Now is the Time: The Impact of Linguistic Time Reference on Corporate Default Risk

Kung-Cheng Ho (School of Finance, Guangdong University of Finance and Economics)

Yiling Chen (Faculty of Business and Economics, University of Hong Kong)

Dezhu Ye (Jinan University)

Cheng Yan (Essex Business School, University of Essex, UK)

Electronic version of an article published as DOI 10.1142/S1094406023500105

Abstract

The research problem This paper assesses whether and how people's perceptions of time — strong future time reference (FTR) versus weak FTR — affect corporate default risk.
Motivation or theoretical reasoning Studies have shown that default risk varies across firms, regions, and countries, highlighting the need for a comprehensive understanding of the contributing factors. Traditional studies focus on how firm-level, industry-level, national and international economic and financial variables shape corporate default risk, but they fail to explain cross-country and cross-regional differences in corporate default risk from the perspective of informal institutions, particularly, language. This study takes the first step to examine whether and how future-oriented language shapes corporate default risk.
The test hypotheses We first tested whether strong-FTR language decreases corporate default risk. We further tested whether the effect of strong-FTR language on default risk depends on firms' level of information transparency. In addition, we tested whether the effect of strong-FTR language on default risk depends on a country's disclosure requirements. Lastly, we tested whether the effect of strong-FTR language on default risk depends on a country's control of corruption.
Target population We find that corporate default risk is significantly higher in regions dominated by speakers of weak-FTR languages, using a comprehensive sample of firms in 36 countries with 180,013 observations spanning from 1988 to 2017.
Adopted methodology Ordinary least square regressions were used in this study.
Analyses Corporate default risk is measured by two proxies of firm probability of default, following Merton [(1974) Journal of Finance, 29(2), 449–470] and Lee and Lin [(2012) Journal of International Financial Markets, Institutions, and Money, 22(4), 973–989]. Our independent variable is Strong FTR, which equals 1 if a language belongs to the strong-FTR language family, as defined by the European Science Foundation's Typology of Languages in Europe (EUROTYP) project. If a language does not require "obligatory [FTR] use in (main clause) prediction-based contexts" [Dahl (2000)Tense and Aspect in the Languages of Europe, O. Dahl (Ed.), pp. 309–328], then we put this language into the weak-FTR group. On the other hand, if a language does have the above-mentioned requirement, then it belongs to the strong-FTR group.
Findings We found that corporate default risk is significantly higher in regions dominated by speakers of weak-FTR languages. Furthermore, the FTR effect on default risk is weakened in countries with stronger formal institutions (e.g., high disclosure quality, greater transparency, and less corruption). Our results introduce a new explanation for heterogeneity in corporate default risk, provide insights about whether language is an economic institution, and adds to research on the effects of languages on economic and financial outcomes.

Now is the time: The impact of linguistic time reference on corporate default risk

Abstract

Speakers of weak future time reference (FTR) languages (e.g., Chinese) do not need to grammatically mark future events, while speakers of strong FTR languages (e.g., English) do. We conjecture that weak FTR languages lead speakers to hold less precise beliefs about the timing and hence are associated with higher corporate default risk. Accordingly, we find that corporate default risk is significantly higher in regions dominated by speakers of weak FTR languages, using a comprehensive sample of firms in 36 countries with 180,013 observations spanning from 1988 to 2017. Furthermore, the FTR effect on default risk is weakened in countries with stronger formal institutions (e.g., high disclosure quality, greater transparency, and less corruption). Our results introduce a new explanation for heterogeneity in corporate default risk, provide insights about whether language is an economic institution, and adds to research on the effects of languages on economic/financial outcomes.

Keywords**:** Language, Default risk, Economic institution, Future-time reference

JEL Classification: F39, G14, G28, G33, M41

1. Introduction

Understanding corporate default risk is crucial in accounting, economics, and finance, particularly in the aftermath of the late 2000s global financial crisis (see, e.g., Jia et al., 2020 and the references therein). Decades of studies have demonstrated that default risk varies significantly across firms, regions, and countries, highlighting the need for a comprehensive understanding of the factors that contribute to this variation. While traditional studies on this topic typically examine how firm-level, industry-level, national and international economic/financial variables (typically the ratios which capture the firm-level characteristics) shape corporate default risk, they fall short in explaining the cross-country and/or cross-region differences in corporate default risk from the perspective of institutions, especially informal institutions.[1] One under-researched potential institution which we focus in this paper is language, as recent studies in linguistics and psychology show that language may affect thinking[2]/decisions and a crucial difference across languages that may be associated with future-oriented behaviors is whether they require grammatically marked future events or not[3]. For instance, speakers of weak future time reference (FTR) languages (e.g., Chinese) do not need to grammatically mark future events, while speakers of strong FTR languages (e.g., English) do. Although a seminal study in economics has demonstrated that cross-national variations in people's perceiving time horizon affect their long-term oriented behaviors (M.K. Chen, 2013), little is known about whether and how these intertemporal trade-offs or cross-national variations in perceiving time horizon affect corporate default risk. To fill this research gap, we present cross-national evidence for the real effect of languages on corporate default risk from the perspective of FTR.

[1] Although North (1991) and Williamson (2000) emphasize the importance of both formal and informal institutions, the traditional accounting, economics and finance literature focuses on formal institutions (e.g., law and investor protection) and pay little attention to informal institutions (Karolyi, 2016). The first widely recognized informal institution is perhaps religion, which has only been gradually accepted since the seminal study of Stulz and Williamson (2003). Recently, culture has become another generally accepted informal institution in accounting, economics and finance (Aggarwal et al., 2016).

[2] For instance, different languages divide the visible spectrum into lexical categories differently, which is referred as ``the variation in basic color terms" in linguistics (see e.g., Berlin and Kay, 1969; Kay and Regier, 2006).

[3] It is the so-called Sapir–Whorf hypothesis, or "linguistic relativity" (for more details, see, e.g., Whorf et al., 1956; Hickmann, 2000; Boroditsky 2001, 2011; Richard and Toffoli, 2009; Ge et al., 2020).

Understanding whether and how strong vs weak FTR languages affect corporate default risk is critical for several reasons. Firstly, a substantial portion of firm defaults can be attributed to firm policies (see Bakshi et al., 2022, for a review) which are themselves shaped by social processes negotiated among members within the firm (e.g., Liang et al., 2018). Language, as a critical tool of communication, plays a critical role in these social processes (e.g., Gotti et al., 2021). It is thus natural to assess the role of language in determining firm default risk. Relatedly, while previous literature has shown the importance of temporal orientation for understanding firm behaviors in general (Flammer and Bansal 2017; Kaplan and Orlikowski 2013), little is known about how languages with different emphases on intertemporal trade-offs shape specific firm outcomes. We make an important advancement in this literature by looking at the impact of strong vs weak FTR languages on the risk of firm default. Lastly, our findings can provide valuable insights to firm managers who aim to reduce their default risk and to policymakers aiming to enhance financial stability and reduce systemic risk. In particular, if our results suggest that firms with a weak FTR language tend to have higher default risk since mangers and other firm members hold less precise beliefs about the timing of negative shocks in the future, then our findings would imply that one way to mitigate default risk in this language environment is to strengthen the beliefs and improve the information precision about future shocks. For policymakers, similarly, one potential takeaway would be to design regulations that incorporate such goals.

Ex ante, the answer to our research question is unclear. For instance, M.K. Chen (2013) proposes two mechanisms via which weak FTR speakers save more: One concern a linguistically-induced bias in time perception (i.e., weak FTR makes the future feel less distant), and the other one concerns the precision of beliefs about time (i.e., weak FTR speakers hold less precise beliefs about the timing of future events than the strong FTR speakers). Based on the mechanism of linguistically-induced bias in time perception, S. Chen et al. (2017) carry over the impact of language FTR from individual behaviors to the corporate domain, and demonstrate that companies hold more precautionary savings if they are located in weak FTR languages-dominated nations. Extending this argument, it is not difficult to conjecture that companies located in nations where weak FTR languages are spoken low corporate default risk, due to higher precautionary cash holding. However, the other mechanism of precision of beliefs about time may overturn this conjecture as companies located in nations where weak FTR languages are spoken also tend to have less precise beliefs about the timing of future shocks, which may exacerbate the classical adverse selection and moral hazard problems[4], and hence have higher corporate default risk. Ultimately, it is an empirical question of which mechanism is dominating.

We construct a comprehensive data set of firms in 36 countries with 180,013 observations spanning from 1988 to 2017 to examine whether the strength of FTR is associated with default risk. We use two measures to capture corporate default risk. One is the risk-neutral probability of default, which stems from the KMV model and Merton (1974). The other one is the physical Probability of Default (Lee et al., 2011; Lee and Lin, 2012; Lee et al., 2013; Ho et al., 2020). After controlling for country-level and firm-level characteristics, firms in countries with strong FTR languages display less default risk than those in countries with weak FTR languages.

Moreover, we introduce more country-level factors as moderators for linguistic effects on default risk. The influences of transparency, disclosure requirements, and level of corruption are investigated. High-quality transparency and information disclosure can mitigate moral hazard and adverse selection by reducing information asymmetry and enhancing a firm's investment efficiency (Chen et al., 2011), whereas corruption can affect the incidence rate of cases of corporate misconduct, such as accounting fraud, that harm a firm's value (Liu, 2016; Xu, 2018).

In addition, we conduct an instrumental variable regression analysis that uses the

[4] On the one hand, the classical adverse selection problem may be execrated by the use of weak-FTR language as outsiders or uninformed investors have less accurate information about the value of a firm, while strong-FTR may mitigate adverse selection. On the other hand, the classical adverse selection problem may also be execrated, as weak-FTR languages offer the opportunity for managers to do bad news hoarding, whereas strong-FTR languages may alleviate the risk of moral hazard via higher precision of beliefs about time. For instance, Brochet et al. (2016) argue that, during verbal disclosure, managers may hide adverse information through non-plain English and erroneous expressions, which affects investor behaviors.

two-stage least squares method to alleviate the potential effect of endogeneity. Following the literature (e.g., La Porta et al., 1997; Pevzner et al., 2015), we use a country's primary religious belief as an instrumental variable, as the primary belief is related to language in that they are both aspects of culture, but primary belief is not suggested to be correlated with default risk *ex ante*.

The present study contributes to the literature in several respects. First, it is among the few studies that provide evidence of the impact of FTR strength, especially in terms of how it affects adverse selection and moral hazard. Previous research has highlighted the influence of strong or weak FTR on cognition and corporate behavior through the mechanism of linguistically-induced bias in time perception (S. Chen et al., 2017). However, the present study investigates another mechanism of FTR and its economic implications-the precision of beliefs about time. Strong FTR languages are associated with higher precision of beliefs about time, which in turn reduces adverse selection and moral hazard. Our analysis reveals that the mechanism of precision of beliefs about time outweighs the mechanism of linguistically-induced bias in time perception regarding corporate default risk. Specifically, weak FTR speakers hold less precise beliefs about the timing of future events than strong FTR speakers, which increases the likelihood of adverse selection and moral hazard.

Secondly, this study sheds light on the global variation in default risk by exploring a fundamental factor that has been largely overlooked in previous research. While many studies have investigated the relationship between default risk and firm-level characteristics such as management quality (Grunert et al., 2005), macroeconomic conditions (Bonfim, 2009; Giesecke et al., 2011; Koopman et al., 2012), and noneconomic factors such as industry type (Chava and Jarrow, 2004), the impact of language on corporate default risk has been largely ignored. As adverse selection and moral hazard crucially affect default risk, our study finds that noninstitutional factors, such as language, play a significant role in determining corporate performance through influencing moral hazard and adverse selection.

The current research aligns with previous studies on incomplete information and investor response. Incomplete information can harm investors by increasing adverse selection and moral hazard. Countries with a language that places a high emphasis on precision and accuracy of beliefs about time, such as those with a strong FTR, may experience lower default risk. However, it's important to note that other factors at the country-level, such as anti-corruption measures and disclosure requirements, can also contribute to a reduction in corporate default risk. In an environment with opaque information, the effect of language on corporate default risk is even more pronounced, as evidenced by the findings of this study.

The remainder of the paper is organized as follows. Section 2 presents our main hypotheses. Section 3 describes the research design and sample. Section 4 presents the empirical results. Section 5 concludes.

2. Literate review and hypothesis development

According to the theory of linguistic relativity (or the Sapir–Whorf hypothesis), languages shape people's perception of the world by influencing their cognitive processing patterns. Applying this hypothesis, Chen (2013) examines whether a language with the obligatory grammatical marking of future events significantly affects the perception of time. Chen (2013) argues that languages with the obligatory grammatical marking of future events would make the future seem more distant, which would cause speakers to discount possible future rewards. The inference can be made that corporations in countries with strong FTR languages engage less in financial fraud (and other behaviors that may compromise the development of a firm) and induce a lower degree of adverse selection and moral hazard. Correspondingly, Graafland and Niels (2020) find that long-term orientation that grammatically marks the future is associated with precautionary activities and higher levels of corporate social responsibility, even when controlling for the country- and firm-level variables, further proving that strong FTR languages are associated with decision-making that is less future-oriented. Sutter et al. (2015) provide experimental support for the effect of FTR on an individual's economic behavior. As discussed, the reduction of financial fraud and the increase in precautionary activities are both conducive to enhancing the stability of a firm's operations and mitigating adverse selection and moral hazard. According to these arguments, the relationship between strong FTR languages and default risk should be positive.

However, language clarity has an association with default risk. Unclear explanations may reduce investors' confidence in their interpretation of disclosure and make the adverse selection more prominent (Bloomfield et al., 1999). Jeanjean et al. (2010, 2015) find that languages such as English that have higher precision when future corporate plans are disclosed can positively affect market reactions by reducing adverse selection. That is, using a strong FTR language can enhance the confidence of investors. Restated, a language with higher precision of beliefs about time can benefit corporations by affecting investor reactions to information disclosure (Brochet et al., 2016). For speakers of weak FTR languages, imprecision of beliefs about time provides more management opportunities to hide unfavorable news or information that does not concern the company when describing the future development of the enterprise, thus increasing asymmetry between management and investors and inducing adverse selection and moral hazard. As imprecision of beliefs about time promotes adverse selection and strong FTR languages have higher precision of beliefs about time, the relationship between such language and default risk should be negative. Thus, Hypothesis 1 is proposed as follows:

Hypothesis 1: The strong FTR language decreases corporate default risk.

Subsequently, three other factors that can affect corporate default risk are examined to discern whether a correlation exists between language and corporate default risk: information transparency, accounting disclosure requirements, and corruption. Transparency can affect corporate default risk through two mechanisms: adverse selection and moral hazard. If the transparency of the market is lower, then investors cannot fully judge the value of companies and are thus only willing to purchase at the average price. This prompts well-run companies to withdraw from the market, leaving it filled with poorly operated companies. Default risk naturally increases because poorly managed companies are more likely to be insolvent. However, companies can ease the adverse selection problem by increasing information transparency (Ho et al., 2020). Increased transparency reduces information asymmetry between informed and uninformed traders, thus reducing inequity between investors and informed traders and mitigating adverse selection and reducing the cost of capital, which can lead to reductions in both costs and default risk (Bushman and Smith, 2003; Hope et al., 2009; Lang, 2011). In addition, higher transparency means more comprehensive management supervision, thus limiting self-interested management behavior and reducing moral hazard. Because transparency can alleviate adverse selection and moral hazard, the effect of language on default risk is moderated or strengthened by transparency. Specifically, if strong FTR languages are associated with higher default risk, transparency curbs this impact; thus, the positive impact of language on default risk is dampened. If strong FTR languages are associated with lower default risk, transparency can serve as a substitute to reduce default risk and thus dampen the effect of strong FTR languages.

Information transparency reflects the accessibility of accounting information, and accounting disclosure requirements reflect information quality. Disclosure requirements benefit both businesses and investors (Ewert et al., 2005; Lambert, 2007). Like transparency, disclosure requirements help resolve adverse selection and moral hazards (Biddle and Hilary, 2006). For example, firms committing to increased levels of the disclosure have lower bid-ask spreads, which reflects reductions in asymmetry and adverse selection (Welker, 1995; Leuz et al., 2000;). Disclosure requirements also affect default risk by reducing moral hazard because they allow less room for self-interested behavior. Thus, accounting disclosure rules can affect corporate operations, such as by increasing investment efficiency (Bushman and Smith, 2001; Healy and Palepu, 2001; Lambert et al., 2007). Because disclosure requirements can reduce corporate default risk by alleviating both adverse selection and moral hazard, the effect of language on default risk is influenced by these requirements. That is, if strong FTR languages are associated with higher default risk, then accounting disclosure requirements curb this impact, thereby dampening the positive impact of language on default risk. However, if strong FTR languages are associated with lower default risk, then disclosure requirements can serve as a substitute, thus reducing default risk and also dampening the effect of such language on default risk. Thus, Hypotheses 2 and 3 are proposed as follows:

Hypothesis 2: The effect of strong FTR language on default risk depends on firms' levels of information transparency.

Hypothesis 3: The effect of strong FTR language on default risk depends on a country's disclosure requirements.

Another noteworthy factor in the study of corporate default risk is corruption. Studies have shown that the legal system affects not only the operation and governance of firms but also the behavior of investors (La Porta et al., 2000; Ball et al., 2000; Rydqvist et al., 2014; Kanagaretnam et al., 2014). In a weak legal system, a corruptive government may weaken investor rights or deter outside investors from participating in corporate governance through loose legal arrangements and keep information opaque by implementing lax accounting standards, resulting in adverse selection and moral hazard, the main causes of default risk (Ho et al., 2019). Furthermore, when corruption prevails among listed companies, chances of fraud increase; this leads to increased operating risks and reduced investor confidence, which both negatively affect default risk (Liu, 2016; Zhang, 2018). Because corruption not only results in poor operation but also increases adverse selection and moral hazard, the effect of FTR languages on default risk is influenced by corruption. That is, if strong FTR languages are associated with higher default risk, then government integrity curbs this impact, thus dampening the positive impact of language on default risk. However, if it is associated with lower default risk, then government integrity has a substitution effect on the language effect, thus alleviating the effect of strong FTR languages. Thus, Hypothesis 4 is proposed as follows:

Hypothesis 4: The effect of strong FTR language on default risk depends on a country's control for corruption.

3. Sample and research design

3.1 Sample construction

Our sample contains 1,084,397 firm-year observations collected from the Thomson Reuters Worldscope database (Datastream). These observations correspond to firms distinguished from ten major industries (INDC2) in 36 countries within 30 years from 1988 to 2017.

To obtain enough observation to calculate the volatility of equity returns and equity beta, only firms whose information on assets is available for five consecutive years and who have more than 200 annual records of daily returns are included. Besides, companies without liability information are excluded from the sample. To avoid problems of outliers, we winsorize all firm-level variables at the 1st to 99th percentiles. Our firm-level and country-level variables are listed in Table 1.

[Insert Table 1 here]

In the final step, we combine country-level statistics with firm-level variables. The information on religion is obtained from Stulz and Williamson (2003), Pevzner et al. (2015), and ICRG (International Country Risk Guide), whereas each country's sub-legal information such as International Accounting Standards score, transparency score, and Anti-corruption index, are collected following Bushman and Smith (2001), Bushman et al. (2004), Kroszner et al. (2006), Christensen et al. (2013), Enikolopov et al. (2014), Leuz and Wysocki (2015), and Ho et al. (2020). After all these steps, our final sample has 180,013 firm-year observations.

3.2 Variable definition

3.2.1 Default risk measure

In the analysis, we provide two PD measurements to capture corporate default risk[5]: one is based on the structural model proposed by Merton (1974), which treats the value of equity as a European call option written based on the underlying assets of a firm (We refer to this PD measure as risk-neutral PD). The value of a firm's total assets is assumed to follow a geometric Brownian motion as follows Equation (1):

$$E = VN(d_1) - e^{-rT}DN(d_2), \tag{1}$$

where V is the value of a firm's assets, E is the market value of a firm's equity, D is the firm level of debt, r is the risk-free rate, T is defined to be one year and $N(.)$ is the cumulative distribution of standard normal. Finally, d_1 and d_2 are two standard normal variables that are defined in Equation (2) and (3), in which σ_A is the standard deviation of return on assets.

$$d_1 = \frac{\ln(\frac{V}{D}) + (r + \frac{\sigma_A^2}{2})(T)}{\sigma_A\sqrt{T}} \tag{2}$$

$$d_2 = d_1 - \sigma_A\sqrt{T} \tag{3}$$

In this standard approach, $N(-d_2)$ is defined as the firm's PD and can be estimated after we gain the necessary information on other variables mentioned previously. To check our results' robustness, we provide an alternative PD measure as physical PD (Lee et al., 2011; Lee and Lin, 2012; Lee et al., 2013; Ho et al., 2020). In this estimation

[5] We also measure corporate stability (low default risk) using Z-scores as in Fazio et al. (2015). The evidence of an alternative proxy of stability has a result similar to that of PD.

procedure, the risk-neutral assumption in Merton's standard approach is relaxed by replacing the risk-free interest rate with the instantaneous expected return on firm assets (μ_A) when calculating default probabilities.

3.2.2 Language measure

When describing future events, the grammatical structures of languages vary (Dahl, 2000). Specifically, a language grammatically marks future events. For example, in English, different tenses are usually expressed through clear grammatical changes. If a speaker wishes to express that it will rain tomorrow, then he or she is more likely to say "it will rain tomorrow" rather than "it rains tomorrow." That is, the tense must be adjusted according to the time of the event (Dahl, 1985, 2000). Another example is French. French has two forms of future tense: *futur proche* and *futur simple*. Speakers of French rely on the inflection of the verb, specifically the ending (i.e., *-ai*, *-as*, *-a*, -*ons*, *-ez*, and *-ont*) in writing and informal speech, whereas *futur proche* takes the form *aller* + the original form of the corresponding action verb and is usually used in spoken language. An example of the opposite is Chinese. When predicting tomorrow's weather, Chinese speakers will say 明天下雨 ("it rains tomorrow"). From a deeper perspective, the Chinese do not strictly divide time, whereas English regards such a division as one criterion for grammatical correctness.[6]

Differences exist even within language families (Dahl, 2000; Thieroff, 2000; Nurse, 2008). Among the Germanic family of languages, for example, English is the only one that requires these tense elements. For example, German speakers naturally express tomorrow's rain in the following words: *Morgen regents* ("it rains tomorrow"). As for Dutch, the form *Zullen* + infinitive is the general future tense, which is its most formal form for referring to future events.

Spanish uses inflectional forms to represent the future (Dahl, 1985, 2000). The first

[6] We thank an anonymous referee for pointing this out.

category is regular inflection. When a verb ends with *-ar*, *-er*, or *-ir*, the original form of the verb is retained. At the end of the verb, Spanish adds *-e*, *-s*, *-emos*, *-is*, or *-n*. The second is irregular inflection. For example, the future tense of "tener" is "tendré." Like Spanish, Portuguese uses inflections when describing future events. The inflections for *ar*/*er*/*ir* are identical, with the following suffixes added to the original form of the verb: *ei*/*s*/*Ø*/*emos*/*eis*/*o*. However, a few verbs have irregular changes: the future form of *dizer* is *direi*.

However, Finnish has a different tense structure. Finnish has no independent future tense, although there are many ways of referring to the future. German uses the present tense with a noticeably greater frequency, although its grammar has two acceptable means of expressing the future (Abbot-Smith and Behrens, 2006). A simple approach to constructing future tense in German is combining *werden* and infinitives. However, in spoken language, the present tense with a time adverbial is often used, such as in *Susanne zieht übermorgen nach Berlin* ("Susanna moves to Berlin the day after tomorrow"), where the present tense is used to describe future events.

In sum, languages differ in how they describe future events. Language can be classified into two categories according to whether the use of future tense is mandatory, as the distinction is between languages that obligatorily mark the future tense and languages that do not obligatorily mark the future tense. For instance, Chen (2013) defines futureless languages as weak FTR languages and their opposites as strong FTR languages.

Throughout the paper, we define whether a language belongs to strong- versus weak-FTR following previous literature on the economic impact of language and on cross-linguistic analyses (e.g., Thieroff, 2000; Bybee et al., 1994; Cyffer et al., 2009; Dahl, 1985; Dahl and Kós-Dienes, 1984; Nurse, 2008, Chen, 2013). In particular, we rely on a specific criterion, developed by the European Science Foundation's Typology of Languages in Europe (EUROTYP) project, in assigning languages into strong- versus weak-FTR groups. If a language does not require "obligatory [FTR] use in (main clause) prediction-based contexts" (Dahl, 2000), then we put this language into the weak-FTR group. On the other hand, if a language does have the above-mentioned requirement, then in our study it belongs to the strong-FTR group.

3.2.3 Control variables

Besides language, we control a wide array of the firm and country characteristics that affect firms' default behaviors. First of all, the firm size is considered because compared to small and medium-sized firms, firms with larger sizes are easier to raise capital with less cost, thus they have lower default risk. Then, firms with higher debt often mean they have a financial burden and greater financial risk. Another factor to be considered is the firm's profitability since firms with a higher profit margin are more likely to repay their loans. The concentration ratio can reflect the dispersion of equity, and it is recognized that the less concentrated equity is favorable for balances of power and make the decision of the company more precise.

As discussed above, we include firm-specific information from DataStream. For instance, *SIZE* is the natural logarithm of total assets, *ROA* is the firm profitability, *TANG* is the firm tangibility ratio, which is calculated as property, plant, and equipment (*PPE*) divided by the book value of total assets, *GROWTH* is the firm growth rate, and *VOL* is the firm volatility computed as the standard deviation of monthly stock return over the prior year (Brogaard et al., 2017). Industry-level includes *HHI* is the concentration ratio, and *LITIGATION* is the high-litigation industry.

Along with firm characteristics, we also control for country-level variables. For instance, *INFLATION* from World Bank is the inflation rate (Fazio et al., 2015), *GGDP* is GDP growth ratios (Meng and Yin, 2019), and *FREEDOM* is the economic freedom of the world (Wang et al., 2016). We construct two indicator variables for the proxy of information availability. TRAN (Bushman, 2004) and ACCOUNT (Leuz and Wysocki, 2015) represent the two aspects of disclosure: one is the transparency score reflecting the information accessibility, while the other one is the accounting score reflecting the accounting disclosure quality. *CPI* is the Anti-corruption index, which reflects the extent of corruption in one country (Ho et al., 2022).

3.3 Research design

To test Hypothesis 1, we estimate the following regression:

$$PD_{it} = \beta_0 + \beta_1 Strong\ FTR_c + \gamma\ X_{it} + \varepsilon_{it} \tag{4}$$

In this baseline regression specification, $PD_{i,t}$ is the default probability of firm *i* in year *t*, which is estimated based on the structural model of Merton (1974) and its modified version developed by Lee et al. (2011), in which a firm's equity value is treated as a European call option. The variable *Strong FTR* measures a country's level of indicator variable equal to 1 if a language obligatorily marks future events *c*. In addition, X_{it} is a vector of control variables that captures various firm- and country-level characteristics, including *SIZE* is the firm size, *LEV* is the firm leverage, *ROA* is the firm profitability, *TANG* is the firm tangibility ratio, *GROWTH* is the firm growth rate, *VOL* is the firm volatility, *HHI* is the concentration ratio, *LITIGATION* is the high-litigation industry, *INFLATION* is the inflation rate, *GGDP* is GDP growth ratios, and *FREEDOM* is economic freedom of the world. In addition to formal institutions (Kim et al., 2021), we also control for national cultural characteristics, such as *POWER DISTANCE, MASCULINITY, UNCERTAINTY AVOIANCE, INDULGENCE*, and *RELIGIOUS*. Finally, we control the industry-, year-, and Language family fixed effect (Roberts and Winters, 2013; Roberts et al., 2015).

To test Hypotheses 2 and 3, we estimate regressions (5) and (6), respectively:

$$\begin{aligned} PD_{it} = {} & \vartheta_0 + \vartheta_1 Strong\ FTR_c + \vartheta_2 TRAN_c + \vartheta_3 TRAN_c \times Strong\ FTR_c \\ & + \gamma\ X_{it} + \varepsilon_{it} \end{aligned} \tag{5}$$

$$\begin{aligned} PD_{it} = {} & \alpha_0 + \alpha_1 Strong\ FTR_c + \alpha_2 ACCOUNT_c \\ & + \alpha_3 ACCOUNT_c \times Strong\ FTR_c + \gamma\ X_{it} + \varepsilon_{it} \end{aligned} \tag{6}$$

To test Hypothesis 4, the following expression is estimated:

$$PD_{it} = \pi_0 + \pi_1 Strong\ FTR_c + \pi_2 CPI_{ct} + \pi_3 CPI_{ct} \times Strong\ FTR_c + \gamma\ X_{it} + \varepsilon_{it} \qquad (7)$$

In regressions (5-7), three dummy variables: *TRAN* (Transparency score), *ACCOUNT* (Accounting score), and *CPI* (Corruption Perception Index), with values equal to 1 if a country's transparency score and accounting score are high. These three binary variables all reflect the condition of information disclosure; one is measured at the firm level, and the other two are measured at the country level.

Our focus in regressions (5-7) is the coefficient between Strong FTR and default risk, which represents the casual effect that strong FTR language has on corporate default risk. In addition, the interactive term in each regression is expected to be opposite to the coefficient between Strong FTR and corporate default risk as the institutional environment serves as a moderating factor.

Finally, to address the concern that default patterns vary across industries (Chava and Jarrow, 2004; Koopman et al., 2012) and the impact of the business cycle, we control for industry, year, and language family fixed effects in all regressions. Heteroscedasticity is also considered, and all standard errors are adjusted and clustered at the country level.

3.4 Summary statistics

Table 2 summarizes country distributions of firm-year observations and some main variables in our sample. The means of risk-neutral and physical PD vary largely across countries, with a low risk-neutral (physical) PD mean of 0.01 (0.01) and a high risk-neutral (physical) PD mean of 0.21 (0.15).

[Insert Table 2 here]

Table 3 presents the distribution of key variables and their average use by country. The mean of Strong FTR is 0.53, indicating that 53% of the sample uses a language that requires the marking of future events. For the other variables of firm-level characteristics, the mean log size of a firm is 20.87, and the mean leverage ratio 0.49. Ownership concentration is at a low level with a median of 0.11.

[Insert Table 3 here]

Table 4 reports the variance inflation factor and correlations of the main variables in this study. Across our full sample, all variables, particularly those representing a country's legal and economic conditions, are not almost perfectly correlated with language, indicating that the evolution of language is not mainly driven by a country's institutions or economic development and that these cannot fully explain the effect of language.

[Insert Table 4 here]

4. Empirical results

4.1 Baseline regression results

Hypothesis 1 is tested by estimating the regression (4), and the results are presented in Table 5. The coefficient β reveals the relationship between strong FTR and default risk, which is negative and significant at any conventional statistical levels (i.e., 10%, 5%, or 1%), suggesting that firms in countries with strong FTR languages have lower default risk than those in countries with weak FTR languages. This effect is not only statistically and economically significant. Other things equal, strong FTR languages decrease risk-neutral and physical PD by 0.0034 and 0.0059, which accounts for 8.5% and 11.8% of their sample means (i.e., 0.04 and 0.05), respectively. These results support Hypothesis 1.

[Insert Table 5 here]

The second column, where the dependent variable is risk-neutral PD, reveals that using a strong FTR language is still negatively related to risk-neutral PD at the 1% significance level. This result provides additional evidence that the use of a strong FTR language is associated with lower default risk. This result is consistent with H1 and indicates that firms in countries with strong FTR languages have lower default risk than those in countries with weak FTR languages.

Of the other variables, firm size is negatively associated with lower default risk. Larger firms are more capable of dealing with insolvency, and leverage level is associated with higher default risk because liabilities increase the possibility that assets are insufficient to repay debt. Furthermore, a higher gross GDP suggests lower default risk because it reflects the degree of economic development and is associated with larger capital markets where firms generally develop well. Finally, firms with a higher ROA, higher growth, and lower PPE ratio tend to have lower default risk.

4.2 Robustness tests

4.2.1 Alternative measure of language FTR

In this section, this paper investigates whether the results hold under different alternative classifications of language FTR. We follow Chen's (2013) classify language into strong- and weak-FTR languages in our baseline regression. Frist, we use long-term orientation (LTO) as our first alternative measure of language FTR. The societal level measure of LTO is from https://www.hofstede-insights.com/product/compare-countries/. Countries represented by LTO culture emphasize the preparation for the future, LTO is represented by but not limited to persistence and thrift (Hofstede and Minkov, 2010; Kong et al., 2021). As shown in the Panel A of Table 6, the results show the LTO for the alternative classifications of language FTR are independent variables. Both the coefficients of LTO are positively associated with risk-neutral (0.0003) and physical PD (0.0004) at the 0.01 significant level and with a larger economic magnitude, indicating that firms in countries with LTO culture have higher default risk.

Moreover, we adopt other continuous variables in the sensitivity analysis. This paper designs two indexes (SENTENCE and VERB) of FTR in each language based on a word frequency analysis of the text of weather forecasts retrieved from the web. Following Chen (2013) and Kim et al. (2021), we use sentence ratio (SENTENCE) and verb ratio (VERB) as our second and third alternative measures of language FTR. Sentence (VERB) is defined as the number of sentences (verbs) that are grammatically future-marked, divided by the total number of sentences (verbs). A higher value indicates a strong FTR. Regression results are presented in Panel B of Table 6. Both SENTENCE and VERB are negatively associated with risk-neutral and physical PD at the 0.01 significant level, suggesting that a higher percentage of grammatically future-marked sentences and verbs lead to default risk activities. Overall, Table 6 suggests the relation between Strong FTR and default risk, supporting our baseline regression.

[Insert Table 6 here]

4.2.2 Alternative measure of default risk

To ensure the robustness of our findings, we re-run our baseline regression with alternative measures of default risk. Previous literature suggests that the distance to default is a good measure of the default risk (Bharath and Shumway, 2008; Brogaard et al., 2017; Ho et al., 2019). Hence, we consider the distance to default measure (*DD1* and *DD2*) with a much better distribution which will no longer distort the coefficients of interest, and two standard normal variables are defined in Equations (2) and (3). The two DD calculation methods are to capture firm's financial stability but with different emphasis. The higher distance to default, the lower probability of bankruptcy.

We re-test Hypothesis 1 by estimating the regression provided in Reg. (4) and the results presented in Table 7. The dependent variables are DD1 and DD2. The coefficient of Strong FTR in DD1 is 0.4893 (t-stat = 7.47), which is statistically significant, suggesting that firms in countries with Strong FTR language engage in more distance to default. In the next column, the coefficient on Strong FTR is 0.4494 (t-stat = 6.84), and the significance level remains the same. The results show that the coefficients of Strong FTR are all significantly positive in the total sample. The influence of Strong FTR language is significant not only statistically but also economically. The result in Table 7 is similar to the results of Table 5, and it further supports our hypothesis

[Insert Table 7 here]

4.2.3 Within-country evidence

4.2.3.1 Within-country evidence: Belgium

Our regression results might be biased because of an omitted variable between Strong FTR language and default risk. Due to the multicollinearity problem, we cannot add country-fixed effects to control the time-invariant country characteristics. The current design works for firms operating in several regions (or just one region) that speak a common language. To address these issues raised, we conduct within-country tests by controlling for all observable and unobservable country-level attributes and directly address the concern about country-level omitted-variable bias (Kong et al., 2021).

Belgium has 11 provinces, and each province uses a dominant language of either Dutch or French. Belgium, therefore, provides an interesting opportunity to examine the effect of language within one country, holding country-specific institutional characteristics fixed. Following Chi et al. (2021), we create a dummy variable Strong FTR language equal to 0 if a firm is located in a Dutch-speaking province and 1 if a firm is located in a French-speaking province. We re-run the regression of Eq. (4) without the country-level controls. The result, reported in Table 8, shows that default risks in Strong FTR regions are significantly lower than in weak FTR regions.

[Insert Table 8 here]

4.2.3.2 Within-country evidence: U.S.

While conducting our cross-country analysis, we made an assumption that managers within the same country speak the same language. However, it's important to acknowledge that this assumption may not hold true in all cases, especially if the managers are immigrants from foreign countries. Their perceptions and decision-making can be influenced by their native languages, which may impact our findings. Additionally, our cross-country evidence may be subject to the omitted variable bias if some unobservable country-specific factors influence both the language FTR and firm default risk. To mitigate these potential concerns, we performed a within-country analysis using only US firms.

To identify the nationalities of US firm CEOs, we followed the methodology used by Kim et al. (2017) and Kong et al. (2021), which involved collecting their birthplaces and determining whether their native language was a Strong-FTR or Weak-FTR language. Specifically, we used Strong-FTR to refer to nationalities whose native language has a strong future-time-reference (FTR), while Weak-FTR refers to nationalities whose native language has a weak FTR. We then used a dummy variable, Strong-FTR CEO, as the main variable of interest in our within-country analysis. This variable takes a value of one if the CEO's nationality is associated with a Strong-FTR language and zero otherwise. In our sample, we found 5,650 observations with Strong-FTR CEOs, and our sample contained 21,326 firm-year observations.

The results of our within-country analysis are presented in columns (1) and (2) of Table 9. As shown in the table, the coefficient of CEO Strong FTR is negative and statistically significant, with the controlling of various firm characteristics, time, industry, and language family fixed effects. This confirms our baseline finding that managers speaking weak-FTR languages are associated with higher firm default risk.

[Insert Table 9 here]

4.2.4 Endogeneity

4.2.4.1 Difference in differences

In cross-country studies like this one, endogeneity is a common concern. It arises when factors that potentially influence the formation of language affect a firm's default behavior but are not adequately controlled for. Although this is always a potential limitation, it is considered unlikely that the results of this research are entirely driven by any omitted variables. First, the analysis explicitly incorporates relevant factors such as GDP, accounting regulations, and the strength of investor protections. The results are robust after including these variables in the regression. Second, as Pevzner et al. (2015) argue, if omitted variables bias the results, then they must be able to account for more than the main relationship between social culture and corporate default risk. For instance, one may argue that fair legislation and effective legal enforcement drive both higher interpersonal language and lower corporate default risk. However, this argument cannot explain why the negative relationship between language and default risk is more significant in countries with laxer accounting standards and firms with less information disclosure.

One non-English speaking country may use English for financial reporting (or vice versa). An event in which a country exogenously changes its FTR language structure would provide an important natural experiment on the effect of language structure on corporate default behavior. Following Chen et al. (2017), this research uses the event of the transfer of sovereignty over Hong Kong from the U.K. (Strong FTR language) to China (Weak FTR language). Hong Kong is an indicator variable for Hong Kong companies, i.e., firms headquartered in Hong Kong, and zero otherwise. In addition, we also follow Chi et al. (2021) to compare default risk in Hong Kong before and after 1997. The sample comprises 22,062 from 1994 to 2001. Table 10 shows the results, and we find that the interaction effect of Hong Kong * post-1997 is significantly positive. They suggest that firms in Hong Kong increased their corporate default risk after the handover. This finding is consistent with H1 claims that strong-FTR language demotes default risk.

[Insert Table 10 here]

4.2.4.2 Instrumental Variables: Two-Stage Least Squares

To alleviate the endogeneity concern and establish a causal interpretation for our finding that default risks are significantly higher in regions with a language that does not involve grammatical marking of time, the two-stage least squares (2SLS) method is employed, in which a country's primary religious belief is used as the driving force shaping the structure of the language used. Following La Porta et al. (1997), Stulz and Williamson (2003), Guiso et al. (2006, 2008), and Pevzner et al. (2015), we take the choice of religious belief as exogenous as it should not be related to a firm's default behavior *ex-ante*. This statement is not intended to suggest that religiosity is exogenous to corporate default risk. In fact, evidence provided by Hilary and Hui (2009), Chen et al. (2016), and Adhikari and Agrawal (2016) indicates that religiosity is negatively related to risk exposure, and religious values may help to constrain opportunistic behavior. However, no evidence suggests a correlation between religion and firm risk, which is reasonable because what matters should be the high ethical standards embodied in religious values, whereas ethical influence should exist regardless of religion. Seven dummy variables are established to capture five world religions (Catholicism, Protestantism, Islam, Buddhism, and Hinduism) and two belief systems common in Hong Kong. These indicator variables (excluding Hinduism) are used as instrumental variables in the first-stage regression.

The 2SLS regression results are presented in Table 11. In Panel A, Model 1 contains the first-stage regression results for the association between language and religion. Consistent with previous studies, overall, the six variables are significantly associated with language, thereby proving the relevancy of religion. Models 2 and 3 report the second-stage regression results, where *language* in the baseline regression is replaced with the predicted value from the first-stage regression. The *language* used has a significantly negative effect on both measures of firm default risk, suggesting a causal relationship between language and firm default probability. In summary, the results are robust to endogeneity alleviation. We also use the generalized method of moments (GMM) panel estimation method to address endogeneity concerns

(Chhaochharia et al., 2012). The results are presented in Panel B of Table 11, and are consistent with those in Table 5.

[Insert Table 11 here]

4.3 Dependence of the effect of language on information asymmetry

In this subsection, Hypothesis 2 is tested to determine whether the effect of language on default risk depends on the level of transparency by estimating the regression expressed as regression (5). In this regression, information on a country's transparency score is used to conduct variable *TRAN* to serve as a proxy for the general condition of information accessibility. The result is presented in Table 12 with two measures of default risk: risk-neutral PD and physical PD.

The negative relationship between language and default risk is still also in the two regressions that measure default risk. The coefficient of proxy of the level of transparency is significantly negative. That is, both the use of a strong FTR language and transparency reduce default risk. Thus, it can be inferred that the effects of the level of transparency and language may act as substitutions for each other. The interaction term is positive, suggesting that the effect of transparency on default risk is dampened when a strong FTR is used. This supports Hypothesis 2 by providing evidence that using a strong FTR language has a substitution effect on transparency in reducing corporate default risk.

[Insert Table 12 here]

Hypothesis 3 is then tested to examine whether the impact of language varies with the level of accounting requirements by estimating the regression expressed as regression (6). In this regression, information on a country's accounting standard is used to construct the variable *ACCOUNT* to serve as a proxy for the level of requirements. The results are presented in Table 13. As in the preceding estimation, corporate default risk was measured as risk-neutral PD and physical PD. The coefficient of language remains negative and significant after controlling for the overall information level of a country and factors that potentially affect the default risk of a corporation.

Considering the coefficient for the level of required accounting disclosure and the interaction term of this level of disclosure and the language structure, the coefficient of the proxy for required accounting disclosure is negative and the interaction term is positive. The negative coefficient indicates that stronger accounting requirements reduce default risk. Because both strong FTR languages and disclosure requirements are negatively correlated to default risk, it can be inferred that they have substitution effects on each other. As predicted, the interaction terms in both models are positive and significant, which means that the effect of language on corporate behavior is decreased when the language is a strong FTR language. This supports Hypothesis 3 by providing evidence that the use of a strong FTR language has a substitution effect for accounting disclosure requirements in reducing corporate default risk.

[Insert Table 13 here]

4.4 Dependence of the effect of language on corruption

Next, by estimating the regression expressed as regression (7), Hypothesis 4 is tested to examine whether the degree of government corruption affects the relationship between language and corporate default risk. This regression uses CPIs obtained from Transparency International to construct the variable CPI to serve as a proxy for the level of corruption in a country. A lower CPI value represents a more corrupt government.

The results are presented in Table 14. In Model 1, where the dependent variable is risk-neutral PD, the coefficient between the use of a strong FTR language and risk-neutral PD is negative and significant, and the coefficient in Model 2, where the dependent variable is physical PD, remains significantly negative, which supports the finding that use of a strong FTR language is associated with lower default risk. In addition, the coefficient of the proxy of corruption is negative, meaning a lower level of corruption reduces corporate default risk. Both strong FTR languages and government integrity are negatively correlated with default risk. The positive and significant coefficients of the interaction terms in Models 1 and 2 support Hypothesis 4 by suggesting that the use of a strong FTR language has a substitution effect on government integrity in reducing corporate default risk.

[Insert Table 14 here]

4.5 Mechanism analysis

As mentioned in the H1, the positive relationship between weak FTR language and the default risk may be conducted by the "moral hazard effect". However, the internal mechanism is still uncertain. Now, we try to analyze the leverage and stock return when a country uses Strong FTR language further to test the Strong FTR language's internal role to increase default risk.

First, a highly mature financial system promotes the rate of capital accumulation, improves the efficiency in the use of physical capital, effectively balances the macro leverage, and stabilizes financial markets (Lee et al., 2011). Thus, in a conventional environment, an advanced financial system helps reduce corporate default risk. However, in a global financial crisis, a well-established financial system makes it more costly for risk averters to repay their non-contingent liabilities (Ho et al., 2020) or quickly downgrade the credit rating of enterprises in operational difficulties, and the default probability of enterprises is further increased. Hence, weak FTR language provides more firms leverage, thus increasing asymmetry between management and investors and inducing adverse selection and moral hazard. Second, previous studies suggest that the firm uses less future tense to offer higher returns since they are riskers (Karapandza, 2016). This research is based on a risk framework, and further investigates the impact of the strong FTR language on the stock return of enterprises at the micro-level. This study replaces the explained variable in Eq. (1) with the *LEV* and the *RET*, and the renewed regression results are presented in Table 15. The above results in Table 15 show that in countries with the Strong FTR language, the stock price return increased significantly, and the firm's leverage has dropped significantly. The above results support the path of "moral hazard effect".

[Insert Table 15 about here]

5. Conclusion

While weak FTR languages do not require speakers to mark future events grammatically, strong FTR languages do. The extant studies have demonstrated that the use of weak- or strong FTR languages affects future-oriented behaviors: individuals' perception of saving and health behaviors (Chen, 2013), earnings management (Kim et al., 2017), precautionary cash holdings (Chen et al., 2017), corporate social responsibility (Liang et al., 2018), corporate innovations (Kong et al., 2021), corporate R&D (Chi et al., 2021), and investment efficiency (Kim et al., 2021). This study differs from the extant ones by selectively focusing on corporate default risk, one of the most important research topics in accounting, economics, and finance in the aftermath of the late 2000s global financial crisis (see, e.g., Jia et al., 2020 and the references therein).

Specifically, we attempt to investigate whether language, as an informal institution, has any impact on corporate default risk by shaping inventors' perception of long-term innovation. Given that strong FTR languages increase the psychological distance from the future and make the timing of future rewards more accurate, inventors speaking strong FTR language perceive a lower present value of innovation projects than their weak FTR language counterparts.

Using a comprehensive sample of firms in 36 countries with 180,013 observations spanning from 1988 to 2017, we show that corporate default risk is significantly higher in regions dominated by speakers of weak FTR languages. Moreover, the FTR effect on default risk is weakened in countries with stronger formal institutions (e.g., high disclosure quality, greater transparency, and less corruption). This provides additional support to the findings that identify a relationship between the use of strong FTR languages and default risk. To alleviate the potential endogeneity problem, we adopt the 2SLS method and find a robust relationship between the use of a strong FTR language and default risk.

We contribute to a small but growing literature on how language affects corporate behaviors and business outcomes. In addition to the existing factors of corporate default risk in the existing literature (see, e.g., Jia et al., 2020 and the references therein), we argue that language can be an informal institution that affects corporate default risk.

Availability of data: Data is available on request from the authors.

References

Adhikari, B. K., and Agrawal, A. (2016). Does local religiosity matter for bank risk-taking? *Journal of Corporate Finance*, 38, 272-293.

Aggarwal, R., Faccio, M., Guedhami, O., & Kwok, C. C. (2016). Culture and finance: An introduction. *Journal of Corporate Finance*, (41), 466–474.

Agrawal, A. K. (2013). The impact of investor protection law on corporate policy and performance: evidence from the blue-sky laws. *Journal of Financial Economics*, 107(2), 417–435.

Bakshi, G., Gao, X., & Zhong, Z. (2022). Decoding default risk: A review of modeling approaches, findings, and estimation methods. *Annual Review of Financial Economics*, 14, 391-413.

Ball, R., Kothari, S. P., and Robin, A. (2000). The effect of international institutional factors on properties of accounting earnings. *Journal of Accounting and Economics*, 29(1), 1-51.

Beaver, W. H. (1966). Financial ratios as predictors of failure. *Journal of Accounting Research*, 4, 71–111.

Berlin, B. & Kay, P. (1969). *Basic color terms: Their universality and evolution*. Berkeley: University of California Press.

Bharath, S. T., and Shumway, T. (2008). Forecasting default with the Merton distance to default model. *Review of Financial Studies*, 21 (3), 1339-1369.

Biddle, G. C., and Hilary, G. (2006). Accounting quality and firm-level capital investment. *Accounting Review*, 81(5), 963–982.

Bloomfield, R. J., Libby, R., and Nelson, M. W. (1999). Confidence and the welfare of less-informed investors. *Accounting Organizations and Society*, 24(8), 623–647.

Bonfim, D. (2009). Credit risk drivers: Evaluating the contribution of firm level information and of macroeconomic dynamics. *Journal of Banking and Finance*, 33(2), 281–299.

Boroditsky, L. (2001). Does language shape thought?: Mandarin and English speakers' conceptions of time. *Cognitive psychology*, *43*(1), 1–22.

Boroditsky, L. (2011). How language shapes thought. *Scientific American*, *304*(2), 62–65.

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

Brogaard, J., Li, D., and Xia, Y. (2017). Stock liquidity and default risk. *Journal of Financial Econo*mics, 124(3), 486–502.

Bushman, R. M., and Smith, A. J. (2001). Financial accounting information and corporate governance. *Journal of Accounting and Economics*, 32 (3), 181-235.

Bushman, R. M., Piotroski, J. D., and Smith, A. J. (2004). What determines corporate transparency? *Journal of Accounting Research*, 42 (2), 207-252.

Bybee, J., Perkins, R. and Pagliuca, W. (1994). The Evolution of Grammar, Chicago: University of Chicago Press.

Chava, S. and Jarrow, R. A. (2004). Bankruptcy prediction with industry effects. *Review* *of Finance*, 8, 517-549.

Chen, F., Hope, O.-K., Li, Q., and Wang, X. (2011). Financial reporting quality and investment efficiency of private firms in emerging markets. *Accounting Review*, 86(4), 1255–1288.

Chen, H., Huang, H. H., Lobo, G. J., and Wang, C. (2016). Religiosity and the cost of debt. *Journal of Banking and Finance*, 70, 70-85.

Chen, M.K., (2013). The effect of language on economic behavior: evidence from savings rates, health behaviors, and retirement assets. *American Economic Review* 103 (2),690–731.

Chen, S., Cronqvist, H., Ni, S. X., and Zhang, F. (2017). Languages and corporate savings behavior. *Journal of Corporate Finance*, 46, 320–341.

Chhaochharia, V., Kumar, A., and Niessen-Ruenzi, A. (2012). Local investors and corporate governance. *Journal of Accounting and Economics*, 52, 42-67.

Chi, J. D., Su, X., Tang, Y., & Xu, B. (2021). Is language an economic institution? Evidence from R&D investment. *Journal of Corporate Finance*, *62*, 101578.

Christensen, H. B., Luzi, H., and Christian, L. (2013). Mandatory IFRS reporting and changes in enforcement. *Journal of Accounting and Economics*, 56 (2-3), 147-177.

Cyffer, N., Ebermann, E., and Ziegelmeyer, G. (2009). Negation Patterns in West African Languages and Beyond. Amsterdam: John Benjamins Publishing Company.

Dahl, Ö and Kós-Dienes, D. (1984). Selected Working Papers from the Tense-Mood-Aspect Project. Stockholm: Institute of Linguistics, University of Stockholm.

Dahl, Ö. (1985). Tense and Aspect Systems. Oxford: Basil Blackwell.

Dahl, Ö. (2000). The grammar of future time reference in European languages. In O. Dahl (Ed.) *Tense and Aspect in the Languages of Europe*, (pp. 309–328). Berlin: Mouton de Gruyter.

Dahl, Ö. (2000). The Grammar of Future Time Reference in European Languages. In Tense and Aspect in the Languages of Europe, edited by Östen Dahl, 309–28. Berlin: Mouton de Gruyter.

Enikolopov, R., Petrova, M., and Stepanov, S. (2014). Firm value in crisis effects of firm-level transparency and country-level institutions. *Journal of Banking and Finance*, 46, 72-84.

Ewert, R., and Wagenhofer, A. (2005). Economic effects of tightening accounting standards to restrict earnings management. *Accounting Review*, 80(4), 1101–1124.

Fazio, D. M., Tabak, B. M., and Cajueiro, D. O. (2015). Inflation targeting: Is it to blame for banking system instability? *Journal of Banking and Finance*, 59, 76-97

Flammer, C. and Bansal, P. (2017) Does a long-term orientation create value? Evidence from a regression discontinuity. *Strategic Management Journal*, 38(9), 1827–1847.

Ge, Y., Kong, X., Dadilabang, G., Ho, K. (2020). The effect of Confucian culture on household risky asset holdings: Using categorical principal component analysis. *International Journal of Finance and Economics,* forthcoming.

Giesecke, K., Longstaff, F. A., Schaefer, S., and Strebulaev, I. (2011). Corporate bond default risk: A 150-year perspective. *Journal of Financial Economics*, 102 (2), 233-250.

Gotti, G., Roberts, S. G., Fasan, M., & Robertson, C. B. (2021). Language in economics and accounting research: The role of linguistic history. *The International Journal of Accounting*, 56(03), 2150015.

Graafland, J., and Niels, N. (2020). Culture and institutions: How economic freedom and long-term orientation interactively influence corporate social responsibility. *Journal of International Business Studies*, 51, 1034-1043.

Grunert, J., Norden, L., and Weber, M. (2005). The role of non-financial factors in internal credit ratings. *Journal of Banking and Finance*, 20 (2), 509-531.

Guiso, L., Sapienza, P., and Zingales, L. (2006). Does culture affect economic outcomes? *Journal of Economic Perspectives*, 20 (2), 23-48.

Guiso, L., Sapienza, P., and Zingales, L. (2008). Social capital as good culture. *Journal of the European Economic Association*, 6 (2-3), 295–320.

Hail, L., and Leuz, C. (2005). International differences in the cost of equity capital: Do legal institutions and securities regulation matter? *Journal of Accounting Research*, 44(3), 485–531.

Hickmann, M. (2000). Linguistic relativity and linguistic determinism: some new directions. *Linguistics*, 38(2), 409-434.

Hilary, G., and Hui, K. W. (2009). Does religion matter in corporate decision making in America? *Journal of Financial Economics*, 93 (2), 45-473.

Ho, K. C., Ma, Z., Yang, L., and Shi, L. (2019). Do anticorruption efforts affect banking system stability? *Journal of International Trade & Economic Development*, 28, 3, 277-298.

Ho, K. C., Yao, C., Zhao, C., and Pan, Z. (2022). Modern health pandemic crises and stock price crash risk. *Economic Analysis and Policy*, 74, 448-463.

Ho, K. C., Yen, H., Gu, Y., and Shi, L. (2020). Does societal trust make firms more trustworthy? *Emerging Markets Review*, 42, 10067.

Hofstede, G., & Minkov, M. (2010). Long- versus short-term orientation: New perspectives. *Asia Pacific Business Review*, 16(4), 493-504.

Jeanjean, T., Lesage, C., and Stolowy, H. (2010). Why do you speak English (in your annual report)? *International Journal of Accounting*, 45(2), 200–223.

Jeanjean, T., Stolowy, H., Erkens, M., and Yohn, T. L. (2015). International evidence on the impact of adopting English as an external reporting language. *Journal of International Business Studies*, 46(2), 180–205.

Jia, Z., Shi, Y., Yan, C., & Duygun, M. (2020). Bankruptcy prediction with financial systemic risk. *European Journal of Finance*, *26*(7-8), 666-690.

Kanagaretnam, K., Lim, C. Y., and Lobo, G. J. (2014). Effects of international institutional factors on earnings quality of banks. *Journal of Banking and Finance*, 39(2), 87–106.

Kaplan, S. and Orlikowski W.J. (2013) Temporal work in strategy making. *Organzation Science*, 24(4), 965–995.

Karapandza, R. (2016). Stock returns and future tense language in 10-K reports. *Journal of Banking and Finance*, 71, 50-61.

Kay, P., & Regier, T. (2006). Language, thought and color: recent developments. *Trends in Cognitive Sciences*, 10(2), 51-54.

Kim, J., Kim, Y., & Zhou, J. (2017). Languages and earnings management. *Journal of Accounting and Economics*, *63*(2-3), 288-306.

Kim, J., Kim, Y., & Zhou, J. (2021). Time encoding in languages and investment efficiency. *Management Science*, *67*(4), 2609-2629.

Kong D, Wang J, Wang Y, & Zhang J. (2021). Language and innovation. *Journal of Business Finance and Accounting*, forthcoming.

Koopman, S. J., Lucas, A. and Schwaab, B. (2012). Dynamic factor models with macro, frailty, and industry effects for US default counts: the credit crisis of 2008. *Journal of Business and Economic Statistics*, 30 (4), 521-532.

Kroszner, R., Laeven, L., and Kingebiel, D. (2006). Banking crises, financial dependence, and growth. *Journal of Financial Economics*, 84 (1), 187-228.

La Porta, R., Lopez-de-Silanes, F., Shleifer, A., and Vishny, R. (1997). Trust in large organizations. *American Economic Review*, 87 (2), 333–338.

La Porta, R., Lopez-de-Silanes, F., Shleifer, A., and Vishny, R. (2000). Investor protection and corporate governance. *Journal of Financial Economics*, 58(12), 3-27.

Lambert, R. A., Leuz, C., and Verrecchia, R. E. (2007). Accounting information, disclosure, and the cost of capital. *Journal of Accounting Research*, 45(2), 385–420.

Lang, M. H., Lins, K. V., and Maffett, M. G. (2012). Transparency, liquidity, and valuation: International evidence on when transparency matters most. *Journal of Accounting Research*, 50(3), 729–774.

Lee, S. C., and Lin, C.T. (2012). Book-to-market equity, operating risk, and asset correlations: Implications for Basel capital requirement. *Journal of International Financial Markets, Institutions, and Money*, 22 (4), 973-989.

Lee, S. C., Lin, C. T., and Yang, C. K. (2011). The asymmetric behavior and procyclical impact of asset correlations. *Journal of Banking and Finance*, 35, 2559-2568.

Lee, S. C., Lin, C. T., and Yu, M. T. (2013). Book-to-market equity, asset correlations, and the Basel capital requirement. *Journal of Business Finance and Accounting*, 40 (7-8), 991-1008.

Leuz, C., and Verrecchia, R. E. (2000). The economic consequences of increased disclosure. *Journal of Accounting Research,* 38, 91–124.

Leuz, C., and Wysocki, P. D. (2015). The economics of disclosure and financial reporting regulation: Evidence and suggestions for future research. *Journal of Accounting Research,* 54(2), 525–622.

Liang, H., Marquis, C., Renneboog, L., & Sun, S. L. (2018). Future-time framing: The effect of language on corporate future orientation. *Organization Science*, *29*(6), 1093-1111.

Liu, X. (2016). Corruption culture and corporate misconduct. *Journal of Financial Economics*, 122(2), 307–327.

Meng, Y., and Yin, C. (2019). Trust and the cost of debt financing. *Journal of International Financial Markets, Institutions and Money*, 59, 58-73.

Merton, R. C. (1974). On the pricing of corporate debt: The risk structure of interest rates. *Journal of Finance*, 29 (2), 449-470.

North, D. C. (1991). Institutions. *Journal of Economic Perspectives*, *5*(1), 97–112.

Nurse, D. (2008). Tense and Aspect in Bantu. New York: Oxford University Press.

Ohlson, J.A. (1980). Financial ratios and the probabilistic prediction of bankruptcy. *Journal of Accounting Research*, 18 (1), 109-131.

Pevzner, M., Xie, F., and Xin, X. (2015). When firms talk, do investors listen? The role of trust in stock market reactions to corporate earnings announcement. *Journal of Financial Economics*, 117 (1), 190-223.

Richard, M., and Toffoli, R. (2009). Language influence in responses to questionnaires by bilingual respondents: A test of the Whorfian hypothesis. *Journal of Business Research*, 6 (10), 987-994.

Roberts, S. G., Winters, J., and Chen, K. (2015). Future tense and economic decisions: Controlling for cultural evolution. *Plos One*, 10(7), e0132145.

Roberts, S., and Winters, J. (2013). Linguistic diversity and traffic accidents: Lessons from statistical studies of cultural traits. *Plos One*, 8(8), e70902.

Rydqvist, K., Spizman, J. D., and Strebulaev, I. A. (2014). Government policy and ownership of equity securities. *Journal of Financial Economics*, 111(1), 70–85.

Stulz, R. M., and Williamson, R. (2003). Culture, openness, and finance. *Journal of Financial Economics*, 70 (3), 313-349.

Sutter, M., Angerer, S., and Glätzle-Rützler, D. (2015). The effect of language on economic behavior: Experimental evidence from children's intertemporal choices. *Cesifo working papers*.

Thieroff, R. (2000). On the Areal Distribution of Tense-Aspect Categories in Europe. In Tense and Aspect in the Languages of Europe, edited by Östen Dahl, 309–28. Berlin: Mouton de Gruyter.

Wang, M., Rieger, M. O., and Hens, T. (2016). How time preferences differ: Evidence from 53 countries. *Journal of Economic Psychology*, 52(52), 115–135.

Welker, M. (1995). Disclosure policy, information asymmetry, and liquidity in equity markets. *Contemporary Accounting Research*, 11(2), 801–827.

Whorf, B. L., Carroll, J. B., and Chase, S. (1956). Language, thought and reality. *American Sociological Review*, 21(5), 643.

Williamson, O. E. (2000). The new institutional economics: Taking stock, looking ahead. *Journal of Economic Literature*, *38*(3), 595–613.

Xu, Y. (2018). Anti-corruption regulation and firm value: Evidence from a shock of mandated resignation of directors in China. *Journal of Banking and Finance*, 92(7), 67–80.

Zhang, J. (2018). Public governance and corporate fraud: Evidence from the recent anti-corruption campaign in China. *Journal of Business Ethics*, 148(2), 375–396.

Table 1. Variable definitions

This table defines each dependent and independent variable used in the empirical analysis.

Variable	Explanation
Risk-neutral PD	Probability of Default (Brogaard et al., 2017).
Physical PD	Objective Probability of Default (Ho et al., 2020).
DD1	distance to default 1: Standard normal variables which are defined in Equation (2) (Ho et al., 2019).
DD2	distance to default 2: Standard normal variables which are defined in Equation (3) (Brogaard et al., 2017).
	Country factor variables
Strong FTR	Indicator variable that equals 0 if a language is a weak future time reference language, that is, it does not differentiate the present and the future obligatorily, and 1 otherwise.
TRAN	An indicator variable equal to one if a county's CIFAR (Center for International Financial Analysis Research) Index of transparency (see, e.g., Bushman et al. 2004) is above the median and zero otherwise.
ACCOUNT	An indicator variable equal to one if a county's accounting index (Leuz and Wysocki, 2015) is above the median and zero otherwise.
CPI	An indicator variable equal to one if a county's CPI index (ICRG) is above the median and zero otherwise.
Catholic, Protestant, Muslim, Buddhist, Hindu, and Indigenous	Catholic, Protestant, Muslim, Buddhist, and Hindu are indicator variables capturing whether a country's primary religious belief one of five religions. Indigenous is an indicator variable capturing Hong Kong's local religious belief.
GGDP	GDP growth ratio (Meng and Yin, 2019).
INFLATION	Inflation rate is the annual rate of change on the consumer price index. Source: World Bank.
FREEDOM	Economic Freedom of the World (EFW) datasets.
POWER DISTANCE	Power distance (Hofstede, 2001).
MASCULINITY	Masculinity (Hofstede, 2001).
UNCERTAINTY AVOIANCE	Uncertainty avoidance (Hofstede, 2001).
INDULGENCE	Indulgence (Hofstede, 2001).
RELIGIOUS	Religious (ICRG).
SENTENCE	The number of sentences that are grammatically future-marked, divided by the total number of sentences (Kim et al., 2021).
VERB	The number of verbs that are grammatically future-marked, divided by the total number of sentence verbs (Kim et al., 2021).
LTO	long-term orientation (Hofstede and Minkov, 2010).
	Industry characteristics
HHI	The Herfindahl-Hirschman index is computed as the sum of squared market shares.

LITIGATION	Dummy variable equal to one if a firm operates in a high-litigation industry (SIC codes 2833–2836, 3570–3577, 3600–3674, 5200–5961, 7370–7374, 8731–8734) and zero otherwise.
	Firm characteristics
SIZE	Natural logarithm of total assets.
LEV	Debt value over the sum of equity and debt values.
ROA	Return on assets: Net income divided by the book value of assets.
TANG	The ratio of property, plant, and equipment (PPE) to the book value of total assets.
GROWTH	The sales growth rate, calculated as the ratio of the difference between sales in the current year and prior year to sales in the prior year.
VOL	Annualized stock return volatility is computed as the standard deviation of weekly stock return.
CASH	Cash / Book Assets.
RD	R & D investment / Book Assets.

Table 2. Sample distribution by country

This table summarizes country distributions of firm-year observations and some main variables in our sample. *Language* is the major language at the country level, ISO is ISO 3166-1 alpha-3, *Religion* is the most major religion in-country, *Risk-neutral PD* is the probability of default under risk-free rate, and *Physical PD* is the probability of default under real asset return rate. All variables are defined in Table 1.

Country	ISO	FTR	*Language*	*Religion*	*Risk-neutral PD*	*Physical PD*
Australia	AUS	Strong	English	Protestant	0.07	0.09
Austria	AUT	Weak	German	Catholic	0.04	0.05
Belgium	BEL	Weak	Dutch	Roman	0.04	0.05
Brazil	BRA	Weak	Portuguese (BR)	Roman	0.08	0.10
Chile	CHL	Strong	Spanish	Roman	0.01	0.01
China	CHN	Weak	Mandarin	Buddhist	0.01	0.01
Colombia	COL	Strong	Spanish	Catholic	0.02	0.02
Czech Republic	CZE	Strong	Spanish	Roman	0.21	0.15
Denmark	DNK	Weak	Danish	Protestant	0.04	0.05
Finland	FIN	Weak	French	Protestant	0.03	0.04
France	FRA	Strong	French	Catholic	0.03	0.05
Germany	DEU	Weak	German	Protestant	0.05	0.08
Greece	GRC	Strong	Greek	Greek	0.06	0.12
Hong Kong	HKG	Weak	Cantonese	Indigenous	0.03	0.04
Hungary	HUN	Strong	Hungarian	Roman	0.04	0.04
India	IND	Weak	Indonesian	Hindu	0.04	0.06
Indonesia	IDN	Strong	Hindi	Muslim	0.06	0.06
Ireland	IRL	Strong	English	Roman	0.19	0.11
Italy	ITA	Strong	Italian	Catholic	0.03	0.04
Japan	JPN	Weak	Japanese	Buddhist	0.02	0.03
Malaysia	MYS	Weak	Malaysian	Muslim	0.04	0.05
Mexico	MEX	Strong	Spanish	Catholic	0.05	0.05
Netherlands	NLD	Weak	Norwegian	Catholic	0.05	0.07
New Zealand	NZL	Weak	Dutch	Protestant	0.03	0.04
Norway	NOR	Weak	Norwegian	Protestant	0.06	0.08
Philippines	PHL	Strong	Tagalog	Catholic	0.04	0.06
Poland	POL	Strong	Polish	Catholic	0.05	0.08
Portugal	PRT	Strong	Portuguese (EU)	Catholic	0.08	0.09
Singapore	SGP	Weak	Mandarin	Buddhist	0.07	0.09
South Korea	KOR	Strong	Korean	Protestant	0.04	0.05
Spain	ESP	Strong	Spanish	Catholic	0.02	0.03
Sweden	SWE	Weak	Swedish	Protestant	0.06	0.07
Thailand	THA	Strong	Thai	Buddhist	0.09	0.09
Turkey	TUR	Strong	Turkish	Muslim	0.01	0.02
UK	GBR	Strong	English	Protestant	0.06	0.07
US	USA	Strong	English	Protestant	0.07	0.08

Table 3. Sample distribution

This table reports descriptive statistics for sample firms. *Risk-neutral PD* is the probability of default under risk-free rate, *Physical PD* is the probability of default under real asset return rate, *Strong FTR* is the major language of country-level equal 1 when language present and the future obligatorily. All variables are defined in Table 1.

	Mean	STD	Q1	Median	Q3
Risk-neutral PD	0.04	0.14	1.E-07	2.E-04	0.01
Physical PD	0.05	0.16	1.E-08	1.E-04	0.02
DD1	4.58	14.55	2.65	3.82	5.49
DD2	4.17	14.59	2.28	3.51	5.21
Strong FTR	0.53	0.50	0.00	1.00	1.00
LTO	60.15	24.60	38.00	61.00	87.00
SENTENCE	46.21	44.82	0.00	34.40	92.90
VERB	42.01	41.08	0.00	28.20	82.20
SIZE	20.87	3.17	18.38	20.88	23.06
LEV	0.49	0.23	0.32	0.50	0.66
ROA	0.05	1.40	0.03	0.08	0.13
TANG	0.30	0.23	0.10	0.26	0.44
GROWTH	0.16	0.63	-0.04	0.06	0.20
VOL	0.10	0.25	0.03	0.05	0.08
CASH	0.12	0.15	0.02	0.07	0.16
RD	0.01	0.11	0.00	0.00	2.E-03
HHI	0.11	0.14	0.03	0.06	0.14
LITIGATION	0.23	0.42	0.00	0.00	0.00
GGDP	0.08	0.02	0.07	0.08	0.09
INFLATION	0.09	0.01	0.09	0.10	0.10
FREEDOM	69.02	9.87	61.90	71.00	76.30
POWER DISTANCE	59.26	19.40	40.00	54.00	77.00
MASCULINITY	62.14	20.37	50.00	62.00	66.00
UNCERTAINTY AVOIANCE	59.67	25.66	36.00	48.00	88.00
INDULGENCE	46.37	17.67	29.00	42.00	68.00
RELIGIOUS	0.05	0.01	0.05	0.06	0.06
TRAN	0.46	0.50	0.00	0.00	1.00
ACCOUNT	0.42	0.49	0.00	0.00	1.00
CPI	0.69	0.46	0.00	1.00	1.00

Table 4. Correlation coefficients between independent variables

This table reports the correlation coefficients between independent variables. Pearson's correlation coefficients are below the diagonal. All variables are defined in Table 1. Numbers in bold are statistically significant at the 10% confidence level.

		VIF	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)	(21)	(22)
(1)	*Strong FTR*	2.88	1.00																					
(2)	*SIZE*	1.67	**-0.33**	1.00																				
(3)	*LEV*	1.25	**-0.01**	**0.11**	1.00																			
(4)	*ROA*	1.01	**-0.01**	**0.04**	**0.02**	1.00																		
(5)	*TANG*	1.20	**-0.03**	**0.13**	**0.02**	**0.01**	1.00																	
(6)	*GROWTH*	1.03	**0.08**	**-0.05**	**-0.05**	0.00	**-0.03**	1.00																
(7)	*VOL*	1.03	**0.06**	**-0.06**	**-0.03**	**-0.01**	**0.02**	**0.02**	1.00															
(8)	*CASH*	1.40	**-0.15**	**-0.05**	**-0.36**	**-0.04**	**-0.25**	**0.04**	0.00	1.00														
(9)	*RD*	1.09	**0.02**	**-0.04**	**-0.05**	**-0.04**	**-0.07**	**0.01**	**0.00**	**0.13**	1.00													
(10)	*HHI*	1.20	**0.07**	**-0.17**	**-0.07**	**-0.01**	**0.04**	**0.02**	**0.04**	**0.02**	0.00	1.00												
(11)	*LITIGATION*	1.10	**0.00**	**-0.02**	**-0.11**	**-0.01**	**-0.20**	**0.03**	**0.01**	**0.14**	**0.11**	**-0.06**	1.00											
(12)	*GGDP*	1.96	**0.08**	**0.02**	**-0.10**	**0.00**	**0.06**	**0.08**	**0.02**	**-0.01**	**-0.03**	**-0.03**	**-0.02**	1.00										
(13)	*INFLATION*	2.19	**-0.27**	**-0.02**	**0.04**	**-0.01**	**-0.13**	**-0.04**	**-0.07**	**0.10**	**0.04**	**-0.12**	**0.07**	**-0.31**	1.00									
(14)	*FREEDOM*	4.21	**0.18**	**-0.26**	**-0.03**	**-0.03**	**-0.11**	**0.03**	**0.01**	**0.04**	**0.05**	**0.02**	**0.03**	**-0.45**	**0.44**	1.00								
(15)	*POWER DISTANCE*	4.88	**-0.13**	**0.07**	**-0.19**	**-0.01**	**0.18**	**0.01**	**0.02**	**0.06**	**-0.06**	**-0.02**	**-0.08**	**0.43**	**-0.36**	**-0.49**	1.00							
(16)	*MASCULINITY*	2.07	**-0.33**	**0.28**	**0.09**	**0.01**	**-0.01**	**-0.06**	**-0.05**	**0.10**	**-0.01**	**-0.30**	**0.02**	**-0.23**	**0.28**	**0.04**	**-0.11**	1.00						
(17)	*UNCERTAINTY AVOIANCE*	3.02	**-0.03**	**0.29**	**0.05**	**-0.01**	**0.04**	**-0.06**	**-0.02**	**0.07**	0.00	**-0.10**	**-0.01**	**-0.40**	**0.26**	**0.13**	**-0.02**	**0.38**	1.00					
(18)	*INDULGENCE*	3.37	**0.17**	**-0.26**	**0.11**	**0.01**	**-0.14**	**0.02**	**0.01**	**-0.14**	**0.04**	**0.07**	**0.04**	**-0.30**	**0.23**	**0.49**	**-0.64**	**-0.12**	**-0.23**	1.00				
(19)	*RELIGIOUS*	2.77	**-0.01**	**-0.11**	**0.02**	**-0.02**	**-0.08**	**0.02**	**-0.02**	**0.11**	**0.04**	**0.01**	**0.06**	**-0.40**	**0.49**	**0.60**	**-0.50**	**0.17**	**0.28**	**0.34**	1.00			
(20)	*TRAN*	7.24	**0.35**	**-0.54**	**-0.08**	**-0.02**	**-0.11**	**0.08**	**0.03**	**-0.03**	**0.05**	**0.12**	**0.02**	**-0.12**	**0.16**	**0.64**	**-0.26**	**-0.40**	**-0.38**	**0.62**	**0.32**	1.00		

(21)	*ACCOUNT*	4.79	**0.61**	**-0.30**	**0.09**	**0.01**	**-0.12**	**0.05**	**0.05**	**-0.18**	**0.03**	**0.02**	**0.04**	**0.04**	**-0.17**	**0.17**	**-0.43**	**-0.29**	**-0.41**	**0.56**	**-0.11**	**0.50**	1.00	
(22)	*CPI*	3.45	**0.03**	**-0.08**	**0.06**	**-0.02**	**-0.13**	0.00	**-0.01**	**0.04**	**0.05**	**-0.01**	**0.05**	**-0.50**	**0.45**	**0.76**	**-0.57**	**0.12**	**0.28**	**0.41**	**0.59**	**0.42**	**0.09**	1.00

Table 5. The effect of language on default probability

This table summarizes the estimation results from regression (4) using either Risk-neutral PD or Physical PD as the dependent variable. All variables are defined in Table 1. We report t-statistics in parentheses, which are based on robust standard errors clustered by country. We use ***, **, * to denote statistical significance at 1%, 5%, and 10% levels, respectively.

Explanatory Variables	OLS regression	
Dependent variable	Model 1	Model 2
	Risk-neutral PD	*Physical PD*
constant	0.0811	0.2365 ***
	(1.05)	(2.53)
Strong FTR	-0.0034 ***	-0.0059 ***
	(-4.36)	(-6.34)
SIZE	-0.0057 ***	-0.0100 ***
	(-54.03)	(-78.34)
LEV	0.0714 ***	0.1098 ***
	(53.94)	(68.43)
ROA	-0.0007 ***	-0.0013 ***
	(-3.66)	(-5.39)
TANG	0.0040 ***	0.0043 ***
	(2.93)	(2.59)
GROWTH	-0.0015 ***	-0.0114 ***
	(-3.24)	(-19.74)
VOL	0.4841 ***	0.4798 ***
	(253.48)	(207.31)
CASH	0.0091 ***	-0.0032
	(4.16)	(-1.19)
RD	-0.0100 ***	0.0049
	(-2.85)	(1.15)
HHI	-0.0043 *	-0.0072 ***
	(-1.72)	(-2.38)
LITIGATION	0.0010	0.0043 ***
	(1.44)	(5.01)
GGDP	-0.0150	-0.1238 ***
	(-0.56)	(-3.78)
INFLATION	-1.2639 ***	-2.1828 ***
	(-20.92)	(-29.81)
FREEDOM	0.0011 ***	0.0010 ***
	(24.88)	(18.93)
POWER DISTANCE	-0.0004 ***	-0.0005 ***
	(-12.96)	(-12.84)
MASCULINITY	-0.0001 ***	-0.0001 ***
	(-4.34)	(-4.18)
UNCERTAINTY AVOIANCE	0.0003 ***	0.0004 ***
	(18.43)	(19.17)
INDULGENCE	0.0000	-0.0003 ***
	(-0.48)	(-8.27)
RELIGIOUS	0.1607 ***	0.3645 ***
	(4.32)	(8.08)
Year fixed effect	Yes	Yes
Industry fixed effect	Yes	Yes
Language family fixed effect	Yes	Yes

Adj. R^2	0.31	0.30
Number of observations	180,013	180,013

Table 6. Robustness tests: Alternative measures of language FTR

This table summarizes the re-estimation results from regression (4) using either Risk-neutral PD or Physical PD as the dependent variable. The alternative measure of language FTR: *LTO* in Panel A; *SENTENCE* and *VERB* in Panel B. All variables are defined in Table 1. We report t-statistics in parentheses, which are based on robust standard errors clustered by country. We use ***, **, * to denote statistical significance at 1%, 5%, and 10% levels, respectively.

Panel A:

Explanatory Variables	OLS regression	
Dependent variable	Model 1	Model 2
	Risk-neutral PD	*Physical PD*
constant	0.0776	0.2264 ***
	(1.01)	(2.42)
LTO	0.0003 ***	0.0004 ***
	(16.02)	(16.20)
Control Variables	Yes	Yes
Year fixed effect	Yes	Yes
Industry fixed effect	Yes	Yes
Language family fixed effect	Yes	Yes
Adj. R^2	0.31	0.30
Number of observations	180,013	180,013

Panel B:

Explanatory Variables	OLS regression			
Dependent variable	Model 1	Model 2	Model 3	Model 4
	Risk-neutral PD	*Physical PD*	*Risk-neutral PD*	*Physical PD*
constant	0.0232	0.1056	0.0279	0.1123
	(0.30)	(1.14)	(0.37)	(1.22)
SENTENCE	-4.E-05 ***	-0.0001 ***		
	(-4.91)	(-8.37)		
VERB			-0.0001 ***	-0.0001 ***
			(-7.62)	(-11.63)
Control Variables	Yes	Yes	Yes	Yes
Year fixed effect	Yes	Yes	Yes	Yes
Industry fixed effect	Yes	Yes	Yes	Yes
Language family fixed effect	Yes	Yes	Yes	Yes
Adj. R^2	0.36	0.33	0.36	0.33
Number of observations	151,859	151,859	151,859	151,859

Table 7. Robustness tests: Alternative measures of distance to default

This table summarizes the re-estimation results from regression (4) using either DD1 or DD2 as the dependent variable. All variables are defined in Table 1. We report t-statistics in parentheses, which are based on robust standard errors clustered by country. We use ***, **, * to denote statistical significance at 1%, 5%, and 10% levels, respectively.

Explanatory Variables	OLS regression	
Dependent variable	Model 1	Model 2
	DD1	*DD2*
constant	-2.5496	-2.6269
	(-0.39)	(-0.40)
Strong FTR	0.4893 ***	0.4494 ***
	(7.47)	(6.84)
Control Variables	Yes	Yes
Year fixed effect	Yes	Yes
Industry fixed effect	Yes	Yes
Language family fixed effect	Yes	Yes
Adj. R^2	0.03	0.04
Number of observations	180,013	180,013

Table 8. Within-country evidence: Belgium.
The table presents a within-country estimation of Strong FTR language on default risk in Belgium. Some regions speak Strong FTR language, and other regions speak Weak FTR language in Belgium. All variables are defined in Table 1. We report t-statistics in parentheses, which are based on robust standard errors clustered by country. We use ***, **, * to denote statistical significance at 1%, 5%, and 10% levels, respectively.

Explanatory Variables	OLS regression	
Dependent variable	Model 1	Model 2
	Risk-neutral PD	*Physical PD*
constant	-0.0226	0.0507
	(-0.27)	(0.56)
Strong FTR	-0.0185 *	-0.0188 *
	(-1.83)	(-1.68)
Control Firm-level Variables	Yes	Yes
Year fixed effect	Yes	Yes
Industry fixed effect	Yes	Yes
Adj. R^2	0.36	0.35
Number of observations	1,546	1,546

Table 9. Within-country evidence: Strong-FTR CEOs in the US
The table presents a within-country estimation of Strong FTR language on default risk in US. Some CEOs have nationalities with Strong FTR language, and other regions speak Weak FTR language in US. All variables are defined in Table 1. We report t-statistics in parentheses, which are based on robust standard errors clustered by country. We use ***, **, * to denote statistical significance at 1%, 5%, and 10% levels, respectively.

Explanatory Variables	OLS regression	
Dependent variable	Model 1	Model 2
	Risk-neutral PD	*Physical PD*
constant	-0.2975	-0.5072
	(-0.00)	(-0.00)
CEO Strong FTR	-0.0424 ***	-0.0504 ***
	(-20.02)	(-19.62)
Control Firm-level Variables	Yes	Yes
Year fixed effect	Yes	Yes
Industry fixed effect	Yes	Yes
Language family fixed effect	Yes	Yes
Adj. R^2	0.40	0.37
Number of observations	21,326	21,326

Table 10. Hong Kong – before and after 1997

This table compares default risk in Hong Kong before and after 1997. The sample comprises 22,062 from 1994 to 2001. Hong Kong is an indicator variable for Hong Kong companies, i.e., firms headquartered in Hong Kong, and zero otherwise. All variables are defined in Table 1. We report t-statistics in parentheses, which are based on robust standard errors clustered by country. We use ***, **, * to denote statistical significance at 1%, 5%, and 10% levels, respectively.

Explanatory Variables	OLS regression	
Dependent variable	Model 1	Model 2
	Risk-neutral PD	*Physical PD*
constant	0.6278 ***	1.0175 ***
	(20.35)	(26.47)
*Hong Kong * Post-1997*	0.0897 ***	0.0991 ***
	(4.72)	(4.18)
Control Variables	Yes	Yes
Year fixed effect	Yes	Yes
Industry fixed effect	Yes	Yes
Language family fixed effect	Yes	Yes
Adj. R^2	0.26	0.31
Number of observations	22,062	22,062

Table 11. Two-stage least square (2SLS) regressions of default probability on language

Table 11 of Panel A summarizes the results from 2SLS regression, where we estimate a fitted value of societal language in the first stage and use either Risk-neutral PD or Physical PD as a dependent variable. Panel B summarizes the results from GMM regression, where we estimate a fitted value of societal language in the first stage and use either Risk-neutral PD or Physical PD as the dependent variable. All variables are defined in Table 1. We report t-statistics in parentheses, which are based on robust standard errors clustered by country. We use ***, **, * to denote statistical significance at 1%, 5%, and 10% levels, respectively.

Panel A: IV-2SLS

	First stage:	Second stage:	
Explanatory Variables	*Strong FTR*	*Risk-neutral PD*	*Physical PD*
Dependent variable	Model 1	Model 2	Model 3
constant	-0.5768 ***	0.1275 *	0.2981 ***
	(-3.67)	(1.65)	(3.19)
Strong FTR_{2SLS}		-0.0176 ***	-0.0249 ***
		(-17.18)	(-20.00)
Roman	-0.7449 ***		
	(-144.27)		
Protestant	0.1077 ***		
	(35.29)		
Muslim	-0.5746 ***		
	(-108.32)		
Buddhist	-1.0106 ***		
	(-307.56)		
Indigenous	-1.1788 ***		
	(-122.52)		
Control Variables	Yes	Yes	Yes
Year fixed effect	Yes	Yes	Yes
Industry fixed effect	Yes	Yes	Yes
Language family fixed effect	Yes	Yes	Yes
Adj. R^2	0.80	0.31	0.30
Number of observations	180,013	180,013	180,013

Panel B: IV-GMM		
Explanatory Variables	GMM regression	
Dependent variable	Model 1 *Risk-neutral PD*	Model 2 *Physical PD*
constant	0.1905 ***	0.4326 ***
	(14.01)	(26.81)
Strong FTR	-0.0177 ***	-0.0279 ***
	(-10.29)	(-14.08)
Control Variables	Yes	Yes
Year fixed effect	Yes	Yes
Industry fixed effect	Yes	Yes
Language family fixed effect	Yes	Yes
Adj. R^2	0.31	0.29
Number of observations	180,013	180,013

Table 12. Cross-country transparency variations in the effect of language on default probability

This table summarizes the estimation results from regression (5) using either Risk-neutral PD or Physical PD as the dependent variable. All variables are defined in Table 1. We report t-statistics in parentheses, which are based on robust standard errors clustered by country. We use ***, **, * to denote statistical significance at 1%, 5%, and 10% levels, respectively.

Explanatory Variables	OLS regression	
Dependent variable	Model 1	Model 2
	Risk-neutral PD	*Physical PD*
constant	0.0354	0.1631 *
	(0.46)	(1.75)
Strong FTR	-0.0051 ***	-0.0046 ***
	(-4.47)	(-3.35)
TRAN	-0.0388 ***	-0.0553 ***
	(-18.77)	(-22.11)
*Strong FTR * TRAN*	0.0140 ***	0.0136 ***
	(7.68)	(6.20)
Control Variables	Yes	Yes
Year fixed effect	Yes	Yes
Industry fixed effect	Yes	Yes
Language family fixed effect	Yes	Yes
Adj. R^2	0.31	0.30
Number of observations	180,013	180,013

Table 13. Cross-country accounting standard score variations in the effect of language on default probability

This table summarizes the estimation results from regression (6) using either Risk-neutral PD or Physical PD as the dependent variable. All variables are defined in Table 1. We report t-statistics in parentheses, which are based on robust standard errors clustered by country. We use ***, **, * to denote statistical significance at 1%, 5%, and 10% levels, respectively.

Explanatory Variables	OLS regression	
Dependent variable	Model 1	Model 2
	Risk-neutral PD	*Physical PD*
constant	0.0960	0.2534 ***
	(1.24)	(2.71)
Strong FTR	-0.0045 ***	-0.0072 ***
	(-3.49)	(-4.56)
ACCOUNT	-0.0205 ***	-0.0231 ***
	(-12.43)	(-11.56)
*Strong FTR * ACCOUNT*	0.0151 ***	0.0169 ***
	(6.98)	(6.45)
Control Variables	Yes	Yes
Year fixed effect	Yes	Yes
Industry fixed effect	Yes	Yes
Language family fixed effect	Yes	Yes
Adj. R^2	0.31	0.30
Number of observations	180,013	180,013

Table 14. Cross-country CPI score variations in the effect of language on default probability

This table summarizes the estimation results from regression (7) using either Risk-neutral PD or Physical PD as the dependent variable. All variables are defined in Table 1. We report t-statistics in parentheses, which are based on robust standard errors clustered by country. We use ***, **, * to denote statistical significance at 1%, 5%, and 10% levels, respectively.

Explanatory Variables	OLS regression	
Dependent variable	Model 1 *Risk-neutral PD*	Model 2 *Physical PD*
constant	0.1115	0.2778 ***
	(1.44)	(2.97)
Strong FTR	-0.0123 ***	-0.0174 ***
	(-9.10)	(-10.65)
CPI	-0.5783 ***	-1.0927 ***
	(-11.28)	(-17.59)
*Strong FTR * CPI*	0.0071 ***	0.0063 ***
	(5.64)	(4.15)
Control Variables	Yes	Yes
Year fixed effect	Yes	Yes
Industry fixed effect	Yes	Yes
Language family fixed effect	Yes	Yes
Adj. R^2	0.31	0.30
Number of observations	180,013	180,013

Table 15. Mechanism analysis

This table summarizes the estimation results using either leverage (*LEV*) or stock return (*RET*) as the dependent variable. All variables are defined in Table 1. We report t-statistics in parentheses, which are based on robust standard errors clustered by country. We use ***, **, * to denote statistical significance at 1%, 5%, and 10% levels, respectively.

Explanatory Variables	OLS regression	
Dependent variable	Model 1	Model 2
	LEV	*RET*
constant	0.6340 ***	0.1052 ***
	(4.61)	(4.51)
Strong FTR	-0.0217 ***	0.0014 ***
	(-15.84)	(5.88)
Control Variables without *LEV*	Yes	Yes
Year fixed effect	Yes	Yes
Industry fixed effect	Yes	Yes
Language family fixed effect	Yes	Yes
Adj. R^2	0.27	0.78
Number of observations	180,013	180,013