

**Tax Avoidance and Disclosure Strategies:
Evidence from Annual Report Tone Ambiguity**

Abstract

We examine the relation between corporate tax avoidance and the ambiguity in the tone of managerial narrative disclosures in firm's annual reports. We find that firms engaging in more tax avoidance issue more ambiguous annual reports. The positive relation is more pronounced for firms with a higher tax-based proprietary information cost or weaker internal monitoring. To mitigate the ambiguity-induced information asymmetry, tax-aggressive firms are found to supplement their mandatory disclosure with earlier and more frequent voluntary disclosure to investors, especially when the tax-based proprietary information cost is low. Our findings indicate that the disclosure strategies of tax-avoiding firms are determined by the trade-off between the benefits of reducing information asymmetry and the costs of tax-based proprietary information.

Keywords: corporate tax avoidance; disclosure tone; mandatory disclosure; voluntary disclosure; textual analysis; readability

JEL Classifications: G32, G34, M41, M42

1. Introduction

Corporate tax avoidance is generally considered a value-enhancing activity since it transfers wealth from the government to corporate shareholders (Desai and Dharmapala 2009; Goh et al. 2016; Edwards et al. 2016; Tang 2019). However, communication with outside investors can become a challenge to the managers of tax-avoiding firms. On the one hand, managers would prefer to disclose more value-relevant information (e.g., strong profitability and growth prospects) to investors to reduce external financing costs. On the other hand, those managers need to obscure information that can potentially reveal tax avoidance strategies when making public financial disclosures.¹ The reason is that tax authorities pay substantial attention to public financial disclosures such as annual 10-K reports to monitor firms' tax avoidance strategies (Bozanic et al. 2017), and that being audited and penalized by tax authorities can be costly for both shareholders and managers. Therefore, managers of tax-avoiding firms face a disclosure trade-off between the benefits of reducing information asymmetry and the tax-based proprietary costs of information revealed to tax authorities. Examining how this trade-off issue affects tax-avoiding firms' financial disclosure strategies appears to be intriguing and important.

The emerging stream of tax studies has started exploring the effect of tax avoidance on corporate information environment (Balakrishnan et al. 2019; Inger et al. 2018; Nguyen 2021). For example, Balakrishnan et al. (2019) find that tax aggressiveness reduces corporate information transparency. The underlying notion is that tax avoidance activities induce organizational and financial complexity and create difficulty for outsiders to interpret the source and persistence of the firm's earnings and cash flows. Using textual analysis, Inger et al. (2018) and Nguyen (2021) explore the mechanism by which tax-avoiding firms complicate their qualitative disclosure to

¹ Bozanic et al. (2017) find firm size, profitability, and foreign exposure are positively related to IRS attention.

hinder the IRS's ability to identify and challenge tax avoidance activity. They find that managers of firms with high levels of tax avoidance issue less-readable tax footnotes or financial statements, as evidenced by higher FOG or BOG indices.² We extend this line of research by examining the relation between tax avoidance and tone ambiguity—a textual property of qualitative disclosure beyond complexity.

In addition to readability, ambiguous text in annual reports can be a source of informational risk and interfere with users' ability to comprehend reports (Ertugrul et al. 2017). Studies have shown that the tone of financial disclosures plays a vital role in shaping corporate information environment. For instance, Loughran and McDonald (2013) show that the tone in the first SEC filing in the initial public offering (IPO) process affects investors' ability to value an IPO. Specifically, they find that firms with a higher proportion of weak modal words in their filings subsequently have higher stock return volatility, suggesting that ambiguous financial disclosures make it difficult for investors to accurately assess a firm's value. More ambiguous tone in 10-K reports is associated with higher cost of bank loans, greater future stock price crash risk (Ertugrul et al. 2017), more corporate cash holdings, and a higher probability of derivatives use (Friebert and Seiler 2017). As information in financial statements can help tax authorities identify questionable items in firms' tax returns and potentially illegitimate tax avoidance activities (Mills and Sansing 2000; Bozanic et al. 2017), we predict that tax-avoiding firms have strong incentives to adopt an ambiguous tone to obscure information that may increase the risk of being audited and challenged (e.g., the firm's true underlying earnings, items generating large book-tax differences).³

² In their studies, the FOG index is measured by the average words per sentence and the percent of 10-K complex words whereas the BOG index is constructed by (1) the average sentence length across the whole document; and (2) the complexity of words in the document related to English style problems (e.g., passive verbs, hidden verbs, overwriting, legal terms, and so on) and word difficulty (e.g., heavy words, abbreviations, and specialist terms).

³ Anecdotal evidence also confirms our prediction. For instance, the sample averages of cash effective tax rates (percentage of weak modal words) for Apple Inc., Amazon.com Inc., HP Inc., DISH Network Corp., and Salesforce.com Inc., are 16.9% (0.676), 14.2% (0.751), 17.1% (0.779), 10.3% (1.078), and 8.6% (0.916), respectively,

Following Loughran and McDonald (2011), we define tone ambiguity as the percentage of weak modal words (e.g., *depending*, *might*, *perhaps*, *somewhat*, and *possible*) in 10-Ks.⁴ We utilize four existing measures of tax avoidance, including permanent book-tax differences, cash effective tax rates, long-run cash effective tax rates, and industry- and size-matched cash effective tax rates. We also construct a first principal component of the above four tax avoidance measures to avoid measurement errors associated with various aggressive tax positions. Using a sample of publicly listed U.S. firms over the period 1994–2017, our baseline finding indicates that firms use weaker modal words in 10-Ks when they avoid more taxes.

We then further examine whether the positive association between tone ambiguity and tax avoidance is driven by the tax-based proprietary costs (i.e., explicit costs if the revealed tax information is used against firms by the tax authority). Prior studies document that tax directors and managers are fully aware of tax-based proprietary costs (Hoopes et al. 2018; Graham et al. 2014). Firms at a higher risk of being audited, having subsidiaries in tax havens, or having foreign operations face a higher tax audit risk and hence are more motivated to obscure tax information.⁵ Consistent with this notion, we find that the positive association between tone ambiguity and tax avoidance is more pronounced for these firms. This finding suggests that it is the concern over the tax-based proprietary information costs that drives managers of tax-avoiding firms to narrate a tone of ambiguity.

over the period 1994–2017. These firms' cash effective tax rates are much lower than the sample mean (median) value of 27.1% (26.0%), while the values of their 10-K tone ambiguity are far greater than the sample mean (median) value of 0.456 (0.448). This suggests that tax aggressive firms use more ambiguous verbiage in 10-Ks as their tax avoidance increases.

⁴ See Appendix B for the ambiguity word list from Loughran and McDonald (2011).

⁵ For example, firms with foreign exposure face greater IRS attention (Bozanic et al. 2017). Tax-avoiding firms tend to withhold disclosure of subsidiaries located in tax havens (Dyrend et al. 2020) and their geographic earnings (Hope et al. 2013) in 10-Ks to mitigate the IRS scrutiny.

One disadvantage of the ambiguous 10-K filings, however, is the increased corporate cost of borrowing (Ertugrul et al. 2017) and valuation uncertainty (Loughran and McDonald 2011). The frequent use of weak modal words exacerbates the information asymmetry between managers and outside investors, consequently generating greater difficulty for investors and analysts to forecast income for tax-aggressive firms. While listed firms face public pressure to pay more tax, they also face pressure from shareholders to maximize after-tax profit. Given that tax avoidance is considered a value-enhancing activity (Hanlon et al. 2014) and investors demand more tax information to understand profitability and growth (Ernst & Young 2011), managers also have incentives to convey such value-relevant information to investors via voluntary disclosure in an environment where managers have more control over the information flow. As such, tax-avoiding firms provide us with an ideal setting to investigate how firms choose between mandatory and voluntary disclosure strategies to balance the cost of information asymmetry against the cost of proprietary information.

We predict and find that tax-avoiding firms tend to engage in earlier and more frequent voluntary disclosure (measured by VOLDISC_DUMMY, VOLDISC_FREQ and VOLDISC_HORIZON). These results suggest that managers believe that tax avoidance itself brings value to investors and that voluntary disclosure is helpful to lower their overall external financing costs, especially when there is a demand from investors for information transparency and when the use of ambiguous tone in mandated disclosure increases firms' cost of capital (Ertugrul et al. 2017). Presumably, managers supplement mandatory disclosure with voluntary disclosure to relay information that is related to the result of successful tax planning or firm performances masked and "downplayed" with ambiguous verbiage in the 10-Ks. Furthermore, an additional test reveals that the positive relation heightens after the implementation of Regulation

Fair Disclosure (Reg FD) in 2000, confirming that tax-aggressive managers use additional voluntary disclosures to communicate with investors for balancing the cost of information asymmetry against the cost of proprietary information.⁶ Our findings are consistent with Balakrishnan et al. (2019) who find that tax-aggressive managers mitigate transparency problems by providing supplemental disclosure. Interestingly, we find that the positive relation is sensitive to tax audit risks. Overall, our findings paint a comprehensive picture of managers' disclosure strategies in the context of tax avoidance.

We also investigate whether monitoring mechanisms affect the extent to which managers of tax-avoiding firms manipulate the tone in 10-K disclosures. Our results show that tax-avoiding firms with strong (weak) board monitoring (proxied by board size, board independence, and director ownership) are less (more) likely to use ambiguous tone in reporting.

To test the validity of our inferences, we conduct a battery of robustness checks. First, we employ a propensity-score matching sample analysis to alleviate the concerns of model misspecification (Shipman et al. 2017). Our results show that the positive relation between tax avoidance and tone ambiguity is not driven by controlled observable firm characteristics that are associated with both tax avoidance and tone ambiguity. Second, to alleviate the endogeneity concern in our baseline regression, we conduct a quasi-natural experiment by exploiting exogenous shocks to tax avoidance through the implementation of "Check-the-Box" regulations in 1997.⁷ Our propensity-score matching difference-in-differences (DID) test shows a significant increase in tax-avoidance-induced tone ambiguity in the post-1997 period, corroborating our baseline results.

⁶ Reg FD was introduced by the Securities and Exchange Commission (SEC) to promote transparency and accountability. Before the passage of Reg FD, publicly-traded firms could release important information in meetings and earnings/forecast calls where shareholders and the general public were excluded. Under Reg FD, public companies need to simultaneously disclose such material nonpublic information to the general public.

⁷ "Check-the-box" regulations, effective January 1, 1997, require U.S. multinationals create new legal entities in order to elect "disregarded entity status", offering an opportunity for U.S. multinationals to circumvent an anti-avoidance provision (see Blouin and Krull (2016) for a detailed discussion).

Finally, our results are robust to alternative proxies for tone ambiguity and controlling for various 10-K readability measures and uncertain tax benefits (UTB).

Our paper makes several important contributions. First, to the best of our knowledge, we are among the first to investigate the association between tax avoidance and annual report tone ambiguity, which is a crucial aspect of narrative financial disclosures in 10-K reports (Loughran and McDonald 2011, 2014). Our evidence that tax-avoiding firms use weak tone to obfuscate information in order to minimize tax scrutiny enriches our understanding of the link between tax avoidance and corporate information environment (Gallemore and Labro 2015; Balakrishnan et al. 2019). While Balakrishnan et al. (2019) find that tax aggressiveness alters the transparency of firms' aggregate reporting environment (beyond just transparency of tax-specific reporting), the channel through which these firms weaken transparency remains unclear. We extend this line of research by focusing on linguistic ambiguity—a mechanism by which tax-avoiding firms interfere with users' ability to comprehend reports through information opacity. Our results suggest that the intentional use of hard-to-read and weak-tone reports to complicate and obfuscate information worsens information asymmetry and affects the quality of overall corporate information environment. Our findings complement prior research that interprets complex language in firms' disclosures as indicative of managerial obfuscation and information asymmetry (Loughran and McDonald 2016; Bushee et al. 2018).

Our study complements and extends concurrent papers examining the relationship between tax avoidance and readability (e.g., Inger et al. 2018; Nguyen 2021) in three aspects. First, we investigate the effect of tax avoidance on managerial obfuscation instead of complexity of qualitative disclosure. While readability and linguistic tone can be correlated, they do capture different attributes of textual disclosure. Second, Ertugrul et al. (2017) find that 10-Ks with low

readability and ambiguous tone are associated with higher costs of capital, and that when firms try to obscure mandated earnings-relevant information, they are more likely to conceal the information in longer and more ambiguous documents than to use complex words. We provide additional evidence that after controlling for the 10-K readability, tax-avoiding firms continue to issue more ambiguous reports to lower their tax-based proprietary information costs. Third, we empirically examine whether tax-avoiding firms use voluntary disclosure as a supplement to alleviate the increased cost of information asymmetry associated with ambiguous mandatory disclosure.

Finally, our paper adds to the disclosure substitution literature. Prior work shows that managers substitute voluntary disclosure for mandatory disclosure when the proprietary cost associated with providing financial statements increases (Heinle et al. 2022) and that managers use voluntary disclosure to mitigate the negative effects of complex financial statements on the information environment (Guay et al. 2016). In a similar vein, Balakrishnan et al. (2019) find that tax-aggressive firms provide more detailed management discussion and analysis (MD&A) sections of the 10-K report and lengthier conference calls to mitigate the cost of low transparency resulting from tax aggressiveness. We extend this line of research by showing that managers of tax-avoiding firms are willing to increase voluntary disclosure through management earnings forecasts, which is another important source of information disclosed to investors. Consistent with the notion that managers are concerned about the increase in the likelihood that tax authorities identify tax planning activities and take action against the firm, we find the positive association between tax avoidance and voluntary disclosure is only prominent when firms face a sufficiently low risk in tax audit. Our new findings suggest that managers are fully aware of the costs of voluntary and mandatory disclosure and that for tax-avoiding firms, the tax-based proprietary information cost is an important factor that determines their strategies of mandatory and voluntary disclosure.

Overall, we believe our paper addresses an important issue that relates a tax phenomenon to an economic phenomenon. Specifically, we provide new insights into tax avoidance and corporate financial disclosure by demonstrating that tax avoidance creates incentives for managers to weigh different disclosure choices to reach an optimal information environment and that the trade-off between the tax-based propriety information costs and external financing costs is a key determinant of tax-avoiding firms' disclosure strategies. Our results have implications for investors, tax authorities, shareholders, and regulators who traditionally rely on firms' quantitative information to evaluate the extent and value of tax avoidance.

The remainder of the paper proceeds as follows. Section 2 reviews the related literature and develops our hypotheses. Section 3 delineates data sources, variable construction, and research methodology. Section 4 presents empirical results. Section 5 concludes.

2. Prior research and hypothesis development

2.1 Tax avoidance

The most apparent benefit of tax avoidance is cash savings. Theoretically, the cash savings lead to improved free cash flows to the firm and benefit shareholders through more dividend distributions and increased firm value. For example, Edward et al. (2016) document that financially constrained firms engage in more tax avoidance to generate additional internal funds. Such activity leads to the increases of between \$3.82 and \$5.15 million in tax savings and between 2.87 and 4.82 percent of operating cash flows. Goh et al. (2016) provide evidence on the negative association between tax avoidance and cost of equity capital, indicating that equity investors require a lower expected rate of return for tax-avoiding firms due to tax savings from corporate tax avoidance. In an international study of 46 countries, Tang (2019) finds that the average relation between tax avoidance and firm value is positive and significant.

While tax avoidance is documented to be value-enhancing, it is not costless. Prior literature suggests that the net benefit of tax avoidance is determined by the size of non-tax costs: e.g., reputational losses, agency costs, penalties, auditing risk, and political costs (Mills 1998; Desai and Dharmapala 2009, Tang 2019). One major non-tax cost is the risk of being audited by tax authorities. Mills (1998) and Mills and Sansing (2000) document that tax avoidance, as reflected by book-tax differences, is positively associated with IRS audit risk. Once a firm's tax positions are challenged and overturned by tax authorities, high direct and indirect costs can arise. Direct costs consist of litigation and related expenses in mounting a defense against the tax authority's challenge, paying back owed taxes, interests, and penalties, and subsequently heightened scrutiny by tax authorities. Indirect costs lay in the potential loss of reputation, which may harm the firm's future cash flows and market value. Hanlon and Slemrod (2009) find that a company's stock price declines when there is news about its involvement in tax shelters. Using a survey of corporate tax executives, Graham et al. (2014) show that reputation concern is cited by 70% of firms as important or very important in deciding whether to avoid taxes. In the same survey, 58% of firms rated the risk of adverse media attention as important or very important. In fact, many firms have chosen to settle with tax authorities to avoid lengthy legal battles that are likely to draw even more adverse media attention and publicity.⁸ Taken together, the risks and costs of getting challenged and overturned weaken the benefits of tax avoidance. Consequently, we expect that managers have incentives to obscure tax avoidance activities to minimize risks and costs of being caught.

2.2 Financial disclosure—linguistic complexity and tone

Financial disclosure is an important means for management to communicate firm performance

⁸ For example, GlaxoSmithKline in 2006 agreed to settle with the IRS for \$3.4 billion—the latest IRS settlement to date, and to abandon its claim seeking a refund of \$1.8 billion in overpaid income taxes. Other large settlements include Merck's \$2.3 billion with the IRS in 2007, AstraZeneca's \$1.1 billion with the IRS in 2011, and Boston Scientific's \$275 million plus interest with the IRS in 2016.

with outside stakeholders. The extant financial disclosure literature shows that managers' reporting and disclosure strategies depend on disclosure rules set by regulators, creditor and investor demand for financial accounting information, the monitoring role of information intermediaries (e.g., auditors and financial analysts), potential litigation costs, contracting and political factors, industry concentration, stock price performance, and corporate governance (see, e.g., reviews by Healy and Palepu 2001; Beyer et al. 2010). All these together shape the corporate information environment.

The content and categories of mandatory disclosures, such as a public company's 10-K report, are highly regulated. As an important dimension, the readability or linguistic complexity of mandatory disclosures is shown to be tied to strategic choices. For instance, Li (2008) finds that the annual reports of firms with lower earnings are more difficult to read, and that firms with annual reports that are easier to read have more persistent positive earnings. Chakrabarty et al. (2018) show that managers with higher risk incentives issue less readable disclosures.

Recent studies suggest that tax authorities acquire public information from mandatory disclosures such as the 10-K reports to corroborate and complement private information gathered from tax returns (Mills and Sansing 2000). Bozanic et al. (2017) document that the IRS downloads the entire 10-Ks in their surveillance effort. By examining categories of information in the 10-K that can be related to the IRS's private information, they find strong association between IRS attention and the amount of detail relating to firms' foreign operations and the richness of narrative detail in the 10-K regarding firms' recent activity. They also find that the IRS downloads increase with the introduction of FIN 48 that requires more contingent tax liability disclosure to investors. In a related study, Hoopes et al. (2012) find a greater likelihood of being audited by the IRS in a given year limits corporate tax avoidance. Therefore, managers of the tax-avoiding firms face a trade-off on whether to provide decision-useful information for outside stakeholders or to conceal

information from the tax authority.

In line with this trade-off view, Balakrishnan et al. (2019) report that tax aggressiveness creates transparency problems, such as larger analysts' forecast errors, greater analysts' forecast dispersion, and a higher level of information asymmetry. To mitigate these transparency problems, tax aggressive firms provide additional disclosure through the Management Discussion & Analysis (MD&A) section of the annual report and conference calls. Adopting textual analysis of financial disclosure, studies find high tax-avoiding firms conceal tax-related information from tax authorities by issuing less readable 10-Ks (Nguyen 2021) and tax footnotes (Inger et al. 2018), partially explaining why tax aggressiveness lessens information transparency.⁹

In addition to readability, ambiguous text in annual reports can be a source of informational risk and interfere with users' ability to comprehend reports. While low readability reflects information complexity, tone ambiguity captures information obfuscation.¹⁰ Ertugrul et al. (2017) document that both readability and tone ambiguity of a firm's financial disclosures are related to managerial information hoarding. We extend tax-related readability studies to linguistic tone.

Loughran and McDonald (2011) show that word classifications derived for non-business disciplines frequently misclassify common words in the financial context. Hence, they compile categories of word lists, including uncertain and weak modal words, to accurately reflect the ambiguous tone of financial disclosures. Weak modal words such as might, possible, and somewhat indicate a lack of confidence. Words that indicate uncertainty, such as approximate, assume, contingent, depend, and indefinite, emphasize imprecision. In tests on the 10-K filing date, Loughran and McDonald (2011) find a positive relation between the use of weak modal and

⁹ Law and Mills (2015) examine the implication of linguistic cues and report that more negative words in annual reports signal tax avoidance.

¹⁰ A key limitation of readability measures is that complex language can simply reflect the provision of complex information instead of obfuscation (Bushee et al. 2018).

uncertain words in corporate 10-K filings and the subsequent stock return volatility. In addition, they link the proportion of weak modal and uncertain words in 10-K filings to negative excess returns and higher abnormal trading volume. Their findings illustrate that ambiguous texts of financial disclosures affect valuation uncertainty.

Ertugrul et al. (2017) discover that firms with a higher proportion of ambiguous words in annual reports face higher costs of bank loans and greater future stock price crash risk. Their findings indicate that the use of vague words of 10-K filings conveys relevant information in assessing a firm's risk level and influences both price and non-price loan terms. Overall, their results provide significant evidence that the ambiguous language of 10-K filings is associated with a firm's information-concealing behavior that increases its information risk and cost of capital.

Managers have an informational advantage over outside regulators and tax authorities. Analogous to protecting a competitive position in product markets, managers of the tax-avoiding firms are incentivized to maintain that information edge and use a weak tone in their mandatory disclosure to increase information processing costs for tax authorities. Therefore, we expect that firms avoiding more taxes provide more ambiguous mandatory disclosure than their counterparts, and propose the following Tax-based Proprietary Information Costs Hypothesis:

H1. Tax avoidance is positively associated with tone ambiguity of mandatory 10-K filings.

Hoopes et al. (2012) find that a greater likelihood of being audited by the IRS in a given year deters firms from pursuing aggressive tax strategies, suggesting that managers perceive IRS audits as costly and burdensome. To mitigate the IRS scrutiny, firms are more likely to withhold disclosure of subsidiaries located in tax haven countries compared to other countries in Exhibit 21 of the Form 10-K (Dyrenge et al. 2020) and their geographic earnings (Hope et al. 2013). Bozanic et al (2017) also document that firms with foreign operations attract more IRS attention. Given

that these firms have higher odds of being flagged by the tax authority for an audit, they have a higher tax-related proprietary cost. If H1 is valid, we expect that managers at these firms face more incentives to strategically use the tone of mandated disclosures to obfuscate financial information and mask tax avoidance activities. Our second hypothesis is developed as follows:

***H2:** The positive association between tax avoidance and tone ambiguity of mandatory 10-K filings is more pronounced for firms with a higher audit probability, with subsidiaries in tax haven countries, or with foreign operations.*

2.3 Disclosure substitution—mandatory vs. voluntary disclosure

Managers have an informational advantage over outside investors regarding firm value and profitability and have competing incentives to disclose information strategically and selectively. Beyer et al. (2010) conduct a decomposition of quarterly stock return variance to study the relative importance of the pillars that shape the accounting-based corporate information environment. The results are surprising. Approximately 66% of the accounting-based information is provided by voluntary disclosures (55% from management forecasts and 21% from earnings pre-announcements). The literature has identified a few motives for managers to provide voluntary disclosures, including the proprietary information costs motive and capital markets transactions motive. Under the proprietary information costs motive, managers' decisions to disclose information are influenced by the concern that such disclosures can damage their competitive position in product markets. Under the capital market transactions motive, managers have incentives to provide voluntary disclosure to mitigate the information asymmetry problem, thereby reducing the firm's cost of external financing (Healy and Palepu 2001). Heile et al. (2022) develop and test a model of voluntary disclosure where managers can choose to withhold (i.e., redact) information from mandatory disclosure. They provide novel evidence that concerns about the

proprietary cost of mandatory disclosure induce managers to redact mandatory disclosure and substitute voluntary disclosure.

One disadvantage of using ambiguous language in 10-K filings is the high cost of capital (Ertugrul et al. 2017) and valuation uncertainty (Loughran and McDonald 2011). The frequent use of weak modal words exacerbates the information asymmetry between managers and outside investors, making it difficult for investors and analysts to forecast income for tax-aggressive firms. A natural alternative to alleviate the adverse consequences of tone ambiguity in mandatory filings is voluntary disclosure, where managers can divulge information strategically with investors. The information conveyed through voluntary disclosure is controlled by managers and could be related to the successful implementation of tax-planning strategies (i.e., shareholders can enjoy a higher after-tax return) or firm performances masked and “downplayed” with ambiguous verbiage in the 10-Ks. Investors demand additional tax information to understand profitability and growth of a firm, as taxes saved are considered a form of profit enhancement (Hanlon et al. 2014).

Lanis et al. (2019) investigate the consequences of tax avoidance on the reputational changes of board directors and CEOs. They find that when firms engage in tax avoidance, both directors and CEOs are rewarded by an increased number of outside board seats, suggesting that tax avoidance improves their reputation. Evidence in prior work shows that firms with a higher level of tax avoidance are more likely to provide tax-related disclosures in the Management Discussion and Analysis (MD&A) section of their 10-Ks (Balakrishnan et al. 2019) and in earnings releases (Schwab 2009). For these reasons, we expect that high tax-avoiding firms will provide more voluntary disclosure than their counterparts to lower external financing costs resulting from the ambiguous disclosure in the 10-Ks. We hence propose our third hypothesis:

H3: Tax-avoiding firms engage in more voluntary disclosure to reduce the cost of information

asymmetry associated with ambiguous mandatory disclosure.

3. Research design

3.1 Sample selection

Our data come from four sources: annual Compustat database, the WRDS SEC Analytics Suite, the CIG files of the First Call Historical Database (FCHD), and the Bog index data¹¹ (Bonsall et al. 2017). To test the relation between tax avoidance and tone ambiguity, we draw firm-level data for publicly listed non-financial and non-utility U.S. firms from the annual Compustat database and the WRDS SEC Analytics Suite. The sample contains 7,924 unique firms representing 55,467 firm-year observations from 1994 to 2017. Missing values of the explanatory variables and tone ambiguity measure reduce the panel used in our baseline model to 25,851 firm-year observations covering 4,221 unique firms.

To examine how tax avoidance is related to voluntary disclosure, we complement the annual Compustat data with management earnings forecast (guidance) data from the CIG files of the First Call Historical Database (FCHD). The merged sample includes 12,380 firm-year observations for 2,711 unique firms between 1997 and 2010. The sample ends in 2010 because First Call ceased to provide guidance data in 2010.

3.2 Variable construction

3.2.1 Measurement of tone ambiguity

Our tone ambiguity measure is the percentage of weak modal words (e.g., depending, might, perhaps, somewhat, and possible) in 10-Ks as defined by Loughran and McDonald (2011). A higher proportion of weak modal words in a financial document indicates a more ambiguous tone and reflects high degrees of uncertainty. Loughran and McDonald (2011) also measure tone

¹¹ Bog index data come from Professor Brian Miller's website <https://kelley.iu.edu/bpm/activities/bogindex.html>.

ambiguity with the percentage of words conveying uncertainty (e.g., approximate, contingency, uncertain, and indefinite). To streamline results, we use the percentage of weak modal words for tone ambiguity in our main results. In robustness checks, we employ the percentage of words conveying uncertainty as an alternative measure of tone ambiguity.¹² Loughran and McDonald (2013) suggest that tone ambiguity creates a weak information environment between a firm and its stakeholders. Bozanic et al. (2017) argue that the IRS takes full advantage of the narrative disclosure in the 10-K, such as the MD&A, Risk Factors, and Business sections, to provide context to the audit process and to complement the private information in tax returns that is mostly quantitative. They further find a strong association between the IRS attention and the richness of narrative detail in the 10-K. Balakrishnan et al. (2019) also find that tax-aggressive firms provide more tax-related disclosures in the Management, Discussion & Analysis (MD&A) section of firms' 10-K reports.¹³ We therefore choose to measure the linguistic tone of the entire narrative disclosure in the 10-K as opposed to only of the tax footnote section.

3.2.2 Measurement of tax avoidance

To capture a broad spectrum of tax avoidance, we adopt a number of tax avoidance proxies that have been suggested in the prior tax literature, including permanent book-tax differences, cash effective tax rates, long-run cash effective tax rates, and industry- and size-matched cash effective tax rates. Following Goh et al. (2016), our first measure is permanent book-tax differences (PBTB). PBTB is defined as the total book-tax differences (BTB) less temporary book-tax differences

¹² The use of uncertain words reflects the general notion of imprecision and risk, whereas the use of weak modal words indicates a lack of confidence (Loughran and McDonald 2011).

¹³ Other sections of firms' 10-Ks (e.g., the Risk Factors section and the Index to Supplementary Data section) could also contain tax-related information.

(TXDI/STR), where TXDI is total deferred tax expense, and STR is statutory tax rate.¹⁴ The permanent book-tax difference is then scaled by lagged total assets.

Our second measure is cash effective tax rate (CETR), which captures tax deferral strategies and is free from possible earnings management. CETR is defined as cash tax paid (TXPD) divided by pre-tax book income (PI) less special items (SPI), following Dyreng et al. (2008). We multiply the variable CETR by -1 so higher values of CETR indicate more tax avoidance.

Since the one-year CETR may not capture long-run tax avoidance, we complement CETR with a long-run cash effective tax rate measure over a three-year period, similar to Dyreng et al. (2008).¹⁵ CETR_3Y is computed as the three-year sum of cash taxes paid (TXPD) (from year $t-2$ to year t) divided by three-year sum of pre-tax income (PI) less special items (SPI). The variable CETR_3Y is multiplied by -1 so higher values of CETR_3Y indicate more tax avoidance.

Our fourth measure of tax aggressiveness is the mean industry- and size-matched CETR minus a firm's CETR (CETR_ADJ) based on Kim and Zhang (2016) and Balakrishnan et al. (2019). The mean industry- and size-matched CETR is the mean CETR for the portfolio of firms in the same quintile of total assets and in the same industry, where size and industry are sorted independently each year and industry is based on the Fama-French 48 industries. A higher value of CETR_ADJ indicates more tax avoidance.

Finally, to mitigate measurement errors and capture all different aspects of tax avoidance, we construct a first principal component of the above four tax avoidance measures (e.g., Kim and Zhang 2016). The higher amount of these measures, the higher level of tax avoidance is.

¹⁴ Total book-tax difference (BTD) is calculated as book income less taxable income, then scaled by lagged total assets (AT). Book income is pre-tax income (PI) in year t . Taxable income is calculated by summing the current federal tax expense (TXFED) and current foreign tax expense (TXFO) and dividing by the statutory tax rate.

¹⁵ The three-year cumulative cash effective tax rate (CETR_3Y) measure also mitigates the effects of volatility in CETR and potential earnings management.

3.3 Regression specification

3.3.1 The relation between tax avoidance and report tone ambiguity

To test H1, we estimate the following baseline model:

$$WEAKMODAL_{i,t} = \beta TAX_AVOIDANCE_{i,t} + \theta' X_{i,t} + \eta_j + \phi_t + \varepsilon_{i,t}, \quad (1)$$

where i , j , and t denote firm, industry and year, respectively. The dependent variable is *WEAKMODAL*, the percentage of weak modal words in 10-Ks (Loughran and McDonald 2011). *TAX_AVOIDANCE* corresponds to the measure of tax avoidance (PBDT, CETR, CETR_3Y, CETR_ADJ, or TA_FACTOR). To test the hypothesis, we focus on the coefficient of *TAX_AVOIDANCE* (β). A positive β indicates tax-avoiding firms are more likely to produce modal weak annual reports. Following prior studies, the vector X represents a set of firm-level controls for the 10-K report disclosure quality used by Li (2008), Lo et al. (2017), and Rushee et al. (2018). These controls include a constant term and market to book (*MTB*), the logarithm of market value of equity ($Ln(MVE)$), firm age (*AGE*), amount of special items (*SPI*), earnings (*EARN*), earnings volatility (*EARNVOL*), stock returns (*SRET*), stock returns volatility (*SRETVOL*), the logarithm of one plus the number of business segments ($Ln(BSEG + 1)$), the logarithm of one plus the number of geographic segments ($Ln(GSEG + 1)$), the logarithm of number of items in Compustat with non-missing values ($Ln(NMITEMS)$), a dummy variable that takes the value of one if a firm is incorporated in Delaware (*DEL*) and zero otherwise, leverage (*LEV*), institutional ownership (*INST*), and a dummy variable for litigation risk that takes the value of one for industries with a high litigation risk and zero otherwise (*LIT*).¹⁶ η_j is industry fixed

¹⁶ In unreported tables, we show that our results are robust to lagged independent variables and the inclusion of earnings management proxied by either total accruals (calculated as ((ACT-lagged ACT) - (CHE- lagged CHE) - (LCT- lagged LCT) + (DLC- lagged DLC) -DP)/lagged AT, where ACT is total current assets, CHE is cash and short-term investments, LCT is total current liabilities, DLC is the current portion of long-term debt, DP is depreciation and amortization, and AT is total assets) or discretionary accruals estimated using the Jones (1991) model, the modified Jones model of Dechow et al. (1995), or the performance-matched model of Kothari et al. (2005). Hadlock and Pierce

effects that capture systematic heterogeneities in financial disclosure practices across industries. ϕ_t refers to year fixed effects that capture any macroeconomic shocks affecting corporate financial disclosures and/or tax avoidance. Detailed variable definitions are presented in Appendix A.

Our H2 predicts that the relation between tax avoidance and tone ambiguity is more pronounced for firms with a high audit probability, having subsidiaries in tax haven countries, or foreign operations. To test H2, we re-estimate our baseline regression (1) by splitting the sample into high- and low-IRS-audit firms, firms with and without tax haven subsidiaries, and firms with and without foreign operations. We measure the probability of an IRS audit (AUDIT_PROB) as lagged *ex post* realizations of actual face-to-face audits scaled by the number of corporate tax returns received, following Hanlon et al. (2014). The measure captures the threat of an IRS audit from a managerial perspective. The data are collected from the TRAC IRS Site. TAX_HAVEN is an indicator variable that equals to one if a firm has at least one subsidiary in a tax haven mentioned in Exhibit 21 of Form 10-K, as defined by Dyreng and Lindsey (2009). MULTI_FIRM is an indicator variable that equals to one if a firm's foreign income (PIFO) is not zero or missing.

3.3.2 The relation between tax avoidance and voluntary disclosure

To test H3, we estimate the following model:

$$VOLDISC_{i,t} = \beta TAX_AVOIDANCE_{i,t} + \theta' X_{i,t-1} + \eta_j + \phi_t + \varepsilon_{i,t}, \quad (2)$$

where VOLDISC is voluntary disclosure. Following the prior literature (e.g., Ball et al. 2012; Baginski et al. 2002), we adopt three measures of voluntary disclosure as the dependent variable: 1) VOLDISC_DUMMY, an indicator variable set equal to one if the firm provides earnings guidance; 2) VOLDISC_FREQ, the number of management forecasts made during a fiscal year

(2010) use a combination of firm size and age to measure financial constraints. Similarly, our baseline specification includes firm size and firm age to capture their effect on tone ambiguity and tax avoidance. But our baseline results remain qualitatively unchanged after explicitly controlling for the index of financial constraints such as the Kaplan and Zingales (1997) index and the Whited and Wu (2006) index.

for the earnings of that fiscal year; and 3) `VOLDISC_HORIZON`, the natural logarithm of one plus the number of days between the management forecast date for a firm's fiscal year earnings and the fiscal period end date. Management earnings forecast (guidance) data come from the CIG files of the First Call Historical Database (FCHD).¹⁷

Following prior studies (Hirst et al. 2008; Chen et al. 2008; Ali et al. 2014; Li and Yang 2016), our control variables include industry concentration (`INDCON`), the natural logarithm of one plus the number of analysts following the stock (`ANALYST`), institutional ownership (`INST`), the natural logarithm of total assets at the fiscal year end (`SIZE`), change in earnings per share from fiscal year $t-1$ to t (`DEARN`), standard deviation of operating earnings over the prior five years (`EARNVOL`), volatility of stock returns (`RVOL`), market-to-book (`MTB`), leverage (`LEV`), equity issuance dummy (`ISS`), and high litigation industry dummy (`LIT`). All control variables are measured at the beginning of the year except for equity issuance and high litigation industry dummy.

3.4 Descriptive statistics

In Panel A of Table 1, we present summary statistics of the key variables employed in our analysis. The average and median percentage of weak modal words (`WEAKMODAL`) are 0.456% and 0.448%, respectively, similar to those in Loughran and McDonald (2011). Descriptive statistics for tax avoidance measures are largely consistent with prior tax avoidance studies (e.g., Dyreng et al. 2008; Goh et al. 2016; Balakrishnan et al. 2019). The descriptive statistics for the control variables and readability measures are largely in line with the previous literature (e.g., Li 2008; Lo et al. 2017; Bonsall et al. 2017). The mean of `VOLDISC_DUMMY` is 0.361, indicating that, on

¹⁷ Management earnings forecasts are an important voluntary disclosure choice (Beyer et al. 2010), similar to alternative voluntary disclosures such as conference calls. Management earnings forecasts are forward-looking disclosures in detailed press releases and often result in significant investor and analyst response because they contain both forecasts of future earnings and value-relevant information such as new products and segment profitability.

average, 36.1% of the sample firms provide at least one management forecast. The average value of 1.274 for VOLDISC_FREQ suggests that, on average, firms make 1.274 management forecasts for the earnings during a fiscal year. VOLDISC_HORIZON (days), the average number of days between the management forecast date for a firm's fiscal year earnings and the fiscal period end date, is 102 days. These figures are consistent with the prior voluntary disclosure literature (see, e.g., Ball et al. 2012; Kim et al. 2018). In Panel B of Table 1, we present correlation matrix for key variables used in our analyses, with Pearson (Spearman) correlations below (above) the diagonal. Notably, the correlations between our tone ambiguity and all tax avoidance measures are significantly positive as predicted. All three proxies for voluntary disclosure are significantly positively correlated with our main tax avoidance measure, TA_FACTOR.

[Table 1 about here]

4. Empirical findings

4.1 Baseline regression results

In this section, we investigate the relation between a firm's tax avoidance and its annual report tone ambiguity. Columns (1) to (4) of Table 2 report the results from pooled OLS regressions after controlling for firm characteristics, industry fixed effects, and year fixed effects.¹⁸ The coefficient estimates on all four tax avoidance proxies are positive and significant. In Column (5), the coefficient estimate on the first principal component of the four tax avoidance measures (TA_FACTOR) is also positive and statistically significant at the 1% level, indicating a positive relation between tax avoidance and tone ambiguity of 10-K filings.¹⁹ The coefficient estimate is

¹⁸ Our baseline result is robust to including industry-year fixed effects that control for industry-specific time trends affecting both firm behavior and disclosure.

¹⁹ In untabulated results, we also find that the positive effect of tax avoidance on tone ambiguity is more pronounced for firms with above industry median tax avoidance, similar to the results found for the effect of tax avoidance on readability in Inger et al. (2018) and Nguyen (2021).

also economically sizable, as a one standard deviation (1.992) increase in a firm's tax avoidance is associated with an increase in tone ambiguity (WEAKMODAL) of 0.006 ($= 0.003 \times 1.992$), which is about 2 percent of the median value of WEAKMODAL.

The sign on the control variables is consistent with prior findings in the 10-K report readability literature (e.g., Li 2008). For instance, growth firms with high market to book ratios face more growth opportunities and uncertainties, and hence they communicate to investors using weaker tone in 10-K filings. Firms with higher stock return volatility and earnings volatility face greater business uncertainty, so they tend to use weaker tone.

[Table 2 about here]

4.2 Impacts of tax-based proprietary information costs

Column (1) of Table 3 reveals that tax-avoiding firms that are subject to more IRS oversight tend to provide more opaque reports using weak modal words. We find no evidence that tax-avoiding firms with low-audit probability are associated with weak tone.

Column (3) of Table 3 shows that the effect of tax avoidance on tone ambiguity is more pronounced in firms with subsidiaries located in tax havens. Column (5) shows the same heightened effect among multinational firms. These results suggest that managers of tax-avoiding firms are cognizant of tax-based proprietary costs of information disclosure in financial reports and that the sensitivity of tone ambiguity to tax avoidance becomes stronger when the risks of being audited and penalized by tax authorities are high. These findings support our inference that managers choose weak tone reporting to lower the potential tax-based proprietary information costs, corroborating our main hypothesis.

[Table 3 about here]

4.3 Tax avoidance and voluntary disclosure

We have thus far shown that tax-avoiding firms are more likely to provide low quality mandatory disclosure due to ambiguous tone. To paint a coherent picture of firms' communication with their stakeholders, we further investigate the effect of tax avoidance on corporate voluntary disclosure.

Table 4 presents results from estimating the association between tax avoidance and voluntary disclosure. Column (1) presents coefficient estimates from a logit regression model for the probability of management issuing earnings forecasts. The coefficient on TA_FACTOR is positive and significant, indicating that firms that avoid taxes are more likely to disclose information to investors voluntarily. This result suggests that managers use voluntary disclosure to mitigate the negative effects of low-quality mandatory financial disclosures on asymmetric information, consistent with the finding that proprietary costs motivate managers to view mandatory and voluntary disclosure as substitutes (Guay et al. 2016; Heinle et al. 2022). Column (3) reports the results of a Tobit regression model for the management earnings forecast frequency. As expected, the coefficient on TA_FACTOR is positive and significant, implying that firms tend to release more information to investors by frequently updating their forecasts as tax avoidance increases. In Column (5), we examine the association between tax avoidance and the horizon of management earnings forecasts. The result suggests that tax avoidance leads to earlier management forecasts, which provide investors with valuable information on a timelier basis.²⁰

In Columns (2), (4) and (6), we include the effect of Reg FD by interacting TA_FACTOR with POST, which is an indicator variable that equals to one after the implementation of Reg FD in 2000. Reg FD was introduced by the Securities and Exchange Commission (SEC) to promote transparency and accountability. Before 2000, publicly traded firms could release important information in meetings and earnings/forecast calls where shareholders and the general public

²⁰ If a firm issues more than one forecast in a year, the date of the earliest one is used.

were excluded. Under Reg FD, any material nonpublic tax-related information that is selectively reported to certain groups of market participants must now be publicly disclosed (Wang 2007). As expected, we find that firms engaging in more tax avoidance tend to increase voluntary disclosure after Reg FD. This evidence complements the finding of Balakrishnan et al. (2019) that managers supplement tax aggressiveness with increased mandatory disclosure via 10-K reports.

[Table 4 about here]

In Table 5, we further test the effect of IRS audit probability on the positive association between tax avoidance and voluntary disclosure. The results show that the positive association is pronounced only when the audit probability is low. This implies that tax-avoiding firms are indeed concerned about potential tax-based proprietary information costs, and the IRS may also pay attention to firms' voluntary disclosures such as periodical press releases which may content both value-relevant and tax-related information.

Overall, our above analyses suggest that tax-avoiding firms opt to voluntarily disclose more frequently and earlier to improve the corporate information environment. This is in line with the idea that tax-avoiding firms choose to reveal more information through voluntary disclosures to reduce their external financing costs. Meanwhile, tax aggressive firms balance the reduced cost of information asymmetry against the increased cost of proprietary information disclosure.

[Table 5 about here]

4.4 Does internal corporate governance affect the way tax-avoiding firms disclose?

In this subsection, we further examine the potential impact of corporate governance via board monitoring on managers' incentives of tax-avoiding firms in manipulating the tone in 10-K disclosures. Tax avoidance can be used for expropriation activities that extend beyond pure tax-saving activities (Desai and Dharmapala 2009; Tang et al. 2017). When tax strategies are intended

to serve the manager's self-interest, managers may be reluctant to provide clarifying disclosure. Prior literature has shown that diversionary tax avoidance more likely occurs for firms with weak corporate governance (Desai and Dharmapala 2009; Tang 2019; Wilson 2009). Accordingly, managers of poorly governed tax-avoiding firms are more motivated to use ambiguous verbiage in the 10-Ks to mask information for self-dealing purposes. Therefore, we expect that stricter internal governance would mitigate diversionary tax avoidance and hence deter tax-avoiding firms from manipulating the tone of 10-K reports. We use board size (BOARD_SIZE), board independence (BOARD_INDEP), and director ownership (DO) to measure the degree of board monitoring and partition our sample according to the sample median value of each internal governance proxy before re-estimating our baseline regression in Equation (1).

4.4.1 Board size

Board involvement in corporate decision-making is an important internal governance mechanism designed to mitigate agency problems.²¹ Following the literature, we define board size (BOARD_SIZE) as the number of directors serving on the board at fiscal year-end. Results in Column (2) of Table 6 show that the positive effect of tax avoidance on tone ambiguity is significant for firms with smaller board size, and the difference in estimated coefficients of TA_FACTOR between the high and low subsamples is significant at the 5% level. This indicates that tax-avoiding firms tend to use weak tone when board monitoring is weak.

4.4.2 Board independence

²¹ Prior research links board size to the effectiveness of board monitoring over management. Larger boards are associated with better board monitoring because of a broader range of expertise and increased likelihood of having independent directors (Klein 2002) and lower earnings management degrees (Xie et al. 2003).

According to the corporate governance literature, greater board independence enhances internal monitoring over managerial rent extraction in protecting shareholder interests.²² Consistent with this idea, in Column (4) of Table 6, we find that the difference in estimated coefficients of TA_FACTOR is significant at the 1% level, suggesting that managers of tax-avoiding firms use less weak tone words to obscure information when board independence is high.

4.4.3 Director ownership

Director ownership (DO) is defined as the proportion of outstanding shares owned collectively by all directors and calculated from data provided in the Institutional Shareholder Services (ISS) Database.²³ Column (6) shows that as tax avoidance increases, firms with greater director ownership are less likely to engage in information hiding activity by manipulating the tone of 10-K reports. In sum, we find that strict internal monitoring deters tax-avoiding firms from using ambiguous tone in financial reports.²⁴

[Table 6 about here]

4.5 Robustness Checks

In this subsection, we conduct a number of robustness checks on the results of our baseline regression model (Model 5 of Table 2) that links tax avoidance to tone ambiguity.

4.5.1 Propensity score matching

We first adopt a nearest-neighbor logit propensity score matching (PSM) methodology developed by Rosenbaum and Rubin (1983) to mitigate the concern with our baseline regression analysis that the observable firm characteristics associated with tax avoidance cause differences in the

²² For instance, Fama and Jensen (1983) show that a high fraction of outside directors serving on the board is positively correlated with corporate governance quality. Agrawal and Knoeber (1996) also argue that the use of outside directors can lead to more effective internal monitoring and find that board independence positively affects corporate value.

²³ Bhagat and Bolton (2013) find a significant positive relationship between corporate governance and director ownership through firm performance.

²⁴ Unlike the internal governance measures, we do not find a significant effect of external governance (e.g., proxied by institutional ownership) on the relation between tax avoidance and tone ambiguity.

association between tax avoidance and tone ambiguity.²⁵ The propensity score matching technique allows us to compare the levels of tone ambiguity of two groups of firms that are similar in terms of several observable characteristics but not in their level of tax avoidance.

We implement the PSM procedure by first creating a dummy variable, High TA_FACTOR, that equals one if TA_FACTOR is higher than the industry-year median, and zero otherwise. We classify observations into treatment (control) group if High TA_FACTOR equals to one (zero). We then estimate propensity scores using a Logit regression model that regresses High TA_FACTOR on the set of control variables used in Table 2. Next, we use the predicted propensity scores of High TA_FACTOR to match (without replacement) each treatment firm with a control firm using the one-to-one closest propensity score with a caliper width of 0.05.²⁶ The procedure yields a sample of 5,265 matched pairs of treatment and control firms.

Panel A of Table 7 shows results of the first-stage Logit regression model and post-match diagnostic regression analysis. We report the results for the full sample in Column (1) and the results for the PSM sample in Column (2). As expected, all of the estimated coefficients for the control variables are insignificant in the PSM sample, suggesting that none of the controls explains the variation in whether tax avoidance affects tone ambiguity. In Panel B, covariate balance tests show that all the mean differences in the variables used to generate the matched sample between the treatment and control firms are statistically insignificant, indicating that the matching is successful. Finally, we re-estimate our baseline specification on the propensity matched sample. Consistent with our main findings in Table 2, a significantly positive coefficient on TA_FACTOR

²⁵ Although we control for a battery of firm characteristics that are identified by the prior literature to relate to tone ambiguity and/or tax avoidance in equation (1), the model specification assumes a linear functional form to control for confounding variables. The propensity score matching (PSM) procedure could also alleviate such an endogeneity concern related to functional form misspecification according to Shipman et al. (2016).

²⁶ Our results are similar if a caliper width of 0.01 is used.

shown in Panel C suggests that tax-aggressive firms issue weaker tone reports than their counterparts after matching for several observed firm characteristics.

[Table 7 about here]

4.5.2 A quasi-natural experiment

One major concern of our baseline regression model is that tax avoidance could be endogenous. Tax avoidance can be influenced by corporate financial and investment policies as well as managerial incentives, which could also affect corporate disclosure decisions. It is also possible that our baseline results arise because tax-avoiding firms are somehow systematically different from their counterparts. That is, the increase in tone ambiguity for tax-avoiding firms that we find may stem from an unobservable factor causing tone to become weaker regardless of the level of tax avoidance. To address these potential endogeneity issues, we conduct a quasi-natural experiment by utilizing 1997 “Check-the-Box” (CTB) regulations on the current effective income tax rates of U.S. multinational firms as an exogenous shock to tax avoidance opportunities, following Balakrishnan et al. (2019) and Nguyen (2021).

The “Check-the-Box” regulations, effective January 1st, 1997, was designed to eliminate the complexity of tax rules and enable firms to choose their organizational form to be treated as a corporation or transparent entity for U.S. tax purposes by making a check-the-box election for a foreign eligible entity. However, this flexibility in entity classification had profound unintended impact on tax planning opportunities. The availability of hybrid entities allowed multinationals to pay very low effective tax rates on their foreign earnings through establishing new foreign legal entities as being disregarded in low-tax jurisdictions, and hence significantly increased tax avoidance (Altshuler and Grubert 2005).

Following Balakrishnan et al. (2019) and Nguyen (2021), our difference-in-differences (DID) sample ranges from 1994 to 2000. We expect that if tax-avoiding firms use weak tone reports to obscure information from tax authorities, then tax-avoiding firms should produce a weaker tone in 10-K fillings than non-tax-avoiding firms after the 1997 issuance of the regulations. We test this conjecture by using the following difference-in-differences (DID) regression model:

$$\begin{aligned} WEAKMODAL_{i,t} = & \beta_1 TREATMENT_i \times POST1997_t + \beta_2 POST1997_t \\ & + \beta_3 TREATMENT_i + \theta' X_{i,t} + \eta_j + \phi_t + \varepsilon_{i,t}, \end{aligned} \quad (3)$$

where *POST1997* is an indicator variable equal to one during the period 1997–2000; *TREATMENT* is a dummy variable equal to one if a firm reported a non-zero pre-tax foreign income (PIFO), and zero otherwise. We expect the difference-in-differences estimate to be positive ($\beta_1 > 0$) and statistically significant. The controls are the same as those used in regression Model (1).

We implement a DID methodology combined with propensity-score matching. First, we use the propensity score matching method to control for observable differences in characteristics between treated firms and control firms. Specifically, we obtain propensity scores by running a Logit regression of *TREATMENT* on the same set of control variables as in Table 2 and measured in the year 1996. We then use the predicted propensity scores of *TREATMENT* to match (without replacement) each treatment firm with a control firm using the one-to-one closest propensity score with a caliper width of 0.05.²⁷ The sample consists of 389 matched pairs of firms. We first present summary statistics for our propensity-score-matched sample in Panel A of Table 8. The figures are similar to those in Panel A of Table 1. Column (1) in Panel B of Table 8 shows our difference-in-differences results of regression Model (3). Our main focus is the coefficient on the interaction

²⁷ Our results are robust to using alternative caliper values of 0.01 and 0.10.

term, $\text{TREATMENT} \times \text{POST1997}$, which captures the difference-in-differences effect on tone ambiguity between treatment and control firms following the adoption of CTB. The coefficient on interaction term is positive and statistically significant ($\beta_1 = 0.037$, $t = 2.22$). The result is also economically significant. Treatment firms on average have an increase in the mean outcome variable (WEAKMODAL) of 0.037 relative to control firms following the implementation of CTB.

In Column (2), we verify the parallel trends assumption of the DID regression model. We replace the POST1997 dummy with three dummies: 1) YEAR1996, a dummy variable equal to one for the year 1996 (the year before CTB), and zero otherwise; 2) YEAR1997, a dummy variable equal to one for the year 1997 (the year of CTB), and zero otherwise; and 3) POST1998, a dummy variable equal to one for the period after CTB, and zero otherwise. As shown in Column (2), the coefficient of $\text{TREATMENT} \times \text{YEAR1996}$ is not statistically significant, while the coefficients of $\text{TREATMENT} \times \text{YEAR1997}$ and $\text{TREATMENT} \times \text{POST1998}$ are statistically significant and positive. This indicates that there is no evidence of differential pre-treatment trends in Tone Ambiguity (WEAKMODAL) between treatment firms and control firms, consistent with the parallel-trends assumption. It also alleviates the concern that our DID result is driven by alternative confounding factors. Putting together, the DID regression results corroborate our earlier findings that tone ambiguity is positively associated with tax avoidance.

[Table 8 about here]

4.5.3 Additional robustness checks

In this subsection, we perform additional robustness checks to confirm the validity of the main findings. First, we control for the readability of narrative disclosures in 10-K reports because previous studies document a negative association between tax avoidance and readability. Second, we control for the effect of uncertain tax benefits, which are related to tax avoidance and

uncertainty. Third, we examine the effect of conforming tax avoidance, which is a type of tax planning with the lowest audit risks. Last, we alternatively measure tone ambiguity with the percentage of words conveying uncertainty as in Loughran and McDonald (2011).

Parallel to our tax-based proprietary information costs explanation, Inger et al. (2018) and Nguyen (2021) suggest that managers have incentives to make reports less readable so that it would be more difficult for tax authorities to process value-relevant financial information and detect tax avoidance activities.

To control for the readability of 10-K reports, we adopt three measures of report readability: BOG, FOG, and FSIZE. BOG is the Bog index, based on Bonsall et al. (2017). FOG is the Gunning fog index, which is defined as 0.4 times the sum of average number of words per sentence and the percentage of 10-K complex words. FSIZE is the natural logarithm of the 10-K file size in megabytes of the SEC EDGAR, based on Loughran and McDonald (2014). A high value of the readability proxy implies less readable text. Columns (1) to (3) of Table 9 show that TA_FACTOR remains positive and statistically significant in its association with tone ambiguity after controlling for readability of the 10-K reports.²⁸

Tax literature shows that uncertain tax positions capture corporate tax shelter activities (Lisowsky et al. 2013) and are positively related to tax avoidance (Dyreng et al. 2019). Since tone ambiguity also reflects uncertainty (Loughran and McDonald 2011), it is possible that the positive association between tone ambiguity and tax avoidance is simply due to the uncertain tax positions taken by firms. In Column (4), we address this concern by controlling for uncertain tax benefits (UTB), calculated as a firm's uncertain tax benefits at the end of a fiscal year (TXTUBEND) scaled

²⁸ The Pearson (Spearman) correlation coefficients between WEAKMODAL and readability measures (BOG, FOG, and FSIZE) are 0.37 (0.39), 0.25 (0.30) and 0.50 (0.55), respectively, all significant at the 1% level.

by total assets (AT) at the fiscal year-end. We find UTB is significantly positively related to tone ambiguity, and the positive and significant effect of TA_FACTOR on tone ambiguity still holds.

Different types of tax avoidance may have different tax risks. Conforming tax avoidance increases or decreases book and tax income simultaneously, which has the lowest tax risk and is hard to be detected. Therefore, firms that engage in such tax avoidance should have fewer incentives to opaque information than their counterparts. Following Badertscher et al. (2019), we measure CONFORM_TAX as the residual of an OLS regression estimated by Fama-French 48 industry and fiscal year combinations.²⁹ The result in Column (5) of Table 9 reveals that conforming tax-avoiding firms don't issue reports with a less ambiguous tone.

Last, we follow Loughran and McDonald (2011) to measure tone ambiguity with the percentage of words conveying uncertainty (e.g., approximate, contingency, uncertain, and indefinite). In Column (6) of Table 9, this alternative measure generates the same qualitative results as our baseline measure.

[Table 9 about here]

5. Conclusion

Corporate tax avoidance is an activity that is traditionally welcomed by shareholders but scrutinized by other stakeholders. In this paper, we empirically examine how the disclosure trade-off between the benefits of reducing information asymmetry and the tax-base propriety costs of information revealed to tax authorities affects tax-avoiding firms' disclosure strategies. First, we find that tax-avoiding firms issue more ambiguous mandatory 10-K filings to mitigate costs of being audited and challenged by tax authorities). Second, we find that holding readability constant,

²⁹ Specifically, we regress the ratio of cash taxes paid to lagged total assets on total book-tax differences (BTD), a dummy variable (NEG) that equals to one for observations with negative book-tax differences (and zero otherwise), the interaction of BTD and NEG, the level of net operating loss carryforwards (NOL), and changes in NOLs.

the relation between tax avoidance and tone ambiguity remains positive. These results suggest that the low information transparency of tax-avoiding firms can be attributed to the intentional use of linguistic complexity and obfuscation in mandatory disclosure. Third, we provide evidence that tax-avoiding firms reveal more information through voluntary disclosure to reduce external financing costs associated with information asymmetry caused by tone ambiguity. However, the disclosure substitution is conditional on tax-based propriety information costs. Therefore, we paint a comprehensive picture of managers' disclosure strategies in the setting of tax avoidance.

Our study links two timely and crucial topics together (i.e., tax avoidance and the quality of information environment). We complement and extend the emerging textual studies on tax avoidance and the literature on disclosure substitution. Understanding the textual information in corporate disclosures is important as qualitative disclosures have informational value beyond traditional quantitative measures. Our findings will be of interest to researchers and investors that seek to better understand how firms trade off the economic benefits from tax avoidance and the potential costs associated with investors having difficulty understanding the financial implications of those choices as well as the potential tax-based propriety information costs.

The finding that board monitoring deters managers from using weak tone reporting to obscure tax avoidance informs the regulators and shareholders of the implication of an effective internal corporate governance mechanism in stakeholder-manager incentive alignment for tax-avoiding firms. Overall, our study provides additional insights into the roles of managerial incentives and monitoring mechanisms in shaping the corporate financial information environment in the context of tax avoidance.

References

- Agrawal, A., & Knoeber, C. R. (1996). Firm performance and mechanisms to control agency problems between managers and shareholders. *Journal of Financial and Quantitative Analysis* 31(3), 377–397.
- Ali, A., Klasa, S., & Yeung, E. (2014). Industry concentration and corporate disclosure policy. *Journal of Accounting and Economics* 58(2–3), 240–264.
- Altshuler, R., & Grubert, H. (2005). The three parties in the race to the bottom: Host governments, home governments and multinational companies. *Florida Tax Review* 7(3), 137–209.
- Badertscher, B., Katz, S., Rego, S., & Wilson, R. (2019). Conforming tax avoidance and capital market pressure. *Accounting Review* 94(6), 1–30.
- Baginski, S. P., Hassell, J. M., & Kimbrough, M. D. (2002). The effect of legal environment on voluntary disclosure: Evidence from management earnings forecasts issued in U.S. and Canadian markets. *Accounting Review* 77(1), 25–50.
- Balakrishnan, K., Blouin, J. L., & Guay, W. R. (2019). Tax aggressiveness and corporate transparency. *Accounting Review* 94(1), 45–69.
- Ball, R., Jayaraman, S., & Shivakumar, L. (2012). Audited financial reporting and voluntary disclosure as complements: A test of the confirmation hypothesis. *Journal of Accounting and Economics* 53(1–2), 136–166.
- Beyer, A., Cohen, D. A., Lys, T. Z., & Walther, B. R. (2010). The financial reporting environment: Review of the recent literature. *Journal of Accounting and Economics* 50(2–3), 296–343.
- Bhagat, S., & Bolton, B. (2013). Director ownership, governance, and performance. *Journal of Financial and Quantitative Analysis* 48(1), 105–135.
- Bonsall, S. B., Leone, A. J., Miller, B. P., & Rennekamp, K. (2017). A plain English measure of financial reporting readability. *Journal of Accounting and Economics* 63(2–3), 329–357.
- Bozanic, Z., Hoopes, J. L., Thornock, J. R., & Williams, B. (2017). IRS attention. *Journal of Accounting Research* 55(1), 79–114.
- Bushee, B. J., Gow, I. D., & Taylor, Daniel J. (2018). Linguistic complexity in firm disclosures: Obfuscation or information? *Journal of Accounting Research* 56(1), 85–121.
- Chakrabarty, B., Seetharaman, A., Swanson, Z., & Wang, X. F. (2018). Management risk incentives and the readability of corporate disclosures. *Financial Management* 47(3), 583–616.

- Chen, S., Chen, X., & Cheng, Q. (2008). Do family firms provide more or less voluntary disclosure? *Journal of Accounting Research* 46(3), 499–536.
- Chen, H., Noronha, G., & Singal, V. (2004). The price response to S&P 500 index additions and deletions: Evidence of asymmetry and a new explanation. *Journal of Finance* 59(4), 1901–1929.
- Chen, T., & Lin, C. (2017). Does information asymmetry affect corporate tax aggressiveness? *Journal of Financial and Quantitative Analysis* 52(5), 2053–2081.
- Dechow, P. M., Sloan, R. G., & Sweeney, A. P. (1995). Detecting earnings management. *Accounting Review* 70(2), 193–225.
- Desai, M. A., & Dharmapala, D. (2009). Corporate tax avoidance and firm value. *Review of Economics and Statistics* 91(3), 537–546.
- Dyreng, S. D., Hanlon, M., & Maydew, E. L. (2008). Long-run corporate tax avoidance. *Accounting Review* 83(1), 61–82.
- Dyreng, S. D., Hanlon, M., & Maydew, E. L. (2019). When does tax avoidance result in tax uncertainty? *Accounting Review* 94(2), 179–203.
- Dyreng, S. D., Hoopes, J., Langetieg, P., & Wilde, J. H. (2020). Strategic subsidiary disclosure. *Journal of Accounting Research* 58(3), 643–692.
- Dyreng, S. D., & Lindsey, B. P. (2009). Using financial accounting data to examine the effect of foreign operations located in tax havens and other countries on US multinational firms' tax rates. *Journal of Accounting Research* 47(5), 1283–1316.
- Dyreng, S. D., Hoopes, J. L., & Wilde, J. H. (2016). Public pressure and corporate tax behavior. *Journal of Accounting Research* 54(1), 147–186.
- Edwards, A., Schwab, C., & Shevlin, T. (2016). Financial constraints and cash tax savings. *Accounting Review* 91(3), 859–881.
- ERNST & YOUNG. (2011). 2011–12 Tax risk and controversy survey.
- Ertugrul, M., Lei, J., Qiu, J., & Wan, C. (2017). Annual report readability, tone ambiguity, and the cost of borrowing. *Journal of Financial and Quantitative Analysis* 52(2), 811–836.
- Fama, E. F., & Jensen, M. C. (1983). Separation of ownership and control. *Journal of Law and Economics* 26(2), 301–325.
- Friberg, R., & Seiler, T. (2017). Risk and ambiguity in 10-Ks: An examination of cash holding and derivatives use. *Journal of Corporate Finance* 45(4), 608–631.

- Gallemore, J., & Labro, E. (2015). The importance of the internal information environment for tax avoidance. *Journal of Accounting and Economics* 60(1), 149–167.
- Graham, J. R., Hanlon, M., Shevlin, T., & Shroff, N. (2014). Incentives for tax planning and avoidance: Evidence from the field. *Accounting Review* 89(3), 991–1023.
- Goh, B. W., Lee, J., Lim, C. Y., & Shevlin, T. (2016). The effect of corporate tax avoidance on the cost of equity. *Accounting Review* 91(6), 1647–1670.
- Guay, W., Samuels, D., & Taylor, D. (2016). Guiding through the Fog: Financial statement complexity and voluntary disclosure. *Journal of Accounting and Economics* 62(2–3), 234–269.
- Hadlock, C., & Pierce, J. (2010). New evidence on measuring financial constraints: Moving beyond the KZ Index. *Review of Financial Studies* 23(5), 1909–1940.
- Hanlon, M., and Heitzman, S. (2010). A review of tax research. *Journal of Accounting and Economics* 50(2–3), 127–178.
- Hanlon, M., Hoopes, J. L., & Shroff, N. (2014). The effect of tax authority monitoring and enforcement on financial reporting quality. *Journal of the American Taxation Association* 36(2), 137–170.
- Hanlon, M., & Slemrod, J. (2009). What does tax aggressiveness signal? Evidence from stock price reactions to news about tax shelter involvement. *Journal of Public Economics* 93(1–2), 126–141.
- Healy, P. M., & Palepu, K. G. (2001). Information asymmetry, corporate disclosure, and the capital markets: A review of the empirical disclosure literature. *Journal of Accounting and Economics* 31(1–3), 405–440.
- Heinle, M. S., Samuels, D., & Taylor, D. J. (2022). Disclosure substitution. *Management Science*, forthcoming.
- Hirst, D. E., Koonce, L., & Venkataraman, S. (2008). Management earnings forecasts: A review and framework. *Accounting Horizons* 22(3), 315–338.
- Hoopes, J. L., Mescall, D., & Pittman, J. A. (2012). Do IRS audits deter corporate tax avoidance? *Accounting Review* 87(5), 1603–1639.
- Hoopes, J. L., Robinson, L. A., & Slemrod, J. B. (2018). Public tax-return disclosure. *Journal of Accounting and Economics* 66(1), 142–162.

- Hope, O., Ma, M., & Thomas, W. B. (2013). Tax avoidance and geographic earnings disclosure. *Journal of Accounting and Economics* 56(2–3), 170–189.
- Inger, K. K., Meckfessel, M. D., Zhou, M., & Fan, W. (2018). An examination of the impact of tax avoidance on the readability of tax footnotes. *Journal of the American Taxation Association* 40(1), 1–29.
- Jones, J. J. (1991). Earnings management during import relief investigations. *Journal of Accounting Research* 29(2), 193–228.
- Kaplan, S., & Zingales, L. (1997). Do investment-cash flow sensitivities provide useful measures of financing constraints? *Quarterly Journal of Economics* 112(1), 169–215.
- Kim, J. B., Shroff, P. K., Vyas, D., & Wittenberg-Moerman, R. (2018). Credit default swaps and managers' voluntary disclosure. *Journal of Accounting Research* 56(3), 953–988.
- Kim, J., & Zhang, L. (2016). Accounting conservatism and stock price crash risk: firm-level evidence. *Contemporary Accounting Research* 33(1), 412–441.
- Klein, A., (2002). Economic determinants of audit committee independence. *Accounting Review* 77(2), 435–452.
- Kothari, S. P., Leone, A., & Wasley, C. E. (2005). Performance matched discretionary accrual measures. *Journal of Accounting and Economics* 39(1), 163–197.
- Lanis, R., Richardson, G., Liu, C., & McClure, R. (2019). The impact of corporate tax avoidance on board of directors and CEO reputation. *Journal of Business Ethics* 160(2), 463–498.
- Law, K. K. F., & Mills, L. F. (2015). Taxes and financial constraints: Evidence from linguistic cues. *Journal of Accounting Research* 53(4), 777–819.
- Li, F. (2008). Annual report readability, current earnings, and earnings persistence. *Journal of Accounting and Economics* 45(2–3), 221–247.
- Li, X., & Yang, H. I. (2016). Mandatory financial reporting and voluntary disclosure: The effect of mandatory IFRS adoption on management forecasts. *Accounting Review* 91(3), 933–953.
- Lisowsky, P., Robinson, L., & Schmidt, A. (2013). Do publicly disclosed tax reserves tell us about privately disclosed tax shelter activity? *Journal of Accounting Research* 51(3), 583–629.
- Lo, K., Ramos, F., & Rogo, R. (2017). Earnings management and annual report readability. *Journal of Accounting and Economics* 63(1), 1–25.

- Loughran, T., & McDonald, B. (2011). When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. *Journal of Finance* 66(1), 35–65.
- Loughran, T., & McDonald, B. (2013). IPO first-day returns, offer price revisions, volatility, and form S-1 language. *Journal of Financial Economics* 109(2), 307–326.
- Loughran, T., & McDonald, B. (2014). Measuring readability in financial disclosures. *Journal of Finance* 69(4), 1643–1671.
- Loughran, T., & McDonald, B. (2016). Textual analysis in accounting and finance: A survey. *Journal of Accounting Research* 54(4), 187–230.
- Mills, L. F. (1998). Book-tax differences and Internal Revenue Service adjustments. *Journal of Accounting Research* 36(2), 343–56.
- Mills, L. F., & Sansing, R. C. (2000). Strategic tax and financial reporting decisions: Theory and evidence. *Contemporary Accounting Research* 17(1), 85–106.
- Nguyen, J. (2021). Tax avoidance and financial statement readability. *European Accounting Review* 30(5), 1043–1066.
- Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika* 70(1), 41–55.
- Schwab, C. M. (2009). The determinants and effects of voluntary book-tax difference disclosures: Evidence from earnings press releases. Working Paper. University of Texas at Austin.
- Shipman, J. E., Swanquist, Q. T., & Whited, R. L. (2017). Propensity score matching in accounting research. *Accounting Review* 92(1), 213–244.
- Tang, T. Y. H. (2019). The value implications of tax avoidance across countries. *Journal of Accounting, Auditing and Finance* 34(4), 615–638.
- Tang, T. Y. H., Mo, P. L. L., & Chan, K. H. (2017). Tax collector or tax avoider? An investigation of intergovernmental agency conflicts. *Accounting Review* 92(2), 247–270.
- Xie, B., Davidson, W., & DaDalt, P. J. (2003). Earnings management and corporate governance: the role of the board and the audit committee. *Journal of Corporate Finance* 9(3), 295–316.
- Wang, I. Y. (2007). Private earnings guidance and its implications for disclosure regulation. *Accounting Review* 82(5), 1299–1332.
- Whited, T.M., & Wu, G. (2006). Financial constraints risk. *Review of Financial Studies* 19(2), 531–559.

Wilson, R. J. (2009). An examination of corporate tax shelter participants. *Accounting Review* 84(3), 969–999.

Appendix A: Variable definitions

This table provides the definition of the variables used in the study.

Definitions with corresponding Compustat item names	
Tone ambiguity measures	
WEAKMODAL	The percentage of weak modal words as defined by Loughran and McDonald (2011). The value is multiplied by 100.
UNCERTAIN	The percentage of words conveying uncertainty as defined by Loughran and McDonald (2011). The value is multiplied by 100.
Tax avoidance measures	
PBTD	Total book-tax differences (BTD) less temporary book-tax differences (TXDI/STR), where TXDI is total deferred tax expense and STR is statutory marginal tax rate. The permanent book-tax difference is then scaled by lagged total assets. Based on Goh et al. (2016). Total book-tax difference (BTD), calculated as book income less taxable income, then scaled by lagged total assets (AT). Book income is pre-tax income (PI) in year t . Taxable income is calculated by summing the current federal tax expense (TXFED) and current foreign tax expense (TXFO) and dividing by the statutory tax rate. Based on Frank et al. (2009) and Goh et al. (2016).
CETR	Cash effective tax rate, measured as cash tax paid (TXPD) divided by pre-tax book income (PI) less special items (SPI). Cash effective tax rate is set as missing when the denominator is zero or negative. We remove observations with missing cash taxes paid (TXPD) or negative pre-tax book income ($PI < 0$). We truncate cash effective tax rate to the range $[0, 1]$. Based on Dyreng et al. (2008). The variable is multiplied by -1 so higher values of CETR indicate more tax avoidance.
CETR_3Y	Three-year cumulative cash effective tax rate, computed as the three-year sum of cash taxes paid (TXPD) (from year $t-2$ to year t) divided by three-year sum of pre-tax income (PI) less special items (SPI). The variable is multiplied by -1 so higher values of CETR indicate more tax avoidance. We require the three-year cumulative pre-tax book income less special items to be positive. Based on Dyreng et al. (2008).
CETR_ADJ	Adjusted CETR, measured as mean industry- and size-matched Cash ETR minus a firm's Cash ETR. The mean industry- and size-matched Cash ETR is the mean Cash ETR for the portfolio of firms in the same quintile of total assets and in the same industry, where size and industry are sorted independently each year and industry is based on the Fama-French 48 industries. Note that a higher value of adjusted Cash ETR indicates more tax avoidance. Based on Kim and Zhang (2016) and Balakrishnan et al. (2019).
TA_FACTOR	The first principal component of the above four tax avoidance measures (see, e.g., Kim and Zhang 2016).
UTB	Uncertain tax benefits, calculated as a firm's uncertain tax benefits at the end of a fiscal year (TXTUBEND) scaled by total assets (AT) at

	the fiscal year-end.
CONFORM_TAX	Conforming tax avoidance, as in Badertscher et al. (2019), calculated as the residual of an OLS regression estimated by Fama-French 48 industry and fiscal year combinations. Specifically, the ratio of cash taxes paid to lagged total assets is regressed on total book-tax differences (BTD), a dummy variable (NEG) that equals to one for observations with negative book-tax differences (and zero otherwise), the interaction of BTD and NEG, the level of net operating loss carryforwards (NOL), and changes in NOLs. Conforming tax avoidance removes the impact of nonconforming tax strategies.

Firm-level control variables

MTB	Market value of equity ($PRCC_F \times CSHO$) plus book value of liability (LT), divided by total assets (AT) at the fiscal year-end.
MVE	Market value of equity ($PRCC_F \times CSHO$) at the fiscal year-end.
AGE	Number of years since a firm first appears in the CRSP monthly stock return file.
SPI	Amount of special items (SPI) divided by total assets (AT).
EARN	Operating earnings (IB) divided by total assets (AT).
EARNVOL	Standard deviation of operating earnings (IB) during the prior five years.
SRET	The stock return during the previous fiscal year.
SRETVOL	Standard deviation of monthly stock returns in the prior year.
BSEG	Number of business segments from the Compustat segment files at the fiscal year-end.
GSEG	Number of geographic segments from the Compustat segment files at the fiscal year-end.
NMITEMS	Number of items in Compustat with non-missing values.
DEL	A dummy variable that takes the value of 1 if a firm is incorporated in Delaware; 0 otherwise.
LEV	Total liabilities (DLTT+DLC) divided by total assets (AT) at the fiscal year end.
INST	Institutional ownership measured as the fraction of a firm's outstanding shares owned by institutional investors; calculated from data provided by Thomson-Reuter's Institutional Holdings (13F) Database.
LIT	A high litigation risk indicator, coded as one for industries with a high litigation risk (Standard Industrial Classification (SIC) codes 2833–2836, 3570–3577, 7370–7374, 3600–3674, 5200–5961, 8731–8734), and zero otherwise.
AUDIT_PROB	Probability of an IRS audit, measured by the <i>ex post</i> realizations of actual face-to-face audits divided by the number of corporate tax returns received, following Hanlon et al. (2014). The data are collected from the TRAC IRS Site.
TAX_HAVEN	An indicator variable that equals to one if a firm has at least one subsidiary in a tax haven mentioned in Exhibit 21 of Form 10-K, as

	defined by Dyreng and Lindsey (2009).
MULTI_FIRM	An indicator variable that equals to one if a firm's foreign income (PIFO) is not zero or missing.
BOARD_SIZE	Number of directors on the board; calculated from data provided in the Institutional Shareholder Services (ISS) Database.
BOARD_INDEP	Percent outside directors is the percentage of outside directors on the board; calculated from data provided in the Institutional Shareholder Services (ISS) Database.
DO	Proportion of outstanding shares owned collectively by all directors; calculated from data provided in the Institutional Shareholder Services (ISS) Database.
INDCON	Industry concentration measured as the sales-based Herfindahl-Hirschman Index (HHI) at the three-digit SIC code industry level.
ANALYST	Number of financial analysts covering the firm. The data come from IBES Summary file.
SIZE	Natural logarithm of total assets (AT) at the fiscal year end.
DEARN	Change in earnings per share from fiscal year $t-1$ to t , scaled by the stock price at the fiscal year end.
RVOL	Return volatility, measured as the standard deviation of daily stock returns.
ISS	An indicator variable that takes a value of one if the firm issues equity during the forecast year, and zero otherwise.

Readability measures

BOG	The Bog Index, based on Bonsall et al. (2017). Bog Index = Sentence Bog + Word Bog – Pep. Sentence Bog refers to sentence length. Word Bog consists of: (1) plain English style problems and (2) word difficulty. Pep identifies writing attributes that facilitate understanding of texts by readers. A higher Bog Index implies a less readable document.
FOG	The Gunning fog index, which is defined as 0.4 times the sum of average number of words per sentence and the percentage of 10-K complex words.
FSIZE	Natural logarithm of the 10-K file size in megabytes of the SEC EDGAR, based on Loughran and McDonald (2014).

Voluntary disclosure measures

VOLDISC_DUMMY	An indicator variable set equal to one if the firm provides earnings guidance, and zero otherwise. Management earnings forecast (guidance) data from the CIG files of the First Call Historical Database (FCHD).
VOLDISC_FREQ	Number of management forecasts made for the earnings during a fiscal year.
VOLDISC_HORIZON	Natural logarithm of one plus the number of days between the management forecast date for a firm's fiscal year earnings and the fiscal period end date.

Appendix B: Ambiguity word list from Loughran and McDonald (2011)

ABEYANCE	IMPRECISION	SELDOM	UNFORSEEN
ABEYANCES	IMPRECISIONS	SELDOMLY	UNGUARANTEED
AMBIGUITIES	INCOMPLETENESS	SPECULATE	UNIDENTIFIABLE
AMBIGUITY	INDEFINITE	SPECULATED	UNIDENTIFIED
AMBIGUOUS	INDEFINITENESS	SPECULATES	UNKNOWN
ANOMALIES	INDETERMINABLE	SPECULATING	UNKNOWNNS
ANOMALOUS	INDETERMINATE	SPECULATION	UNOBSERVABLE
ANOMALOUSLY	INEXACT	SPECULATIONS	UNPLANNED
ANOMALY	INEXACTNESS	SPECULATIVE	UNPREDICTABILITY
ARBITRARILY	MAYBE	SPECULATIVELY	UNPREDICTABLE
ARBITRARINESS	MIGHT	SPORADIC	UNPREDICTABLY
ARBITRARY	NONASSESSABLE	SPORADICALLY	UNPREDICTED
BELIEVE	PERHAPS	SUDDEN	UNPROVED
BELIEVED	PRECAUTION	SUDDENLY	UNPROVEN
BELIEVES	PRECAUTIONARY	UNCERTAIN	UNQUANTIFIABLE
BELIEVING	PRECAUTIONS	UNCERTAINLY	UNQUANTIFIED
CAUTIOUS	PRESUMABLY	UNCERTAINTIES	UNRECONCILED
CAUTIOUSLY	PRESUME	UNCERTAINTY	UNSEASONABLE
CAUTIOUSNESS	PRESUMED	UNCLEAR	UNSEASONABLY
CONCEIVABLE	PRESUMES	UNCONFIRMED	UNSPECIFIC
CONCEIVABLY	PRESUMING	UNDECIDED	UNSPECIFIED
CONFUSES	PRESUMPTION	UNDEFINED	UNTESTED
CONFUSING	PRESUMPTIONS	UNDESIGNATED	UNUSUAL
CONFUSINGLY	REINTERPRET	UNDETECTABLE	UNUSUALLY
CONFUSION	REINTERPRETATION	UNDETERMINABLE	VAGARIES
COULD	REINTERPRETATION	UNDETERMINED	VAGUE
DOUBT	REINTERPRETED	UNDOCUMENTED	VAGUELY
DOUBTED	REINTERPRETING	UNEXPECTED	VAGUENESS
DOUBTFUL	REINTERPRETS	UNEXPECTEDLY	VAGUENESSES
DOUBTS	REVISE	UNFAMILIAR	VAGUER
HIDDEN	REVISED	UNFAMILIARITY	VAGUEST
IMPRECISE	RUMORS	UNFORECASTED	

Table 1. Descriptive statistics and correlation matrix

This table presents descriptive statistics and correlation matrix of key variables of interest for the sample of firms included in our study. The sample covers firm-year observations with non-missing values for all control variables from 1994–2017. Panel A provides mean, standard deviations, three quartiles, and the number of observations of variables employed in the analysis. Panel B presents Pearson correlations of key variables in the lower triangle and Spearman correlations in the upper triangle. The p -values are below the correlation coefficients. All variables are defined in the Appendix. Data on tax avoidance measures and firm characteristics are collected from the merged Compustat-CRSP database for 1994–2017. Annual report readability data come from <https://kelley.iu.edu/bpm/activities/bogindex.html> for the period 1994–2017. Management earnings forecast (guidance) data from the CIG files of the First Call Historical Database (FCHD). Variables are winsorized at the 1% and 99% levels. Details on the construction of all variables are provided in Appendix A.

<i>Panel A. Descriptive Statistics</i>						
Variable	Mean	Median	Std. Dev.	25 th percentile	75 th percentile	<i>N</i>
Tone ambiguity measure						
WEAKMODAL	0.456	0.448	0.231	0.354	0.690	25,851
Tax avoidance measures						
PBTD	0.022	0.011	0.048	0.001	0.028	25,851
CETR	-0.271	-0.260	0.183	-0.355	-0.148	25,851
CETR_3Y	-0.274	-0.271	0.158	-0.348	-0.177	25,851
CETR_ADJ	-0.008	0.000	0.163	-0.084	0.097	25,851
TA_FACTOR	0.043	0.078	1.992	-0.889	1.108	25,851
Firm characteristics						
MTB	1.997	1.627	1.212	1.226	2.317	25,851
MVE	4,932	769	14,432	199	2,795	25,851
AGE	24.274	19.000	17.744	11.000	33.000	25,851
SPI	-0.005	0.000	0.015	-0.007	0.000	25,851
EARN	0.159	0.146	0.075	0.108	0.197	25,851
EARNVOL	0.016	0.012	0.014	0.007	0.020	25,851
SRET	0.143	0.069	0.491	-0.157	0.332	25,851
SRETVOL	0.246	0.206	0.161	0.154	0.284	25,851
BSEG	3.112	3.000	1.384	2.000	4.000	25,851
GSEG	3.297	3.000	1.912	2.000	4.000	25,851
NMITEMS	318	331	43	279	352	25,851
DEL	0.617	1.000	0.486	0.000	1.000	25,851
LEV	0.188	0.165	0.169	0.018	0.304	25,851
INST	0.608	0.666	0.293	0.385	0.848	25,851
LIT	0.293	0.000	0.455	0.000	1.000	25,851
Voluntary disclosure measures						

VOLDISC_DUMMY	0.361	0.000	0.480	0.000	1.000	12,380
VOLDISC_FREQ	1.274	0.000	2.091	0.000	2.000	12,380
VOLDISC_HORIZON (days)	102	0.000	145	0.000	299	12,380
Voluntary disclosure controls						
INDCON	0.166	0.130	0.133	0.078	0.206	12,380
ANALYST	9.208	7.000	8.899	2.000	13.000	12,380
INST	0.581	0.628	0.279	0.368	0.808	12,380
SIZE	6.407	6.361	1.743	5.180	7.581	12,380
DEARN	0.017	0.008	0.063	-0.007	0.024	12,380
EARNVOL	0.041	0.026	0.047	0.015	0.048	12,380
RVOL	0.122	0.108	0.062	0.077	0.151	12,380
MTB	2.025	1.644	1.245	1.226	2.373	12,380
LEV	0.173	0.151	0.158	0.014	0.284	12,380
ISS	0.035	0.000	0.184	0.000	0.000	12,380
LIT	0.305	0.000	0.460	0.000	1.000	12,380
Report readability measures						
BOG	82.589	83.000	7.238	78.000	88.000	25,176
FOG	19.779	19.725	1.131	19.061	20.416	25,850
FSIZE	5.053	1.290	7.728	0.350	6.590	25,851
Alternative tax avoidance measures						
UTB	0.009	0.005	0.012	0.001	0.012	9,864
CONFORM_TAX	-0.001	-0.006	0.029	-0.020	0.012	23,078
Alternative tone ambiguity measure						
UNCERTAIN	1.052	1.265	0.349	0.991	1.510	25,850

<i>Panel B. Pearson/Spearman Correlation Matrix</i>															
Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
1 WEAKMODAL		0.13	0.14	0.17	0.05	0.13	0.00	-0.01	0.83	0.39	0.30	0.55	0.08	0.10	0.09
		0.00	0.00	0.00	0.00	0.00	0.85	0.21	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2 PBTD	0.11		0.14	0.16	0.09	0.19	0.19	0.18	0.13	0.09	0.02	0.12	0.03	0.04	0.04
	0.00		0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00
3 CETR	0.11	0.21		0.74	0.86	0.95	0.07	-0.49	0.14	0.18	0.10	0.13	0.04	0.05	0.05
	0.00	0.00		0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
4 CETR_3Y	0.14	0.20	0.63		0.61	0.85	0.09	-0.36	0.17	0.22	0.12	0.17	0.08	0.08	0.09
	0.00	0.00	0.00		0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
5 CETR_ADJ	0.04	0.18	0.90	0.55		0.90	0.05	-0.48	0.04	0.07	0.04	0.01	0.03	0.03	0.03
	0.00	0.00	0.00	0.00		0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.01	0.00	0.00
6 TA_FACTOR	0.11	0.29	0.95	0.79	0.92		0.08	-0.46	0.13	0.16	0.09	0.11	0.05	0.06	0.06
	0.00	0.00	0.00	0.00	0.00		0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
7 UTB	0.07	0.14	0.09	0.08	0.06	0.09		-0.02	0.06	0.15	0.08	0.02	0.09	0.11	0.14
	0.00	0.00	0.00	0.00	0.00	0.00		0.07	0.00	0.00	0.00	0.06	0.00	0.00	0.00
8 CONFORM_TAX	0.00	0.05	-0.35	-0.25	-0.35	-0.35	-0.05		0.00	-0.11	-0.07	-0.01	0.00	0.00	0.00
	0.63	0.00	0.00	0.00	0.00	0.00	0.00		0.82	0.00	0.00	0.17	0.99	0.95	0.99
9 UNCERTAIN	0.82	0.09	0.12	0.14	0.03	0.11	0.09	0.00		0.39	0.18	0.56	0.12	0.15	0.13
	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.62		0.00	0.00	0.00	0.00	0.00	0.00
10 BOG	0.37	0.09	0.14	0.15	0.04	0.13	0.12	-0.11	0.40		0.54	0.41	0.13	0.15	0.14
	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00		0.00	0.00	0.00	0.00	0.00
11 FOG	0.25	0.02	0.07	0.08	0.02	0.06	0.06	-0.07	0.12	0.51		0.34	0.15	0.16	0.16
	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00		0.00	0.00	0.00	0.00
12 FSIZE	0.50	0.06	0.11	0.14	0.02	0.10	-0.03	-0.04	0.54	0.41	0.32		0.24	0.29	0.27
	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00		0.00	0.00	0.00
13 VOLDISC_DUMMY	0.07	-0.02	0.05	0.07	0.04	0.05	0.02	-0.03	0.12	0.14	0.14	0.25		0.96	0.96
	0.00	0.07	0.00	0.00	0.00	0.00	0.43	0.01	0.00	0.00	0.00	0.00		0.00	0.00
14 VOLDISC_FREQ	0.10	0.00	0.06	0.08	0.05	0.06	0.05	-0.03	0.16	0.15	0.13	0.31	0.82		0.96
	0.00	0.96	0.00	0.00	0.00	0.00	0.01	0.02	0.00	0.00	0.00	0.00	0.00		0.00
15 VOLDISC_HORIZON	0.08	-0.02	0.05	0.08	0.04	0.06	0.02	-0.03	0.13	0.14	0.15	0.27	0.99	0.84	
	0.00	0.09	0.00	0.00	0.00	0.00	0.36	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Table 2. Baseline results

This table presents estimates of baseline OLS regressions that examine the relationship between a firm's tax avoidance and its 10-K report tone ambiguity. The dependent variable is WEAKMODAL, which is the percentage of weak modal words as defined by Loughran and McDonald (2011). PBTD is the permanent book-tax difference. CETR is the cash effective tax rate. CETR_3Y is the three-year cumulative cash effective tax rate. CETR_ADJ is the mean industry- and size-matched Cash ETR. TA_FACTOR is the first principal component of the above four tax avoidance measures. Robust *t*-statistics clustered at the firm level are reported in parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively. Details on the construction of all variables are provided in Appendix A.

Dependent variable: WEAKMODAL	(1)	(2)	(3)	(4)	(5)
PBTD	0.056* (1.75)				
CETR		0.023*** (2.77)			
CETR_3Y			0.027** (2.43)		
CETR_ADJ				0.021** (2.36)	
TA_FACTOR					0.003*** (2.67)
MTB	0.012*** (4.23)	0.012*** (4.30)	0.012*** (4.25)	0.012*** (4.29)	0.012*** (4.26)
Ln(MVE)	0.000 (0.20)	0.000 (0.15)	0.000 (0.12)	0.000 (0.20)	0.000 (0.14)
AGE	-0.002*** (-9.94)	-0.002*** (-9.92)	-0.002*** (-9.91)	-0.002*** (-9.92)	-0.002*** (-9.91)
SPI	-0.197** (-2.20)	-0.150* (-1.69)	-0.161* (-1.82)	-0.157* (-1.76)	-0.153* (-1.72)
EARN	-0.217*** (-5.22)	-0.215*** (-5.16)	-0.214*** (-5.14)	-0.215*** (-5.16)	-0.214*** (-5.15)
EARNVOL	0.718*** (3.97)	0.744*** (4.12)	0.746*** (4.13)	0.744*** (4.12)	0.742*** (4.11)
SRET	-0.006** (-2.46)	-0.007*** (-2.70)	-0.005** (-2.28)	-0.006*** (-2.62)	-0.006*** (-2.65)
SRETVOL	0.076*** (5.33)	0.075*** (5.29)	0.074*** (5.27)	0.075*** (5.34)	0.074*** (5.24)
Ln(BSEG +1)	-0.049*** (-7.17)	-0.049*** (-7.10)	-0.049*** (-7.08)	-0.049*** (-7.11)	-0.049*** (-7.08)
Ln(GSEG +1)	-0.021*** (-4.00)	-0.021*** (-3.95)	-0.020*** (-3.94)	-0.021*** (-3.95)	-0.021*** (-3.95)
Ln(NMITEMS)	-0.035 (-0.76)	-0.033 (-0.71)	-0.032 (-0.69)	-0.033 (-0.71)	-0.033 (-0.70)
DEL	0.027***	0.027***	0.027***	0.027***	0.027***

	(4.45)	(4.43)	(4.44)	(4.43)	(4.43)
LEV	-0.045***	-0.048***	-0.048***	-0.047***	-0.048***
	(-3.17)	(-3.35)	(-3.37)	(-3.31)	(-3.37)
INST	0.068***	0.067***	0.067***	0.067***	0.067***
	(5.96)	(5.86)	(5.86)	(5.86)	(5.86)
LIT	0.056***	0.057***	0.057***	0.057***	0.056***
	(5.34)	(5.35)	(5.35)	(5.35)	(5.33)
Industry fixed effects	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes
No. of observations	25,851	25,851	25,851	25,851	25,851
Adj. R^2	0.48	0.48	0.48	0.48	0.48

Table 3. Impact of tax-based proprietary information costs

This table presents OLS regression results of whether the association between tax avoidance and 10-K report tone ambiguity varies with tax-based proprietary costs. The dependent variable is WEAKMODAL, which is the percentage of weak modal words as defined by Loughran and McDonald (2011). Tax-based proprietary costs are proxied by IRS audit probability (AUDIT_PROB), tax haven (TAX_HAVEN), and multinational firms (MULTI_FIRM). AUDIT_PROB is the lagged likelihood of an IRS audit. TAX_HAVEN is an indicator variable that equals to one if a firm has at least one subsidiary in a tax haven mentioned in Exhibit 21 of Form 10-K, as defined by Dyreng and Lindsey (2009). MULTI_FIRM is an indicator variable that equals to one if a firm's foreign income (PIFO) is not zero or missing. In columns (1) and (2), the sample is partitioned according to the sample median value of AUDIT_PROB. Robust *t*-statistics clustered at the firm level are reported in parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively. Details on the construction of all variables are provided in Appendix A.

Dependent variable: WEAKMODAL	AUDIT_PROB		TAX_HAVEN		MULTI_FIRM	
	(1) High	(2) Low	(3) Yes	(4) No	(5) Yes	(6) No
TA_FACTOR	0.004** (2.50)	0.002 (1.44)	0.004** (2.34)	0.002 (1.52)	0.006*** (3.46)	0.000 (0.22)
MTB	0.017*** (4.24)	0.002 (0.70)	0.020*** (5.34)	0.006 (1.61)	0.014*** (3.95)	0.009** (2.26)
Ln(MVE)	0.001 (0.18)	0.011*** (2.84)	-0.002 (-0.44)	0.004 (1.33)	0.002 (0.60)	-0.001 (-0.35)
AGE	-0.001*** (-6.01)	-0.003*** (-9.46)	-0.001*** (-5.19)	-0.002*** (-7.74)	-0.002*** (-7.28)	-0.002*** (-7.44)
SPI	-0.113 (-0.95)	-0.207* (-1.67)	-0.110 (-0.80)	-0.134 (-1.03)	-0.042 (-0.36)	-0.240* (-1.84)
EARN	-0.261*** (-4.46)	-0.170*** (-3.50)	-0.292*** (-4.93)	-0.192*** (-3.58)	-0.303*** (-5.99)	-0.143** (-2.57)
EARNVOL	1.225*** (4.63)	0.416** (1.98)	1.504*** (4.78)	0.452** (2.25)	1.388*** (4.73)	0.236 (1.13)
SRET	-0.007* (-1.90)	-0.005 (-1.62)	-0.011*** (-2.67)	-0.003 (-0.82)	-0.008** (-2.29)	-0.003 (-1.03)
SRETVOL	0.100*** (4.54)	0.061*** (3.59)	0.104*** (4.22)	0.066*** (3.93)	0.108*** (5.24)	0.052*** (2.87)
Ln(BSEG +1)	-0.039*** (-4.81)	-0.062*** (-6.25)	-0.049*** (-4.99)	-0.045*** (-4.82)	-0.049*** (-5.43)	-0.045*** (-4.48)
Ln(GSEG +1)	-0.020*** (-2.83)	-0.021*** (-3.24)	-0.020*** (-2.67)	-0.022*** (-3.21)	-0.025*** (-3.27)	-0.012 (-1.62)
Ln(NMITEMS)	-0.077 (-1.32)	-0.046 (-0.71)	-0.072 (-0.99)	-0.025 (-0.41)	-0.095 (-1.42)	0.014 (0.23)
DEL	0.018** (2.38)	0.033*** (4.26)	0.035*** (3.96)	0.019** (2.54)	0.028*** (3.41)	0.024*** (3.14)
LEV	0.002	-0.084***	-0.021	-0.054***	-0.048**	-0.053***

	(0.11)	(-4.45)	(-0.92)	(-3.10)	(-2.27)	(-3.07)
INST	0.034**	0.059***	0.051***	0.068***	0.063***	0.072***
	(2.47)	(3.94)	(2.78)	(4.63)	(4.14)	(5.07)
LIT	0.054***	0.060***	0.040***	0.067***	0.056***	0.061***
	(3.77)	(4.36)	(2.72)	(4.61)	(4.41)	(3.78)
<i>p</i> -value of difference for TA_FACTOR		0.019		0.024		0.002
Industry fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
No. of observations	12,407	12,431	9,028	12,999	13,581	12,740
Adj. R^2	0.47	0.48	0.48	0.44	0.50	0.46

Table 4. Tax avoidance and voluntary disclosure

This table examines the association between tax avoidance and voluntary disclosure. In columns (1) and (2), the dependent variable is an indicator variable set equal to one if the firm provides earnings guidance. In columns (3) and (4), the dependent variable is the number of management forecasts made during a fiscal year for the earnings of that fiscal year. In columns (5) and (6), the dependent variable is the natural logarithm of one plus the number of days between the management forecast date for a firm's fiscal year earnings and the fiscal period end date. POST is an indicator variable that equals to one after the implementation of Regulation Fair Disclosure (Reg FD) in 2000. All control variables are measured at the beginning of the year except for Issuance and Litigation industry dummy. Values of *t*-statistics that are based on robust standard errors and firm-level clustering are reported in parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively. Details on the construction of all variables are provided in Appendix A.

Dependent variable	VOLDISC_DUMMY		VOLDISC_FREQ		VOLDISC_HORIZON	
	(1) Logit	(2) Logit	(3) Tobit	(4) Tobit	(5) Tobit	(6) Tobit
TA_FACTOR	0.047** (2.24)	-0.043 (-0.97)	0.045*** (3.35)	-0.005 (-0.39)	0.049*** (2.67)	-0.012 (-0.46)
TA_FACTOR × POST		0.108** (2.19)		0.069*** (3.21)		0.082** (2.52)
INDCON	0.525 (1.30)	0.537 (1.33)	0.480 (1.45)	0.495 (1.50)	0.578 (1.46)	0.595 (1.50)
Ln (1+ANALYST)	0.522*** (7.07)	0.518*** (7.02)	0.250*** (5.22)	0.247*** (5.16)	0.457*** (6.89)	0.453*** (6.84)
INST	1.288*** (6.11)	1.300*** (6.17)	0.510*** (3.13)	0.518*** (3.18)	1.147*** (5.49)	1.156*** (5.53)
SIZE	0.067 (1.53)	0.068 (1.56)	0.141*** (3.89)	0.141*** (3.90)	0.074* (1.73)	0.075* (1.75)
DEARN	1.550*** (3.86)	1.536*** (3.82)	0.856*** (3.79)	0.852*** (3.77)	1.208*** (3.79)	1.203*** (3.77)
EARNVOL	-4.021*** (-4.66)	-4.077*** (-4.70)	-2.351*** (-4.58)	-2.378*** (-4.62)	-3.205*** (-4.60)	-3.238*** (-4.63)
RVOL	-3.323*** (-5.64)	-3.299*** (-5.60)	-2.056*** (-5.76)	-2.030*** (-5.68)	-2.992*** (-6.07)	-2.961*** (-6.00)
MTB	-0.020 (-0.57)	-0.017 (-0.49)	-0.002 (-0.10)	-0.001 (-0.03)	-0.018 (-0.54)	-0.016 (-0.48)
LEV	1.185*** (4.58)	1.189*** (4.60)	0.778*** (3.78)	0.782*** (3.80)	1.224*** (4.97)	1.229*** (4.99)
ISS	0.393*** (2.60)	0.387** (2.56)	0.873*** (4.86)	0.867*** (4.82)	0.611*** (3.89)	0.604*** (3.85)
LIT	0.052 (0.36)	0.049 (0.34)	0.042 (0.40)	0.040 (0.38)	0.026 (0.19)	0.023 (0.17)
Industry fixed effects	Yes	Yes	Yes	Yes	Yes	Yes

Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
No. of observations	12,378	12,378	12,380	12,380	12,380	12,380
Pseudo R^2	0.23	0.23	0.07	0.07	0.06	0.06

Table 5. Tax avoidance and voluntary disclosure: Effect of IRS audits

This table examines the effect of IRS audit probability on the association between tax avoidance and voluntary disclosure. In columns (1) and (2), the dependent variable is an indicator variable set equal to one if the firm provides earnings guidance. In columns (3) and (4), the dependent variable is the number of management forecasts made during a fiscal year for the earnings of that fiscal year. In columns (5) and (6), the dependent variable is the natural logarithm of one plus the number of days between the management forecast date for a firm's fiscal year earnings and the fiscal period end date. The sample is partitioned according to the sample median value of AUDIT_PROB, which is the lagged likelihood of an IRS audit. Values of *t*-statistics that are based on robust standard errors and firm-level clustering are reported in parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively. Details on the construction of all variables are provided in Appendix A.

Dependent variable	VOLDISC_DUMMY		VOLDISC_FREQ		VOLDISC_HORIZON	
Partition by AUDIT_PROB	(1) High	(2) Low	(3) High	(4) Low	(5) High	(6) Low
TA_FACTOR	0.024 (0.77)	0.092*** (3.25)	0.029 (0.92)	0.053*** (4.24)	0.036 (1.11)	0.060*** (3.11)
INDCON	0.305 (0.57)	0.698 (1.39)	0.099 (0.14)	0.474 (1.61)	0.365 (0.62)	0.388 (0.94)
Ln (1+ANALYST)	0.401*** (4.02)	0.651*** (6.67)	0.153 (1.57)	0.292*** (6.72)	0.433*** (4.23)	0.496*** (7.00)
INST	0.691** (2.53)	1.440*** (4.34)	0.396 (1.40)	0.599*** (3.69)	0.746** (2.55)	1.355*** (4.72)
SIZE	0.028 (0.46)	0.211** (2.49)	0.154** (2.29)	0.113*** (3.21)	0.030 (0.46)	0.101 (1.62)
DEARN	0.987 (1.62)	1.585*** (2.89)	0.911 (1.22)	0.542** (2.50)	1.083* (1.74)	0.997*** (2.90)
EARNVOL	-5.733*** (-4.23)	-3.381*** (-3.14)	-5.885*** (-4.96)	-1.520*** (-2.95)	-6.357*** (-4.71)	-2.177*** (-3.05)
RVOL	-5.006*** (-5.73)	-1.549** (-1.99)	-3.475*** (-4.17)	-1.241*** (-3.54)	-5.199*** (-5.70)	-1.092** (-2.14)
MTB	-0.073 (-1.52)	0.078* (1.71)	0.066 (1.40)	-0.001 (-0.03)	-0.072 (-1.44)	0.056 (1.41)
LEV	1.032*** (3.20)	1.442*** (3.61)	0.999** (2.51)	0.537*** (2.80)	1.149*** (3.29)	1.201*** (4.08)
ISS	0.398** (2.54)	0.710 (1.11)	0.832*** (3.52)	0.638** (2.45)	0.482*** (2.99)	0.675 (0.95)
LIT	-0.067 (-0.36)	0.235 (1.13)	-0.007 (-0.03)	0.077 (0.77)	-0.076 (-0.36)	0.115 (0.78)
<i>p</i> -value of difference for TA_FACTOR		0.004		0.055		0.056
Industry fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes

No. of observations	5,520	5,624	5,521	5,684	5,521	5,684
Pseudo R^2	0.18	0.25	0.08	0.07	0.06	0.06

Table 6: Impact of monitoring

This table presents OLS results of whether the association between tax avoidance and 10-K report tone ambiguity varies with internal board monitoring. The dependent variable is WEAKMODAL, which is the percentage of weak modal words as defined by Loughran and McDonald (2011). The degree of board monitoring is measured by board size (BOARD_SIZE), board independence (BOARD_INDEP), and director ownership (DO). BOARD_SIZE is the number of directors on the board. BOARD_INDEP is the percentage of outside directors on the board. DO is the proportion of outstanding shares owned collectively by all directors. The sample is partitioned according to the sample median value of each internal governance proxy. Other control variables are the same as in Table 2. Robust *t*-statistics clustered at the firm level are reported in parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively. Details on the construction of all variables are provided in Appendix A.

Dependent variable: WEAKMODAL	BOARD_SIZE		BOARD_INDEP		DO	
	(1)	(2)	(3)	(4)	(5)	(6)
	High	Low	High	Low	High	Low
TA_FACTOR	0.002 (0.64)	0.007*** (2.79)	0.004* (1.76)	0.006*** (2.78)	0.002 (0.73)	0.007*** (2.70)
MTB	0.014*** (3.03)	0.009* (1.92)	0.019*** (3.96)	0.010** (2.22)	0.006 (1.17)	0.015*** (3.20)
Ln(MVE)	-0.002 (-0.42)	-0.001 (-0.29)	-0.007* (-1.89)	0.003 (0.61)	0.005 (0.87)	-0.008* (-1.96)
AGE	-0.001*** (-3.66)	-0.002*** (-6.10)	-0.001*** (-4.55)	-0.002*** (-5.57)	-0.002*** (-5.01)	-0.001*** (-3.44)
SPI	-0.254 (-1.57)	-0.033 (-0.17)	-0.150 (-0.93)	-0.145 (-0.91)	-0.161 (-0.88)	-0.118 (-0.74)
EARN	-0.175** (-2.25)	-0.326*** (-4.76)	-0.237*** (-3.58)	-0.295*** (-4.57)	-0.324*** (-4.29)	-0.209*** (-2.88)
EARNVOL	1.282** (2.56)	1.108*** (2.99)	1.329*** (3.25)	1.470*** (4.17)	1.271*** (3.02)	1.108*** (2.60)
SRET	-0.003 (-0.51)	-0.005 (-0.91)	-0.009* (-1.71)	-0.003 (-0.55)	-0.002 (-0.32)	-0.009 (-1.51)
SRETVOL	0.069* (1.96)	0.209*** (4.69)	0.104*** (3.44)	0.107*** (3.20)	0.179*** (4.48)	0.143*** (3.61)
Ln(BSEG +1)	-0.043*** (-3.56)	-0.063*** (-5.44)	-0.049*** (-4.14)	-0.048*** (-4.17)	-0.047*** (-3.67)	-0.059*** (-4.61)
Ln(GSEG +1)	-0.026*** (-2.60)	-0.016* (-1.82)	-0.024*** (-2.76)	-0.019* (-1.91)	-0.031*** (-2.96)	-0.014 (-1.56)
Ln(NMITEMS)	-0.108 (-1.09)	-0.046 (-0.59)	-0.178* (-1.96)	-0.034 (-0.42)	-0.065 (-0.65)	-0.097 (-1.10)
DEL	0.023** (2.08)	0.031*** (2.99)	0.019* (1.77)	0.030*** (3.05)	0.029** (2.55)	0.024** (2.20)
LEV	-0.037 (-1.20)	-0.071*** (-2.73)	-0.092*** (-3.17)	0.005 (0.18)	-0.083*** (-2.90)	-0.024 (-0.81)
INST	0.057**	0.010	0.053**	0.028	0.027	0.059**

LIT	(2.28) 0.041* (1.95)	(0.45) 0.052*** (2.77)	(2.39) 0.044** (2.31)	(1.29) 0.060*** (3.45)	(1.12) 0.065*** (2.99)	(2.21) 0.028 (1.59)
<i>p</i> -value of difference for TA_FACTOR		0.019		0.008		0.008
Industry fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
No. of observations	7,123	5,523	7,549	6,528	5,999	5,998
Adj. R^2	0.45	0.54	0.48	0.48	0.48	0.49

Table 7. Robustness checks: Propensity score matching

This table presents a propensity-score-match sample analysis of the baseline regression model (Model 5 of Table 2) investigating the effect of tax avoidance on 10-K report tone ambiguity. We match each treatment observation (i.e., a firm with high tax avoidance) with a control observation (i.e., a firm with low tax avoidance) that is similar in terms of several observable characteristics but not in its level of tax avoidance. Panel A shows results of the first-stage Logit regression model and post-match diagnostic regression analysis. The dependent variable, High TA_FACTOR, is a dummy variable that equals one if TA_FACTOR is higher than the industry-year median, and zero otherwise, following Shipman et al. (2017). We use the predicted propensity scores of High TA_FACTOR to match (without replacement) each treatment firm with a control firm using the one-to-one closest propensity score with a caliper width of 0.05. The procedure yields a sample of 5,265 matched pairs of treatment and control firms. Panel B presents the results of a covariate balance test, which assesses whether the average values of observable characteristics are similar across treatment and control firms. Panel C. reports estimates of our baseline regression model on the propensity-score-matched sample. Robust *t*-statistics clustered at the firm level are reported in parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively. Details on the construction of all variables are provided in Appendix A.

<i>Panel A. First-stage Logit regression model</i>		
Dependent variable: High TA_FACTOR	(1) Pre-matched sample	(2) Post-matched sample
MTB	0.109*** (4.68)	0.037 (1.14)
Ln(MVE)	0.091*** (5.10)	-0.022 (-0.94)
AGE	-0.003* (-1.82)	-0.001 (-0.33)
SPI	-10.829*** (-10.98)	0.159 (0.10)
EARN	-3.780*** (-11.69)	0.219 (0.46)
EARNVOL	4.790*** (2.94)	0.583 (0.25)
SRET	0.260*** (8.75)	0.024 (0.50)
SRETVOL	1.382*** (9.43)	0.301 (1.40)
Ln(BSEG +1)	-0.191*** (-3.28)	0.056 (0.71)
Ln(GSEG +1)	0.060 (1.45)	-0.019 (-0.33)
Ln(NMITEMS)	-0.757* (-1.69)	0.538 (0.87)
DEL	0.067	-0.049

	(1.42)	(-0.80)
LEV	1.011***	0.032
	(7.23)	(0.16)
INST	0.189**	-0.025
	(2.11)	(-0.20)
LIT	0.337***	-0.008
	(4.16)	(-0.08)
Industry fixed effects	Yes	Yes
Year fixed effects	Yes	Yes
No. of observations	25,851	10,530
Pseudo R^2	0.04	0.00

Panel B. Covariate balance test using the propensity-score-matched sample

	(1)	(2)	(3)
	High TA_FACTOR firms (N= 5,265)	Low TA_FACTOR firms (N= 5,265)	P-value of difference in mean
MTB	1.995	1.954	0.171
MVE	5,738	5,684	0.861
AGE	25.319	25.719	0.246
SPI	-0.005	-0.005	0.825
EARN	0.157	0.155	0.122
EARNVOL	0.016	0.015	0.171
SRET	0.125	0.116	0.284
SRETVOL	0.237	0.235	0.210
BSEG	3.126	3.124	0.961
GSEG	3.352	3.375	0.533
NMITEMS	319.389	319.386	0.997
DEL	0.601	0.612	0.247
LEV	0.180	0.182	0.484
INST	0.621	0.627	0.258
LIT	0.367	0.367	0.950

Panel C. Regression using the propensity-score-matched sample

Dependent variable: WEAKMODAL	(1) OLS
TA_FACTOR	0.004** (2.42)
MTB	0.013*** (3.48)
Ln(MVE)	-0.000 (-0.14)
AGE	-0.002***

SPI	(-8.15) -0.097 (-0.72)
EARN	-0.229*** (-4.16)
EARNVOL	0.585** (2.58)
SRET	-0.009** (-2.20)
SRETVOL	0.082*** (3.87)
Ln(BSEG +1)	-0.046*** (-5.36)
Ln(GSEG +1)	-0.018*** (-2.68)
Ln(NMITEMS)	-0.043 (-0.67)
DEL	0.032*** (4.54)
LEV	-0.036* (-1.94)
INST	0.055*** (3.84)
LIT	0.057*** (4.51)
Industry fixed effects	Yes
Year fixed effects	Yes
No. of observations	10,530
Adj. R^2	0.48

Table 8. Robustness checks: Difference-in-Differences (DID) analysis

This table presents the results for the difference-in-differences regressions examining the effect of tax avoidance on 10-K report tone ambiguity surrounding the implementation of Check-the-Box (CTB) regulation in 1997. Panel A shows summary statistics for propensity-score-matched sample. Panel B presents the results of investigating the DID regression model and verifying the parallel trends assumption. Following Balakrishnan et al. (2019) and Nguyen (2021), TREATMENT is a dummy variable equal to one if a firm reported a non-zero pre-tax foreign income (PFI), and zero otherwise. POST1997 (POST1998) is a dummy variable equal to one during the period 1997–2000 (1998–2000), and zero otherwise. The sample period is 1994–2000. To control for observable differences in characteristics between treated firms and control firms, we use the propensity score matching method. First, we obtain propensity scores by running a Logit regression of TREATMENT on the same set of control variables as in Table 2 and measured in the year 1996. Second, we use the predicted propensity scores of TREATMENT to match (without replacement) each treatment firm with a control firm using the one-to-one closest propensity score with a caliper width of 0.05. The sample consists of 389 matched pairs of firms. YEAR1996 (YEAR1997) is a dummy variable equal to one for the year 1996 (1997), and zero otherwise. Robust *t*-statistics clustered at the firm level are reported in parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively. Details on the construction of all variables are provided in Appendix A.

<i>Panel A. Summary statistics for propensity-score-matched sample</i>						
Variable	Mean	Median	Std. Dev.	25 th percentile	75 th percentile	<i>N</i>
Tone ambiguity measure						
WEAKMODAL	0.347	0.309	0.164	0.228	0.428	1,427
Firm characteristics						
MTB	0.347	0.309	0.164	0.228	0.428	1,427
MVE	1.897	1.481	1.299	1.113	2.150	1,427
AGE	1,526	228	4,456	65.857	998	1,427
SPI	20.889	16.000	14.495	9.000	29.000	1,427
EARN	-0.004	0.000	0.017	-0.001	0.000	1,427
EARNVOL	0.163	0.155	0.070	0.113	0.204	1,427
SRET	0.018	0.014	0.014	0.009	0.022	1,427
SRETVOL	0.119	0.029	0.573	-0.244	0.321	1,427
BSEG	0.266	0.211	0.238	0.162	0.287	1,427
GSEG	2.666	2.000	1.077	2.000	3.000	1,427
NMITEMS	2.496	2.000	1.016	2.000	3.000	1,427
DEL	256	254	14.484	246	268	1,427
LEV	0.573	1.000	0.495	0.000	1.000	1,427
INST	0.200	0.174	0.170	0.036	0.329	1,427
LIT	0.432	0.460	0.242	0.229	0.629	1,427

Panel B. Difference-in-differences regressions with propensity-score-matched sample

Dependent variable: WEAKMODAL	(1)	(2)
TREATMENT	-0.021 (-1.21)	-0.035* (-1.68)
TREATMENT × POST1997	0.037** (2.22)	
TREATMENT × YEAR1996		0.029 (1.36)
TREATMENT × YEAR1997		0.044* (1.77)
TREATMENT × POST1998		0.055** (2.51)
MTB	0.013 (1.53)	0.013 (1.53)
Ln(MVE)	0.008 (1.22)	0.008 (1.24)
AGE	-0.001** (-2.45)	-0.001** (-2.48)
SPI	-0.126 (-0.49)	-0.118 (-0.46)
EARN	-0.284*** (-2.86)	-0.286*** (-2.86)
EARNVOL	1.002** (2.10)	1.002** (2.10)
SRET	-0.006 (-0.76)	-0.006 (-0.82)
SRETVOL	0.030 (0.89)	0.030 (0.87)
Ln(BSEG +1)	-0.010 (-0.61)	-0.010 (-0.57)
Ln(GSEG +1)	-0.011 (-0.64)	-0.011 (-0.63)
Ln(NMITEMS)	-0.029 (-0.21)	-0.035 (-0.26)
DEL	0.005 (0.38)	0.005 (0.37)
LEV	-0.043 (-1.22)	-0.043 (-1.22)
INST	0.113*** (2.96)	0.112*** (2.96)
LIT	0.065*** (2.90)	0.065*** (2.92)
Industry fixed effects	Yes	Yes
Year fixed effects	Yes	Yes

No. of observations	1,427	1,427
Adj. R^2	0.17	0.17

Table 9. Additional robustness checks

This table presents further robustness checks for the baseline OLS regressions that examine the relationship between a firm's tax avoidance and its 10-K report tone ambiguity. The dependent variable in columns (1)-(5) is WEAKMODAL, which is the percentage of weak modal words as defined by Loughran and McDonald (2011). The dependent variable in column (6) is UNCERTAIN, which is the percentage of words conveying uncertainty as defined by Loughran and McDonald (2011). TA_FACTOR is the first principal component of the following four tax avoidance measures: PBTB, CETR, CETR_3Y, and CETR_ADJ. Annual 10-K report readability is measured by BOG, FOG, and FSIZE. BOG is the Bog index, based on Bonsall et al. (2017). FOG is the Gunning fog index, which is defined as 0.4 times the sum of average number of words per sentence and the percentage of 10-K complex words. FSIZE is the natural logarithm of the 10-K file size in megabytes of the SEC EDGAR, based on Loughran and McDonald (2014). A high value of the readability proxy implies less readable text. UTB is uncertain tax benefits, calculated as a firm's uncertain tax benefits at the end of a fiscal year (TXTUBEND) scaled by total assets (AT) at the fiscal year-end. Based on Badertscher et al. (2019), CONFORM_TAX is the conforming tax avoidance that removes the impact of nonconforming tax strategies. It is calculated as the residual of an OLS regression estimated by Fama-French 48 industry and fiscal year combinations. Specifically, the ratio of cash taxes paid to lagged total assets is regressed on total book-tax differences (BTD), a dummy variable (NEG) that equals to one for observations with negative book-tax differences (and zero otherwise), the interaction of BTD and NEG, the level of net operating loss carryforwards (NOL), and changes in NOLs. Robust *t*-statistics clustered at the firm level are reported in parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively. Details on the construction of all variables are provided in Appendix A.

Dependent variable	WEAKMODAL					UNCERTAIN
	(1)	(2)	(3)	(4)	(5)	(6)
TA_FACTOR	0.003** (2.35)	0.003*** (2.64)	0.003*** (2.62)	0.004** (2.00)		0.003** (2.21)
BOG	0.002*** (6.06)					
FOG		0.011*** (5.53)				
FSIZE			0.010*** (3.43)			
UTB				0.719*** (2.68)		
CONFORM_TAX					-0.072 (-0.81)	
MTB	0.013*** (4.59)	0.012*** (4.24)	0.013*** (4.45)	0.010** (2.25)	0.015*** (6.44)	0.009** (2.51)
Ln(MVE)	-0.001 (-0.31)	-0.001 (-0.26)	-0.001 (-0.48)	-0.007** (-2.03)	0.001 (0.30)	0.003 (0.82)
AGE	-0.002*** (-9.28)	-0.002*** (-9.76)	-0.002*** (-9.88)	-0.002*** (-8.72)	-0.002*** (-10.15)	-0.002*** (-9.35)
SPI	-0.111 (-1.24)	-0.130 (-1.47)	-0.147* (-1.66)	-0.337** (-2.47)	-0.189*** (-2.77)	-0.314** (-2.52)

EARN	-0.196*** (-4.83)	-0.202*** (-4.86)	-0.210*** (-5.04)	-0.193*** (-3.17)	-0.247*** (-6.77)	-0.263*** (-5.02)
EARNVOL	0.724*** (4.00)	0.745*** (4.15)	0.717*** (3.99)	0.853*** (2.70)	0.628*** (4.99)	0.344 (1.38)
SRET	-0.006** (-2.57)	-0.006** (-2.52)	-0.006*** (-2.61)	-0.001 (-0.17)	-0.006*** (-3.05)	-0.005 (-1.62)
SRETVOL	0.070*** (4.98)	0.072*** (5.08)	0.072*** (5.13)	0.038* (1.79)	0.009* (1.70)	0.098*** (5.14)
Ln(BSEG +1)	-0.052*** (-7.50)	-0.049*** (-7.16)	-0.050*** (-7.38)	-0.057*** (-5.65)	-0.052*** (-7.22)	-0.066*** (-7.05)
Ln(GSEG +1)	-0.020*** (-3.89)	-0.020*** (-3.92)	-0.021*** (-4.07)	-0.022*** (-2.92)	-0.018*** (-3.43)	-0.001 (-0.10)
Ln(NMITEMS)	-0.029 (-0.62)	-0.030 (-0.65)	-0.038 (-0.82)	-0.072 (-0.81)	-0.021 (-0.46)	0.091 (1.40)
DEL	0.025*** (4.22)	0.026*** (4.38)	0.026*** (4.40)	0.039*** (4.37)	0.037*** (6.14)	0.022*** (2.72)
LEV	-0.050*** (-3.49)	-0.052*** (-3.65)	-0.054*** (-3.78)	-0.065*** (-2.93)	-0.034** (-2.47)	-0.072*** (-3.65)
INST	0.063*** (5.51)	0.064*** (5.63)	0.066*** (5.75)	0.057*** (3.94)	0.056*** (4.93)	0.099*** (6.29)
LIT	0.054*** (5.11)	0.055*** (5.27)	0.056*** (5.31)	0.049*** (3.29)	0.058*** (5.29)	0.048*** (3.52)
Industry fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
No. of observations	25,176	25,850	25,851	9,864	23,078	25,850
Adj. R^2	0.48	0.48	0.48	0.30	0.49	0.54