The Impact of Linguistic Distance and Financial Reporting Readability on Foreign Holdings of U.S. Stocks

Kristian D. Allee
Lisa S. Anderson
Michael J. Crawley
University of Arkansas

August 2018

Abstract: Using a sample of 75 countries from U.S. Treasury regulatory data, we show that foreign investors invest less (more) in U.S. stocks (U.S. Treasuries) when they are from countries with greater linguistic distances and when U.S. financial reports are more difficult to read. This suggests that linguistic distance and financial reporting readability act as significant frictions for foreign investors when investing in equities, even in U.S. the market where foreign investors should have the greatest ability, resources, and economic incentives to overcome language translation and readability issues. Our results on U.S. Treasuries are consistent with a "substitution effect" where foreign investors who want to invest in U.S. stocks, but are sufficiently deterred by linguistic distance and financial reporting readability frictions, appear to seek the relative simplicity of investing in Treasury securities.

Keywords: Linguistic distance; financial reporting readability; foreign investors; U.S. stocks; U.S. Treasuries

JEL Classifications: M41, F21, G15

We thank workshop participants at the University of Arkansas, the University of Kansas, and the Oklahoma State University Accounting Research Conference for their helpful comments.

I. Introduction

Foreign investment is an essential component of the depth and liquidity of the United States (U.S.) security markets (Stulz 1999a; Stulz 1999b). Approximately one-third of U.S. corporate stocks are owned by foreign investors, and foreign investment can facilitate growth and lower the cost of equity for U.S. firms (Bump 2017; Rosenthal 2017; Henry 2000a; Henry 2000b; Lizardo and Mollick 2009). However, investors often face economic frictions (e.g. capital controls) and other impediments (e.g. a "home bias") when considering investing abroad (Kang and Stulz 1997; Karolyi and Stulz 2003; Chan et al. 2005; Lundholm et al. 2014). We examine whether linguistic distance and the readability of U.S. financial reports are additional market frictions which affect foreign investment in U.S. stocks.

Linguistic distance reflects how different one language is from another (Isphording and Otten 2013). We use the simple and atheoretic "language acquisition approach" from the linguistics literature to measure the linguistic distance between English and 43 other languages (Chiswick and Miller 2005; Hart-Gonzalez and Lindemann 1993). In additional analyses, we utilize an alternative "phonetic similarity approach" to measure linguistic distance (Isphording 2014; Isphording and Otten 2013) and find similar results. Financial reporting readability refers to the ease with which a reader can process written text (Bonsall et al. 2017). We create an aggregate measure of U.S. financial reporting readability based on the firm filing-level "Bog Index" created by Bonsall et al. (2017) which captures plain English attributes of disclosure (e.g. avoiding passive voice, weak verbs, overused words, complex words, and jargon). Linguistic distance and financial reporting readability are related constructs in our international setting because both may increase foreign investors' information processing costs, and in turn, reduce the amount of U.S. stocks they are willing to hold.

We demonstrate that linguistic distance and financial reporting readability are both negatively associated with foreign holdings of U.S. stocks. That is, foreign investors invest less in U.S. stocks when they are from countries with greater linguistic distances and when U.S. financial reports are more difficult to read. Furthermore, we show that linguistic distance and financial reporting readability interact. Specifically, our results show that foreign investors from high linguistic distance countries facing unreadable financial reports reduce their investments less than foreign investors from low language distance countries facing the same unreadable financial reports. This suggests that foreign investors are sensitive to linguistic distance and financial reporting readability but are reluctant to reduce their investments in U.S. stocks below a certain floor (perhaps because their next best option involves either investing in their home country which reduces global diversification or investing in another country where they will still likely face a linguistic distance and financial reports that are difficult to read).

We find that these results are robust to (1) the inclusion of controls for a country's demographic, macroeconomic, tax, financial reporting, legal, religious, educational, financial literacy, capital control, and cultural environments, (2) the use of an alternative linguistic distance measure, (3) excluding English speaking countries, (4) the use of monthly or annual foreign holdings data, (5) excluding offshore banking center countries, (6) separate analyses of the world's largest and most important economies (i.e. G-20 countries) and less developed economies, (7) estimating changes specifications designed to reduce the potential for non-stationarity or unobserved time-invariant heterogeneity, and (8) estimating our models excluding one country at a time to rule out influence by any single country in our sample. These robustness tests demonstrate consistency and support for a negative association between both linguistic distance and financial reporting readability and foreign holdings of U.S. stocks.

If foreign investors are investing less in U.S. stocks in response to linguistic distance and difficulty reading financial reports, then foreign investors could be using that capital in numerous other ways. For example, foreign investors could choose to consume (rather than invest), invest in their home country, or invest in other foreign countries. We explore U.S. Treasuries as one possible alternative investment vehicle for foreign investors who are investing less in U.S. stocks in response to linguistic distance and difficulty reading financial reports. U.S. Treasuries allow foreign investors to access U.S. financial markets without facing the same linguistic distance and financial reporting readability frictions confronted when investing in individual equity securities because U.S. Treasuries do not have financial statements to analyze. We find that linguistic distance and financial reporting readability are positively associated with foreign holdings of U.S. Treasuries, and the interaction of linguistic distance and financial reporting readability is negatively associated with foreign holdings of U.S. Treasuries. These mirror images of the U.S. stock results are consistent with a "substitution effect." In other words, foreign investors who want to invest in the U.S., but are sufficiently deterred by the linguistic distance and financial reporting readability frictions associated with stock ownership, appear to seek the relative simplicity of investing in U.S. Treasuries.

Our study contributes to multiple literatures including the international literature which examines frictions in cross-border financial market transactions. We contribute to this literature by using a large sample of 75 countries to show that linguistic distance and financial reporting readability influence foreign holdings of U.S. stocks. Demonstrating that linguistic distance and financial reporting affects foreign investment in the U.S. is important because the U.S. is the largest and most competitive capital market in the world with a strong information environment and no capital controls (Bradshaw et al. 2004; Ahearne et al. 2004; Stulz and Wasserfallen

1995). Moreover, English is the dominant *lingua franca* ("language of business") around the world (Fredriksson et al. 2006; Neely 2012). Thus, our results suggest that linguistic distance and financial reporting readability persist as significant frictions, even in the market where foreign investors should have the greatest ability, resources, and economic incentives to overcome language translation and readability issues. Moreover, we are the first to examine the impact of linguistic distance and financial reporting readability on foreign ownership of U.S. Treasuries. Our results suggest that linguistic distance and financial reporting readability increase the demand for U.S. Treasuries (a unique result given that frictions generally reduce demand).

Second, we contribute to the recent literature on financial reporting readability. Prior research shows that firms opportunistically alter the readability of their financial reports (Li 2008), and these actions adversely impact investors (You and Zhang 2009; Miller 2010; Rennekamp 2012; Lawrence 2013), analysts (Lehavy et al. 2011; Bozanic and Thevenot 2015), and credit rating agencies (Bonsall and Miller 2017). Our results suggest that financial reporting readability acts as an international friction which affects the investment behavior of foreign investors—a large but previously overlooked population in this literature. Additionally, our results suggest that financial reporting readability can influence the demand for assets other than corporate stocks, including U.S. Treasuries that do not have associated financial statements.

Finally, our results may be of interest to regulators. Policy makers can do little to change linguistic distances, but securities regulators may be able to influence the readability of U.S. financial reports. The Securities and Exchange Commission (SEC) has long been concerned that firms fail to make their filings readable and understandable (Firtel 1999). The SEC issued plain English disclosure guidelines in 1998 (SEC 1998), and the SEC has considered using a readability measure to assess firm compliance (Loughran and McDonald 2014). Our results

suggest that a lack of financial reporting readability imposes a unique information externality on foreign investors and that the consequences differ depending on linguistic distance (i.e. foreign investors are not homogenous with respect to translation costs). Therefore, regulators may wish to more explicitly consider the ramifications of financial reporting readability for this economically important community, particularly given trends towards less readable financial reports and more foreign ownership of U.S. stocks.[1]

This study is organized as follows. Section II discusses prior research and motivates our hypotheses. Section III describes our data sources and sample. Section IV describes our methodology. Section V presents the results, and Section VI concludes.

II. Background and Hypothesis Development

Linguistic Distance

A large literature explores how language and other cultural differences act as market frictions. Within the accounting literature, Lundholm et al. (2014) show that foreign firms with listings on U.S. stock exchanges write more readable financial reports than their U.S. counterparts to lessen the psychological distance between the firm and prospective U.S. investors. The finance literature concludes that investors have a propensity to avoid foreign stocks. That is, investors display a "home bias" where they overweight domestic stocks, potentially due to informational deficiencies and psychological biases induced by differences in language and culture (see Karolyi and Stulz 2003 for a review). Relatedly, Baik et al. (2013) show that increases in foreign institutional ownership in U.S. stocks are negatively correlated with future returns, consistent with foreign investors facing a "liability of foreignness" that

[1] According to the Tax Policy Center, 35% of U.S. stocks were held by foreign investors in 2016 compared to 10% in the early 1980s (Bump 2017; Rosenthal 2017).

impairs their ability to predict returns. In economics, anecdotal evidence and academic research suggests that language barriers can impact bilateral trade volumes, the literacy and assimilation of immigrants, and a country's ability to attract high-skilled workers (Isphording and Otten 2013; Isphording 2014; Khazan 2013). Finally, prior research suggests that individuals from countries where the language does not distinguish between the present and future tense (e.g. a weak future time reference such as "It rains tomorrow" compared to the stronger "It will rain tomorrow") save more, smoke less, practice safer sex, and are less obese (Chen 2013), and managers from such countries are less likely to engage in earnings management (Kim et al. 2017).

We contribute to this literature by examining the association between linguistic distance and foreign investors' demand for U.S. stocks and U.S. Treasuries. At the construct level, linguistics researchers generally agree that linguistic distance represents how different one language is from another. However, languages are complex (e.g. they differ with respect to vocabulary, grammar, syntax, written form), and linguistics researchers do not agree on the best way to measure linguistic distance (Chiswick and Miller 2005; Isphording and Otten 2013).

There are at least three methods to measure linguistic distance. First, some studies adopt an anthropologic approach and count the number of branches on language trees that two languages share (see Guiso et al. 2009 for an example). The underlying theory is that all languages descend from a small number of languages, and language diversity is the result of prehistoric *Homo sapien* migration "out of Africa". According to this view, languages can be represented in the form of a tree (like genealogical trees). Languages that share more (fewer) branches are deemed to have a lower (greater) linguistic distance (Ginsburgh and Weber 2016).[2]

[2] A related approach relies on the high correlation between genetic and linguistic evolution. Chen et al. (1995) use DNA samples from 130 populations to construct genetic trees. Distances on the genetic trees are then used to proxy for linguistic distance. See Demirbag et al. (2007) for an application of this approach in the economics literature.

Second, other studies use a phonetic approach to measure linguistic distance (e.g. Isphording 2014; Isphording and Otten 2013). Researchers identify words having the same meaning in two different languages (e.g. “hello” and “hola”) and use linguistics software to measure how much the pronunciations differ between the pair. Languages with a greater (lower) average pronunciation similarity across a set of words are assumed to have more (fewer) “cognates” (common ancestries of words) and a lower (greater) linguistic distance.

Finally, Chiswick and Miller (2005) use a language acquisition approach. The idea is to measure how difficult it is to learn to speak another language. Conveniently, the U.S. Department of State collects data on how difficult it is for native English speakers to learn other languages. The U.S. Department of State’s School of Language Studies (SLS) administers language training protocols in over 70 languages to native English-speaking U.S. government employees (e.g. diplomats). At the end of each protocol, the SLS administers standardized tests to assess the employee’s reading and speaking proficiency. Higher (lower) average test scores indicate a lower (greater) linguistic distance. Data for a sample of SLS test subjects was initially published by Hart-Gonzalez and Lindemann (1993) and first used as a measure of linguistic distance in the academic literature by Chiswick and Miller (2005).

We use the Chiswick and Miller (2005) measure of linguistic distance as our primary measure for three reasons. First, the language acquisition approach is simple and atheoretic. The approach does not rely on assumptions about the functional form of language trees, the validity of theories about prehistoric migration patterns, or the appropriate way to measure differences in pronunciation. Second, the Chiswick and Miller (2005) measure is the only proxy measuring how difficult it was for humans to process and respond to communication in another language. Finally, the Chiswick and Miller (2005) measure is publicly available for many languages.

One criticism of the Chiswick and Miller (2005) measure is that linguistic distances only exist relative to English (Isphording 2014), but this restriction does not apply in our study because we only require a proxy for the distance between English and other languages. Another concern about the Chiswick and Miller (2005) measure is the lack of symmetry (i.e. the difficulty for a native English speaker to learn Spanish may not necessarily be the same as the difficulty for a native Spanish speaker to learn English). Hence, we use an alternate symmetric measure of linguistic distance using the phonetic approach as a robustness check (see Section V).

We expect linguistic distance to increase foreign investors' information processing costs, and in turn, reduce the amount of U.S. stocks they are willing to hold. However, there are multiple reasons why we may fail to find evidence supporting this prediction. First, the U.S. is the largest and most competitive capital market in the world with a strong information environment and no capital controls (Bradshaw et al. 2004; Ahearne et al. 2004; Stulz and Wasserfallen 1995). Second, English is the dominant language of business around the world (Fredriksson et al. 2006; Neely 2012). Third, tens of millions of foreign citizens have studied at universities where instruction is delivered in English, including many who study business and go on to be investors (Tietze 2004). Overall, foreign investors investing in the U.S. are likely to be relatively sophisticated and may have sufficient abilities, resources, and economic incentives to overcome translation issues. Therefore, we state our first hypothesis in null form:

H1: Linguistic distance has no impact on foreign holdings of U.S. stocks.

Financial Reporting Readability

Firm financial reports are an important element of the U.S. information environment (Kothari 2001), and both foreign and domestic investors attempting to process the information within financial statements face a non-trivial task. Attempts to quantify the difficulty of processing written English text date back to the Fog Index created by Gunning (1952). However,

the use of readability measures in accounting and finance research was limited until relatively recently. Li (2008) shows that longer and less readable 10-K filings are associated with lower profitability and earnings persistence, consistent with managers manipulating the readability of financial reports to obfuscate poor performance. Subsequent research shows that such opportunistic behavior can affect the decision usefulness of financial statements for investors (You and Zhang 2009; Miller 2010; Rennekamp 2012; Lawrence 2013), analysts (Lehavy et al. 2011; Bozanic and Thevenot 2015), and credit rating agencies (Bonsall and Miller 2017).

The empirical proxies for financial reporting readability have evolved as the literature has matured. Most early studies use the Gunning (1952) Fog Index which measures average sentence length and the percentage of complex words (i.e. words with more than three syllables). However, Loughran and McDonald (2014) note that financial reports contain a large percentage of words with more than three syllables but that are not difficult to understand for most financial statement users (e.g. depreciation, liability). As such, Loughran and McDonald (2014) propose using the file size of the firm's electronic Form 10-K filing as an alternative readability measure. More recently, Bonsall et al. (2017) use computational linguistics software to create a "Bog Index" designed to capture plain English attributes of disclosure (e.g. avoiding passive voice, weak verbs, overused words, complex words, and jargon).

We use the Bonsall et al. (2017) Bog Index as the starting point for our measure of aggregate U.S. financial reporting readability for several reasons. First, the Bog Index uses a dictionary which avoids overstating word difficulty based on syllable counts. Second, the Bog Index captures the most dimensions of the SEC's plain English guidelines. Third, Bonsall et al. (2017) use an experiment and validate that human subjects find high Bog score documents

harder to read (controlling for word count, file size, formatting, and Fog score). Finally, Bog scores for a large sample of 10-K filings over multiple years are publicly available.

We expect financial reporting readability to increase foreign investors' information processing costs, and in turn, reduce the amount of U.S. stocks they are willing to hold. However, there are multiple reasons why we may fail to find evidence supporting this prediction. First, Guay et al. (2016) find that firms with less readable financial reports engage in more voluntary disclosure which suggests managers mitigate the negative effects of unreadable financial reports. Second, foreign investors investing in the U.S. are likely to be relatively sophisticated and may have sufficient abilities, resources, and economic incentives to minimize the impact of readability frictions. Therefore, we state our second hypothesis in null form:

H2: Financial reporting readability has no impact on foreign holdings of U.S. stocks.

Linguistic Distance and Financial Reporting Readability

Linguistic distance and financial reporting readability may interact and influence the amount of U.S. stocks that foreign investors hold. On one hand, a foreign investor from a country with a high linguistic distance facing unreadable U.S. financial reports may further reduce her U.S. stock holdings due to the compounding of multiple frictions (i.e. the interaction could amplify the main effects of linguistic distance and financial reporting readability).

On the other hand, if a foreign investor with a high linguistic distance facing unreadable U.S. financial reports invests less in U.S. stocks, then the question becomes where does she direct those funds? Investing those funds in her home country will reduce her global diversification and prevent her from enjoying the levels of investor protection, financial reporting quality, and financial market development available in the U.S. Alternatively, she's unlikely to escape language translation and financial reporting readability issues if she invests in another foreign country. For example, a South Korean investor will also face a high language

distance if she decides to invest those funds in French stocks, and French financial reports may be difficult to read as well. Consequently, U.S. stocks may still be her first best option for at least a minimum portion of her portfolio. More formally, a foreign investor from a high linguistic distance country facing unreadable financial reports may reduce her investments less than a foreign investor from a low language distance country facing the same unreadable financial reports (e.g. the interaction could attenuate any main effect of financial reporting readability).

Ultimately, the impact of the interaction between linguistic distance and financial reporting readability is an empirical question, and we state our final hypothesis in null form:

H3: The interaction of linguistic distance and financial reporting readability has no impact on foreign holdings of U.S. stocks.

III. Data and Sample

Foreign Ownership of U.S. Stocks

The U.S. Treasury Department monitors the ownership of U.S. assets by foreign investors. The Treasury International Capital (TIC) reporting system measures cross-border financial positions and helps policy makers monitor systemic risks in the financial system (Brandner et al. 2012). Commercial banks, securities dealers, and other financial institutions are required to file various monthly and quarterly reports with district U.S. Federal Reserve Banks. Reporting is mandatory for U.S. entities and foreign entities pursuant to the International Investment and Trade in Services Survey Act (22 U.S.C. 3101 et seq.), and penalties can be imposed for non-compliance (U.S. Treasury 2016). TIC Form Securities Long-Term (SLT) requires each reporting entity to aggregate the ownership of U.S. stocks (at nominal current market value) by the country of citizenship of its account holders by month.[3] An institution's

[3] For example, assume Deutsche Bank has 5 account holders who own a total of $100 million of U.S. stocks at December 31, 2015. The account holders include two German investors which own $22 million each, two Swiss investors which own $10 million each, and one French investor owning $36 million. Deutsche Bank's TIC Form

monthly report does not include the identities of its account holders or any other account-level information (i.e., which stocks they hold), and individual reports are not available to the public. However, reported values for all institutions with foreign holders of U.S. stocks are aggregated by country. The resulting total ownership in U.S. stocks for investors from country *i* at the end of month *j* ($CS_{i,j}$) is available continuously beginning in December 2011.[4] We deflate holdings by the annual price index from U.S. Bureau of Economic Analysis (BEA) National Income and Product Account (NIPA) Table 1.1.4 to convert *CS* values to millions of real 2010 U.S. dollars.[5]

Linguistic Distance

TIC data on foreign holdings of U.S. stocks are at the country level. Thus, we require a measure of linguistic distance by country. We begin with the measure of linguistic distance between English and 43 other languages based on how difficult it was for a sample of English-speaking U.S. State Department employees to learn those languages (Chiswick and Miller 2005; Hart-Gonzalez and Lindemann 1993). The individual linguistic distance scores are publicly available, see Table 1 of Chiswick and Miller (2005). We create a mapping from languages to countries by weighting the linguistic distance of the languages spoken within a country by the percentage of the population speaking each language.[6] Finally, we subtract each value from 4 so that the resulting time-invariant measure of linguistic distance for country *i* (LD_i) varies from 0 to 3 with higher values denoting greater linguistic distance.

Financial Reporting Readability

SLT for December 31, 2015 would show aggregate German ownership of U.S. stocks of $44 million, aggregate Swiss ownership of $20 million, and aggregate French ownership of $36 million.

[4] See http://ticdata.treasury.gov/Publish/fslt-dec2013.pdf for the official TIC Form SLT from the U.S. Treasury. See Table 1D in Section B.2 at https://www.treasury.gov/resource-center/data-chart-center/tic/Pages/ticsec2.aspx for aggregated U.S. stock holdings by country-month.

[5] See BEA NIPA Table 1.1.4 at https://www.bea.gov/iTable/index_nipa.cfm. Results are quantitatively and qualitatively similar using nominal or real holdings values.

[6] The percentage of the population speaking each language within a country is from the Central Intelligence Agency (CIA) Factbook at https://www.cia.gov/library/publications/the-world-factbook/fields/2098.html.

TIC data on foreign holdings of U.S. stocks are at a monthly frequency, and thus we require a monthly aggregate measure of U.S. financial reporting readability. We begin with the Bog scores for firm-level Form 10-K filings created by Bonsall et al. (2017).[7] We create a proxy for aggregate U.S. financial reporting readability by calculating an equal weighted average of the Bog scores for all firms filing a 10-K over the past 12 months from month *j*-12 to month *j*-1.[8] Finally, we subtract the sample mean of the time series from each month's value.[9] The resulting aggregate measure of financial reporting readability for month *j*-1 (BOG_{j-1}) varies from -1.01 to 2.24 with higher values denoting less readability.

Sample

The TIC data described above provides monthly aggregate holdings of U.S. stocks by country for 125 countries. Requiring control variable data (discussed in the next section) reduces our sample to 75 countries. Our sample of 75 countries is quite large compared to other studies in the international literature. For example, Lundholm et al. (2014) examine the disclosure behavior of firms from 45 countries who list on a U.S. stock exchange. Moreover, the countries in our sample exhibit significant cross-sectional variation in size and economic importance.[10]

Table 1 presents linguistic distances and average U.S. stock holdings over the sample period for the 75 countries in our sample (countries are presented in descending order by linguistic distance). The largest linguistic distances are Japan and South Korea (both *LD* = 3), Hong Kong (*LD* = 2.75), and China (*LD* = 2.58). These are areas where "logographic" languages are dominant. Logographic languages like Japanese, Korean (to some extent), and Mandarin

[7] Firm filing-level Bog data are available at https://kelley.iu.edu/bpm/activities/bogindex.html.
[8] Results are quantitatively and qualitatively similar using an aggregated *BOG* measure weighted by firm size.
[9] "Centering" (i.e. de-meaning) helps facilitate interpretation of the regression main effects in the presence of an interaction term. See Section V for further discussion.
[10] Our sample includes all possible G-20 countries. The G-20 is a group of the largest and most important economies which promotes international financial stability. Our sample excludes the U.S. (by construction) and the European Union (TIC data tracks holdings for the individual countries) but includes the other 18 members.

Chinese use symbols to represent words or morphemes (components of words).[11] This use of symbols stands in contrast to English which uses an alphabetic system where separate letters represent phonemes (basic sounds).[12] The countries with the next largest linguistic distances (e.g. Kuwait, Saudi Arabia, Egypt, Morocco, and Lebanon with *LD* = 2.5) generally share Arabic as the dominant language. Arabic is an alphabetic language, but it uses its own alphabet (rather than the Latin alphabet used in English). Unsurprisingly, the countries with the lowest linguistic distances (United Kingdom, Ireland, Australia, New Zealand, Trinidad and Tobago, and Jamaica where *LD* = 0) are or were British settlements where English is the official language.

Table 1 also shows that the largest holders of U.S. stocks over the sample period are the United Kingdom ($668 billion) and Canada ($649 billion). These countries also have two of the lowest linguistic distances (consistent with H1). Rounding out the top five largest holders of U.S. stocks are Luxembourg ($438 billion), Japan ($348 billion), and Switzerland ($300 billion). The presence of Luxembourg and Switzerland amongst the largest holders is somewhat surprising given their small populations, but both countries are large financial centers. Similarly, the absence of China from the top five holders is somewhat surprising. However, China has the largest average holdings of U.S. Treasuries over the sample period ($1.15 Trillion) suggesting that Chinese investors have a relative preference for U.S. debt securities over U.S. equities. We

[11] The Korean writing system, Hangul, is an "alphabetic syllabary" which employs an alphabet, a syllabary, and a logography. Korean text often uses Hangul mixed with Chinese characters in a manner which aids or hinders reading depending on the familiarity of the reader with Chinese logographs (Taylor 1980).

[12] Other systems exist including pictographic systems (e.g. ancient Aztec) and syllabaries (e.g. Cherokee). While Korean has syllabic elements, we do not have any "pure" pictographic or syllabic languages in our sample. However, some such languages have historical significance. For example, U.S. armed forces used "Code Talkers" to transmit coded messages based on Native American languages (e.g. Navajo and Choctaw) during World War II. Native American languages were useful because non-native speakers found them difficult to understand, few people outside of the associated tribes knew the languages, and virtually no books had been published in those languages. The Japanese never cracked these codes (Arbuckle 2017), a testament to the power of linguistic distance.

examine the impact of financial centers and the effect of linguistic distance and financial reporting readability on U.S. Treasury holdings later in the paper.

Figure 1 plots the time-series of aggregate foreign holdings of U.S. stocks for the 75 countries in our sample. The solid line shows that foreign holdings of U.S. stocks are increasing over the sample period, but the trend is not monotonic. Total foreign holdings of U.S. stocks for all countries in the TIC data set (the 75 countries for which we have all control variable data) as of December 31, 2016 were $5.9 ($4.8) trillion. Overall, our sample covers 81.6% of total foreign holdings. Figure 1 also plots the time series of *BOG*. The increasing trend suggests that U.S. financial reports have become less readable over time, consistent with Bonsall et al. (2017).

IV. Methodology

We empirically test H1, H2, and H3 simultaneously using the following linear regression where the unit of observation is country-month:

$$CS_{i,j} = \beta_0 + \beta_1 LD_i + \beta_2 BOG_{j-1} + \beta_3 LD_i * BOG_{j-1} + \beta_4 CS_{i,j-12} + \beta_5 EQUITY_{j-1} + \beta_6 TREAS_{j-1} + \beta_7 OIL_{j-1} + \beta_8 POP_{i,t-1} + \beta_9 BILAT_{i,t-1} + \beta_{10} GDPPC_{i,t-1} + \beta_{11} \Delta GDPPC_{i,t-1} + \beta_{12} INF_{i,t-1} + \beta_{13} G20_i + \beta_{14} TAX_i + \beta_{15} OFFSHORE_i + \beta_{16} FINSTD_i + \beta_{17} COMMON_i + \beta_{18} PROTECT_i + \beta_{19} DIST_i + \beta_{20} CULTURE_i + \varepsilon_{i,j} \quad (1)$$

Variables not discussed in Section III are defined below. All variables are also defined in the Appendix. The Δ operator denotes the percentage change from year t-2 to year t-1. The β_1, β_2, and β_3 coefficients test the null for H1, H2, and H3, respectively. We also estimate a reduced form of model (1) excluding the interaction.

With respect to the control variables, we include $CS_{i,j-12}$ (aggregate U.S. stocks held by investors from country i from 12 months ago). We expect the β_4 coefficient to be positive to the extent that foreign investment levels are "sticky". For example, investors in a country may not all

jointly adjust their positions in response to changes in macroeconomic conditions due to transaction costs and the potential for large price impacts by moving simultaneously.

The next group of control variables (*EQUITY*, *TREAS*, and *OIL*) are proxies for global and U.S. economic conditions. $EQUITY_{j-1}$ and $TREAS_{j-1}$ are monthly returns on the Center for Research in Securities Prices (CRSP) value-weighted index and 30-year U.S. Treasury bonds over the past 12 months from month *j*-12 to month *j*-1 from the CRSP database, respectively. We expect the β_5 coefficient on *EQUITY* to be positive because foreign investors should have more investable funds when the global economy is strong, and prior studies show that foreign investors often "chase" returns (Lin and Swanson 2003). U.S. stock returns and U.S. Treasury bond returns are typically negatively correlated, but our sample includes periods of unique monetary policy actions by the U.S. Federal Reserve (i.e. quantitative easing) which resulted in simultaneous rallies in both the equities and Treasuries markets. As such, we make no predictions for the β_6 coefficient on *TREAS*. Monthly spot prices of Brent crude oil at the end of month *j*-1 (OIL_{j-1}) are from the Federal Reserve Economic Database.[13] We expect the β_7 coefficient on *OIL* to be positive because oil prices tend to be pro-cyclical, and foreign investors should have more money to invest when global macroeconomic conditions are strong.

The next group of controls (*POP*, *BILAT*, *GDPPC*, *ΔGDPPC*, *INF*, and *G20*) capture demographic and macroeconomic conditions in the investor's home country. Annual population in millions ($POP_{i,t-1}$), gross domestic product per capita in real 2010 U.S. dollars ($GDPPC_{i,t-1}$), and inflation ($INF_{i,t-1}$) for country *i* at the end of year *t*-1 (i.e. the most recent calendar year ending before month *j*) are available from the World Bank.[14] We make no predictions about the sign of the β_8, β_{10}, or β_{11} coefficients because investors from larger, more efficient, and faster

[13] See https://fred.stlouisfed.org/series/MCOILBRENTEU.
[14] See https://data.worldbank.org/indicator.

growing economies may have more funds to invest, but they may invest less in the U.S. because there are more investment opportunities in their home country. $BILAT_{i,t-1}$ denotes bilateral trade (i.e. the sum of imports and exports) between country *i* and the U.S. in year *t*-1 from the United Nations Commodity Trade Statistics Database.[15] $G20_i$ is a dummy variable that equals 1 if country *i* is a member of the G-20, and $G20_i$ equals 0 otherwise. We expect the β_9 and β_{13} coefficients to be positive because greater engagement with the U.S. (via bilateral trade or via participation in G-20 treaties) could lead to greater U.S. equity investment. We expect the β_{12} coefficient on *INF* to be positive because prior research demonstrates that investing in other countries is one way to hedge against home country inflation (Sayek 2009).

The next group of control variables (*TAX*, *OFFSHORE*, and *FINSTD)* are time-invariant measures of the investor's home country tax and financial environments. Estimates of the average corporate tax rate (including corporate income taxes, payroll taxes, retirement contributions, and other taxes) (TAX_i) for country *i* are available from the World Economic Forum.[16] We expect the β_{14} coefficient to be negative, consistent with research which shows that high corporate tax rates are associated with lower foreign investment (Clark 2008). $OFFSHORE_i$ is a dummy variable that equals 1 if the U.S. Treasury has identified country *i* as an "offshore banking center", and $OFFSHORE_i$ equals 0 otherwise.[17] We expect the β_{15} coefficient to be positive if banks in offshore financial centers classify some foreign account holders as domestic

[15] See https://comtrade.un.org/. Results are quantitatively similar bifurcating *BILAT* into exports and imports.
[16] See http://reports.weforum.org/global-competitiveness-index-2017-2018/competitiveness-rankings/.
[17] Recent data leaks (e.g. "Wikileaks", the "Panama Papers", and the "Paradise Papers") reveal that wealthy individuals often use financial institutions in certain countries as intermediaries to evade taxes and maintain privacy, and such behavior can affect basic economic statistics (Haldevang 2017; Enrich 2018; Pacaud 2017). The U.S. Treasury monitors various countries to identify any improper behavior. The only countries identified by the U.S. Treasury as offshore financial centers in our sample are Luxembourg, Switzerland, Belgium, and the United Kingdom. All other countries identified by the U.S. Treasury as offshore financial centers do not have the necessary data to be included in our sample (e.g. Cyprus, Cayman Islands, etc.).

investors.[18] Finally, $FINSTD_i$ is a measure of the strength of the financial reporting and auditing standards in country *i* from the World Economic Forum. $FINSTD_i$ values are integers ranging from 1 to 7 where higher (lower) values denote stronger (weaker) standards. We expect the β_{16} coefficient to be negative, consistent with research showing that investors in countries with stronger reporting and auditing standards have less incentive to invest abroad (Aggarwal et al. 2005; Bradshaw et al. 2004; Khurana and Michas 2011; Baik et al. 2013).

The next set of control variables (*COMMON* and *PROTECT*) are time-invariant measures of the investor's home country legal environment. $COMMON_i$ is a dummy variable that equals one if country *i* is a former British colony or is listed as a common law country in Table 2 of LaPorta et al. (1998), and $COMMON_i$ equals zero otherwise. We expect the β_{17} coefficient to be positive, consistent with individuals from common law countries preferring to invest in another common law country, such as the U.S. (Khurana and Michas 2011; Bae et al. 2008a; Bae et al. 2008b). $PROTECT_i$ is a measure of the strength of investor protections in country *i* from the World Economic Forum. $PROTECT_i$ values are integers from 1 to 7 where higher (lower) values denote stronger (weaker) protections. We expect the β_{18} coefficient to be negative, consistent with research which shows that investors in countries with stronger investor protections have less incentive to invest abroad (Aggarwal et al. 2005; Khurana and Michas 2011; Baik et al. 2013).

The final set of control variables (*DIST* and *CULTURE*) are time-invariant measures of the cultural environment in an investor's home country. $DIST_i$ is the distance between the capital of country *i* and New York City in kilometers.[19] The geographic distance between two countries

[18] For example, a Russian investor may have a numbered account at a Swiss bank (i.e. an account that does not include any identifying information about the account holder). If the Russian investor uses funds from the numbered account to invest in U.S. stocks, then the Swiss bank may classify the account holder as Swiss (rather than Russian) on TIC Form SLT to mask the account holder's nationality.

[19] Pairwise distance searches are from https://www.timeanddate.com/worldclock/distance.html.

can be an element of the "psychic distance" between the two cultures (Lopez-Duarte and Vidal-Suarez 2010; Lundholm et al. 2014). Prior research shows that greater distances are negatively correlated with the bilateral trade of physical goods (Anderson 2011) and investment in financial securities (Portes et al. 2001; Portes and Rey 2005; Ferreira and Matos 2008; Baik et al. 2013). As such, we expect the β_{19} coefficient on *DIST* to be negative. $CULTURE_i$ is the sum of six sub-measures of culture for country *i* based on Hofstede (2001).[20] We make no predictions on the sign of the β_{20} coefficient because it is not clear how the cultural sub-categories will combine to influence foreign investment in U.S. stocks.

V. Results

Descriptive Statistics

Table 2 presents descriptive statistics for the variables in regression model (1) and variables for changes specifications later in our study. Requiring non-missing control variables results in a sample of 3,582 country-month observations from December 2012 to December 2016. The mean *CS* (ΔCS) value of $63,224 (1.87) indicates that U.S. stock holdings averaged $63 billion and increased by 1.87% of GDP per year. Holding levels and growth are right skewed, and skewness is addressed through the inclusion of the control variables. For example, China's holdings are significantly larger than the median holdings, but so is China's population, GDP per capita growth, etc. The mean (median) linguistic distance is 1.65 (1.75) which approximates the linguistic distance of Switzerland (Germany). The mean (median) *BOG* value

[20] The sub-measures include *power distance* (higher values denote a greater deference to people in power), *individualism* (higher [lower] values denote that the individual [group] is the most important social unit), *masculinity* (higher values denote emphasis on ambition, wealth acquisition, and differentiated gender roles whereas lower values denote emphasis on nurturing behaviors, sexual equality, environmental awareness, and more fluid gender roles), *uncertainty avoidance* (higher values denote a culture that feels more threatened by ambiguity), *long-term orientation* (higher values denote a culture more prepared to delay short-term gratification), and *indulgence* (higher values denote a culture more prepared to allow relatively free gratification of natural human desires). Sub-measure scores are available from https://www.hofstede-insights.com/product/compare-countries/.

of 0.16 (-0.15) approximates 0 after centering the variable. The mean Δ*BOG* value suggests that aggregate financial reporting readability has decreased by an average of 0.67% per year.

With respect to the control variables, Table 2 shows that equity (30-year U.S. Treasury) returns averaged 12.44% (5.19%) per year. The average oil price was $79.84 per barrel, and oil prices declined by an average of 17.28% per year over the sample period. The average country's population was 72.58 million and grew at 1.00% per year. G-20 countries constitute 25% of our sample. The average country's bilateral trade with the U.S. is $47 billion and grew at 1.72% per year. The average country's GDP per capita was $23,149 and grew at an average of 1.69% per year. The average country's inflation was 3.10% per year. The average country's corporate tax rate is 40.87%. Offshore banking center (common law) countries constitute 6% (23%) of our sample. The average country capital is separated from New York City by 8,019 kilometers. The mean *FINSTD* and *PROTECT* values of 5.07 and 6.07, respectively, are more difficult to interpret because the variables are simply integers from 1 to 7. Similarly, the mean *CULTURE* value of 298.44 has no real economic connotation because of the arbitrary scoring methodology. Overall, our descriptive statistics appear reasonable and consistent with prior studies.

Table 3 presents univariate correlations. Correlations significant at the 5% level are in bold, and we discuss only the Pearson correlations between the primary variables of interest for brevity. The significantly negative correlation between *CS* and *LD* of -0.28 is consistent with investors from countries with larger linguistic distances investing less in U.S. stocks. The significantly positive correlation between *CS* and *BOG* of 0.04 is inconsistent with foreign investors holding less U.S. stocks when U.S. financial reports are more difficult to read. However, this positive correlation is likely the result of both variables trending upward (see Figure 1). We control for time trends in the regressions by including year fixed effects.

Primary Results

Table 4 presents the results from regression model (1). We estimate regression model (1) with year fixed effects and cluster the standard errors by country. The first column of Table 4 presents the results of regression model (1) estimated without the interaction of *LD* and *BOG*. The advantage of omitting the interaction term is the interpretation of the main effect coefficients. However, the disadvantage of omitting the interaction is that if the interaction term is significant in the "true" model, then omitting the interaction term can introduce omitted correlated variable bias (Greene 2017). With respect to the variables of interest, the β_1 coefficient on *LD* is negative and significant at the 10% level. This result suggests rejection of the null hypothesis for H1 in favor of the alternative that investors from countries with greater linguistic distances hold less U.S. stocks. Interpreting the magnitude of the *LD* coefficient is complicated because linguistic distance is fixed for a country over time. Nevertheless, our results suggest that a one unit change in *LD* (which would be equivalent to Italians suddenly speaking Arabic) is associated with a reduction in U.S. stock holdings by an economically significant $638 million. Next, the β_2 coefficient on *BOG* is not statistically significant which indicates failure to reject the null hypothesis for H2 in the reduced model.

The second column of Table 4 presents the results of regression model (1) estimated with the interaction term. The β_1 coefficient on *LD* remains negative and significant (now at the 1% level), again suggesting that investors from countries with greater linguistic distances hold less U.S. stocks. Next, the β_2 coefficient on *BOG* of -8,207.01 is negative and significant at the 1% level. This result suggests rejection of the null hypothesis for H2 in favor of the alternative that foreigners invest less in U.S. stocks when U.S. financial reports are more difficult to read. Interpreting the magnitude of the main effect coefficients requires holding all other variables

constant (including the interaction). Thus, the magnitude of the β_2 coefficient in the presence of the interaction term reflects the impact of a one-unit change in *BOG* when *LD*=0. Therefore, the results suggest that a one unit change in *BOG* (approximately 1.5 years of change at the average growth rate) reduces holdings of U.S. stocks for investors in English-speaking countries where *LD*=0 (i.e. the U.K., Ireland, Australia, New Zealand, Trinidad and Tobago, and Jamaica) by an economically significant $8.2 billion.

Finally, the β_3 coefficient on the interaction of *LD* and *BOG* of 5,237.35 is significantly positive. The significant coefficient on the interaction term helps explain the lack of significance of the β_2 coefficient on *BOG* in the reduced model. More specifically, the β_2 and β_3 coefficients have opposing signs in the full model, but Table 3 shows that *BOG* and *LD**BOG* are positively correlated (r=0.92, p<0.01). Thus, excluding the interaction term introduces omitted correlated variable bias. Notably, the β_3 coefficient on the interaction term is less in absolute magnitude than the coefficient on *BOG* which suggests the interaction attenuates but does not fully eliminate the impact of financial reporting readability. More formally, the negative β_3 coefficient suggests rejection of the null hypothesis for H3 in favor of the alternative that foreign investors from high linguistic distance countries facing unreadable financial reports reduce their investments less than foreign investors from low language distance countries facing the same unreadable financial reports. Overall, foreign investors appear to be sensitive to linguistic distance and financial reporting readability but seem reluctant to reduce their investments in U.S. stocks below a certain floor (perhaps because their next best option involves either investing in their home country which reduces global diversification or investing in another country where they may continue to face a linguistic distance and financial reports that are difficult to read).

The second column of Table 4 also shows that all control variables in the full model are significant at the 5% level except *PROTECT*. Moreover, the sign of the coefficients on each control variable for which we offer a directional prediction is consistent with our predictions. Finally, the adjusted R^2 of 0.980 suggests our model explains nearly all variation in foreign holdings of U.S. stocks. The high R^2 value and the presence of a lagged dependent variable raises the possibility of non-stationarity (e.g. a unit root). As described previously, we estimate changes models later in the paper to reduce the potential for non-stationarity and unobserved time-invariant heterogeneity, and all results are quantitatively and qualitatively similar.

Robustness Checks

The results in Table 4 are consistent with linguistic distance and financial reporting readability reducing foreign investors' willingness to hold U.S. stocks. Table 5 presents the results from a variety of robustness tests. Panel A of Table 5 replicates regression model (1) utilizing an alternative linguistic distance from Isphording (2014). As discussed in Section II, the Isphording (2014) measure has the advantage of being symmetric (although the phonetic approach relies on a computer algorithm to measure pronunciation similarity). The coefficients on *LD* and *BOG* remain negative and significant while the interaction term remains positive and significant. Additionally, the control variables all behave similarly to the main results in Table 4. Overall, inferences are similar regardless of which linguistic distance measure is used.

Next, Demirbag et al. (2007) observe that most studies examining the consequences of language use a simple dummy variable for a common language which implicitly treats all foreign languages the same (e.g. Lundholm et al. 2014; Baik et al. 2013). We exclude the United Kingdom, Ireland, Australia, New Zealand, Trinidad and Tobago, and Jamaica where $LD = 0$, thereby creating a sample where a common language dummy variable would equal zero. We also

subtract 0.33 (Canada's *LD* value) from each remaining country's *LD* value so that *LD* can equal zero (i.e. Canada becomes the reference country) and we can interpret the main effect in the presence of the interaction term. Panel B of Table 5 shows that the coefficients on *LD* and *BOG* remain negative and significant while the interaction term remains positive and significant. This suggests that linguistic distance and financial readability matter even amongst countries with strictly positive linguistic distances. More importantly, these results suggest that other studies should consider using measures of linguistic distance rather than simple dummy variables to allow the economic consequences of language barriers to differ across languages.

Our next tabulated robustness check uses only annual data. As described above, the TIC holdings data are monthly while many of the control variables (e.g. macroeconomic data from the World Bank) are annual. In order to eliminate this inconsistency, we use only TIC holdings data at the end of January of each year. Using January data minimizes the time between the TIC holdings data and the associated control variables and yields a maximum of four observations per country (e.g. one observation per country for 2013-2016). However, regression model (1) contains an intercept, four variables that are the same for every country for a given month (*BOG*, *EQUITY*, *TREAS*, and *OIL*), and is estimated using year fixed effects. Thus, when using annual data, we must suppress the intercept and eliminate the year fixed effects (and instead cluster both by country and year) or else there would be insufficient degrees of freedom to estimate the model. Panel C of Table 5 shows that the coefficients on *LD* and *BOG* remain negative and significant while the interaction term remains positive and significant. Thus, our full sample results are robust to the use of annual data.

Finally, we estimate regression model (1) augmented with several additional control variables. Specifically, we add (1) the percentage of the population of country *i* speaking a

language identified as having a weak future time reference in Chen (2013), (2) an estimate of the total home bias of investors in country *i* from Cooper et al. (2017), (3) a measure of the quality of primary education in country *i* from the World Economic Forum, (4) the variance in exchange rates between country *i*'s local currency and U.S. dollars, (5) a numeric variable equal to |1| {2} [3] if the International Monetary Fund classifies country *i* as having |"Open"| {"Gate"} ["Wall"] capital controls in place which limit foreign investment by their citizens, (6) the scaled number of immigrants from country *i* living in the U.S. as of 2010 from the Migration Policy Institute, (7) the percentage of adults in country *i* that are financially literate according to the S&P Global FinLit Survey, (8) the most recent monthly percentage change in the local stock index of country *i* using data from Thomson Reuters, and (9) the percentage of country *i* identifying as religious from the Pew Center. We do not include these variables in our primary regression specification in Table 4 because inclusion of these variables reduces our sample size by 53% (including the exclusion of economically important countries such as China) and introduces multicollinearity (e.g. education quality is highly correlated with GDP per capita). Nevertheless, the coefficients on *LD* and *BOG* remain negative and significant while the interaction coefficient remains positive and significant.

We also perform a variety of untabulated robustness checks. First, we re-estimate regression model (1) excluding all offshore banking center countries. The coefficients on *LD* and *BOG* remain negative and significant while the interaction coefficient remains positive and significant. Thus, the full-sample results do not appear to be driven by the presence of offshore financial centers. Second, we estimate regression model (1) separately for G-20 countries and non-G-20 countries. The coefficients on *LD* and *BOG* are negative and significant in both sub-samples, and the interaction coefficient is positive and significant in both sub-samples. This

suggests that our results hold for the largest and most developed countries as well as the smaller and less developed countries. Finally, we re-estimate regression model (1) excluding one country at a time. The coefficients on *LD* and *BOG* are negative and significant and the interaction coefficient is positive and significant in all 75 specifications, suggesting no one country in our sample drives our results.

Changes Analyses

One econometric concern with the levels specifications presented so far is that regression model (1) is an autoregressive model. The coefficient on the lagged holdings variable is statistically significant in Table 4, which suggests that including the lag term is necessary to prevent autocorrelation in the residuals. However, the coefficient on the lagged holdings variable is statistically indistinguishable from one which raises concerns about a unit root (i.e. non-stationarity) and potentially inflates the model's R^2. Another concern with regression model (1) is the potential for unobserved time-invariant heterogeneity. One solution for both concerns is the use of changes specifications.[21] We estimate two separate changes analyses. First, we subtract the lagged holdings variable from both sides of model, and the dependent variable becomes ($CS_{i,j} - CS_{i,j\text{-}12}$). The resulting model is a "pseudo changes" model in that the dependent variable is an unscaled changes variable, but the remaining independent variables are unaltered (generally as levels variables). In a second changes specification, we scale the change variable by GDP to reduce heteroskedasticity and mitigate the impact of outliers (e.g. avoid small denominator problems). The dependent variable is ($CS_{i,j} - CS_{i,j\text{-}12}$)/$GDP_{i,t\text{-}1}$, and we also transform the independent variables to changes variables where possible. Specifically, Δ*BOG*,

[21] Another common approach for the potential presence of unobserved time-invariant heterogeneity is a fixed effects model. Unfortunately, estimating a model with country fixed effects is problematic in our setting because one of the primary variables of interest (*LD*) and multiple control variables are all time-invariant for a given country. Additionally, introducing fixed effects into an autoregressive model results in biased parameters (Nickell 1981).

ΔOIL, ΔPOP, and $\Delta BILAT$ represent percentage changes from twelve months prior. *EQUITY*, *TREAS*, and $\Delta GDPPC$ are already changes variables, and the other independent variables remain as levels variables because they are time-invariant.

Table 6 presents the results of the changes analyses. The first column of Table 6 presents the results of the pseudo changes model. The coefficients on *LD* and *BOG* remain negative and significant while the interaction coefficient remains positive and significant. Moreover, the magnitude of these coefficients closely approximates those in Table 4, and the behavior of all control variables also matches those in the primary analysis. Finally, the adjusted R^2 drops to 0.282, consistent with moving from a levels model to a changes model. The second column of Table 6 presents the specification using the scaled changes variable. The negative coefficient on *LD* suggests that a one-unit change in linguistic distance reduces a country's holdings of U.S. stocks by an economically meaningful 0.63% of the country's GDP (when ΔBOG equals zero). Similarly, the negative coefficient on ΔBOG suggests that a one-unit increase in *BOG* reduces a country's holdings of U.S. stocks by 2.11% of GDP (for countries where $LD=0$). The coefficient on the interaction term of 0.97% is significant but smaller in magnitude than the coefficient on *BOG*, suggesting that the interaction attenuates but does not fully eliminate the impact of a change in financial reporting readability on linguistically distant countries. While neither specification in Table 6 is a "pure" changes analysis (e.g. because *LD* is time-invariant), the inferences from the changes analyses are consistent with those from the levels analyses.

In untabulated analyses, we also re-estimate the four robustness tests in Table 5 (using an alternative linguistic distance measure, excluding English speaking countries, using annual data, and including additional controls) in both pseudo changes and full changes for a total of eight additional specifications. For seven of those eight specifications, the results are quantitatively

and qualitatively similar to the changes results reported in Table 6 (statistically significant negative coefficients on ΔLD and ΔBOG and a significantly positive interaction coefficient). When using annual data in full changes, the sample size decreases by 92% but we continue to find a significantly negative ΔLD coefficient and a significantly positive interaction coefficient (similar to all prior results). The coefficient on ΔBOG is not statistically significant, potentially due to a lack of power in the smaller sample.

Additional Analysis

The results in Tables 4 through 6 suggest that foreign investors reduce their holdings of U.S. stocks in response to linguistic distance and difficulty reading U.S. financial reports. A natural follow-up question is "Where do foreign investors then direct those funds?" Foreign investors' options include, but are not limited to, (1) consuming rather than investing, (2) re-allocating their U.S. equity investments (e.g. divesting from U.S. firms with less readable financial reports and investing in U.S. firms with more readable financial reports), and (3) investing in other foreign countries. Data limitations prevent us from examining these particular alternatives. Specifically, we do not have foreign citizen consumption data, our dependent variable would be unaffected if foreign investors merely sell some of one U.S. stock and buy an equal dollar amount another, and we only have foreign investment positions for U.S. securities. However, we can explore one potential alternative investment vehicle for foreign investors. In our final analysis, we examine whether foreign investors direct a portion of their spare investment funds towards U.S. Treasuries.

We re-estimate regression model (1) using country *i*'s holdings of U.S. Treasuries in millions of real 2010 U.S. dollars at the end of month *j* from TIC Table 1D as the dependent

variable ($US_{i,j}$). Table 7 presents the results.[22] The coefficients on *LD* and *BOG* are positive and significant, and the coefficient on the interaction term is negative and significant. These results are the mirror image of the U.S. stock results in Table 4 and suggest a "substitution effect." In other words, foreign investors who want to invest in U.S. stocks, but are sufficiently deterred by linguistic distance and financial reporting readability frictions, appear to seek the relative simplicity of investing in U.S. Treasuries. These results are meaningful because they suggest that linguistic distance and financial reporting readability increase the demand for U.S. Treasuries (a unique result given that frictions generally reduce demand). Moreover, these results suggest that financial reporting readability can influence the demand for assets other than corporate stocks, including U.S. Treasuries that do not have accompanying financial statements.

Overall, our results suggest that linguistic distance and financial reporting readability are negatively (positively) associated with foreign holdings of U.S. stocks (U.S. Treasuries). We acknowledge the limitations of our aggregate data as a caveat to our results and inferences. For example, we cannot identify investor-level characteristics others than nationality (e.g. whether the investor is an individual, institutional investor, firm, or government or whether the investor is sophisticated or unsophisticated). Similarly, we are unable to identify holdings characteristics (e.g. whether foreign investors tend to own stocks with easy or difficult to read financial reports).

VI. Conclusion

We demonstrate that linguistic distance and financial reporting readability are both negatively associated with foreign holdings of U.S. stocks. Our results suggest that foreign investors are sensitive to linguistic distance and U.S. financial reporting readability but are reluctant to reduce their investments in U.S. stocks below a certain floor (perhaps because their

[22] The coefficient on the lagged holdings variable is less than one ($p<0.01$) indicating a unit root is not present. Thus, we present only the levels specification.

next best option involves either investing in their home country which reduces global diversification or investing in another country where they will still likely face a linguistic distance and financial reports that are difficult to read). Additionally, we find that linguistic distance and financial reporting readability are positively associated with foreign holdings of U.S. Treasuries. This suggests that foreign investors who want to invest in U.S. stocks, but are sufficiently deterred by linguistic distance and financial reporting readability frictions, appear to seek the relative simplicity of investing in U.S. Treasuries.

These results should be of interest to researchers and policy makers. First, our results suggest that linguistic distance and financial reporting readability persist as significant frictions, even in the U.S. market where foreign investors should have the greatest ability, resources, and economic incentives to overcome these issues. Second, our results suggest that linguistic distance and financial reporting readability increase the demand for U.S. Treasuries (a novel result considering that U.S. Treasuries do not have associated financial statements and that frictions generally reduce demand). Finally, our study has implications for future research. A growing literature examines the relation between firm-level financial reporting outputs and macroeconomic outcomes (see Konchitchki and Patatoukas 2014; Crawley 2015; Gallo et al. 2016; Nallareddy and Ogneva 2017; Shivakumar and Urcan 2017). Relatedly, our results extend to the aggregate level and suggest that linguistic distance and financial reporting readability can alter the absolute and relative demand for U.S. stocks and U.S. Treasuries. As such, future research may investigate whether changes in U.S. Treasury rates (i.e., the cost of capital for the U.S. government) and the cost of capital of U.S. firms influence firms' investment opportunity sets and total macroeconomic output.

References

Aggarwal, R., L. Klapper, and P.D. Wysocki. 2005. Portfolio Preferences of Foreign Institutional Investors. *Journal of Banking and Finance* 29: 2919-2946.

Ahearne, A.G., W.L. Griever, and F.E. Warnock. 2004. Information Costs and Home Bias: An Analysis of US Holdings of Foreign Equities. *Journal of International Economics* 62: 313-336.

Anderson, J.E. 2011. The Gravity Model. *Annual Review of Economics* 3: 133-160.

Arbuckle, A.Q. 2017. The Navajo Code Talkers that Helped the U.S. Win WWII. *Mashable* November 11, 2017.

Bae, K.-H., R.M. Stulz, and H. Tan. 2008a. Do Local Analysts Know More? A Cross-Country Study of the Performance of Local Analysts and Foreign Analysts. *Journal of Financial Economics* 88: 581-606.

Bae, K.-H., H. Tan, and M. Welker. 2008b. International GAAP Differences: The Impact on Foreign Analysts. *The Accounting Review* 83: 593-628.

Baik, B., J.-K. Kang, J.-M. Kim, and J. Lee. 2013. The Liability of Foreignness in International Equity Investments: Evidence from the US Stock Market. *Journal of International Business Studies* 44: 391-411.

Bonsall IV, S.B., A.J. Leone, B.P. Miller, and K. Rennekamp. 2017. A Plain English Measure of Financial Reporting Readability. *Journal of Accounting and Economics* 63: 329-357.

Bonsall IV, S.B., and B.P. Miller. 2017. The Impact of Narrative Disclosure Readability on Bond Ratings and the Cost of Debt Capital. *Review of Accounting Studies* 22: 608-643.

Bozanic, Z., and M. Thevenot. 2015. Qualitative Disclosure and Changes in Sell-Side Financial Analysts' Information Environment. *Contemporary Accounting Research* 32: 1595-1616.

Bradshaw, M.T., B.J. Bushee, and G.S. Miller. 2004. Accounting Choice, Home Bias, and U.S. Investment in Non-U.S. Firms. *Journal of Accounting Research* 42: 795-841.

Brandner, E., F. Cai, and R. Judson. 2012. Improving the Measurement of Cross-Border Securities Holdings: The Treasury International Capital SLT. *Federal Reserve Bulletin* 98: 1-28.

Bump, P. 2017. The Asterisk that Accompanies the Gains in the Stock Market: A Third of the Shares are Owned by Foreigners. *The Washington Post* October 24, 2017.

Chan, K., V. Covrig, and L. Ng. 2005. What Determines the Domestic Bias and Foreign Bias? Evidence from Mutual Fund Equity Allocations Worldwide. *The Journal of Finance* 60: 1495-1534.

Chen, M.K. 2013. The Effect of Language on Economic Behavior: Evidence from Savings Rates, Health Behaviors, and Retirement Assets. *American Economic Review* 103: 690-731.

Chen, J., R.R. Sokal, and M. Ruhlen. 1995. Worldwide Analysis of Genetic and Linguistic Relationships of Human Populations. *Human Biology* 67: 595-612.

Chiswick, B.R., and P.W. Miller. 2005. Linguistic Distance: A Quantitative Measure of the Distance Between English and Other Languages. *Journal of Multilingual and Multicultural Development* 26: 1-11.

Clark, W.S. 2008. Tax Effects on Foreign Direct Investment. *Organization for Economic Co-Operation and Development*.

Cooper, I.A., P. Secru, and R. Vanpee. 2018. A Measure of Pure Home Bias. *Review of Finance* 22: 1469-1514.

Crawley, M.J. 2015. Macroeconomic Consequences of Accounting: The Effect of Accounting Conservatism on Macroeconomic Indicators and the Money Supply. *The Accounting Review* 90: 987-1011.

Demirbag, M., K.W. Glaister, and E. Tatoglu. 2007. Institutional and Transaction Cost Influences on MNEs' Ownership Strategies of Their Affiliates: Evidence from an Emerging Market. *Journal of World Business* 42: 418-434.

Enrich, D. 2018. A Swiss Banker Helped Americans Dodge Taxes. Was It a Crime? *The New York Times* January 6, 2018.

Ferreira, M.A., and P. Matos. 2008. The Colors of Investors' Money: The Role of Institutional Investors Around the World. *Journal of Financial Economics* 88: 499-533.

Firtel, K.B. 1999. Plain English: A Reappraisal of the Intended Audience of Disclosure Under the Securities Act of 1933. *Southern California Law Review* 851: 851-898.

Fredriksson, R., W. Barner-Rasmussen, and R. Piekkari. 2006. The Multinational Corporation as a Multilingual Organization: The Notion of a Common Corporate Language. *Corporate Communications: An International Journal* 11: 406-423.

Gallo, L.A., R.N. Hann, and C. Li. 2016. Aggregate Earnings Surprises, Monetary Policy, and Stock Returns. *Journal of Accounting and Economics* 62: 103-120.

Ginsberg, V., and S. Weber. 2016. *The Palgrave Handbook of Economics and Language*. London, U.K. Palgrave MacMillan.

Greene, W.H. 2017. *Econometric Analysis*. New York, N.Y. Pearson.

Guay, W., D. Samuels, and D. Taylor. 2016. Guiding Through the Fog: Financial Statement Complexity and Voluntary Disclosure. *Journal of Accounting and Economics* 62: 234-269.

Guiso, L., P. Sapienza, and L. Zingales. 2009. Cultural Biases in Economic Exchange? *The Quarterly Journal of Economics* 124: 1095-1131.

Gunning, R. 1952. *The Technique of Clear Writing*. New York, N.Y. McGraw-Hill International Book Company.

Haldevang, M. 2017. Why We Can't Trust Basic Economic Figures. *Quartz* November 28, 2017.

Hart-Gonzalez, L. and S. Lindemann. 1993. Expected Achievement in Speaking Proficiency. School of Language Studies, Foreign Services Institute, Department of State (mimeo).

Henry, P.B. 2000a. Do Stock Market Liberalizations Cause Investment Booms? *Journal of Financial Economics* 58: 301-334.

Henry, P.B. 2000b. Stock Market Liberalization, Economic Reform, and Emerging Market Equity Prices. *The Journal of Finance* 55: 529-564.

Hofstede, G. 2001. Culture's consequences: International differences in work-related values (2 ed.). Beverly Hills: Sage.

Isphording, I.E. 2014. Disadvantages of Linguistic Origin: Evidence from Immigrant Literacy Scores. *Economics Letters* 123: 236-239.

Isphording, I.E., and S. Otten. 2013. The Costs of Babylon—Linguistic Distance in Applied Economics. *Review of International Economics* 21: 354-369.

Kang, J.-K., and R. Stulz. 1997. Why is there a Home Bias? An Analysis of Foreign Portfolio Equity Ownership in Japan. *Journal of Financial Economics* 46: 3-28.

Karolyi, G.A., and R. Stulz. 2003. Are financial assets priced locally or globally? *Handbook of the Economics of Finance.* 1: 975-1020. Elsevier.

Khazan, O. 2013. Language Distance: The Reason Immigrants Have Trouble Assimilating. *The Atlantic* May 7, 2013.

Khurana, I.K., and P.N. Michas. 2011. Mandatory IFRS Adoption and the U.S. Home Bias. *Accounting Horizons* 25: 729-753.

Kim, J., Y. Kim, and J. Zhou. 2017. Languages and Earnings Management. *Journal of Accounting and Economics* 63: 288-306.

Konchitchki, Y., and P.N. Patatoukas. 2014. Accounting Earnings and Gross Domestic Product. *Journal of Accounting and Economics* 57: 76-88.

Kothari, S.P. 2001. Capital Markets Research in Accounting. *Journal of Accounting and Economics* 31: 105-231.

La Porta, R., F. Lopez-de-Silanes, A. Shleifer, and R. Vishny. 1998. Law and Finance. *Journal of Political Economy* 106: 1113-1155.

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

Lehavy, R., F. Li, and K. Merkley. 2011. The Effect of Annual Report Readability on Analyst Following and the Properties of Their Earnings Forecasts. *The Accounting Review* 86: 1087-1115.

Li, F. 2008. Annual Report Readability, Current Earnings, and Earnings Persistence. *Journal of Accounting and Economics* 45: 221-247.

Lin, A.Y., and P.E. Swanson. 2003. The Behavior and Performance of Foreign Investors in Emerging Equity Markets: Evidence from Taiwan. *International Review of Finance* 4: 189-210.

Lizardo, R.A., and A.V. Mollick. 2009. Do Foreign Purchases of U.S. Stocks Help the U.S. Stock Market? *Journal of International Financial Markets, Institutions, and Money* 19: 969-986.

Lopez-Duarte, C., and M.M. Vidal-Suarez. 2010. External Uncertainty and Entry Mode Choice: Cultural Distance, Political Risk and Language Diversity. *International Business Review* 19: 575-588.

Loughran, T., and B. McDonald. 2014. Measuring Readability in Financial Disclosures. *The Journal of Finance* 69: 1643-1671.

Lundholm, R.J., R. Rogo, and J.L. Zhang. 2014. Restoring the Tower of Babel: How Foreign Firms Communicate with U.S. Investors. *The Accounting Review* 89: 1453-1485.

Miller, B.P. 2010. The Effects of Reporting Complexity on Small and Large Investor Trading. *The Accounting Review* 85: 2107-2143.

Nallareddy, S., and M. Ogneva. 2017. Predicting Restatements in Macroeconomic Indicators using Accounting Information. *The Accounting Review* 92: 151-182.

Neeley, T. 2012. Global Business Speaks English. *Harvard Business Review*.

Nickell, S. 1981. Biases in Dynamic Models with Fixed Effects. *Econometrica* 49: 1417-1426.

Pacaud, J. 2017. One Man's Fight Against the Swiss Offshore Banking System. *The Economist* December 23, 2017.

Portes, R., H. Rey, and Y. Oh. 2001. Information and Capital Flows: The Determinants of Transactions in Financial Assets. *European Economic Review* 45: 783-796.

Portes, R., and H. Rey. 2005. The Determinants of Cross-Border Equity Flows. *Journal of International Economics* 65: 269-296.

Rennekamp, K. 2012. Processing Fluency and Investors' Reactions to Disclosure Readability. *Journal of Accounting Research* 50: 1319-1354.

Rosenthal, S.M. 2017. Slashing Corporate Taxes: Foreign Investors Are Surprise Winners. *Tax Notes* October 23, 2017.

Sayek, S. 2009. Foreign Direct Investment and Inflation. *Southern Economic Journal* 76: 419-443.

Securities and Exchange Commission. 1998. *A Plain English Handbook: How to Create Clear SEC Disclosure Documents.* Washington, D.C.

Shivakumar, L., and O. Urcan. 2017. Why Does Aggregate Earnings Growth Reflect Information about Future Inflation? *The Accounting Review* 92: 247-276.

Stulz, R. 1999a. Globalization, Corporate Finance, and the Cost of Capital. *Journal of Applied Corporate Finance* 12: 8-25.

Stulz, R. 1999b. International Portfolio Flows and Security Markets. Working Paper, Ohio State University.

Stulz, R., and W. Wasserfallen. 1995. Foreign Equity Investment Restrictions, Capital Flight, and Shareholder Wealth Maximization: Theory and Evidence. *The Review of Financial Studies* 8: 1019-1057.

Taylor I. 1980. The Korean writing system: An alphabet? A syllabary? A logography? *Processing of Visible Language*. Nato Conference Series. Springer, Boston, MA

Tietze, S. 2004. Spreading the Management Gospel – in English, Language and Intercultural Communication. *Language and Intercultural Communication* 4: 175-189.

United States Treasury, Federal Reserve Bank of New York. 2016. *Foreign Portfolio Holdings of U.S. Securities.* New York, NY.

You, H., and X.-J. Zhang. 2009. Financial Reporting Complexity and Investor Underreaction to 10-K Information. *Review of Accounting Studies* 14: 559-586.

Appendix

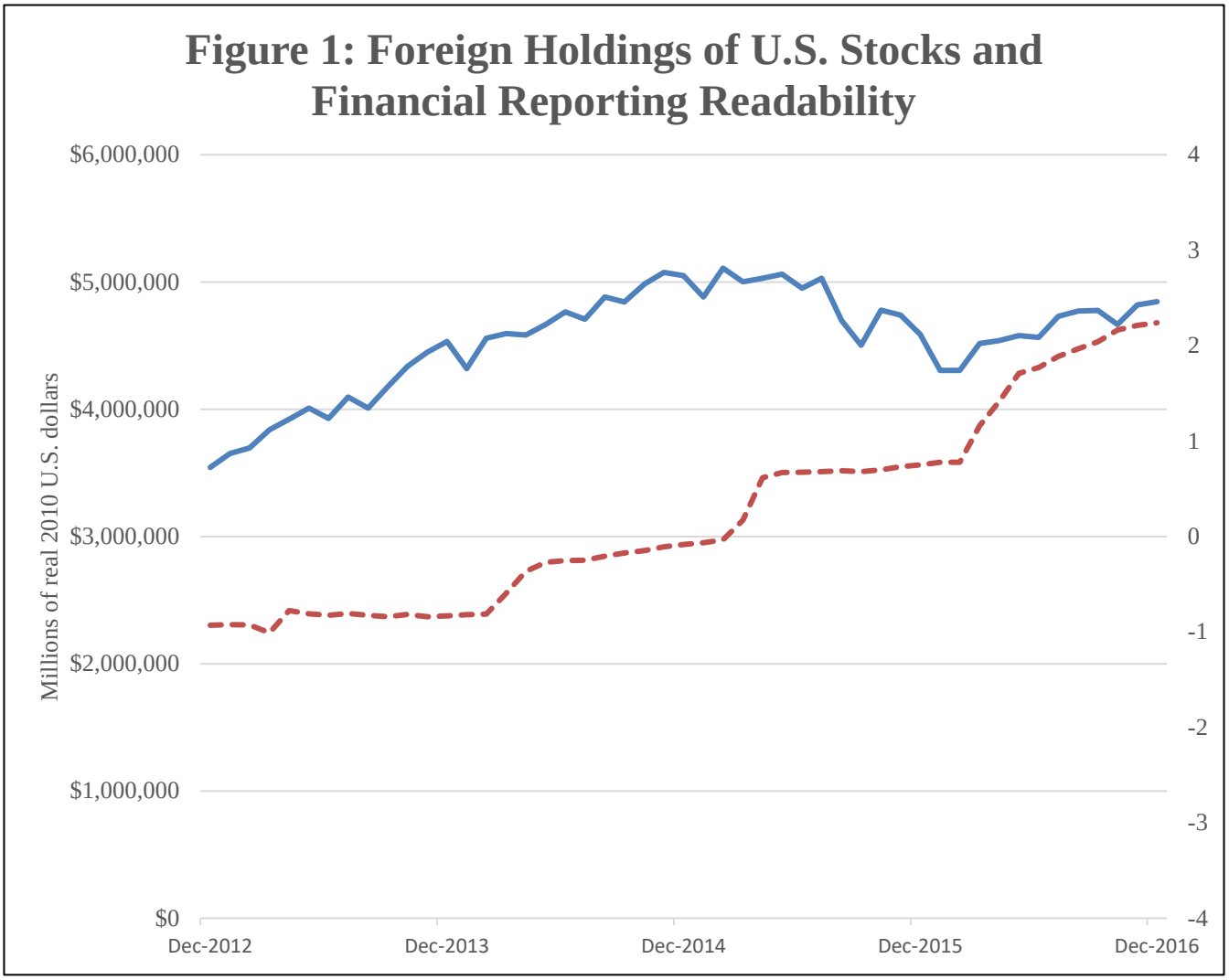
Variable Definitions

$CS_{i,j}$	Total holdings of U.S. stocks by investors in country i at the end of month j from U.S. Treasury International Capital Table 1D.
$US_{i,j}$	Total holdings of U.S. Treasuries by investors in country i at the end of month j from U.S. Treasury International Capital Table 1D.
LD_i	Linguistic distance of country i based on Chiswick and Miller (2005). See Section III for further discussion.
BOG_{j-1}	Aggregate U.S. financial reporting readability at the end of month j-1 based on Bonsall et al. (2017). Higher values denote less readability. See Section III for further discussion.
$EQUITY_{j-1}$	Monthly returns on the Center for Research in Securities Prices (CRSP) value-weighted index from month j-12 to month j-1 from the CRSP database.
$TREAS_{j-1}$	Monthly returns on 30-year U.S. Treasury bonds from month j-12 to month j-1 from the CRSP database.
OIL_{j-1}	Monthly spot price of Brent crude oil at the end of month j-1 from the Federal Reserve Economic Database.
$POP_{i,t-1}$	Population of country i (in millions) at the end of year t-1 (i.e. the most recent calendar year ending before month j) from the World Bank.
$BILAT_{i,t-1}$	Bilateral trade (i.e. the sum of imports and exports) between country i and the U.S. in year t-1 from the United Nations Commodity Trade Statistics Database.
$GDPPC_{i,t-1}$	Gross domestic product per capita in real 2010 U.S. dollars for country i in year t-1 from the World Bank.
$\Delta GDPPC_{i,t-1}$	Percentage change in gross domestic product per capita in real 2010 U.S. dollars for country i from year t-2 to t-1 from the World Bank.
$INF_{i,t-1}$	Inflation for country i in year t-1 from the World Bank.
$G20_i$	Indicator variable equal to 1 if country i is a member of the G-20 and 0 otherwise.
TAX_i	Average corporate tax rate (including corporate income taxes, payroll taxes, retirement contributions, and other taxes) for country i from the World Economic Forum.
$OFFSHORE_i$	Indicator variable equal to 1 if the U.S. Treasury has identified country i as an "offshore banking center" and 0 otherwise.
$FINSTD_i$	The strength of the financial reporting and auditing standards in country i from the World Economic Forum. Values are integers from 1 to 7, and higher (lower) values denote stronger (weaker) standards.
$COMMON_i$	Indicator variable that equals 1 if country i is a former British colony or is listed as a common law country in Table 2 of LaPorta et al. (1998) and 0 otherwise.
$PROTECT_i$	The strength of investor protections in country i from the World Economic Forum. Values are integers from 1 to 7, and higher (lower) values denote stronger (weaker) investor protections.
$DIST_i$	Distance between the capital of country i and New York City in kilometers.
$CULTURE_i$	Sum of six sub-measures of culture for country i based on Hofstede (2001). The sub-measures include *power distance* (higher values denote a greater deference to people in power), *individualism* (higher [lower] values denote that the individual [group] is the most important social unit), *masculinity* (higher values denote emphasis on ambition, wealth acquisition, and differentiated gender roles whereas lower values denote emphasis on nurturing behaviors, sexual equality, environmental awareness, and more fluid gender roles), *uncertainty avoidance* (higher values denote a culture that feels more threatened by ambiguity), *long-term orientation* (higher values denote a culture more prepared to delay short-term gratification), and *indulgence* (higher values denote a culture more prepared to allow relatively free gratification of natural human desires).

Appendix (cont.)

Variable Definitions

$WEAKFTR_i$	The percentage of the population of country i speaking a language identified as having a weak future time reference in Chen (2013) Appendix B Table 1.
$HOMEBIAS_i$	An estimate of the total home bias of investors in country i from Cooper et al. (2017) Table 3.
$EDUC_i$	A measure of the quality of primary education in country i ranging from 1 (low) to 7 (high) from the World Economic Forum.
$ERVAR_{i,t-1}$	The variance in the weekly exchange rate between the local currency in country i and U.S. dollars over year t-1. Weekly exchange rates are from OANDA.
$CAPCONTROL_i$	A numeric variable equal to 1 (2) [3] if the International Monetary Fund classifies country i as having "Open" ("Gate") ["Wall"] capital controls in place which limit foreign investment by their citizens.
$IMMIGRANT_{i,t-1}$	The number of immigrants from country i living in the U.S. as of 2010 from the Migration Policy Institute, scaled by the population of country i in year t-1.
$FINLIT_i$	The percentage of adults in country i that are financially literate according to the S&P Global FinLit Survey.
$HOMERET_{i,j-1}$	The percentage change in the local stock index of country i from month j-2 to j-1. Stock index data are from Thomson Reuters.
$RELIGIOUS_i$	The percentage of country i identifying as religious from the Pew Center.



This figure plots the time series of two monthly data series. The sample period begins in December 2012 and ends in December 2016. The solid line denotes total foreign holdings of U.S. stocks from U.S. Treasury International Capital Table 1D in millions of real 2010 U.S. dollars. The dashed line denotes an aggregate measure of financial reporting readability based on Bonsall et al.'s (2017) plain English "Bog Index" for firm-level 10-K filings (higher values denote less readability). See Section III for further discussion.

Table 1

Linguistic Distance and Holdings of U.S. Stocks by Country

Country	Linguistic Distance	Average U.S. Stock Holdings (millions of real 2010 U.S. dollars)	Country	Linguistic Distance	Average U.S. Stock Holdings (millions of real 2010 U.S. dollars)
Japan	3.00	348,413	Uruguay	1.75	1,901
Korea, South	3.00	52,804	Costa Rica	1.75	725
Hong Kong	2.75	65,555	Ecuador	1.75	460
China, Mainland	2.58	249,541	Russia	1.75	428
Kuwait	2.50	119,362	Dominican Republic	1.75	288
Saudi Arabia	2.50	53,798	Guatemala	1.75	265
Malta	2.50	576	El Salvador	1.75	200
Lebanon	2.50	520	Latvia	1.75	165
Egypt	2.50	350	Honduras	1.75	71
Morocco	2.50	79	Ukraine	1.75	54
Vietnam	2.50	31	Lithuania	1.75	42
Pakistan	2.31	133	Switzerland	1.66	300,365
Greece	2.25	3,595	Luxembourg	1.63	437,787
Sri Lanka	2.25	32	France	1.50	131,213
Israel	2.05	33,353	Italy	1.50	22,709
Finland	2.00	17,042	Brazil	1.50	5,701
Thailand	2.00	2,629	Portugal	1.50	2,634
Poland	2.00	2,371	Belgium	1.35	29,767
Czech Republic	2.00	1,819	Netherlands	1.25	181,628
Philippines	2.00	1,657	Singapore	1.25	127,982
Slovenia	2.00	902	United Arab Emirates	1.25	82,318
Hungary	2.00	643	Malaysia	1.25	9,720
Indonesia	2.00	333	Iceland	1.25	637
Turkey	2.00	332	Tanzania	1.25	13
Croatia	2.00	201	South Africa	1.13	6,819
Bulgaria	2.00	65	India	1.13	1,822
Slovakia	2.00	50	Norway	1.00	165,887
Serbia	2.00	15	Sweden	1.00	99,409
Germany	1.75	127,822	Romania	1.00	53
Denmark	1.75	53,777	Kenya	0.63	48
Mexico	1.75	35,117	Canada	0.33	649,310
Chile	1.75	23,230	United Kingdom	0.00	667,998
Colombia	1.75	10,707	Ireland	0.00	189,036
Panama	1.75	9,866	Australia	0.00	153,122
Spain	1.75	9,855	New Zealand	0.00	10,960
Peru	1.75	8,425	Trinidad and Tobago	0.00	2,045
Austria	1.75	7,990	Jamaica	0.00	147
Argentina	1.75	4,398			

This table presents a measure of linguistic distance by country and the country's average holdings of U.S. stocks (in millions of real 2010 U.S. dollars) from U.S. Treasury International Capital Table 1D from December 2012 to December 2016. Linguistic distance captures how different one language is from another. We use a measure of linguistic distance between English and 43 other languages created by Chiswick and Miller (2005). Higher scores denote greater linguistic distance. We map from languages to countries by weighting the linguistic distance of the languages spoken within a country by the percentage of the population speaking each language from the Central Intelligence Agency Factbook. Countries are presented in descending order of linguistic distance.

Table 2

Descriptive Statistics

Variable	N	Mean	Standard Deviation	Q1	Median	Q3
CS	3,582	63,224.07	134,075.34	305.07	4,138.70	58,308.56
Δ*CS*	3,582	1.87	14.44	-0.01	0.06	0.76
LD	3,582	1.65	0.70	1.25	1.75	2.00
BOG	3,582	0.16	1.01	-0.82	-0.15	0.73
Δ*BOG*	3,582	0.67	0.51	0.12	0.78	0.99
*LD*BOG*	3,582	0.26	1.81	-1.22	-0.11	1.31
*LD*ΔBOG*	3,582	1.10	1.03	0.17	1.08	1.82
EQUITY	3,582	12.44	9.68	4.86	14.17	19.80
TREAS	3,582	5.19	12.76	-5.33	4.94	14.42
OIL	3,582	79.84	29.99	47.76	97.09	108.90
Δ*OIL*	3,582	-17.28	20.39	-39.03	-8.69	-1.07
POP	3,582	72.58	217.30	5.60	16.14	52.23
Δ*POP*	3,582	1.00	217.30	0.35	0.91	1.52
G20	3,582	0.25	0.43	0.00	0.00	1.00
BILAT	3,582	47,274.18	108,061.18	3,601.39	13,695.13	39,358.02
Δ*BILAT*	3,582	1.72	16.90	-5.29	1.19	7.54
GDPPC	3,582	23,149.09	22,281.34	5,670.71	13,681.00	39,226.35
Δ*GDPPC*	3,582	1.69	2.90	0.39	1.50	3.28
INF	3,582	3.10	5.47	0.85	1.97	4.36
TAX	3,582	40.87	14.90	30.90	39.70	49.00
OFFSHORE	3,582	0.06	0.23	0.00	0.00	0.00
FINSTD	3,582	5.07	0.78	4.50	5.00	5.70
COMMON	3,582	0.23	0.42	0.00	0.00	0.00
PROTECT	3,582	6.07	1.16	5.50	6.00	7.00
DIST	3,582	8,018.72	3,685.13	5,850.00	6,924.00	10,873.00
CULTURE	3,582	298.44	56.09	257.00	307.00	335.00

This table presents descriptive statistics for the regression variables. The sample period begins in December 2012 and ends in December 2016. See Sections III and IV for variable definitions.

Table 3

Univariate Correlations

		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)
(1)	*CS*		**-0.28**	**0.04**	0.01	-0.03	0.03	**-0.04**	**0.08**	**0.53**	**0.60**	**-0.19**	**0.33**	**-0.21**	**0.54**	**0.47**	**0.24**	**0.18**	**-0.13**	**0.25**
(2)	*LD*	**-0.22**		-0.01	**0.06**	0.01	-0.01	0.02	**0.09**	-0.01	**-0.25**	0.01	-0.03	**0.08**	**-0.17**	**-0.29**	**-0.49**	**-0.26**	**0.16**	**0.09**
(3)	*BOG*	**0.06**	-0.02		**0.92**	**-0.72**	**0.36**	**-0.88**	0.01	0.01	0.03	**-0.11**	0.02	0.01	0.01	0.01	0.01	0.01	-0.02	0.01	0.00
(4)	*LD*BOG*	**0.08**	**-0.10**	**0.91**		**-0.66**	**0.34**	**-0.81**	0.02	0.02	0.00	**-0.12**	0.02	0.01	0.00	-0.01	**-0.03**	**-0.04**	0.02	0.00
(5)	*EQUITY*	**-0.05**	0.02	**-0.75**	**-0.75**		**-0.34**	**0.85**	-0.01	-0.01	-0.02	**0.06**	-0.02	-0.01	-0.01	-0.01	-0.01	0.02	-0.01	0.00
(6)	*TREAS*	0.03	-0.01	**0.41**	**0.40**	**-0.41**		**-0.40**	0.01	0.01	0.01	-0.01	0.01	0.00	0.01	0.01	0.00	-0.01	0.00	0.00
(7)	*OIL*	**-0.05**	0.02	**-0.90**	**-0.83**	**0.79**	**-0.32**		-0.01	-0.02	-0.03	**0.09**	-0.03	-0.01	-0.01	-0.01	-0.01	0.02	-0.01	0.01
(8)	*POP*	0.02	**0.13**	**0.04**	0.03	**-0.04**	0.02	**-0.04**		**0.45**	**-0.20**	**0.05**	**0.41**	**0.31**	**-0.06**	**-0.17**	**0.07**	-0.03	**0.23**	**0.06**
(9)	*BILAT*	**0.68**	**-0.09**	0.03	**0.05**	**-0.04**	0.02	**-0.03**	**0.55**		**0.08**	**-0.08**	**0.52**	**0.10**	0.01	**0.13**	**0.07**	**0.08**	**-0.11**	**0.26**
(10)	*GDPPC*	**0.78**	**-0.25**	0.03	**0.05**	-0.02	0.01	-0.02	**-0.36**	**0.29**		**-0.30**	0.02	**-0.25**	**0.47**	**0.70**	**0.10**	**0.31**	**-0.10**	**0.25**
(11)	*INF*	**-0.42**	0.02	**-0.17**	**-0.16**	**0.10**	**-0.05**	**0.15**	**0.19**	**-0.17**	**-0.49**		**0.12**	**0.44**	**-0.09**	**-0.30**	**-0.07**	**-0.12**	**0.04**	**-0.10**
(12)	*G20*	**0.34**	**-0.04**	0.02	0.03	-0.02	0.01	-0.02	**0.63**	**0.62**	**0.09**	**0.09**		**0.30**	0.00	-0.02	-0.01	**0.17**	**0.15**	**0.44**
(13)	*TAX*	**-0.17**	**0.04**	0.01	0.00	-0.01	0.00	-0.01	**0.42**	**0.10**	**-0.22**	**0.19**	**0.23**		**-0.10**	**-0.27**	**-0.27**	**-0.16**	**-0.09**	**0.26**
(14)	*OFFSHORE*	**0.33**	**-0.22**	0.01	0.03	-0.01	0.00	-0.01	**-0.08**	**0.15**	**0.32**	**-0.16**	0.00	**-0.13**		**0.31**	0.01	**-0.06**	**-0.14**	**0.31**
(15)	*FINSTD*	**0.65**	**-0.31**	0.01	**0.04**	-0.01	0.00	-0.01	**-0.27**	**0.35**	**0.65**	**-0.28**	-0.03	**-0.24**	**0.30**		**0.22**	**0.32**	-0.03	**0.09**
(16)	*COMMON*	**0.15**	**-0.35**	0.01	**0.05**	-0.01	0.00	-0.01	**-0.16**	**0.07**	**0.09**	0.00	-0.01	**-0.29**	0.01	**0.21**		**0.35**	**0.25**	**-0.20**
(17)	*PROTECT*	**0.38**	**-0.23**	-0.02	0.00	0.02	-0.01	0.02	0.01	**0.29**	**0.44**	**-0.20**	**0.18**	**-0.19**	**-0.08**	**0.32**	**0.33**		**0.20**	**0.16**
(18)	*DIST*	**-0.05**	**0.28**	0.00	-0.02	0.00	0.00	0.00	**0.30**	**0.04**	**-0.13**	**0.10**	**0.16**	**-0.08**	**-0.17**	**-0.11**	**0.21**	**0.17**		-0.01
(19)	*CULTURE*	**0.30**	0.02	0.00	0.00	0.00	0.00	**0.00**	**0.27**	**0.30**	**0.36**	**-0.23**	**0.47**	**0.31**	**0.31**	**0.08**	**-0.20**	**0.12**	-0.02	

This table reports univariate correlations for the levels variables used in the regressions (changes variables are omitted for brevity). Pearson (Spearman) correlations are presented above (below) the diagonal. The sample period begins in December 2012 and ends in December 2016. Correlations significant at the 5% level are in bold. See Sections III and IV for variable definitions.

Table 4

Linguistic Distance, Financial Reporting Readability, and Foreign Holdings of U.S. Stocks

	Predicted Sign	Reduced Model	Full Model
Intercept	?	-3,655.55	-3,177.86
		(-0.57)	(-0.50)
LD_i	?	-637.99 *	-1,529.90 ***
		(-1.70)	(-3.69)
BOG_{j-1}	?	415.19	-8,207.01 ***
		(0.33)	(-4.83)
$LD_i * BOG_{j-1}$	?		5,237.35 ***
			(7.35)
$CS_{i,j-12}$	+	1.00 ***	1.00 ***
		(71.21)	(76.84)
$EQUITY_{j-1}$	+	403.39 ***	404.58 ***
		(6.26)	(6.41)
$TREAS_{j-1}$	?	145.02 ***	144.85 ***
		(4.70)	(4.78)
OIL_{j-1}	+	105.16 **	107.25 **
		(2.11)	(2.20)
$POP_{i,t-1}$	?	-9.33 ***	-9.32 ***
		(-2.87)	(-2.75)
$BILAT_{i,t-1}$	+	0.03 **	0.02 **
		(2.48)	(2.30)
$GDPPC_{i,t-1}$	?	0.21 ***	0.21 ***
		(3.96)	(4.13)
$\Delta GDPPC_{i,t-1}$	?	73.05	206.64 **
		(0.81)	(2.30)
$INF_{i,t-1}$	+	271.94 ***	313.77 ***
		(5.22)	(5.93)
$G20_i$	+	5,954.08 ***	5,558.98 ***
		(5.11)	(4.70)
TAX_i	-	-142.03 ***	-142.94 ***
		(-5.71)	(-5.93)
$OFFSHORE_i$	+	19,500.69 ***	18,182.00 ***
		(5.79)	(5.77)
$FINSTD_i$	-	-784.50 *	-904.97 **
		(-1.96)	(-2.26)
$COMMON_i$	+	5,064.21 ***	5,055.93 ***
		(7.49)	(7.16)
$PROTECT_i$	-	-45.51	-30.77
		(-0.11)	(-0.08)
$DIST_i$	-	-0.35 ***	-0.37 ***
		(-3.72)	(-3.81)
$CULTURE_i$	?	-15.66 ***	-12.09 ***
		(-3.34)	(-2.72)
N		3,582	3,582
Adjusted R^2		0.980	0.980
Year Fixed Effects		Yes	Yes

This table reports the results from cross-sectional regressions of $CS_{i,j}$ (total holdings of U.S. stocks by investors in country i at the end of month j in millions of real 2010 U.S. dollars) on a measure of country i's linguistic distance, an aggregate measure of U.S. financial reporting readability at month j-1, and a vector of controls. See Sections III and IV for variable definitions. The sample period begins in December 2012 and ends in December 2016. t-statistics in parentheses are based on standard errors clustered by country. ***, **, and * represent two-tailed significance at the 1%, 5%, and 10% levels, respectively.

Table 5

Robustness Checks

	Panel A	Panel B	Panel C	Panel D
	Alternate Linguistic Distance Measure	Strictly Positive Linguistic Distance Countries	Annual Data	Additional Control Variables
Intercept	1,390.77	-550.52		47,063.02 ***
	(0.21)	(-0.10)		(3.35)
LD_i	-67.60 ***	-1,432.61 **	-2,967.05 *	-4,892.62 ***
	(-4.72)	(-2.01)	(-2.20)	(-4.80)
BOG_{i-1}	-13,003.28 ***	-6,769.42 ***	-6,293.46 *	-18,887.20 ***
	(-6.39)	(-3.62)	(-2.05)	(-6.84)
$LD_i * BOG_{i-1}$	168.79 ***	4,865.88 ***	4,948.28 **	12,593.01 ***
	(8.88)	(4.58)	(2.41)	(11.12)
$CS_{i,j-12}$	1.00 ***	0.99 ***	1.01 ***	0.98 ***
	(80.11)	(71.55)	(14.34)	(60.44)
$EQUITY_{j-1}$	428.42 ***	318.11 ***	1,133.87 **	560.36 ***
	(6.51)	(5.31)	(2.94)	(5.81)
$TREAS_{j-1}$	153.16 ***	123.36 ***	263.10 ***	208.54 ***
	(4.85)	(4.48)	(4.89)	(4.34)
OIL_{j-1}	114.29 **	98.17 **	-28.16	124.27
	(2.25)	(2.30)	(-0.19)	(1.55)
$POP_{i,t-1}$	-10.47 ***	-8.77 **	-8.25	-6.42 ***
	(-3.15)	(-2.47)	(-1.42)	(-2.89)
$BILAT_{i,t-1}$	0.02 **	0.03 ***	0.01	0.07 ***
	(2.19)	(3.98)	(0.19)	(3.07)
$GDPPC_{i,t-1}$	0.21 ***	0.29 ***	0.20	0.25 ***
	(3.98)	(6.19)	(1.60)	(3.21)
$\Delta GDPPC_{i,t-1}$	326.18 ***	284.51 ***	59.87	736.73 ***
	(3.25)	(3.26)	(0.33)	(3.18)
$INF_{i,t-1}$	263.60 ***	397.09 ***	245.82 **	248.21 **
	(5.10)	(9.85)	(2.84)	(2.26)
$G20_i$	5,506.72 ***	3,192.44 ***	6,636.10 ***	5,995.26 ***
	(4.35)	(2.78)	(6.33)	(3.44)
TAX_i	-120.05 ***	-145.97 ***	-124.73	-286.48 ***
	(-5.30)	(-6.79)	(-1.63)	(-5.95)
$OFFSHORE_i$	17,347.85 ***	9,171.22 ***	16,873.42	22,389.46 ***
	(5.63)	(4.48)	(1.44)	(7.03)
$FINSTD_i$	-1,329.04 ***	-852.17 **	-458.60	-7,286.43 ***
	(-2.76)	(-2.28)	(-0.47)	(-5.46)
$COMMON_i$	4,991.85 ***	4,530.35 ***	5,157.01 **	8,106.66 ***
	(5.88)	(4.48)	(4.08)	(4.00)
$PROTECT_i$	-353.61	-1,003.41 ***	-174.58	-288.01
	(-0.83)	(-2.85)	(-0.25)	(-0.39)
$DIST_i$	-0.33 ***	-0.22 *	-0.30 **	-0.46 *
	(-2.98)	(-1.81)	(-2.35)	(-1.87)
$CULTURE_i$	-11.48 **	-4.07	-8.83	-40.54 ***
	(-2.48)	(-1.16)	(-0.57)	(-3.69)
$WEAKFTR_i$				1,619.70
				(1.24)
$HOMEBIAS_i$				-6,289.43
				(-1.59)
$EDUC_i$				-1,869.22 ***
				(-2.68)
$ERVAR_{i,t-1}$				0.00
				(-0.16)
$CAPCONTROL_i$				1,015.58
				(1.09)
$IMMIGRANT_{i,t-1}$				-296,247.24 ***
				(-3.09)
$FINLIT_i$				111.95 **
				(2.16)
$HOMERET_{i,j-1}$				179.23 ***
				(2.63)
$RELIGIOUS_i$				-262.99 ***
				(-5.04)
N	3,382	3,282	286	1,700
Adjusted R^2	0.981	0.980	0.984	0.985
Year Fixed Effects	Yes	Yes	No	Yes

This table reports the results from cross-sectional regressions of $CS_{i,j}$ (total holdings of U.S. stocks by investors in country i at the end of month j in millions of real 2010 U.S. dollars) on a measure of country i's linguistic distance, an aggregate measure of U.S. financial reporting readability at month j-1, and a vector of controls. See Sections III and IV for variable definitions. The sample period begins in December 2012 and ends in December 2016. t-statistics in parentheses are based on standard errors clustered by country. ***, **, and * represent two-tailed significance at the 1%, 5%, and 10% levels, respectively.

Table 6

Changes Analyses

	Dependent Variable		Dependent Variable	
	$(CS_{i,j} - CS_{i,j-12})$		$(CS_{i,j} - CS_{i,j-12})/GDP_{i,t-1}$	
Intercept	1,460.79		3.40	**
	(0.37)		(2.27)	
LD_i	-1,543.78	***	-0.63	**
	(-3.69)		(-2.34)	
BOG_{j-1}	-8,771.12	***		
	(-6.29)			
ΔBOG_{j-1}			-2.11	***
			(-2.82)	
$LD_i * BOG_{j-1}$	5,208.95	***		
	(7.03)			
$LD_i * \Delta BOG_{j-1}$			0.97	***
			(3.06)	
$EQUITY_{j-1}$	440.39	***	0.14	***
	(7.30)		(4.10)	
$TREAS_{j-1}$	128.97	***	0.02	
	(4.64)		(0.79)	
OIL_{j-1}	82.77	***		
	(2.80)			
ΔOIL_{j-1}			0.02	
			(1.41)	
$POP_{i,t-1}$	-9.34	***		
	(-2.72)			
$\Delta POP_{i,t-1}$			0.78	***
			(5.44)	
$BILAT_{i,t-1}$	0.03	**		
	(2.45)			
$\Delta BILAT_{i,t-1}$			0.02	**
			(1.97)	
$GDPPC_{i,t-1}$	0.21	***		
	(6.41)			
$\Delta GDPPC_{i,t-1}$	183.63	**	-0.12	
	(2.27)		(-1.42)	
$INF_{i,t-1}$	308.67	***	0.17	***
	(5.93)		(5.82)	
$G20_i$	5,668.62	***	-0.61	*
	(4.07)		(-1.92)	
TAX_i	-144.23	***	-0.12	***
	(-6.00)		(-6.19)	
$OFFSHORE_i$	18,614.59	***	18.66	***
	(4.62)		(5.18)	
$FINSTD_i$	-898.06	**	0.96	***
	(-2.26)		(5.60)	
$COMMON_i$	5,128.99	***	-0.97	**
	(6.46)		(-2.23)	
$PROTECT_i$	-34.12		-0.64	***
	(-0.09)		(-3.25)	
$DIST_i$	-0.37	***	0.00	***
	(-3.44)		(-2.97)	
$CULTURE_i$	-12.73	**	0.00	
	(-2.37)		(1.52)	
N	3,582		3,582	
Adjusted R^2	0.282		0.151	
Year Fixed Effects	No		No	

This table reports the results from cross-sectional regressions of changes in total holdings of U.S. stocks by investors in country *i* in millions of real 2010 U.S. dollars from month *j* -12 to month *j* on a measure of country *i*'s linguistic distance, the level or percentage change in an aggregate measure of U.S. financial reporting readability from month *j* -12 to month *j* , and a vector of controls. See Sections III and IV for variable definitions. The sample period begins in December 2012 and ends in December 2016. *t* -statistics in parentheses are based on standard errors clustered by country. ***, **, and * represent two-tailed significance at the 1%, 5%, and 10% levels, respectively.

Table 7

Linguistic Distance, Financial Reporting Readability, and Foreign Holdings of U.S.

Intercept	-6,549.40	
	(-0.86)	
LD_i	1,322.25	***
	(2.64)	
BOG_{j-1}	1,619.06	*
	(1.74)	
$LD_i * BOG_{j-1}$	-1,992.48	***
	(-3.87)	
$US_{i,j-12}$	0.98	***
	(157.31)	
$EQUITY_{j-1}$	-44.90	
	(-0.68)	
$TREAS_{j-1}$	114.97	***
	(3.47)	
OIL_{j-1}	73.92	
	(1.40)	
$POP_{i,t-1}$	9.73	***
	(2.90)	
$BILAT_{i,t-1}$	0.00	
	(-0.42)	
$GDPPC_{i,t-1}$	0.15	***
	(4.72)	
$\Delta GDPPC_{i,t-1}$	582.81	***
	(6.44)	
$INF_{i,t-1}$	93.58	*
	(1.89)	
$G20_i$	1,838.19	
	(1.23)	
TAX_i	-82.09	**
	(-2.23)	
$OFFSHORE_i$	3,667.50	
	(0.70)	
$FINSTD_i$	-862.91	**
	(-2.04)	
$COMMON_i$	3,876.65	***
	(5.12)	
$PROTECT_i$	103.44	
	(0.41)	
$DIST_i$	-0.28	***
	(-3.90)	
$CULTURE_i$	2.50	
	(0.40)	
N	3,582	
Adjusted R^2	0.987	
Year Fixed Effects	Yes	

This table reports the results from a cross-sectional regression of $US_{i,j}$ (total holdings of U.S. Treasuries by investors in country i at the end of month j in millions of real 2010 U.S. dollars) on a measure of country i's linguistic distance, an aggregate measure of U.S. financial reporting readability at month j-1, and a vector of controls. See Sections III and IV for variable definitions. The sample period begins in December 2012 and ends in December 2016. t-statistics in parentheses are based on standard errors clustered by country. ***, **, and * represent two-tailed significance at the 1%, 5%, and 10% levels, respectively.