경영학박사학위논문

Essays on Disclosure Activity and Corporate Behavior

공시와 기업의 행태에 관한 연구

2021 년 8 월

서울대학교 대학원
경영학과 경영학 전공
최 가 영

Essays on Disclosure Activity and Corporate Behavior

공시와 기업의 행태에 관한 연구

지도교수 백 복 현

이 논문을 경영학 박사학위논문으로 제출함

2021년 4월

서울대학교 대학원

경영학과 경영학 전공

최 가 영

최가영의 박사학위논문을 인준함

2021년 7월

위 원 장	이 우 종
부위원장	백 복 현
위 원	최 선 화
위 원	최 원 석
위 원	조 현 권

Abstract

Essays on Disclosure Activity and Corporate Behavior

Choi, Ga-Young
College of Business Administration
The Graduate School
Seoul National University

This dissertation is comprised of two essays on corporate disclosure activity and corporate behavior. The first essay, entitled "*Perceived Tax Audit Risk and Corporate Tax Behavior: Using Footnotes in 10-K Filings*", examines the association between firms' tax audit risk and their tax behavior. Utilizing Latent Dirichlet Allocation (LDA), I measure the tax audit risk perceived by each firm as the extent of tax audit discussion in firms' 10-K filings. I find that firms comply more with tax laws when they are subject to stronger tax audit risks. I validate that the tax audit discussion in 10-K filings captures tax audit risk rather than a rhetoric of managers by showing the correlation with the IRS audit probability. The results are robust when including both tax audit topic and the IRS audit probability, implying that firms' tax audit discussion delivers incremental information to the IRS audit rate. Also, I find that there exists heterogeneity in the effect of tax audit risk across firms. The result is attenuated when there are stronger external monitoring mechanisms and when the firm operates in foreign jurisdictions.

The second essay, entitled "*The Effect of Financial Reporting Readability on Foreign Investors' Information Acquisition Activity*", examines whether and how financial reporting

readability enhances foreign investors' information acquisition activity. I contend that more readable financial reports facilitate foreign investors' information collection activities, and this in turn increases foreign investors' investment in U.S. stocks. As anticipated, I find that foreign investors more frequently access 10-K filings that are more readable. Their frequent access leads to higher foreign institutional ownership of U.S stocks in the subsequent period. I also find that the positive association between foreign access and financial reporting readability is stronger when readers are from countries of which language shares more similarities with English. Overall, findings suggest that financial reporting readability plays an important role in facilitating foreign investment, further highlighting the benefit of making financial reports more readable.

Keywords: Tax enforcement, Tax compliance, XBRL, Machine learning, Language, Information acquisition, Foreign ownership, Institutional investors, Readability, Information disadvantage, Home bias, BOG index

Student Number: 2017-37066

Table of Contents

Essay 2. The Effect of Financial Reporting Readability on Foreign Investors' Information Acquisition Activity

List of Tables and Figures

Essay 1. Perceived Tax Audit Risk and Corporate Tax Behavior: Using Footnotes in 10-K Filings

Essay 2. The Effect of Financial Reporting Readability on Foreign Investors' Information Acquisition Activity

Essay 1

Perceived Tax Audit Risk and Corporate Tax Behavior

: Using Footnotes in 10-K Filings

1. Introduction

Tax is one of the significant factors that may affect firms' future liability. For instance, in their 10-K filing of 2020, Apple states "The Company is subject to laws and regulations worldwide, changes to which could increase the Company's costs and individually or in the aggregate adversely affect the Company's business.", and they explicitly list tax as one of the risk factors as such laws and regulations. As a major component of the tax system, tax enforcement[1] is effective in securing tax compliance (Allingham and Sandmo 1972; Alm et al. 1992; Slemrod and Yitzhaki 2002) and exerts substantial risks to tax payers by monitoring and potentially challenging their tax behavior. However, few archival research exists examining tax enforcement in corporate setting due to the limitation of firm-level data.

Moreover, although firms in general are exposed to the risk of tax enforcement[2], the extent to which each firm perceives risk from it varies, and consequently the effect of tax enforcement on a firm's tax behavior may vary as well. Notably, large firms are continuously under tax examinations by the IRS, but their perception on the tax enforcement in the current year may deviate from that in past years. This raises the need to look into the firm-year specific risk from the tax audit. In this research, I examine how much risks firms perceive from tax authorities and, more importantly, how firms react to the tax audit risk.

---

[1] In this research, tax enforcement, tax audit and tax assessment are used interchangeably.

[2] For example, the detailed selection process of the IRS audit is known to be a black box. However, all of the public and private firms filing tax returns to the IRS can be the subject of tax audit.

It is generally unobservable of which firms are under tax audit and how much risk they perceive from it unless they disclose. I capture the extent of firms' discussion on tax audit issue (hereafter, tax audit topics) using Latent Dirichlet Allocation (LDA), a machine-learning based topic classification[3]. Using the tax audit topics as the proxy for tax audit risk, I investigate the extent to which firms perceive risks from tax authorities and how they react to this risk by examining the association between the extent of tax audit discussion and their tax behavior. If the tax audit discussion captures firms' perceived tax risk and firms change their tax behavior accordingly, I would expect a positive association between the extent of tax audit discussion and tax compliance. If the tax audit discussion does not capture firms' perceived tax audit risk or firms are not sensitive to it, there will be no association between the two. Using GAAP effective tax rate (ETR) and cash effective tax rate (CETR), I find that firms on average reduce their GAAP effective tax rate (ETR) and cash effective tax rate (CETR) when they discuss tax audit issues more[4].

Next, if firms' tax behavior varies across firms by the extent of tax audit risk they perceive, then the next question is what drives the heterogeneity among

[3] The motivation of using the topic modeling is similar to that of readability. I cannot directly measure whether the tax footnote related to tax audit is more or less readable because the tax footnote is a set of various tax-related issues. Instead, I identify how much of the tax footnote is attributable to tax audit topic, which are similar to the length measure used as an alternative proxy for readability. For instance, Loughran and McDonald (2014) suggests that file size, which is directly associated with the length of a document can be a good proxy for readability.

[4] It is a joint test because the underlying latent factor of the association between the extent of tax audit discussion and tax behavior is the tax audit risk the firms perceive. However, such association is important because if the tax footnote is a boilerplate or firms do not react to the tax audit risk, there will be no association between the two. As such, the test shows whether the tax audit risk measured by LDA is a good proxy for perceived tax audit risk and whether firms change their tax behavior in accordance with the risk.

them. I investigate the factors associated with firms' perception on tax audit risk. Even when the tax authority imposes a similar level of tax audit risk on firms, the extent to which each firm perceives risks from it may vary, thus affecting their tax compliance level differently. Considering that tax authorities themselves are subject to limited resources, it is important to understand which firm characteristics are associated with the deterrent effect of tax audit to secure effective and efficient tax enforcement. The cross-sectional tests can also provide a hint on the mechanism behind the tax audit risk. Therefore, I explore factors that may be associated with the sensitivity of deterrent effect.

The first set of cross-sectional tests investigates the effect of institutions on the firms' perception of tax audit risk. Considering that both institutional ownership and tax audit act as a monitoring mechanism toward firms, the effect of tax audit would be less pronounced for firms with high level of institutional ownership (Hoopes et al. 2012). Consistent with the conjecture, I find that firms perceive less risks from tax enforcement when they have a high level of institutional ownership. It is also possible that firms with high institutional ownership do not affect the association between tax audit and tax behavior, since firms' noncompliance of tax laws and its resulting tax savings can benefit the institutional ownership by enhancing the after-tax income (Khan et al. 2017). In prior literature, mixed evidence exists on whether and how institutional ownership affects corporate tax decision (Khurana and Moser 2013; Khan et al. 2017) and I expect the test can provide a hint on the shareholders' tax preferences and its association with another external monitor.

Secondly, I test whether the association of tax audit risk and tax behavior is different between domestic and multinational firms. Given that international transactions, such as income shifting across countries or tax haven operations, act as important means for tax strategy (Clausing 2009; Klassen and Laplante 2012), domestic firms may not invest in tax strategy as much as multinational firms do. As such, domestic firms may react less strongly to the tax audit risk than multination firms because they do not have enough exposures to risky tax strategies in the first place. On the contrary, the greater opportunities multinational firms have in implementing tax strategies may enable them to find ways to mitigate the tax audit risk without cutting the investments on the tax strategies. Consistent with the latter, I find that multinational firms are less sensitive to the tax audit than domestic firms.

While the aforementioned results indicate that the tax-footnote based measure provides a reasonable proxy for the perceived tax risk, it is also likely that the managers use the tax footnote as rhetoric that justifies uncertainties of tax outcome, rather than show the effect of tax audit risk on their tax strategy. I validate the tax audit topic by investigating the correlation with the IRS audit probability from the Transactional Records Access Clearinghouse (TRAC)[5]. The Pearson correlation between the tax audit topic and the TRAC measure is 0.20 and is statistically significant at 1% level. Moreover, when regressing the tax behavior on both tax audit topic and TRAC measure, I find that the coefficient on tax audit topic remains statistically significant, indicating that the topic-based measure is effective in capturing the tax audit risk incremental to the TRAC measure.

[5] There is no firm-level measure of tax audit and prior literature mostly relies on the TRAC measure which is based on the size group-year level. Refer to Section 3 for detailed explanation on the measures.

This study contributes to the literature as follows. First, my research adds to the tax enforcement studies. Although prior literature finds that tax enforcement is effective in securing the tax compliance for individuals and small businesses (representatively, Allingham and Sandmo 1972), there have been few empirical findings on whether firms change their tax behavior when they are subject to heightened tax audit risks. This paucity of research is in part attributable to the difficulties in capturing the tax audit risk at firm-level. This study fills the void in the literature by showing that firm-specific tax audit risks are associated with more compliance with tax law in the current period. The cross-sectional results also expand the understanding by reporting that taxpayers of certain characteristics are more or less sensitive to the tax audit risk.

Second, it contributes to disclosure literature. Prior studies report mixed evidence on whether firms' footnote disclosure of 10-K filings provides useful information to its users. I find that tax footnotes entail information about the risks from tax authorities. Not all firms perceive tax audit as a credible threat that is significant enough to change their behavior. I identify the perceived tax audit risk using a machine learning methodology called LDA. Rather than obfuscating tax related information or merely justifying the uncertainty of tax outcome, the tax audit topic provides useful information that can be of help to information users in assessing firms' tax strategies.

Third, this study provides policy implications. Tax authorities are subject to limited budget. For example, IRS has experienced budget curtailment, especially after 2010, which led to a shrink in the scope of tax audit and eventually a decline in the tax revenues collected (Nessa et al. 2020). The finding that firms' perceived

tax audit risk differs by their characteristics implies that tax audit itself may not be sufficient enough to fully mitigate the horizontal inequity among firms[6]. Tax authorities can refer to these factors to facilitate more efficient and effective tax audit.

However, I acknowledge that there exist limitations in this study. First, the results do not exclude the possibility that firms' disclosure policy is associated with firms' tax behavior. Since the measure relies on firms' tax footnotes in 10-K Filings, the incentives to disclose or not to disclose the tax audit may be associated with firms' tax behavior[7]. I try to validate the measure by showing the positive and significant correlation between tax audit topic and the IRS audit probability. Moreover, when regressing tax compliance on both tax audit topics and IRS audit probability, the results are robust, implying that tax audit topic provides incremental information of tax audit risk. Second, the U.S. firms are subject to the IRS audit as well as state-level audit and audit from foreign tax authorities. My measure captures the tax audit topic as a whole, rather than specifying each tax audits. The premise is that tax audits from each jurisdiction exerts similar effect on firms' tax behavior. Yet, it is still possible that different rationale exists behind each tax audit and the scope of jurisdiction where firms file their tax return (e.g., whether firms operate

[6] Tax principles require taxpayers of similar taxable income have similar level of tax burden ("horizontal equity") and those of different taxable income face with different level of tax burden ("vertical equity"). The ability of firms in evading taxes incurs horizontal inequity in that firms with similar income results in different tax outcome. Tax enforcement alleviates this inequity by deterring tax evasion.

[7] Specifically, the tax audit topic does not identify firms that perceive their tax audit risk high while not discussing about it in their tax footnotes. However, since my measure focuses on how firms "perceive" the tax audit risk, whether firms are under tax audit or not is not a critical issue, although most cases of high measure overlap with the actual tax audit. Firms that expect to be under tax audit in the future or that have been tax audit in the past may also discuss about it in their tax footnotes if they perceive it as a risk.

only in the U.S. or in various jurisdictions around the world) may be associated with tax behavior.

The remainder of this paper proceeds as follows. Section 2 reviews related literature and develops hypothesis. Section 3 discusses the data, research design, and descriptive statistics. Section 4 reports the empirical results and Section 5 concludes.

2. Related Literature and Hypothesis Development

2.1. The Effect of Tax Enforcement

Tax enforcement is one of the major components that constitutes tax system – law, regulation, and procedures (Slemrod and Gillitzer 2014). A tax system assigns detailed procedures such as sending letters to taxpayers, mandating information reporting[8], and conducting face-to-face tax audit, to enforce tax liabilities to taxpayers. Among these various types of enforcement tool, tax audit is the most powerful one that monitors firms' tax evasion. The main objective of tax audit is to review taxpayers' financial information and tax payments to ensure that the taxpayer is reporting and paying taxes in accordance with laws and regulations. Prior literature finds evidence that tax audit is effective in deterring the noncompliance of taxpayers. Allingham and Sandmo (1972) show that individual taxpayers determine the extent of tax noncompliance by the tax rate, probability of detection, penalty, and risk preferences. Taxpayers evade taxes only when a positive outcome is expected, and the optimal level of tax evasion diminishes as the

[8] For example, FATCA, disclosing foreign subsidiaries (Exhibit 21), etc. (Hoopes, Robinson, and Slemrod 2018)

probability of detection rises (Slemrod 2019). In a corporate setting, Hoopes et al. (2012) find using the TRAC database that tax avoidance of U.S. public firms is deterred by higher probability of tax audit.

Tax enforcement also influences corporate governance. Heightened tax enforcement constrains resources that managers can divert and facilitates more to outside shareholders, leading to an increase in firm value and performances and a decline in cost of equity capital (Desai, Dyck, and Zingales 2007; Mironov 2013; El Ghoul, Guedhami, and Pittman 2011). Also, tax enforcement has a spillover effect on firms' information environment (Hanlon, Hoopes, and Shroff 2014; Bauer, Fang, and Pittman 2020). When a firm is under tax examination, tax authorities have access to a vast document including confidential data of the firm, which incurs a monitoring role not only to tax liabilities but also to the overall information environment.

Prior studies suggest that tax enforcement plays a significant role for firms, especially for small businesses. The risk aversion of individual and small-sized firms incurs them to be more deterred by tax enforcement, since it is costlier for them to evade taxes compared to large businesses[9] (Allingham and Sandmo 1972; Slemrod 2019). Small businesses suffer from lack of sufficient outside monitors and well-organized financial reporting, and the effect of tax enforcement on governance and information environment can be more pronounced for these firms. Benefited from enhanced transparency, small and midsized enterprises (SME) experience an increase in their bank lending (Gallemore and Jacob 2020). Small

[9] The separation of ownership and control makes firms risk neutral, implying that tax enforcement may not be effective for them compared to small businesses.

firms that are subject to tax audit are likely to have some real changes, such as revenue declines and more conversion to the S corporations (Belnap et al. 2020).

2.2. Textual Information in Corporate Disclosure

Textual information has doubled in its length during the past 20 years (Dyer et al. 2017). Textual information is one of the crucial sources of information in corporate disclosures along with quantitative data. For example, firms provide detailed information regarding the financial numbers in the footnotes, manager's interpretation and expectation on the performance in MD&A, risk factors in section 1A of 10-K filings, etc. Even though the length of text disclosure continues to hike, there exists mixed evidence on the usefulness of text disclosure.

A line of literature suggests that a firm's text disclosure contains information that is useful to investors. Firms' risk disclosures (i.e., Section 1A of 10-K filings) reflect firm-specific risk types and firms under higher risk discuss more risk factors (Campbell et al. 2014). The text disclosure is also useful in capturing the financial constraints of firms (Bodnaruk, Loughran and McDonald 2015). On the other hand, another line of literature contradicts the informativeness of text disclosures. The text information contains boilerplate and redundant sentences and the information becomes less specific (Loughran and McDonald 2014; Cazier and Pfeiffer 2015; Hope et al. 2016; Dyer et al. 2017). Managers have incentives to obfuscate their disclosure so that the bad news is less likely to be incorporated into the price (Li 2008). Managers may write a long disclosure to obscure or attribute its bad news or to distract the investors from the news (Bloomfield 2008). Managerial discretions over text disclosure imply that it lacks

sufficient informativeness by failing to represent substances.

2.3. Hypothesis Development

Tax audit is a powerful enforcement tool that significantly affects firms' tax behavior. In the seminal work of Allingham and Sandmo (1972), taxpayers comply more with laws and regulations when the probability of detection is high. Following literature finds consistent evidence with this deterrent effect of tax enforcement. Firms reduce tax avoidance under the high level of IRS audit probability (Hoopes et al. 2012). Managers regards ''Risk of detection and challenge by the IRS'' as the third-ranked reason not to be engaged in tax planning (Graham et al. 2017). However, the effectiveness of tax audit may vary across firms. Some firms can be more sensitive to tax audits, which means that they are more likely to be compliant with tax laws and regulations when subject to the tax audit compared to others under tax audit. To the extent that firms perceive a credible threat from tax audits, they would change their tax behavior in a way more compliant with tax laws and regulations. Moreover, large firms are subject to continuous enforcement by the Internal Revenue Service (IRS)[10]. The IRS continuously monitors large firms to see whether they are compliant with tax laws and regulations. This raises the possibility that the existence of tax audit itself might not be enough in explaining the actual

[10] The IRS monitors large firms as the Large Corporate Compliance (LCC) program which replaced the Coordinated Industry Case (CIC) program in 2019. The LCC program is designed to identify non-compliant corporate taxpayers in a more objective way using data analytics. Previous program, the Coordinated Industry Case (CIC), allows the IRS to identify the largest taxpayers and apply an increased scrutiny to large firms.

threat firms perceive from tax enforcement.

Firms describe their risk in their disclosures, and tax audit is one of the risk the firms are exposed to. Tax audit is often regarded as an unfavorable event that can challenge the tax position of the firm and unveiling this bad news can incur substantial litigation costs or reputation costs to managers (Skinner 1994). Firms' text disclosure reveals their substances, such as the risk factors (Campbell et al. 2014) or the financial constraint (Bodnaruk et al. 2015). Similarly, discussions about tax audit in the tax footnote may present the threat, which likely change the firms' tax risk, enforced by tax authority insiders perceive. On the contrary, the extent of discussion related to tax audit may not capture firms' perceived tax threat. When managers expect a higher threat and consequent negative outcome from unveiling their tax audit, they may not clearly disclose. Rather, they may obscure, attribute, or use longer discussion (Bloomfield 2008). In addition, literature shows that firms use more complex languages when the underlying information is more complex (Bushee et al. 2018). Firms may discuss tax audit because the nature of it is complex, regardless of the perceived threat. These raise the possibility that the tax footnote does not provide information on firms' perceived threat.

Considering the above discussion, it is *ex ante* not clear whether the corporate tax audit topic from the firm's 10-K filings is associated with more tax compliance. I test the extent to which a firm perceives a threat from tax audit and how the firm reacts to it in a combination with the following hypothesis:

H1: *Firms' tax audit topic is not associated with tax behavior*.

3. Data, Research Design, and Descriptive Statistics

3.1. Measuring the Tax Enforcement Risk

One of the obstacles in conducting corporate tax enforcement research is that there is no publicly available firm-level data that directly measure corporate tax audits. Prior literature often uses the IRS audit probability from the Transactional Records Access Clearinghouse (TRAC), a non-profit research organization at Syracuse University (Guedhami and Pittman 2008; El Ghoul et al. 2011; Hoopes et al. 2012; Hanlon et al. 2014; Bauer et al. 2017; Gallemore and Jacob 2020; Yost and Shu 2020). Audit probability from TRAC is not a firm-level variable but it varies by size group and calendar year. Firms' assets are categorized into 12-group[11]. TRAC reports the number of audits completed in year *t* divided by the total number of returns filed in each size-year group in year *t-1*.

Several limitations exist in the IRS audit probability from TRAC. Size-time variation is very crude and it entails innate measurement errors. Also, large firms in the U.S. are often included in the IRS' Large Corporate Compliance (LCC)[12] and are subject to continuous audit by the IRS. However, the audit probability measure does not incorporate such cases. In addition, it counts only the audit completed in a certain year. The IRS audit may last over a year, especially for large firms, and the beginning year and the ending year of the IRS audit may not always coincide with

[11] Firm asset was used to be categorized into 8 groups by 2009. From 2010, the largest size group ($250 Million Assets or more) is segmented into 5 groups. Since the sample period of this study begins in 2011, it is available to use the 12-group categorization. But I use both the 8-group and 12-group categorizations to compare with prior literature.

[12] LCC program has replaced the Coordinated Industry Case (CIC) program in 2019.

each other. It implies that the IRS audit probability from TRAC has limitation in capturing the actual timing of audit. In addition, firms are subject to tax audit from multiple tax authorities. The IRS and the state's Department of Revenue both have the authority to conduct the audits and foreign tax authorities also may audit MNCs' tax returns. Thus, relying only on the IRS audit may misguide the tax audit risk firms perceive.

To overcome the drawbacks, I develop a firm-level measure of tax audit risk based on the publicly available financial statements. Firms often disclose that they are being audited by tax authorities. Gleason and Mills (2002) report that firms often miss disclosing their contingent liabilities from tax audit, even though the amount is significant. Firms may fail to disclose the exact numerical data related to tax audit. However, I find that many firms disclose that they are being audited by tax authorities in their tax footnotes. Specifically, it partly attributes to the implementation of eXtensible Business Reporting Language (XBRL), which has incurred an increase in firm's footnote disclosure (Blankespoor 2019). Utilizing XBRL, I collect quantitative information in firm's tax footnotes of 10-K filings. Then, I identify the topics constituting each firm's tax footnote using a machine learning technique called Latent Dirichlet Allocation (LDA) (Blei et al. 2003).

LDA is a topic modeling methodology that can answer "what is being said" in disclosures (Huang et al. 2018). LDA views that a document is a combination of topics. It assumes that a topic is a probability distribution of words. LDA uncovers the latent topic distribution of a document by iteratively tracing back how a document is constituted until it finds the best set of variables regarding the topic and word distributions. In conducting the LDA analysis, the researcher has to

choose how many topics to discover from a document. I tried several sets of topic numbers until I decide to identify 20 topics[13]. Each topic distribution reports the extent to which a specific topic is discussed in a firm's 10 K filing. Among 20 topics, 3 topics are found to be associated with tax audit[14]. I aggregate the ratio of 3-topic distribution and use it as *Tax_audit* proxy in this research.

One of the strengths of *Tax_audit* is that it is a firm-specific measure. It directly measures how much the firm perceives its own tax audit risk. Another strength is that it is a contemporaneous proxy that captures the exact timing of tax audit. Also, I find that multinational firms disclose not only the tax audit by the IRS, but also the audit from the state or foreign tax authorities. Considering that a significant portion of U.S. public firms operates in foreign jurisdictions, this clearly provides an advantage over other measures.

3.2. Sample

The sample consists of U.S. public firms from 2011 to 2019. I exclude financial institutions and utilities (SIC Codes 4900-4999 and 6000-6999), since tax enforcement implication may differ in those firms. The sample period starts from 2011, because XBRL reporting has become widely available to the public since then[15]. The denominator of ETR is pre-tax earnings and it requires to be positive to calculate the ETR. Thus, the sample consists of firms reporting non-negative

[13] Results are qualitatively similar when choosing 10 topics.
[14] Refer to the Appendix 1 for the most probable words that constitute each topic.
[15] XBRL regulation has been implemented in April 2009. The implementation process consists of three phases, from Phase 1 requiring the largest accelerated filers in 2009 to Phase 3 requiring every U.S. public firm of filing requirement to adopt XBRL regulation in 2011. Since the phase categorization is based on firm size, using observations before 2011 would be subject to sample selection. Thus, I decide the sample to begin in 2011.

earnings. Firm-specific tax enforcement variable is constructed using tax footnotes in the 10-K filings from XBRL data. I obtain financial information from Compustat. Detailed sample selection procedures are described in Table 1. After requiring non-missing data for effective tax rate, tax enforcement, and control variables, final sample consists of 10,751 firm-years. For cross-sectional analysis, I manually identify the institutional ownership, which further limits the sample size of effective tax rate to 8,405 firm-years.

<Insert Table 1 here>

3.3. Regression Design

To examine whether and to what extent corporate tax audit risk deters firms' noncompliance of tax laws, the following model is estimated:

$$Tax\_compliance_{i,t} = \beta_0 + \beta_1 Tax\_audit_{i,t} + \gamma_k \sum_k Control_{k,i,t} + \epsilon_{i,t} \qquad (1)$$

where the subscript *i* and *t* index firms and years, respectively. To capture a firm's tax compliance level, I use *GAAP_ETR* and *CASH_ETR*. I use contemporaneous measure of ETR rather than 3-year average ETR, since I am interested in the yearly change in tax audit risk as well as cross-sectional variations[16]. Rather than using tax sheltering measures, I choose ETR to capture the overall tax compliance level. The variable of interest is *Tax_audit*, which proxies for the firm's tax audit risk. *Tax_audit* is a continuous measure that ranges from 0 to 1. Rather

[16] For one-year ETR, a mismatch between the numerator and denominator occurs. This can occur when taxes are paid or tax expenses are reported based on different period earnings. I find that the results are qualitatively similar when using 3-year ETR measures.

than capturing whether the firm is under tax audit or not, this measure incorporates firms' perceived risk of tax audit.

Variables that are known to be associated with a firm's tax avoiding activities are controlled (Dyreng et al. 2008; Hoopes et al 2012). Firm size can be either positively or negatively associated with tax avoidance. In the one hand, firm size proxies for the political cost and that larger firms comply more with taxes since they are under stronger scrutiny by the government or market (Zimmerman 1983). On the other hand, sizable firms have enough resources to engage in tax avoiding activities so that there is a positive association between the two (Siegfried 1974; Porcano 1986). To control for these effects, I include *SIZE*. I also include Market to book ratio (*MB*) and return-on-assets (*ROA*) to capture the effect of growth and profitability on tax avoidance. Leverage (*Lev*) and R&D expenditures (*RnD*) are expected to be positively associated with tax avoidance, since interest and R&D expenses reduce taxable income and that firms cab further decrease their tax expenses from R&D credits. Firms have less tax burden when they have tax loss carryforwards (*NOL*) or report foreign income (*For_d*). Sales growth (*Sgrowth*) is also included to account for the volatile operation that can increase the tax burden.

3.4. Descriptive Statistics

Panel A of Table 2 presents the descriptive statistics of variables used in the regressions. The variable of interest, *Tax_audit*, has a mean value of 0.33, implying that firms, in average, discuss tax audit about 33% of tax footnotes[17].

[17] The value is comparable to the IRS audit rate measure from TRAC that prior literature relies on. For example, Hoopes et al. (2012) report the mean value of audit probability as 29%. More recent research by Bauer et al. (2020) presents the mean audit rate as 30.51%. The variable from TRAC shows the likelihood the

Tax_audit does not measure whether the firm is under the audit in year t. Rather, firms can discuss about tax audit if they perceive it as a threat. Thus, the minimum value of *Tax_audit* is non-zero and the standard deviation is higher than the IRS audit probability. The IRS audit probability shows the mean value of 0.19 when using 8-group of size (*TRAC_audit*) and 0.24 when using 12-group of size (*TRAC_audit_detail*), which are comparable to the studies that use this measure. The mean value is higher when segmenting the large firm group in detail, since large firms are more likely to be audited by the IRS. In addition, the standard deviation of TRAC_audit_detail is 0.18, which is much higher than that of TRAC_audit (0.08), implying that the IRS audit probability escalates for some large firms and varies significantly among them. Panel B of Table 2 reports the pairwise Pearson correlations of variables used in the tests. The correlation between *Tax_audit* and ETR is all positive and significant (0.12 for *CASH_ETR* and 0.13 for *GAAP_ETR*), showing that firms pay taxes more and report higher tax expenses when they perceive heightened tax audit risks. For the correlation between *Tax_audit* and the IRS audit probability, a detailed explanation is provided in the validation test part.

<Insert Table 2 here>

4. Empirical Results

4.1. The effect of tax audit risk on tax compliance

firm is under tax audit by the IRS. The variable I use in this research and the TRAC variable capture different aspect of firm's audit risk. But, since two proxies are all related to firm's tax audit risk, showing that the variable structures are comparable can be meaningful.

I test the effect of firms' tax audit risk on their tax compliance level. Results from estimating Eq. (1) are presented in Table 3. In column (1), I estimate tax compliance as GAAP effective tax rate. The coefficient on *Tax_audit* is positive and significant at 1% level (0.073***; t=6.07). In column (2), *CASH ETR*[18] is used as the dependent variable. The coefficient on *Tax_audit* is also positive and statistically significant at 1% level (0.088***; t=6.10) The results in columns (1) and (2) show that firms under heightened risk of tax audit not only report more tax expenses, but they actually pay more taxes. *GAAP ETR* incorporates permanent tax planning and is not affected by tax deferrals. On the other hand, firms can change the level of *CASH ETR* by deferring earnings (or accelerating expenses) for tax purposes and thus *CASH ETR* incorporates both temporary and permanent tax planning. Both temporary and permanent tax plannings are restrained when the firm is subject to high tax audit risks.

Control variables' coefficients are largely consistent with prior literature. Size reports significantly negative coefficients in both columns (Coef. = -0.005*** (t=-3.09) in Col. (1); -0.003* (t=-1.80) in Col. (2)), showing that larger firms have more resources to engage in tax avoidance (Siegfried 1974; Porcano 1986). Coefficient on Lev is negative in both columns as predicted, but it is statistically significant only for *GAAP ETR* result (Coef. = -0.035** (t=-2.14) in Col. (1)). *MB* does not show any statistical significance and *ROA* is found to be negatively

[18] I use the one-year ETR measure in estimating Eq.(1). However, ETR entails innate volatility in it, especially for *CASH ETR*. This occurs when taxes are paid in a different period when they are expensed. Thus, I also use 3-year average ETR from year *t-2* to year *t* following Dyreng et al.(2008). I find the results are qualitatively similar.

associated with ETR. As predicted, coefficients of *RnD*, *NOL*, and *Sgrowth* are all negative and statistically significant in both columns.

In sum, table 3 demonstrates that firms comply with tax laws more when they explain tax audit information more in their footnotes in 10-K filings. Firms are less likely to engage in both temporary and permanent tax planning when they are subject to higher risk of tax audit. Thus, hypothesis is rejected.

<Insert Table 3 here>

4.2. Validation Test

I measure *Tax_audit* as the extent to which firms discuss tax audit topics in their tax footnotes of 10-K filings. The intention is to capture how much risks firms perceive from tax audit, assuming that firms express their risk in the textual disclosure. However, managers may use the text disclosure as a rhetoric that justifies their uncertain tax status. I validate the measure by showing the correlation with the IRS audit probability that previous literature uses. Although the two measures capture different aspects of tax audit, the IRS audit probability can be a reference in that it is the *ex post* tax audit outcome. Panel A of Table 4 shows that *Tax_audit* is positively correlated with the IRS tax audit rate. In Panel B, I regress firms' tax compliance on both *Tax_audit* and IRS audit probability from TRAC. Columns (1) and (2) use the IRS audit probability that use 8-group size categorization in comparison with Hoopes et al. (2012). As in Hoopes et al. (2012), the coefficients of the IRS audit probability (*TRAC_audit*) are significant and positive. Moreover, *Tax_audit* also presents significant and positive coefficients, suggesting that firms' tax audit discussion in 10-K filings provides incremental

information. When using detailed TRAC measure that relies on 12-group size categorization in Columns (3) and (4), only *Tax_audit* shows significantly positive coefficients. This implies that the TRAC measure is insufficient in capturing the tax audit risk among large firms.

<Insert Table 4 here>

4.3. Cross-sectional tests

4.3.1. The role of institutional investors

One of the mechanisms that tax audit affects the firm is through governance. As Desai et al. (2007) maintain, managerial diversion decreases when the firm is audited by tax authorities, resulting in better governance. If tax authorities act as an external monitoring mechanism toward firms, the effect would be attenuated when there already exists another external monitor. I test whether the deterrent effect from Table 3 is less pronounced when there is high level of institutional investors. I conduct subsample tests estimating Eq. (1) with low and high groups of institutional ownership.

In Table 5, I examine the deterrent effect of tax audit by the level of institutional ownership. Results show that the coefficient on *Tax_audit* is larger in low group. The difference is statistically significant for *CASH ETR* result (Diff. = -0.079***; (p=0.01)) in columns (3) and (4). For *GAAP ETR* results (columns (1) and (2)), the difference is not significant but is still negative. The results present that firms lack of sufficient external monitors are more likely to be compliant with tax codes when they are under higher risk of tax audit. Alternative explanation is that institutional ownership favors tax savings from noncompliance and prevent the

firm from being more compliant even when the firm is under stronger risk of tax audit (Khan et al. 2017). If this is the case, the deterrent effect of tax audit affects the firms with low institutional ownership. Overall, Table 5 suggests that the monitoring role of tax audit incurs deterrent effect and the effect is more pronounced when there lacks alternative external monitor.

<Insert Table 5 here>

4.3.2. The role of foreign operations

While domestic firms in the U.S. are audited mainly by Internal Revenue Service (IRS), multinational firms (MNCs) are subject to audits from both the IRS and foreign tax authorities. If MNCs find tax audits from foreign jurisdiction to be significant, they can include such information in their tax footnotes as well, resulting in high value of *Tax_audit* measure. However, it is ex ante not clear whether MNCs would behave in a manner that domestic firms do.

In Table 6, I conduct sub-sample analyses with groups of domestic firms and MNCs. It shows that tax audit deters noncompliance for both domestic firms and MNCs. However, the magnitudes of *Tax_audit* of domestic firms (Coef. = 0.101*** (t=6.17) in Column (1); Coef. = 0.114*** (t=6.23) in Column (3)) are almost double than those of MNCs firms (Coef. = 0.040** (t=2.24) in Column (2); Coef. = 0.050** (t=2.28) in Column (4)). MNCs engage in complex transactions, such as profit shifting across jurisdictions or using transfer pricing, that makes auditing the MNCs much more difficult compared to domestic firms (Bustos et al. 2019). MNCs may not be compliant with tax laws because they know their activities are difficult to be detected.

<Insert Table 6 here>

In sum, Table 5 and Table 6 suggest that there exists heterogeneity in terms of the deterrent effect of tax audit. I expect the results deliver policy implications. Tax noncompliance is problematic because it is related to the tax revenues that government generates. The findings that the deterrent effect is not effective for firms with high level of institutions and multinational firms as much as it is for the other firms presents that tax authorities have to apply different strategies in conducting audits of certain firms. Notably, the IRS has experienced budget cuts in the past years, which raises the need to conduct more efficient tax audits.

4.4. Robustness checks

In estimating the effect of tax audit, it is presumed that my tax audit proxy fully measures the extent of firms perceive their own tax audit risk. Since the measure is generated using the footnote contents in 10-K filings, it is possible that firms mechanically include tax audit contents even when they do not perceive the risk of tax audit to be high. In this case, tax audit measure does not vary among time periods in each firm but would vary cross-sectionally only. If cross-sectional variation of tax footnote disclosures also relates to the difference in tax compliance level among firms, then the results are driven by the disclosure policy and not by the tax audit risk. To mitigate the concern, I conduct empirical tests with firm fixed effect. The results are qualitatively similar with the main analyses.

<Insert Table 7 here>

5. Conclusion

I examine the effect of tax audit risk on corporate tax compliance behavior. Utilizing a topic modeling methodology (LDA), I develop a firm-level measure of tax audit risk based on the textual information from firms'10-K filings. The tax audit risk presents the extent of tax audit related topics discussed in the tax footnotes. I validate the measure by showing the correlation with the IRS audit probability from TRAC database and provide robust results when including both the tax audit topic and the IRS audit probability. The empirical results present that firms comply more with tax laws when they are under heightened risk of tax audit. In the cross-sectional tests, I suggest factors associated with the deterrent effect of tax audit – institutional ownership and foreign operation. The motoring role of tax audit is attenuated when there are alternative external monitors such as institutional ownership and when the firm operates in foreign jurisdictions.

This study broadens the knowledge of tax enforcement. The results confirm the deterrent effect of corporate tax audit. Moreover, the tax audit topic measures the perceived tax audit risk, especially for large firms. Considering that the tax evasion of large firms is highly influential to the market and the government, this study contributes to the recent literature that begins to focus on small businesses regarding the effect of tax enforcement. In addition, under the constraint resources of tax authorities, understanding the factors associated with tax enforcement can be of help in securing effective and efficient tax audits.

REFERENCES

Allingham, M. G., and A. Sandmo. 1972. Income tax evasion: A theoretical analysis. *Journal of Public Economics* 1 (3-4): 323-338.

Alm, J., G. H. McClelland, W. D. Schulze. 1992. Why do people pay taxes? *Journal of Public Economics* 48 (1): 21-38.

Bauer, A. M., J. Fang, and J. Pittman. 2020. How Aggressive Tax Planning Facilitates the Diversion of Corporate Resources: Evidence from Path Analysis. *Contemporary Accounting Research* 37 (3): 1882-1913.

Blankespoor, E. 2019. The impact of information processing costs on firm disclosure choice: Evidence from the XBRL mandate. *Journal of Accounting Research* 57 (4): 919-967.

Blei, D. M., A. Y. Ng, M. I. Jordan. 2003. Latent Dirichlet allocation. *Journal of Machine Learning Research* 3 (Jan): 993–1022.

Bloom, R. 2008. Discussion of "Annual report readability, current earnings, and earnings persistence". *Journal of Accounting and Economics* 45(2-3): 248-252.

Bodnaruk, A., T. Loughran and B. McDonald. 2015. Using 10-K Text to Gauge Financial Constraints. *Journal of Financial and Quantitative Analysis* 50 (4): 623-646.

Bushee, N., I. D. Gow, and D. Talyer. 2018. Linguistic Complexity in Firm Disclosures: Obfuscation or Information? *Journal of Accounting Research* 56 (1): 85-121.

Bustos, S., D. Pomeranz, J. Vila-Belda, and G. Zucman. 2019. Challenges of Monitoring Tax Compliance by Multinational Firms: Evidence from Chile. *AEA Papers and Proceedings* 109: 500-505.

Cazier, R. A., Pfeiffer, R. J. 2015. Say again? Assessing redundancy in 10-K disclosures. Working Paper.

Clausing, K. A. 2009. Multinational Firm Tax Avoidance and Tax Policy. *National Tax Journal* 62(4): 703-725.

Desai, M. A., A. Dyck, and L. Zingales. 2007. Theft and taxes. *Journal of Financial Economics* 84 (3): 591–623.

Dyer, T., M. Lang, L. Stice-Lawrence. 2017. The evolution of 10-K textual disclosure: Evidence from Latent Dirichlet Allocation. *Journal of Accounting and Economics* 64 (2-3): 221-245.

Dyreng, S., M. Hanlon, and E. L. Maydew. 2008. Long-Run Corporate Tax Avoidance. *The Accounting Review* 83 (1): 61-82.

El Ghoul, S., O. Guedhami, and J. Pittman. 2011. The role of IRS monitoring in equity pricing in public firms. *Contemporary Accounting Research* 28 (2): 643-674.

Gallemore, J., and M. Jacob. 2020. Corporate Tax Enforcement Externalities and

the Banking Sector. *Journal of Accounting Research* 58 (5): 1117-1159.
Gleason, C. A., and L. F. Mills. 2008. Evidence of differing market responses to beating analysts' targets through tax expense decreases. *Review of Accounting Studies* 13: 295-318.
Graham, J. R., J. S. Raedy, and D. A. Shackelford. 2012. Research in accounting for income taxes. *Journal of Accounting and Economics* 53 (1-2): 412-434.
Graham, J. R., M. Hanlon, T. Shevlin, and N. Shroff. 2014. Incentives for tax planning and avoidance: Evidence from the field. *The Accounting Review* 89 (3): 991-1023.
Graham, J. R., M. Hanlon, T. Shevlin, and N. Shroff. 2017. Tax Rates and Corporate Decision-making. *The Review of Financial Studies* 30(9): 3128–3175.
Guedhami, O. and J. Pittman. 2008. The importance of IRS monitoring to debt pricing in private firms. *Journal of Financial Economics* 90 (1): 38-58.
Hanlon, M., J. L. Hoopes, and N. Shroff. 2014. The effect of tax authority monitoring and enforcement on financial reporting quality. *Journal of the American Taxation Association* 36 (2): 137-170.
Hoopes, J. L., D. Mescall, and J. A. Pittman. 2012. Do IRS audits deter corporate tax avoidance? *The Accounting Review* 87 (5): 1603-1639.
Huang A. H., R. Lehavy, A. Y. Zang, and T. Zheng. 2018. Analyst Information Discovery and Interpretation Roles: A Topic Modeling Approach. *Management Science* 64 (6): 2833-2855.
Khan, M., S. Srinivasan, and L. Tan. 2017. Institutional Ownership and Corporate Tax Avoidance: New Evidence. *The Accounting Review* 92 (2): 101-122.
Khurana, I., and W. Moser. 2013. Institutional shareholders' investment horizons and tax avoidance. *Journal of the American Taxation Association* 35 (1): 111–134.
Klassen, K. and Laplante S. 2012. Are U.S. Multinational Corporations Becoming More Aggressive Income Shifters? *Journal of Accounting Research* 50 (5): 1245-1285.
Lev, B., and D. Nissim. 2004. Taxable income, future earnings, and equity values. *The Accounting Review* 79 (4): 1039-1074.
Loughran, T. and B. McDonald. 2014. Measuring readability in financial disclosures. *Journal of Finance* 69 (4): 1643-1671.
Mironov, M. 2013. Taxes, theft, and firm performance. *The Journal of Finance* 68 (4): 1441-1472.
Porcano, T. 1986. Corporate tax rates: Progressive, proportional, or regressive. *The Journal of the American Taxation Association* 7: 17–31.
Siegfried, J. J., 1974. Effective average U.S. corporation income tax rates. *National Tax Journal* 27 (2): 245–259.
Skinner, D. 1994. Why firms voluntarily disclose bad news. *Journal of Accounting Research*, 32 (1), 38–60.

Slemrod, J. 2019. Tax compliance and enforcement. *Journal of Economic Literature* 57 (4): 904-954.
Slemrod, J. and S. Yitzhaki. 2002. Tax avoidance, evasion, and administration. Handbook of Public Economics 3: 1423-1470.
Slemrod, J. and C. Gillitzer. 2014. Insights from a Tax-systems Perspective. *CESifo Economic Studies* 60 (1): 1-31.
Slemrod, J., M. Blumenthal, and C. Christian. 2001. Taxpayer response to an increased probability of audit: evidence from a controlled experiment in Minnesota. *Journal of Public Economy* 79 (3): 455-483.
Yost, B., and S. Shu 2020. The Effect of Tax Enforcement on Managers' Self-Dealing: Evidence from Insider Trading. Working Paper.
Zimmerman, J. 1983. Taxes and firm size. *Journal of Accounting and Economics* (5): 119–149.

TABLE 1
Sample selection

Sample (firm-year)	Observations
U.S. firms-years from Compustat (2011-2019) with non-missing asset excluding financial firms and utilities	44,698
Less: Missing XBRL data	(20,938)
Less: Missing financial variables	(5,900)
Less: Negative and zero pre-tax income	(7,109)
Final sample (firm-year)	**10,751**

This table shows the sample selection procedure. Initial sample consists of the U.S. firm-years with non-missing assets from 2011 to 2019, excluding financial firms and utilities. After removing observations with missing XBRL data that is necessary to capture my variable of interest (*Tax_audit*), observations with missing financial variables and those with negative pre-tax income, 10,751 observations comprise the final sample. Firm-years that report negative pre-tax income or zero pre-tax income are deleted to calculate the ETR.

TABLE 2
Summary table of variables

Panel A: Descriptive statistics

Variable	N	Mean	SD	P25	Median	P75
Tax_audit	10,751	0.33	0.2	0.18	0.3	0.45
TRAC_audit	10,751	0.24	0.18	0.12	0.2	0.3
TRAC_audit_de	10,751	0.19	0.08	0.14	0.19	0.27
CASH_ETR	10,751	0.24	0.22	0.07	0.21	0.33
GAAP_ETR	10,751	0.28	0.2	0.15	0.29	0.37
Size	10,751	6.93	2.06	5.63	7.01	8.32
MB	10,751	1.35	0.63	0.91	1.24	1.65
Lev	10,751	0.19	0.18	0.01	0.17	0.32
ROA	10,751	0.1	0.07	0.06	0.09	0.13
RnD	10,751	0.03	0.05	0	0	0.03
NOL	10,751	0.18	0.52	0	0.02	0.11
For_d	10,751	0.53	0.5	0	1	1
Sgrowth	10,751	0.13	0.29	0.01	0.07	0.18

This table presents summary statistics for the test samples. Variable definitions are in Appendix 3, and all continuous variables are winsorized at the 1st and 99th percentiles.

Panel B. Pearson correlation

		[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]	[11]	[12]	[13]
[1]	*Tax_audit*	1												
[2]	*TRAC_audit*	**0.22**	1											
[3]	*TRAC_audit_de*	**0.20**	**0.57**	1										
[4]	*CASH_ETR*	**0.12**	**0.07**	**0.02**	1									
[5]	*GAAP_ETR*	**0.13**	**0.10**	**0.02**	**0.37**	1								
[6]	*Size*	**0.15**	**0.48**	**0.77**	**0.05**	**0.02**	1							
[7]	*MB*	0.02	0.01	**0.08**	**-0.03**	**-0.02**	**0.13**	1						
[8]	*Lev*	0.01	**0.15**	**0.22**	**-0.02**	**-0.02**	**0.43**	**0.19**	1					
[9]	*ROA*	**0.11**	**0.04**	0.01	**-0.02**	0.01	-0.01	**0.43**	**-0.09**	1				
[10]	*RnD*	**-0.17**	**-0.09**	**-0.10**	**-0.06**	**-0.09**	**-0.17**	**0.19**	**-0.28**	0.02	1			
[11]	*NOL*	**-0.20**	**-0.24**	**-0.18**	**-0.17**	**-0.15**	**-0.30**	**0.03**	**-0.11**	**-0.09**	**0.27**	1		
[12]	*For_d*	**-0.07**	**0.20**	**0.24**	**0.13**	0.01	**0.36**	**0.13**	**0.06**	**0.04**	**0.10**	**-0.11**	1	
[13]	*Sgrowth*	**-0.13**	**-0.06**	**-0.08**	**-0.17**	**-0.11**	**-0.10**	**0.10**	-0.02	**0.12**	**0.11**	**0.09**	**-0.14**	1

This table reports the Pearson correlations matrix among the variables used in the empirical analysis. Bold figures indicate statistical significance at $p < 0.05$. Variable definitions are in Appendix 3, and all continuous variables are winsorized at the 1st and 99th percentiles.

Table 3
The effect of tax audit risk on tax compliance

Dependent Variable	(1) *GAAP ETR*	(2) *CASH ETR*
Tax_audit	0.073***	0.088***
	(6.07)	(6.10)
Size	-0.005***	-0.003*
	(-3.09)	(-1.80)
MB	0.002	-0.007
	(0.41)	(-1.23)
Lev	-0.035**	-0.016
	(-2.14)	(-0.75)
ROA	-0.043	-0.148***
	(-1.14)	(-3.39)
RnD	-0.172**	-0.194**
	(-2.37)	(-2.58)
NOL	-0.055***	-0.061***
	(-10.48)	(-11.30)
For_d	0.006	0.037***
	(1.00)	(5.38)
Sgrowth	-0.050***	-0.086***
	(-6.22)	(-10.60)
Constant	0.266***	0.203***
	(6.60)	(4.36)
Observations	10,751	10,751
Adjusted R^2	0.074	0.101
Fixed Effects	Year Ind	Year Ind
Cluster	Firm	Firm

This table contains estimates of OLS regression following Eq. (1) on the effect of tax audit on corporate tax compliance.

$$Tax\_compliance_{i,t} = \beta_0 + \beta_1 Tax\_audit_{i,t} + \gamma_k \sum_k Control_{k,i,t} + \epsilon_{i,t} \quad (1)$$

Dependent variable is the level of tax compliances, captured as *GAAP_ETR* in Column (1) and *CASH_ETR* in Column (2). The variable of interest is *Tax_audit*, which captures the firm-level perceived tax audit risk. *Tax_audit* is measured as the extent to which firms discuss tax audit in their tax footnotes of 10-K filings. Standard errors are clustered by firm. ***, **, * indicate statistical significance at 1%, 5%, and 10%, respectively Variable definitions are in Appendix 3, and all continuous variables are winsorized at the 1st and 99th percentiles.

Table 4
Validation test

Panel A. Pearson correlation

	Tax_audit	*TRAC_audit*	*TRAC_audit_detailed*
Tax_audit	1		
TRAC_audit	0.22***	1	
TRAC_audit_detailed	0.20***	0.57***	1

Panel B. The effect of tax audit topic and TRAC measures on tax compliance

	(1)	(2)	(3)	(4)
	Using TRAC measure		Using *detailed* TRAC measure	
Dependent Variable	*GAAP ETR*	*CASH ETR*	*GAAP ETR*	*CASH ETR*
Tax_audit	0.074***	0.089***	0.074***	0.088***
	(6.20)	(6.16)	(6.11)	(6.11)
TRAC_audit	0.358***	0.268***	-0.040	-0.025
	(6.02)	(3.79)	(-1.51)	(-0.85)
Size	-0.011***	-0.008***	-0.002	-0.001
	(-6.09)	(-3.69)	(-0.76)	(-0.49)
MB	0.002	-0.007	0.002	-0.007
	(0.48)	(-1.18)	(0.48)	(-1.18)
Lev	-0.039**	-0.019	-0.039**	-0.019
	(-2.40)	(-0.90)	(-2.38)	(-0.88)
ROA	-0.044	-0.149***	-0.045	-0.149***
	(-1.17)	(-3.41)	(-1.19)	(-3.41)
RnD	-0.181**	-0.201***	-0.174**	-0.196***
	(-2.45)	(-2.66)	(-2.38)	(-2.59)
NOL	-0.050***	-0.057***	-0.055***	-0.060***
	(-9.36)	(-10.37)	(-10.25)	(-11.07)
For_d	0.004	0.036***	0.005	0.037***
	(0.67)	(5.15)	(0.88)	(5.27)
Sgrowth	-0.051***	-0.087***	-0.050***	-0.086***
	(-6.32)	(-10.64)	(-6.24)	(-10.62)
Constant	0.226***	0.174***	0.259***	0.199***
	(5.48)	(3.61)	(6.37)	(4.23)
Observations	10,751	10,751	10,751	10,751
Adjusted R^2	0.078	0.102	0.074	0.101
Fixed Effects	Year Ind	Year Ind	Year Ind	Year Ind
Cluster	Firm	Firm	Firm	Firm

This table shows the validation test of the variable of interest, *Tax_audit*. Panel A presents the Pearson correlation with the TRAC data. TRAC provides the IRS audit probability on firm size group and calendar year level. *TRAC_audit* is constructed based on 8-group category of firm size and *TRAC_audit_detailed* is constructed based on 12-group category of firm size. In Panel B, I regress tax compliance (ETR) on both *Tax_audit* and the IRS audit probability (*TRAC_audit, TRAC_audit_detailed*) from TRAC following the equation below.

$$Tax\_compliance_{i,t} = \beta_0 + \beta_1 Tax\_audit_{i,t} + \beta_2 TRAC\_audit_{i,t} + \gamma_k \sum_k Control_{k,i,t} + \epsilon_{i,t}$$

Standard errors are clustered by firm. ***, **, * indicate statistical significance at 1%, 5%, and 10%, respectively
Variable definitions are in Appendix 3, and all continuous variables are winsorized at the 1st and 99th percentiles.

Table 5
The role of institutional ownership on tax compliance

	(1)	(2)	(3)	(4)
Dependent Variable	*GAAP ETR*		*CASH ETR*	
Institutional Own.	Low	High	Low	High
Tax_audit	0.083***	0.059***	0.112***	0.033
	(4.68)	(3.26)	(5.32)	(1.45)
Size	-0.007***	-0.012***	-0.008***	-0.004
	(-2.85)	(-4.14)	(-3.17)	(-1.26)
MB	-0.004	0.007	0.007	-0.025***
	(-0.49)	(0.88)	(0.72)	(-2.64)
Lev	-0.066**	-0.011	-0.068**	0.034
	(-2.47)	(-0.43)	(-2.09)	(1.03)
ROA	-0.052	0.034	-0.153**	-0.059
	(-0.86)	(0.51)	(-2.28)	(-0.72)
RnD	-0.140	-0.345***	-0.253**	-0.281**
	(-1.29)	(-3.03)	(-2.27)	(-2.29)
NOL	-0.050***	-0.035**	-0.060***	-0.054***
	(-6.52)	(-2.23)	(-7.53)	(-3.20)
For_d	0.018**	-0.032***	0.042***	0.004
	(1.97)	(-4.00)	(4.08)	(0.37)
Sgrowth	-0.046***	-0.059***	-0.091***	-0.100***
	(-3.69)	(-3.47)	(-8.89)	(-5.15)
Constant	0.312***	0.401***	0.260***	0.313***
	(6.68)	(12.51)	(4.47)	(7.82)
Diff. test on *Tax_audit*				
$\Delta\chi^2(1)$	-0.024		-0.079***	
(p-value)	(0.34)		(<0.01)	
Observations	4,204	4,201	4,204	4,201
Adjusted R^2	0.082	0.042	0.116	0.075
Fixed Effects	Year Ind	Year Ind	Year Ind	Year Ind
Cluster	Firm	Firm	Firm	Firm

This table contains estimates of OLS regression following Eq. (1) on the effect of tax audit on tax compliance after partitioning the sample into low and high groups of institutional ownership.

$$Tax\_compliance_{i,t} = \beta_0 + \beta_1 Tax\_audit_{i,t} + \gamma_k \sum_k Control_{k,i,t} + \epsilon_{i,t} \quad (1)$$

Dependent variable is the level of tax compliances, captured as *GAAP_ETR* in Column (1) and *CASH_ETR* in Column (2). The variable of interest is *Tax_audit*, which captures the firm-level perceived tax audit risk. *Tax_audit* is measured as the extent to which firms discuss tax audit in their tax footnotes of 10-K filings. Standard errors are clustered by firm. ***, **, * indicate statistical significance at 1%, 5%, and 10%, respectively Variable definitions are in Appendix 3, and all continuous variables are winsorized at the 1st and 99th percentiles.

Table 6
Differential effect on domestic firms and MNCs

	(1)	(2)	(3)	(4)
Dependent Variable	*GAAP ETR*		*CASH ETR*	
Domestic vs. MNC	Domestic	MNC	Domestic	MNC
Tax_audit	0.101***	0.040**	0.114***	0.050**
	(6.17)	(2.24)	(6.23)	(2.28)
Size	0.003	-0.012***	0.003	-0.009***
	(1.38)	(-6.54)	(1.02)	(-3.52)
MB	-0.005	0.017**	-0.011	0.007
	(-0.76)	(2.39)	(-1.52)	(0.87)
Lev	-0.031	-0.056**	-0.034	0.002
	(-1.30)	(-2.53)	(-1.15)	(0.06)
ROA	0.056	-0.200***	0.050	-0.508***
	(1.20)	(-3.15)	(1.02)	(-6.34)
RnD	-0.029	-0.385***	-0.089	-0.356***
	(-0.27)	(-4.32)	(-0.83)	(-3.44)
NOL	-0.058***	-0.039***	-0.060***	-0.052***
	(-9.61)	(-3.35)	(-10.02)	(-4.44)
Sgrowth	-0.041***	-0.060***	-0.073***	-0.104***
	(-4.40)	(-3.45)	(-8.57)	(-5.17)
Constant	0.209***	0.385***	0.147**	0.342***
	(3.80)	(9.34)	(2.47)	(4.63)
Diff. test on *Tax_audit*				
$\Delta\chi^2(1)$	-0.061**		-0.064**	
(p-value)	(0.01)		(0.02)	
Observations	5,084	5,667	5,084	5,667
Adjusted R^2	0.133	0.047	0.145	0.064
Fixed Effects	Year Ind	Year Ind	Year Ind	Year Ind
Cluster	Firm	Firm	Firm	Firm

This table contains estimates of OLS regression on the effect of tax audit on tax compliance after partitioning the sample into domestic firms and MNCs. Domestic firms are defined as firms not reporting foreign income and MNCs are firms that report foreign income.

$$Tax\_compliance_{i,t} = \beta_0 + \beta_1 Tax\_audit_{i,t} + \gamma_k \sum_k Control_{k,i,t} + \epsilon_{i,t} \quad (1)$$

Dependent variable is the level of tax compliances, captured as *GAAP_ETR* in Column (1) and *CASH_ETR* in Column (2). The variable of interest is *Tax_audit*, which captures the firm-level perceived tax audit risk. *Tax_audit* is measured as the extent to which firms discuss tax audit in their tax footnotes of 10-K filings. Standard errors are clustered by firm. ***, **, * indicate statistical significance at 1%, 5%, and 10%, respectively Variable definitions are in Appendix 3, and all continuous variables are winsorized at the 1st and 99th percentiles.

Table 7
Robustness Test

Dependent Variable	(1) *GAAP ETR*	(2) *CASH ETR*
Tax_audit	0.088***	0.069***
	(4.16)	(3.04)
Size	-0.021*	0.027***
	(-1.90)	(2.64)
MB	0.003	-0.013
	(0.29)	(-1.46)
Lev	0.042	0.083**
	(1.24)	(2.43)
ROA	-0.184***	-0.470***
	(-2.90)	(-6.50)
RnD	0.235	0.107
	(1.12)	(0.48)
NOL	-0.083***	-0.030***
	(-5.51)	(-2.58)
For_d	-0.054***	-0.058***
	(-4.81)	(-4.81)
Sgrowth	-0.022**	-0.064***
	(-1.98)	(-5.57)
Constant	0.441***	0.106
	(5.61)	(1.48)
Observations	10,751	10,751
Adjusted R^2	0.032	0.044
Fixed Effects	Year Firm	Year Firm
Cluster	Firm	Firm

This table contains estimates of OLS regression following Eq. (1) on the effect of tax audit on corporate tax compliance.

$$Tax\_compliance_{i,t} = \beta_0 + \beta_1 Tax\_audit_{i,t} + \gamma_k \sum_k Control_{k,i,t} + \epsilon_{i,t} \quad (1)$$

Instead of including industry fixed effect, firm fixed effect is included. Standard errors are clustered by firm. ***, **, * indicate statistical significance at 1%, 5%, and 10%, respectively Variable definitions are in Appendix 3, and all continuous variables are winsorized at the 1st and 99th percentiles.

Appendix 1
Highest-Probability Words in The Audit-related Topics

Topic Label	Words
Tax audit1	IRS, filed, received, audit, assessment, internal_revenue_service_irs
Tax audit2	examinations, authorities, penalties_related, components
Tax audit3	audits, audit, matters, adjustments, could, reasonably_possible

Based on the textual data of U.S. firms' 10-K filings from 2011 to 2019, I extract the topics consisting each firm's tax footnote using a deep learning-based topic modeling methodology, Latent Dirichlet allocation (Blei et al. 2003). Each topic accounts for the extent to which a specific topic is discussed in a tax footnote of firm's 10-K filing, ranging from 0 to 1. I extract 20 topics and 3 topics among them are found to be related to the tax audit. Extracted topics are not significantly different when choosing 10 topics, instead of 20. For the variable used in the analyses, *Tax_audit* is generated by aggregating [*Tax audit1*, *Tax audit2*, *Tax audit 3*].

Appendix 2
Examples of Tax Footnotes with High and Low Scores of Tax Audit

Panel A. Tax Footnotes that record a *high* score of tax audit

THE BOSTON BEER COMPANY, INC. / CIK = 949870 / datadate = 2013-12-28
Tax_audit = 0.962
I. Income Taxes • Significant components of the provision (benefit) for income taxes are as follows: • The Company's reconciliations to statutory rates are as follows: • Significant components of the Company's deferred tax assets and liabilities are as follows at: • The Company's practice is to classify interest and penalties related to income tax matters in income tax expense. Interest and penalties included in the provision for income taxes amounted to $0.4 million, $0.1 million and $0.4 million for fiscal years 2013, 2012 and 2011, respectively. Accrued interest and penalties amounted to $0.3 million and $0.7 million at December 28, 2013 and December 29, 2012, respectively. • A reconciliation of the beginning and ending amount of unrecognized tax benefits is as follows: • Included in the balance of unrecognized tax benefits at December 28, 2013 and December 29, 2012 are potential net benefits of $0.5 million and $1.1 million, respectively, that would favorably impact the effective tax rate if recognized. Unrecognized tax benefits are included in accrued expenses in the accompanying consolidated balance sheets and adjusted in the period in which new information about a tax position becomes available or the final outcome differs from the amount recorded. • **In May 2013, the Internal Revenue Service (the "IRS") commenced an examination of the Company's 2010 and 2011 consolidated corporate income tax returns. The examination was still in process as of December 28, 2013.** • **The Company's state income tax returns remain subject to examination for three or four years depending on the state's statute of limitations. The Company is being audited by one state as of December 28, 2013. In addition, the Company is generally obligated to report changes in taxable income arising from federal income tax audits.** • **It is reasonably possible that the Company's unrecognized tax benefits may increase or decrease in 2014 if there is a completion of certain income tax audits**; however, the Company cannot estimate the range of such possible changes. The Company does not expect that any potential changes would have a material impact on the Company's financial position, results of operations or cash flows.

Panel B. Tax Footnotes that record a *low* score of tax audit

SUNESIS PHARMACEUTICALS INC / CIK = 1061027 / datadate = 2011-12-31
Tax_audit = 0
13. Income Taxes No provision for income taxes was recorded in the periods presented due to tax losses incurred in each period. The income tax provision differs from the amount computed by applying the statutory income tax rate of 34% to pre-tax loss as follows (in thousands): Deferred income taxes reflect the net tax effects of loss and credit carry-forwards and temporary differences between the carrying amounts of assets and liabilities for financial reporting purposes and the amounts used for income tax purposes. Significant components of the Company's deferred tax assets for federal and state income taxes are as follows (in thousands): Realization of the deferred tax assets is dependent upon future taxable income, if any, the amount and timing of which are uncertain. Accordingly, the net deferred tax assets have been fully offset by a valuation allowance. The net valuation allowance increased by approximately $10.3 million, $8.9 million and $7.7 million during the years ended December 31, 2011, 2010 and 2009, respectively. As of December 31, 2011, the Company had federal net operating loss carry-forwards of $278.2 million and federal research and development tax credit carry-forwards of $6.5 million. If not utilized, the federal net operating loss and tax credit carry-forwards will expire at various dates beginning in 2018. As of December 31, 2011, the Company had state net operating loss carry-forwards of $169.3 million, which begin to expire in 2012, and state research and development tax credit carry-forwards of $5.9 million, which do not expire. Utilization of these net operating loss and tax credits carry-forwards may be subject to a substantial annual limitation due to the ownership change rules under Section 382 of the Internal Revenue Code of 1986, as amended (the “Code"). The limitations are applicable if an “ownership change”

Appendix 3
Variable Definitions

Variable	Definition
Tax_audit	Perceived tax audit risk, measured as the extent of which firms discuss tax-audit-related information in tax footnotes of 10-K filings.
TRAC_audit	The IRS audit probability rate from TRAC constructed by 8-size group and calendar year level.
TRAC_audit_detailed	The IRS audit probability rate from TRAC constructed by 12-size group and calendar year level.
Size	Natural logarithm of total assets.
MB	The ratio of market value of equity to the book value of equity.
Lev	The ratio of total long-term debt to total assets.
ROA	The ratio of operating income before depreciation to total assets of beginning of the fiscal year.
RnD	The ratio of R&D expenses scaled by total assets. Missing values are adjusted as zero.
NOL	The ratio of unused tax loss carryforwards scaled by total assets.
For_d	An indicator variable that equals one if the firm reports foreign income, and zero otherwise.
Sgrowth	The ratio of the change in sales to the sales of last year.
Inst_own	The ratio of institutional ownership to the market capitalization (in percentage).

Essay 2

The Effect of Financial Reporting Readability on Foreign Investors' Information Acquisition Activity

1. Introduction

Foreign portfolio investment is a significant source of liquidity for the U.S. capital markets. According to the U.S. Department of the Treasury, foreign holdings of U.S. securities (i.e., corporate debt, U.S. treasury, and equity) have steadily increased from $12.4 trillion in 2011 to $19.4 trillion in 2018.[1] When breaking down the foreign holdings by security type, equity holdings amount to $3.8~8.1 trillion which accounts for 21~27 percent of the U.S. market capitalization during 2011~2018. While foreign investors play a significant role in the U.S. equity market, it is well documented that foreign investors are at an information disadvantage compared to domestic investors (Allee, Anderson, and Crawley 2019, Dvořák 2005) and they have a tendency to prefer equity from their own geographical vicinity, known as home bias (Coval and Moskowitz 1999).

Prior research proposes several factors that may contribute to this enduring and stable phenomenon, including geographic distance from investee firms (Butler 2008), differences in accounting practices (Bae et al. 2008), culture (Ahern et al. 2015) and language (Grinblatt and Keloharju 2001; Lundholm et al. 2014; Lundholm et al. 2017). Given the importance of foreign investment in the U.S market, it is surprising that little attention is paid to the role of financial reporting readability in determining foreign investors' investment in empirical research. I fill this void in the literature.

This study examines the impact of financial reporting readability on foreign investors' information acquisition activities and their choice of equity holdings. Unlike the factors prior studies have identified to influence foreign investors'

[1] https://ticdata.treasury.gov/Publish/shla2018r.pdf

information disadvantage such as culture and language, financial reporting readability is mainly prerogative of firm management. Although the accounting standards and disclosure rules are intended to ensure the consistency of financial reporting across firms, firm management has considerable latitude in financial reporting, e.g., length and complexity of the reports. Consistent with this view, prior research suggests that there is a significant variation in the readability of annual reports, i.e., 10-K filings (Li 2008, Bonsall et al. 2017), and economic consequences of the readability are significant. For example, firms with high financial reporting readability exhibit more persistent earnings (Li 2008), greater investment efficiency (Biddle et al. 2009), higher firm value (Hwang and Kim 2017), or lower cost of debt (Bonsall and Miller 2017). These findings suggest that 10-K filings with high readability are more informative and useful.

I posit that financial reporting readability impacts foreign investors' information acquisition activity for U.S. stocks. Prior research indicates that language gap is a significant barrier to information acquisition and deters information transmission and processing (Brochet et al. 2016; Tenzer et al. 2014; Slangen 2011). Beatie (1977) defines 'reading' as "the ability to deduce the core meaning of a sentence without necessarily knowing the specific meaning of each word." According to Beatie (1977), understanding of written foreign language is not a word-by-word decoding process. Hence, even though foreign investors understand financial terminologies and have access to language aid such as computerized translation or dictionary, reading a complex English text can be challenging. This suggests that language gap will limit foreign investors' ability to extract and infer value-relevant information from financial reports. Therefore, while

low financial reporting readability increases information processing costs in general (Bonsall and Miller 2017), this tendency is likely disproportionately larger for foreign investors than domestic investors. Based on this argument, I anticipate that low financial reporting readability will have a greater adverse impact on foreign investors' information collection activity than that of domestic investors.

To measure information collection activities, I use investors' 10-K search activities. Investors access 10-K filings in the EDGAR database to collect company information (Loughran and McDonald 2017) and the Securities and Exchange Commission (SEC) makes the IP-access log files available to the public. I obtain from the log files of IP addresses that access the firm's 10-K filing within seven days subsequent to the 10-K filing release. I then identify the origins of the IPs based on the first three parts of the IP addresses[2] and count the number of accesses from the U.S. and foreign countries. I construct foreign access as the ratio of the number of accesses from foreign countries to the number of accesses from the U.S. multiplied by 100.[3] Following prior research, I measure readability with the inverse of the BOG index (Bonsall et al. 2017).

Using all firms with 10-K filings in the period 2002 to 2016, I find compelling evidence that foreign access is positively associated with readability of 10-K filings. This finding suggests that easy writing attracts foreign investors by

[2] The IP address has format xxx.xxx.xxx.xxx, where each x stands for a number. To protect the exact identity of the IP address, the SEC hides the last three digits of the IP address. I use the first three parts (xxx.xxx.xxx) to identify the location. The location is identified using the dataset provided by http://www.ip2location.com.

[3] I use the ratio rather than frequencies because my research question is about the *relative sensitivity* of foreign investors' information acquisition activities to financial reporting readability. It also helps me control for investors' general interest in the firms that are not specific to foreign investors.

facilitating their information collection activities. To explore the economic consequences of financial reporting readability, I next examine whether foreign investors' information collection activities lead to foreign investment in the subsequent periods. To the extent that high financial readability attracts foreign investors by alleviating foreign investors' information disadvantages over domestic investors, I expect that there exists a positive association between foreign investors' access to these firms' 10-K filings and foreign holdings of the firm shares. Consistent with my conjecture, I find a positive association between foreign access and foreign institutional holdings in the following year. The association is both statistically and economically significant – a one standard deviation increase in the foreign access is associated with 3.64 percent increase in the foreign institutional ownership in the following year. I also examine foreign investors' search behavior after accessing the focal firm's 10-K and find that they are more likely to search another firm's 10-K, presumably for supplementary information, when the focal firm's 10-K is less readable. This finding provides further evidence that financial reporting readability influences foreign investors' information search activities. Collectively, these results suggest that financial reporting readability helps lower foreign investors' information disadvantage and this in turn attracts foreign investors' equity investment.

I conduct additional analysis to gain further insight into the relation between financial reporting readability and foreign access. First, I explore the moderating role of language similarity by examining whether foreign investors' mother tongue influences the relation between financial reporting readability and their access to the reports. It is an empirical question how language difference influences the extent

to which financial reporting readability helps foreign investors' information collection activity. On the one hand, the readability may be more beneficial when the difference between the investor's mother tongue and English is larger, i.e., it substitutes foreign investors' English proficiency. On the other hand, it may complement foreign investors' familiarity with English when processing information in 10-K filings (Schulz 1981). [4] I find that the positive impact of financial reporting readability on foreign access is salient when investors are from countries with high language similarity (i.e., foreign investors are more familiar with English), while the effect is muted when investors are from countries with low language similarity with English. I find a similar result when I use language diversification (i.e., extent of multilingualism) as an alternative measure of foreign investors' familiarity with English. The findings suggest that foreign investors' benefit from easy financial report is restricted to those from countries with high language similarity with English – foreign investors with severe language gap are unable to benefit from the readability.

Second, I examine whether such association is pronounced in the subsamples of firms. Due to the information asymmetry between domestic and foreign investors, foreign investors likely select stocks with low information asymmetries such as safe and visible stocks (Kang and Stulz 1997). If more readable 10-K filings mitigate foreign investors' information asymmetry that limits their research on risky and less visible stocks, I expect to observe a significant

[4] She posits that "if the foreign language reader is presented with an insufficient number of familiar concepts and contexts … he reverts to a word by word decoding process which contributes neither to the development of global reading comprehension, enjoyment of the text, nor to the encouragement of continued reading in the foreign language (an ellipsis added)."

association between financial reporting readability and foreign access in all firms. However, if higher readability mitigates the information asymmetry of stocks that foreign investors likely have interest, I expect to observe a significant association between financial statement readability and foreign access only for firms with low risk and high visibility. My result supports the latter. It suggests that improved readability does not fully address foreign investors' information disadvantages over domestic investors for all firms. Instead, it appears to mitigate incrementally the information gap between foreign and domestic investors in the subsamples of firms that foreign investors are likely to target.

I perform a battery of sensitivity tests and find similar inferences. First, I use the FOG index instead of the BOG index for financial reporting readability and find consistent results. Second, I use the frequency of accesses instead of the relative ratio for foreign access. The result shows that financial reporting readability is associated with accesses from foreign countries but not with domestic accesses. Third, I relate financial reporting readability to accesses from English speaking foreign countries only and find a weaker association. This finding further suggests that foreign investors' preference for firms with high reporting readability is in part attributable to their language gap.

I make two important contributions to the literature. First, I expand the literature on readability (Li 2008; Biddle et al. 2009; Hwang and Kim 2017; Bonsall and Miller 2017) and provide another dimension of empirical evidence in support of the argument that readability of financial reporting reduces the cost to process information (Miller 2010; Lehavy et al. 2011; Lawrence 2013). I find that language similarity amplifies the relation between readability and foreign accesses. Second,

I add to the literature on the presence of home bias and the underlying cause of the bias (Coval and Moskowitz 1999; Grinblatt and Kolaharju 2001). Previous studies present evidence on a significant informational gap between foreign and domestic investors (Hau 2001; Chan et al. 2008). This information gap also imposes a limitation on foreign investments (Kang and Stulz 1997). While these findings improve the understanding of the underlying cause of home bias, the literature presents limited evidence about remedies for such gaps. My evidence highlights the benefit of marking financial statements more readable by documenting that readability of 10-Ks helps foreign investors' information acquisition activities, which in turn draws their investment to the U.S. stock market.

The remainder of this paper proceeds as follows. Section 2 reviews related literature and develops hypothesis. Section 3 discuss the data, research design, and descriptive statistics. Section 4 reports the main analysis results. Section 5 discuss additional analysis and robustness tests, and Section 6 concludes.

2. Related Literature and Hypothesis Development

My study is related to research that investigates financial reporting readability and research that assesses investor geography-based information advantages. I briefly review this literature and develop my hypothesis below.

2.1. Financial Reporting Readability

Complex and less readable disclosures are costlier to process and hence less informative. Li (2008) conjectures that managers with bad news have an incentive to obfuscate the information and thus are likely to issue a less readable annual report.

Consistent with this conjecture, he finds that firms of which annual reports (10-K filings) have poor readability tend to have lower current profitability and less persistent earnings in the future. Since Li (2008), researchers have documented other evidence in support of the management obfuscation hypothesis. For example, lower readability of financial report decreases investment efficiency (Biddle et al. 2009), decreases firm value (Hwang and Kim 2017), and increases cost of debt (Bonsall and Miller 2017). Besides managerial obfuscation, complex business models and reporting standards can also make annual reports less readable (Guay et al. 2016).

Regardless of the underlying sources, poor readability makes annual reports less informative and hence less useful in lowering information asymmetry between firm managers and market participants (e.g., investors and financial analysts). Consistent with low readability of financial reports being associated with opaque information environments, prior research finds that financial report complexity discourages stock trading of small investors who are in disadvantage of information processing relative to large investors (Miller 2010). Similarly, compared to sophisticated investors, individual investors invest less in firms with complex financial reports (Lawrence 2013). Readability of financial reports seems to influence analyst forecast quality as well as analyst following (Lehavy et al. 2011).

2.2. Information Disadvantages of Foreign Investors over Domestic Investors

Investors prefer domestic stocks to foreign stocks for their portfolios, a tendency referred to as “home bias”. An explanation for home bias is that domestic investors have information advantages over foreign investors (Chan et al. 2008).

Prior research suggests that such information disparity has a negative impact on foreign investors' trading performance. For example, domestic investors earn higher returns than do foreign investors in Indonesia (Dvořák 2005) and Germany (Hau 2001). In the Chinese market, discount on the share class open only to foreign investors (Class B) is larger than discount on the share class open only to domestic investors (Class A) (Chan et al. 2008). Home bias also limits foreign investors' portfolio selection. Foreign investors in the Japanese markets are biased toward bigger, successful, less levered, and lower risk stocks (Kang and Stulz 1997). Using mutual fund data from 26 developed or developing countries, Chan et al. (2005) show that mutual funds allocate significantly higher fraction of their investment to their domestic markets. These findings are consistent with the notion that foreign investors are at an informational disadvantage and endure higher costs for collecting and processing information about domestic firms.

The geography-based information disparity influences financial analysts as well. Domestic analysts provide more accurate and precise forecasts than foreign analysts (Bae et al. 2008), and U.S. analysts forecast less accurately for cross-listed firms especially when the language of the cross-listed firm is more difficult for English speakers (Cho et al. 2020). These results suggest that even analysts who are experts in analyzing information and predicting future firm performance face similar information asymmetries.

Researchers have identified several factors that may influence foreign investors' information disadvantage and hence their preference for domestic stock. Such factors include culture, geographic distance, and language (Lerner 1995, Coval and Moskowitz 1999, Grinblatt and Keloharju 2001, Ivkovic and Weisbenner

2005). Chan et al. (2005) suggest that stock market development and investors' linguistic and geographical familiarity with firms contribute to mutual funds' bias toward stocks of their home countries. Beugelsdijk and Frijns (2010) show that cultural aspects, such as individualism or uncertainty avoidance, affect wealth allocation between domestic and foreign investors. Geographic distance significantly hinders information flows despite the advances in information technology. Local banks' interest rates and fees are lower than those of non-local banks, implying that local banks may possess information advantages (Butler 2008). Finnish investors prefer stocks of adjacently located firms (Grinblatt and Keloharju 2001). Language dissimilarity is another important factor. Lundholm et al. (2017) show that U.S. investors avoid firms located in French speaking providences of Canada, and Hau (2001) shows that investors located in non-German speaking cities show lower profits from their trades in the German market.

My study relates financial reporting readability to the aforementioned line of research in capital markets. I contend that information access to 10-K reports is a cleaner measure of information search activities and therefore provides direct evidence about the impact of reporting readability on investors' information search activities, which in turn influences their portfolio-decision.

2.3. Hypothesis Development

Prior research documents that understanding linguistic cues is critical to investment decisions. Investors gather information about a firm's future performance from news articles (Tetlock et al. 2008) and distinguish stale news from fresh news (Tetlock 2011). Lack of understanding of such cues might

jeopardize investors' performance. For example, individual investors often misinterpret the linguistic tone of management forecasts, resulting in unintended wealth transfer to large institutional investors (Baginski et al. 2018). Deciphering linguistic cues is more challenging when the disclosures are complex or investors face a language gap. Brochet et al. (2016) show that market response to a conference call is weaker when the firm frequently uses erroneous expressions and non-plain English, and that such finding is more pronounced when there are language gaps between the firm and participants of the conference call.

Annual reports (Form 10-K) are an important source of information about firms' performance and financial status and are intended to help investors in their investment decisions. The SEC has encouraged registrant firms to prepare their annual reports using plain language to improve the readability (SEC 1998). Consistent with readability of written disclosures influencing users' decision making, Rennekamp (2012) shows, in an experiment setting, that small investors react to news more strongly when financial statements are more readable. Similarly, Elliott et al. (2015) finds in a lap experiment that investors are more likely to invest when the information is written in a more concrete language. In addition, annual reports with higher readability mitigate information asymmetry, resulting in improved analysts' forecast accuracy (Lehavy et al. 2011), lower cost of capital (Bonsall and Miller 2017), and higher firm value (Hwang and Kim 2017). These studies suggest that when investors find a firm's financial statements less readable, they are likely to have higher information processing costs.

I argue that the adverse impact of poor readability on information acquisition would be more pronounced for foreign investors, because they are at an

informational disadvantage over domestic investors and have a language gap that domestic investors do not have. Related to this conjecture, Lundholm et al. (2014) find that foreign companies cross-listed in the U.S. market improve the readability of their annual reports and provide more numerical data to attract U.S. investors. Extending Lundholm et al. (2014), this study investigates whether readability of U.S firms' annual reports has an asymmetric effect on foreign investors versus domestic investors in their information search activities. The foregoing discussion leads to the following hypothesis in the alternate form:

H1: Low readability of 10-K filings is more likely to decrease foreign investors' information acquisition activity than that of domestic investors.

3. Data, Research Design, and Descriptive Statistics

3.1. Data

The sample begins with all the U.S. firms with 10-K filings during 2002 to 2016.[5] After matching accounting and market variables from Compustat and CRSP, I delete observations in the financial industry (SIC Code 6000-6999) because foreign investors' search for financial firms may be determined by different factors. I delete observations with missing or negative total assets from the sample, resulting in 92,480 firm-year observations. I merge this data with the EDGAR log files that contains the history of investor access to 10-K filings. I delete firm-year

[5]I start my sample period from 2002 because, from 2003, the U.S. Securities and Exchange Commission (SEC) made the history of investors' access to 10-K filings available to the public in its Electronic Data Gathering, Analysis, and Retrieval (EDGAR) system. Since firms disclose their 10-K filings after their fiscal year ends, the IP access log files in year t+1 correspond to the access to the 10-K filing of fiscal year t. Thus, the sample periods are from 2002 to 2016.

observations with multiple 10-K filings on the same date, because these observations may introduce noise to my investor access measure. Finally, I require non-missing data for the EDGAR access and readability measures, arriving at the final sample of 38,539 firm-year observations. When I limit the sample to firms with at least one foreign access, the sample size decreases to 23,868 observations. For the analysis foreign institutional ownership, I obtain information from the Factset/Lionshares. Table 1 reports the detailed sample selection procedure.

<Insert Table 1 here>

3.2. Key Variable Definition

3.2.1. Foreign Access to 10-K filings (*FOREIGN_ACCESS*)

I identify foreign access using log files provided by EDGAR database.[6] The SEC provides log files containing all IP addresses that accessed filings in the EDGAR database from 2003 to 2017.[7] The format of an IP address is xxx.xxx.xxx.xxx, where x is a number between 0 and 9. The log files provide IP addresses after they mask three last digits to protect personal information. I identify location of the IP addresses using dataset provided by IP2Location LITE data.[8] I delete IP accesses by robot activities (e.g., web-crawling activities) – Appendix A reports the details. Based on the locations of IP addresses, I classify investor accesses into accesses from foreign countries and those from domestic investors. In

[6] https://www.sec.gov/dera/data/edgar-log-file-data-set.html
[7] The SEC provides log files up to 2017 as of the end of 2020.
[8] http://www.ip2location.com

addition, I exclude accesses from countries[9] where English is the primary language, because there is no language gap in those countries. For the analysis, I use the ratio of the number of foreign accesses to the number of domestic accesses for each 10-K filing multiplied by 100 (*FOREIGN_ACCESS*) because I am interested in the effect of financial reporting readability on foreign investors relative to the effect on domestic investors. It also helps address the possibility that the frequency of foreign access is driven by investors' general interest level, i.e., a correlated omitted variable in the estimation.

3.2.2. Foreign Investors' Additional Information Search (*COSEARCH*)

Alongside the volume of searches of firm's 10-K filing, I use the sequence of searches as another variable that can capture investors' information acquisition activities. Lee et al. (2015) proposes that sequence of searches in EDGAR reflects investors' perception about the firms they access. They find that firms accessed in sequence have fundamental firm characteristics in common. When a firm's 10-K filing is complex and difficult to understand, investors may supplement their understanding by referring to peers' financial reports (co-search, henceforth). I use co-search as a proxy for costs incurred in information acquisition and processing. Specifically, I first identify foreign IPs that search for at least two firms in a given day. Next, for each IP (i.e., foreign investor), I count the total number of accesses after searching for 10-K filing of firm *i* (focal firm). Finally, I average the IP-level post searches by firm-year. *COSEARCH* is the natural logarithm of one plus the

[9] Specifically, accesses from Australia, Canada, Hong Kong, Ireland, New Zealand, United Kingdom, Virgin Islands-British and Virgin Islands-US are excluded when measuring foreign accesses.

number of post searches.

3.2.3. Financial Reporting Readability (*BOG*)

To capture the readability of 10-K filings, I use BOG index following Bonsall et al. (2017).[10] BOG index is a multifaceted measure of readability calculated by *StyleWriter* software. The main strength of BOG Index is that it captures plain English writing attributes (e.g., passive voice, hidden verbs, legal terms, etc.). The SEC adopted the Securities Act Rule 421(d) that requires firms to write their disclosures in plain English and provides a handbook on how to write a disclosure in plain English (SEC 1998). BOG index directly incorporates aspects of plain English suggested in the handbook by constructing the measure with three components: (1) sentence length, (2) word features including the plain English style and word difficulty, and (3) writing attributes that help readers better understand texts.

Among numerous measures of financial statement readability (e.g., FOG index, Flesch Score, etc.), I believe BOG index is most suitable in this research for the following reasons. First, it is in line with the plain English concept proposed by the SEC (1998). Second, it overcomes some of the shortcomings of other readability proxies. BOG index is more advanced in capturing the word difficulty compared to FOG index by calculating it based on general vocabulary[11]. File size is often used to proxy for readability (Loughran and McDonald 2014). However, it is noisier than

[10] https://kelley.iu.edu/bpm/activities/bogindex.html

[11] Loughran and McDonald (2014) point out that the FOG index is poorly applicable in financial context, since it measures the complexity of words based on syllable counts. This results in common business words (e.g., corporation, depreciation, etc.) being complex.

BOG index in capturing readability considering that attachments unrelated to 10-K text (e.g., separate exhibits, HTML, XML, PDF) may increase the file size. Thus, I use BOG index in the main analysis and provide results using FOG index in the robustness test.

3.3. Research Design

To test the impact of financial reporting readability on foreign investors' information acquisition activities as opposed to the activities of domestic investors, I use following OLS model:

$$FOREIGN\_ACCESS_{i,t} = \beta_0 + \beta_1 BOG_{i,t} + \gamma_k \sum_k Control_{k,i,t} + \epsilon_{i,t} \qquad (1)$$

The dependent variable, *FOREIGN_ACCESS* is the number of foreign access divided by the number of U.S. access. *BOG* is BOG index as described previously and an inverse measure of financial reporting readability. That is, a larger value of *BOG* indicates lower readability. I control for firm attributes that may affect foreign investors' access to 10-K filings. Control variables include factors the prior research has identified as being associated with foreign investors' interests and their investment decisions. I control for firm size (*SIZE*), because foreign investors tend to prefer large firms (Falkenstein 1996; Gompers and Metrick 2001; Ferreira and Matos 2008). Considering that investors are likely to be interested in profitable firms, I include *ROA* and *LOSS* to control for such a tendency. I also control for market-to-book ratio (*MB*), since domestic and foreign investors show different preferences toward value stocks (Ferreira and Matos 2008; Baik et al. 2013). Foreign investors are known to favor firms paying less dividends,

firms that are less levered, and firms with large cash holding (Dahlquist and Robertsson 2001). I control for Dividend yield (*DIV*), Leverage (*LEV*), and cash holding (*CASH*) in the analyses to account for such investors' behavior. In addition, I control for whether the firm reports foreign incomes (*FI*), because foreign investors may be more interested in firms with greater amount of foreign income (*FO*) than firms operating only in the U.S. I also control for *CAPX* and *INTANGIBLE* to capture foreign investors' interests over the firms' risk associated with their investment decisions. The litigation risk (*LITIGATION*) is included in the regression because the physical remoteness and information asymmetry might induce foreign investors to be more sensitive to the legal issues. Appendix B provides the detailed variable descriptions. Provided that the adverse effect of poor financial reporting is more salient for foreign investors, I expect that the coefficient on *BOG* (β_1) will be negative.

3.4. Descriptive Statistics

Table 2 provides the descriptive statistics of variables used in the analyses. The mean (median) frequency of access from foreign investors (*FOREIGN_ACCESS*) constitutes 13.5 percent (6.3 percent) of the access from domestic (U.S.) investors. Specifically, the sample firms had on average 2.7 foreign accesses and 21.8 domestic (U.S.) accesses to their 10-K with seven days following the 10-K filing date. BOG index, the variable of interest, ranges from 70 to 103 with a mean of 85.507. It shows that 31.7 percent of firms in my sample report foreign income (*FI*) and the average foreign institutional ownership (*FORINST*) is 3.914 percent. *COSEARCH* is the number of 10-K filings of other firms that the foreign

investors access after downloading the focal firm's 10-K reports. The mean (median) *COSEARCH* of the sample firms is 10.4 (5.9).

<Insert Table 2 here>

3.5. Correlation Matrix

Table 3 reports Pearson correlations among the variables used in the analyses. Inconsistent with the prediction, *FOREIGN_ACCESS* is positively associated with *BOG* in the univariate level. *FOREIGN_ACCESS* is significantly associated with various firm characteristics. Foreign access ratio (*FOREIGN_ACCESS*) is negatively related with firm size (*SIZE*), market-to-book (*MB*), leverage (*LEV*), profitability (*ROA*), and capital expenditures (*CAPX*). Loss dummy (*LOSS*), litigious industry (*LITIGATION*), foreign income dummy (*FI*) and cash ratio (*CASH*) are positively associated with the ratio of foreign access to domestic access. In addition, *FOREIGN_ACCESS* is positively associated with foreign institutional ownership, suggesting that foreign interest leads to foreign investment in the following period.

<Insert Table 3 here>

4. Main Analysis Results

4.1. Test of H1

I test H1 by regressing the *FOREIGN_ACCESS* on financial reporting readability (*BOG*) as described in Equation (1). Columns (1) and (2) report regression estimates for the full sample, and columns (3) and (4) report the estimates for the subsample of firms that have at least one foreign access within seven days

of 10-K filing disclosure. Table 4 reports the results.

<Insert Table 4 here>

The coefficients on *BOG* (β_1), across all four columns, are negative and significant (p-value < 0.05 or smaller). This is consistent with H1 that fewer foreign investors access a firm's 10-K filing, when its 10-K report has lower readability. The effect of financial reporting readability is also economically significant. As *BOG* increases by a one standard deviation, the foreign access decreases by 41 percent relative to the domestic access for the full sample (Model 2) and by 95 percent for the subsample with at least one foreign access (Model 4). The result suggests that an increase in readability of financial report facilitates information acquisition of foreign investors, thereby mitigating the information asymmetry between foreign and domestic investors.

The coefficients on *SIZE* are significantly negative in columns (2) and (4). Considering the dependent variable is the ratio of foreign access to domestic access, I can interpret the result as foreign investors being less likely to seek larger firms than domestic investors do. It is also possible that foreign investors easily get information about large firms without searching for firms' 10-K filings. The coefficients of *MB* are significantly negative in both columns, indicating that foreign investors are more likely to seek value firms. *LEV* presents negative coefficients in both columns consistent with the prior literature, implying that foreign investors are reluctant to search for highly levered firms. The coefficient on *LOSS* is negative and significant only for the subsample of firms with at least one foreign access. The coefficients on *FI* in columns (2) and (4) show that foreign

investors are more likely to search for firms reporting foreign income. The coefficients on *CASH* and *INTANGIBLE* in column (2) indicate that foreign investors more frequently access firms with less cash and less intangibles assets, yet their coefficients are no longer significant in column (4). Consistent with the prior literature, *DIV* presents a significant and negative coefficient in column (2) but exhibits an insignificant coefficient in column (4).

4.2. Does the frequent access lead to foreign investments?

Having documented that 10-K filings with high readability are associated with more frequent foreign access, I next examine whether foreign accesses lead to foreign holdings of firm shares in the subsequent year. To the extent that financial reporting readability helps mitigate the information gap between foreign and domestic investors, I expect that an increase in the frequency of foreign access to 10-K filings will lead to larger foreign holdings. I test this prediction by regressing foreign institutional ownership on foreign access to 10-K filings (*FREIGN_ACCESS*). I use one-year lag value of *FOREIGN_ACCESS* in the regression to address the concern about reverse causality, i.e., foreign access would be higher simply because the foreign institution invested in those stocks. Besides all control variables used in Equation (1), I also control for lagged institutional ownership to mitigate the effect of firm-specific interests the foreign institutions may have. Table 5 reports the regression estimates.

<Insert Table 5 here>

Table 5 column (1) shows the regression result using the full sample and column (2) shows the result for the subset of firms with at least one foreign access.

In both columns, the coefficients on $FOREIGN\_ACCESS_{t-1}$ are positive and significant, indicating that larger frequency of foreign access leads to higher foreign institutional ownership in the subsequent period. Based on the coefficient of 0.002 on $FOREIGN\_ACCESS_{t-1}$, a one standard deviation increases in foreign access (18.204) is associated with an increase in the foreign institutional ownership by 3.64 percent. This effect is economically significant given that the average foreign institutional ownership in the sample is 3.91 percent.

Overall, the results in Tables 4 and 5, collectively, suggest that financial reporting readability, through its effect on foreign access, can facilitate foreign investors' information acquisition from the EDGAR database.

4.3. Post-reading behavior

My hypothesis assumes that financial reporting readability influences the efficiency of foreign investors' information acquisition from 10-K filings. To further examine this assumption, I examine whether readability of financial report affects investors' post-reading behavior. Lee et al. (2015) suggest that firms that are sequentially searched in EDGAR share similar characteristics, such as growth rate and R&D expenditures. When a firm's 10-K filing is complex and difficult to understand, investors may search other related-information that may clarify the complex information (i.e., co-search). Using co-search as a proxy for costs of information acquisition and processing, I examine how foreign-investors' co-search behavior varies with financial reporting readability. My measure of investors' post-reading behavior (*COSEARCH*) indicates the number of 10-K filings of other firms that investors search for after accessing the focal firm's 10-K filing. Table 6 reports

regression estimates where *COSEARCH* is regressed on focal firms' BOG Index (*BOG*).

<Insert Table 6 here>

The result suggests that low financial reporting readability significantly increases foreign investors' information acquisition cost. The coefficients on *BOG* are 0.005 ($p < 0.01$) for the full sample and 0.002 ($p < 0.05$) for the sub-sample of firm years with non-zero foreign access. In other words, foreign investors search more firms after reading ′less readable 10-K filing of focal firm. This is consistent with foreign investors finding another similar firm's 10-K filing to supplement their lack of understanding of focal firm's 10-K filing. It is also possible that foreign investors shift their interest from less readable 10-K to more readable ones. In either case, this finding provides further insight on foreign investors' preference for firms with higher financial reporting readability.

5. Additional Analysis and Robustness Tests

5.1. Language similarity

The analysis shows that foreign investors more frequently access 10-K filings with higher readability. I have attributed in part their 10-K filing search for firms with high readability to language gap. In this section, I examine this premise by using the variation in foreign investors' language similarity. Prior research suggests that foreign investors are at information disadvantages and have a language gap compared to domestic investors. However, it is unclear how the benefit of higher financial reporting readability will vary with the extent of their language similarity to English. On the one hand, financial reporting readability may

be more beneficial when the language gap is larger, i.e., it substitutes foreign investors' English proficiency. On the other hand, its benefit will be larger when the language gap is smaller, i.e., financial reporting readability may complement foreign investors' familiarity with English when processing information in 10-K filings. Therefore, I am agnostic about the direction of the relationship. Instead, I expect that there will be a systematic variation in the relationship between financial reporting readability and foreign access across the level of language similarity, if language contributes to the relationship.

I employ two language characteristics to identify the level of foreign investors' exposure to English. First, I employ the language similarity (typology similarity) between English and other languages (Borin and Saxena 2013). The typology similarity captures structural similarity between English and foreign languages. Since the measure is language specific, I first identify the country of each IP and match the language to the IP based on their countries. I then divide all languages into two groups: high and low typology similarity. Finally, I count the number of IP accesses for each firm-group. As an alternative measure of the language gap, I use the language diversity (linguistic diversity index, hereafter LDI) from SIL international.[12] Higher diversity indicates that the people in the country are more likely to be multilingual, and I relate it to greater exposure to English. Similar to the language similarity, I categorize all IPs into two groups: high and low diversity groups-based on the LDI. Then, I count the number of IP accesses for each firm-group.

<Insert Table 7 here>

[12] https://en.wikipedia.org/wiki/Linguistic_diversity_index

Using the number of IP access by groups, I estimate equation (1) for each group. Panel A of Table 7 shows the regression results using the typology similarity as a partitioning variable. I find that the coefficients on *BOG* are negative and significant only in columns (2) and (4) where the dependent variable (*FOREIGN_ACCESS*) is measured based on accesses from countries of which language has a high level typological similarity to English. I do not find such pattern for foreign access from countries with low language similarity. The results are similar when I use language diversity as a measure of foreign investors' exposure to English (Panel B of Table 7). I find *BOG* is significantly associated with accesses from foreign countries with high level of language diversity. Overall, these results are consistent with financial reporting readability complementing, rather than substituting, foreign investors' English proficiency in their information acquisition from 10-K filing.

5.2. Does readability of 10-Ks affect portfolio selection for foreign investors?

Findings in Tables 4 and 5 show that foreign investors access 10-K filings more frequently when the filings are easier to read, and that more frequent foreign access is associated with higher foreign ownership in the following period. These findings support the view that higher readability of 10-K filings mitigates the information gap between domestic and foreign investors. Foreign investors' information disadvantage is in twofold. First, it limits their portfolio selection. Foreign investors are biased toward their domestic market (Chan et al. 2005) and the foreign markets with same languages (Grinblatt and Keloharju 2001; Lundholm et al. 2017). When investors invest in a foreign country, they prefer more visible

and less risky firms (Kang and Stulz 1997). Second, for the same stock, foreign investors do not possess as much information as the local investors have (Chan et al. 2008). In this section, I examine whether financial reporting readability enable foreign investors to widen the range of portfolio selection.

Based on Kang and Stulz (1997), I choose firm visibility and risk as the criteria foreign investors use in selecting their portfolio firms. If high financial reporting readability encourages foreign investors to research firms that they usually do not include in their portfolio (i.e., low visibility and/or high risk firms), I will observe a significant impact of the readability on their access to 10-K filings of these firms. I measure firms' visibility with size and the number of geographic segments, and firm risk with R&D expenditures and the absolute value of discretionary accruals (Dechow et al. 1995). I classify firms into high visibility (risk) group and low visibility (risk) group based on the sample median and estimate Equation (1) separately for each subgroup. Table 8 reports the results.

<Insert Table 8 here>

Panel A of Table 8 shows cross-sectional analyses using the proxies for visibility. I find that the association between *BOG* and *FOREIGN_ACCESS* is negative and significant for high visibility firms (t-values in the range of -2.47 ~ -4.54), but the association is insignificant for low visibility firms (t-values in the range of 0.36 ~ -1.74). In other words, an easy writing of 10-K filings does not necessarily attract foreign investors when the firm is less visible. I find similar results when partitioning the sample into low- and high-risk groups. In Panel B of Table 8, the coefficients on *BOG* are negative and significant only for low-risk firms (i.e., low R&D expenditure or low absolute value of discretionary accruals). Overall,

the results in Table 8 suggest that high readability does not necessarily widen foreign investors' portfolio selection. Instead, it helps foreign investors process information for those firms that fit into their portfolio selection criteria.

5.3. Robustness tests

I perform a battery of robustness tests. First, I use the FOG index (Gunning 1952) used frequently in the previous literature (e.g., Li 2008) as an alternative measure of financial reporting readability to examine whether findings are robust to alternative measure of readability. Second, I use the natural logarithm of the number of IP accesses as an alternative dependent variable. In the main analysis, I use the ratio of foreign access to domestic access multiplied by 100. This likely mitigates the effect of general interests to the firm, but taking the ratio may bias the results when domestic and foreign investors search for completely different stocks for their investment decisions. Third, I re-estimate the results after obtaining foreign accesses from English-speaking foreign countries only. The aforementioned results indicate that foreign investors search for 10-K filings with high readability, possibly due to their language gap. However, the language constraint is not binding for foreign investors from English speaking countries. Therefore, if the association between financial reporting readability and foreign access is attributable to foreign investors' language gap, I may not observe such association (or much weaker association at least) when I limit foreign access to English speaking countries. I report the results of these robustness tests in Table 9.

<Insert Table 9 here>

Panel A shows regression results using the FOG index. Consistent with the

main results in Table 4, the coefficients on *FOG* are negative and significant across all four columns, supporting *H1*. Panel B provides regression results using the number of accesses as the dependent variable. I count the number of domestic accesses (accesses from IP within the U.S.) and the number of foreign accesses (accesses from IP outside the U.S.). Using these two counts as dependent variables, I estimate Equation (1). In columns (1) and (3) of Panel B, the coefficients on *BOG* are significant when I use the number of foreign accesses as the dependent variable. However, when I use the number of domestic accesses as a dependent variable in columns (2) and (4), the coefficients on *BOG* are insignificant. This result is consistent with H1 that readability affects reading experience of foreign investors more strongly than that of domestic investors.

Panel C of Table 9 shows regression results using the IP access from English speaking foreign countries. The results in columns (1) and (2) show that the coefficients on *BOG* are not significant at the 5 percent level for the full sample. When I use the subsample of firms having at least one foreign access (columns (3) and (4)), the coefficients on *BOG* are significant. However, juxtaposition of the coefficient on *BOG* in column (4) and that in column (4) of Table 4 offers an interesting comparison: the magnitude of the former (-0.062) is lower than half the magnitude of the latter (-0.139). Overall, the results show that absent the language gap, the readability affects investors' information acquisition less significantly. This finding is consistent with the results in Panel B of Table 8, showing that domestic investors may not consider readability as important as foreign investors do when searching for the 10-K filings.

6. Conclusion

Existing studies suggest that foreign investors are at information disadvantage compared to domestic investors, and this information disadvantage impedes foreign investors' investment in domestic stocks. In this paper, I investigate whether readability of 10-K filings mitigates foreign investors' information disadvantage. I find that foreign investors access 10-K filings more frequently when the readability of 10-K filings is higher. I also find that more frequent access by foreign investors leads to higher foreign institutional ownership in the subsequent period, and that foreign investors are more likely to access 10-K filings of other firms after they access a less readable 10-K. These findings, together, suggest that making financial statements more readable helps foreign investors acquire and process information in 10-filings, thereby increasing foreign investment through reduced information disadvantage over domestic investors.

I further find evidence that language similarity has an interactive effect on the relation between readability and foreign access. I find that the association is more pronounced when language similarity increases. This result is consistent with the view that financial reporting readability does not substitute but complement foreign investors' English proficiency in their acquisition of information from 10-K filings. While readability improves foreign investors' information acquisition, I also find that readability does not widen the search area of foreign investors who generally focus more on safer firms (i.e., more visible and less risky firms). Findings are robust to various sets of sensitivity tests. The association between readability and foreign access is robust when I use the FOG index and the raw number of accesses instead of the BOG index and the ratio of access, respectively.

Combined, the results emphasize the importance of financial readability by showing that more readable financial statements facilitate foreign investors to increase information acquisition activity and investment of U.S. stocks.

Findings in this research contribute to the literature on reporting readability (Miller 2010; Lehavy et al. 2011; Lawrence 2013; Bonsall and Miller 2017). Reporting readability affects several aspects of investors' decision because higher readability reduces information acquisition and processing costs, enabling investors to acquire information more easily. This research also extends the literature on home bias by showing additional evidence that language is the root cause of the bias and suggesting that improved information environments reduce investors' home bias.

REFERENCES

Ahern, K. R., D. Daminelli, and C. Fracassi. 2015. Lost in translation? The effect of cultural values on mergers around the world. *Journal of Financial Economics* 117 (1): 165–189.

Allee, K. D., L. S. Anderson, and M. J. Crawley. 2019. The impact of linguistic distance and financial reporting readability on foreign holdings of U.S. stocks. Working Paper.

Ashbaugh-Skaife, H., D. W. Collins, and W. R. Kinney. 2007. The discovery and reporting of internal control deficiencies prior to SOX-mandated audits. *Journal of Accounting and Economics* 44 (1–2): 166–192.

Bae, K.-H., R. M. Stulz, and H. Tan. 2008. Do local analysts know more? A cross-country study of the performance of local analysts and foreign analysts. *Journal of Financial Economics* 88 (3): 581–606.

Bae, K.-H., H. Tan, and M. Welker. 2008. International GAAP differences: The impact on foreign analysts. *The Accounting Review* 83 (3): 593-628.

Baik, B., J. Kang, J. Kim, and J. Lee. 2013. The liability of foreignness in international equity investments: Evidence from the U.S. stock market. *Journal of International Business Studies* 44 (4): 391-411.

Baginski, S. P., E. Demers, A. Kausar, and Y. J. Yu. 2018. Linguistic tone and the small trader. *Accounting, Organizations and Society* 68-69 (July): 21–37.

Beatie, B. A. 1977. "Reading German": A self-tutorial skills-learning package. *The Modern Language Journal* 61 (8): 410–415.

Beugelsdijk, S., and B. Frijns. 2010. A cultural explanation of the foreign bias in international asset allocation. *Journal of Banking & Finance* 34 (9): 2121–2131.

Biddle, G., G. Hilary, and R. Verdi. 2009. How does financial reporting quality relate to investment efficiency? J*ournal of Accounting and Economics* 48 (2-3): 112–31.

Bonsall IV, S. B., A. J. Leone, B. P. Miller, and K. Rennekamp. 2017. A plain English measure of financial reporting readability. *Journal of Accounting and Economics* 63 (2–3): 329–357.

Bonsall IV, S. B., and B. P. Miller. 2017. The impact of narrative disclosure readability on bond ratings and the cost of debt. *Review of Accounting Studies* 22 (2): 608–643.

Borin, L., and A. Saxena. 2013. *Approaches to measuring linguistic differences*. Vol. 265. Walter de Gruyter.

Bradshaw, M., B. Bushee, and G. Miller. 2004. Accounting choice, home bias, and U.S. investment in non-U.S. firms. *Journal of Accounting Research* 42: 795–841.

Brochet, F., P. Naranjo, and G. Yu. 2016. The capital market consequences of language barriers in the conference calls of non-U.S. firms. *Accounting Review* 91 (4): 1023–1049.

Bushee, B., and C. Noe. 2000. Corporate disclosure practices, institutional investors, and stock return volatility. *Journal of Accounting Research* 38: 171–202.

Butler, A. W. 2008. Distance still matters: Evidence from municipal bond underwriting. *Review of Financial Studies* 21 (2): 763–784.

Chan, K., V. Covrig, and L. Ng. 2005. What determines the domestic bias and foreign bias? Evidence from mutual fund equity allocations worldwide. *The Journal of Finance* 60 (3): 1495–1534.

Chan, K., A. J. Menkveld, and Z. Yang. 2008. Information asymmetry and asset prices: Evidence from the China foreign share discount. *The Journal of Finance* 63 (1): 159–196.

Cho, H., V. Muslu, and M. Koo. 2020. Lost in translation? Analysts' forecasts of cross-listed firms. *Journal of International Accounting Research* (forthcoming)

Coval, J. D., and T. J. Moskowitz. 1999. Home bias at home: Local equity preference in domestic portfolios. *Journal of Finance* 54 (6): 2045-2073.

Dahlquist, M., and G. Robertsson. 2001. Direct foreign ownership, institutional investors, and firm characteristics. *Journal of Financial Economics* 59 (3): 413-440.De Franco. G., O. K. Hope., D. Vyas., and Y. Zhou. 2015. Analyst report readability. *Contemporary Accounting Research* 32 (1): 76-104.

Dechow. P. M., R. G. Sloan., and A. P. Sweeney. 1995. Detecting earnings management. *The Accounting Review* 70 (2): 193-225.

Dvořák, T. 2005. Do domestic investors have an information advantage? Evidence from Indonesia. *The Journal of Finance* 60 (2): 817–839.

Elliott, W. B., K. M. Rennekamp, and B. J. White. 2015. Does concrete language in disclosures increase willingness to invest? *Review of Accounting Studies* 20 (2): 839–865.

Falkenstein, E. G. 1996. Preferences for stock characteristics as revealed by mutual fund portfolio holdings. *The Journal of Finance* 51 (1): 111-135.

Ferreira, M. A., and P. Matos. 2008. The colors of investors' money: The role of institutional investors around the world. *Journal of Accounting and Economics* 88 (3): 499-533.

Guay, W., D. Samuels., and D. Taylor. 2016. Guiding through the Fog: Financial statement complexity and voluntary disclosure. *Journal of Accounting and Economics* 62 (2-3): 234-269.

Gunning, R. 1952. The technique of clear writing. McGraw-Hill International Book Co., New York, NY.

Gompers, P. A., and A. Metrick. 2001. Institutional investors and equity prices. *The Quarterly Journal of Economics* 116 (1): 229-259.

Grinblatt, M., and M. Keloharju. 2001. How Distance, Language, and Culture Influence Stockholdings and Trades. *The Journal of Finance* 56 (3): 1053–1073.

Hau, H. 2001. Location matters: An examination of trading profits. *The Journal of Finance* 56 (5): 1959–1983.

Hwang, B.-H., and H. H. Kim. 2017. It pays to write well. *Journal of Financial Economics* 124 (2): 373–394.

Ivkovic, Z., and S. Weisbenner. 2005. Local does as local is: Information content of the geography of individual investors' common stock investments. *Journal of Finance* 60 (1): 267-306.

Kang, J.-K., and R. Stulz. 1997. Why is there a home bias? An analysis of foreign portfolio equity ownership in Japan. *Journal of Financial Economics* 46 (1): 3–28.

Lawrence, A. 2013. Individual Investors and Financial Disclosure. *Journal of Accounting and Economics* 56 (1): 130-147.

Lee, C. M. C., P. Ma., and C. C. Y. Wang. 2015. Search-based peer firms: Aggregating investor perceptions through internet co-searches. *Journal of Financial Economics* 116 (2): 410-431

Lehavy, R., F. Li, and K. Merkley. 2011. The effect of annual report readability on analyst following and the properties of their earnings forecasts. *The Accounting Review* 86 (3): 1087–1115.

Lerner, J. 1995. Venture capitalists and the oversight of private firms. *The Journal of Finance* 50 (1). 301-318.

Li, F. 2008. Annual report readability, current earnings, and earnings persistence. *Journal of Accounting and Economics* 45 (2–3): 221–247.

Loughran, T., and B. McDonald. 2014. Measuring readability in financial disclosures. *Journal of Finance* 69 (4): 1643–1671.

Loughran, T., and B. McDonald. 2017. The use of EDGAR filings by investors. *Journal of Behavioral Finance* 18 (2): 231–248.

Lundholm, R. J., R. Rogo, and J. L. Zhang. 2014. Restoring the Tower of Babel: How foreign firms communicate with U.S. investors. *Accounting Review* 89 (4): 1453–1485.

Lundholm, R., N. Rahman, and R. Rogo. 2017. The foreign investor bias and its linguistic origins. *Management Science* 64 (9): 4433-4450.

Miller, B. P., 2010. The effects of disclosure complexity on small and large investor trading. *The Accounting Review* 85 (6): 2107-2143.

Rennekamp, K. 2012. Processing fluency and investors' reactions to disclosure readability. *Journal of Accounting Research* 50 (5): 1319–1354.

Ryans, J. 2017. Using the EDGAR log file data set. *SSRN Electronic Journal*.

Securities and Exchange Commission. 1998. A Plain English Handbook: How to Create Clear SEC Disclosure. SEC Office of Investor Education and Assistance Washington. District of Columbia.

Schulz, R. A. 1981. Literature and readability: Bridging the gap in foreign language reading. *The Modern Language Journal* 65 (1): 43–53.

Slangen, A. H. L. 2011. A communication-based theory of the choice between greenfield and acquisition entry. *Journal of Management Studies* 48 (8): 1699–1726.

Tenzer, H., M. Pudelko, and A. Harzing. 2014. The impact of language barriers on trust formation in multinational teams. *Journal of International Business Studies; Basingstoke* 45 (5): 508–535.

Tetlock, P. C. 2011. All the news that's fit to reprint: Do investors react to rtale information? *Review of Financial Studies* 24 (5): 1481–1512.

Tetlock, P. C., M. Saar-Tsechansky, and S. Macskassy. 2008. More than words: Quantifying language to measure firms' fundamentals. *The Journal of Finance* 63 (3): 1437–1467.

Table 1
Sample Selection Procedure

	Firm-year observations
U.S. firms from 2002 to 2016 in *Compustat*	140,715
Less: Financial industry (SIC Codes 6000-6999)	(21,847)
Less: Missing or negative total assets	(26,388)
	92,480
Less: Unable to use the EDGAR data	(39,380)
Less: Disclosing more than one 10-K filings on a same date	(62)
Less: Missing variables used in the analyses	(14,499)
Full Sample	38,539
Less: Zero access from foreign IPs	(14,671)
Sample with at least one foreign access	23,868

Table 1 presents the sample selection procedure. Beginning from the U.S. firms available in the Compustat from 2002 to 2016, I follow the above procedure that leaves me a final sample of 38,539 firm-year observations. When I further restrict the sample to have at least one foreign access, the sample size reduces to 23,868 firm-year observations.

Table 2
Descriptive Statistics

Variable	N	Mean	SD	Min	Q1	Median	Q3	Max
FOREIGN_ACCESS (%)	38,539	13.563	22.107	0	0	6.25	16.667	133.333
BOG	38,539	85.507	6.824	70	81	85	90	103
FACCESS_N (raw)	38,539	2.683	4.922	0	0	1	3	33
DACCESS_N (raw)	38,539	21.836	25.243	1	7	14	27	155
SIZE	38,539	5.633	2.283	0.726	3.953	5.622	7.286	10.863
MB	38,539	1.325	0.737	0.212	0.832	1.169	1.624	4.334
LEV	38,539	0.153	0.177	0	0	0.087	0.267	0.686
ROA	38,539	-0.035	0.547	-3.903	-0.024	0.095	0.164	0.57
CAPX	38,539	0.045	0.059	-0.019	0.011	0.026	0.054	0.346
LOSS	38,539	0.406	0.491	0	0	0	1	1
LITIGATION	38,539	0.268	0.443	0	0	0	1	1
FI	38,539	0.317	0.465	0	0	0	1	1
CASH	38,539	0.239	0.257	0	0.04	0.138	0.359	0.973
INTANGIBLE	38,539	0.17	0.204	0	0.002	0.083	0.276	0.794
DIV	38,539	0.008	0.02	0	0	0	0.006	0.127
COSEARCH (raw)	38,539	10.404	5.952	0	7.4	10.321	13.357	31
FORINST (%)	23,305	3.914	4.755	0	0.445	2.162	5.702	24.585
RET (%)	23,305	0.17	0.647	-0.837	-0.213	0.077	0.385	3.226

Table 2 provides the descriptive statistics for the variables used in the analyses. All continuous variables are winsorized at 1% of each tail. See appendix B for variable definitions.

Table 3
Pearson Correlation Matrix

		FOREIGN_ACCESS	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]	[11]	[12]	[13]
[1]	*BOG*	**0.04**													
[2]	*SIZE*	**-0.05**	**0.04**												
[3]	*MB*	**-0.02**	**0.12**	0.00											
[4]	*LEV*	**-0.06**	**0.02**	**0.44**	**0.09**										
[5]	*ROA*	**-0.06**	**-0.29**	**0.40**	**-0.13**	**0.15**									
[6]	*CAPX*	**-0.05**	**-0.14**	**0.13**	**-0.02**	**0.17**	**0.14**								
[7]	*LOSS*	**0.02**	**0.20**	**-0.38**	**0.02**	**-0.06**	**-0.60**	**-0.06**							
[8]	*LITIGATION*	**0.04**	**0.23**	**-0.21**	**0.14**	**-0.20**	**-0.28**	**-0.15**	**0.20**						
[9]	*FI*	**0.03**	**-0.03**	**0.37**	**0.05**	**0.02**	**0.26**	**-0.09**	**-0.29**	**-0.07**					
[10]	*CASH*	**0.04**	**0.30**	**-0.39**	**0.27**	**-0.40**	**-0.50**	**-0.23**	**0.30**	**0.40**	**-0.12**				
[11]	*INTANGIBLE*	-0.01	**0.07**	**0.23**	**0.01**	**0.21**	**0.16**	**-0.25**	**-0.11**	**-0.07**	**0.19**	**-0.31**			
[12]	*DIV*	0.01	**-0.11**	**0.25**	**-0.05**	**0.11**	**0.17**	0.01	**-0.20**	**-0.16**	**0.06**	**-0.18**	**-0.01**		
[13]	*FORINST*	**0.05**	**0.08**	**0.52**	**0.12**	**0.17**	**0.14**	**0.05**	**-0.16**	**-0.04**	**0.21**	**-0.07**	**0.12**	**0.09**	
[14]	*COSEARCH*	**-0.10**	**-0.07**	**0.02**	**-0.08**	**-0.03**	**0.09**	**0.03**	**-0.07**	**-0.04**	**0.03**	**-0.05**	**-0.02**	**0.03**	**-0.05**

Table 3 presents the Pearson correlation for the variables used in the analyses. The sample period is from 2002 through 2016. Correlations in bold are significant at $p < 0.05$. See appendix B for variable definitions.

Table 4
Readability and Foreign Access

	Dependent Variable = *FOREIGN_ACCESS*			
	Full Sample		Foreign Access > 0	
BOG	-0.133***	-0.060**	-0.357***	-0.139***
	(-3.00)	(-2.19)	(-7.70)	(-4.01)
SIZE		-0.559**		-2.448***
		(-2.40)		(-7.91)
MB		-1.003**		-1.945***
		(-2.87)		(-4.36)
LEV		-5.057***		-3.705*
		(-3.87)		(-2.07)
ROA		-0.421		-0.180
		(-1.40)		(-0.42)
CAPX		-3.893		-1.962
		(-1.40)		(-0.39)
LOSS		-0.238		-0.935**
		(-1.02)		(-2.63)
LITIGATION		-0.030		-0.414
		(-0.06)		(-0.63)
FI		1.379***		1.596***
		(3.95)		(3.14)
CASH		-1.917**		-1.735
		(-2.55)		(-1.35)
INTANGIBLE		-2.883*		-2.710
		(-2.14)		(-1.70)
DIV		-12.002**		-5.442
		(-2.15)		(-0.75)
Constant	24.935***	24.814***	52.698***	52.775***
	(6.56)	(6.79)	(13.13)	(13.22)
Observations	38,539	38,539	23,868	23,868
Adjusted R^2	0.064	0.071	0.051	0.095
Industry and Year Fixed	Yes	Yes	Yes	Yes

Table 4 presents the regression result of Equation (1) to test whether readability of 10-K filings is positively associated with foreign access. The dependent variable is *FOREIGN_ACCESS*, the ratio of foreign access to domestic access multiplied by 100. The variable of interest is *BOG*. Columns (1) and (2) report the results of full sample. Columns (3) and (4) show the results of sample with at least one foreign access. See Appendix B for variable definitions. *t*-statistics are in parentheses. Standard errors are clustered at both yearly and firm level. ***, **, and * denote two-sided significance at the 1%, 5%, and 10% level, respectively.

Table 5
Foreign Access and Foreign Ownership

	Dependent Variable: Foreign Institutional Ownership	
	Full Sample	Foreign Access > 0
FOREIGN ACCESS$_{t-1}$	0.002**	0.002**
	(2.21)	(2.94)
SIZE	0.268***	0.275***
	(10.52)	(9.59)
MB	0.317***	0.310***
	(7.90)	(6.63)
LEV	-0.409	-0.427
	(-1.73)	(-1.47)
ROA	0.271**	0.245
	(2.28)	(1.44)
CAPX	1.516**	2.560***
	(2.74)	(3.04)
LOSS	-0.116**	-0.164***
	(-2.56)	(-3.69)
LITIGATION	-0.015	0.011
	(-0.25)	(0.15)
FI	0.004	0.022
	(0.12)	(0.50)
CASH	0.709***	0.803***
	(6.82)	(8.05)
INTANGIBLE	0.435***	0.465***
	(3.90)	(3.10)
DIV	-4.025***	-4.800***
	(-3.90)	(-4.53)
RET	0.059	0.079
	(1.65)	(1.40)
FORINST$_{t-1}$	0.850***	0.857***
	(75.85)	(74.19)
Constant	-1.436***	-1.532***
	(-7.25)	(-7.57)
Observations	23,305	14,874
Adjusted R^2	0.784	0.796
Industry and Year Fixed	Yes	Yes

Table 5 reports the result of whether the frequent access by foreign investors leads to foreign investment in the subsequent period. The dependent variable is foreign institutional ownership at year *t*. The variable of interest is *FOREIGN_ACCESS*$_{t-1}$, the ratio of foreign access to domestic access at year *t*-1 multiplied by 100. Column (1) reports the results of full sample. Column (2) shows the results of sample with at least one foreign access. See Appendix B for variable definitions. *t*-statistics are in parentheses. Standard errors are clustered at both yearly and firm level. ***, **, and * denote two-sided significance at the 1%, 5%, and 10% level, respectively.

Table 6
Readability and Investors' Co-search Behavior

	Dependent Variable: *COSEARCH*	
	Full Sample	Foreign Access > 0
BOG	0.005**	0.002**
	(2.83)	(2.15)
SIZE	0.073**	0.017
	(2.42)	(1.12)
MB	-0.042**	-0.014
	(-2.95)	(-1.26)
LEV	-0.081*	-0.027
	(-1.78)	(-0.78)
ROA	0.100**	0.052
	(2.67)	(1.51)
CAPX	-0.221**	-0.280**
	(-2.15)	(-2.92)
LOSS	0.020	-0.003
	(1.20)	(-0.21)
LITIGATION	-0.001	-0.033**
	(-0.07)	(-2.37)
FI	-0.001	0.018
	(-0.06)	(1.20)
CASH	0.026	-0.066*
	(0.59)	(-1.83)
INTANGIBLE	-0.115**	-0.085**
	(-2.73)	(-2.65)
DIV	-0.412	0.028
	(-1.41)	(0.15)
Constant	1.465***	1.999***
	(5.01)	(11.90)
Observations	38,539	23,868
Adjusted R^2	0.415	0.592
Industry and Year Fixed	Yes	Yes

Table 6 shows the co-search behavior after foreign investors search for the focal firm's 10-K. The dependent variable is the log value of one plus the average number of access the foreign investors search for after accessing the focal firm's 10-K. The variable of interest is *BOG*. Column (1) reports the results of full sample. Column (2) shows the results of sample with at least one foreign access. See Appendix B for variable definitions. *t*-statistics are in parentheses. Standard errors are clustered at both yearly and firm level. ***, **, and * denote two-sided significance at the 1%, 5%, and 10% level, respectively.

Table 7
The Effects of Language Similarity

Panel A. Language difficulty for foreign investors

	Full Sample		Foreign Access > 0	
	Low Typology Similarity	*High* Typology Similarity	*Low* Typology Similarity	*High* Typology Similarity
BOG	0.007	-0.060***	0.013	-0.119***
	(0.88)	(-3.50)	(1.14)	(-4.83)
SIZE	-0.137*	-0.047	-0.591***	-0.964***
	(-2.10)	(-0.46)	(-5.48)	(-6.01)
MB	-0.277	-0.246	-0.536*	-0.609**
	(-1.74)	(-1.61)	(-2.10)	(-2.95)
LEV	-1.153**	-3.180***	-0.945	-2.761**
	(-2.19)	(-5.61)	(-1.13)	(-2.66)
ROA	-0.093	-0.346*	0.019	-0.329
	(-1.36)	(-1.88)	(0.11)	(-1.26)
CAPX	0.015	-1.224	0.951	0.990
	(0.01)	(-0.98)	(0.35)	(0.34)
LOSS	-0.019	0.251	-0.193	0.135
	(-0.20)	(1.63)	(-0.97)	(0.61)
LITIGATION	0.242	-0.309	0.357	-0.705
	(1.33)	(-1.25)	(1.06)	(-1.67)
FI	0.235**	0.577***	0.283*	0.561*
	(2.21)	(3.36)	(1.90)	(1.94)
CASH	-0.470	-1.023**	-0.565	-1.255
	(-1.13)	(-2.25)	(-0.65)	(-1.66)
INTANGIBLE	-0.952**	-0.889	-1.298*	-0.445
	(-2.19)	(-1.67)	(-1.83)	(-0.58)
DIV	-5.025**	-8.456*	-6.368*	-9.794
	(-2.66)	(-2.11)	(-2.00)	(-1.55)
Constant	3.413***	12.007***	8.351***	28.091***
	(3.05)	(6.75)	(4.59)	(14.31)
Observations	38,539	38,539	20,565	20,565
Adjusted R^2	0.060	0.050	0.070	0.057
Industry and Year Fixed	Yes	Yes	Yes	Yes

Panel B. Language Diversity

	Full Sample		Foreign Access > 0	
	Foreign Access from *Low* Diversity	Foreign Access from *High* Diversity	Foreign Access from *Low* Diversity	Foreign Access from *High* Diversity
BOG	-0.027*	-0.028**	-0.049**	-0.061***
	(-2.02)	(-2.58)	(-2.48)	(-3.36)
SIZE	0.173*	-0.371***	-0.348**	-1.233***
	(2.06)	(-4.06)	(-2.88)	(-7.87)
MB	-0.030	-0.540**	-0.226	-1.005***
	(-0.27)	(-2.58)	(-1.63)	(-3.08)
LEV	-2.763***	-1.741**	-2.714***	-1.296
	(-6.03)	(-2.58)	(-3.44)	(-1.20)
ROA	-0.502***	0.101	-0.762**	0.546**
	(-3.32)	(0.81)	(-2.63)	(2.16)
CAPX	-1.238	0.353	0.078	2.517
	(-1.27)	(0.23)	(0.04)	(0.73)
LOSS	0.073	0.128	-0.164	0.051
	(0.60)	(0.77)	(-0.82)	(0.17)
LITIGATION	0.004	-0.075	-0.077	-0.290
	(0.02)	(-0.37)	(-0.21)	(-0.86)
FI	0.410**	0.461**	0.354	0.590**
	(2.85)	(2.78)	(1.44)	(2.36)
CASH	-0.856**	-0.584	-1.054	-0.638
	(-2.61)	(-1.18)	(-1.74)	(-0.68)
INTANGIBLE	-0.829	-1.109*	-0.747	-1.149
	(-1.73)	(-2.14)	(-1.07)	(-1.30)
DIV	-6.029**	-8.279**	-7.015	-10.631**
	(-2.21)	(-2.75)	(-1.71)	(-2.18)
Constant	6.207***	9.625***	15.237***	21.958***
	(4.54)	(6.60)	(9.52)	(10.66)
Observations	38,539	38,539	20,565	20,565
Adjusted R^2	0.047	0.062	0.043	0.074
Industry and Year Fixed	Yes	Yes	Yes	Yes

Table 7 presents the result of a test to examine whether the association between readability and foreign access is more pronounced for foreign investors who are proficient in English. I use language similarity measure (typology similarity) in panel A and language diversity (LDI) in panel B to proxy for the language gap. To measure the dependent variable, I first categorize the languages that the foreign investors use in the sample into high and low group by the typology similarity (panel A) and LDI (panel B). Then, I count the foreign access by each group. In panel A, the dependent variable in columns (1) and (3) is the ratio of foreign access that shows *low* typology similarity to domestic access. The dependent variable in columns (2) and (4) is the ratio of foreign access that shows *high*

typology similarity to domestic access. In panel B, the dependent variable in columns (1) and (3) is the ratio of foreign access that shows *low* diversity to domestic access. The dependent variable in columns (2) and (4) is the ratio of foreign access that shows *high* diversity to domestic access. The variable of interest if the readability measure, *BOG*. Columns (1) and (2) report the results of full sample. Columns (3) and (4) show the results of sample with at least one foreign access. See Appendix B for variable definitions. *t*-statistics are in parentheses. Standard errors are clustered at both yearly and firm level. ***, **, and * denote two-sided significance at the 1%, 5%, and 10% level, respectively.

.

Table 8
Subsample Analyses

Panel A. Firm Visibility

	Full Sample				Foreign Access > 0			
	Size		Number of Geographic Segment		*Size*		Number of Geographic Segment	
	Small	*Big*	*Low*	*High*	*Small*	*Big*	*Low*	*High*
BOG	0.012	-0.103**	-0.014	-0.162***	-0.038	-0.156***	-0.040	-0.279***
	(0.36)	(-2.47)	(-0.47)	(-3.11)	(-0.60)	(-3.47)	(-0.87)	(-4.54)
SIZE	-0.782*	-0.112	-0.911***	-0.159	-3.345***	-1.744***	-2.937***	-1.904***
	(-2.00)	(-0.57)	(-3.39)	(-0.65)	(-7.75)	(-5.69)	(-8.66)	(-5.40)
MB	-1.414***	-0.702	-0.989**	-1.191**	-2.811***	-1.572**	-1.761***	-2.507***
	(-3.23)	(-1.53)	(-2.89)	(-2.41)	(-5.72)	(-2.80)	(-4.24)	(-3.83)
LEV	-4.702***	-3.312**	-5.400***	-3.457	-5.083**	-1.686	-4.405*	-2.186
	(-3.03)	(-2.71)	(-3.41)	(-1.71)	(-2.19)	(-1.01)	(-1.84)	(-1.02)
ROA	-0.249	0.879	-0.228	-2.459*	0.375	1.222	0.141	-2.455*
	(-0.77)	(0.80)	(-0.91)	(-1.87)	(0.89)	(1.09)	(0.36)	(-1.89)
CAPX	2.928	-8.762**	-3.942	1.366	9.264	-9.755*	-2.928	8.465
	(0.85)	(-2.47)	(-1.10)	(0.18)	(1.57)	(-1.78)	(-0.49)	(0.82)
LOSS	0.113	-0.907**	-0.567*	-0.395	0.235	-2.039***	-1.500**	-0.730
	(0.35)	(-2.31)	(-1.79)	(-0.91)	(0.39)	(-4.33)	(-2.71)	(-1.48)
LITIGATION	0.051	-0.260	-0.168	0.427	-0.111	-0.630	-0.577	0.006
	(0.10)	(-0.34)	(-0.35)	(0.54)	(-0.12)	(-0.72)	(-0.67)	(0.01)
FI	-0.397	2.134***	2.011***	-0.488	-0.053	2.231***	2.364***	-0.462
	(-0.82)	(5.22)	(4.05)	(-1.07)	(-0.08)	(3.90)	(3.12)	(-0.80)
CASH	-2.471**	0.450	-2.472**	-1.711	-2.847	1.078	-3.298**	0.954
	(-2.38)	(0.36)	(-2.71)	(-1.41)	(-1.72)	(0.65)	(-2.21)	(0.57)
INTANGIBLE	-1.858	-3.977**	-2.909*	-2.282	-2.813	-3.796**	-3.983*	0.042
	(-1.24)	(-2.91)	(-1.88)	(-1.19)	(-1.37)	(-2.40)	(-2.02)	(0.02)
DIV	-15.055	-4.216	-5.698	-16.415	-21.247	4.793	1.672	-8.249
	(-1.32)	(-0.56)	(-0.90)	(-1.33)	(-1.30)	(0.46)	(0.19)	(-0.60)
Constant	19.741***	24.316***	22.358***	32.976***	48.920***	47.406***	46.983***	62.461***

	(7.00)	(5.41)	(6.23)	(6.38)	(9.52)	(11.30)	(11.20)	(9.80)
Observations	19,334	19,204	24,416	14,121	10,302	13,566	14,294	9,574
Adjusted R^2	0.087	0.069	0.078	0.064	0.071	0.074	0.107	0.087
Industry and Year Fixed	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Panel B. Firm Risk

	Full Sample				Foreign Access > 0			
	R&D		Absolute value of discretionary accruals		R&D		Absolute value of discretionary accruals	
	Low	*High*	*Low*	*High*	*low* similarity	*high* similarity	*low* similarity	*high* similarity
BOG	-0.102***	0.001	-0.060*	-0.046	-0.157***	-0.096	-0.149***	-0.106*
	(-3.03)	(0.01)	(-1.78)	(-1.19)	(-3.20)	(-1.52)	(-3.48)	(-1.79)
SIZE	-0.736***	-0.303	-0.379	-0.739**	-2.760***	-2.046***	-2.157***	-2.876***
	(-3.23)	(-1.22)	(-1.46)	(-2.81)	(-7.53)	(-7.51)	(-6.69)	(-7.64)
MB	-0.911**	-1.117**	-0.994**	-1.197**	-1.653***	-2.178***	-1.673**	-2.465***
	(-2.72)	(-2.40)	(-2.26)	(-2.96)	(-3.42)	(-4.11)	(-2.89)	(-4.52)
LEV	-4.231**	-5.393***	-4.509**	-6.004***	-1.908	-5.503***	-4.127*	-3.674
	(-2.65)	(-4.34)	(-2.50)	(-3.89)	(-0.75)	(-4.01)	(-1.87)	(-1.47)
ROA	-0.221	-0.566*	-0.855	-0.194	-0.468	-0.044	-0.272	0.219
	(-0.45)	(-2.05)	(-0.71)	(-0.50)	(-0.57)	(-0.13)	(-0.17)	(0.44)
CAPX	-4.629	-4.013	-8.626	0.063	-2.335	-4.033	-10.864	5.864
	(-1.19)	(-0.65)	(-1.66)	(0.03)	(-0.38)	(-0.41)	(-1.55)	(0.95)
LOSS	-0.355	-0.179	-0.889*	0.496	-1.438**	-0.407	-1.477**	0.303
	(-1.08)	(-0.51)	(-1.85)	(1.56)	(-2.46)	(-0.77)	(-2.47)	(0.59)
LITIGATION	1.806*	-0.692	-0.137	-0.092	1.576	-1.090	-0.628	-0.091
	(2.09)	(-1.27)	(-0.24)	(-0.12)	(1.24)	(-1.54)	(-1.11)	(-0.08)
FI	2.530***	0.233	1.149**	1.847***	3.033***	0.330	1.374*	2.257***
	(5.09)	(0.49)	(2.26)	(4.34)	(3.77)	(0.57)	(1.93)	(3.84)
CASH	-0.836	-2.552**	-1.396	-2.213**	0.236	-2.201	-0.670	-2.260

	(-0.86)	(-2.34)	(-1.16)	(-2.18)	(0.14)	(-1.20)	(-0.39)	(-1.39)
INTANGIBLE	-4.473**	-2.191	-3.149**	-2.288	-4.522**	-1.812	-2.317	-2.910*
	(-2.53)	(-1.55)	(-2.19)	(-1.46)	(-2.20)	(-0.97)	(-1.26)	(-1.78)
DIV	-8.697	-15.703	-11.632	-19.223**	1.563	-22.998*	-1.755	-15.161
	(-1.29)	(-1.51)	(-1.06)	(-2.45)	(0.15)	(-1.87)	(-0.13)	(-1.07)
Constant	28.137***	19.797***	24.397***	24.286***	54.534***	48.558***	52.057***	52.153***
	(9.19)	(3.66)	(5.72)	(6.66)	(14.56)	(7.21)	(11.54)	(9.99)
Observations	19,160	19,377	18,854	16,294	11,613	12,251	12,137	9,526
Adjusted R^2	0.068	0.075	0.073	0.070	0.112	0.081	0.088	0.093
Industry and Year Fixed	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Table 8 reports the result of cross-sectional analyses to see whether readable 10-K filings affect the portfolio selection for foreign investors. In panel A, I use firm size and the number of geographic segments to proxy for the firm visibility. The dependent variable is $FOREIGN\_ACCESS_t$, the ratio of foreign access to domestic access multiplied by 100. The variable of interest is BOG_t. In panel B, I use R&D expenditures and the absolute value of discretionary accruals to capture the firm risk. Columns (1), (2) and (3) present the results of full sample. Columns (4), (5) and (6) show the results of sample with at least one foreign access. See Appendix B for variable definitions. *t*-statistics are in parentheses. Standard errors are clustered at both yearly and firm level. ***, **, and * denote two-sided significance at the 1%, 5%, and 10% level, respectively.

Table 9. Robustness Tests

Panel A. *FOG* index

	Dependent Variable = *FOREIGN_ACCESS*			
	Full Sample		Foreign Access > 0	
FOG	-0.566**	-0.412**	-1.291***	-0.718***
	(-2.75)	(-2.56)	(-6.57)	(-3.39)
SIZE		-0.577**		-2.504***
		(-2.50)		(-7.97)
MB		-0.935**		-1.825***
		(-2.76)		(-4.20)
LEV		-5.385***		-4.352**
		(-3.93)		(-2.39)
ROA		-0.259		0.041
		(-0.88)		(0.09)
CAPX		-3.978		-2.212
		(-1.44)		(-0.45)
LOSS		-0.284		-1.036**
		(-1.17)		(-2.80)
LITIGATION		-0.042		-0.474
		(-0.09)		(-0.74)
FI		1.391***		1.654***
		(3.92)		(3.23)
CASH		-2.097**		-2.285*
		(-2.78)		(-1.89)
INTANGIBLE		-3.004**		-3.071*
		(-2.22)		(-1.97)
DIV		-11.775*		-4.934
		(-2.11)		(-0.69)
Constant	24.904***	28.098***	47.992***	55.827***
	(6.02)	(5.83)	(12.04)	(10.38)
Observations	38,108	38,108	23,606	23,606
Adjusted R^2	0.064	0.071	0.048	0.095
Industry and Year Fixed	Yes	Yes	Yes	Yes

Panel B. Raw number of access

	Dependent Variable			
	Number of Foreign Access	Number of Domestic Access	Number of Foreign Access	Number of Domestic Access
	Full Sample		Foreign Access > 0	
BOG	-0.003**	-0.000	-0.005***	-0.001
	(-2.32)	(-0.39)	(-3.46)	(-0.95)
SIZE	0.125***	0.230***	0.106***	0.229***
	(15.07)	(19.15)	(12.37)	(17.99)
MB	0.054***	0.112***	0.042***	0.120***
	(5.32)	(9.36)	(3.68)	(9.53)
LEV	-0.387***	-0.258***	-0.363***	-0.282***
	(-7.73)	(-5.65)	(-8.03)	(-4.88)
ROA	-0.053***	-0.033***	-0.050***	-0.054***
	(-5.26)	(-3.12)	(-5.58)	(-4.06)
CAPX	-0.505***	-0.074	-0.555***	-0.181
	(-3.90)	(-0.49)	(-3.35)	(-1.02)
LOSS	0.059***	0.131***	0.065***	0.148***
	(3.65)	(7.80)	(5.26)	(7.39)
LITIGATION	0.028	0.048**	0.035	0.055**
	(1.33)	(2.42)	(1.51)	(2.38)
FI	0.047***	-0.038**	0.029*	-0.045**
	(3.60)	(-2.94)	(1.76)	(-2.79)
CASH	-0.136***	-0.068	-0.124***	-0.107**
	(-4.36)	(-1.61)	(-3.31)	(-2.52)
INTANGIBLE	-0.130**	-0.039	-0.133***	-0.061
	(-2.96)	(-1.07)	(-3.10)	(-1.65)
DIV	-0.526*	-0.004	-0.224	0.051
	(-1.90)	(-0.02)	(-1.00)	(0.15)
Constant	0.443***	1.286***	1.155***	1.617***
	(3.79)	(12.60)	(11.24)	(13.28)
Observations	38,539	38,539	23,868	23,868
Adjusted R^2	0.348	0.493	0.253	0.439
Industry and Year Fixed	Yes	Yes	Yes	Yes

Panel C. English speaking foreign country

	Dependent Variable: Number of Foreign Access / Number of Domestic Access			
	English Speaking Foreign Countries		Non-English Speaking Countries	
BOG	-0.039*	-0.022	-0.133***	-0.060**
	(-2.00)	(-1.22)	(-3.00)	(-2.19)
SIZE		-0.120		-0.559**
		(-1.04)		(-2.40)
MB		-0.188		-1.003**
		(-1.51)		(-2.87)
LEV		-2.347***		-5.057***
		(-4.23)		(-3.87)
ROA		-0.505**		-0.421
		(-2.50)		(-1.40)
CAPX		-2.025		-3.893
		(-1.19)		(-1.40)
LOSS		0.142		-0.238
		(0.85)		(-1.02)
LITIGATION		-0.523		-0.030
		(-1.75)		(-0.06)
FI		0.920***		1.379***
		(4.69)		(3.95)
CASH		-0.674		-1.917**
		(-1.12)		(-2.55)
INTANGIBLE		-0.118		-2.883*
		(-0.24)		(-2.14)
DIV		-20.169***		-12.002**
		(-4.46)		(-2.15)
Constant	24.935***	24.814***		24.814***
	(6.56)	(6.79)		(6.79)
Observations	38,539	38,539	38,539	38,539
Adjusted R^2	0.064	0.071	0.064	0.071
Industry and Year Fixed	Yes	Yes	Yes	Yes

Table 9 reports the robustness test of Equation (1). In panel A, I use *FOG* instead of *BOG* as the readability measure. In panel B, the dependent variable is replaced from *ACCESS* to *FACCESS_N* in columns (1) and (3) and *DACCESS_N* in columns (2) and (4). In panel C, I use countries that use English as their primary languages when calculating the foreign access in columns (1) and (2). Columns (3) and (4) show the results of countries that does not use English as their primary languages. See Appendix B for variable definitions. *t*-statistics are in parentheses. Standard errors are clustered at both yearly and firm level. ***, **, and * denote two-sided significance at the 1%,

5%, and 10% level, respectively.

APPENDIX A
Robot Access

I define robot access following Ryans (2017). Ryans (2017) suggests three ways of identifying robot access as following.

1. *Rpv* assumes that IPs from human activity do not download more than 25 items per minute, more than three firm per minute, and more than 500 items per day.
2. *LMpv* assumes that if an IP downloads more than 50 filings in a day, then the IP is a robot access.
3. *DRTpv* assumes that IPs from human activity do not download more than 1,000 items per day and 5 items per minute.

I construct all three measures, and identify an IP as a robot access if the access from IP satisfies at least one of the three conditions.

APPENDIX B
Variables Description

Variable	Definition
FOREIGN_ACCESS	The ratio of the EDGAR access from foreign users to that from domestic (U.S.) users multiplied by 100 (in percentage).
BOG	Readability measure of firm's financial reports based on Bonsall et al. (2017).
FOG	Readability measure of firm's financial reports based on Gunning (1952).
FACCESS_N	Natural logarithm of one plus the number of foreign access.
DACCESS_N	Natural logarithm of the number of domestic (U.S.) access.
SIZE	Natural logarithm of total assets.
MB	The ratio of market value of equity to the book value of equity.
LEV	The ratio of total long-term debt to total assets.
ROA	The ratio of operating income before depreciation to total assets of beginning of the fiscal year.
CAPX	The ratio of net capital expenditure to total assets.
LOSS	An indicator variable that equals one if the firm reports negative net income in the fiscal year.
LITIGATION	An indicator variable that equals one if the firm's industry is a litigious industry (SIC code is 2833-2836, 3570-3577, 3600-3674, 5200-5961, and 7370), and zero otherwise (Ashbaugh-Skaife et al. 2007).
FI	An indicator variable that equals one if the firm reports foreign income, and zero otherwise.
CASH	The ratio of cash and short-term investments to total assets.
INTANGIBLE	The ratio of total intangible assets to total assets.
DIV	The ratio of dividends to the market capitalization.
FORINST	The ratio of foreign institutional ownership to the market capitalization (in percentage).
COSEARCH	The log value of one plus the average number of access that the EDGAR users search for after accessing the focal firm's financial report.
RET	Cumulated monthly returns over 12-month during the

	fiscal year.

국문초록

공시와 기업의 행태에 관한 연구

본 논문은 기업의 공시 활동과 기업의 행태에 관한 두 개의 독립적인 연구로 구성되어 있다. 첫 번째 논문은 사업보고서 공시의 주석 내용을 통해 파악한 세무조사 위험과 기업의 세무 행태 간 연관성을 분석하였다. 미국 국세청은 기업의 규모 별로 매년 세무조사가 이루어진 비율을 공개하는 데, 이는 기업-연도별 자료가 아니므로 이용하는 데에 한계가 있고, 기업이 실제로 세무조사에 어느 정도의 위협을 느끼는지 알기 부족하다. 이에 본 연구는 세무조사 위험을 기업이 세무조사에 대해 느끼는 기업-연도 수준의 위협으로 보고, 잠재 디히클레 할당(LDA)을 이용하여 법인세비용 관련 주석에서 세무조사 주제를 다루는 비중으로 측정하였다. 분석 결과, 기업이 세무 관련 주석에서 세무조사에 관해 많은 논의를 할수록 납세순응도가 높은 것을 발견했다. 이는 기업이 세무조사에 대해 큰 위험을 느낄수록 세무조사에 관한 많은 내용을 공시하고, 그에 따라 납세순응적인 행태를 보이는 것이라 해석할 수 있다. 본 연구의 관심변수인 세무위험 주제는 미국 국세청의 기업 규모-연도별 실제 세무조사 비율과 양의 상관관계를 가지며, 특히 세무조사 비율을 통제한 후에도 세무위험 주제는 유의한 강건성을 보이는 것으로 나타났다. 기업이 세무조사에 대해 느끼는 위협은 기업 특성 별로 그 민감도가 다르게 나타났다. 기관투자자와 같은 외부 감독 기구의 비중이 높을 경우, 그리고 다국적 기업일 경우 세무조사에 의한 감독 기능이 약화되어, 세무조사 위험이 납세순응도에 미치는 영향 역시 약화된 양상을 보였다.

두번째 논문은 기업의 재무보고 가독성이 외국인 투자자의 정보습득 행위에 미치는 영향에 대해 연구하였다. 영어가 모국어가 아닌 외국인 투자자는 언어 차이로 인해 미국인 투자자보다 기업의 사업보고서를 이해하는 데 어려움이 있을 수 있다. 이에 본 연구는 사업보고서 상 가독성이 높을수록 외국인 투자자들이 그러한 기업의

사업보고서에 더 많이 접속하는 것으로 나타났다. 또한 이러한 가독성이 높은 보고서에 대한 관심은 실제 투자로도 이어져, 이듬해 외국인 기관투자자의 비율이 높아짐을 발견하였다. 재무보고의 가독성과 외국인 투자자의 관심도 사이의 양의 관계는 특이하게도 외국인 투자자의 언어와 영어 간의 유사도가 클수록 더욱 강화된 양상을 보였다. 또한 외국인 투자자는 가독성이 낮은 재무보고를 접속한 후, 부족한 정보격차를 보완하기 위해 더 많은 기업들의 재무보고를 후에 접속하는 것으로 나타났다. 본 연구는 기업의 재무보고 가독성이 외국인 투자를 촉진하는 데 중요한 역할을 함을 보여줌으로써, 기업이 가독성이 높은 재무보고를 작성하는 것의 이점을 보여준 데에 의의가 있다.

주요어: 세무조사, 납세순응, 머신러닝, 정보습득, 기관 투자자, 외국인 투자자, 가독성, 자국편의.

학번: 2017-37066