7-21-2021

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

Giorgio Gotti
The University of Texas Rio Grande Valley

Seán G. Roberts

Marco Fasan

Cole B. J. Robertson

Recommended Citation

Gotti, G., Roberts, S.G., Fasan, M. and Robertson, C.B., 2021. Language in economics and accounting research: the role of linguistic history. The International Journal of Accounting, 56(03), p.2150015. https://doi.org/10.1142/S1094406021500153

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

Giorgio Gotti (*)
University of Texas at El Paso
e-mail: ggott@utep.edu

Sean Roberts
University of Bristol
e-mail: sean.roberts@bristol.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

(*) 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 (the linguistic relativity or Sapir-Whorf hypothesis). For example, the way a language allows one to talk about the future might influence future-oriented decisions (Chen, 2013), such as a company's earnings management (Kim et al., 2017; Cheng et al. 2019). However, languages have historical relations which lead to shared features, meaning that 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 for controlling for linguistic history. We then provide an empirical demonstration of how the Galton's problem can bias results in an investigation of whether a company's earnings management behaviour is predicted by structural features of the language of its employees. 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 a 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, or Sapir-Whorf, hypothesis is the suggestion that differences between languages can affect speakers' thoughts and behaviours (Whorf, 1956). A growing body of research has found that such interlinguistic differences may indeed affect people's perception and cognition (Evans & Levinson, 2009). 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 colour perception (Berlin & Kay, 1991; Roberson et al., 2008). The hypothesis has recently been taken up by economists who have sought to use interlanguage differences 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 (Kim et al., 2017; Hooghiemstra et al. 2019; Cheng et al., 2019). The objective of this study is to argue that a greater consideration of linguistic history is needed in these cross-cultural statistical studies. By linguistic history, we mean the historical relationships between languages, such as all Indo-European languages deriving form a single common ancestor. Related languages inherit similar linguistic features, creating clusters of non-independent languages. Our contribution is an overview of the problem and an empirical demonstration that controlling for linguistic history can make a difference to the inferences that are drawn from cross-cultural studies of language and economic behaviour.

The last few years have seen a growing interest in the Linguistic Savings Hypothesis (LSH), which suggests that interlinguistic differences in the grammaticalisation of Future Time Reference (FTR) may affect people's patience—their willingness to delay present gratification and instead wait for future rewards (Chen, 2013). It has been suggested that the interlinguistic difference most germane to time preferences is whether or not grammatical marking of future time is obligatory. To get a feel for this difference, consider the following two sentences:

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

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

The English sentence (1) uses the future auxiliary ‘will’ to indicate that the event will happen in the future. It would be ungrammatical in most contexts in English to use the unmarked present tense to speak about future events, e.g. ‘Tomorrow it rains.’ No such restrictions exist in the Germanic languages (other than English), Finnish, Estonian, or Mandarin. Indeed, in many languages around the world, speakers are free to use constructions similar to (2). The suggestion is that speakers of these languages might be more likely to perceive those events as being closer to them in time,[1] and because people tend to discount value as a factor of temporal distance (see Frederick et al., 2002), this is hypothesised to cause speakers of these languages to assign a higher value to future events and thus be more likely to make future-oriented decisions (Chen, 2013).

Chen (2013) used Dahl’s (2000) FTR typology to create a categorical variable which classified languages as either ‘weak-FTR’ or ‘strong-FTR’. Using data from the World Values Survey (Inglehart et al., 2014), he found that speakers of ‘weak-FTR’ languages (which commonly use the present tense in future time reference) tend to smoke less, be less obese, engage in safer sex practices, retire with more money, and save more. These are all signs of future-orientation.

Despite the apparent predictive power of future time reference, Chen’s original analysis has been criticised by linguists in at least two ways (see Roberts et al., 2015b). First, it fails to give adequate treatment to the linguistic complexities of future time reference, and, secondly, it did not control for non-independence in the data due to relatedness between languages, which may have caused the models reported to produce biased estimates. Indeed, Chen collaborated with two linguists to show that the

1 Chen (2013) actually suggested that there might be two distinct mechanisms which would cause people who spoke such languages to assign a higher value to future events—first, that they would perceive those events to be closer to them in time (because we associate the present tense with the immediacy of the present moment), and, secondly, because they might estimate the temporal distance of future events less precisely, which—provided that their discounting curve is strictly convex—would cause them to value the events less highly. See Chen (2013) for further explanation.

original relationship between FTR and savings behaviour was not significant when controlling for linguistic history (Roberts et al., 2015b). Despite this, studies continue to use the original binary variable and replicate these potential problems. For instance, Kim et al. (2017) show that the extent of a company's earnings management can be predicted by whether the main language of the country in which they trade has 'strong' or 'weak' FTR. In the next section we describe in detail why the linguistic issues matter and what can be done to address them. We then demonstrate the importance of this issue through an empirical investigation of the relationships between language and earnings management with and without controls for linguistic history. We show that the relationship between earnings management and FTR is largely driven by linguistic history, and end by discussing the implications for the field.

2. Literature review and hypothesis development

Prior studies of Whorfian economics

Since Chen (2013)'s study, a number of other studies have used the 'weak- vs. strong-FTR' categorisation to explain other economic variables. For instance, Hübner and Vannoorenberghe (2015) find that FTR predicts inflation rates (which may be sensitive to future-orientation) at the national level in a worldwide sample. Liang et al. (2014) find that national measures of sustainability and corporate responsibility, as well as institutional measures of corporate social responsibility, are negatively related to obligatory FTR marking ('strong-FTR' languages). Using data from the Swiss Household Panel, Guin (2016) reports that French-speaking ('strong-FTR') Swiss households saved less and overspent more that their German-speaking ('weak-FTR') counterparts. In a broader analysis of effects of long-term-orientation on educational outcomes, Figlio et al. (2016) find home use of 'weak-FTR' languages among first-generation immigrants in Florida predicted positive performance in math, reading, likelihood of graduation and retention, and fewer disciplinary incidents and absences, which mirrors the effect of long-term orientation more generally. Chen et al. (2017) find that companies in 'weak-FTR' countries tend to keep more precautionary cash reserves, indicating that they are making more long-term-oriented decisions.

There are also some experimental results. Angerer et al. (2015) presented a simple intertemporal choice task to schoolchildren in the Italian-German bilingual area of South Tyrol. They found that speakers of German ('weak-FTR') were 46% more likely to opt to delay gratification than speakers of Italian ('strong-FTR'). Thoma and Tytus (2018) conducted a series of experiments in which they conducted a temporal choice task among speakers of Chinese, German, Danish, Spanish and English, which they argued demonstrated increasing 'FTR-strength', in the order given. Confusingly, their findings showed a relationship between FTR-strength and temporal choice that runs in the opposite direction from most published studies, as speakers of Chinese were the least likely to choose a larger reward later, and speakers of English the most likely. They also conducted a linguistic priming task where English participants were given an intertemporal choice framed in either a syntactically marked (use of 'will') or a syntactically unmarked (no use of 'will') condition. They found no effect, though this manipulation does not appear to be particularly ecologically valid in languages in which the unmarked condition is ungrammatical. Similarly, Chen et al. (2019) performed an intertemporal choice experiment with Chinese speakers where the questions were phrased either with or without an explicit future tense marker. They also found no evidence that this manipulation caused a change in participants' economic decisions. The strongest causal evidence to date that linguistic factors are driving the observed differences in intertemporal choice comes from Pérez and Tavits (2017). They conducted 1,200 interviews in either Estonian (which does not have a future tense) or Russian (which does) with Estonian-Russian bilinguals in Estonia. After the interview, participants were asked whether they supported the addition of a new tax to help protect the environment. They found that participants interviewed in Estonian were significantly more likely to express support for the tax, indicating that they may have valued the outcome (e.g. environmental protection) more highly. However, all these studies have significant issues with how the linguistic structures are treated, to which we now turn.

Linguistic Issues

Organising languages into typological clusters is complicated, and a binary variable does not accurately capture all their nuances. Consider the following: even though Chen (2013) classifies German and English as 'weak' and 'strong' FTR languages, respectively, the future tense system in both languages is very similar. The future tense in English is formed by either the present progressive (*going to* + *infinitive*), or by the modal auxiliary *will* + *infinitive*. Equivalently to the English *will*, the German future is formed by the modal auxiliary *werden* + *infinitive*. The main difference is that in German it is possible—but not necessary—to drop *werden* some of the time. Calling one 'weak' and the other 'strong' conflates English with languages like French, which exhibits obligatory inflectional marking of the future tense,[2] and conflates German with languages like Finnish, which hardly exhibits any marking of FTR at all. In fact, even in Romance languages with strict inflectional FTR rules, such as Spanish and French, it is often possible to use the present tense in reference to the future, provided that the future is close and well known (Dahl, 2000).

It gets more complicated. We can refer to the future using three epistemological categories: *prediction*, *intention* and *scheduling* (Dahl, 2000). These categories differ in their epistemic valence: an intention is more well-known than a prediction, and less well-known than a schedule. Moreover, different grammatical strategies for referring to the future are not evenly distributed across these categories. When Chen (2013) borrowed Dahl's (2000) typology, the paper also borrowed Dahl's definition of 'futureless' languages as those which do not require obligatory grammatical marking *in prediction contexts*. This may not be the only salient typological criterion when investigating psycholinguistic effects. Consider that English speakers must use the future tense in prediction contexts, but can use the unmarked present in scheduling contexts ('The train *arrives* at 6 p.m.') and the present progressive in intention contexts ('I

2 To modify the present tense sentence *il pleut* ('it rains') to refer to the future, French speakers are required to inflectionally modify the main verb: *il pleuvera* ('it *will* rain').

am going for a walk tomorrow'). However, it is currently unknown how Whorfian effects apply and interact across different contexts.

Secondly, modal constructions (such as the English auxiliary verbs *would, could, should,* or *must*) and adverbial/adjectival modifiers (such as *surely, possibly,* and *probably*) are extremely common in future time reference contexts. This is important because these constructions, when used in prediction contexts, tend to express shades of certainty (e.g. 'It *will surely* rain later today.') vs. possibility (e.g. 'It *could probably* rain later today'). This may be highly relevant to the LSH because the proportional use of such expressions as a fraction of total future time referencing sentences is likely to differ between languages. Such expressions could either reflect cultural differences regarding underlying beliefs about the likelihood of a future event coming to pass, or, conversely, could come to influence such beliefs over time (in, for instance, a language in which speakers used a relatively large number of such constructions). Clearly these constructions are a salient typological consideration for any investigation of the effects of interlinguistic differences on future-orientation.

Finally, even within languages, speakers' usage habits differ greatly (see Hughes et al., 2013; Cacoullos & Walker, 2009), such that some speakers of a given 'weak-FTR' language might nearly always use the future tense, while others might hardly ever do so. This does not mean that there cannot be meaningful interlanguage differences, but it does mean that comparisons between populations at the language level may overlook substantive individual differences in language usage habits which might affect between-individual differences in patience.

The complexities above cast doubt on whether the binary 'strong' versus 'weak' distinction is a valid measure. However, the variation between speakers and languages could be harnessed: future research into the LSH might refocus on investigating whether individual usage habits predict economic behaviour, or at least try to get direct measures of future time reference usage (such as the proportion of future tense usage in cross-linguistic corpora from Chen, 2013). As well as making use of meaningful individual differences, this would help address the issue of cultural confounding, which we address in the next section.

Linguistic history

In general, if data in a sample are not independent, then 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 a 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 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 and cover 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, 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 neighbouring languages over long periods of time. This leads to *areal patterns*: languages within the same geographical regions tend to be similar. These historic process of inheritance and borrowing lead 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 behaviour (see Spolaore & Wacziarg, 2013, 2014; Alesina & Giuliano, 2015).

[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 which capture 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 conservative, with a change only happening once every 10,000 years of independent evolution. Roberts et al. (2015b) estimated that the binary future time reference variable also showed 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).

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. Random effects have been used in linguistics and psychology to capture non-independencies between observations (Clark, 1973; Baayen et al., 2008), including controlling for linguistic history by entering language family as a random effect (e.g. Roberts et al., 2015b). Areal effects can be controlled for in the same way by including the geographical area xxx. When applying this method to Chen's (2013) original data, the effect of FTR on savings behaviour disappears (Roberts et al., 2015b). Similarly, Chen et al. (2017) study the relationship between FTR and corporate savings behaviour and include robustness tests where language family or continent are included as fixed effects. However, we note that 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. As a minor note, grouping 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.

Studies that compare 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 labour 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) that have 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 (possible 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.

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 (AAM), 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 AAM practices than countries whose main language has weak FTR (the same result as Kim et al., 2017).

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

Any statistical relationship might disappear when adding additional controls due to a lack of power rather than a lack of a real relationship. However, our prediction is specifically about linguistic variables. We expect non-linguistic predictors to survive controls for linguistic history, hence an additional hypothesis:

H3: When controlling for linguistic history, the relationship between AAM and non-linguistic measures of economic behaviour will not be diminished.

3. 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.

Sample and measures

We measure earnings management following Kothari et al. (2005); thus, we proxy for accrual-based earnings management (AAM) by performance-matched discretionary accruals. We use the following variables as controls:

- Investor protection score, based on the anti-director index from Djankov et al. (2008) (invpro);
- Power distance index, based on Hofstede (2001) (pd);
- Individualism/collectivism score, based on Hofstede (2001) (indiv);
- Masculinity/femininity score, based on Hofstede (2001) (mas);
- Uncertainty avoidance score, based on Hofstede (2001) (ua);
- Long-/short-term orientation score, based on Hofstede (2001) (lto);
- Indulgence, based on Hofstede (2001) (indul);
- Country GDP growth rate (ggr);
- Company size, measured as the natural logarithm of total assets adjusted for inflation rate (size);

- Book value of common equity divided by common value of equity (btm);
- Leverage, measured as short- and long- term debt divided by total assets (lev);
- Return on assets, measured as income before extraordinary items divided by total assets (roa);
- Dummy variable that takes one for firm-year observations with actual annual EPS greater than or equal to consensus analyst earnings forecast, zero otherwise (meet);
- Dummy variable that takes one for firm-year observations with negative income before extraordinary items, zero otherwise (loss).

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.

3.1. *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 there being 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. All continuous variables in the regressions were scaled and centered to have a mean of 0 and a standard deviation of 1.

[Insert Table 2 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]

Modelling

The main analysis uses mixed-effects modelling in R (R Core Team, 2013) using the lme4 package (Bates et al., 2015). 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, ggr, size, btm, lev, roa, meet* and *loss*.

Since AAM 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 AAM 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.). 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 (using the R package *MCMCglmm*, Hadfield, 2010) on Indo-European data, predicting AAM 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, 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.

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 recursively partitions data into bins according to a series of yes/no questions. It is used in many applications, including to explore patterns in linguistic data (Majid et al., 2018; Roberts et al., 2015a; Tagliamonte & Baayen, 2012). The decision tree predicted AAM 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 AAM, 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 at https://github.com/seannyD/FTRAccountingStudy.

4. 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 for 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 fit of the model 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 AAM disappears when controlling for linguistic history. In contrast, and in accordance with H3, most non-linguistic predictors of AAM remained significant or became stronger when controlling for language family (table 3). For example, the effect of a company's book value remained roughly the same (from $\beta = -0.036$ to $\beta = -0.044$) 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 non-significant.

[Insert Table 3 and Figure 4 here]

The same results and inferences were obtained using the alternative methods. When using a Gaussian distribution 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, $\beta = 0.02$, 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 AAM (H2, $\beta = 0.95$ [-0.41, 2.32], ESS = 1069, $p = 0.18$). In contrast, LTO was a significant predictor of earnings management (AAM) under the same test (H3, $\beta = -0.44$ [-0.90, -0.02], ESS = 8875, $p = 0.048$). The decision tree did not select the FTR variable to predict earnings management (AAM), 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, $\beta = 0.13$, SE = 0.03, $p < 0.0001$) except when clustering robust standard errors by language family (H2, $\beta = 0.13$, SE = 0.08, $p = 0.13$). In contrast, nine other non-linguistic predictors remained significant under this test (H3, see Table 4).

[Insert Table 4 here]

5. 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 behaviour. 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 which reflect inheritance of features from ancestor languages. They have also identified areal patterns, caused by features being borrowed between 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 Kim et al. (2017), suggests that grammatical rules about referring to the future (FTR) are significantly associated with earnings management, consistently with the results obtained by the original authors. Once we control for linguistic history, though, the association between future tense and earnings management becomes non-significant. Controlling for linguistic history increases the standard error but also reduces the magnitude of the estimate for FTR. This is not the case for non-linguistic variables like 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, inference in 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 behaviour within cultures.

Tables and figures

Figure 1

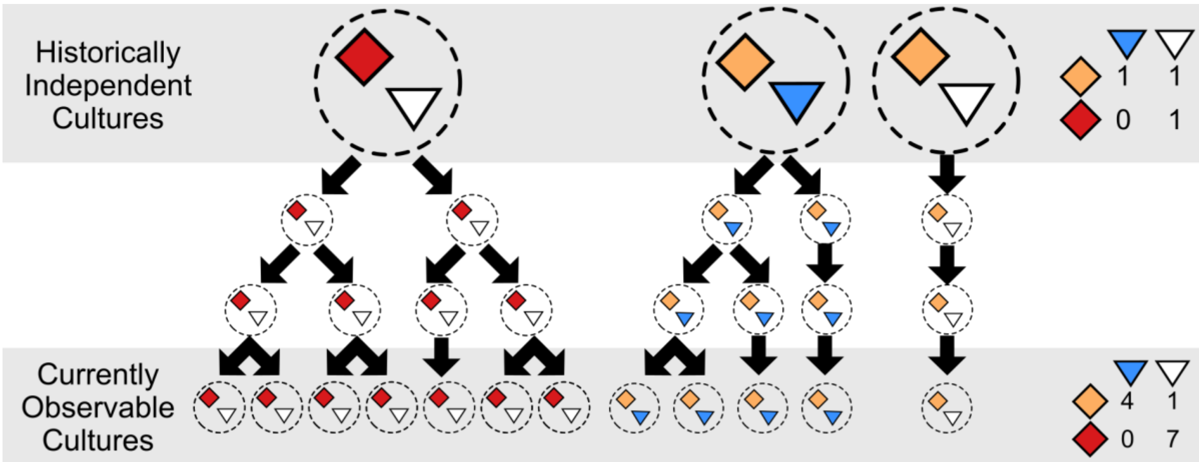
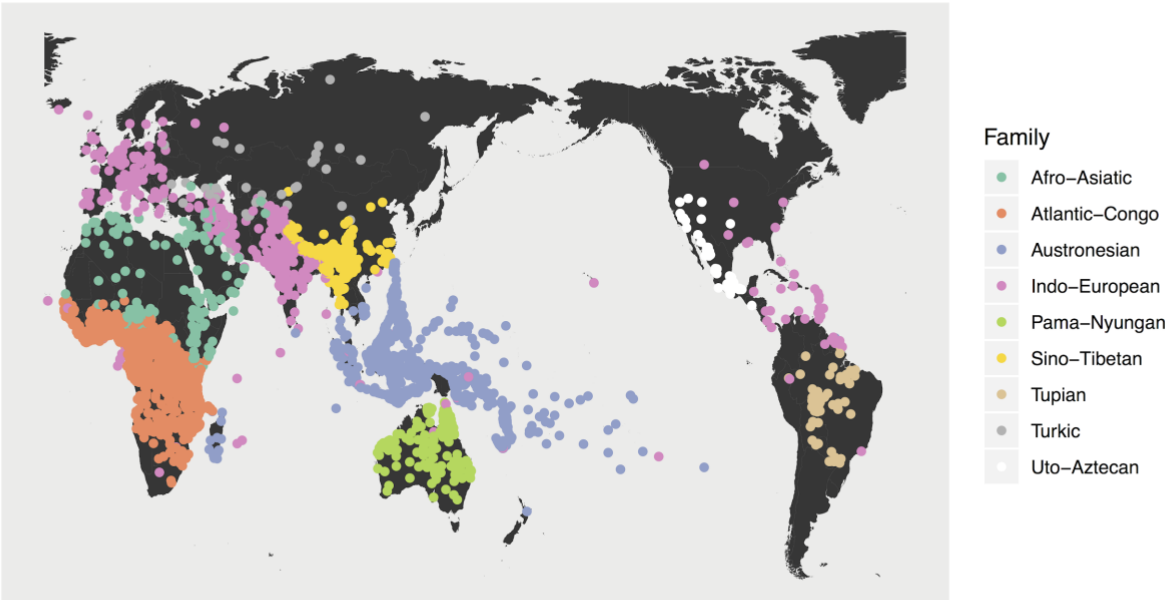


Fig. 1. An illustration of how cultural inheritance can lead to spurious correlations (adapted from Roberts, Winters & Chen, 2015). At the top are three independent historical cultures, each of which has a bundle of various traits which are represented as coloured shapes. Each trait is causally independent of the others. On the right is a contingency table for the colours of triangles and diamonds. Originally, there is no particular relationship between the colour of triangles and the colour 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



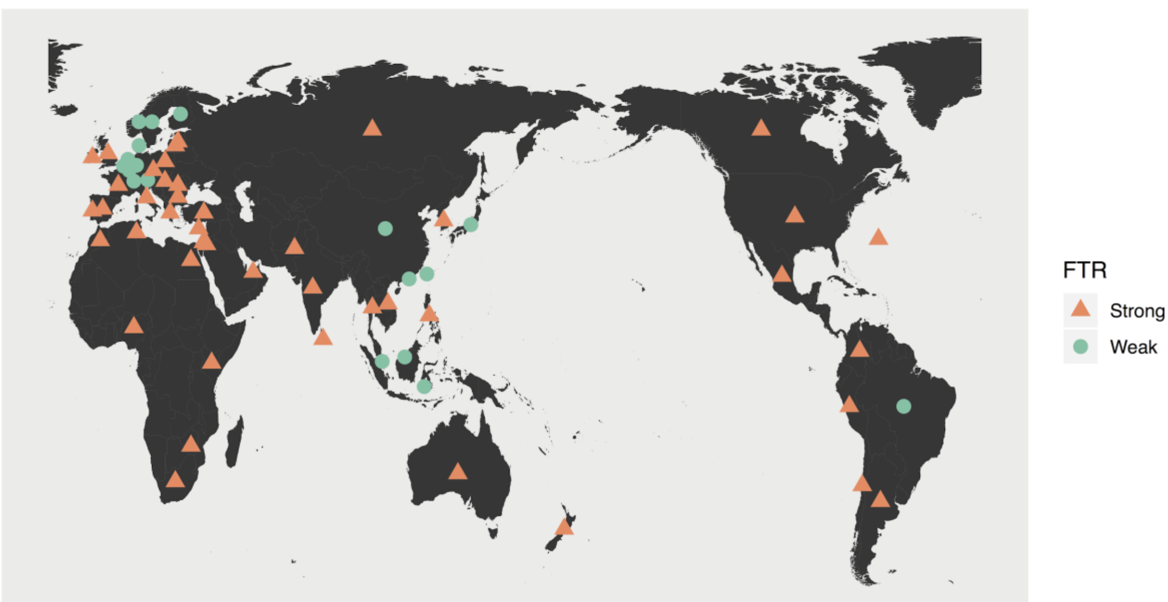


Figure 2: Top: Distribution of some of the world's largest language families, including relatively recent creoles and pidgins. Data from Glottolog (Hammarström et al., 2018). Bottom: Map of countries and their main languages in the study, labelled by the FTR value of the main language.

Figure 3

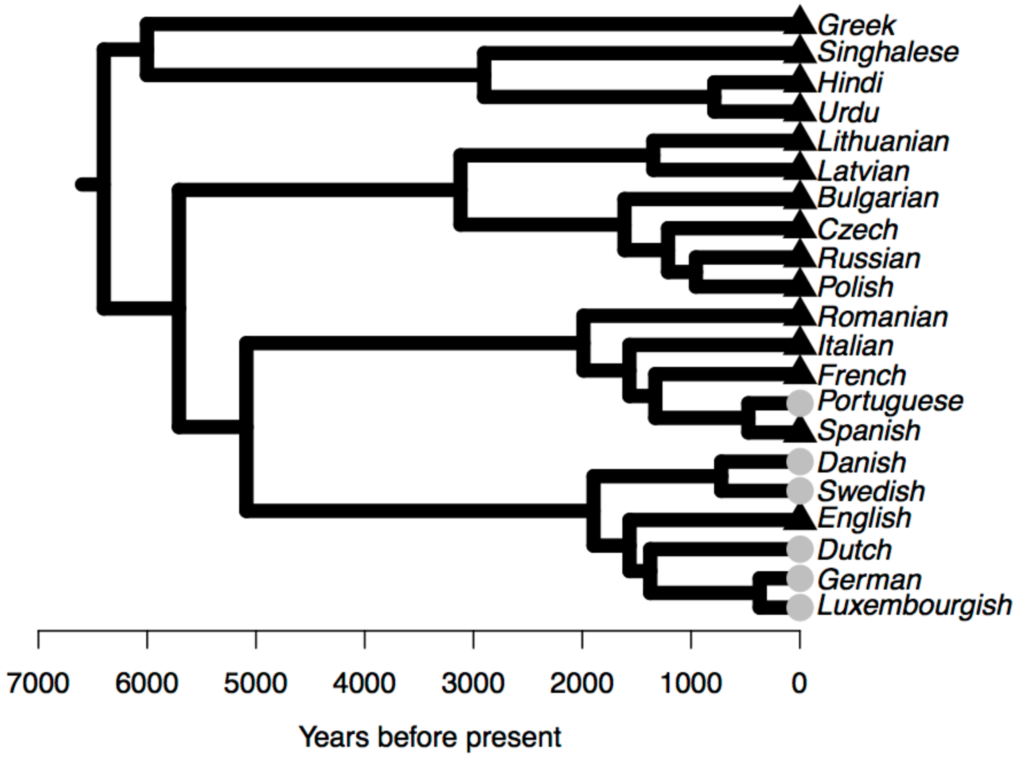


Fig. 3. The phylogenetic tree used in the analysis (Bouckaert et al., 2012). The tips show the FTR value of each language (black triangle = strong, gray circle = weak).

Figure 4

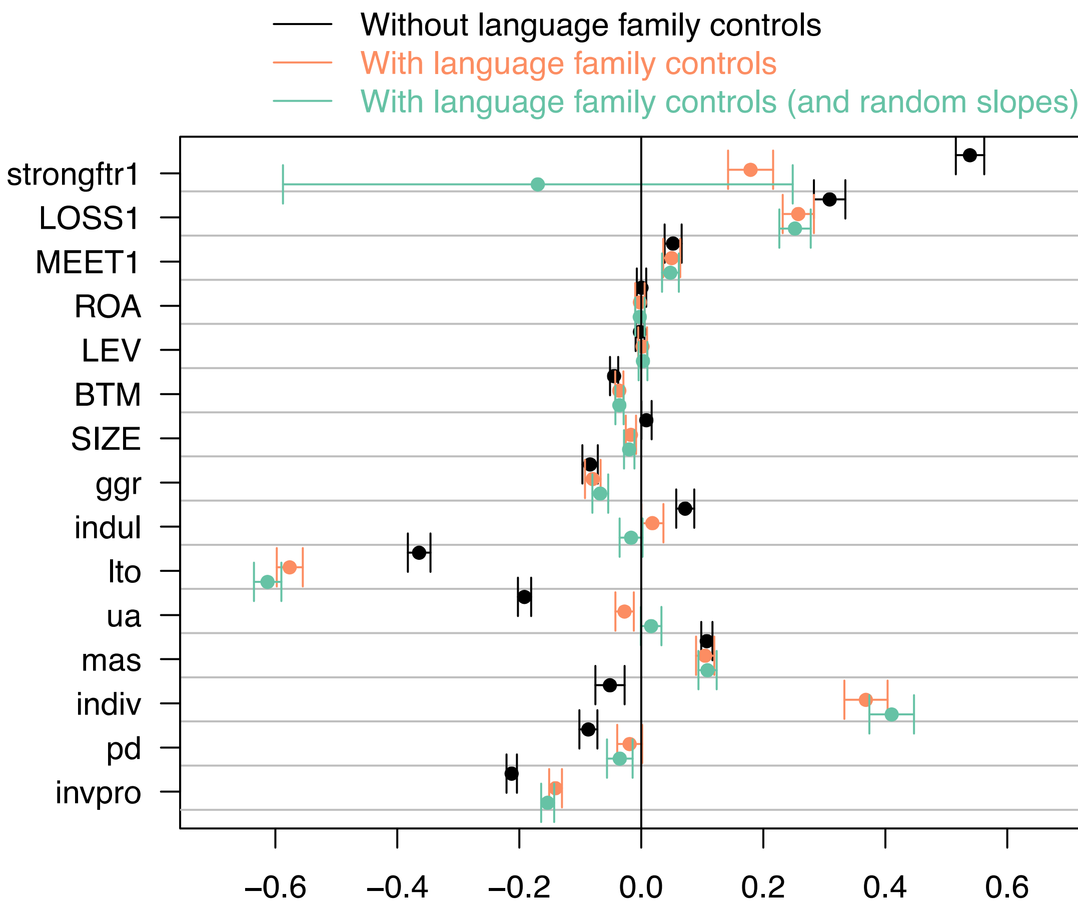


Fig. 4. Model estimates (β values) for different variables in the model, without controls for language family (black), with controls (orange) and with additional random slopes (green). Error bars show 95% confidence intervals for the estimate.

Table 1

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)

Table 1. List of language families for which dated phylogenies are available (https://d-place.org/phylogenys).

Table 2

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

Table 3

Variable	No controls for language history	With controls for language history	Robust?
Power distance	-0.087 ($p < 0.0001$)	-0.035 ($p = 0.001$)	Yes
Masculinity	0.11 ($p < 0.0001$)	0.11 ($p < 0.0001$)	Yes
Long-term orientation	-0.36 ($p < 0.0001$)	-0.61 ($p < 0.0001$)	Yes
GDP growth rate	-0.084 ($p < 0.0001$)	-0.067 ($p < 0.0001$)	Yes
Company size	0.0084 ($p = 0.049$)	-0.02 ($p < 0.0001$)	Yes
Book value	-0.045 ($p < 0.0001$)	-0.036 ($p < 0.0001$)	Yes
Leverage	-0.0021 ($p = 0.57$)	0.0029 ($p = 0.43$)	Yes
Return on assets	0.00038 ($p = 0.92$)	-0.0024 ($p = 0.54$)	Yes
Individualism	-0.051 ($p < 0.0001$)	0.41 ($p < 0.0001$)	No
Uncertainty avoidance	-0.19 ($p < 0.0001$)	0.016 ($p = 0.061$)	No
Indulgence	0.072 ($p < 0.0001$)	-0.017 ($p = 0.084$)	No

Table 3. Estimates of how variables predict earnings management with and without controls for linguistic history. 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 4

VARIABLES	(1) **AAM_scaled** **no control**	(2) **AAM_scaled** **with language family clustered se**
Effect of FTR	**0.132*****	**0.132**
	(39.618)	**(1.667)**
Investor protection	-0.099***	-0.099**
	(-50.273)	(-2.587)
Power distance	0.010***	0.010
	(7.039)	(0.232)
Individualism	0.021***	0.021
	(7.494)	(0.518)
Masculinity	0.062***	0.062**
	(41.122)	(2.669)
Uncertainty avoidance	-0.044***	-0.044
	(-28.515)	(-1.605)
Long term orientation	-0.111***	-0.111**
	(-40.732)	(-2.827)
Indulgence	0.025***	0.025
	(12.241)	(1.549)
GDP growth	-0.072***	-0.072*
	(-28.949)	(-1.932)
Size	0.036***	0.036***
	(9.909)	(5.299)
BTM	-0.020***	-0.020*
	(-9.796)	(-2.240)
Leverage	-0.005	-0.005*
	(-1.384)	(-2.083)
ROA	0.009**	0.009
	(2.178)	(1.629)
meet	0.030***	0.030***
	(4.809)	(4.557)
loss	0.170***	0.170***
	(9.9929)	(6.241)
Constant	0.225***	0.225
	(24.902)	(1.443)
Observations	94,707	94,707
R-squared	0.073	0.073
Adj. R-squared	0.073	0.073

Table 4.

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.

Angerer, S., Glatzle-Rutzler, D., Lergetporer, P., Sutter, M., 2015. Donations, risk attitudes and time preferences: A study on altruism in primary school children. J. Econ. Behav. Organ. 115, 67–74.

Baayen, R.H., Davidson, D.J., Bates, D.M., 2008. Mixed-effects modeling with crossed random effects for subjects and items. J. Mem. Lang. 59(4), 390–412.

Bates, D., Maechler, M., Bolker, B., Walker, S., 2015. Fitting linear mixed-effects models using lme4. J. Stat. Softw. 67(1), 1–48.

Berlin, B., Kay, P., 1991. Basic Color Terms: Their Universality and Evolution. Univ. of California Press, Berkeley.

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: Hopcroft, R.L. (Ed.), Oxford Handbook of Evolution, Biology, and Society. Oxford University Press, Oxford.

Bouckaert, R.R., Bowern, C., and Atkinson, Q.D., 2018. The origin and expansion of Pama-Nyungan languages across Australia. Nat. Ecol. Evol. 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., Atkinson, Q.D., 2012. Computational phylogenetics and the internal structure of Pama-Nyungan. Language, 88(4), 817–845

Bowern, C., 2018. Computational phylogenetics. Annu. Rev. Linguist. 4, 281–296.

Cacoullos, R.T., Walker, J.A., 2009. The present of the English future: Grammatical variation and collocations in discourse. Language 85(2), 321–354.

Chacon, T.C., List, J.M., 2015. Improved computational models of sound change shed light on the history of the Tukanoan languages. J. Lang. Relatsh. 3, 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, M.K. (2013). The effect of language on economic behavior: Evidence from savings rates, health behaviors, and retirement assets. Am. Econ. Rev. 103(2), 690–731.

Chen, S., Cronqvist, H., Ni, S., Zhang, F., 2017. Languages and corporate savings behavior. J. Corp. Finance 46, 320–341.

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.

Cheng, A., Kim, J., Kim, Y., Zhou, J. (2019) Languages and Tax Avoidance. *SSRN Electronic Journal*.

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

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

Donohue, M., Wichmann, S., Albu, M., 2008. Typology, areality, and diffusion. Ocean. Linguist. 47(1), 223–232.

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

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. Behav. Brain Sci. 32(5), 429–448

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

Frederick, S., George, L., O'Donoghue, T., 2002. Time discounting and time preference: A critical review. J. Econ. Lit. 40(2), 351–401.

Gay, V., Hicks, D.L., Santacreu-Vasut, E., Shoham, A., 2018. Decomposing culture: an analysis of gender, language and labor supply in the household. Rev. Econ. Househ. 16(4), 879–909.

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.

Grollemund, R., Branford, S., Bostoen, K., Meade, A., Venditti, C. Pagel, M., 2015. Bantu expansion shows habitat alters the route and pace of human dispersals. Proc. Natl. Acad. Sci. U.S.A. 112(43), 13296–13301.

Guin, B., 2016. Culture and household saving. SSRN. http://dx.doi.org/10.2139/ssrn.2698872, retrieved December 1, 2018

Hadfield, J.D., 2010. MCMC methods for multi-response generalized linear mixed models: The MCMCglmm R package. J. Stat. Softw. 33(2), 1–22. http://dx.doi.org/10.18637/jss.v033.i02, retrieved December 1, 2018

Hammarström, H., Forkel, R., Haspelmath, M., 2018. Glottolog 3.3. Max Planck Institute for the Science of Human History, Jena. http://glottolog.org

Hofstede, G., 2001. Culture's Consequences: Comparing Values, Behaviors, Institutions, and Organizations Across Nations. SAGE, Thousand Oaks, CA.

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. J. Evol. Biol. 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, doi: 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, London/New York.

Hubner, M., Vannoorenberghe, G., 2015. Patience and long-run growth. Econ. Lett. 137, 163–167.

Inglehart, R., Haerpfer, C., Moreno, A., Welzel, C., Kizilova, K., Diez-Medrano, J., Lagos, M., Norris, P., Ponarin, E., Puranen, B., et al. (Eds.), 2014. World Values Survey: Round Four—Country-Pooled Datafile. www.worldvaluessurvey.org/WVSDocumentationWV4.jsp. JD Systems Institute, Madrid.

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

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

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. Proc. Biol. Sci. 270(1668), 2703–2710.

Kim, J., Kim, Y., Zhou, J., 2017. Languages and earnings management. J. Account. Econ. 63 (2–3), 288–306.

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. R. Soc. Open Sci. 5(3), 171504.

Kothari, S., Leone, A., Wasley, C., 2005. Performance matched discretionary accrual measures. J. Account. Econ. 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. Front. Psychol. 4, 142.

Lee, S., Hasegawa, T., 2011. Bayesian phylogenetic analysis supports an agricultural origin of Japonic languages. Proc. Biol. Sci. 278(1725), 3662–9.

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

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., Bowerman, M., Kita, S., Haun, D.B.M., Levinson, S.C., 2004. Can language restructure cognition? The case for space. Trends Cogn. Sci. 8(3), 108–114.

Majid, A., Roberts, S.G., Cilissen, R., 2018 The differential coding of perception in the world's languages. Proc. Natl. Acad. Sci. U.S.A. 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.

Pagel, M., 1997. Inferring evolutionary processes from phylogenies. Zool. Scr. 26, 331–348.

Pereltsvaig, A., Lewis, M.W., 2015. The Indo-European Controversy. Cambridge University Press, Cambridge.

Pérez, E.O., Tavits, M., 2017. Language shapes people's time perspective and support for future-oriented policies. Am. J. Pol. Sci. 61(3), 715–727.

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.

Roberts, S.G., Torreira, F., Levinson, S.C., 2015a. The effects of processing and sequence organization on the timing of turn taking: a corpus study. Front. Psychol. 6, 509.

Roberts, S.G., Winters, J., Chen, K., 2015b. 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. Bus. Soc. 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 *Handbook of economic growth* (Vol. 2, pp. 121-176). Elsevier.

Tagliamonte, S.A., Baayen, R.H., 2012. Models, forests, and trees of York English: Was/were variation as a case study for statistical practice. Lang. Variat. Change 24, 135–178.

Thoma, D., Tytus, A.E., 2018. How cross-linguistic differences in the grammaticalization of future time reference influence intertemporal choices. Cogn. Sci. 42, 974–1000.

Verkerk, A., 2013. Scramble, scurry and dash: The correlation between motion event encoding and manner verb lexicon size in Indo-European. Lang. Dyn. 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, Cambridge MA.