit of real-time financial reporting is that much less time is needed to debate certain measurement, disclosure and presentation issues for regulators. However, Debreceny et al. (2010) found that a quarter of the initial XBRL filings by the 400 largest companies contain errors. Half of these errors were due to filers overlooking the debit or credit attribute that underlies the calculation relationship. A quarter of these errors occurred because one or more values of a calculation relationship were missing or extraneous. These errors can be overcome by validation and quality management techniques. To prevent these errors in the future, Debreceny et al. (2010) suggested for audited XBRL filings. Other risks related to XBRL are failing of compliance to the Edgar Filer Manual and for instance missing deadlines due to the additional work of tagging data points. Hence, the SEC granted companies a two-year modified liability period regarding their XBRL filings. Thus, a company would not be liable for the inaccuracies occurred in its XBRL filings for two years after starting to comply with the XBRL mandate if the company made a good faith effort to comply with the mandate. This modified liability period has expired for all companies after October 31, 2014 (PwC, 2011).

2.6 XBRL and Auditing

The risks that are related with the accuracy of XBRL emphasize the potential need for internal and external auditing of XBRL filings. Srivastava & Kogan (2010) argue that even if the SEC does not require third party auditing of XBRL instance documents, it is in the best interest for the public that these filings are being audited by external auditors. They argue that for real-time reporting to be effective, it is necessary to be followed with real-time assurance. According to Rezaee et al. (2001) XBRL-GL provides the ability for auditors to conduct continuous auditing[4]. Financial data can be tagged at the transaction level and once this data have been registered in the enterprise systems, auditors are able to directly retrieve information from the enterprise databases. Traditional financial statements will not be needed anymore. Other benefits of continuous auditing are: (1) electronically tracing the information systems of a company from the journal entry to the financial statement and (2) minimize error-prone activities by obtaining data directly from the information system (Perdana et al. 2015). Srivastava & Kogan (2010) developed a conceptual framework consisting of a set of assertions determining the quality of an XBRL document[5]. They believe that a conceptual framework just as with traditional financial auditing offers the most effective and efficient way to provide assurance on XBRL filings[6]. The framework states that the main assertion for an XBRL instance document is that it is a true representation of the electronic document filed with the SEC. Two main deficiencies can exist within an XBRL filing: first, data-deficiencies, meaning that the facts (data) that are marked-up are not accurate and second, meta-data deficiency meaning that the mark-up itself is incorrect (including mark-up deficiencies in the instance document and mark-up deficiencies in the XBRL taxonomy). The researchers argue that some of these assertions in the conceptual framework can be easily validated using XBRL automated software. However, other assertions will still require human analysis of intermediate level of expertise. The role of internal auditors will change too, as they should help companies understand the risks that are associated with XBRL by evaluating whether these risks have been appropriately addressed and whether the process is producing high quality XBRL filings. Finally, internal auditors should advise on how to improve the XBRL filing process (PwC, 2011).

2.7 Developments of XBRL in The Netherlands and Europe

The Netherlands have made a conscious effort to adopt XBRL. It needs to be noted that whilst a short analysis will be done on the developments of XBRL in The Netherlands, the scope of the thesis remains focused on the U.S. and that the quantitative analysis in this thesis uses U.S.

[4] See Appendix I for an illustration of the technical aspect of continuous auditing.
[5] See Appendix I for the conceptual framework for XBRL instance document as proposed by Srivastava & Kogan (2010).
[6] Note that this initial conceptual framework lays on the premise that financial statements have been audited accurately and hence, can be used as evidence for the audit of XBRL statements. In the future, when traditional financial statements do not longer exist, this conceptual framework needs to be revisited.

data. The institution that is responsible for XBRL[7] and its infrastructure in The Netherlands is Logius. Logius is a department of the Ministry of Internal Affairs that is responsible for the structuring, development and implementation of services that relates to an e-government. Logius facilitates the portals that are necessary for digital filings. SBR language is introduced in The Netherlands by governmental institutions, such as, tax authorities, the Chamber of Commerce and Statistics Netherlands and some banks that coorporate with SBR banks on credit reports (Belastingdienst, sd). The SBR taxonomy is captured in The Dutch Taxonomy Architecture (Nederlandse Taxonomie Architectuur). The first XBRL tax return was issued in 2011. From 2013 onwards companies were mandated to increasingly start filing their tax declarations, statutory accounts and credit reports with XBRL. For instance, the Dutch parliament has mandated that starting from 2016 small companies needs to file their financial accounts to the Chamber of Commerce electronically, from 2017 onwards medium-sized companies follow and starting from 2019 all private-companies are obliged to file their financial accounts in XBRL (Nitchman, 2015).

The European Commission has mandated, in its revised Transparency directive, that all companies (including public companies) are required to file their financial statements in an electronic format, such as, XBRL from 1 January 2020. The European Commission deems that a harmonized electronic standard will make reporting easier and facilitates accessibility, analysis and comparability of reports. Other benefits of a harmonized electronic standard are better information supply for investors, more transparency for civil society and reduced administrative burden for the issuers (European Comission, 2013). According to Deloitte Netherlands (2013) entrepreneurs can experience major benefits of which the most important one is time-efficiency. An estimated 50-80% of data can be re-used by different institutions making the processes to prepare the different reports much quicker. Despite the major benefits, XBRL has not yet reached its full potential in The Netherlands. The problem is in the supply and demand of the necessary software. According to Van Ardenne (2007) Dutch software developers are not willing to develop new suitable software for SBR because it is costly and the issuers of the XBRL filings (the companies) do not care enough about electronic filing mechanisms to invest in these types of software. However, Van Ardenne (2007) argues that once the initial investment is made, convergence of all XBRL documents should be relatively easy. This will yield benefits for developers and issuers. Developers can offer more products

[7] Specifically SBR, (i.e. Standard Business Reporting language, a specific language of XBRL for business purposes).

against lower costs and issuers are able to reduce administrative burden due to the increased software supply.

2.8 Summary

The SEC required a phased-in approach for mandatory XBRL adoption for all companies. Starting from 15 June 2009, large accelerated filers with a public float of more than $5 billion and reporting in U.S. GAAP were the first to file using XBRL. Furthermore, the European Union recently decided on adopting mandatory XBRL filings starting from 2020. Therefore, it is of great interest for regulators, companies, information intermediaries and investors what the effect is of XBRL. The tagging of financial data in an uniform language is the essence of XBRL. The uniformity and tagging of data at the elementary level is what makes XBRL of value. However, XBRL filings still contain errors and therefore, should be audited just like traditional financial statements. Consequently, the framework for auditing XBRL filings might differ from auditing traditional financial statements. In addition, the role of internal auditors might change into advisors of risks related to XBRL. Furthermore, recent developments of XBRL in The Netherlands and within the European Union also points toward mandatory adoption of XBRL.

3 Literature Review and Hypothesis Development

3.1 Introduction

This section starts with an analysis of existing literature on XBRL. First, a general overview of the existing literature is provided. Then, this section provides an in-depth discussion of the XBRL literature that is related to the information environment. Consequently, the theories on information asymmetry, agency theory and the extended functional fixation hypothesis are discussed. Furthermore, insider trading profitability as a proxy for information asymmetry shall be explained. The theories are linked to XBRL and insider trading profitability providing theoretical frameworks for hypothesis development, which is discussed in the final part of this section.

3.2 XBRL's Benefits in Literature

XBRL has been developed to solve the difficulties in exchanging and distributing financial information between different systems (Perdana et al., 2015). A broad body of researches have been performed on XBRL since the introduction of the concept in the beginning of the 21^{st} century. The beginning research on XBRL mainly focused on the conceptual explanation of how XBRL aids in financial data and information exchange. Later on, research focused more on the implementation of XBRL and the empirical investigations thereof (Perdana., 2015). In the broad spectrum of XBRL research several streams of studies exist: (1) XBRL's effect on accounting and good corporate governance; (2) XBRL's impact on auditing; (3) XBRL's impact on perceptual factors and decision making; (4) XBRL's adoption; (5) XBRL's technical development; and (6) educational related aspects of XBRL (Perdana et al. 2015). This thesis analyzes the effect of XBRL on the information environment which consists of the literature streams accounting, auditing and the effect of XBRL on investor's decision making. These first three streams of literature are classified by Perdana et al. (2015) into one main literature theme: XBRL's impact on business. Bonson (2001) identified different parties that could benefit from XBRL: companies, financial analysts, investors, regulators, data aggregators or financial information providers and software vendors. Companies benefit from XBRL because they only have to prepare their financial statements once. Afterwards users can download the financial statements in different formats and the financial statements can also be translated in multiple languages. Furthermore, XBRL boosts the transition to IAS for companies that report according to local standards (Bonson, 2001). Financial analysts, investors and regulators benefit from XBRL because it expands the possibilities of using the information from financial statements and XBRL allows for automatic analysis of financial statements. In addition, XBRL provides this group the capability to receive the financial information customized to their preferred formats (Bonson, 2001). Data aggregators and financial information providers benefit from XBRL because automatic data aggregation can lead to a reduction of preparation costs and errors (Bonson, 2001). Software vendors benefit from XBRL because XBRL increases congruence with other applications (Bonson, 2001).

From the parties that are identified by Bonson (2001), this thesis specifically focuses on external information users, which are analysts and investors. These users rely on financial statements to acquire information. Thus, reliable accounting processes and information systems are of high importance for these users to acquire information on the company because these processes and systems provide input for financial statements. Perdana et al. (2015) argues that XBRL benefits accounting processes in three ways: (1) XBRL integrates accounting processes and information supply chains; (2) XBRL improves accounting data; and (3) XBRL supports organizations trying to achieve good corporate governance. These XBRL benefits aid external investors because it results in more reliable accounting processes and information systems. Indeed, prior studies have documented XBRL's ability to improve accounting data, information integrity and information quality (e.g., Wagenhofer 2003; Bonson et al. 2008; Madden 2011; Vasarhelyi et al. 2012). Furthermore, Bonson et al. (2008) suggests that XBRL enables the development of both EBR and SBR, which in turn improves the reliability, accuracy and comparability of financial data. In experimental settings, researchers have found that XBRL increases transparency for investors contributing to their decision-making process. In addition, XBRL is documented to increase user's access to financial information, help investors to make better informed investment decisions and facilitate continuous and real-time auditing (Baldwin & Trinkle, 2011).

Table 1: Summary of XBRL Benefits

XBRL Benefits
• Single format preparation of financial statements, but offers end-user customization • Expands possibilities to use financial statements • Allows for automatic analysis of financial statements • Integrates accounting processes and information supply chain • Improves accounting data • Supports organizations to achieve good corporate governance

Table 1 summarizes the benefits of XBRL

3.3 Information Asymmetry and Agency Problem

Healy & Palepu (2001) argue that both entrepreneurs and investors would like to do business with each other. However, two problems occur between both parties: first, the 'information problem' and second, the 'agency problem'. The information or the 'lemons' problem occurs due to information asymmetry and conflicting incentives between entrepreneurs and investors. The following example illustrates the information problem: in a certain situation half of the businesses is 'good' and the other half is 'bad'. Both investors and entrepreneurs are rational and they analyze the value of the investments based on their own information. If investors cannot differentiate between 'good' businesses and 'bad' businesses then entrepreneurs can state that their 'bad' businesses are as valuable as the 'good' ones. When investors realize that entrepreneurs can benefit from the information asymmetry then they will value both 'good' and 'bad' businesses at the average level. This results in rationally undervaluing 'good' businesses and overvaluing 'bad' businesses.

Healy & Palepu (2001) state that the 'agency' problem occurs once the investors have invested their funds in the businesses. In agency theory information asymmetry is explained using the principal-agent relationship. Eisenhardt (1989) argues that due to (a) the different goals of the agent and the principal (self-interest of the agent); and (b) given that the principal does not know whether the agent behaves as agreed information asymmetry arises. For instance, the ones who run the businesses are the entrepreneurs and investors typically do not want to play an active role in the management of the business. Therefore, when investors have invested their funds in the businesses then the self-interested entrepreneur has incentives to use the investors' funds to his own benefits (Healy & Palepu, 2001).

One solution to this problem is corporate disclosure. Corporate disclosure is necessary for an efficient capital market (Healy & Palepu, 2001). Corporate disclosure is positively associated with efficiencies in the capital market since corporate disclosure provides investors with a widely available pool of knowledge for investment decisions. However, if the corporate disclosure is insufficient then information asymmetry arises, resulting in less efficient capital markets (e.g. Healy & Palepu, 2001; Yoon et al., 2011). Frankel & Li (2004) state that the demand for reducing information asymmetry has led to the creation of the 1934 Securities Act. Furthermore, Benston (1973) states that the argument for the 1934 Securities Act was to provide a fair and efficient capital market. Indeed, corporate disclosure enforced by regulators, such as, the SEC reduces information asymmetry in the capital markets (Greenstein & Sami, 1994; Hagerman & Healy, 1992; Leuz & Verrecchia, 2000). Diamond (1985) explains in his equilibrium model of information disclosure that all traders are better off when companies release public information due to: (1) savings of real resources that would have devoted to private information acquisition if public information would not have been released; and (2) improvement in risk sharing since public information makes the believes of traders more homogeneous and reduces the magnitude of speculation. Corporate disclosure can be improved by XBRL. Since XBRL increasing the reliability, comparability, accuracy and transparency of financial data (e.g., Wagenhofer 2003; Bonson et al. 2008; Madden 2011; Vasarhelyi et al. 2012).

3.4 XBRL and Information Asymmetry: Empirical Research

Hodge et al. (2004) conducted an analysis on an XBRL enabled search-engine in the context of recognition versus disclosure of stock-compensation. They found that non-professional financial statement users who had access to XBRL technology were better able to acquire and integrate information in their decision-making process. Similar results are found by Ahmadpour & Bodaghi (2010). Arnold et al. (2012) investigated the impact of tagged qualitative business presentations on non-professional and professional investors. The results indicate that the tagged presentation facilitates incorporation of information on risk into investment decisions for both non-professional as professional investors. Despite these findings, concerns exist relating to the information reported in the XBRL financial statements, such as, filing errors and comparability of XBRL financial statements. Errors were found in the filings up to 1 September, 2009, however, the amount of errors decreased between SEC's voluntary XBRL filings program and the mandatory program (Debreceny et al. 2010; Bartley et al. 2011). The following studies focused on bid-ask spread, trading volume and price volatility as proxies for information asymmetry: Yoon et al. (2011) found that information asymmetry decreased in Korea after

XBRL adoption, whereas Blankespoor et al. (2014) and Cong et al. (2014) concluded the contrary for the U.S. market. Kim et al., (2012) found that mandatory XBRL adoption in the U.S. has the ability to decrease information risk and decrease information asymmetry through increased transparency of financial information. Liu et al. (2017) conducted research on the impact of XBRL in a European setting, specifically Belgium and concluded that liquidity increased whilst information asymmetry decreased after XBRL adoption. Furthermore, Liu et al. (2017) found that the increase for liquidity was higher for relatively large firms. Yen & Wang (2015) investigated the earnings surprises of a sample of XBRL filers around 10-Q and 10-K filing dates and found a positive relationship between XBRL adoption and market reaction to earnings surprises for Phase II and Phase III filers, suggesting that XBRL enables information access and processing. Liu et al. (2014a) conducted research on mandatory XBRL adoption in China and found that the expected benefits from XBRL might be hindered due to the implementation and development of the technology. Liu et al. (2014b) found that the mandatory adoption of XBRL was positively related with analyst following, resulting from a decrease in information processing costs and increased transparency of a firm due to XBRL. Furthermore, a positive association was found between analyst accuracy and mandatory XBRL adoption, suggesting increased quality of disclosure. Another study done by Liu et al. (2014c) found a decrease in analyst forecast accuracy in the early adoption period of XBRL in China. However, it needs to be noted that law enforcement affects the quality of disclosure. Furthermore, this negative association could also be the result of the early adoption of XBRL. Thus the implementation and development of XBRL technology could affect the results (Liu et al, 2014a). Evidence from post-earnings announcement drift showed that the adoption of XBRL was associated with a decrease in post-earnings announcement drift (Efendi et al. 2014). Empirical evidence on the impact of XBRL on information asymmetry shows positive and negative results and has not reached consensus yet. Furthermore, a potential trade-off between company-specific XBRL financial data and cross-industry comparability of XBRL financial data exists, resulting in less comparable cross-industry data (Zhu & Hu, 2011). It is possible that due to these reasons the evidence from empirical researches on the effect of XBRL on information asymmetry are mixed. As the discussion on the effect of XBRL continues, this thesis aims to add value to the existing literature by finding empirical evidence of XBRL on information asymmetry from a unique perspective: it is the first thesis that uses insider trading as a proxy for information asymmetry in the XBRL context.

3.5 Insider Trading as Proxy for Information Asymmetry

Following Frankel & Li (2004) this thesis defines insiders as the company's "management", such as, directors, managers, officers and presidents. The theory behind insider trading profits as proxy for information asymmetry is as follows: insiders can profit from value-relevant information when they trade on this before public disclosure results in full incorporation into stock prices. Meaning that when information asymmetry between insiders and outsiders increases then insiders will have more value-relevant information compared to outsiders, resulting in increased insider profits (Frankel & Li, 2004). Furthermore, Kyle's (1985) seminal paper predicts a positive relationship between information asymmetry and the abnormal profits resulting from insider trading. The model of Baiman & Verrechia (1996) shows that insider profits decrease when disclosure increases.

Huddart & Ke (2007) explain that insider trading consists of two aspects. First, the information advantage of insiders over other market participants. Therefore, insider trading may be driven by the insider's desire for profit from their superior information about the company's prospect. The other aspect of insider trading is the insider's specific information at a certain point of time which cannot be observed. However, the actual trading profits are able to be observed. These trading profits are driven by two components: (1) the uncertainty of the market regarding the firm's value and (2) the precision of the insider's information advantage compared to outsiders. Since insider trading profits are a result of the uncertainty of the firm's value and the precision of insiders' information advantage, temporarily mispricing of securities by outsiders can lead to insider trading gains. The higher the uncertainty the more an insider can profit from his perfect information. However, the insider can only profit from his perfect information when the precision of this information is higher than the outsider's information. Hence, the better the information advantage compared to the outsider the higher the profits resulting from insider trading. Aboody et al. (2005) find in their empirical study that asymmetric information is positively associated with insider trading profits, whilst Frankel & Li (2004) explain that various factors limit the ability of insiders to create insider gains. Uninformed traders can alter their trading behavior or even leave the market. Uninformed traders can acquire information by themselves or via intermediaries. For instance, high information asymmetry can make private information acquisition more profitable and thus increases the analyst following of the firm (Barth et al., 2001). Increased private information acquisition results in decreased profits from trading and hence, in equilibrium information acquirers earn a normal rate of return on their trades. Furthermore, competitiveness among insiders themselves can also decrease insider gains. In addition, organizational economics argue that organizations want to maximize their profits so that greater benefits can be shared. However, when management is involved in insider trading then this could prompt them to make business decisions that are less efficient. Hence, management can be punished for the inefficient decisions they make when engaged in insider trading, which reduces the incentive for insider trading (Frankel & Li, 2004). Corporate policies or governmental regulations may too restrict insider trading. Prohibitions on insider trading are placed to protect investors against information asymmetry in financial markets (Replogle, 2011). U.S. federal laws, such as, The Securities Act of 1933, The Securities and Exchange Act of 1934 and The Sarbanes-Oxley Act of 2002 are all legislation that affect insider trading. Despite these factors some researches find that insider trading is still lucrative. Insider trading that is done on material nonpublic information that the trader was aware of when the trader was performing his or her duties at the company is considered illegal. However, some forms of insider trading, when it is done without taking advantage of material nonpublic information and properly reported to SEC, are considered legal. The difference between illegal and legal insider trading is the degree of materiality of nonpublic information. Suggesting that legal insider trading still contain some degree of nonpublic information, in other words, information asymmetry between the insider and the outsider (Replogle, 2011). Indeed, previous studies found that abnormal returns follow insider trading (e.g., Burgstahler & Eames, 2006; Degeorge et al., 1999; Matsumoto, 2002).

Figure 1 illustrates the theoretical framework for the first hypothesis. The benefits of XBRL leads to increased reliability, accuracy, comparability and transparency of financial data, which improves corporate disclosure. Better corporate disclosure reduces information asymmetry. Hence, based on the previous literature review on XBRL and its potential benefits, this thesis expects that XBRL is negatively associated with information asymmetry thus insider trading profits.

H1: XBRL is negatively associated with information asymmetry

Figure 1: Illustration of H1

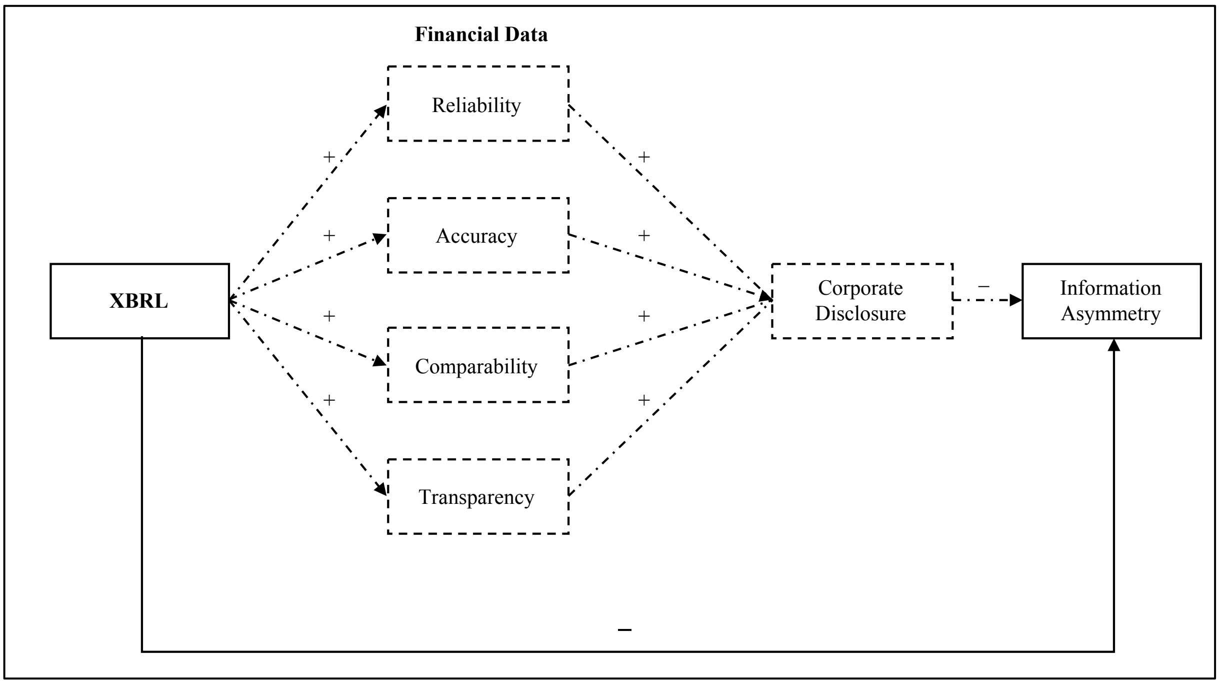


Figure 1 illustrates and explains the theoretical relation between XBRL and information asymmetry. Hypothesis 1 expects a negative association between XBRL and information asymmetry. This negative association is explained by theory and evidence from prior literature of the effect of XBRL on financial data and consequently on corporate disclosure

3.6 Investor Sophistication and Insider Trading Profits

The extended functional fixation hypotheses (EFFH) as proposed by Hand (1990) suggests that investors are heterogeneous when reacting on different accounting policies. The EFFH argues that a firm's stock price can sometimes be set by more-sophisticated investors (investors who have more knowledge about accounting data and the interpretation thereof) and other times set by less-sophisticated investors (investors with lesser knowledge on financial data). A sophisticated investor is a natural person or an entity that qualifies under Securities Act of 1933[8] (SEC, 2014). Bartov et al. (2000) defines a sophisticated investor as "a wealthy individual or institutions, such as, pension plan, insurance company, charitable institution or bank. Individuals must earn over $200,000 per year or own over $1 million in net assets". Following prior research sophisticated investors are classified as institutionalized investors (e.g., Hand, 1990; Walther, 1997; Ali et al., 2000; Bartov et al., 2000). Indeed, prior research suggest that institutional investors have better means compared to individual or less sophisticated investors to analyze financial information and corporate disclosure. Hence, they are more capable in pricing the securities than individual investors. Furthermore, institutional investors can be

[8] The term "Accredited investor" is defined under the Securities Act of 1933.

quasi-insiders or outsiders, this depends on their trading behavior and the extent of ownership they have within the company. Thus on average institutional investors have superior information compared to less sophisticated investors. It needs to be noted that insiders have by definition the most information about the firm (Piotroski & Roulstone, 2004). Therefore, Piortroski & Roulstone (2004) suggest that the information gap between insiders and institutional investors is smaller than between insiders and less sophisticated investors. Furthermore, Bushee & Goodman (2007) find that institutional investors consistently trade on private information. Therefore, the information asymmetry between institutional investors and insiders is expected to be of lesser degree than the information asymmetry between individual investors and insiders. As insiders are expected to make profits from superior information advantage over other investors, it is expected that the greater the information advantage, the greater the difference in trading profit will be. Hence, the difference in trading profits between insiders and individual investors is larger than the difference in trading profits between insiders and institutional investors.

The benefits of XBRL (as discussed in *Section 3.2*) are expected to democratize the capital market, making financial information more accessible (Blankespoor et al. 2014). Therefore, it is expected that XBRL filings aid individual investors better in analyzing financial information data than institutional investors, who already have more means to analyze financial data. Hence, XBRL is expected to improve the degree of information asymmetry between individual investors and insiders, and individual investors and institutional investors. Hence, this thesis expects that XBRL democratizes the financial market. Thus the information gap between institutional investors and individual investors decreases after initial XBRL 10-K filings as illustrated by figure 2.

H2: XBRL is associated with a decrease of information asymmetry between individual and institutional investors

Figure 2: Illustration of H2

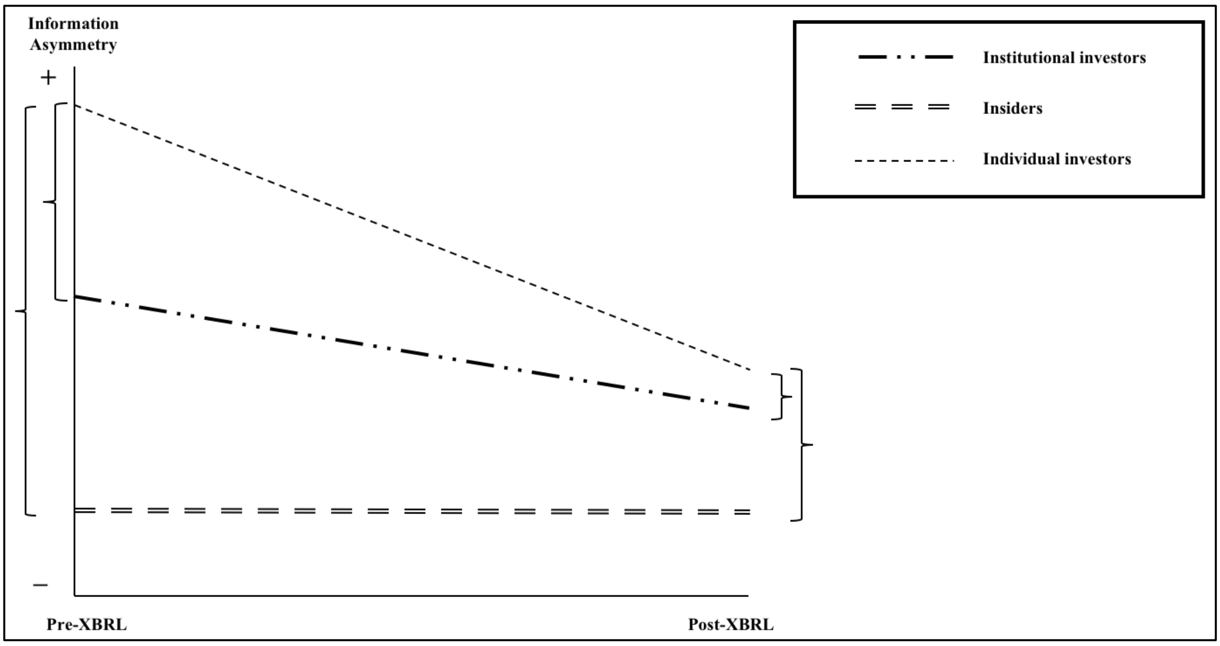


Figure 2 illustrates the hypothetical effect of XBRL on information asymmetry between insiders, institutional investors and individual investors. The gaps between insiders and outsiders (institutional and individual investors) decreases after initial XBRL filings. The information asymmetry between institutional investors and individual investors decreases after initial XBRL filings

3.7 Contribution

In the wake of the mandatory adoption of XBRL in the U.S. and the efforts of the European Commission and the Dutch government to adopt XBRL, one important question arises: does XBRL adoption yield any empirical benefit? Although XBRL has, in theory, great potential for users of financial statements, implementation was costly (Van Ardenne, 2007). Furthermore, XBRL does not come without deficiencies (Srivastava & Kogan, 2010). This thesis contributes to the discussion on the value of XBRL and aims to empirically validate the benefits of XBRL. This thesis is the first to explore the consequences of XBRL on information asymmetry with respect to insider trading and contributes to the existing literature by providing the first empirical study of the effect of XBRL on insider trading. Furthermore, the results of this question benefit several parties. The results on this research question can be used as motivation for future regulation, not only regarding XBRL regulation but also regarding regulation on insider trading. Regulators can use the conclusion of the study as an argument to enforce the use of XBRL especially when adopting XBRL leads to a decrease of (unfavorable) insider trading. Investors and other stakeholders can benefit from these results because based on the degree of effectiveness of XBRL they can decide whether and how to allocate resources, such as, human capital and other investments to XBRL. Investors and other stakeholders have more incentive to use XBRL Web services if it is proven that it will convey more value-relevant information. Furthermore, since structured data is perceived as the future of financial reporting, auditing and other financial services, the results of this research question can be valuable to the development of XBRL and other types of structured data in the future.

3.8 Summary

This thesis focuses on the effect of XBRL on the information environment. Prior studies have suggested many benefits of XBRL. XBRL is argued to improve reliability, comparability, transparency and accuracy of financial data. This leads in turn to improved corporate disclosure. Better corporate disclosure is a solution to information asymmetry and agency problems. Despite the theoretical benefits, empirical research on the effect of XBRL on information asymmetry provided mixed results. Related research found that insider trading profitability is associated with increased information asymmetry. Hence, this thesis hypothesize that XBRL is negatively associated with information asymmetry. This thesis contributes to the existing literature as being the first study to shed light on the effect of XBRL on insider trading profitability as a proxy for information asymmetry. Furthermore, XBRL is believed to democratize the financial markets. According to the EFFH financial markets consist of heterogenous investors each with a different level of knowledge and access to financial information. This thesis expects that the introduction of XBRL reduces the information gap between individual investors, institutional investors and insiders. This thesis further analyzes the effect of XBRL on the information content of 10-K filings on institutional and individual investors.

4 Measurement and Methodology

4.1 Introduction

This section explains the research design to test the hypotheses (as discussed in *Section 3.5 and 3.6*). Furthermore, the dependent, independent and control variables will be discussed. Then this section elaborates on the predictive framework and validity of the hypotheses. Following this analysis the potential endogeneity problems will be discussed.

4.2 Research Design

4.2.1 XBRL and insider trading profitability

To examine the specific impact of XBRL on insider gains, this thesis uses a difference-in-difference model:

$$CAR_{0,180}^{i} = \alpha_0 + \alpha_1 XBRL + \alpha_2 POST + \alpha_3 POST * XBRL + \alpha_i CONTROLS + \varepsilon \qquad (1)$$

CAR represents the cumulative abnormal returns in the [0, 180] after an insider trading event (*i*). It captures insider trading profitability: a positive *CAR* indicates higher profitability from insider trades.

XBRL is an indicator variable that takes the value one if the firm was required to adopt XBRL and zero for control firms.

POST is an indicator variable that is equal to one if the firm's fiscal year ends between 15 June, 2009 and 14 June, 2010. Filings of which the fiscal year is between 15 June, 2008 and 14 June, 2009 are classified as zero.

*POST*XBRL* captures the effect of XBRL on insider gains. When the coefficient is positive (negative) and significant then XBRL is associated with an increase (decrease) of insider trading profitability.

CONTROLS represent the control variables: *N_An* (number of analysts), *Lsize* (log of firm size), *M_B_ratio* (market to book ratio), *RND* (research and development expenditures), and *STD_RET* (standard deviation of stock return).

Section 4.3 provides an in-depth discussion of these variables and their measurement.

4.2.2 XBRL and investor sophistication

For the second hypothesis the following model is used:

$$CAR_{0,180}^{i} = \beta_0 + \beta_1 XBRL + \beta_2 ISH + \beta_3 ISH * XBRL + \beta_i CONTROLS + \varepsilon \qquad (2)$$

CAR represents the cumulative abnormal returns in the [0, 180] after an insider trading event (*i*). It captures insider trading profitability: a positive *CAR* indicates higher profitability from insider trades.

XBRL is an indicator variable that takes the value one if the firm was required to adopt XBRL and zero for control firms.

ISH indicates investor sophistication as the percentage of the firm's total shares outstanding that are held by institutional holdings. Investor sophistication increases (decreases) when the percentage of the firm's total shares outstanding held by institutional holdings increases (decreases).

*ISH*XBRL* is an interaction variable that captures the interaction effect of institutional holdings and XBRL on insider trading profitability. When the coefficient is positive (negative) and significant then institutional holdings from XBRL firms are associated with an increase (decrease) of insider trading profitability.

CONTROLS represent the control variables: *N_An* (number of analysts), *Lsize* (log of firm size), *M_B_ratio* (market to book ratio), *RND* (research and development expenditures), and *STD_RET* (standard deviation of stock return).

Section 4.3 provides an in-depth discussion of these variables and their measurement.

4.3 Variable Explanation

4.3.1 Dependent variable

The dependent variable insider trading profits is measured using cumulative abnormal return as the proxy for insider trading profits. Fama (1998) state that CAR is a popular and reliable measurement of long term returns. The SEC requires corporate insiders (e.g. company's officers and directors) that own the company's securities to file a statement with the SEC regarding these securities. Form 3,4, and 5 are the corresponding files. Form 3 needs to be filed during an initial registration of an insider's ownership. Form 4 is ought to be filed whenever the security's ownership changes. Form 5 is filed to report any transaction that should have been reported earlier using Form 4. These insider trading data are collected through the Thomson Reuters

Insider Data Filing Feed via WRDS. Following Frankel & Li (2004) several filters are applied to the insider transactions to obtain informative transactions. Non-management insider transactions are deleted since this thesis focuses on information asymmetry between "management" and investors. Furthermore, share prices of less than 2 dollar and transactions with less than 100 shares are deleted to eliminate noise (Lakonishok & Lee, 2001).

The remaining transactions are netted at company level for each transaction day. The net transaction is a sale when the transaction is negative or a purchase if the transaction is positive. To calculate the cumulative abnormal returns: first, the abnormal return (AR) is estimated for each trading day over the 180 calendar days after the transaction date for each net transaction. ARs are computed over a six-month horizon since insiders get penalties for profits earned on transactions made less than six months subsequent to prior transactions; Rule Section 16(b) for the Securities and Exchange Act of 1934 (Agrawal & Jaffe, 1995). The AR is computed as the difference between the firm's stock return and the value-weighted market return (Agrawal & Nasser, 2012). The AR of stock *i* on day *t* is calculated as follows:

$$AR_{it} = r_{it} - r_{mt} \tag{1}$$

r_{it} is the return for firm *i* on day *t* and r_{mt} is defined as the return of the CRSP value-weighted stock index on day *t*.

Then for each net transaction the cumulative abnormal return (CAR) is computed over days [0,180]. This is computed as follows:

$$CAR^{i}_{0,180} = \sum_{t=0}^{t180} AR_{it} \tag{2}$$

Since negative CARs indicate losses, negative CARs following insider sales indicate that insiders were able to avoid to experience (more) losses from the stocks. In other words, insiders gained from selling the stock. Thus negative CARs following insider sales are multiplied by negative one. The CARs are obtained through Eventus event-study software via WRDS. The estimation options that were used to calculate the CARs in the Daily Cross-Sectional Analysis in Eventus are as follows: the estimation period ranges from minimum 3 days to maximum 255 days that ends 10 days before the event date, where the transaction date is the event date [day 0].

4.3.2 Independent variables

The following independent variables are included in the first model:

XBRL is an independent and indicator variable that equals one for companies which filed their initial 10-K XBRL filings for the fiscal year ending on or after 15 June 2009 and zero otherwise. The 10-K filings are downloaded from EDGAR (as discussed in *Section 5.2*).

POST is an independent and indicator variable that equals one for each 10-K filing on or after 15 June 2009 and zero otherwise.

*POST*XBRL* is the interaction variable and the variable of interest. *POST*XBRL* captures the effect of XBRL on insider gains. When the coefficient is positive (negative) and significant then XBRL is associated with an increase (decrease) of insider trading profitability.

In the second model the following independent variables are included:

XBRL is an independent and indicator variable that equals one for companies which filed their initial 10-K XBRL filings for the fiscal year ending on or after 15 June 2009 and zero otherwise. The 10-K filings are downloaded from EDGAR (as discussed in *Section 5.2*).

ISH measures the investor sophistication. Bartov et al. (2000) state that institutional investor holdings of a stock has been greatly used in the literature as a proxy for investor sophistication. One reason why institutional holdings have been used as a proxy for investor sophistication is because institutional holdings have better resources in gathering and analyzing information. Hence, institutional holdings have an advantage over individual investors. Data on investor sophistication is obtained from the Thomson Reuter Stock Ownership database.

*ISH*XBRL* is an interaction variable that captures the interaction effect of institutional holdings and XBRL on insider trading profitability. When the coefficient is positive (negative) and significant then institutional holdings from XBRL firms are associated with an increase (decrease) of insider trading profitability.

The difference in the coefficients of *ISH* and *ISH*XBRL* shows the effect of XBRL on the effect of institutional holdings on CAR.

4.3.3. Control variables

In addition to the matching of firms on industry and firm size, and industry and analyst following, this thesis includes analyst following, market capitalization and other control variables that are known to be correlated with information asymmetry.

Number of analysts: Following Blankespoor et al. (2014) this thesis includes the number of analysts to control for the information environment of the firm. For fiscal year 2009 the maximum amount of analysts for the fiscal year is obtained through the I/B/E/S database via WRDS.

Size: Similar to e.g. Blankespoor et al. (2014) and Lakonishok & Lee (2001) firm size is included as control variable. Firm size affects insider trading activities as it is more beneficial for insiders to time an index of small stocks than of large stocks. Firm size is defined as the market capitalization of the firm in the most recent fiscal year and data is obtained through AuditAnalytics via WRDS.

Market-to-book ratio: Similar to Lakonishok & Lee (2001) market-to-book ratio is included as control variable, since market-to-book ratios can affect insider trading behavior regarding the sale and purchase of the stocks. The fiscal year 2009 market-to-book ratio is obtained through Compustat via WRDS.

R&D expenditure: R&D expenditure is an indicator variable that equals one if the firm reports R&D expenditures in the current fiscal year and zero otherwise. This thesis includes R&D expenditures since, firms with R&D expenditures have greater information asymmetry than firms without R&D expenditures (Aboody & Lev., 2000; Frankel and Li, 2004; Aboody & Lev., 2000); Blankespoor et al., 2014). Data is obtained through Compustat via WRDS.

Standard deviation of stock return: Similar to e.g. Bettis et al. (2000) and Frankel & Li (2004) this thesis controls for the standard deviation of stock return. Data for daily stock return over 180 days until one day prior to the event date are obtained from the TAQ database then the standard deviation of these stock returns are computed in STATA.

4.4 Predictive Validity and Endogeneity

4.4.1 Predictive validity framework

Figure 3 and 4 provide the predictive validity as description for the hypotheses testing process. According to Libby et al. (2002) no theory can be tested directly. Instead the theory can be tested by evaluating the relation between the operational definition of the theoretical concepts. By using these predictive validity frameworks researches can be designed efficiently and effectively. An efficient research is a research that is effective as possible using the least amount of resources. An effective research is a research that supplies evidence of adequate internal validity that convinces readers to believe the results of the hypothesis tests. Furthermore, an effective research provides enough external validity that it supports a significant part of the financial accounting issue of interest (Libby et al., 2002).

The first link (figure 3 and figure 4) dives into the specification of a good research question and hypothesis development. The hypotheses must have external validity. External validity refers to the degree that results can be generalized beyond the tasks, measurement methods or participants employed in the research. The hypotheses employed in this research are based on an in-depth analysis of existing literature and theory (as discussed in *Section 3*). Therefore, this thesis deems that the hypotheses as depicted in the first link of figure 3 and figure 4 are sound and externally valid. Furthermore, Libby et al. (2002) suggests that a theory that describes a causal relation increases the external validity. The theoretical framework of the hypotheses in this thesis imply causality between the independent and dependent variable, which also results in improved external validity.

Link 2 and 3 (figure 3 and figure 4) are relevant for the internal validity of the predictive validity framework. Internal validity refers to the degree that one can be certain that the observed effects of the research are the result of the independent variable. According to Libby et al. (2002) an internally valid research requires that each independent variable is manipulated, so that only one theoretical antecedent changes. This thesis tests two different hypotheses each with a distinctive theoretical concept and independent variable. Thus by testing these two separate hypotheses, this thesis meets this requirement for internal validity. However, the independent variables of both hypotheses are measured instead of manipulated, which according to Libby et al. (2002) lowers the internal validity due to correlated-omitted variables. It is impractical and outside the scope of this thesis to manipulate the effect of XBRL on large scale, in such circumstances it is justified to measure the independent variable instead of manipulate them (Libby et al., 2002). Another important aspect of link 2 and 3 is construct validity. Construct validity refers to the degree to which a test measures what it intends to measure (Cronbach & Meehl, 1955). This thesis chose the measurement of the independent and dependent variables based on an extensive research of prior studies and their methodologies (see *Section 4.3*). Hence, it is expected that the constructs indeed measures the independent and dependent variables as intended.

Link 4 estimates the relation between the operational independent variable and the dependent variable through statistical tests and modeling (see S*ection 4.2*). Link 5 captures other potentially influential variables (exogeneous) that might affect the relation between the independent variable and the dependent variable (Libby et al., 2002). This research controls for these exogeneous variables by adding them to the research design as control variables and employing a difference-in-difference research design that matches each subject with a control subject (see *Section 4.3.3*).

Figure 3: Predictive validity framework of H1

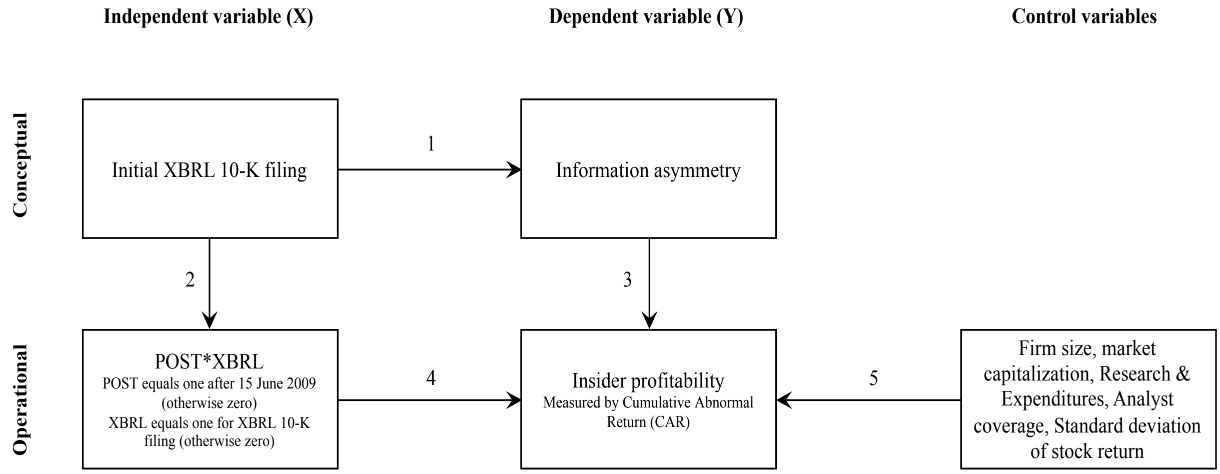


Figure 3 depicts the predictive validity framework of H1

Figure 4: Predictive validity framework of H2

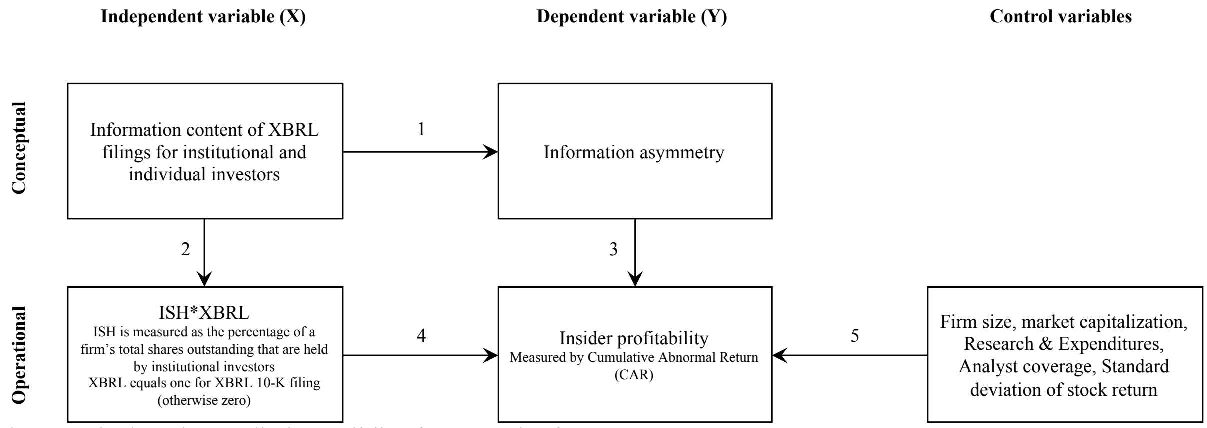


Figure 4 depicts the predictive validity framework of H2

4.4.2 Endogeneity issues

Endogeneity refers to the issue where an explanatory variable is correlated with the residuals (*Section 5.4.2* tests for endogeneity issues in the datasets used for hypothesis testing). Endogeneity may occur due to various reasons, some common sources of endogeneity are: omitted variables bias, measurement error in the independent variable, reverse causation or simultaneity and autocorrelation (in time-series) (Wooldridge, 2009).

Omitted variables bias: Frankel & Li (2004) suggest that return variance can be a correlated-omitted variable in the information environment. Return variance may be correlated with insider trading profitability. When managers can accurately (on average) predict the sign of future return then an increase in return variance may imply an increase in the manager's expected profitability. Bettis et al. (2000) reported that return variance is positively correlated with insider trading profitability. Furthermore, the firm's market-to-book ratio may also affect the information environment as firms with high market-to-book ratios have more unreported assets, leading to increased information asymmetry (e.g. Bettis et al., 2000; Frankel & Li, 2004). In addition, a firm's research and development expenditures may also be a correlated-omitted variable (e.g. Dai et al., 2012; Frankel & Li, 2004; Aboody & Lev., 2000). Aboody & Lev (2000) argue that the uniqueness of each firm's R&D expenditures make it difficult for outsiders to assess the productivity and value of a firm's R&D because it is difficult to benchmark with other companies. Hence, increasing the information asymmetry. Firm size can potentially affect insider trading profitability as insiders have a relative advantage in timing an index of small stocks than for large stocks (Lakonishok and Lee, 2001). Furthermore, the analyst coverage (number of analysts) can affect the information environment, since firms with higher analyst coverage may have richer information environments. Thus leading to a decrease of information asymmetry (Blankespoor et al., 2014). Despite controlling for these correlated-omitted variables (see *Section 4.3.3*), this thesis does not rule out that other potential correlated-omitted variables exist.

Measurement error in the independent variable: Measurement error may occur due to coding error or reporting error (Wooldridge, 2009). Since the values of the independent variable (*CAR*) are directly obtained from Eventus, no coding errors are expected. However, potential reporting error from the database cannot be eliminated.

Reverse causation: Reverse causation occurs when the dependent variable affects the independent variable. The mandatory adoption of XBRL is determined by the SEC for all large accelerated filers with a public float that is larger than 5 billion dollars. Hence, no reverse causation is possible for these adopters. However, reverse causation could exist with voluntary XBRL filers as they may signal better corporate disclosure by doing so. By removing firms that filed XBRL 10-K filings before the mandatory adaption period it is assumed that there is no reverse causation in the research design.

Autocorrelation: Autocorrelation occurs when the explanatory variable at time t are correlated with t_{-1}. The independent variables *XBRL* and *POST* cannot be correlated with t_{-1} as XBRL was adopted at a specific point in time (15, June 2009), *POST* is an indicator variable for the pre- or post- period, which also does not lead to autocorrelation. However, the control variables may be autocorrelated as size, number of analysts, market-to-book ratio, standard deviation of stock return and R&D expenditures could be correlated with prior year values.

4.5 Summary

Using the predictive validity framework the external, internal and construct validity of this research has been assessed. Due to the extensive literature review, this thesis deems that the research design meets the requirement as proposed by Libby et al. (2002). A difference-in-difference test using an OLS regression is used to test the first hypothesis. For the second hypothesis, this thesis utilizes again OLS regression to estimate the effect of XBRL. In addition, potential endogeneity issues and their sources are discussed. Endogeneity arising from omitted variables, measurement error, reverse causation and autocorrelation might be present in the dataset. This shall be analyzed in *Section 5.4.2*.

Table 2: Variable description

Variable	Variable description	Variable definition
CAR	Cumulative abnormal return	Measured 180 days after the insider trading date using CRSP value-weighted stock index. Negative CARs following insider sales are multiplied by negative one
XBRL	XBRL adopters	Indicator variable equal to one if the firm initially adopted XBRL in the first year of mandatory adoption and zero otherwise.
POST	Post period	Indicator variable equal to one if the 10-K filing occurred between une 15, 2009, and 14 June, 3020 while filings that occur in the year prior are coded as zero.
*POST*XBRL*	Interaction variable	Interaction variable that captures the effect of XBRL on insider trading profitability
ISH	Institutional holdings	The percentage of a firm's total outstanding shares that are held by institutional investors. Data is obtained through Thomson Reuters.
*ISH*XBRL*	Interaction variable	Interaction variable that captures the effect of XBRL for institutional investors on insider trading profitability.
Controls		
M_B_ratio	Market to book ratio	Measured as market value divided by total shareholder equity of the current fiscal year, using Compustat
N_An	Number of analysts	The maximum number of analysts from the consensus analyst forecast of the current fiscal year, obtained from I/B/E/S.
RND	R&D expenditures	Indicator variable equal to one if the firm reported R&D expenditures in the current fiscal year, otherwise zero. Data is obtained from Compustat.
Lsize	Log of firm size	Measured as the log of the market capitalization of the current fiscal year and obtained via AuditAnalytics.
STD_RET	Standard deviation of stock return	Measured over 180 days until 1 day prior to the insider transaction date. Daily stock return is obtained via TAQ. Standard deviation of the stock return is calculated using STATA.

Table 2 provides the description and definition of the dependent, independent and control variables used in the two models for hypothesis testing

5 Sample Selection and Data

5.1 Introduction

This section provides an explanation of the sample selection process for XBRL firms, the matching process with the control groups and how the insider transaction data are obtained. Then the descriptive statistics of the samples are provided. Furthermore, this section analyzes the correlation matrixes of the samples. Finally, this section explains how the data is tested for accordance with the OLS assumptions.

5.2 Sample Selection

5.2.1 XBRL firms

All 10-K XBRL filings for the fiscal period ending on or after 15 June, 2009 to 15 June, 2010 were downloaded from EDGAR. This research restricts to 10-K filings following Blankespoor et al., (2014). 477 10-K XBRL filings were identified. Next, previous XBRL filers were reduced from the sample as the filing for fiscal period 2009 was not their initial XBRL filing. Then all filers that were not categorized as large accelerated filers[9] were deleted from the sample. Furthermore, all filings without the necessary data for both pre- and post- periods were also eliminated, resulting in 191 large accelerated first-time XBRL filers (table 3).

This thesis follows Blankespoor et al. (2014) to establish the pre-XBRL and post-XBRL period. The pre-XBRL period starts from 15 June, 2008 until 14 June, 2009. The post-XBRL period ranges from 15 June, 2009 until 14 June, 2010. In addition, this thesis choses this cut-off period as the amount of errors in XBRL filings decreased between SEC's voluntary XBRL filings program and the mandatory program (Debreceny et al. 2010; Bartley et al. 2011).

Consistent with Blankespoor et al. (2014), samples are created with two control groups. The first control group is chosen based on industry type using the one-digit SIC code and firm size measured by market capitalization. By creating two samples this thesis aims to add additional robustness to the hypothesis testing results. The XBRL firms for each one-digit SIC industry type were counted and the equivalent amount of firms are chosen based on the same one-digit SIC and largest market capitalization as selection criteria. Panel A of table 4 shows the industry breakdown of the XBRL sample. Since companies with a public float larger than 5 billion

[9] For this categorization data from AuditAnalytics were used to classify large accelerated filers.

dollars were required to adopt XBRL in the initial phase, it is not surprising to see that the average firm size of both control groups are smaller than the average firm size of the XBRL group (Panel B of table 4). The second control group is selected based on industry type and the number of analysts. The XBRL firms for each one-digit SIC industry type were counted and the equivalent amount of firms were selected based on the same one-digit SIC and the highest number of analysts to create a control group with better information environments (Blankespoor et al., 2014). Panel B of table 4 shows that the mean number of analysts of both industry control groups (15.9 analysts and 17.3 analysts, respectively) are close to the XBRL group (18.3 analysts).

5.2.2 Insider transactions

For each firm in the matched sample, all insider transactions are obtained from the Thomson Reuter database via WRDS. These insider trading data are collected through the Thomson Reuters Insider Data Filing Feed via WRDS. Following Frankel & Li (2004) several filters are applied to the insider transactions in order to obtain informative transactions. Non-management insider transactions are deleted, since this thesis focuses on information asymmetry between “management” and investors. Furthermore, share prices of less than 2 dollar and transactions with less than 100 shares are deleted to eliminate noise (Lakonishok & Lee, 2001). The remaining transactions are netted at company level for each transaction day. For the first model the Industry – Size and Industry – Analyst sample contain respectively, 10,194 and 11,599 transactions. For the second model the Industry – Size and Industry – Analyst sample contain respectively, 6,153 and 6,350 transactions.

Table 3: Sample selection

XBRL 10-K's for fiscal periods between 6/15/2009 and 6/14/2010	477
Less filings of firms that filed 10-K's in prior period	(4)
Less filings that do not have the necessary data for both pre- and post-periods	(282)
Total initial XBRL 10-K filings in year of adoption	191

Table 3 provides the sample selection process of the XBRL firms

Table 4: Industry breakdown and descriptive statistics

Panel A — industry breakdown of XBRL sample

One-digit SIC	Industry name	Number of XBRL firms	Percent of total
1	Mining and construction	30	15.71%
2	Light manufacturing and chemicals	24	12.57%
3	Heavy manufacturing	47	24.61%
4	Transportation and public utilities	30	15.71%
5	Wholesale and retail trade	21	10.99%
6	Finance, insurance, and real estate	21	10.99%
7	Services	15	7.85%
8	Health services	2	1.05%
9	Unclassified	1	0.52%
	Total	191	100.00%

Panel B — descriptive statistics across samples

		XBRL	All non-XBRL	Industry and size	Industry and analysts
Number of firms		191	689	191	191
Number of observations				382	382
Market capitalization					
(in millions)	(mean)	28,800	2,840	19,200	18,100
	(median)	13,200	1,110	6,180	5,370
Assets (in millions)	(mean)	47,275	3,747	27,435	26,091
	(median)	12,555	1,466	6,470	5,211
Number of analysts	(mean)	18.3	8.5	15.9	17.3
	(median)	17	7	15	16

Panel A provides the distribution of the XBRL firms across industries as identified by the one-digit SIC code. Panel B provides the descriptive statistics across the XBRL and non-XBRL firms for the Industry – Size and Industry – Analyst samples

5.3 Descriptive Statistics

Table 5 reports the descriptive statistics across all samples. Panel A and C are the Industry – Size sample to test hypothesis one and two, respectively. Panel B and D are the Industry – Analyst sample to test hypothesis one and two, respectively. The Industry – Size samples contain less insider transactions than the Industry – Analyst samples (10,194 and 6,153 against 11,599 and 6,350, respectively). Furthermore, the mean and median *CAR* is lower for the Industry – Size samples than for the Industry – Analyst samples. Barth et al. (2001) suggested that increased information asymmetry may lead to increased profitability of information acquisition, thus increased number of analysts following the stock. This reasoning could explain the higher mean and median *CAR* for the Industry – Analyst sample.

The mean *CAR* is 0.13, 0.15, 0.10 and 0.11 for Panel A, B, C and D, respectively, which is the abnormal returns that insiders experience on average. Panel C and D only contain the insider transactions of the post period. The mean *CAR* of Panel C and D are both lower than the *CAR* of Panel A and B, suggesting that for both non-XBRL and XBRL firms the *CAR* is lower in the post period than in the combined pre and post period sample. This also indicates that the higher amount of post period transactions in Panel A and B (indicated by *POST*) deflates the mean *CAR* in these samples.

Across all samples there are slightly more insider transactions from non-XBRL firms than from XBRL firms as the mean of *XBRL* is lower than 0.50 in each Panel. This may suggest that insiders of XBRL firms are less involved with insider trading. In Panel A and B the mean of *POST* is respectively, 0.58 and 0.57, thus, in both samples more insider transactions occur in the post period than in the pre period.

Panel A and B show that the mean of the interaction variable *POST*XBRL* is 0.27 and 0.26 respectively, indicating that in both samples there are less insider trades for XBRL firms in the post period.

The Industry – Size samples (Panel A and C) have a mean (median) *size* of 19,300 (6,180) and 19,100 (6,410) million dollars, respectively, which is larger than the Industry – Analyst samples (Panel B and D), with a mean (median) *size* of 18,100 (6,000) and 18,200 (5,830) million dollars, respectively. Since, the Industry – Size samples are matched according to the largest firm size (as discussed in *Section 5.2.1*) it is not surprising that the mean and median firm size is larger in the Industry – Size samples.

The maximum value of market-to-book ratio is exceptionally high across the samples (44,843.56). After further examination of the data (Appendix II), it appears that these values are outliers at the 95th percentile. This may lead to inflation of the mean. Another outlier is at the 1st percentile. Hence, *M_B_ratio* is winsorized at the 1st and 95th percentile. Resulting in a mean of 3.13, 3.13, 3.72 and 3.82, respectively in Panel A, B, C and D (table 6). The market-to-book ratio in the post period (Panel C and D) is higher than in the combined pre- and post-period (Panel A and B). The difference might be due to the financial crisis that caused stock prices to fall in 2008 (Dwyer, 2009).

Panel A and B both have a mean *RND* of 0.49, indicating that less insider transactions stem from firms that reported R&D expenses than from firms that did report R&D expenses. In Panel C and D the mean *RND* equals 0.51 and 0.55 respectively, thus more insider transactions stem from firms that reported R&D expenses than firms that did not.

The mean standard deviation of stock return equals 0.0291, 0.0311, 0.0225 and 0.0237 for Panel A, B, C and D, respectively. The mean *STD_RET* is higher for those samples that have a higher *CAR*. This is consistent with Frankel & Li's (2004) notion that the standard deviation of stock return is positively related with insider trading profitability.

Table 5: Descriptive statistics across all samples

	Min.	Max.	Mean	Median	Std. dev.
Panel A: sample model 1 insider transactions Industry — Size (N = 10,194)					
CAR (%)	-2.2581	3.2294	0.1321	0.0976	0.3508
XBRL	0	1	0.4765	0	0.4995
POST	0	1	0.5820	1	0.4933
*POST*XBRL*	0	1	0.2737	0	0.4459
Size (in million $)	261	322000	19300	6180	39300
N_An	1	39	15.93	15	7.4268
M_B_ratio	-688.46	44843.56	6.9533	2.77	443.24
RND	0	1	0.4901	0	0.4999
STD_RET (%)	0.0064	0.1467	0.0291	0.0253	0.0154
Panel B: sample model 1 insider transactions Industry — Analyst (N = 11,599)					
CAR (%)	-2.0350	3.2294	0.1544	0.1218	0.3800
XBRL	0	1	0.4600	0	0.4984
POST	0	1	0.5729	1	0.4947
*POST*XBRL*	0	1	0.2608	0	0.4391
Size (in million $)	32.7	322000	18100	6000	38100
N_An	2	39	17.14	16	6.4033
M_B_ratio	-166.16	44843.56	11.12	2.77	587.33
RND	0	1	0.4908	1	0.4999
STD_RET (%)	0.0064	0.1389	0.0311	0.0273	0.0160
Panel C: sample model 2 insider transactions Industry — Size (N = 6,153)					
CAR (%)	-1.2036	1.7521	0.1016	0.0761	0.2845
XBRL	0	1	0.4735	0	0.4993
ISH (%)	0.0002	2.1998	0.7486	0.7718	0.2083
*ISH*XBRL*	0	1.0693	0.3562	0	0.3866
Size (in million $)	261	322000	19100	6410	39200
N_An	1	39	15.67	15	7.2351
M_B_ratio	-688.46	44843.56	10.00	2.85	570.93
RND	0	1	0.5099	1	0.4999
STD_RET (%)	0.0064	0.1028	0.0225	0.0204	0.0099
Panel D: sample model 2 insider transactions Industry — Analyst (N = 6,350)					
CAR (%)	-1.5197	1.7521	0.1087	0.0842	0.2987
XBRL	0	1	0.4620	0	0.4986
ISH (%)	0.0037	2.1998	0.7770	0.7948	0.2127
*ISH*XBRL*	0	1.0693	0.3477	0	0.3855
Size (in million $)	32.7	322000	18200	5830	38800
N_An	2	39	16.83	16	6.6295
M_B_ratio	-163.62	44843.56	18.11	2.97	793.56
RND	0	1	0.5453	1	0.4980
STD_RET (%)	0.0064	0.1028	0.0237	0.0220	0.0101

Table 5 provides the descriptive statistics of across all samples. All variables are explained in table 2

Table 6: Descriptive statistics of *M_B_ratio* winsorized at 1st and 95th percentile

	Min.	Max.	Mean	Median	Std. dev.
Panel A: sample model 1 insider transactions Industry — Size (N = 10,194)					
M_B_ratio	-5.7929	8.3427	3.1324	2.770043	2.2440
Panel B: sample model 1 insider transactions Industry — Analyst (N = 11,599)					
M_B_ratio	-1.3913	11.6448	3.5041	2.7700	2.8268
Panel C: sample model 2 insider transactions Industry — Size (N = 6,153)					
M_B_ratio	1.8824	16.0009	3.7213	2.8473	2.7288
Panel D: sample model 2 insider transactions Industry — Analyst (N = 6,350)					
M_B_ratio	0.3390	12.0415	3.8179	2.9701	2.8776

Table 6 provides the descriptive statistics across all samples of *M_B_ratio* winsorized at the 1st and 95th percentile, since the outliers of *M_B_ratio* appear at the 1st and 95th percentile (see *Section 5.3* and Appendix II)

5.4 Data

5.4.1 Correlation matrixes

Table 7 provides the correlation coefficients across all samples. This table shows that no coefficient is higher than 0.7 in Panel A and B, indicating that no multicollinearity is present. Multicollinearity is also tested for by computing the variance inflation factor (VIF). For Panel A and B the VIF equals 1.90 and 1.89, respectively, which is smaller than 10, proving that no multicollinearity is present[10]. However, for Panel C and D, the table shows that the correlation coefficient between *ISH* and *ISH*XBRL* equals 0.9704 and 0.9707, respectively. Thus it might be possible that multicollinearity exist. However, the VIF equals 7.54 and 7.32, respectively for Panel C and D. Despite that the VIF is much higher than the samples of the first model, it is still lower than 10. Hence, multicollinearity is no concern in the samples.

The interaction variable *POST*XBRL* is significantly and negatively correlated with CAR. This finding is consistent with the expectation that XBRL is negatively associated with insider trading profitability. Furthermore, contrary to prior research, R&D expenditures is negatively correlated with *CAR* across all samples. The log of size (*Lsize*) is significantly and negatively correlated with *CAR*, this finding supports the argument that insiders' profits tend to be more significant in smaller firms (Lakonishok and Lee, 2001). *RND* is negatively correlated with *CAR*, this finding contradicts the findings of Aboody & Lev (2000) that firms that report R&D expenditures experience greater information asymmetry. Frankel & Li (2004) found similar negative correlation coefficients in their correlation analyses, however the results were not significant. An alternative explanation may be that the private information of insiders of firms

[10] See Appendix II for the test results.

that report R&D expenditures is actually less precise (Huddart & Ke, 2007). *STD_RET* is positive and significantly correlated with *CAR*, which is consistent with Frankel & Li's (2004) findings.

Table 7: Pearson correlation across all samples

Panel A: Pearson correlations model 1 Industry – Size

	CAR	*XBRL*	*POST*	*POST*XBRL*	*Lsize*	*N_An*	*M_B_ratio*	*RND*	*STD_RET*
CAR	1.0000								
XBRL	-0.0644***	1.0000							
POST	-0.1105***	-0.0148	1.0000						
*POST*XBRL*	-0.1146***	0.6434***	0.5202***	1.0000					
Lsize	-0.0567***	0.4697***	0.0194**	0.3125***	1.0000				
N_An	-0.0030	0.3089***	-0.0243**	0.1727***	0.4996***	1.0000			
M_B_ratio	-0.0104	-0.0074	0.0089	-0.0041	0.0004	-0.0084	1.0000		
RND	-0.0767***	0.1223***	0.0194*	0.0842***	0.2251***	0.3416***	-0.0092	1.0000	
STD_RET	0.2539***	-0.0142	-0.5054***	-0.2864***	-0.2006***	0.0374***	-0.0094	-0.1655***	1.0000

Panel B: Pearson correlations model 1 Industry – Analyst

	CAR	*XBRL*	*POST*	*POST*XBRL*	*Lsize*	*N_An*	*M_B_ratio*	*RND*	*STD_RET*
CAR	1.0000								
XBRL	-0.0800***	1.0000							
POST	-0.1042***	-0.0110	1.0000						
*POST*XBRL*	-0.1207***	0.6436***	0.5129***	1.0000					
Lsize	-0.0743***	0.5280***	0.0482***	0.3587***	1.0000				
N_An	-0.0420***	0.1621***	-0.0290***	0.0856***	0.3790***	1.0000			
M_B_ratio	-0.0107	-0.0123	0.0122	-0.0074	0.0008	-0.0191**	1.0000		
RND	-0.0822***	0.0325***	-0.0020	0.0243***	0.0514***	0.3339***	-0.0104	1.0000	
STD_RET	0.2159*	-0.0539***	-0.5021***	-0.3096***	-0.2447***	-0.0016	-0.0133	-0.1557***	1.0000

Panel C: Pearson correlations model 2 Industry – Size

	CAR	*XBRL*	*ISH*	*ISH*XBRL*	*Lsize*	*N_An*	*M_B_ratio*	*RND*	*STD_RET*
CAR	1.0000								
XBRL	-0.1235***	1.0000							
ISH	-0.1746***	0.0122	1.0000						
*ISH*XBRL*	-0.1332***	0.9704***	0.1201***	1.0000					
Lsize	-0.0090	0.4504***	-0.2165***	0.3667***	1.0000				
N_An	0.0016	0.2815***	0.1224***	0.3031***	0.4937***	1.0000			
M_B_ratio	-0.0144	-0.0100	-0.0319**	-0.0096	-0.0001	-0.0123	1.0000		
RND	-0.0725***	0.1088***	0.2078***	0.1301***	0.1799***	0.2882***	-0.0127	1.0000	
STD_RET	0.2645***	-0.0592***	0.0039	-0.0319**	-0.2678***	0.0130	-0.0095	-0.1401***	1.0000

Panel D: Pearson correlations model 2 Industry – Analyst

	CAR	*XBRL*	*ISH*	*ISH*XBRL*	*Lsize*	*N_An*	*M_B_ratio*	*RND*	*STD_RET*
CAR	1.0000								
XBRL	-0.1368***	1.0000							
ISH	-0.1115***	-0.1170***	1.0000						
*ISH*XBRL*	-0.1452***	0.9707***	-0.0089	1.0000					
Lsize	-0.0365***	0.5161***	-0.3266***	0.4423***	1.0000				
N_An	0.0173	0.1236***	0.1115***	0.1521***	0.3797***	1.0000			
M_B_ratio	-0.0151	-0.0165	-0.0459***	-0.0160	0.0008	-0.0255**	1.0000		
RND	-0.0420***	0.0324***	0.2347***	0.0585***	0.0585***	0.2744***	-0.0176	1.0000	
STD_RET	0.1942***	-0.1658***	0.0808***	-0.1354***	-0.3734***	-0.0216*	-0.0144	-0.0925***	1.0000

Table 7 provides information on the Pearson Correlation between the variables of interest.
*, **, *** reflect the statistical significance of the coefficients at the 10, 5, and 1% level, respectively

5.4.2 Testing OLS assumptions[11]

First, the initial regression is ran for each model. All models are significant with an F-statistic that has a P-value of 0.000 across all models. The *M_B_ratio* is winsorized at the 1st and 95th percentile (as discussed in *Section 5.3*). The winsorized regressions increased the adjusted R^2 and the R^2 of the Industry – Size samples but has slightly decreased the adjusted R^2 and the R^2 of the Industry – Analyst samples.

Furthermore, the normality of errors is checked for the initial samples. The Shapiro Wilk test is used to test the normality of errors. The P-value for each sample is 0.000, thus the H_0 is rejected. The errors are not normally distributed.

In addition, the initial data is checked for heteroscedasticity using the Breusch-Pagan / Cook-Weisberg test. However, since the errors are not normally distributed, the White's general test for heteroscedasticity is also used. The P-value for each sample size for both tests are 0.000. The H_0 is rejected, thus the data is heteroscedastic. To control for heteroscedasticity, White's standard errors are computed for the final regression[12].

Finally, endogeneity as discussed in *Section 4.4.2* is checked for by regressing the predicted residuals with the explanatory variables. Across all models the P-value of the coefficients are 1.000, thus insignificant. Furthermore, the F-statistics report a P-value of 1.000, which indicates that the model is insignificant. Hence, it is concluded that endogeneity is not present across all models.

5.5 Summary

The descriptive statistics did not show any outliers except for the relatively high values of *M_B_ratio*. After winsorizing *M_B_ratio* at the 1st and 95th percentile the data did not show any abnormalities. Except for the sign of the *RND* variable, other variables in the correlation matrixes had a sign that was consistent with prior literature. Furthermore, the OLS assumptions were tested. Heteroskedasticity occurred across all datasets and the residuals of all regressions were non-normal distributed. To solve heteroskedasticity this thesis computed White's standard errors for the final regression (see *Section 6.2, 6.3* and Appendix II). Endogeneity is of no concern across all samples. Hence, this thesis proceeds with hypothesis testing.

---

[11] For a step-by-step explanation please refer to Appendix II.

[12] The robust regression command in Stata is similar to computing White's standard errors.

6 Results

6.1 Introduction

This section discusses the regression results of the initial XBRL 10-K filings on CAR. Then this section proceeds with discussing the regression results of the interaction of initial XBRL 10-K filings and investor sophistication on CAR. After discussing the regression results, this section continues with analyzing the results and the implications of the regression results on the research question.

6.2 Regression Results of the Initial XBRL 10-K Filing

Panel A of table 8 reports the regression with only the independent variables. *POST*XBRL* is significant at the 5% and 10% level for the Industry – Size and Industry – Analyst sample, respectively. The coefficient equals -0.0364 and -0.0399 for the Industry – Size and Industry – Analyst sample, respectively. Indicating that the interaction variable is negatively associated with *CAR*. Thus XBRL firms, after filing the initial XBRL 10-K, are associated with a decrease of 3.64 and 3.99 percentage points of CAR. This suggests that, consistent with the first hypothesis, insiders of XBRL firms experience less profitability compared to insiders of non-XBRL firms. The adjusted R^2 (R^2) are 0.0169 (0.0172) and 0.0179 (0.0181) for the Industry – Size and Industry – Analyst sample, respectively. Indicating that the model explains 1.69% and 1.79% of the variation of the dependent variable.

In Panel B of table 8 the control variables are added to the model. The R^2 increased to 6.98% and 5.66% for the Industry – Size and Industry – Analyst sample, respectively[13]. The models are both significant with a P-value of 0.000 for the F-statistics. This shows that the linear model utilized in Panel B has a greater effect on the variation of the dependent variable. Hence, Panel B depicts a more precise model. The interaction variable *POST*XBRL* is negative and significant at the 5% and 10% level for the Industry – Size and Industry – Analyst sample respectively. Hence, the null-hypothesis is rejected. Thus consistent with the first hypothesis XBRL is negatively associated with information asymmetry. The coefficients experienced a decrease compared to the independent variables only model. The coefficients are equal to -0.0291 for the Industry – Size sample, indicating that after companies filed their 10-Ks in XBRL

[13] The adjusted R^2 is not provided in the Stata output as the robust regression does not allow for the adjusted R^2 as output. Please refer to appendix II for the adjusted R^2 of the initial regression.

the *CAR* of insider trades decreased with 2.91 percentage points for the Industry – Size sample. The Industry – Analyst sample shows a decrease of 2.51 percentage points of the *CAR*. The negative association between XBRL and information asymmetry is consistent with the findings of prior research (e.g. Liu et al., 2017; Blankespoor et al., 2014; Yoon et al., 2011). However, the results show that *XBRL* firms whether in the pre- or post- period are significantly negatively associated with *CAR* and the interaction variable *POST*XBRL* has become less negative in the post period. Thus it seems that the negative effect of XBRL firms on *CAR* has been compensated after the XBRL firms filed their first 10-K filing. The coefficient of *POST* is significant the Industry – Size sample, but not in the Industry – Analyst sample. In both samples the coefficient of *POST* is positive. This indicates that in the Industry – Size sample the CAR of both XBRL and non-XBRL firms experienced a significant increase in the post period. The insignificance and marginal coefficient of *POST* in the Industry – Analyst sample may be explained by the number of analysts that follow the stock. Indeed, the coefficient of *N_An* is negative and significant at 1%. In contrast with the correlation results (as discussed in *Section 5.4.1*) the coefficients of *Lsize* are positive and significant at 1% level for both samples. *M_B_ratio* and *RND* are significant and negatively associated with *CAR*. *STD_RET* is positively and significantly associated with *CAR*.

6.3 Regression results of the initial XBRL 10-K filing and investor sophistication

Table 9 reports the regression with only the independent variables. *ISH*XBRL* is not significant for both Industry – Size and Industry – Analyst samples. The coefficients of *ISH*XBRL* is in both samples smaller than the *ISH* coefficient, suggesting that XBRL decreases the effect of institutional shareholders on *CAR*. The adjusted R^2 (R^2) are 0.0442 (0.0447) and 0.0340 (0.0345) for the Industry – Size and Industry – Analyst sample, respectively. Indicating that the models explain 4.42% and 3.40% of the variation of the dependent variable.
In Panel B of table 9 the control variables are added to the model. The R^2 increased to 13.56% and 7.47% for the Industry – Size and Industry – Analyst sample, respectively[14]. The models are both significant with a P-value of 0.000 for the F-statistics. This shows that the linear model utilized in Panel B has a greater effect on the variation of the dependent variable. Hence, Panel B depicts a more precise model. The interaction variable *ISH*XBRL* remains insignificant in the Industry – Size sample. However, the coefficient of *ISH*XBRL* becomes significant at the

[14] The adjusted R^2 is not provided in the Stata output as the robust regression does not allow for the adjusted R^2 as output. Please refer to appendix II for the adjusted R^2 of the initial regression.

5% level in the Industry – Analyst sample. The coefficient of *ISH* is significant at 1% and negative in both samples. In the Industry – Analyst sample the interaction coefficient (-0.0895) has increased 0.0795 compared to the *ISH* coefficient (-0.1690). Meaning that without XBRL each percentage point increase in institutional holdings results in a decrease of *CAR* of 16.90 percentage points. However, insiders of XBRL firms only experience a decrease of 7.95 percentage points of *CAR* from an increase in institutional holdings. The null for the second hypothesis is rejected at the 5% level for the Industry – Analyst sample. Thus XBRL decreases the negative effect of institutional shareholders on *CAR*. This could be explained due to the superior resources of institutional shareholders to acquire information that dampens profitability of insiders. However, the decrease in insider trading profitability of XBRL-firms is to a lesser extent contributed to institutional shareholders. Indicating that institutional investors lose a part of their superior information due to XBRL. Thus consistent with the second hypothesis, XBRL democratizes the financial markets by making more information accessible to individual investors. O'Brien & Bhushan (1990) suggested analysts convey firm information to institutional investors. Hence, the difference in significance of the interaction variable between the two samples may be explained because of the matching of the maximum number of analysts in the Industry – Analyst sample. Thus institutional investors whom have superior information advantage due to the number of analysts following the firm, have more superior information 'to lose' when this information becomes widely accessible due to XBRL.

XBRL is in both samples negatively associated with *CAR* and significantly negative associated with *CAR* in the Industry – Size sample. Since these samples only contain data on the post period, this finding further supports the notion (as discussed in *Section 6.2*) that *XBRL* firms have a negative effect on *CAR*. Furthermore, *ISH* is in both samples significantly and negatively associated with *CAR*, this is consistent with Piotroski & Roulstone's (2004) notion that institutional investors experience less information asymmetry.

Again, in contrast with the correlation results (as discussed in *Section 5.4.*1) the coefficients of *Lsize* are positive and significant at 1% level for both samples. *M_B_ratio* is significantly and negatively associated with *CAR* in the Industry – Size sample. *STD_RET* is positively and significantly associated with *CAR*.

Table 8: Results of regressing CAR of insider trading events on XBRL initial filing

Panel A: regression results without control variables

Dependent variable	Cumulative Abnormal Return (CAR)			
	Industry – Size		Industry – Analyst	
	Coefficient	P-value	Coefficient	P-value
Main variables				
XBRL	-0.0253**	(0.017)	-0.0391***	(0.000)
POST	-0.0618***	(0.000)	-0.0623***	(0.000)
*POST*XBRL*	-0.0364***	(0.009)	-0.0399***	(0.005)
Intercept	0.1901***	(0.000)	0.2185***	(0.000)
N	10,194		11,599	
F-statistic	60.32	(0.000)	72.54	(0.000)
Adj. R2	0.0169		0.0179	
R2	0.0172		0.0181	

Panel B: regression results with control variables

Dependent variable	Cumulative Abnormal Return (CAR)			
	Industry – Size		Industry – Analyst	
	Coefficient	P-value	Coefficient	P-value
Main variables				
XBRL	-0.0372***	(0.006)	-0.0531***	(0.000)
POST	0.0267**	(0.019)	0.0034	(0.773)
*POST*XBRL*	-0.0291**	(0.047)	-0.0251*	(0.088)
Control variables				
Lsize	0.0119***	(0.001)	0.0115***	(0.001)
N_An	0.0004	(0.503)	-0.0017***	(0.003)
M_B_ratio	-0.0040***	(0.002)	-0.0007**	(0.048)
RND	-0.0216***	(0.003)	-0.0342***	(0.000)
STD_RET	5.9342***	(0.000)	4.9865***	(0.000)
Intercept	-0.2804***	(0.001)	-0.1796**	(0.023)
N	10,194		11,599	
F-statistic	45.52	(0.000)	54.78	(0.000)
Adj. R2				
R2	0.0698		0.0566	

Panel A of table 8 provides the result of regression cumulative abnormal return (CAR) on an XBRL 10-K filer indicator, a pre-/post- period indicator and their interaction variables. Panel B provides the result including the control variables. For the regression 191 XBRL firms and two non-XBRL control groups (industry-size and industry-analyst matched firms) are used for 2008 and 2009. *CAR* (0, +180) is obtained starting from the insider transaction date (day 0) of the sample firms. Two-sided p-values are provided in parentheses to the right the coefficients. All variables are defined in table 2. *M_B_ratio* is winsorized at the 1st and 95th percentile. White's standard errors are computed for the regression of Panel B

*, **, and *** Significance at the 10, 5, and 1% or lower levels, respectively

Table 9: Results of regressing CAR of insider trading events on the period after initial XBRL filing

Panel A: regression results without control variables

Dependent variable	Cumulative Abnormal Return (CAR)			
	Industry – Size		Industry – Analyst	
	Coefficient	P-value	Coefficient	P-value
Main variables				
XBRL	-0.0979***	(0.003)	-0.0716**	(0.037)
ISH	-0.2437***	(0.000)	-0.1760***	(0.000)
*ISH*XBRL*	-0.0408	(0.333)	-0.0231	(0.599)
Intercept	0.3149***	(0.000)	0.2856***	(0.000)
N	6,153		6,350	
F-statistic	96.33	(0.000)	75.86	(0.000)
Adj. R2	0.0442		0.0340	
R2	0.0447		0.0345	

Panel B: regression results with control variables

Dependent variable	Cumulative Abnormal Return (CAR)			
	Industry – Size		Industry – Analyst	
	Coefficient	P-value	Coefficient	P-value
Main variables				
XBRL	-0.0669**	(0.029)	-0.0220	(0.476)
ISH	-0.2111***	(0.000)	-0.1690***	(0.000)
*ISH*XBRL*	-0.0196	(0.617)	-0.0895**	(0.022)
Control variables				
Lsize	0.0178***	(0.000)	0.0113***	(0.009)
N_An	0.0015**	(0.019)	0.0016**	(0.010)
M_B_ratio	-0.115***	(0.000)	-0.0019	(0.255)
RND	-0.0103	(0.158)	2.22E-06	(1.000)
STD_RET	7.6448***	(0.000)	5.9064***	(0.000)
Intercept	-0.2678***	(0.009)	-0.1356	(0.187)
N	6,153		6,350	
F-statistic	82.78	(0.000)	63.95	(0.000)
Adj. R2			0.0735	
R2	0.1356		0.0747	

Panel A of table 9 provides the result of regression cumulative abnormal return (CAR) on an XBRL 10-K filer indicator, a variable that proxies for investor sophistication by measuring the percentage of institutional holdings (*ISH*), and their interaction variable. Panel B provides the result including the control variables. For the regression 191 XBRL firms and two non-XBRL control groups (industry-size and industry-analyst matched firms) are used for 2008 and 2009. *CAR* (0, +180) is obtained starting from the insider transaction date (day 0) of the sample firms. Two-sided p-values are provided in parentheses to the right the coefficients. All variables are defined in table 2. *M_B_ratio* is winsorized at the 1st and 95th percentile. White's standard errors are computed for the regression of Panel B

*, **, and *** Significance at the 10, 5, and 1% or lower levels, respectively

6.4 Analysis and Implications of the Regression Results

The results of the first hypothesis showed that the coefficient of *POST*XBRL* is significantly and negatively associated with *CAR* across both Industry – Size sample and Industry – Analyst sample. The coefficient of *POST*XBRL* (P-value) are -0.0291 (0.047) and -0.0251 (0.088), respectively. Thus for firms that have filed the initial 10-K filing, the *CAR* of insider trades decreased with 2.91 and 8.8 percentage points for the Industry – Size sample and Industry – Analyst sample, respectively. Hence, the null hypothesis is rejected and this thesis finds that indeed, XBRL is negatively associated with information asymmetry.

The results of the second hypothesis showed that the coefficient of *ISH*XBRL* is only significant at the 5% level for the Industry – Analyst sample. This could be explained due to the matching of the control group. As the number of analysts was the criteria, and analysts are a source of superior information of institutional investors (O'Brien & Bhushan (1990), it may be that the institutional investors of stocks with the highest number of analysts 'lose' more superior information because of XBRL. The coefficient equals -0.0895 and has increased compared to the coefficient of *ISH* (-0.1690). The difference between the coefficients is 0.0795. Indicating that XBRL reduces the superior information advantage of institutional investors by 7.95 percentage points. The null hypothesis can be rejected at 5% significance level only for the Industry – Analyst sample.

To answer the research question, first it is observed that XBRL is associated with a decrease of insider trading profitability. Thus the decrease of the information gap between insiders and outsiders is associated with XBRL. When looking more closely to the information environment, outsiders can be distinguished into institutional investors and individual investors. This thesis shows that for stocks with a large number of analysts XBRL is associated with a decrease of the information gap between institutional investors and individual investors. This finding provides a more in-depth insight into the effect of XBRL on the information environment.

The findings of this thesis hold several implications for regulators, firms, investors and analysts. First, as XBRL proves to be associated with decreased information asymmetry, regulators across the globe have more incentive to adopt mandatory XBRL. Furthermore, since XBRL also shows to reduce insider trading profitability, regulators who wish to decrease insider trading also have more incentive to adopt XBRL. In addition, as XBRL is already associated with a reduction of information asymmetry without any external auditing or other practices to increase the informativeness and trustworthiness of XBRL filings, it may be that with proper frameworks and auditing regulations, XBRL even has more potential to improve the information environment. According to Frankel & Li (2004) firms benefit from decreased insider trading thus adopting XBRL could dampen the insider trading profits and in turn lead to a decrease of insider trading. In addition, improved corporate disclosure is a solution to the 'lemons' problem and firms who wants to signal the value of their business may also want to adopt XBRL. Investors may wish that XBRL becomes mandatory as it improves the information environment and allows for better informed decisions for the investors. Finally, the implications for analysts may be mixed, as less sophisticated investors have more means to analyze data themselves, the demand for analysts might decrease. However, since XBRL leads to better customizable financial statements the efficiency of financial statement analyses may increase. Furthermore, since XBRL might result in continuous reporting in the future, analysts as professionals in their field might experience increased demand to analyze the continuous input of financial information.

This thesis contributed to the theory by proving more in-depth evidence on the dynamics of the information market. Indeed, investors are heterogeneous and do not have the same resources to access and analyze financial information. This results in different degrees of information asymmetry in the information environment. Researchers should take into account this heterogeneity and distinguish between individual investors and institutional investors. Furthermore, as XBRL is believed to improve corporate disclosure. Future research needs to take into consideration the effect of XBRL on different information environments.

6.5 Summary

The regression model to test the first hypothesis reports a negative coefficient for the variable of interest (*POST*XBRL*) at the 5% and 10% significance level for the Industry – Size and Industry – Analyst sample. Hence, the null-hypothesis is rejected, thus, consistent with H1 XBRL is negatively associated with information asymmetry. To test the second hypothesis, the *ISH* variable is compared with the interaction variable *ISH*XBRL*. The regression model shows for both variables a negative coefficient across both samples. *ISH* is significant at 1% for both samples. However, *ISH*XBRL* is significant only in the Industry – Analyst sample at 5%. Despite the difference in significance, both *ISH*XBRL* coefficients show a decrease in the absolute value of the coefficients compared to the *ISH* coefficients. This indicates that *XBRL* democratizes the information environment and leads to a decreased information gap between institutional investors and individual investors. The null for the second hypothesis is rejected at the 5% level for the Industry – Analyst sample. These findings have practical implications for regulators, firms, investors and analysts. Furthermore, this thesis provided empirical evidence on the dynamics of the information environment.

7 Conclusions

7.1 Summary and Conclusions

The SEC adopted mandatory XBRL since 15 June, 2009. The SEC introduced a phased-in approach. All large accelerated filers with a worldwide public float larger than 5 billion dollars had to adopt XBRL. Furthermore, the European Union has decided to adopt mandatory XBRL filings starting from 2019.

XBRL is an electronic language that requires tagging of financial data. Due to XBRL financial data become more flexible and better customizable to the user's preferences. Prior research documented the benefits of XBRL, such as, increased transparency, increased accuracy, increased reliability and increased comparability of financial data. However, implementation of XBRL was costly and the filings do not come without errors (Hannon, 2006; Du et al., 2013). Hence, some scholars argued that XBRL filings need to be audited just like traditional financial statements (Srivastava & Kogan, 2010)

Many researchers wondered whether empirical evidence exist on the benefits of XBRL and numerous researches have been conducted (e.g., Wagenhofer 2003; Bonson et al. 2008; Madden 2011; Vasarhelyi et al. 2012; Blankespoor et al. 2014). Answers to this question is very important and relevant as it provides insight in the effectiveness of mandatory XBRL adoption in the U.S., which in turn could help regulators across the world to make better informed decisions. This thesis contributed to the discussion of the effectiveness of XBRL by conducting an empirical analysis on the effect of XBRL on information asymmetry. Specifically, this thesis answered the following research question:

RQ: How does XBRL reduces information asymmetry with respect to insider trading?

This thesis is the first to use insider trading profitability as a proxy for information asymmetry. This thesis hypothesized that due to better corporate disclosure XBRL leads to decreased information asymmetry, thus decreased insider trading profitability. Furthermore, it is expected that XBRL democratizes the information environment thus leading to a decreased information gap between the best informed investors (insiders), medium informed investors (institutional investors) and the least informed investors (individual investors).

A difference-in-difference test is conducted for both hypotheses. The sample consists of 191 XBRL firms that were directly downloaded from the EDGAR database. Then all XBRL firms are matched with a control group. Two control groups were created, one matched on industry and size and the other matched on industry and number of analysts. For all firms in the final sample, insider trades are obtained from the Thomson Reuter Insider Filing Feed. All insider trades from the same firm are netted per trading day. Then the cumulative abnormal return (CAR) over 180 calendar days starting from the insider trading date were obtained from Eventus. The CAR of net insider trading sales was multiplied by negative one to obtain absolute CAR values for negative CARs after a sale. For the first hypothesis, an indicator variable *POST*XBRL* is included, which is the variable of interest that measures the effect of XBRL on *CAR*. For the second hypothesis, an indicator variable *ISH*XBRL* is included, which shows the interaction effect of XBRL on institutional holdings (as proxy for large investors). In both regressions, the following control variables were included: log of firm size, number of analysts, market-to-book ratio, indicator variable for R&D expenses and the standard deviation of stock return.

The results of the first hypothesis rejects the null hypothesis at 1% and this thesis finds that indeed, XBRL is negatively associated with information asymmetry.

The results of the second hypothesis showed that the coefficient of *ISH*XBRL* is only significant at the 5% level for the Industry – Analyst sample. This could be explained due to the matching of the control group. Analysts are a source of superior information of institutional investors (O'Brien & Bhushan (1990), it may be that the institutional investors of stocks with the highest number of analysts 'lose' more superior information because of XBRL. The null hypothesis is rejected at 5% significance level only for the Industry – Analyst sample.

The answer to the research question is that XBRL is negatively associated with information asymmetry. More specifically, XBRL reduces information asymmetry by democratizing the information market. The findings of this thesis holds several implications for regulators, firms, investors and analysts. XBRL is a means for regulators to improve and democratize the information market. Firms can use XBRL to signal better corporate disclosure. XBRL reduces insider trading profitability thus it may decrease the incentives for future insider trading. Investors benefit from improved corporate disclosure due to XBRL and analysts may experience greater demand when continuous reporting becomes reality.

The implications for theory is that future research needs to take into consideration the effect of XBRL on different information environments. Furthermore, research should also distinguish between individual investors and institutional investors.

7.2 Limitations and Recommendations for Future Research

The limitations of this thesis relates to a lack of robustness check. Many different measurement of insider trading profitability exists in the literature, such as, BHAR or Carhart four-factor model (Dai et al. 2012). Future research might conduct analyses on other measurements of insider trading profitability. Furthermore, this research only tested data after the first phase of the phased-in approach. The results of this thesis are related to short-term empirical data and no inferences can be drawn for the long-term effects of XBRL. Further research might analyze the implications of XBRL on the long-term. Another interesting finding is the difference in the results of the two different samples for the second hypothesis. Further research may investigate whether the difference really stems from the matching of the control group or whether there are other correlated-omitted variables that this thesis did not foresee.

References

Aboody, D., & Lev, B. (2000). Information asymmetry. R&D and insider gains. The Journal of Finance, 55(6), 2747-2766.

Agrawal, A. and J.F. Jaffe (1995) 'Does Section 16b Deter Insider Trading by Target Managers?', Journal of Financial Economics 39(2–3): 295-319

Agrawal, A. and Nasser, T. (2012) 'Insider Trading in Takeover Targets', Journal of corporate finance (Amsterdam, Netherlands) 18(3): 598-625.

Ahmadpour, A., & Bodaghi, A. (2010). The effects of XBRL on financial transparency. International Journal of Information Science and Management, 65-76.

AICPA. (sd). Benefits and Potential Uses of XBRL. Opgeroepen op 07 14, 2017, van AICPA: https://www.aicpa.org/InterestAreas/FRC/AccountingFinancialReporting/XBRL/Pages/BenefitsandPotentialUsesofXBRL.aspx

Ali, A., Hwang, L., & Trombley, M. (2000). Accruals and Future Stock Returns: Tests of Naïve Investor Hypothesis. Journal of Accounting, Auditing, and Finance, 15, 161-181.

Alles, M., Kogan, A., & Vasarhelyi, M. (2004, January). Real-time Reporting and Assurance: Has their Time Come? ICFAI Reader.

Arnold, V., Bedard, J. C., Phillips, J. R., & Sutton, S. G. (2012). The impact of tagging qualitative financial information on investor decision making: Implications for XBRL. International Journal of Accounting Information Systems, 2-20.

Baiman, S., & Verrechia, R. (1996). The relation among capital markets, financial disclosure, production efficiency, and insider trading. Journal of Accounting Research, 34, 1-22.

Barth, M. E., Kasznik, R., & McNichols, M. F. (2001). Analyst coverage and intangible assets. Journal of accounting research, 39(1), 1-34.

Baldwin, A. A., Brown, C. E., & Trinkle, B. S. (2006). XBRL: An Impacts of Framework and Research Challenge. Journal of Emerging Technologies in Accounting, 97-116.

Bartley, J., Chen, Y., & Taylor, E. (2011). A comparison of XBRL filings to corporate 10–Ks: Evidence from the voluntary filing program. Accounting Horizons, 25(2), 227-245.

Bartov, E., Radhakrishnan, S., & Krinsky, I. (2000). Investor Sophistication and Patterns in Stock Returns after Earnings Announcements. The Accounting Review, 75, 43-63.

Belastingdienst. (sd). Wat is SBR? Opgeroepen op 07 16, 2017, van Belastingdienst: https://www.belastingdienst.nl/wps/wcm/connect/bldcontentnl/belastingdienst/intermediairs/aangifte_doen/standard_business_reporting/wat_is_sbr/

Bettis, J., Coles, J. and Lemmon, M. (2000) 'Corporate Policies Restricting Trading by Insiders', Journal of Financial Economics 57: 191.

Blankespoor, E., Miller, B. P., & White, H. D. (2014, December). Initial evidence on the market impact of the XBRL mandate. Review of Accounting Studies, 19(4), 1468-1503.

Bonson, E. (2001). The role of XBRL in Europe. The International Journal of Digital Accounting Research, 1(2), 101-110.

Bonson, E., Cortijo, V., & Escobar, T. (2008). The role of XBRL in Enhanced Business Reporting (EBR). Journal of Emerging Technology in Accounting, 5(1), 161-173.

Burgstahler, D., & Eames, M. (2006). Management of earnings and analysts' forecasts to achieve zero and small positive earnings surprises. Journal of Business Finance & Accounting, 5-6, 633-652.

Bushee, B., & Goodman, T. (2007). Which institutional investors trade based on private information about earnings and returns? Journal of Accounting Research, 45(2), 289-321.

Collins, D. W., Gong, G., & Hribar, P. (2003). Investor Sophistication and the Mispricing of Accruals. Review of Accounting Studies, 8, 251-276.

Cong, Y., Hao, J., & Zou, L. (2014). The Impact of XBRL Reporting on Market Efficiency. Journal of Information Systems, 28(2), 181-207.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. Psychological bulletin, 52(4), 281.

Dai, L., Fu, R.,Kang, J and Lee, I. (2012) 'Corporate Governance and Insider Trading', Working Paper, KAIST College of Business.

Debreceny, R., Chandra, A., Cheh, J., Guithues-Amrhein, D., Hannon, N., Hutchinson, P., & et al. (2005). Financial reporting in XBRL on the SEC's EDGAR system: a critique and evaluation. Journal of Information Systems, 19(2), 191-210.

Debreceny, R., Farewell, S., Piechokie, M., Felden, C., & Gräning, A. (2010). Does it add up? Early evidence on the data quality of XBRL filings to the SEC. J. Account. Public Policy, 29, 296-306.

Degeorge, F., Patel, J., & Zeckhauser, R. (1999). Earnings management to exceed thresholds. Journal of Business, 72(1), 1-33.

Diamond, D. W. (1985). Optimal Release of Information by Firms. The Journal of Finance, 40(4).

Du, H., Vasarhelyi, M. A., & Zheng, X. (2013). XBRL Mandate: Thousands of Filing Errors and So What? Journal of Information Systems, 27(1), 61-78.

Dwyer, G. P. (2009, September). Stock Prices in the Financial Crisis. Retrieved July 09, 2018, from https://www.frbatlanta.org/cenfis/publications/notesfromthevault/0909

Efendi, J., Park, J. D., & Smith, L. M. (2014). Do XBRL filings enhance informational efficiency? Early evidence from post-earnings announcement drift. Journal of Business Research, 67, 1099-1105.

European Comission. (2013, June 12). Revised Directive on transparency requirements for listed companies (Transparency Directive) – frequently asked questions . Brussels, Belgium: European Comission.

Fama, E.F., 1998, 'Market Efficiency, Long-Term Returns, and Behavioral Finance',Journal of Financial Economics 49: 283-306.

Frankel, R., & Li, X. (2004). Characteristics of a firm’s information environment and the information asymmetry between insiders and outsiders. Journal of Accounting and Economics, 37, 229-259.

Hand, J. (1990). A Test of the Extended Functional Fixation Hypothesis. The Accounting Review, 65, 740-763.

Hannon, N. (2006, April). Does XBRL Cost Too Much. Strategic Finance, pp. 59-60.

Harris, T. S., & Morsfield, S. (2012). An Evaluation of the Current State and Future of XBRL and Interactive Data for Investors and Analysts. Colombia University.

Huddart, S. J., & Ke, B. (2007). Information Asymmetry and Cross-sectional Variation in Insider Trading. Contemporary Accounting Research, 195-232.

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

Lakonishok, J., & Lee, I. (2001). Are insider trades informative? . The Review of Financial Studies, 14(1), 79-111.

Liu, C., Sia, C., O'Farell, G., & Teo, H. (2014). The impact of XBRL adoption in P.R. China. Decision Support Systems, 59, 242-249.

Liu, C., Wang, D., & Yao, L. (2014). An empirical study of XBRL's impact on analyst forecast behaviour. Journal of Accounting and Public Policy, 33(1), 69-82.

Liu, C., Yao, L., Sia, C., & Wei, K. (2014). The impact of early XBRL adoption on analysts' forecast accuracy: empirical evidence from China. Electronic Markets, 24(1), 47-55.

Madden, P. (2011). Greater accountability, less red tape: The Australian standard business reporting experience. International Journal of E-business Research, 7(2), 1-10.

Matsumoto, D. (2002). Management's incentives to avoid negative earnings surprises. The Accounting Review, 77(3), 483-514.

Nitchman, D. (2015, November 29). New Mandate in The Netherlands. Opgeroepen op July 16, 2017, van XBRL: https://www.xbrl.org/news/new-mandate-in-the-netherlands/

O'Brien, P. C., & Bhushan, R. (1990). Analyst following and institutional ownership. Journal of Accounting Research, 55-76.

Perdana, A., Robb, A., & Rohde, F. (2015). An Integrative Review and Synthesis of XBRL Research in Academic Journals. Journal of Information Systems, 29(1), 115-153.

Pinsker, R., & Li, S. (2008). Costs and benefits of XBRL adoption: early evidence. Communications of the ACM, 51(3), 47-50.

PwC. (2011). XBRL reporting risk and the role of internal audit.

Repogle, R. (2011, January 6). *Insider Trading in the United States* [PDF]. Association of Corporate Counsel.

Rezaee, Z., Elam, R., & Sharbatoghlie, A. (2001). Continuous auditing: The audit of the future. Managerial Auditing Journal, 16(3), 150-158.

Srivastava, R., & Kogan, A. (2010). Assurance on XBRL instance document: A conceptual framework of assertions. International Journal of Accounting Information Systems, 261-273.

Van Ardenne, R. (2007). Nederland haalt de XBRL buit nog niet binnen. Batavia XBRL BV.

Vasarhelyi, M., Chan, D., & Krahel, J. (2012). Consequences of XBRL standardization on financial statement data. Journal of Information Systems, 26(1), 155-167.

Wagenhofer, A. (2003). Economic consequence of internet financial reporting. Schmalenbach Business Review, 55(4), 262-279.

Wallagh, K., & Van den Ende, D. (2013). Dutch Business Benefit from XBRL. Deloitte. iBR.

Walther, B. (1997). Investor Sophistication and Market Earnings Expectations. Journal of Accounting Research, 35, 157-192.

Wooldridge, J. M. (2009). Introductory econometrics a modern approach. USA: South-Western Cengage Learning.

Wu, Wei, Information Asymmetry and Insider Trading (February 19, 2016). Fama-Miller Working Paper; Chicago Booth Research Paper No. 13-67. Available at SSRN: https://ssrn.com/abstract=2323537 or http://dx.doi.org/10.2139/ssrn.2323537

Yoon, H., Zo, H., & Ciganek, A. P. (2011). Does XBRL adoption reduce information asymmetry? Journal of Business Research, 64(2), 157-163.

Zhu, H., & Wu, H. (2011). Interoperability of XBRL financial statements in the U.S. International Journal of E-business Research, 7(2), 19-33.

Appendix I

Figure 5: Example of XBRL tagging

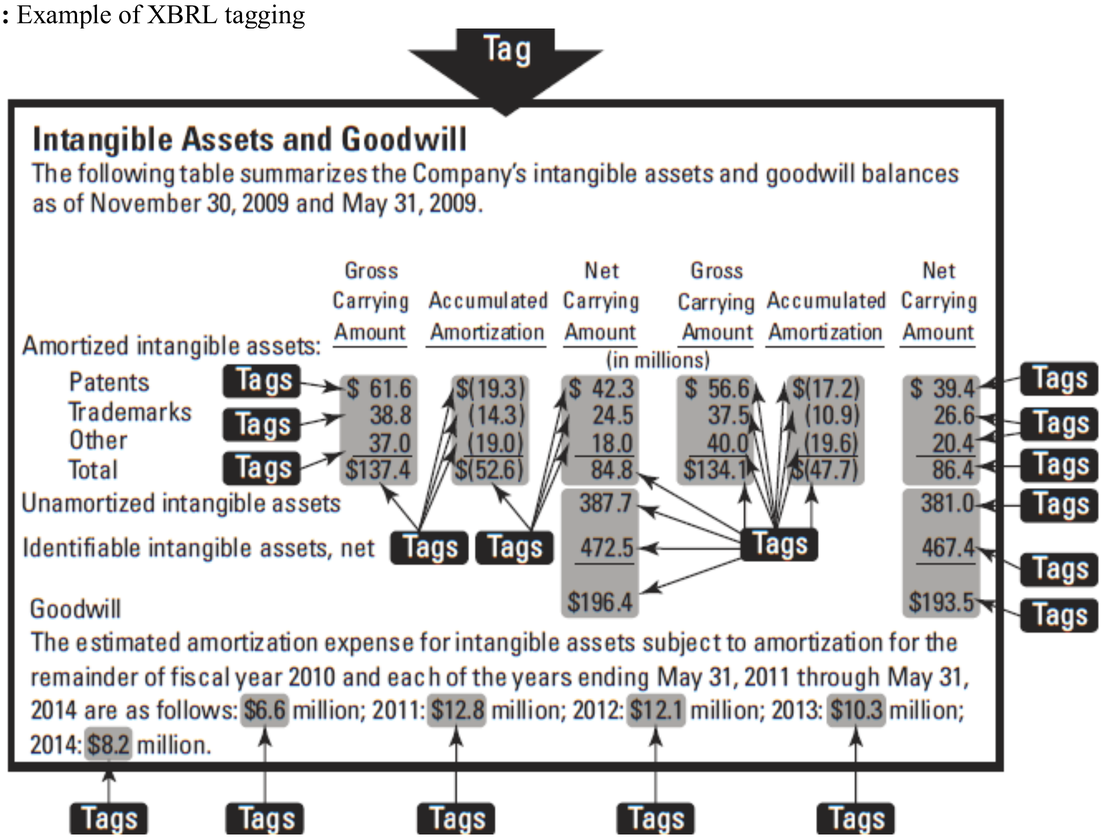


Figure 5 provides an example of XBRL tagging: Block tagging applies a single tag to a block of text. Detail tagging applies tags to every data point.[15]

15 Source: XBRL for Dummies by Ed Tittel

Figure 6: Illustration of continuous auditing

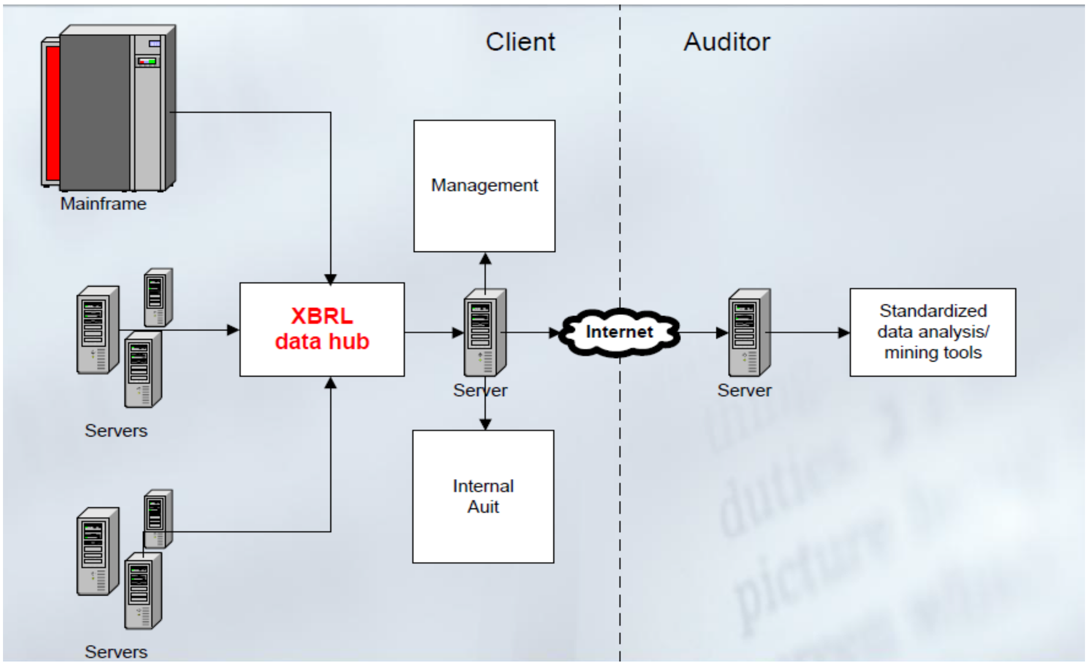


Figure 6 provides an illustration of the technical aspect of continuos auditing

Figure 7: Illustration of the XBRL audit framework

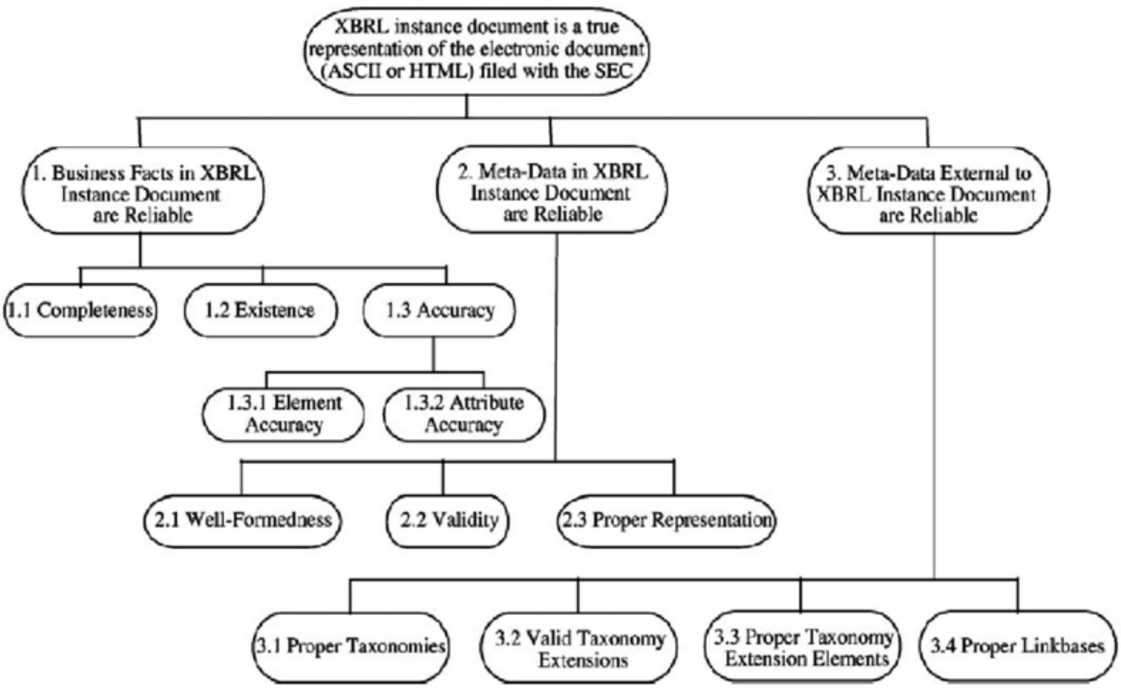


Figure 7 provides the XBRL audit framework as proposed by Srivastava & Kogan (2010)

Appendix II

II.i Model 1 Industry – Size

Table 10 shows the results of the initial regression. The R^2 is 0.0690 and the adjusted R^2 is 0.0683. The F-statistic equals 94.34 and the P-value is 0.000, indicating that the model is significant. I winsorize *M_B_ratio* at the 1st and 95th percentile, because of the outliers that are identified in the descriptive statistics (see *Section 5.3*). The regression is ran with the winsorized variable, I find that the R^2 and adjusted R^2 have increased to 0.0691 and 0.0698 respectively. The F-statistic for the winsorzied model is 95.6 and the P-value is 0.000. Then I check whether the residuals are normally distributed. Figure 8 shows the normal probability plot, the quantiles of normal probability plot and the kernel density plot. I also perform the Shapiro Wilk test to check for normal distribution of residuals (figure 9). The P-value of the test is 0.0000. Thus, the null-hypothesis is rejected and the residuals are non-normal. Furthermore, I check for heteroscedasticity by performing the Breusch-Pagan / Cook-Weisberg test (Figure 11). However, as the residuals are not normally distributed I also use the White's general test for heteroscedasticity (figure 12). For both tests the P-value is 0.000, thus the H_0 is rejected, meaning that the residuals are heteroscedastic. I check for multicollinearity by calculating the variance inflation factor (VIF) (figure 13). As the VIF equals 1.90 which is smaller than 10, I conclude that there is no multicollinearity in this model. I then check for endogeneity by regressing the predicted residuals on the explanatory variables of the model (figure 14). The coefficient of these explanatory variables are not significant (P-value of 1.000), therefore I conclude that endogeneity is not present in this model. Since, heteroskedasticity is present in the model I decide to run a robust regression command in Stata which is similar to computing White's standard errors. Table 12 shows the F-statistic of the model has decreased (50.76) but the model is still significant (P-value 0.000).

Table 10: Comparison of results of the initial regression and the winsorized sample of M1 (Industry – Size)

Dependent variable	Cumulative Abnormal Return (CAR)			
	Initial regression		Winsorized	
	Coefficient	P-value	Coefficient	P-value
Main variables				
XBRL	-0.0377***	(0.001)	-0.0372***	(0.001)
POST	0.0270***	(0.008)	0.0267***	(0.009)
*POST*XBRL*	-0.0294**	(0.031)	-0.0291**	(0.032)
Control variables				
Lsize	0.0125***	(0.000)	0.0119***	(0.001)
N_An	0.0002	(0.781)	0.0004	(0.499)
M_B_ratio (m)	-7.25E-06	(0.335)	-0.0040***	(0.001)
RND	-0.02760***	(0.000)	-0.0216***	(0.004)
STD_RET	6.0034***	(0.000)	5.9342***	(0.001)
Intercept	-0.3052***	(0.000)	-0.2804***	(0.001)
N	10,194		10,194	
F-statistic	94.34	(0.000)	95.6	(0.000)
Adj. R2	0.0683		0.0698	
R2	0.0690		0.0691	

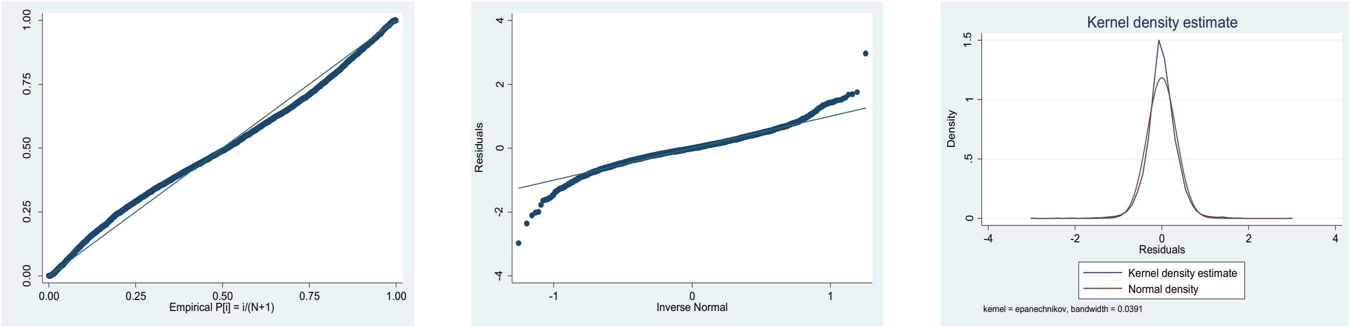
*, **, *** indicate statistical significance at the 10, 5 and 1% level, respectively
The `regress` command is used for this regression

Table 11: Detailed summary statistics of *M_B_ratio*

	Percentiles	Smallest
1%	-5.792871	-688.4559
5%	.8822425	-688.4559
10%	1.10534	-688.4559
25%	1.71258	-688.4559
50%	2.770043	
		Largest
75%	4.073565	103.9233
90%	6.650457	103.9233
95%	8.342748	103.9233
99%	17.87984	44843.56

The `summarize, detail` command is used for this regression

Figure 8: Normal probability plot (a), Quantiles of normal probability plot (b) and Kernel density plot (c) of M1 (Industry – Size)



a b c

The `pnorm (var), qnorm (var) and kdensity (var), normal` command are used for these plots.

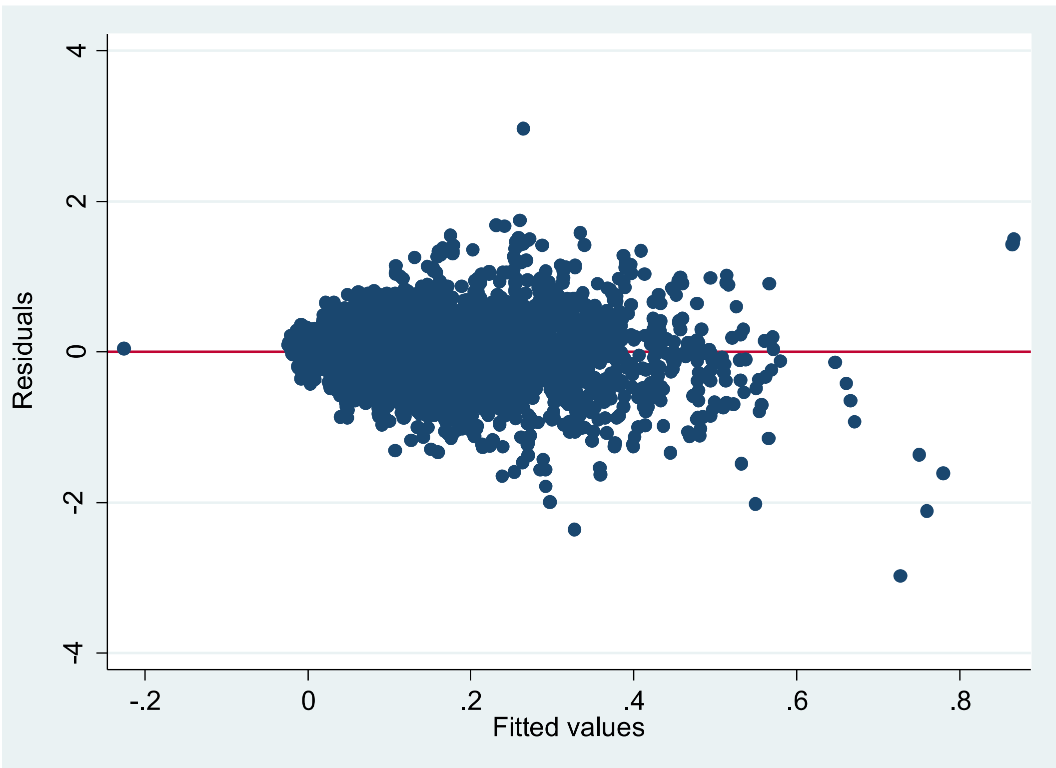
Figure 9: Shapiro Wilk test of M1 (Industry – Size)

Shapiro-Wilk W test for normal data

Variable	Obs	W	V	z	Prob>z
e	10,194	0.97039	149.870	13.418	0.00000

The `swilk (var)` command for this test.

Figure 10: Plot of residuals and fitted values



The `rvfplot, yline(0)` command is used for this test.

Figure 11: Breusch-Pagan / Cook-Weisberg test for M1 (Industry – Size)

```
Breusch-Pagan / Cook-Weisberg test for heteroskedasticity
         Ho: Constant variance
         Variables: fitted values of CAR

         chi2(1)      =  3049.17
         Prob > chi2  =   0.0000
```

The `estat hettest` command is used for this test.

Figure 12: The White's general test for heteroscedasticity for M1 (Industry – Size)

```
White's test for Ho: homoskedasticity
         against Ha: unrestricted heteroskedasticity

         chi2(37)     =   1675.13
         Prob > chi2  =    0.0000

Cameron & Trivedi's decomposition of IM-test
```

Source	chi2	df	p
Heteroskedasticity	1675.13	37	0.0000
Skewness	57.20	8	0.0000
Kurtosis	18.25	1	0.0000
Total	1750.59	46	0.0000

The `imtest, white` command is used for this test.

Table 12: Comparison of results of the initial regression and the robust regression of M1 (Industry – Size)

Dependent variable	Cumulative Abnormal Return (CAR)			
	Initial regression		Robust regression	
	Coefficient	P-value	Coefficient	P-value
Main variables				
XBRL	-0.0377***	(0.001)	-0.0377***	(0.005)
POST	0.0270***	(0.008)	0.0270**	(0.018)
*POST*XBRL*	-0.0294**	(0.031)	-0.0294*	(0.045)
Control variables				
Lsize	0.0125***	(0.000)	0.0125***	(0.001)
N_An	0.0002	(0.781)	0.0002	(0.786)
M_B_ratio	-7.25E-06	(0.335)	-0.0000***	(0.000)
RND	-0.02760***	(0.000)	-0.0276***	(0.000)
STD_RET	6.0034***	(0.000)	6.0034***	(0.000)
Intercept	-0.3052***	(0.000)	-0.3052***	(0.001)
N	10,194		10,194	
F-statistic	94.34	(0.000)	50.76	(0.000)
Adj. R2	0.0683			
R2	0.0690		0.0690	

*, **, *** indicate statistical significance at the 10, 5 and 1% level, respectively
The `regress, robust` command is used for this regression

Figure 13: VIF for M1 (Industry – Size)

Variable	VIF	1/VIF
POSTXBRL	3.31	0.302374
XBRL	2.75	0.363422
POST	2.26	0.441906
Lsize	1.69	0.590641
N_An	1.52	0.658875
STD_RET	1.50	0.665543
RND	1.18	0.848278
M_B_ratio	1.00	0.999637
Mean VIF	1.90	

The `vif` command is used for this test.

Table 13: Regression of predicted residuals on explanatory variables for M1 (Industry – Analyst)

Dependent variable	Predicted residuals (*e*)	
	Coefficient	P-value
Main variables		
XBRL	1.84E-10	1.000
ISH	7.32E-11	1.000
*ISH*XBRL*	-1.03E-10	1.000
Control variables		
Lsize	-7.56E-11	1.000
N_An	2.18E-11	1.000
M_B_ratio	3.21E-14	1.000
RND	-5.22E-11	1.000
STD_RET	2.91E-09	1.000
Intercept	1.21E-09	1.000
N	10,194	
F-statistic	0.00	1.000
Adj. R2	-0.0008	
R2	0.0000	

*, **, *** indicate statistical significance at the 10, 5 and 1% level, respectively
The `regress` command is used for this test

Figure 14: Spearman correlation of M1 (Industry – Size)

	CAR	XBRL	POST	POSTXBRL	Lsize	N_An	M_B_ra~o	RND	STD_RET
CAR	1.0000								
XBRL	-0.0473*	1.0000							
POST	-0.0987*	-0.0224*	1.0000						
POSTXBRL	-0.1022*	0.6437*	0.5171*	1.0000					
Lsize	-0.0370*	0.5514*	0.0195*	0.3635*	1.0000				
N_An	0.0286*	0.2909*	-0.0213*	0.1659*	0.5066*	1.0000			
M_B_ratio	-0.0458*	-0.0335*	0.1013*	0.0264*	0.1168*	0.2369*	1.0000		
RND	-0.0627*	0.1195*	0.0089	0.0807*	0.1978*	0.3099*	0.3823*	1.0000	
STD_RET	0.2219*	-0.0195*	-0.5317*	-0.3107*	-0.2164*	0.0825*	-0.1904*	-0.1125*	1.0000

*, indicates significance at the 5% level

The `spearman` command is used for this test

II.ii Model 1 Industry – Analyst

Table 14 shows the results of the initial regression. The R^2 is 0.0559 and the adjusted R^2 is 0.0552. The F-statistic equals 94.34 and the P-value is 0.000, indicating that the model is significant. I winsorize *M_B_ratio* at the 1st and 95th percentile, because of the outliers that are identified in the descriptive statistics (see *Section 5.3*). The regression is ran with the winsorized variable, I find that the R^2 and adjusted R^2 have slightly decreased to 0.0558 and 0.0551, respectively. The F-statistic for the winsorzied model is 85.62 and the P-value is 0.000. Then I check whether the residuals are normally distributed. Figure 15 shows the normal probability plot, the quantiles of normal probability plot and the kernel density plot. I also perform the Shapiro Wilk test to check for normal distribution of residuals (figure 16). The P-value of the test is 0.0000. Thus, the null-hypothesis is rejected and the residuals are non-normal. Furthermore, I check for heteroscedasticity by performing the Breusch-Pagan / Cook-Weisberg test (figure 17). However, as the residuals are not normally distributed I also use the White's general test for heteroscedasticity (figure 18). For both tests the P-value is 0.000, thus the H_0 is rejected, meaning that the residuals are heteroscedastic. I check for multicollinearity by calculating the variance inflation factor (VIF) (figure 20). As the VIF equals 1.89 which is smaller than 10, I conclude that there is no multicollinearity in this model. I then check for endogeneity by regressing the predicted residuals on the explanatory variables of the model (table 17). The coefficient of these explanatory variables are not significant (P-value of 1.000), therefore I conclude that endogeneity is not present in this model. Since, heteroskedasticity is present in the model I decide to run a robust regression command in Stata which is similar to computing White's standard errors. Table 16 shows the F-statistic of the model has decreased (64.42) but the model is still significant (P-value 0.000).

Table 14: Comparison of results of the initial regression and the winsorized sample of M1 (Industry – Analyst)

Dependent variable	Cumulative Abnormal Return (CAR)			
	Initial regression		Winsorized	
	Coefficient	P-value	Coefficient	P-value
Main variables				
XBRL	-0.0532***	(0.000)	-0.0529***	(0.000)
POST	0.0022	(0.21)	-0.0019	(0.850)
*POST*XBRL*	-0.0248*	(0.076)	-0.0245*	(0.079)
Control variables				
Lsize	0.0116***	(0.000)	0.0117***	(0.000)
N_An	-0.0078***	(0.004)	-0.0018***	(0.004)
M_B_ratio	0.0000	(0.260)	0.0007	(0.602)
RND	-0.0361***	(0.000)	-0.0396***	(0.000)
STD_RET	4.9800***	(0.000)	4.9961***	(0.000)
Intercept	-0.1814**	(0.012)	-0.1860**	(0.011)
N	11,599		11,599	
F-statistic	85.75	(0.000)	85.62	(0.000)
Adj. R2	0.0552		0.0551	
R2	0.0559		0.0558	

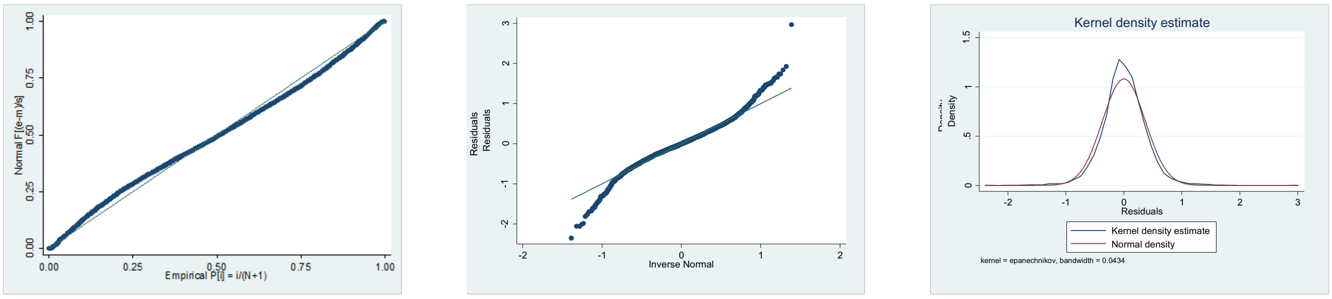
*, **, *** indicate statistical significance at the 10, 5 and 1% level, respectively
The `regress` command is used for this regression

Table 15: Detailed summary statistics of *M_B_ratio*

	Percentiles	Smallest
1%	-1.391292	-166.1596
5%	.7619807	-166.1596
10%	1.019149	-166.1596
25%	1.611524	-166.1596
50%	2.770043	
		Largest
75%	4.155347	103.9233
90%	7.465258	103.9233
95%	11.6448	44843.56
99%	21.15806	44843.56

The `summarize, detail` command is used for this regression

Figure 15: Normal probability plot (a), Quantiles of normal probability plot (b) and Kernel density plot (c) of M1 (Industry – Analyst)



a b c

The `pnorm (var), qnorm (var) and kdensity (var), normal` command are used for these

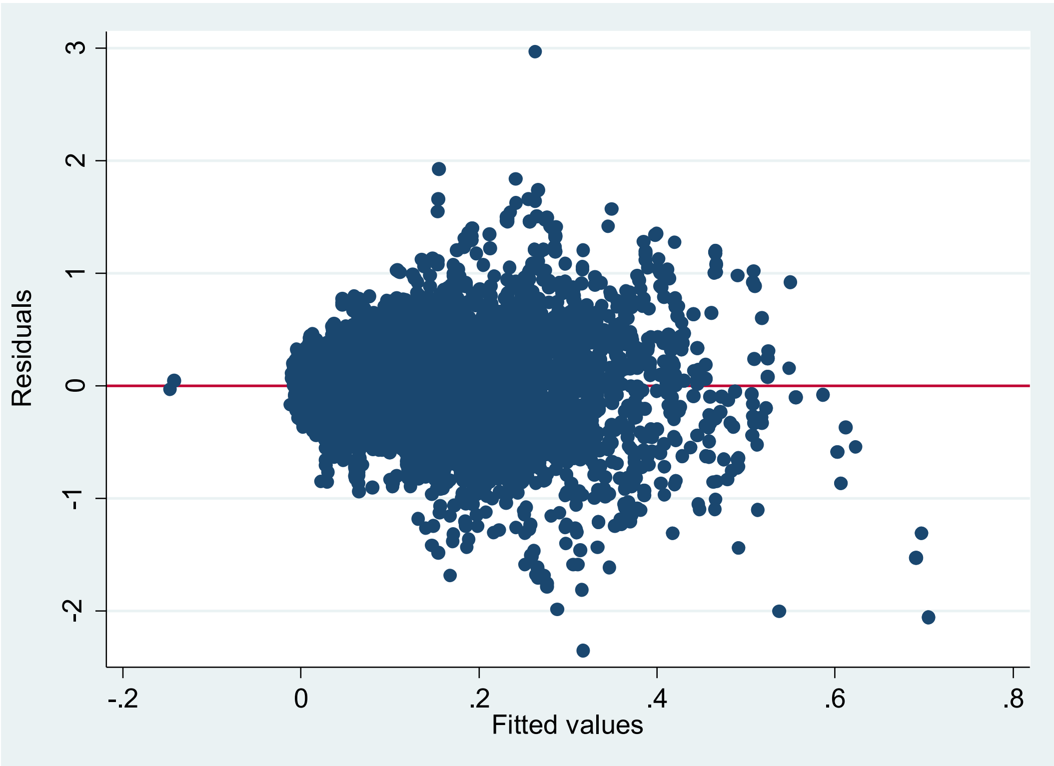
Figure 16: Shapiro Wilk test of M1 (Industry – Analyst)

Shapiro-Wilk W test for normal data

Variable	Obs	W	V	z	Prob>z
e	11,599	0.97662	132.221	13.127	0.00000

The `swilk (var)` command for this test.

Figure 17: Plot of residuals and fitted values



The `rvfplot, yline(0)` command is used for this test.

Figure 18: Breusch-Pagan / Cook-Weisberg test for M1 (Industry – Analyst)

```
Breusch-Pagan / Cook-Weisberg test for heteroskedasticity
         Ho: Constant variance
         Variables: fitted values of CAR

         chi2(1)      =  2686.23
         Prob > chi2  =   0.0000
```

The `estat hettest` command is used for this test.

Figure 19: The White's general test for heteroscedasticity for M1 (Industry – Analyst)

```
White's test for Ho: homoskedasticity
         against Ha: unrestricted heteroskedasticity

         chi2(37)     =  1798.17
         Prob > chi2  =   0.0000

Cameron & Trivedi's decomposition of IM-test
```

Source	chi2	df	p
Heteroskedasticity	1798.17	37	0.0000
Skewness	57.19	8	0.0000
Kurtosis	41.48	1	0.0000
Total	1896.84	46	0.0000

The `imtest, white` command is used for this test.

Table 16: Comparison of results of the initial regression and the robust regression of M1 (Industry – Analyst)

Dependent variable	Cumulative Abnormal Return (CAR)			
	Initial regression		Robust regression	
	Coefficient	P-value	Coefficient	P-value
Main variables				
XBRL	-0.0532***	(0.000)	-0.0532***	(0.000)
POST	0.0022	(0.21)	0.0022	(0.855)
*POST*XBRL*	-0.0248*	(0.076)	-0.0248*	(0.093)
Control variables				
Lsize	0.0116***	(0.000)	0.0116***	(0.001)
N_An	-0.0078***	(0.004)	-0.0018***	(0.001)
M_B_ratio	0.0000	(0.260)	0.0000***	(0.000)
RND	-0.0361***	(0.000)	-0.0361***	(0.000)
STD_RET	4.9800***	(0.000)	4.9800***	(0.000)
Intercept	-0.1814**	(0.012)	-0.1814**	(0.023)
N	11,599		10,599	
F-statistic	85.75	(0.000)	64.42	(0.000)
Adj. R2	0.0552			
R2	0.0559		0.0559	

*, **, *** indicate statistical significance at the 10, 5 and 1% level, respectively

The `regress, robust` command is used for this test.

Figure 20: VIF for M1 (Industry – Analyst)

Variable	VIF	1/VIF
POSTXBRL	3.21	0.311959
XBRL	2.84	0.352313
POST	2.19	0.456019
Lsize	1.77	0.565596
STD_RET	1.52	0.656871
N_An	1.36	0.732990
RND	1.19	0.840900
M_B_ratio	1.00	0.999043
Mean VIF	1.89	

The `vif` command is used for this test.

Table 17: Regression of predicted residuals on explanatory variables for M1 (Industry – Analyst)

Dependent variable	Predicted residuals (*e*)	
	Coefficient	P-value
Main variables		
XBRL	1.21E-10	1.000
POST	-2.97E-10	1.000
*POST*XBRL*	-3.41E-10	1.000
Control variables		
Lsize	6.38E-11	1.000
N_An	-1.31E-11	1.000
M_B_ratio	-2.52E-14	1.000
RND	1.58E-10	1.000
STD_RET	-8.19E-09	1.000
Intercept	-7.69E-10	1.000
N	11,599	
F-statistic	0.00	1.000
Adj. R2	-0.0007	
R2	0.0000	

*, **, *** indicate statistical significance at the 10, 5 and 1% level, respectively
The `regress` command is used for this test

Figure 21: Spearman correlation of M1 (Industry – Analyst)

	CAR	*XBRL*	*POST*	*POST*XBRL*	*Lsize*	*N_An*	*M_B_ratio*	*RND*	*STD_RET*
CAR	1.0000								
XBRL	-0.0762*	1.0000							
POST	-0.0957*	-0.0194*	1.0000						
*POST*XBRL*	-0.1175*	0.6448*	0.5085*	1.0000					
Lsize	-0.0542*	0.5832*	0.0572*	0.3947*	1.0000				
N_An	-0.0147	0.1405*	-0.0294*	0.0753*	0.3612*	1.0000			
M_B_ratio	-0.0238*	-0.1282*	0.1005*	-0.0358*	0.0052	0.2453*	1.0000		
RND	-0.0775*	0.0299*	-0.0157	0.0202*	0.0287*	0.2899*	0.4564*	1.0000	
STD_RET	0.2136*	-0.0933*	-0.5369*	-0.3587*	-0.3039*	0.0434*	-0.1966*	-0.0997*	1.0000

*, indicates significance at the 5% level

The `spearman` command is used for this test

II.iii Model 2 Industry – Size

Table 18 shows the results of the initial regression. The R^2 is 0.1235 and the adjusted R^2 is 0.1223. The F-statistic equals 108.20 and the P-value is 0.000, indicating that the model is significant. I winsorize *M_B_ratio* at the 1^{st} and 95^{th} percentile, because of the outliers that are identified in the descriptive statistics (see *Section 5.3*). The regression is ran with the winsorized variable, I find that the R^2 and adjusted R^2 have increased to 0.1356 and 0.1344, respectively. The F-statistic for the winsorzied model is 120.43 and the P-value is 0.000. Then I check whether the residuals are normally distributed. Figure 22 shows the normal probability plot, the quantiles of normal probability plot and the kernel density plot. I also perform the Shapiro Wilk test to check for normal distribution of residuals (rigure 23). The P-value of the test is 0.0000. Thus, the null-hypothesis is rejected and the residuals are non-normal. Furthermore, I check for heteroscedasticity by performing the Breusch-Pagan / Cook-Weisberg test (figure 25). However, as the residuals are not normally distributed I also use the White's general test for heteroscedasticity (figure 26). For both tests the P-value is 0.000, thus the H_0 is rejected, meaning that the residuals are heteroscedastic. I check for multicollinearity by calculating the variance inflation factor (VIF) (figure 27). As the VIF equals 7.54 which is smaller than 10, I conclude that there is no multicollinearity in this model. I then check for endogeneity by regressing the predicted residuals on the explanatory variables of the model. The coefficient of these explanatory variables are not significant (P-value of 1.000), therefore I conclude that endogeneity is not present in this model (table 21). Since, heteroskedasticity is present in the model I decide to run a robust regression command in Stata which is similar to computing White's standard errors. Table 20 shows the F-statistic of the model has decreased (87.61) but the model is still significant (P-value 0.000).

Table 18: Comparison of results of the initial regression and the winsorized sample of M2 (Industry – Size)

Dependent variable	Cumulative Abnormal Return (CAR)			
	Initial regression		Winsorized	
	Coefficient	P-value	Coefficient	P-value
Main variables				
XBRL	-0.0638*	(0.067)	-0.0669*	(0.053)
ISH	-0.2220***	(0.000)	-0.2111***	(0.000)
*ISH*XBRL*	-0.0249	(0.571)	-0.0196	(0.654)
Control variables				
Lsize	0.0196***	(0.000)	0.0178***	(0.000)
N_An	-0.0008	(0.191)	0.0015**	(0.013)
M_B_ratio (w)	-9.10E-06	(0.123)	-0.01155***	(0.000)
RND	-0.0060	(0.408)	-0.0103	(0.167)
STD_RET	8.0436***	(0.000)	8.0436***	(7.645)
Intercept	-0.3291***	(0.001)	-0.2678***	(0.005)
N	6,153		6,153	
F-statistic	108.20	(0.000)	120.43	(0.000)
Adj. R2	0.1223		0.1344	
R2	0.1235		0.1356	

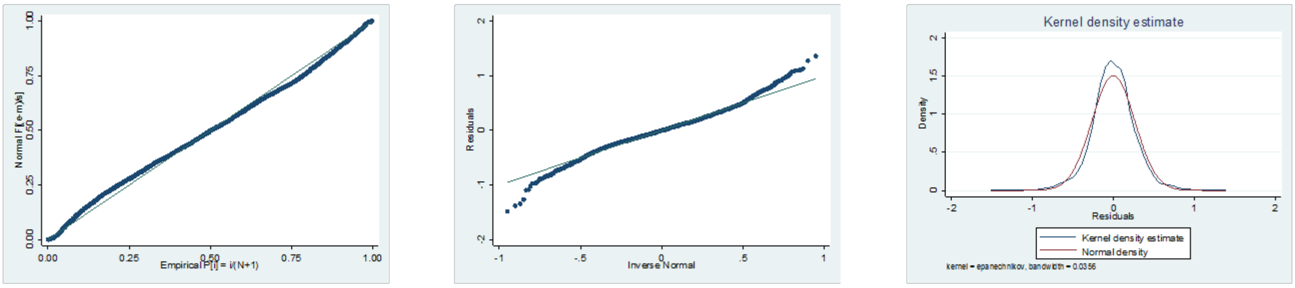
*, **, *** indicate statistical significance at the 10, 5 and 1% level, respectively
The `regress` command is used for this regression

Table 19: Detailed summary statistics of *M_B_ratio*

	Percentiles	Smallest
1%	-3.555318	-688.4559
5%	1.003936	-688.4559
10%	1.221935	-688.4559
25%	1.882418	-688.4559
50%	2.847281	
		Largest
75%	4.260368	103.9233
90%	6.650457	103.9233
95%	8.342748	103.9233
99%	16.00091	44843.56

The `summarize, detail` command is used for this regression

Figure 22: Normal probability plot (a), Quantiles of normal probability plot (b) and Kernel density plot (c) of M2 (Industry – Size)



a b c

The `pnorm (var), qnorm (var) and kdensity (var), normal` command are used for these plots.

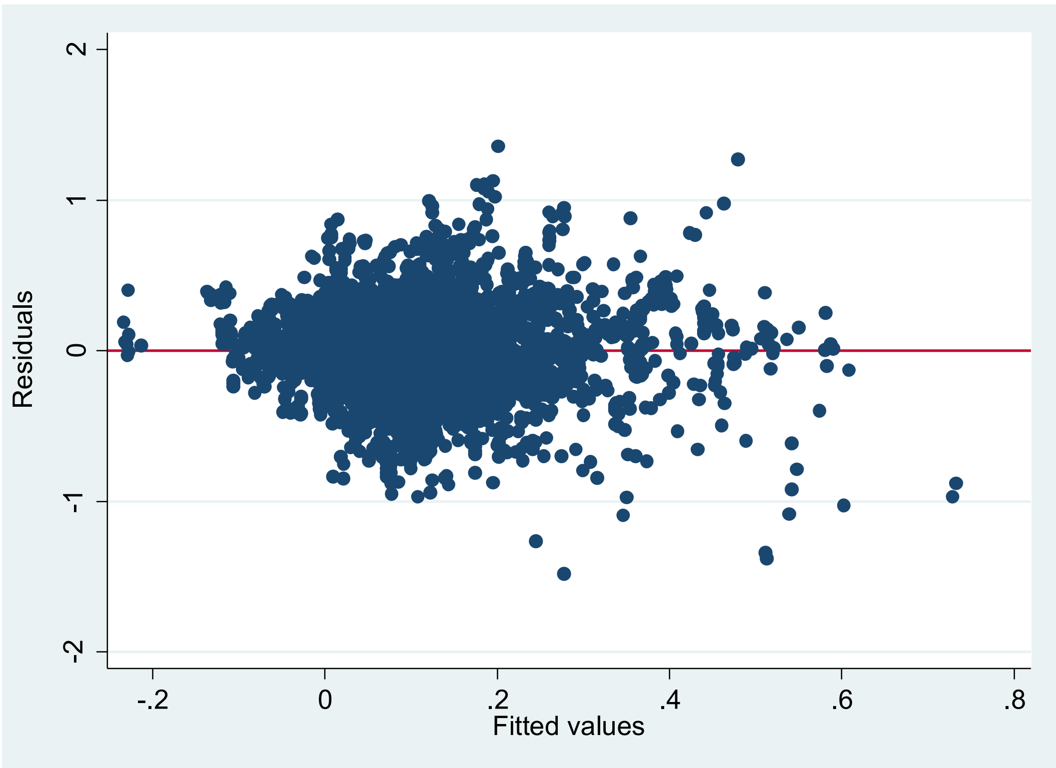
Figure 23: Shapiro Wilk test of M2 (Industry – Size)

Shapiro-Wilk W test for normal data

Variable	Obs	W	V	z	Prob>z
e	6,153	0.98418	51.522	10.409	0.00000

The `swilk (var)` command is used for this test.

Figure 24: Plot of residuals and fitted values



The `rvfplot, yline(0)` command is used for this test.

Figure 25: Breusch-Pagan / Cook-Weisberg test for M2 (Industry – Analyst)

```
Breusch-Pagan / Cook-Weisberg test for heteroskedasticity
         Ho: Constant variance
         Variables: fitted values of CAR

         chi2(1)      =   492.47
         Prob > chi2  =   0.0000
```

The `estat hettest` command is used for this test.

Figure 26: The White's general test for heteroscedasticity for M2 (Industry – Analyst)

```
White's test for Ho: homoskedasticity
         against Ha: unrestricted heteroskedasticity

         chi2(39)     =   1084.99
         Prob > chi2  =    0.0000

Cameron & Trivedi's decomposition of IM-test
```

Source	chi2	df	p
Heteroskedasticity	1084.99	39	0.0000
Skewness	75.55	8	0.0000
Kurtosis	46.42	1	0.0000
Total	1206.95	48	0.0000

The `imtest, white` command is used for this test.

Table 20: Comparison of results of the initial regression and the robust regression of M2 (Industry – Size)

Dependent variable	Cumulative Abnormal Return (CAR)			
	Initial regression		Robust regression	
	Coefficient	P-value	Coefficient	P-value
Main variables				
XBRL	-0.0638*	(0.067)	-0.0638**	(0.038)
ISH	-0.2220***	(0.000)	-0.2230***	(0.000)
*ISH*XBRL*	-0.0249	(0.571)	-0.0249	(0.530)
Control variables				
Lsize	0.0196***	(0.000)	0.0196***	(0.000)
N_An	-0.0008	(0.191)	0.0008	(0.235)
M_B_ratio	-9.10E-06	(0.123)	-0.0000***	(0.000)
RND	-0.0060	(0.408)	-0.0060	(0.388)
STD_RET	8.0436***	(0.000)	8.0436***	(0.000)
Intercept	-0.3291***	(0.001)	-0.3291***	(0.002)
N	6,153		6,153	
F-statistic	108.20	(0.000)	87.61	(0.000)
Adj. R2	0.1223			
R2	0.1235		0.1235	

*, **, *** indicate statistical significance at the 10, 5 and 1% level, respectively
The `regress, robust` command is used for this regression

Figure 27: VIF for M2 (Industry – Size)

Variable	VIF	1/VIF
XBRL	26.44	0.037818
ISHXBRL	25.39	0.039390
Lsize	2.18	0.459116
N_An	1.63	0.612200
ISH	1.36	0.736071
RND	1.16	0.862013
STD_RET	1.15	0.872824
M_B_ratio	1.00	0.998372
Mean VIF	7.54	

The `vif` command is used for this test.

Table 21: Regression of predicted residuals on explanatory variables for M2 (Industry – Analyst)

Dependent variable	Predicted residuals (*e*)	
	Coefficient	P-value
Main variables		
XBRL	-7.16E-11	1.000
ISH	6.44E-10	1.000
*ISH*XBRL*	-2.15E-10	1.000
Control variables		
Lsize	5.87E-11	1.000
N_An	7.95E-12	1.000
M_B_ratio	-3.28E-15	1.000
RND	-2.79E-10	1.000
STD_RET	1.25E-08	1.000
Intercept	-1.84E-09	1.000
N	6,153	
F-statistic	0.00	1.000
Adj. R2	-0.0013	
R2	0.0000	

*, **, *** indicate statistical significance at the 10, 5 and 1% level, respectively
The `regress` command is used for this test

Figure 28: Spearman correlation of M2 (Industry – Analyst)

	CAR	*XBRL*	*ISH*	*ISH*XBRL*	*Lsize*	*N_An*	*M_B_ratio*	*RND*	*STD_RET*
CAR	1.0000								
XBRL	-0.0933*	1.0000							
ISH	-0.1518*	-0.0218	1.0000						
*ISH*XBRL*	-0.1176*	0.9358*	0.1799*	1.0000					
Lsize	0.0057	0.5226*	-0.2755*	0.4019*	1.0000				
N_An	0.0508*	0.2644*	0.1532*	0.2901*	0.5033*	1.0000			
M_B_ratio	-0.0752*	-0.0398*	0.1723*	-0.0121	0.0938*	0.2312*	1.0000		
RND	-0.0640*	0.1124*	0.1731*	0.1314*	0.1523*	0.2453*	0.3821*	1.0000	
STD_RET	0.2226*	-0.0679*	0.1501*	-0.0114	-0.2827*	0.0964*	-0.1585*	-0.0960*	1.0000

*, indicates significance at the 5% level
The `spearman` command is used for this test

II.iv Model 2 Industry – Analyst

Table 22 shows the results of the initial regression. The R^2 is 0.0748 and the adjusted R^2 is 0.0736. The F-statistic equals 64.04 and the P-value is 0.000, indicating that the model is significant. I winsorize *M_B_ratio* at the 1st and 95th percentile, because of the outliers that are identified in the descriptive statistics (see *Section 5.3*).The regression is ran with the winsorized variable, I find that the R^2 and adjusted R^2 have slightly decreased to 0.0747 and 0.0735, respectively. The F-statistic for the winsorized model is 63.95 and the P-value is 0.000. Then I check whether the residuals are normally distributed. Figure 29 shows the normal probability plot, the quantiles of normal probability plot and the kernel density plot. I also perform the Shapiro Wilk test to check for normal distribution of residuals (figure 30). The P-value of the test is 0.0000. Thus, the null-hypothesis is rejected and the residuals are non-normal. Furthermore, I check for heteroscedasticity by performing the Breusch-Pagan / Cook-Weisberg test (figure 32). However, as the residuals are not normally distributed I also use the White's general test for heteroscedasticity (figure 33). For both tests the P-value is 0.000, thus the H_0 is rejected, meaning that the residuals are heteroscedastic. I check for multicollinearity by calculating the variance inflation factor (VIF) (figure 34). As the VIF equals 7.32 which is smaller than 10, I conclude that there is no multicollinearity in this model. I then check for endogeneity by regressing the predicted residuals on the explanatory variables of the model. The coefficient of these explanatory variables are not significant (P-value of 1.000), therefore I conclude that endogeneity is not present in this model (table 25). Since, heteroskedasticity is present in the model I decide to run a robust regression command in Stata which is similar to computing White's standard errors. Table 24 shows the F-statistic of the model has decreased (62.99) but the model is still significant (P-value 0.000).

Table 22: Comparison of results of the initial regression and the winsorized sample of M2 (Industry – Analyst)

Dependent variable	Cumulative Abnormal Return (CAR)			
	Initial regression		Winsorized	
	Coefficient	P-value	Coefficient	P-value
Main variables				
XBRL	-0.0246	(0.502)	-0.0220	(0.548)
ISH	-0.1718***	(0.000)	-0.1690***	(0.000)
*ISH*XBRL*	-0.0860*	(0.059)	-0.0894*	(0.050)
Control variables				
Lsize	0.0121***	(0.002)	0.0113***	(0.004)
N_An	-0.0016**	(0.017)	0.0016**	(0.010)
M_B_ratio (w)	-7.32E-06	(0.105)	-0.0019	(0.164)
RND	-0.0034	(0.657)	2.22E-06	(1.000)
STD_RET	5.9745***	(0.000)	5.9064***	(0.000)
Intercept	-0.1555*	(0.083)	-0.1356	(0.137)
N	6,350		6,350	
F-statistic	64.04	(0.000)	63.95	(0.000)
Adj. R2	0.0736		0.0735	
R2	0.0748		0.0747	

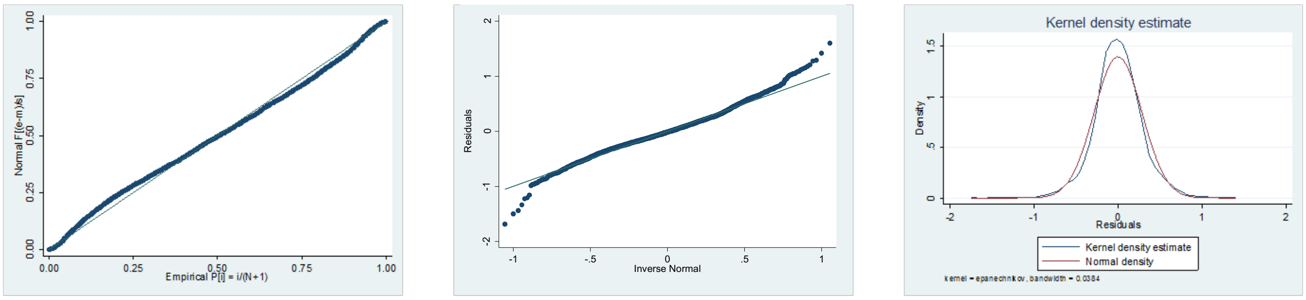
*, **, *** indicate statistical significance at the 10, 5 and 1% level, respectively
The `regress` command is used for this regression

Table 23: Detailed summary statistics of *M_B_ratio*

	Percentiles	Smallest
1%	.3390249	-163.622
5%	.9933265	-163.622
10%	1.220748	-163.622
25%	1.93626	-44.39538
50%	2.970129	
		Largest
75%	4.327537	103.9233
90%	7.873199	103.9233
95%	12.04145	44843.56
99%	30.0794	44843.56

The `summarize, detail` command is used for this regression

Figure 29: Normal probability plot (a), Quantiles of normal probability plot (b) and Kernel density plot (c) of M2 (Industry – Analyst)



a b c

The `pnorm (var)`, `qnorm (var) and kdensity (var), normal` command are used for these plots.

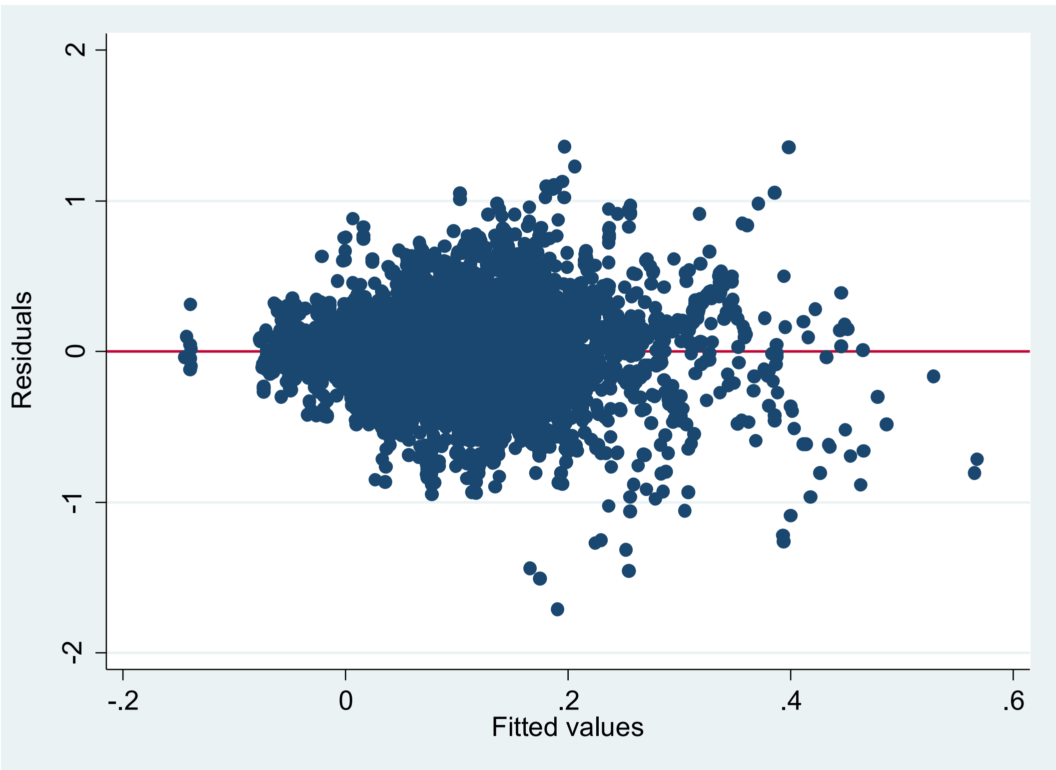
Figure 30: Shapiro Wilk test of M2 (Industry – Analyst)

Shapiro-Wilk W test for normal data

Variable	Obs	W	V	z	Prob>z
e	6,378	0.98574	47.941	10.229	0.00000

The `swilk (var)` command for this test.

Figure 31: Plot of residuals and fitted values



The `rvfplot, yline(0)` command is used for this test.

Figure 32: Breusch-Pagan / Cook-Weisberg test for M2 (Industry – Analyst)

```
Breusch-Pagan / Cook-Weisberg test for heteroskedasticity
         Ho: Constant variance
         Variables: fitted values of e

         chi2(1)      =    38.52
         Prob > chi2  =   0.0000
```

The `estat hettest` command is used for this test.

Figure 33: The White's general test for heteroscedasticity for M2 (Industry – Analyst)

```
White's test for Ho: homoskedasticity
         against Ha: unrestricted heteroskedasticity

         chi2(39)     =   1207.33
         Prob > chi2  =    0.0000

Cameron & Trivedi's decomposition of IM-test
```

Source	chi2	df	p
Heteroskedasticity	1207.33	39	0.0000
Skewness	81.21	8	0.0000
Kurtosis	37.26	1	0.0000
Total	1325.81	48	0.0000

The `imtest, white` command is used for this test.

Table 24: Comparison of results of the initial regression and the robust regression of M2 (Industry – Analyst)

Dependent variable	Cumulative Abnormal Return (CAR)			
	Initial regression		Robust regression	
	Coefficient	P-value	Coefficient	P-value
Main variables				
XBRL	-0.0246	(0.502)	-0.0246	(0.426)
ISH	-0.1718***	(0.000)	-0.1717***	(0.000)
*ISH*XBRL*	-0.0860*	(0.059)	-0.0860**	(0.027)
Control variables				
Lsize	0.0121***	(0.002)	0.0121***	(0.006)
N_An	-0.0016**	(0.017)	0.0016**	(0.019)
M_B_ratio	-7.32E-06	(0.105)	-0.0000***	(0.000)
RND	-0.0034	(0.657)	-0.0034	(0.642)
STD_RET	5.9745***	(0.000)	5.9745***	(0.000)
Intercept	-0.1555*	(0.083)	-0.1555	(0.134)
N	6,350		6,350	
F-statistic	64.04	(0.000)	62.99	(0.000)
Adj. R2	0.0736			
R2	0.0748		0.0748	

*, **, *** indicate statistical significance at the 10, 5 and 1% level, respectively

The `regress, robust` command is used for this test.

Figure 34: VIF for M2 (Industry – Analyst)

Variable	VIF	1/VIF
XBRL	25.93	0.038567
ISHXBRL	24.01	0.041654
Lsize	2.29	0.437102
N_An	1.44	0.693087
ISH	1.42	0.702365
STD_RET	1.22	0.820626
RND	1.18	0.845709
M_B_ratio	1.05	0.948422
Mean VIF	7.32	

The `vif` command is used for this test.

Table 25: Regression of predicted residuals on explanatory variables for M2 (Industry – Analyst)

Dependent variable	Predicted residuals (*e*)	
	Coefficient	P-value
Main variables		
XBRL	-2.54E-10	1.000
ISH	7.07E-11	1.000
*ISH*XBRL*	1.99E-10	1.000
Control variables		
Lsize	3.48E-11	1.000
N_An	1.52E-11	1.000
M_B_ratio	1.57E-14	1.000
RND	2.22E-10	1.000
STD_RET	2.19E-08	1.000
Intercept	-1.66E-09	1.000
N	6,350	
F-statistic	0.00	1.000
Adj. R2	-0.0013	
R2	0.0000	

*, **, *** indicate statistical significance at the 10, 5 and 1% level, respectively

The `regress` command is used for this test

Figure 35: Spearman correlation of M2 (Industry – Analyst)

	CAR	*XBRL*	*ISH*	*ISH*XBRL*	*Lsize*	*N_An*	*M_B_ratio*	*RND*	*STD_RET*
CAR	1.0000								
XBRL	-0.1163*	1.0000							
ISH	-0.1171*	-0.1578*	1.0000						
*ISH*XBRL*	-0.1370*	0.9396*	0.0348*	1.0000					
Lsize	-0.0327*	0.5666*	-0.3970*	0.4582*	1.0000				
N_An	0.0477*	0.1019*	0.1602*	0.1417*	0.3564*	1.0000			
M_B_ratio	-0.0253*	-0.1122*	0.1779*	-0.0799*	-0.0217	0.2261*	1.0000		
RND	-0.0404*	0.0374*	0.1868*	0.0630*	0.0346*	0.2290*	0.4060*	1.0000	
STD_RET	0.2049*	-0.2000*	0.1801*	-0.1395*	-0.4191*	0.0478*	-0.1337*	-0.0535*	1.0000

*, indicates significance at the 5% level
The `spearman` command is used for this test