20-00088
12/15/2020

LANGUAGE IN ECONOMICS AND ACCOUNTING RESEARCH: THE ROLE OF LINGUISTIC HISTORY

Giorgio Gotti (*)
University of Texas at El Paso
Visiting Research Fellow, Ca' Foscari University of Venice
e-mail: ggott@utep.edu

Seán G. Roberts
Cardiff University
University of Bristol
e-mail: robertss55@cardiff.ac.uk

Marco Fasan
Ca' Foscari University of Venice
e-mail: marco.fasan@unive.it

Cole Robertson
University of Oxford
cole.robertson@psy.ox.ac.uk

Declaration of conflict of interest
All authors declare no conflicts of interest.

Funding
SR was supported by a Leverhulme early career fellowship (ECF-2016–435). Otherwise, this research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

(*) corresponding author. Full postal address: Department of Accounting and Information Systems, Business Room 260, 500 W University, El Paso, TX 79902

Language in Economics and Accounting Research: The Role of Linguistic History

Abstract

This paper investigates whether a consideration of linguistic history is important when studying the relationship between economic and linguistic behaviours. Several recent economic studies have suggested that differences between languages can affect the way people think and behave (*linguistic relativity* or *Sapir-Whorf hypothesis*). For example, the way a language obliges one to talk about the future might influence intertemporal decisions, such as a company's earnings management. However, languages have historical relations that lead to shared features—they do not constitute independent observations. This can inflate correlations between variables if not dealt with appropriately (Galton's problem). We discuss this problem and provide an overview of the latest methods to control for linguistic history. We then provide an empirical demonstration of how Galton's problem can bias results in an investigation of whether a company's earnings management behavior is predicted by structural features of its employees' language. We find a strong relationship when not controlling for linguistic history, but the relationship disappears when controls are applied. In contrast, economic predictors of earnings management remain robust. Overall, our results suggest that careful consideration of linguistic history is important for distinguishing true causes from spurious correlations in economic behaviors.

Keywords: Institutions; languages; earnings management; linguistic history

JEL classification: D83, M41, Z10

1. Introduction

The linguistic relativity hypothesis (also known as the Sapir-Whorf hypothesis) suggests that properties of a language can affect speakers' perceptions, thoughts, and behaviours, so that differences among these elements may be explained by differences between languages (Whorf et al., 1956). A growing body of research has found that such cross-linguistic differences may indeed affect speakers' cognition (Evans & Levinson, 2009). For instance, researchers have found that differences between languages affect perception in domains as diverse as olfactory perceptual categories (Majid & Burenhult, 2014), spatial cognition (Majid et al., 2004), internal temporal representations (Lai & Boroditsky, 2013), and color perception (Berlin & Kay, 1991; Roberson et al., 2008). The hypothesis has recently been taken up by economists, who have used notions of linguistic relativity to explain a variety of economic behaviours (Chen, 2013; Chen et al., 2017; Gay et al., 2018; Hübner & Vannoorenberghe, 2015; Jakiela & Ozier, 2018; Liang et al., 2014; Shoham & Lee, 2018), including accountancy practice (Chen et al., 2019; Kim et al., 2017; Hooghiemstra et al., 2019). Much of this research is motivated by the theoretical framework of the so-called "linguistic savings hypothesis," which predicts that speakers of languages that oblige the grammatical marking of future time reference (FTR)[1] will tend to make different intertemporal decisions than speakers of languages that do not (Chen, 2013).

Intertemporal decisions are decisions that involve the balancing of present versus future costs and rewards. One prototypical example is the decision to *invest* or *save*, as individuals deciding to save must balance the present loss of resources against future rewards entailed by the accrual of interest, appreciation of investments, and others (Green & Myerson, 2004). Most people tend to devalue future outcomes to some degree. For instance, given the choice of receiving $10 now or $10 in five years, most would choose the former because they would have devalued (or *temporally discounted*) the latter as a function of the length of time until its receipt

[1] By FTR, we mean any statement that involves talking about future events, while we use *future tense* to refer to the linguistic structures that some languages oblige speakers to use when they make FTR, e.g., in English *will*, *shall*, and *be going to*.

(Green & Myerson, 2004). However, the extent to which people temporally discount is subject to extensive individual differences. Ask one person whether they would prefer to receive \$10 now or \$20 in a month, and they will take \$10 now, while someone else might choose to make the future-oriented decision to wait for \$20 (Green & Myerson, 2004). Future orientation is therefore usually measured by giving participants a series of binary choices like those above; participants who temporally discount to a lesser extent tend to make more future-oriented decisions to wait for future rewards, that is, because they perceive such rewards as more valuable than participants who discount more.

In deciding whether to manage earnings, managers must make such intertemporal decisions. As accruals reverse over time, companies "borrow" earnings from past or future periods to improve earnings reported in the current period (see Healy & Whalen, 1999). Consistent with this idea, Brochet et al. (2014) suggest that earnings management is a short-term-oriented practice, as firms are often found to manipulate earnings in order to meet short-term capital market benchmarks.

Economists have steadily contributed to a body of research finding that cross-linguistic differences predict a range of behaviors implicating future orientation. The cross-linguistic difference hypothesized to impact future orientation is the extent to which languages oblige speakers to grammatically mark FTR. For instance:

(1) **English**: Tomorrow, it *will* rain.

(2) **German**: Morgen *regnet* es. (Tomorrow, it rains.)

In English, a speaker is obliged to use *will* or another modal verb (*could, may, might, should*) when they make predictions about the future, whereas German speakers are free to use the present tense. In this framework, languages like English are called “strong-FTR” languages, while those like German are referred to as “weak-FTR” languages (Chen, 2013).

The linguistic savings hypothesis predicts that weak-FTR speakers should be more likely to make future-oriented decisions than strong-FTR speakers. The hypothesis suggests that this may be the case for either of two reasons. First, habitual use of the present tense for FTR in weak-FTR languages is suggested to “collapse” future time with present time, causing weak-FTR speakers to perceive future events as more temporally proximal than strong-FTR speakers. Given the general tendency to discount rewards as a function of temporal distance, weak-FTR speakers are suggested to perceive future rewards as more valuable and, therefore, make more future-oriented decisions (Chen, 2013). Second, failure to grammatically disambiguate future from present time in weak-FTR languages might cause speakers to represent the future temporal locations less precisely, which would also lead to higher future orientation (for more explanation, see Chen, 2013; Robertson & Roberts, 2020).

A number of studies have found that the FTR strength of a language predicts a variety of behaviours for which future orientation appears relevant. For instance, Hübner and Vannoorenberghe (2015) found that FTR predicts inflation rates (which may be sensitive to future orientation) at the national level in a worldwide sample. Liang et al. (2014) found that national measures of sustainability and corporate responsibility, as well as institutional measures of corporate social responsibility, were negatively related to obligatory FTR marking. Using data from the Swiss Household Panel, Guin (2016) found that French-speaking (strong-FTR) Swiss households saved less and overspent more that their German-speaking (weak-FTR) counterparts. In a broader analysis of the effects of long-term orientation on educational outcomes, Figlio et al. (2016) found the home use of weak-FTR languages among first-generation immigrants in Florida predicted positive educational performance. Additionally, Chen et al. (2017) found that companies in weak-FTR countries tend to keep more precautionary cash reserves, indicating that they are making more long-term-oriented decisions. Finally, there is emerging evidence that FTR grammaticization affects firms' earnings management practices. Particularly, Kim et al. (2017) found that companies from countries with a strong-FTR main language engage in more short-term-oriented accounting practices such as accrual-based earnings management. Hooghiemstra et al. (2019) analyzed the effect of board characteristics (including the presence of directors from weak or strong FTR countries) on earnings management. They found that the presence of strong FTR directors is positively related to earnings management, consistent with the idea that strong-FTR language speakers suffer from short-termism.

However, most of these studies—including Kim et al. (2017)—do not include statistical controls for shared linguistic history, despite that in at least one replication study, the effect of language FTR was inconclusive when such controls were added (Roberts, Winters & Chen, 2015). In fact, controls for history are an important consideration not only for studies relating earnings management and FTR, but to any cross-cultural study of accounting behavior (Naroll, 1965; Roberts & Winters, 2013). The objective of this study is therefore to demonstrate that a greater consideration of linguistic history is needed in these cross-cultural statistical studies. Our contribution is an overview of the problem and an empirical demonstration that controlling for linguistic history can make a difference to the inferences drawn from cross-cultural studies of language and economic behavior.

1.1. Linguistic History

Linguistic history refers to the historical relationships between languages, such as all Indo-European languages deriving from a single common ancestor. Such historical processes of inheritance can result in non-independence between languages from the same language family, as related languages inherit similar linguistic features from shared historical antecedents.

In general, when data in a sample are not independent, this can inflate correlations between their traits. In biology, this is known as Galton's problem, named after Francis Galton, who noticed that correlations between morphological traits of closely related species might be misleading. A parallel problem in economics might be counting each trial in an experiment as an independent data point when many trials came from the same participant, or counting each subsidiary branch of a larger company as an independent data point when it is known that the larger company sets policies and prices across all its stores. Failing to account for non-independencies can lead to effects such as Simpson's paradox.

[Insert Figure 1 here]

Galton's problem also applies to languages because of their historical interrelationships. Glottolog (Hammarström et al., 2018, http://glottolog.org), an online reference repository of languages, lists about 7,000 extant languages. All languages are thought to derive from a single population in the distant past (estimates range from 100,000 to one million years ago). Over time, the population diverged, inheriting the language of their ancestors. Due to isolation or various other sociolinguistic effects, differences gradually arose between languages, leading to the development of new languages. Some languages diverged a long time ago and are strikingly different, and some diverged in the recent past and still retain clear similarities (e.g., Spanish and Portuguese diverged within the last 500 years). Historical linguists reconstruct the history of these divergences, using similarities between words and grammatical structures across languages as clues about shared history, similar to the way the genetic history of biological species is reconstructed by comparing genetic sequences. Glottolog lists about 236 *language families*, collections of languages where there is evidence of historical relationships. There is a broad consensus regarding language families and which languages belong to them, though there are some disputes and also language “isolates”—languages for which there are no currently known links to language families. About half of the world’s languages belong to the five largest families (Atlantic-Congo, Austronesian, Indo-European, Sino-Tibetic, and Afro-Asiatic; see Fig. 1).

The imbalance in language family size means that most samples of languages in empirical studies tend to heavily sample a small number of language families while underrepresenting many others. These large families have expanded rapidly within the last 10,000 years or so, covering vast areas. For example, the Indo-European language family originated from a single language spoken between 5,000 and 10,000 years ago (the exact details are heavily debated; see Bouckaert et al., 2012; Pereltsvaig & Lewis, 2015) and diverged into around 500 languages spread across Europe, the Middle East, and India. Languages change slowly, so Indo-European languages still share many similarities in their vocabulary and grammar due to inheritance. Languages can also “borrow” words and grammatical features from multiple neighboring languages over long periods of time. This leads to *areal patterns*: languages within the same geographical regions tend to be similar. This historic process of inheritance and borrowing leads to non-independencies between languages. Since many economic behaviours can change much more quickly, this may not apply to economic variables, though some studies also show long-term effects of culture on economic behavior (see Alesina & Giuliano, 2015; Spolaore & Wacziarg, 2013, 2014).

[Insert Figure 2 here]

More recently, various methods borrowed from molecular genetics have allowed linguists to identify the dates and geographical locations of historical divergences (see Bowern, 2018, for a review), though the accuracy of the methods are debated (e.g., Donohue et al., 2008; Pereltsvaig & Lewis, 2015). Historical relationships between languages are represented as phylogenetic trees with estimated branch lengths, and these are available for some language families through databases such as D-PLACE (see Table 1). Several analyses represent more complex relationships beyond single binary trees by using samples of many thousands of trees, capturing the distribution of relationships between languages. There is currently no consensus on how language families are connected to each other historically, though there are attempts to reconstruct these relationships (e.g., Jaeger, 2018).

[Insert Table 1 here]

Economic behaviours can be highly reactive to current conditions and change from year to year, reducing the historical dependency between groups. However, certain linguistic features can have strong phylogenetic signals: they are robustly transmitted from generation to generation and can be conserved for long time periods. Dunn et al. (2011) showed that the grammatical word order of basic sentences is highly conserved, with a change only happening once every 10,000 years of independent evolution. Roberts, Winters, et al. (2015) estimated that the binary future time reference variable also shows a strong phylogenetic signal. This means that a single change in an ancestor language a long time ago can cause all of its child languages to have the same features. In other words, grammatical features of language are often not historically independent. Furthermore, the relationship between cultural features can be different for different language families (Dunn et al., 2011).

1.2. Controlling for Galton's Problem in Cross-Cultural Studies

How can these issues be addressed? Answering this question depends on the sample of data that is available. If the languages all belong to a single family, then it may be feasible to use phylogenetic trees to represent the historical relationships between languages. Phylogenetic regression techniques effectively weight the observations by their historical relatedness (Pagel, 1997; Verkerk, 2013). Another approach is to use estimates of similarities between languages, represented as distance matrices (see Hua et al., 2018). Recent advances have also allowed linguists to reconstruct how cultural features change and co-evolve over time (see Blute & Jordan, 2018). However, we suspect that this will be of limited use to economists, since economic variables change at vastly greater rates than languages.

If dated trees are unavailable for the language in question, or if the sample includes languages from multiple language families, then it may be better to use a multilevel modelling approach. Mixed-effects modelling allows the fitting of random effects in addition to main effects.[2] Random effects have been used in linguistics and psychology to capture non-independencies between observations (Baayen et al., 2008; Clark, 1973), including controlling for linguistic history by entering language family as a random effect (e.g., Roberts, Winters, et al., 2015). Areal effects can be controlled for in the same way by including the geographical area as a random effect. G rouping languages under continents has been used as a rough estimate for areal effects, but more relevant measures include the geographic areas from the Autotyp database (Bickel et al., 2017), which are defined to represent areas of known contact between languages. When applying controls for language family and area to Chen's (2013) original data, the effect of FTR on savings behavior disappears (Roberts, Winters, et al., 2015). Similarly, Chen et al. (2017) study the relationship between FTR and corporate savings behavior, including robustness tests where language family or continent are included as fixed effects. However, we note that

[2] We note that there is some difference in terminology between fields. Here we use "random effects" to mean estimates that are calculated at the group level (estimated by shrinkage), while "fixed effects" are calculated for each observation (estimated by least squares, Gelman & Pardoe, 2004).

controlling for Galton's problem should not necessarily be seen as an additional kind of robustness test. The controls are necessary to get an unbiased estimate of the strength of the correlation in the first place. We also note that Chen et al. (2017) did not control for both historical and areal effects at the same time, a test which makes conceptual sense.

Studies comparing linguistic variables to other cultural variables may also face a problem of multilevel data. Many economic studies measure economic variables that are not directly related to individual languages. For example, Chen (2013) was based on survey responses from individual participants who declared their primary language, and Kim et al. (2017) used economic data based on companies that were linked to languages through the country in which they traded. In these cases, multilevel modelling allows these relationships to be explicitly coded into the regression. In a mixed-effects modelling framework, linguistic variables can also be given random slopes, reflecting the possible differences in evolution between families. For these reasons, multilevel approaches like mixed-effects regression are perhaps the most flexible option for dealing with Galton's problem.

An alternative approach is used by Jakiela and Ozier (2018), who study the relationship between grammatical gender and female labor force participation. They suggest that using language family as a grouping factor is too coarse. Instead, they identify clades (sub-trees within a language family) with identical linguistic values. They then permute the linguistic values between these clades to create a baseline against which to compare the true correlation. We have reservations about this method, since it is not clear to us what this baseline represents (possibly, worlds where particular changes did or did not happen, but the point at which they might occur is fixed). Instead, it might be feasible to simulate alternative histories directly using the full phylogenetic tree and estimates of the likelihood of change over time.

A simpler approach to evaluating the role of historical language evolution is to run an OLS regression and cluster standard errors by language family. This approach accounts for correlations occurring in observations within the same language family and, although not as good as the mixed-effects regression approach (which models both within and between language family correlations and provides an unbiased estimate of standard errors), can give us estimates of standard errors that are less unbiased than simply ignoring linguistic history.

1.3. Hypotheses

The main question of this paper is whether controlling for shared history is crucial when investigating relationships between language and accounting behaviours. In the next section, we perform an empirical test of the relationship between future-tense marking and accrual-based earnings management (AEM), reflecting a study by Kim et al. (2017). Kim et al. find that companies from countries whose main language has strong FTR engage in more short-term-oriented accounting practices such as accrual-based earnings management. The study uses the FTR variable from Chen (2013) but does not control for linguistic history. In the analysis below, we extend Kim et al.'s analysis to a new dataset from a wider range of countries. We test three hypotheses:

H1: When not controlling for linguistic history, countries whose main language has strong FTR will have stronger indices of AEM practices than countries whose main language has weak FTR.

This is simply a replication of Kim et al. (2017).

H2: When controlling for linguistic history, the relationship between FTR and AEM will disappear.

However, any given statistical relationship might disappear when including additional controls due to a lack of power rather than a lack of a real relationship. To demonstrate that this is not the case, we also include a third hypothesis. Our prediction in H2 is specifically about linguistic variables. We expect nonlinguistic predictors to survive controls for linguistic history, since they are subject to the same cultural transmission processes and therefore Galton's problem.

H3: When controlling for linguistic history, the relationship between AEM and nonlinguistic measures of economic behavior will not be diminished.

That is, H3 is included to demonstrate that the nonsignificance of FTR is not due to a lack of power: other, nonlinguistic variables will not be affected by controlling for linguistic history.

2. Method

The aim of the empirical investigation is to test whether controlling for linguistic history changes the inference one might make from a study of the relationship between future tense and earnings management.

2.1. Sample and Measures

We test our hypotheses using a global panel data set of companies publicly listed during the period 1993–2013. Financial companies (SIC codes 6000–6999), utility companies (SIC codes 4900–4999), and firms that lack relevant accounting information data to construct the variables are excluded from the sample. Our primary data source for firms' financial performance is Compustat North America and Compustat Global. All continuous financial variables are winsorized at 0.5% and 99.5% of the distributions to reduce the influence of outliers. Data related to analyst forecast are obtained from I/B/E/S. We then identify the official language FTR available from Chen (2013) to match the country-level FTR measures with firm-level financial variables. Other country-level legal and cultural indices are hand collected from the relevant publications and websites. We measure earnings management following Kothari et al. (2005); thus, we proxy for accrual-based earnings management (*AEM*) by performance-matched discretionary accruals. We present the descriptive statistics for the sample in Table 2. We use the following variables as controls (see the Appendix for details):

- Investor protection score (*INVPRO*), based on the anti-director index from Djankov et al. (2008)
- Power distance index (*PD*), based on Hofstede (2001)
- Individualism/collectivism score (*INDIV*), based on Hofstede (2001)
- Masculinity/femininity score (*MAS*), based on Hofstede (2001)
- Uncertainty avoidance score (*UA*), based on Hofstede (2001)
- Long-/short-term orientation score (*LTO*), based on Hofstede (2001)
- Indulgence (*INDUL*), based on Hofstede (2001)
- Country GDP growth rate (*GGROWTH*)
- Company size (*LNSIZE*), measured as the natural logarithm of total assets adjusted for inflation rate
- Book value of common equity divided by common value of equity (*BTM*)
- Leverage (*LEV*), measured as short- and long-term debt divided by total assets
- Return on assets (*ROA*), measured as income before extraordinary items divided by total assets
- Dummy variable (*MEET*) that takes one for firm-year observations with actual annual EPS greater than or equal to consensus analyst earnings forecast, zero otherwise

- Dummy variable (*LOSS*) that takes one for firm-year observations with negative income before extraordinary items, zero otherwise

[Table 2]

These measures mirror the methodology of Kim et al. (2017), though we note that the aim is not to replicate that study in a strict sense, only to demonstrate the methods and importance of controlling for linguistic history.

2.1.1. Linguistic Data

A main language was associated with each country, based on the data obtained by Kim et al. (2017). A main language family was assigned based on the official or de facto languages of the country, according to Glottolog (Hammarström et al., 2018). Four countries (South Africa, Kenya, Nigeria, and Zimbabwe) were excluded due to presence of several main languages from different language families. The final data had 94,707 observations from 50 countries, representing 35 main languages and nine main language families (see Table 3). All continuous variables in the regressions were scaled and centered to have a mean of 0 and a standard deviation of 1.

[Insert Table 3 here]

The phylogenetic tree from Bouckaert et al. (2012) was used to estimate more fine-grained distances between Indo-European languages. The phylogeny includes branch lengths estimated by Bayesian phylogenetic estimation. Patristic distances between languages on the phylogeny (number of years of independent evolution since the last common ancestor) were used as a measure of historical independence. The economic measures were collapsed under the associated main language, resulting in 21 data points that could be linked to the phylogenetic tree.

[Insert Figure 3 here]

2.1.2. Modelling

The main analysis uses mixed-effects modelling using the *lme4* package (see Bates et al., 2015) within the statistical software framework R (R Core Team, 2013). Language family was used as a random effect to control for the historical dependencies between languages. In this approach, estimation of the significance of a dependent variable is typically done by comparing the fit of a model with and without that dependent variable (Baayen et al., 2008).

We fit two sets of models. The first has no controls for linguistic history, closely mirroring the analysis of Kim et al. (2017). The second introduces controls for linguistic history. All models included random intercepts for year and industry and the following independent variables: *INVPRO, PD, INDIV, MAS, UA, LTO, INDUL, GGROWTH, SIZE, BTM, LEV, ROA, MEET* and *LOSS*.

Since *AEM* is an absolute value, and few companies score highly on this measure, the distribution is far from normal (skewness = 7.63). To address this, the models were fit using a gamma distribution and a log-transformed AEM variable (see supporting materials).

In order to check that our results are robust, we also use four alternative methods. The first is to test that the same conclusions are reached when assuming a Gaussian distribution (as in Kim et al., 2017). The second additional method uses a more fine-grained measure of linguistic history. Although it is often the only data available, language family is a coarse measure of linguistic relatedness: it represents the distance between English and Urdu as the same as the distance between English and Dutch. To address this, we performed a phylogenetic regression (Grafen, 1989) on Indo-European data, predicting AEM by FTR alone. This method represents linguistic history as a binary-branching tree with branch lengths representing the amount of time that has passed since languages split (see Figure 3). For Indo-European languages, a dated tree is available, which provides a continuous measure of historical distance between languages. The phylogenetic regression uses the historical distances between languages to produce an expected covariance matrix. Estimates for the parameters of the model were then estimated by MCMC sampling (using the R package *MCMCglmm*, Hadfield, 2010).

The models above assume a simple, linear relationship between the independent and dependent variables. To explore the possibility of more complex relationships, we fit a binary decision tree to the data. A decision tree is a machine-learning technique that aims to find the optimal way to classify data into homogenous groups via a set of yes/no questions (e.g., Strobl et al., 2009). The results of this method indicate how well each variable predicts the outcome. However, rather than linear relationships for the whole data (as in a linear regression), predictions can be different for different groups. Decision trees are used in many applications, including exploring patterns in linguistic data (Majid et al., 2018; Roberts, Torreira et al., 2015; Tagliamonte & Baayen, 2012). The decision tree is predicted AEM by the same variables as in the mixed-effects model above, with random intercepts for language family, year, and industry (using the REEMtree package, Sela & Simonoff, 2011). The relative influence of each variable in predicting the dependent variable can be captured by the *variable importance* measure. If FTR is a good predictor of AEM, we would expect it to appear on the decision tree and have relatively high variable importance.

Finally, we use OLS regression with robust standard errors clustered by language family. The full data and analysis scripts are available online.[3]

[3] https://github.com/seannyD/FTRAccountingStudy

3. Results

In accordance with H1, without controls for linguistic history, there was a strong main effect of FTR ($\beta = 0.53$, std. err. = 0.01, $t = 45.5$, $p < 0.0001$), and the inclusion of FTR significantly improved the fit of the model (log likelihood difference = 110, df = 1, $\chi^2 = 210.38$, $p < 0.0001$). We note that this association is stronger than that of Kim et al. (β for *weak* FTR = -0.02, $t = -4.25$), which may reflect a larger range of countries in the current sample and differences in modelling assumptions. The model fit was significantly improved by adding a random intercept (log likelihood difference = 1305, $\chi^2 = 2609.9$, $p < 0.0001$) and a random slope for FTR by language family (log likelihood difference = 87, $\chi^2 = 173.9$, $p < 0.0001$). That is, the estimated AAC varies between families, and the size of the effect of FTR varies between families. With a random intercept by main language family, the effect of FTR was much weaker ($\beta = 0.17$) and, with a random slope, was not significant ($\beta = -0.17$, std. error = 0.02, $t = -0.8$, $p = 0.42$). That is, we find support for H2 that the relationship between FTR and AEM disappears when controlling for linguistic history. In contrast, and in accordance with H3, most nonlinguistic predictors of AEM remained significant or became stronger when controlling for language family (Table 4). For example, the effect of a company's book value remained roughly the same (from $\beta = -0.045$ to $\beta = -0.036$), and the effect of long-term orientation increased (from $\beta = -0.36$ to $\beta = -0.61$). This shows that it is not just a lack of power that makes FTR nonsignificant.

[Insert Table 4 and Figure 4 here]

The same results and inferences were obtained using the alternative methods. When a normal distribution was used to model the data (as in Kim et al.), there was a strong main effect of FTR without controls for linguistic history (H1, $\beta = 0.15$, log likelihood difference = 110, $\chi^2 =$

210.4 , $p < 0.001$), but the effect disappears when including random intercepts for language family (H2, β = 0.02, log likelihood difference = 0.8, $\chi^2 = 1.66$, $p = 0.2$). In the phylogenetic analysis of Indo-European data, FTR was not a significant predictor of AEM (H2, β = 0.95 [-0.41, 2.32], ESS = 1069, $p = 0.18$). In contrast, LTO was a significant predictor of AEM under the same test (H3, β = -0.44 [-0.90, -0.02], ESS = 8875, $p = 0.048$). The decision tree did not select the FTR variable to predict earnings management, in line with H2. However, other cultural variables (e.g., individualism, LTO, and indulgence) were rated as highly important, in line with H3. When using OLS regression, FTR was a significant predictor (H1, β = 0.13, SE = 0.03, $p < 0.0001$) except when clustering robust standard errors by language family (H2, β = 0.13, SE = 0.08, $p = 0.13$). In contrast, nine other nonlinguistic predictors remained significant under this test (H3, see Table 5).

[Insert Table 5 here]

4. Discussion and Conclusion

This study draws from linguistics in order to contribute to the growing body of accounting and economics literature studying the effect of languages on individual or company behavior. Work from historical linguistics suggests that the properties of languages are not historically independent. Historical linguists, drawing from methodologies used in the field of genetic history, have grouped languages into families, reflecting the inheritance of features from ancestor languages. They have also identified areal patterns caused by features shared among languages within the same geographical regions. This interdependence among languages can inflate correlation between variables (Galton's problem) if not dealt with appropriately.

To demonstrate the issue, we carried out an empirical investigation of earnings management and future tense and thus show empirically that controlling for linguistic history matters. Our base model, which mirrors that of Kim et al. (2017), suggests that grammatical rules on referring to the future (FTR) are significantly associated with earnings management, consistent with the results obtained by Kim et al. (2017). Once we control for linguistic history, though, the association between future tense and earnings management becomes nonsignificant. Controlling for linguistic history increases the standard error but also reduces the magnitude of the estimate for FTR. This is not the case for nonlinguistic variables such as long-term orientation, which remain robust to controls for linguistic history.

The empirical results highlight the need for language-focused accounting and economics studies to control for linguistic history. There are several methodologies for doing this, some of which have been used in the empirical analysis above. Minimally, we recommend mixed-effects modelling, using language family as a random effect to control for the historical dependencies between languages. More fine-grained control can be done using a phylogenetic regression within language families with known historical relations (e.g., Indo-European). However, making inferences using large-scale, cross-cultural data is hard, and we suggest that future studies should take a range of different approaches. This includes treating the complexity of linguistic variables appropriately and taking advantage of variation in language use and economic behavior within cultures.

Tables and Figures

Figure 1

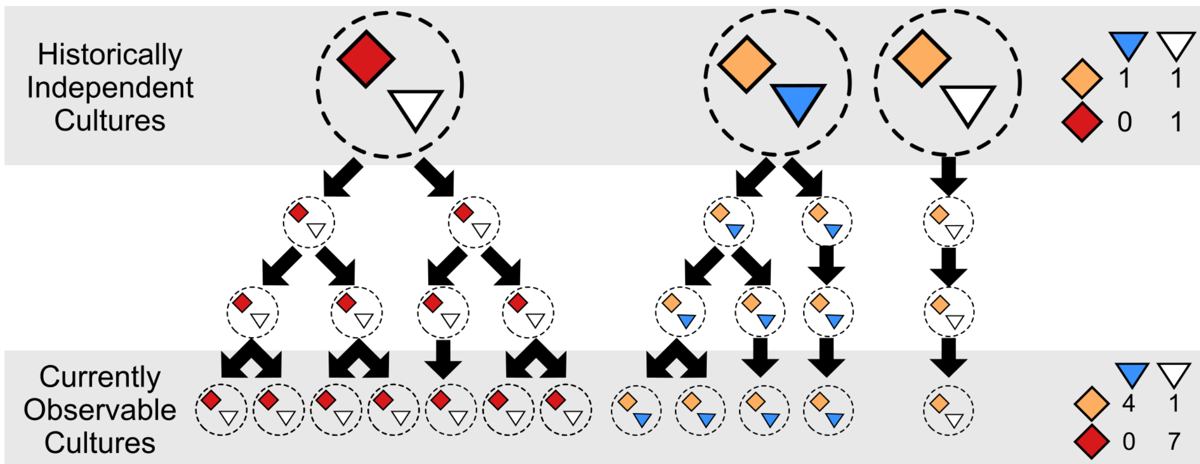

Fig. 1 An illustration of how cultural inheritance can lead to spurious correlations (adapted from Roberts, Winters, et al., 2015). At the top are three independent historical cultures, each of which has a bundle of various traits represented as colored shapes. Each trait is causally independent of the others. On the right is a contingency table for the colors of triangles and diamonds. Originally, there is no particular relationship between the color of triangles and the color of diamonds. However, over time, these cultures split into new cultures. Along the bottom of the graph are the currently observable cultures. We now see that a pattern has emerged in the raw numbers (blue triangles occur with orange diamonds, and white triangles occur with red diamonds). The mechanism that brought about this pattern is simply that the traits are inherited together; there is no causal mechanism whereby blue triangles are more likely to cause orange diamonds.

Figure 2

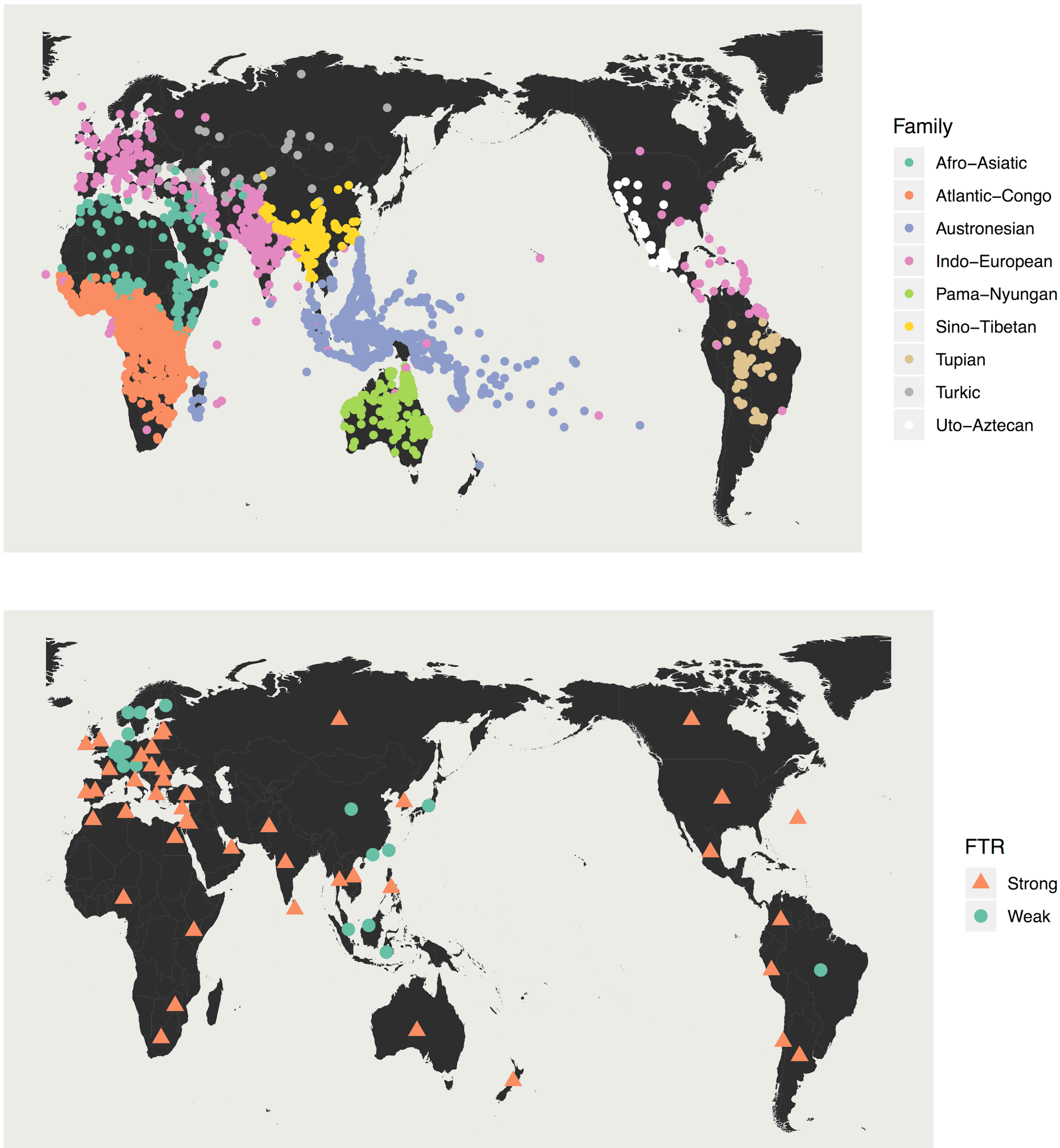

Fig. 2 Top: Distribution of some of the world's largest language families, including relatively recent creoles and pidgins. Data from Glottolog (Hammarström et al., 2018). Bottom: Map of countries and their main languages in the study, labelled by the FTR value of the main language.

Figure 3

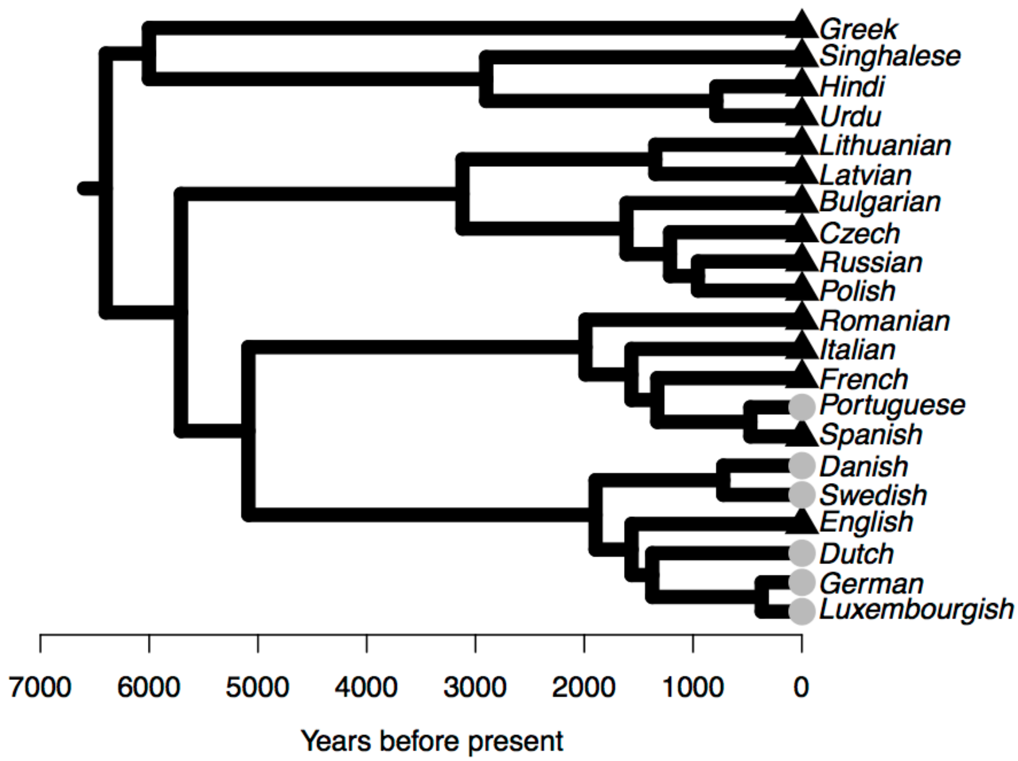

Fig. 3 The phylogenetic tree used in the analysis (Bouckaert et al., 2012). The tips show the FTR value of each language (black triangle = strong, gray circle = weak).

Figure 4

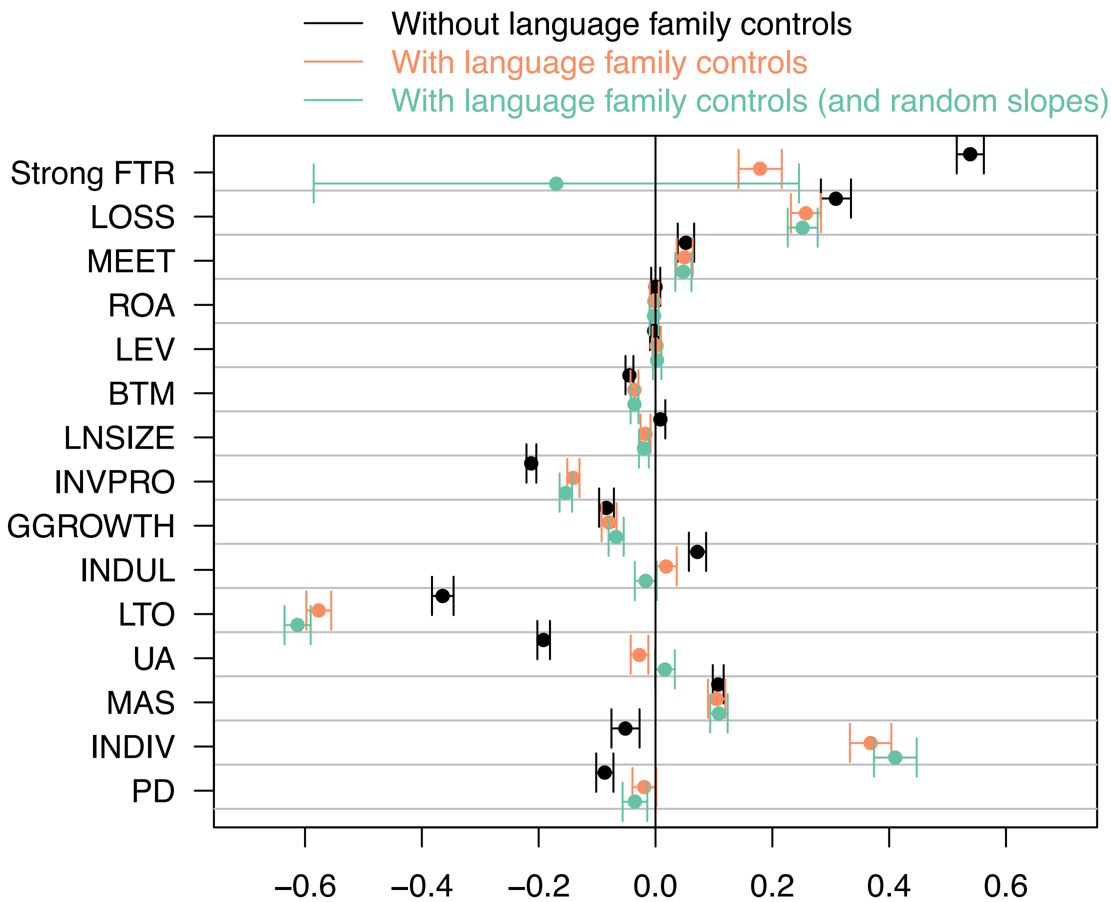

Fig. 4 Model estimates (β values) for different variables in the model, without controls for language family (black), with controls (orange), and with additional random slopes (green). Error bars show 95% confidence intervals for the estimate.

Table 1. List of language families for which dated phylogenies are available*

Family	**Reference**
Austronesian	Gray et al. (2009)
Bantu	Grollemund et al. (2015)
Dene-Yenesian	Sicoli & Holton (2014)
Dravidian	Kolipakam et al. (2018)
Indo-European	Bouckaert et al. (2012); Chang et al. (2015)
Japonic	Lee & Hasegawa (2011)
Pama-Nyungan	Bouckaert et al. (2018); Bowern & Atkinson (2012)
Semitic	Kitchen et al. (2009)
Tukanoan	Chacon & List (2015)
Tupi-Guarani	Michael et al. (2015)
Uralic	Honkola et al. (2013)
Uto-Aztecan	Dunn et al. (2011)

*https://d-place.org/phylogenys

Table 2. Descriptive statistics of the key variables used in this study

Variable	N	Mean	Std Dev	25%	Median	75%
AEM	54300	0.933	2.439	0.0565	0.155	0.540
$StrongFTR$	118492	0.594	0.491	0	1	1
PD	118492	50.50	17.22	39	40	60
$INDIV$	118492	63.35	27.57	46	71	91
MAS	118492	61.75	18.98	52	62	66
UA	118492	55.59	22.58	44	46	75
LTO	118492	53.74	26.63	26	51	87
$INDUL$	118492	53.96	17.13	42	63	68
$INVPRO$	118492	3.569	1.110	3	3.500	4.500
$GGROWTH$	118492	3.039	3.136	1.596	2.667	4.411
$LNSIZE$	118492	5.130	1.969	3.724	4.977	6.408
BTM	118492	0.848	1.210	0.311	0.558	0.964
LEV	118492	0.211	0.189	0.037	0.182	0.330
ROA	118492	0.0103	0.166	0.004	0.038	0.076
$LOSS$	118492	0.226	0.418	0	0	0
$MEET$	118492	0.506	0.500	0	1	1

Table 3. List of countries in the sample with corresponding official languages and the language families to which they belong

Country Code	Country Name	Official Language	FTR	Family
EGY	Egypt	Arabic	Strong	Afro-Asiatic
JOR	Jordan	Arabic	Strong	Afro-Asiatic
MAR	Morocco	Arabic	Strong	Afro-Asiatic
IDN	Indonesia	Indonesian	Weak	Austronesian
MYS	Malaysia	Malaysian	Weak	Austronesian
PHL	Philippines	Tagalog	Strong	Austronesian
AUS	Australia	English	Strong	Indo-European
AUT	Austria	German	Weak	Indo-European
BEL	Belgium	Dutch	Weak	Indo-European
BGR	Bulgaria	Bulgarian	Strong	Indo-European
BRA	Brazil	Portuguese	Weak	Indo-European
CAN	Canada	English	Strong	Indo-European
CHE	Switzerland	Swiss German	Weak	Indo-European
CHL	Chile	Spanish	Strong	Indo-European
COL	Colombia	Spanish	Strong	Indo-European
CZE	Czech Republic	Czech	Strong	Indo-European
DEU	Germany	German	Weak	Indo-European
DNK	Denmark	Danish	Weak	Indo-European
ESP	Spain	Spanish	Strong	Indo-European
FRA	France	French	Strong	Indo-European
GBR	United Kingdom	English	Strong	Indo-European
GRC	Greece	Greek	Strong	Indo-European
IND	India	Hindi	Strong	Indo-European
IRL	Ireland	English	Strong	Indo-European
ITA	Italy	Italian	Strong	Indo-European
LTU	Lithuania	Lithuanian	Strong	Indo-European

Table 3 (Continued)

LUX	Luxembourg	Luxembourgish	Weak	Indo-European
LVA	Latvia	Latvian	Strong	Indo-European
MEX	Mexico	Spanish	Strong	Indo-European
NLD	Netherlands	Dutch	Weak	Indo-European
NOR	Norway	Norwegian	Weak	Indo-European
NZL	New Zealand	English	Strong	Indo-European
PAK	Pakistan	Urdu	Strong	Indo-European
PER	Peru	Spanish	Strong	Indo-European
POL	Poland	Polish	Strong	Indo-European
PRT	Portugal	Portuguese,	Strong	Indo-European
ROU	Romania	Romanian	Strong	Indo-European
RUS	Russia	Russian	Strong	Indo-European
SWE	Sweden	Swedish	Weak	Indo-European
USA	United States of America	English	Strong	Indo-European
JPN	Japan	Japanese	Weak	Japonic
KOR	South Korea	Korean	Strong	Koreanic
CHN	China	Mandarin	Weak	Sino-Tibetan
HKG	Hong Kong	Cantonese	Weak	Sino-Tibetan
SGP	Singapore	Mandarin	Weak	Sino-Tibetan
TWN	Taiwan	Mandarin	Weak	Sino-Tibetan
THA	Thailand	Thai	Strong	Tai-Kadai
TUR	Turkey	Turkish	Strong	Turkic
FIN	Finland	Finnish	Weak	Uralic
HUN	Hungary	Hungarian	Strong	Uralic

Table 4. Estimates of how variables predict earnings management with and without controls for linguistic history

Variable	No controls for language history	With controls for language history	Robust?
PD	-0.087 ($p < 0.0001$)	-0.035 ($p = 0.001$)	Yes
MAS	0.11 ($p < 0.0001$)	0.11 ($p < 0.0001$)	Yes
LTO	-0.36 ($p < 0.0001$)	-0.61 ($p < 0.0001$)	Yes
GGROWTH	-0.084 ($p < 0.0001$)	-0.067 ($p < 0.0001$)	Yes
LNSIZE	0.0084 ($p = 0.049$)	-0.02 ($p < 0.0001$)	Yes
BTM	-0.045 ($p < 0.0001$)	-0.036 ($p < 0.0001$)	Yes
LEV	-0.0021 ($p = 0.57$)	0.0029 ($p = 0.43$)	Yes
ROA	0.00038 ($p = 0.92$)	-0.0024 ($p = 0.54$)	Yes
INDIV	-0.051 ($p < 0.0001$)	0.41 ($p < 0.0001$)	No
UA	-0.19 ($p < 0.0001$)	0.016 ($p = 0.061$)	No
INDUL	0.072 ($p < 0.0001$)	-0.017 ($p = 0.084$)	No

Values are model estimates (beta values) with p-values from model comparison tests in brackets. The final column shows whether the effect is robust to controls for linguistic history.

Table 5. OLS regressions

Variables	(1) **AEM** **no control**	(2) **AEM** **with language family clustered se**
StrongFTR	**0.132***** **(39.618)**	**0.132** **(1.667)**
INVPRO	-0.099*** (-50.273)	-0.099** (-2.587)
PD	0.010*** (7.039)	0.010 (0.232)
INDIV	0.021*** (7.494)	0.021 (0.518)
MAS	0.062*** (41.122)	0.062** (2.669)
UA	-0.044*** (-28.515)	-0.044 (-1.605)
LTO	-0.111*** (-40.732)	-0.111** (-2.827)
INDUL	0.025*** (12.241)	0.025 (1.549)
GGROWTH	-0.072*** (-28.949)	-0.072* (-1.932)
LNSIZE	0.036*** (9.909)	0.036*** (5.299)
BTM	-0.020*** (-9.796)	-0.020* (-2.240)
LEV	-0.005 (-1.384)	-0.005* (-2.083)
ROA	0.009** (2.178)	0.009 (1.629)
MEET	0.030*** (4.809)	0.030*** (4.557)
LOSS	0.170*** (9.9929)	0.170*** (6.241)
Constant	0.225*** (24.902)	0.225 (1.443)
Observations	94,707	94,707
R-squared	0.073	0.073
Adj. R-squared	0.073	0.073

Column 1: OLS regression with robust standard errors.

Column 2: OLS regression with robust and clustered for language family standard errors, to control for language history.

Robust t-statistics in parentheses

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$

References

Alesina, A., & Giuliano, P. (2015). Culture and institutions. *Journal of Economic Literature*, *53*(4), 898–944.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*(4), 390–412.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48.

Berlin, B., & Kay, P. (1991). *Basic color terms: Their universality and evolution.* University of California Press.

Bickel, B., Nichols, J., Zakharko, T., Witzlack-Makarevich, A., Hildebrandt, K., Rießler, M., Bierkandt, L., Zúñiga, F., & Lowe, J. B. (2017). *The AUTOTYP typological databases.* Version 0.1.0. https://github.com/autotyp/autotyp-data/tree/0.1.0 (retrieved Dec 1, 2018)

Blute, M., & Jordan, F. (2018). The evolutionary approach to history. In R. L. Hopcroft (Ed.), *Oxford Handbook of Evolution, Biology, and Society (*ed. Hopcroft, R. J. *621–642)*. Oxford University Press.

Bouckaert, R. R., Bowern, C., & Atkinson, Q. D. (2018). The origin and expansion of Pama-Nyungan languages across Australia. *Nature Ecology & Evolution, 2*, 741–749.

Bouckaert, R. R., Lemey, P., Dunn, M., Greenhill, S. J., Alekseyenko, A. V., Drummond, A. J., Gray, R. D., Suchard, M. A., & Atkinson, Q. D. (2012). Mapping the origins and expansion of the Indo-European language family. *Science 337*(6097), 957–960.

Bowern, C. (2018). Computational phylogenetics. *Annual Review of Linguistics, 4*, 281–296.

Bowern, C., & Atkinson, Q. D. (2012). Computational phylogenetics and the internal structure of

Pama-Nyungan. *Language, 88*(4), 817–845

Brochet, F., Miller, G., Srinivasan S. (2014). Do Analysts follow managers who switch companies? An analysis of relationships in the capital markets. *The Accounting Review*, 89(2), 451-482

Chacon, T. C., & List, J. M. (2015). Improved computational models of sound change shed light on the history of the Tukanoan languages. *Journal of Language Relationship, 13*(3-4), 177–203.

Chang, W., Cathcart, C., Hall, D., & Garrett, A. (2015). Ancestry-constrained phylogenetic analysis supports the Indo-European steppe hypothesis. *Language 91*(1), 194–244.

Chen, J. I., He, T. S., & Riyanto, Y. E. (2019). The effect of language on economic behavior: Examining the causal link between future tense and time preference in the lab. *European Economic Review*, *120*, 103307.

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

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

Clark, H. H. (1973). The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior, 12*(4), 335–359.

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

Djankov, S., La Porta, R., Lopez-de-Silanes, F., & Shleifer, A. (2008). The law and economics of self-dealing. *Journal of Financial Economics, 88*(3), 430–465.

Donohue, M., Wichmann, S., & Albu, M. (2008). Typology, areality, and diffusion. *Oceanic Linguistics, 47*(1), 223–232.

Dunn, M., Greenhill, S. J., Levinson, S. C., & Gray, R. D. (2011). Evolved structure of language shows lineage-specific trends in word-order universals. *Nature, 473*(7345), 79–82.

Evans, N., & Levinson, S. (2009). The myth of language universals: Language diversity and its importance for cognitive science. *Behavioral and Brain Sciences, 32*(5), 429–448

Figlio, D., Giuliano, P., Ozek, U., & Sapienza, P. (2016). *Long-term orientation and educational performance* (NBER Working Paper No. 22541).

Gay, V., Hicks, D. L., Santacreu-Vasut, E., & Shoham, A. (2018). Decomposing culture: An analysis of gender, language and labor supply in the household. *Review of Economics of the Household, 16*(4), 879–909.

Gelman, A., & Pardoe, I. (2004*). Average predictive effects for models with nonlinearity, interactions, and variance components.* Technical report, Department of Statistics, Columbia University.

Grafen, A. (1989). The phylogenetic regression. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences, 326*(1233), 119–157.

Gray, R. D., Drummond, A. J., & Greenhill, S. J. (2009). Language phylogenies reveal expansion pulses and pauses in Pacific settlement. *Science, 323*(5913), 479–483.

Green, L., & Myerson, J. (2004). A Discounting Framework for Choice With Delayed and Probabilistic Rewards. *Psychological Bulletin*, *130*(5), 769–792. https://doi.org/10.1037/0033-2909.130.5.769.

Grollemund, R., Branford, S., Bostoen, K., Meade, A., Venditti, C., & Pagel, M. (2015). Bantu expansion shows that habitat alters the route and pace of human dispersals. *Proceedings of* *the National Academy of Sciences*, *112*(43), 13296-13301.

Guin, B. (2016). *Culture and household saving* (SSRN Paper). http://dx.doi.org/10.2139/ssrn.2698872

Hadfield, J. D. (2010). MCMC methods for multi-response generalized linear mixed models: The MCMCglmm R package. *Journal of Statistical Software, 33*(2), 1–22. http://dx.doi.org/10.18637/jss.v033.i02

Hammarström, H., Forkel, R., & Haspelmath, M. (2018). *Glottolog 4.3*. Max Planck Institute for the Science of Human History, Jena. http://glottolog.org. Accessed [01/09/2020].

Healy, P. M., & Wahlen, J. M. (1999). A review of the earnings management literature and its implications for standard setting. *Accounting horizons, 13*(4), 365-383.

Hofstede, G. (2001). *Culture's consequences: Comparing values, behaviors, institutions, and organizations across nations*. SAGE Publications.

Honkola, T., Vesakoski, O., Korhonen, K., Lehtinen, J., Syrjänen, K., & Wahlberg, N. (2013). Cultural and climatic changes shape the evolutionary history of the Uralic languages. *Journal of Evolutionary Biology, 26*(6), 1244–1253.

Hooghiemstra, R., Hermes, N., Oxelheim, L., & Randøy, T. (2019). Strangers on the board: The impact of board internationalization on earnings management of Nordic firms. *International Business Review, 28*(1), 119–134.

Hua, X., Greenhill, S., Cardillo, M., Schneemann, H., & Bromham, L. (2018). The ecological drivers of variation in global language diversity. *BioRxiv*. https://doi.org/10.1101/426502

Hughes, A., Trudgill, P., & Watt, D. (2013). *English accents and dialects: An introduction to social and regional varieties of English in the British Isles.* Routledge.

Hübner, M., & Vannoorenberghe, G. (2015). Patience and long-run growth. *Economics*

Letters, 137, 163–167.

Jaeger, G. (2018). Global-scale phylogenetic linguistic inference from lexical resources. *Scientific Data, 5*(1), 180–189.

Jakiela, P., & Ozier, O. (2018). *Gendered language* (Policy Research Working Paper 8464). World Bank Group.

Kim, J., Kim, Y., & Zhou, J. (2017). Languages and earnings management. *Journal of Accounting and Economics*, *63*(23), 288–306.

Kitchen, A., Ehret, C., Assefa, S., & Mulligan, C. J. (2009). Bayesian phylogenetic analysis of Semitic languages identifies an Early Bronze Age origin of Semitic in the Near East. *Proceedings of the Royal Society B: Biological Sciences*, *276*(1668), 2703–2710.

Kolipakam, V., Jordan, F. M., Dunn, M., Greenhill, S. J., Bouckaert, R., Gray, R. D., & Verkerk, A. (2018). A Bayesian phylogenetic study of the Dravidian language family. *Royal Society Open Science, 5*(3), 171504.

Kothari, S., Leone, A., & Wasley, C. (2005). Performance matched discretionary accrual measures. *Journal of Accounting and Economics,* 39(1), 163–197.

Lai, V. T., & Boroditsky, L. (2013). The immediate and chronic influence of spatio-temporal metaphors on the mental representations of time in English, Mandarin, and Mandarin-English speakers. *Frontiers in Psychology, 4,* 142.

Lee, S., & Hasegawa, T. (2011). Bayesian phylogenetic analysis supports an agricultural origin of Japonic languages. *Proceedings of the Royal Society B: Biological Sciences*, *278*(1725), 3662–3669.

Liang, H., Marquis, C., Renneboog, L., & Li Sun, S. (2014). *Future-time framing: The effect of language on corporate future orientatio*n (ECGI Finance Working Paper No. 412/2014).

Majid, A., Bowerman, M., Kita, S., Haun, D. B. M., & Levinson, S. C. (2004). Can language restructure cognition? The case for space. *Trends in Cognitive Sciences*, *8*(3), 108–114.

Majid, A., & Burenhult, N. (2014). Odors are expressible in language, as long as you speak the right language. *Cognition, 130*, 266–270. doi: 10.1016/j.cognition.2013.11.004

Majid, A., Roberts, S. G., Cilissen, R. (2018) The differential coding of perception in the world's languages. *Proceedings of the National Academy of Sciences*. *115*(45), 11369–11376.

Michael, L., Chousou-Polydouri, N., Bartolomei, K., Donnelly, E., Wauters, V., Meira, S., & O'Hagan, Z. (2015). A Bayesian phylogenetic classification of Tupi-Guarani. *LIAMES* 15(2), 1–36.

Naroll, R. (1965). Galton's problem: The logic of cross-cultural analysis. *Social Research*, 428-451.

Pagel, M. (1997). Inferring evolutionary processes from phylogenies. *Zoologica Scripta*, *26*(4), 331–348.

Pereltsvaig, A., & Lewis, M. W. (2015). *The Indo-European controversy.* Cambridge University Press.

R Core Team (2013). R: *A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/.

Roberson, D., Pak, H., & Hanley, J. R. (2008). Categorical perception of colour in the left and right visual field is verbally mediated: Evidence from Korean. *Cognition* 107(2), 752–762.

Robertson, C., & Roberts, S. G. (2020). Not when but if: Modality and future time reference in English, Dutch and German. *PsyArXiv*. https://doi.org/10.31234/osf.io/gdr8e.

Roberts, S. G., Torreira, F., & Levinson, S. C. (2015). The effects of processing and sequence organization on the timing of turn taking: A corpus study. *Frontiers in Psychology, 6*, 509.

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

Roberts, S. G., Winters, J., & Chen, K. (2015). Future tense and economic decisions: Controlling for cultural evolution. *PLOS ONE 10*(7), e0132145. doi:10.1371/journal.pone.0132145

Sela, R. J., & Simonoff, J. S. (2011). REEMtree: Regression Trees with Random Effects. R package version 0.90.3. https://cran.r-project.org/web/packages/REEMtree

Shoham, A., & Lee, S. M. (2018). The causal impact of grammatical gender marking on gender wage inequality and country income inequality. *Business & Society*, *57*(6), 1216–1251.

Sicoli, M. A., & Holton, G. (2014). Linguistic phylogenies support back-migration from Beringia to Asia. *PLOS ONE, 9*(3), e91722.

Spolaore, E., & Wacziarg, R. (2013). How deep are the roots of economic development? *Journal of Economic Literature*, *51*(2), 325–69.

Spolaore, E., & Wacziarg, R. (2014). Long-term barriers to economic development. In P. Aghion & S. Durlauf (Eds.), *Handbook of economic growth* (Vol. 2, pp. 121–176). Elsevier.

Strobl, C., Malley, J., & Tutz, G. (2009). An introduction to recursive partitioning: Rationale, application, and characteristics of classification and regression trees, bagging, and random forests. *Psychological Methods, 14*, 323. doi: 10.1037/a0016973

Tagliamonte, S. A., & Baayen, R. H. (2012). Models, forests, and trees of York English: Was/were variation as a case study for statistical practice. *Language Variation and Change*, *24*(2), 135–178.

Verkerk, A. (2013). Scramble, scurry and dash: The correlation between motion event encoding and manner verb lexicon size in Indo-European. *Language Dynamics and Change*, *3*(2), 169–217.

Whorf, B. L., Carroll, J. B., & Chase, S. (1956). Language, thought and reality: Selected writings of Benjamin Lee Whorf. *MIT Press*.

Appendix: Variable Definitions

$StrongFTR$	Dummy variable equal to one for firm-year observations from countries where the official languages are classified as having strong FTR by Chen (2013) suggested by Dahl (2000), zero otherwise
AEM	Absolute value of accrual-based earnings management, i.e., the absolute value of the difference between discretionary accruals [i.e., the residuals estimated from Equation (1)] for firm i and discretionary accruals for its performance-matched firm (with the closest ROA and in the same country-industry-year)
$INVPRO$	Investor protection score based on anti-director index from Djankov et al. (2008)
$GGROWTH$	Country GDP growth rate (data source: http://data.worldbank.org/indicator)
$INDIV$	Individualism/collectivism index score based on Hofstede (2001) (data source: http://geert-hofstede.com/countries.html)
UA	Uncertainty avoidance index score based on Hofstede (2001) (data source: http://geert-hofstede.com/countries.html)
MAS	Masculinity/femininity index score based on Hofstede (2001) (data source: http://geert-hofstede.com/countries.html)
PD	Power distance index score based on Hofstede (2001) (data source: http://geert-hofstede.com/countries.html)
LTO	Long-/short-term orientation index score based on Hofstede (2001) (data source: http://geert-hofstede.com/countries.html)
$LNSIZE$	Natural log of total assets adjusted for inflation rate
BTM	Book value of common equity divided by market value of equity
LEV	Short- and long-term debt divided by total assets
$LOSS$	Dummy variable that takes one for firm-year observations with negative income before extraordinary items, zero other wise
ROA	Income before extraordinary items, divided by total assets
$MEET$	Dummy variable that takes one for firm-year observations with actual annual EPS greater than or equal to consensus analyst earnings forecast, zero otherwise