Common Ancestry, Uncommon Findings: Revisiting Cross-Cultural Research in Economics[*]

Boris Gershman[†]
American University

Tinatin Mumladze[‡]
American University

November 2024

Abstract

Empirical research on culture and institutions in economics often relies on cross-cultural data to examine historical or contemporary variation in traits across ethnolinguistic groups. We argue that this work has not adequately addressed the well-known problem of cultural non-independence due to common ancestry and show how phylogenetic regression, along with newly available global language trees, can be used to directly account for this issue. Our analysis focuses on Murdock's *Ethnographic Atlas*, a widely used database of preindustrial societies, with broader implications for any cross-cultural study. First, we show that various economic, institutional, and cultural characteristics in the *Atlas* exhibit phylogenetic signal – they tend to be more similar among societies with closer ancestral ties – highlighting the non-independence of observations. Second, through simulations in a sample resembling the *Ethnographic Atlas*, we demonstrate that phylogenetic correlation leads to severe inefficiency of the standard OLS estimator compared to phylogenetic generalized least squares (PGLS). Furthermore, although clustered standard errors partially mitigate the problem, OLS estimation yields unacceptably high type I error rates, frequently detecting a statistically significant relationship where none exists. Finally, we revisit some of the recently published results in a phylogenetic regression framework. In many specifications, PGLS estimates differ markedly from their OLS counterparts, indicating a smaller magnitude and weaker statistical significance of relevant coefficients.

Keywords: Common ancestry, Cross-cultural analysis, Culture, Cultural non-independence, Ethnographic Atlas, Institutions, Phylogenetic comparative methods

JEL Classification Numbers: C10, O10, N30, Z12, Z13

---

[*] We are grateful to seminar and conference participants at the 2024 ASREC conference at Chapman University, 2024 conference on deep-rooted factors in comparative development at Brown University, 2024 IRES graduate workshop at Chapman University, American University, George Mason University, University of Maryland, Baltimore County, and Williams College. Quamrul Ashraf, James Fenske, Erik Kimbrough, Nathan Larson, Stelios Michalopoulos, Jonathan Schulz, and David Weil contributed useful comments.

[†] Department of Economics, American University, 4400 Massachusetts Avenue NW, Washington, DC 20016-8029 (e-mail: boris.gershman@american.edu).

[‡] Department of Economics, American University, 4400 Massachusetts Avenue NW, Washington, DC 20016-8029 (e-mail: tm3538a@american.edu).

1 Introduction

In recent years, there has been a surge of empirical research on culture, institutions, and long-run economic development (Gershman, 2017; Nunn, 2020; Lowes, 2023). This work frequently relies on cross-cultural data to explore variation in relevant traits across "societies," which are typically represented by ethnolinguistic groups. A critical issue in such statistical analyses is the non-independence of observations due to shared ancestry and cultural evolutionary history of human societies around the globe (Mace and Pagel, 1994). We argue that this well-known problem has not been adequately addressed by economists and advance the use of phylogenetic comparative methods (PCMs) as a direct and effective approach to account for it.

PCMs are a set of techniques designed to explicitly incorporate phylogenetic trees, representing ancestral relationships between units of observation, into statistical analysis. Originally developed by evolutionary biologists for cross-species comparisons, PCMs have since been adopted by anthropologists and applied in cross-cultural studies (Nunn, 2011). Despite their widespread use in evolutionary anthropology, PCMs have been largely overlooked by economists working with similar datasets and facing the same empirical challenges. This paper aims to bridge the methodological gap between disciplines, offering a guide for applying phylogenetic regression in a typical empirical setting and formally comparing its performance to standard approaches in economics.

Although our insights are relevant for any research employing cross-cultural data, we focus on the *Ethnographic Atlas* (EA), a global sample of more than 1,200 preindustrial societies that is widely used in economics (Murdock, 1967; Lowes, 2021). To approximate the ancestral relationships between these societies, we leverage the newly available global phylogenies of languages (Bouckaert et al., 2022). After constructing a matched dataset – linking ethnographic information from the EA to the corresponding tips of the global language trees – we conduct phylogenetic signal tests to measure the tendency of closely related societies to share similar traits. Our findings show that all commonly examined economic, institutional, and cultural characteristics exhibit a substantial phylogenetic signal, consistent with non-independence of societies due to shared ancestry.

Next, we introduce phylogenetic regression, a key framework for analyzing associations between variables when observations are phylogenetically correlated (Revell, 2010). In this setting, the covariance matrix of the error terms is computed based on a given language tree, capturing ancestral relationships between societies, and an assumed cultural evolutionary process occurring along the branches of this tree. The resulting covariance matrix typically includes non-zero off-diagonal elements, which reflect the non-independence of observations due to common ancestry, and can be used to obtain more accurate estimates of regression parameters. Specifically, the phylogenetic generalized least squares (PGLS) estimator reweights observations based on their phylogenetic relatedness (Symonds and Blomberg, 2014).

We conduct a series of simulations to compare the performance of PGLS to the ordinary least squares (OLS) estimator. Specifically, we use the matched sample of 1,184 societies from the EA to repeatedly generate two variables arising from *independent* evolution of two respective traits on global language trees. We then estimate the relationship between these two variables, which are, by construction, unrelated to each other. Consistent with basic econometric theory, we find that the non-independence of observations leads to severe inefficiency of the OLS estimator, with its variance exceeding that of PGLS by more than 200 times in the bivariate regression model. Although controlling for continent, country, and language fixed effects improves the efficiency of the OLS estimator, PGLS remains more accurate in the vast majority of simulations.

Furthermore, in our simulations, OLS produces unacceptably high type I error rates, incorrectly identifying a statistically significant association between two unrelated variables. In the simplest bivariate regression case, the error rate reaches an alarming 90%. Clustering standard errors by continent, country, or language family – a common practice in economics – helps to reduce this problem, but type I error rates remain elevated, ranging from 21% to 69% depending on model specification and clustering category. In contrast, PGLS consistently yields type I error rates near the nominal 5% level.

Informed by the simulation results, we revisit some of the recently published research. Specifically, we reanalyze a variety of regression models from 10 papers that employ the EA data, comparing OLS estimates to their PGLS counterparts. We detect a substantial phylogenetic signal in the error terms of most regressions, which persists even after including control variables and various fixed effects. This highlights the importance of accounting for phylogenetic correlation across model specifications. We find that PGLS estimates tend to be substantially smaller in magnitude than OLS estimates, though this difference diminishes after including spatial and language family fixed effects. PGLS estimation typically reduces the statistical significance of relationships compared to OLS specifications with clustered standard errors. Overall, PGLS estimates are much more stable and less sensitive to model specification than OLS, and they remain robust to alternative global language phylogenies. The discrepancy between PGLS and OLS results remains, but is somewhat smaller in studies focused on Africa compared to those using global samples.

This paper builds a connection between empirical research on culture across disciplines and contributes primarily to two strands of literature. First, we make a methodological contribution to the growing body of work on culture and comparative development in economics. While economists spearheaded many innovations in addressing causal identification and other statistical challenges, the issue of non-independence due to common ancestry has largely escaped their attention. Attempts to account for it in regression analyses have mostly been limited to clustering standard errors by language family or introducing fixed effects to capture cultural relatedness. We argue that these approaches are conceptually inadequate and demonstrate through simulations that they cannot fully address the issue. Instead, we advocate for the use of PCMs, which, to the best of our knowledge, have not been adopted by economists yet, with the likely sole exception of our recent study on headhunting in Austronesia (Gershman and Mumladze, 2024).

It is particularly surprising that the issue of cultural non-independence has been overlooked, given the vibrant research on ancestry, cultural proximity, and diversity in this field (Spolaore and Wacziarg, 2016; Ginsburgh and Weber, 2020). By contrast, substantial research has been conducted on the problem of spatial non-independence (Conley, 2010; Conley and Kelly, 2024), and statistical adjustments for spatial correlation have become routine in empirical applications. We argue that research employing cross-cultural data should adopt similarly rigorous strategies to account for phylogenetic correlation.

A recent paper by Claessens et al. (2023) highlights both spatial and cultural non-independence in cross-country studies, demonstrating that many findings in this literature, including those by economist, are not robust when controlling for these dependencies using global proximity matrices. While this paper is conceptually aligned with ours, its exclusive focus on countries is methodologically limiting. Since countries are typically multiethnic and multilingual entities, they cannot be uniquely mapped onto language trees, which makes PCMs – specifically designed to model cultural dynamics on phylogenies – not properly applicable. Additionally, since country borders are often endogenous and causal identification with cross-country data is generally problematic, economists have increasingly employed more disaggregated units of analysis, such as ethnic groups and their homelands (Michalopoulos and Papaioannou, 2018). Our focus on societies rather than countries is in line with this recent trend in the literature and enables a proper application of phylogenetic regression and related methods.

Second, we both draw from and contribute to the extensive literature in evolutionary anthropology and related fields that employs PCMs to study cultural evolution (Mace and Zhang, 2023). As pioneers in cross-cultural studies, anthropologists have been acutely aware of the phylogenetic non-independence problem for over a century and have led the adoption of methods from evolutionary biology to address it. Most recent research in this area has focused on societies representing specific language clusters, such as the Austronesian, Bantu, or Indo-European families, for which reliable phylogenies are available. Due to the challenge of constructing a comprehensive tree of world languages, global-scale studies have been rare and mostly limited to small samples of societies (Ringen et al., 2019; Surowiec et al., 2019; Šaffa et al., 2022). We extend the scope of analysis by using the newly available global language trees and the EA to create a large matched dataset suitable for many applications. Rather than focusing on a specific research question, we examine multiple societal characteristics and revisit some of the published research within a phylogenetic regression framework. Furthermore, inspired by the approach common in the evolutionary biology literature, we conduct a series of simulations to emphasize the importance of accounting for phylogenetic non-independence specifically in this commonly used global sample of societies.

The remainder of this paper is organized as follows. Section 2 briefly discusses the issue of non-independence due to shared ancestry. Section 3 provides a detailed description of the data used in our study. Section 4 introduces phylogenetic signal tests and applies them to the EA data. Section 5 sets up the phylogenetic regression framework and presents our simulation exercise. Section 6 revisits a selection of published results after accounting for phylogenetic non-independence of observations. Section 7 concludes.

2 Non-independence due to shared ancestry

The non-independence of human societies as units of observation, along with the resulting statistical challenges in cross-cultural research, has long been recognized in the social sciences. In one of the earliest quantitative studies of its kind, Tylor (1889) examined correlations between various customs related to the “laws of marriage and descent” by tabulating ethnographic data from several hundred societies around the world. Although this was a pioneering work in its own right, it gained additional prominence due to a brief but influential comment by Francis Galton, then president of the Royal Anthropological Institute of Great Britain and Ireland, in response to Tylor’s presentation. The relevant part of Galton’s remarks, published in the discussion section following the main article, is as follows:

> *It was extremely desirable for the sake of those who may wish to study the evidence for Dr. Tylor’s conclusions, that full information should be given as*

> *to the degree in which the customs of the tribes and races which are compared together are independent. It might be, that some of the tribes had derived them from a common source, so that they were duplicate copies of the same original. Certainly, in such an investigation as this, each of the observations ought, in the language of statisticians, to be carefully "weighted."*

In a single paragraph, Galton raised the issue of non-independence, suggesting that a "common source" might give rise to similar bundles of traits across different societies, and proposed weighting observations as a potential solution. Since the original comment may be interpreted in terms of both spatial diffusion and inheritance from a common ancestor, the term "Galton's problem" usually refers to challenges involving both spatial and phylogenetic dependence in cross-cultural data. However, our primary focus throughout this paper is on the latter.

Despite its long history, it took about a century to develop formal statistical tools to address this issue. The first approach sought to reduce non-independence of observations by sampling societies that were not closely related, focusing on groups from distinct language families and geographic regions. The most well-known example of this strategy is the Standard Cross-Cultural Sample (SCCS) of Murdock and White (1969), a subsample of 186 societies selected from the EA to maximize cultural independence. One obvious drawback of this approach is the loss of information resulting from the radical reduction in sample size, even if some of the excluded data points are, in Galton's terms, "copies of the same original." More critically, since *all* human societies are related to some extent, subsampling alone cannot fully resolve the problem, and observed similarities across language clusters may still reflect shared ancestry (Eff, 2004; Mace and Zhang, 2023).

The second approach to addressing Galton's problem emerged in anthropology during the 1980s, drawing inspiration from spatial autoregressive (SAR) models developed earlier by geographers (Dow et al., 1984). In this framework, cultural non-independence was formalized as a network autocorrelation issue and modeled through "interaction" matrices. Unlike in cases of purely spatial correlation, the proposed matrices incorporated weights based on both geographic and linguistic proximity, with the latter capturing shared ancestry. These matrices were then used to define an autoregressive process in the error term and integrated into the estimation procedure following standard SAR models. Although this approach gained some traction in the literature, it was criticized for using arbitrary weights matrices not grounded in a well-specified evolutionary process (Rohlf, 2001), similar to how SAR models were criticized for lacking clear interpretation and structural foundation (Conley, 2010).

A more principled approach to Galton's problem was developed through close collaboration between anthropologists and evolutionary biologists. Given the process of biological evolution and speciation, the issue of non-independence due to common ancestry is particularly significant in cross-species analysis: the co-occurrence of multiple traits across taxa may reflect inheritance rather than a functional relationships between them. A watershed moment in evolutionary biology came with the seminal paper by Felsenstein (1985), who developed the method of phylogenetically independent contrasts to address this issue. His approach combined phylogenetic trees as formal representations of ancestral relationships among species with explicit models of trait evolution along the branches of such trees. This contribution spurred a rapid development of other phylogenetic comparative methods, including specialized regression techniques, which rely on trees and evolutionary models to account for interdependence of species in statistical analyses and infer patterns of trait coevolution (Harvey and Pagel, 1991; Harmon, 2019).

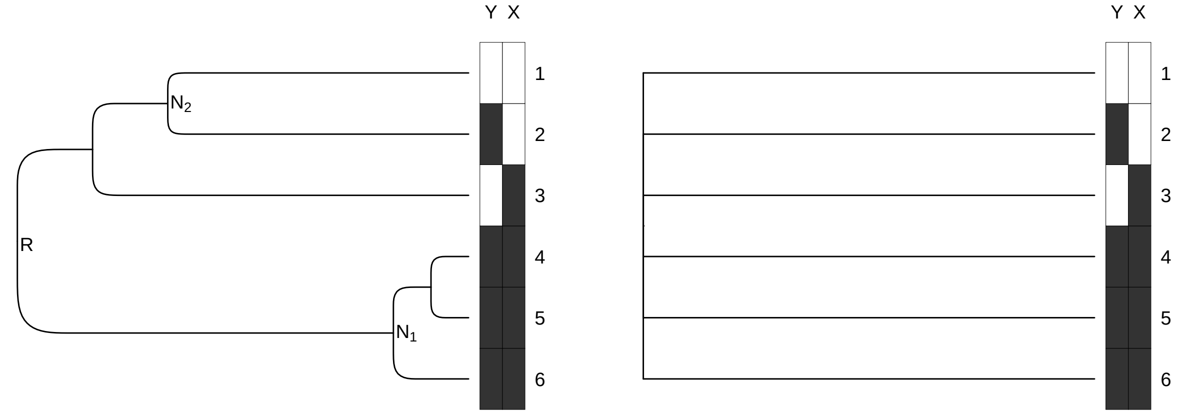
Drawing on the parallels between biological and cultural evolution, Mace and Pagel (1994) proposed the use of phylogenetic comparative methods in cross-cultural studies, sparking a new wave of research in quantitative anthropology (Gray and Watts, 2017; Mace and Zhang, 2023). This approach replaces phylogenetic trees – tracing relationships between biological organisms based primarily on molecular data – with linguistic trees that map ancestral relationships between human societies. Trait evolution is then modeled as occurring along the branches of these linguistic trees, ultimately producing the observed distribution of trait values at the tips representing contemporary societies.[1] This combination of tree structure and assumed evolutionary process can be used to estimate how similar trait values are expected to be across societies, providing information to account for phylogenetic non-independence in regression or other analyses.

Figure 1 illustrates the basic idea behind this approach. It displays the distribution of two binary traits, Y and X, in a sample of six societies labeled from 1 to 6 and arranged at the tips of two linguistic trees. In this example, both traits are present in societies 4, 5, and 6, absent in society 1, and only one trait is present in societies 2 and 3. While the sample size of six is too small to make reliable inference regarding the correlation between these traits, the raw data suggest a tendency toward co-occurrence, as seen in 4 out of 6 societies.

However, this tentative evidence looks different when considering the ancestral relationships between societies. Panel (a) of the figure presents a realistic scenario where societies exhibit varying degrees of relatedness, as reflected in the structure of the linguistic tree,

---

[1] See appendix B for a glossary of terms related to phylogenetic trees and used throughout the paper.



(a) Non-independence due to common ancestry (b) Complete independence

Figure 1: Trait distribution and ancestral relationships among six societies.

Notes: Societies are labeled from 1 to 6. Shaded cells indicate the presence of binary traits Y and X in respective societies. Panel (a) represents a realistic case, where observations are non-independent due to varying degrees of phylogenetic relatedness as captured by the linguistic tree. R marks the root of the tree, N_1 and N_2 are internal nodes corresponding to the MRCAs of societies 4, 5, 6 and 1, 2, respectively. Panel (b) represents an extreme case of complete phylogenetic independence, where ancestral relationships are captured by a star phylogeny.

with branch lengths representing the passage of time. In this case, societies 4, 5, and 6 are close relatives, having shared a common ancestor relatively recently at node N_1. Conversely, all three are only distant relatives of societies 1, 2, and 3, as their most recent common ancestor (MRCA) is located at the root of the tree, R. Societies 1 and 2, sharing their MRCA at node N_2, are more closely related than societies 1 and 6, but not as closely as 4 and 6, since node N_1 is much younger than N_2.

With this information on shared ancestry supplementing the raw data on trait distribution, Galton's problem becomes evident: the co-occurrence of traits Y and X in societies 4, 5, and 6 may well be due to the inheritance of this bundle from their MRCA at N_1 and its persistence, rather than any functional relationship between Y and X. The corresponding observations do not represent three independent bits of information. In contrast, societies 2 and 3, where Y and X do not co-occur, have evolved independently for a longer period and are less likely to share an ancestral trait bundle. From this perspective, the apparent support for the co-occurrence hypothesis in the raw data diminishes. As Galton suggested, observations should be reweighted to account for their non-independence, which is precisely what the phylogenetic regression approach accomplishes. Specifically, as explained in section 4, optimal weights are derived from a covariance matrix based on the relevant language tree and the assumed evolutionary process.

Panel (b) of figure 1 illustrates an extreme scenario in which observations are completely phylogenetically independent. While the trait distribution remains the same, the ancestral relationships between societies are represented by a star phylogeny: all lineages diverge at the root of the tree and evolve entirely independently throughout the evolutionary period. Because all societies are equally unrelated to one another, no weights are needed to account for shared ancestry, making the evidence in favor of the co-occurrence hypothesis stronger than in the case presented in panel (a). However, since all human societies are related to varying degrees, the tree depicted in panel (b) is not realistic.

Despite the many similarities between biological and cultural evolution (Mace and Holden, 2005), there are also important distinctions that may pose challenges for the application of PCMs in the social sciences. In particular, the assumed primacy of vertical cultural transmission, occurring along the branches of linguistic trees, has been widely debated in the literature. While horizontal transmission of traits is rare in biological evolution, it is much more common in the context of culture, though its relative importance likely varies by trait (Mace and Zhang, 2023). When horizontal transmission occurs across branches of a language tree, it no longer provides a fully accurate representation of all relevant historical relationships between societies (Nunn, 2011).

The extent to which horizontal cultural transmission undermines the validity of PCMs has been explored in several simulation studies.[2] Nunn et al. (2006) developed a model that allows for two modes of transmission, assuming that horizontal transfers applied to both modeled traits simultaneously. Their simulations revealed that, in this setting, PCMs exhibited elevated type I error rates, comparable to those in standard regression analysis. However, subsequent research argued that the assumption of paired horizontal transmission was overly restrictive, showing that PCMs are substantially more robust and outperform non-phylogenetic approaches in more realistic scenarios where traits may only be borrowed one at a time (Currie et al., 2010). The debate continues, with new methods being explored in both evolutionary biology and anthropology to simultaneously model vertical and horizontal transmission of traits in phylogenetic networks with reticulations (Gray et al., 2010; Bastide et al., 2018).

---

[2] A related issue is the impact of horizontal transmission on the accuracy of linguistic trees reconstructed from data such as language vocabularies. Simulations by Greenhill et al. (2009) have shown that Bayesian phylogenetic methods are robust to realistic levels of borrowing between languages when reconstructing tree topologies, although they are slightly more sensitive when estimating lineage divergence dates.

As the preceding discussion makes clear, the issue of non-independence due to shared ancestry has been central in anthropology, with various approaches proposed to address it. In contrast, economists working with cross-cultural datasets and facing similar statistical challenges have lagged behind, primarily focusing on spatial correlation. Apart from occasional use of the phylogenetic version of the SAR model (Gershman, 2015), the predominant approach to addressing phylogenetic non-independence has been to cluster standard errors in regression models. This method allows for an arbitrary form of interdependence among the errors terms within specific categories of societies. In practice, these categories are based either on spatial units, such as continents and countries, or on language groupings borrowed from classifications of societies available in the EA.

Although clustering standard errors is a step in the right direction, it does not fully address the issue, as we formally demonstrate through simulations in section 5. The fundamental problem, as discussed earlier, is that all human societies share common ancestry and essentially represent a single large cluster. The imposition of strict and somewhat arbitrary categories, such as continent, country, or language family, eliminates cross-cluster correlations – an implausible assumption, particularly if the resulting clusters are small in size. On the other hand, increasing cluster size would necessarily reduce the number of clusters, which poses a threat to statistical inference (Cameron and Miller, 2015). Moreover, cluster-robust inference becomes unreliable when cluster sizes are highly variable, especially in the presence of a few unusually large clusters (MacKinnon et al., 2023). This is particularly relevant when clustering by language family since the resulting groups vary dramatically in size, both in general and in specific datasets such as the EA, as shown in section 3.

Another issue with cluster-robust inference is that it does not leverage the correlation structure within clusters, even though it could potentially improve the efficiency of estimation. Generally, clustering only affects standard errors, while point estimates remain unchanged. In contrast, the phylogenetic regression approach utilizes information about the ancestral relationships between societies contained in the language tree. When combined with a specified evolutionary process, this information allows to estimate the covariance matrix of the error terms and obtain more precise coefficient estimates via phylogenetic generalized least squares.

3 Data

3.1 The *Ethnographic Atlas*

Although the challenge of accounting for shared ancestry is present in any cross-cultural setting, we limit our analysis to the *Ethnographic Atlas* (Murdock, 1967). The EA is a global ethnographic dataset that contains detailed information on more than 1,200 preindustrial societies worldwide. Its primary sources reflect fieldwork conducted mostly between the mid-19th and mid-20th centuries. We utilize the version of the EA available through D-PLACE, a publicly accessible database of cultural, linguistic, and environmental diversity (Kirby et al., 2016). The EA contains approximately 100 variables describing basic economic activities, social organization, institutional arrangements, and cultural practices and beliefs, typically coded on categorical or ordinal scales.

With the rise of empirical work on culture, institutions, and long-run economic development in the past two decades, the EA has become a key data source for economists working on these topics (Lowes, 2021). Some of this research uses the EA to explain variation in historical characteristics, often by linking them to exogenous environmental features. Other studies leverage EA data to explore the role of deep-rooted factors in explaining contemporary outcomes such as economic activity, conflict intensity, and cultural values. While the EA faced criticism for containing significant measurement error, recent analysis linking EA variables to analogous measures in contemporary societies has found strong positive correlations (Bahrami-Rad et al., 2021). This validation exercise supports the reliability of the EA as a source of ethnographic information.

Most empirical analyses use the EA variables "as is," though they are sometimes combined into composite indices to capture complex social traits such as kinship ties (Enke, 2019; Schulz et al., 2019) and gender inequality norms (Guarnieri and Tur-Prats, 2023). Scholars have also combined the EA with data on population distribution by ethnicity to construct country-level measures of ancestral characteristics of modern societies (Alesina et al., 2013; Giuliano and Nunn, 2018). While our analysis focuses on a subset of commonly used EA variables, the overall approach is readily applicable to any selection of traits and societies.

An obvious benefit of the EA is its size and global scale. However, a consequence of such extensive coverage is that many societies in this dataset are closely related, posing challenges for statistical inference. As mentioned in the previous section, in the economics literature employing the EA, the most common approach to account for non-independence due to shared ancestry is through clustered standard errors. The EA offers several "linguistic affiliation" categories – namely language continent, family, and subfamily – that have been used for this purpose. While a meaningful improvement, clustering does not fully address the underlying issue, and a more comprehensive accounting of ancestral relationships between societies can be achieved through a direct application of PCMs. To make such an analysis feasible, it is necessary to combine the EA with a global phylogeny of languages.

3.2 Global language trees

Building a tree representing ancestral relationships among all world languages is a highly challenging task that has long been a central goal of comparative linguistics. We utilize state-of-the-art estimates of global language supertrees recently produced by Bouckaert et al. (2022). The authors apply Bayesian phylogenetic inference techniques to generate global time-calibrated phylogenies for 6,635 extant languages. Their approach integrates multiple sources of information which fall into five main categories.[3]

First, the authors use the *Glottolog* 4.0 database (Hammarström et al., 2019) to classify all languages into primary families (or designate them as isolates) and obtain the rough structure of their relationships within families. The search of the tree space is then constrained only to phylogenies consistent with these established classifications. Although *Glottolog* provides a "skeleton" for further analysis, it does not specify the relationships between language families or isolates, nor does it offer a fully resolved and dated branching structure within families. Second, language location information, combined with a phylogeographic model, provides another input for establishing likely relationships between families and languages. Third, historical, linguistic, and archaeological evidence is used to date the breakup of major families and sub-clades. Fourth, the authors integrate results from previous Bayesian analyses that provide posterior distributions of dated phylogenies for nine language groups derived from lexical data.[4] Finally, genetic and archaeological evidence on human migration out of Africa is used to date the deepest relationships between languages.

Importantly, rather than providing a single estimate for a global tree of languages, Bouckaert et al. (2022) present a range of possibilities reflecting a substantial degree of uncertainty present in their model and priors. Specifically, the posterior sample from their Bayesian analysis includes 902 likely supertrees. Each of these trees is rooted, time-

---

[3] See the original paper for details on assumptions and modeling techniques.

[4] These include the Austronesian, Bantu, Dravidian, Indo-European, Pama-Nyungan, Semitic, Sino-Tibetan, Turkic, and Uralic families or subfamilies.

calibrated, ultrametric, and bifurcating (meaning that each internal node gives rise to exactly two descendants). Across this sample, the tree height (or crown age) ranges from 75,000 to 120,000 years. Most language diversification events are estimated to be relatively recent, with more than 95% of all language splits occurring within the last 5,000 years. The majority of extant languages are relatively young and 95% are estimated to be less than 3,039 years old. These patterns generally align with the profound impact of agricultural expansions during the Holocene on global linguistic diversity.

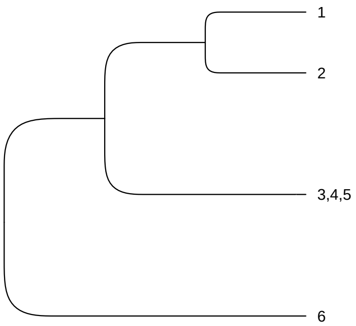
We account for phylogenetic uncertainty by conducting our analyses for each tree in the posterior sample, thereby directly exploring the robustness of our results to different tree estimates. Alternatively, often for illustration purposes, scholars rely on representative summary trees derived from the entire posterior sample of candidates. A common choice is the maximum clade credibility (MCC) summary tree which contains the most frequently occurring clades (Heled and Bouckaert, 2013).

Overall, this newly available dataset improves upon recent attempts to build global phylogenies of languages and societies (Jäger, 2018; Duda and Zrzavý, 2019) in several ways, including the integration of multiple relevant data sources, extensive coverage of languages, and direct accounting for phylogenetic uncertainty. Importantly, the authors highlight the suitability of their phylogenies for cross-cultural research employing PCMs (Bouckaert et al., 2022).

3.3 Matching EA societies to the global language trees

To begin using the EA data in conjunction with the global phylogenies described above, each society in the *Atlas* must be matched to its corresponding language. This mapping is facilitated by the fact that both D-PLACE and Bouckaert et al. (2022) use glottocodes – unique identifiers from the *Glottolog* database – to categorize each EA society and language, respectively.

During the matching process, we lost 87 of the 1,291 EA societies because they were matched to languages that are now extinct, and thus excluded by Bouckaert et al. (2022). Another 7 societies were dropped because they correspond to languages classified as colonial, creole, pidgin, or mixed and also excluded from the global phylogenies. A further 13 societies were discarded because they could not be reliably matched to any glottocode covered in the global supertrees. The remaining 1,184 cases include 954 one-to-one glottocode-based matches and 230 societies that speak different dialects of 80 unique languages. These



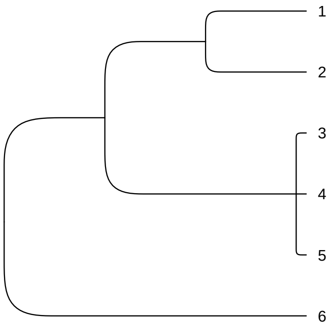


Figure 2: Modifying a language tree to accommodate dialects.

Notes: The left panel shows the original tree, where three societies (3, 4, and 5) that speak dialects of the same language are all matched to a single tree tip. The right panel displays the modified tree, with a new node added to represent the MRCA of the three dialects. This node forms a polytomy, branching into three separate tips, each corresponding to a different society on the modified language tree.

dialects are not assigned distinct glottocodes in the global language tree and are therefore represented by a single tree tip.[5]

The application of PCMs requires that each observation corresponds to a unique tip of the language tree. For the 230 relevant cases, rather than arbitrarily selecting one society for each group of related dialects, we preserved all EA observations by slightly modifying the global language trees. Specifically, for each set of dialects matching the same glottocode, we created a new internal node that branches into multiple dialects, each represented by a separate tree tip. Since there is no precise information on divergence times between dialects of the same language, the age of each new node (the hypothesized MRCA of the dialects) was determined as follows. First, for each tree in the available sample, we computed the minimum age among MRCAs for each pair of distinct languages. We then assumed that the age of the MRCAs for each group of dialects was equal to 10% of that value, reflecting the idea that dialects diverged much more recently than any two distinct languages. Cases with more than two dialects per glottocode resulted in polytomies, that is, nodes giving rise to more than two branches simultaneously.

Our approach is illustrated in figure 2. The left panel shows the initial matching of six societies to the tree tips. Societies 1, 2, and 6 are each matched to distinct languages, while societies 3, 4, and 5 speak dialects of the same language and are all matched to the

---

[5] Approximately half of these observations (108 in total) form pairs of dialects. Another 90 societies form groups of 3 to 9 dialects of the same language. The two outliers are Ute-Southern Paiute (glottocode `utes1238`), whose dialects are spoken by 15 EA societies, and Shoshoni (glottocode `shos1248`), whose dialects are spoken by 17 societies. Both belong to the North American branch of the Uto-Aztecan language family.

same tree trip. The right panel displays the modified tree, where an additional node is created to represent the MRCA of the three dialects spoken by societies 3, 4, and 5. This node forms a polytomy, with three branches leading to separate tips, each corresponding to a different society on the modified language tree.

Our final matched sample of EA societies includes 1,184 observations. Their corresponding languages belong to 126 primary families, based on *Glottolog* 4.0 classification, with 64 of these families represented by only a single society. The five largest language families in our sample are Atlantic-Congo (310 societies), Austronesian (130), Afro-Asiatic (101), Uto-Aztecan (59), and Indo-European (46). Additionally, 23 societies were matched to language isolates that do not belong to any family. We use this matched sample as an input for various analyses in the remainder of the paper.

As part of any preliminary investigation, the relevant society-level data can be visualized at the tips of a language tree. For example, figure 3 illustrates the distribution of two binary characteristics, state centralization and beliefs in moralizing high gods, in a sample of Austronesian societies from the EA, arranged at the tips of a pruned MCC tree. A simple visual inspection suggests that the distribution of both traits is not random with respect to the phylogeny. For instance, consider the clade of Polynesian societies located at the top of the tree. According to the MCC tree, all languages in this group are close relatives that share a MRCA between 1,000 and 2,000 years ago. Nearly all Polynesian societies, with one exception, exhibit the same bundle of traits: centralized states are present, but beliefs in moralizing high gods are absent.

More generally, the tendency of closely related societies to exhibit similar trait values is referred to as phylogenetic signal (Blomberg et al., 2003), which reflects the non-independence of observations due to shared ancestry. While a graphical representation of data on a language tree can be informative, the presence of phylogenetic signal should be assessed through formal statistical tests.

4 Phylogenetic signal in the *Ethnographic Atlas*

A variety of tests can be used to detect and quantify phylogenetic signal in observed traits (Münkemüller et al., 2012). Some of these tests, such as adaptations of Moran's I statistic (Gittleman and Kot, 1990; Abouheif, 1999), rely solely on the measurement of phylogenetic proximities based on tree estimates and are not benchmarked to any particular model of evolution. However, the most commonly used tests are conducted with reference to the strength of phylogenetic signal under Brownian motion (BM), the basic evolutionary model

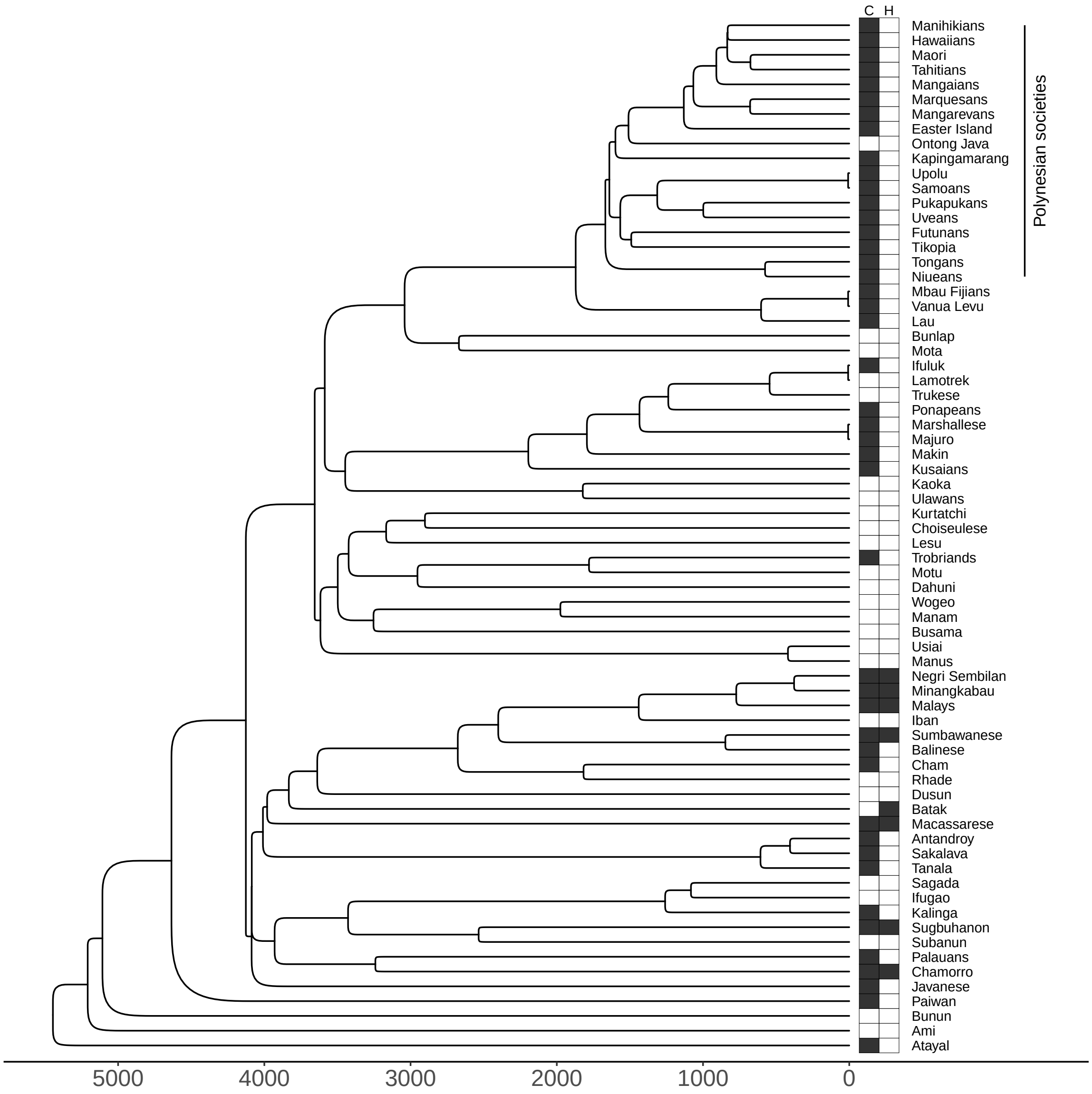


Figure 3: State centralization and moralizing high gods in Austronesia.

Notes: The tree is derived by extracting the portion corresponding to Austronesian societies in the EA from the global MCC tree of Bouckaert et al. (2022). Society names are listed at the tips. Shaded cells indicate the presence of state centralization (C), defined as the existence of one or more levels of jurisdictional hierarchy beyond the local community, and beliefs in moralizing high gods (H) in respective societies. The horizontal axis represents time measured in years before present.

widely employed across applications. Given the fundamental importance of BM in PCMs, we begin by briefly introducing this model and summarizing its main properties. We then describe two standard phylogenetic signal tests for continuous and binary traits and apply them to the EA data.

4.1 Brownian motion on a phylogenetic tree

The Brownian motion (BM) model of trait evolution is a continuous-time random walk process in which instantaneous changes in trait values are drawn from a normal distribution with mean 0 and variance σ^2 (Felsenstein, 1973). Under BM, variance accumulates linearly over time, so after t units of time since the start of an evolutionary process, the trait value is normally distributed with variance $\sigma^2 t$.

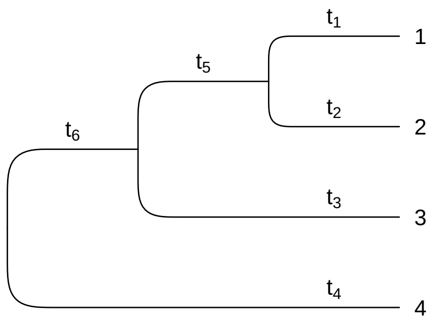
When modeling BM on a phylogenetic tree, one needs to account for occasional population splits that ultimately give rise to new societies. At the time of each split, the two daughter populations have identical trait values, but from that point forward, the trait evolves independently in each population according to the BM process. Consistent with its basic definition, BM on a phylogeny has the following characteristics: 1) trait changes along any given tree branch are independent of both previous evolutionary history and changes on any other branches; 2) larger changes are more likely on longer branches, with variance of change proportional to branch length; 3) the rate of evolution (σ^2) is constant over time, regardless of proximity to extant societies at the tips of the tree (Nunn, 2011).

More formally, consider an ultrametric phylogenetic tree of height t with n tips (societies) indexed by $i = 1, \ldots, n$. Denote the total number of branches connecting the tree root to tip i as T_i, and the length of each branch on this path as t_{ij}, $j = 1, \ldots, T_i$. If trait y evolves on this phylogeny according to the BM process with a constant evolutionary rate σ^2, then

$$y_i = \alpha + \sum_{j=1}^{T_i} \nu_{ij}.$$

Here, α is the (ancestral) state of trait y at the root and $\nu_{ij} \sim \mathcal{N}(0, \sigma^2 t_{ij})$ are independent value changes experienced at each tree branch. It follows, in particular, that $y_i \sim \mathcal{N}(\alpha, \sigma^2 t)$ since $\sum_{j=1}^{T_i} t_{ij} = t$ for all i in an ultrametric tree.

In a hypothetical case where all lineages diverge immediately at the tree root, meaning societies evolve completely independently throughout their entire history, the joint distribution of trait values is a multivariate normal such that $\boldsymbol{y} \sim \mathcal{N}(\alpha \cdot \boldsymbol{\imath}, \sigma^2 t \cdot \mathbf{I})$, where $\boldsymbol{y}$ is an $n \times 1$ vector of trait values across societies, $\boldsymbol{\imath}$ is an $n \times 1$ vector of ones, and $\mathbf{I}$ is an $n \times n$ identity matrix. In a more realistic scenario, where societies share common ancestry and their paths from the tree root to the tips partly overlap, the variance-covariance matrix of $\boldsymbol{y}$ is no longer diagonal. Indeed, if two societies share a most recent common ancestor at time $\bar{t}$, the covariance between their trait values is $\sigma^2 \bar{t}$, reflecting the length of their shared path on the phylogeny multiplied by the evolutionary rate (Freckleton et al., 2002). As a result, $\boldsymbol{y} \sim \mathcal{N}(\alpha \cdot \boldsymbol{\imath}, \sigma^2 \cdot \mathbf{V})$, where the phylogenetic variance-covariance matrix $\mathbf{V}$ has a



$$\mathbf{V} = \begin{pmatrix} t_4 & t_6 + t_5 & t_6 & 0 \\ t_6 + t_5 & t_4 & t_6 & 0 \\ t_6 & t_6 & t_4 & 0 \\ 0 & 0 & 0 & t_4 \end{pmatrix}$$

Figure 4: A simple tree and the corresponding phylogenetic covariance matrix under BM.

special structure: its diagonal elements are all equal to t, the length of the full path from root to tip for each society, while its off-diagonal elements are equal to the shared path lengths for corresponding society pairs (Harmon, 2019).

For illustration, consider the simple phylogenetic tree shown in figure 4, which represents ancestral relationships between four societies. Each branch of this tree is labeled with its estimated length. To compute the off-diagonal elements of the $\mathbf{V}$ matrix for this tree under the BM process, we find the lengths of the shared evolutionary paths between each pair of societies. In our example, societies 1 and 2 are the closest relatives and share the longest common path consisting of two branches with a total length of $t_6 + t_5$. Society 3 is more distantly related to 1 and 2, sharing only one branch of length t_6 with both. Society 4 evolved completely independently from the others, resulting in zero off-diagonal elements in the fourth row and column of $\mathbf{V}$. Since the tree is ultrametric, the diagonal elements of $\mathbf{V}$ are all equal to t_4, the total tree height. The covariance matrix $\mathbf{C}$ for the vector of trait values is then obtained by multiplying $\mathbf{V}$ by the evolutionary rate σ^2.

The core idea behind the BM process is that evolutionary change occurs due to numerous small, random forces. However, it is incorrect to equate BM with the neutral model of evolution. As shown in the evolutionary biology literature, the BM process aligns not only with the simplest case of pure genetic drift but also with various models of adaptive evolution that involve selection (Hansen and Martins, 1996; Harmon, 2019). The popularity of the BM model stems from its simplicity, convenient statistical properties, and its ability to adequately capture trait evolution across a broad range of scenarios, including adaptive evolution. Generally, the BM model is best viewed as describing the distribution of trait changes rather than any specific mechanism producing these changes (Nunn, 2011).

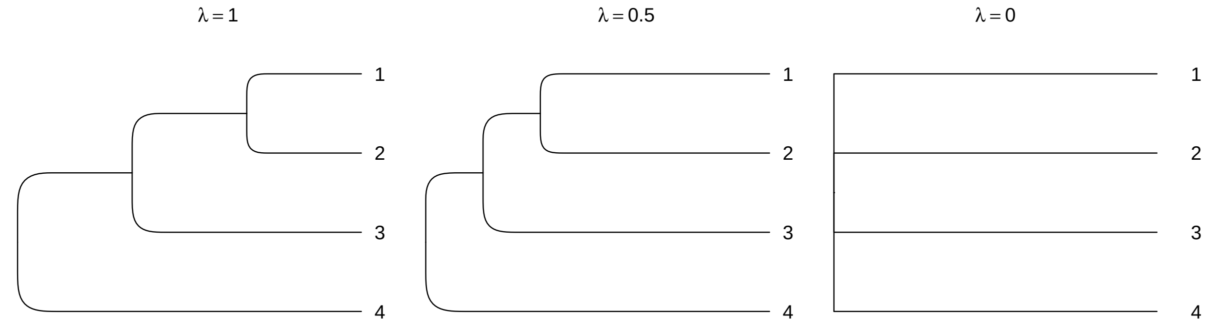


Figure 5: Pagel's λ and rescaled phylogenetic trees.

4.2 Phylogenetic signal tests: Pagel's λ and the D statistic

Pagel's λ is one of the most commonly used tests for detecting phylogenetic signal in continuous traits (Pagel, 1999). It measures the extent to which the observed distribution of trait values is consistent with the outcome of a BM evolutionary process on a given phylogeny. Formally, λ is defined as a multiplier of the off-diagonal elements of the phylogenetic covariance matrix $\mathbf{V}$, typically constrained between 0 and 1 (Freckleton et al., 2002). A value of $\lambda = 0$ corresponds to complete phylogenetic independence, while $\lambda = 1$ indicates an evolutionary process identical to BM. Intermediate values of λ reflect weaker phylogenetic signal compared to BM.

Pagel's λ can be interpreted as a scaling factor for a phylogenetic tree. Specifically, all internal branches are shortened by a factor of λ, while terminal branches are extended accordingly, as illustrated in figure 5 for the simple tree introduced earlier and shown again on the left panel. The middle panel corresponds to the case of $\lambda = 0.5$, where all internal branches are halved in length, shortening the shared evolutionary paths of societies 1, 2, and 3 and allowing for a longer period of independent evolution on terminal branches. This reduces phylogenetic dependence and increases the likelihood of greater divergence in trait values. The right panel illustrates the extreme case of $\lambda = 0$, where all branches become terminal, collapsing the original tree into a star phylogeny. In this scenario, societies evolve completely independently from each other, and phylogenetic signal is absent.

Under BM on a rescaled tree, the vector of trait values $\boldsymbol{y} \sim \mathcal{N}(\alpha \cdot \boldsymbol{\imath}, \sigma^2 \cdot \mathbf{V}(\lambda))$, where $\mathbf{V}(\lambda)$ is the original phylogenetic covariance matrix $\mathbf{V}$ with off-diagonal elements multiplied by λ. Given observed trait values at the tips and a phylogeny, λ can be estimated using a maximum likelihood approach (Freckleton et al., 2002). Standard likelihood ratio tests can then be applied to evaluate relevant hypotheses such as complete phylogenetic independence ($\lambda = 0$) or whether the signal strength is consistent with a BM process on the original, unscaled tree ($\lambda = 1$).[6]

Pagel's λ, based on a BM process for continuous traits, is not well-suited for evaluating phylogenetic signal in binary traits. In this case, the most commonly employed test is the D statistic of Fritz and Purvis (2010).

The D statistic is based on the standardized sum of differences in trait values between sister clades for a given phylogeny. The term "sister clades" refers to any two societies or groups of societies that diverge from the same tree node. For two sister societies, a simple difference in the respective tip values is calculated. In cases involving clades, nodal values are first estimated as the weighted means of descendant node or tip values, with weights inversely proportional to the branch lengths leading to them. The difference between these nodal values represents the sister-clade difference. Intuitively, the sum of all sister-clade differences in a tree will be smaller for traits that show strong phylogenetic dependence compared to traits that are more randomly distributed across the phylogeny. However, this sum depends on both the trait's prevalence and the size and shape of the phylogeny.

To make phylogenetic signal measures comparable across datasets, the sum of sister-clade differences is standardized using two reference points, namely the expected sums for phylogenetically random traits and for traits evolving under a BM threshold model. The former is estimated by repeatedly randomly shuffling observed trait values at the tree tips, whereas the latter is estimated by simulating continuous traits evolving under BM on the given phylogeny. For each simulation, a simple threshold model is applied, where societies with trait values above a chosen threshold are assigned a score of 1, and those below are assigned a score of 0. The threshold is selected to match the observed binary trait prevalence in the actual data. The sums of sister-clade differences are then calculated for the simulated binary traits, and their mean serves as the reference point under the BM threshold model.

The D statistic is calculated as $D = (S - \bar{S}_{BM})/(\bar{S}_R - \bar{S}_{BM})$, where S is the observed sum of sister-clade differences and $\bar{S}_R$ and $\bar{S}_{BM}$ are the reference averages from random reshuffling and BM threshold model simulations, respectively. With this standardization, $D = 1$ corresponds to phylogenetic independence, while $D = 0$ indicates that trait values are phylogenetically related exactly as predicted by the BM threshold model of evolution. Values between 0 and 1 reflect an intermediate level of phylogenetic signal below the BM benchmark. Unlike Pagel's λ, D values may fall outside the $[0, 1]$ interval, with negative

---

[6] Functions for estimating λ and testing these hypotheses in a maximum likelihood framework are available in `R` packages like `phylosignal` and `phylotools`.

values indicating stronger phylogenetic dependence than expected under BM, and values above 1 corresponding to overdispersion (when closely related societies show *greater* trait divergence that expected by chance). The statistical significance of the difference between the observed D value and the benchmark cases of 1 and 0 can be assessed via a randomization test by comparing D to the simulated reference distributions.[7]

4.3 Detecting phylogenetic signal in the ethnographic data

We conducted phylogenetic signal tests on a selected set of variables from the EA, focusing on characteristics with relatively few missing observations that have been commonly used in the economics literature. We categorized these variables into three broad groups: basic economy and social organization, institutions, and culture.[8] From a broader perspective, these features may all be considered as evolving cultural traits.

The first category includes four subsistence production variables: reliance on agriculture, reliance on gathering, the practice of intensive agriculture, and plow use. The social organization variables are class stratification and the degree of settlement complexity (sedentism). Following common empirical approaches, we treat reliance on agriculture, reliance on gathering, and settlement complexity as continuous variables, and the others as binary. The second category covers four societal characteristics often regarded as institutions in the economics literature: indigenous slavery, local democracy, partible land inheritance, and an index of state centralization. The latter is measured by the number of jurisdictional hierarchy levels beyond the local community and is treated as a continuous variable. The third category consists of four binary cultural traits, namely bride price, polygamy, belief in moralizing high gods, and matrilineal descent, along with a continuous kinship tightness index developed by Enke (2019).

To account for phylogenetic uncertainty, we performed signal tests across the 902 available global language trees, as described in section 3. Along with summary statistics for the tree sample, we also report results based on the MCC summary tree. For each variable, we use all non-missing observations and prune the language trees accordingly.

Our findings, detailed in table 1, indicate that all examined characteristics exhibit some degree of phylogenetic dependence. The mean D statistic values for binary traits range from 0.21 to 0.72. Formal randomization tests reject the null hypothesis of phylogenetic independence at the 1% significance level across all trees and variables, with a single

---

[7] A function for estimating D and testing relevant hypotheses is available in the `caper` package for `R`. The baseline number of permutations is 1,000, as recommended in Fritz and Purvis (2010).

[8] See appendix A for detailed definitions of all variables used in our analysis.

Table 1: Phylogenetic signal in the *Ethnographic Atlas*.

	D statistic				Pagel's λ				
Variable	Mean	St.dev.	IQR	MCC	Mean	St.dev.	IQR	MCC	N
Basic economy									
Intensive agriculture	0.42	0.01	0.01	0.41					1089
Plow use	0.21	0.01	0.01	0.19					1084
Class stratification	0.58	0.01	0.01	0.57					1013
Reliance on agriculture					0.95	0.01	0.01	0.94	1183
Reliance on gathering					0.94	0.01	0.02	0.92	1183
Settlement complexity					0.95	0.01	0.02	0.92	1088
Institutions									
Indigenous slavery	0.33	0.01	0.01	0.32					1027
Local democracy	0.72	0.01	0.01	0.71					853
Partible land inheritance	0.65	0.01	0.01	0.65					762
Jurisdictional hierarchy					0.93	0.02	0.02	0.87	1058
Culture									
Bride price	0.37	0.01	0.01	0.36					1168
Polygamy	0.41	0.01	0.01	0.40					1154
High gods	0.33	0.01	0.02	0.32					704
Matrilineal descent	0.45	0.01	0.02	0.42					1169
Kinship tightness					0.93	0.01	0.02	0.90	1184

Notes. The mean, standard deviation, and interquartile range (IQR) values are based on the estimates for 902 language trees. D statistic and Pagel's λ are computed for binary and continuous traits, respectively. MCC columns report estimates for the maximum clade credibility summary tree. In each test, the null hypothesis of phylogenetic independence is rejected at least at the 5% significance level.

exception.[9] For continuous traits, the mean values of Pagel's λ are all above 0.9, indicating a strong phylogenetic signal. Formal likelihood-ratio tests reject the null of phylogenetic independence at the 1% significance level for every language tree.

The variation in test statistics across trees is generally low, with both standard deviations and interquartile range values around 0.01 to 0.02. The signal measures for the MCC summary tree are similar to the averages across the 902 trees, with the largest discrepancy observed for the jurisdictional hierarchy variable. In this case, formal tests again reject the null of phylogenetic independence. We also tested signal strength relative to BM benchmarks, and the null hypothesis was rejected across the board, indicating a weaker signal than implied by the BM process.

---

[9] Specifically, for the plow use variable, this null hypothesis is rejected for 99% of the trees at the 5% significance level but only for 2% of the trees at the 1% significance level.

We tested the robustness of our results by sequentially excluding the three largest language families from our sample. While the test statistics varied somewhat in the reduced samples compared to the baseline, the discrepancies were not substantial, and our qualitative results remained unchanged.

Overall, our analysis reveals a phylogenetic signal across all considered economic, institutional, and cultural characteristics of societies in the EA rejecting the hypothesis of phylogenetic independence. In other words, societies that are more closely related through shared ancestry tend to have more similar trait values across multiple domains. This constitutes *prima facie* evidence of the non-independence of observations in the EA and motivates the use of PCMs in cross-cultural analyses using this dataset.

5 Phylogenetic dependence in regression analysis

5.1 The phylogenetic regression

The linear regression model continues to be a workhorse tool in many cross-cultural analyses that explore the relationships between two or more characteristics of interest. The standard textbook version of this model, in matrix notation, is as follows:

$$\boldsymbol{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon},$$

where $\boldsymbol{y}$ is an $n \times 1$ vector of values for the dependent variable, $\mathbf{X}$ is an $n \times k$ matrix containing the values of k independent variables, $\boldsymbol{\beta}$ is a $k \times 1$ vector of coefficients, and $\boldsymbol{\varepsilon}$ is an $n \times 1$ vector of error terms. Throughout the paper, we maintain the exogeneity assumption $\mathbb{E}[\boldsymbol{\varepsilon}|\mathbf{X}] = 0$. Although this assumption is likely to be violated in many cross-cultural estimation settings, our goal is to focus exclusively on the issue of phylogenetic dependence.

The simplest homoskedastic linear regression model further assumes that the covariance matrix of the error terms is $\mathbb{V}[\boldsymbol{\varepsilon}|\mathbf{X}] = \sigma^2\mathbf{I}$. This corresponds to an i.i.d. sampling framework, where the error terms are uncorrelated across observations and the off-diagonal elements of the covariance matrix are zero. In this setting, the Gauss-Markov theorem states that the OLS estimator $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{y}$ is efficient, meaning it has the smallest covariance matrix (in the positive semidefinite sense), equal to $\sigma^2(\mathbf{X}'\mathbf{X})^{-1}$, in the class of linear unbiased estimators.

When observations represent societies, such as ethnolinguistic groups, the i.i.d. assumption is violated in particular due to the shared ancestry of all human populations around the globe. In this case, the covariance matrix of $\boldsymbol{\varepsilon}$ is not diagonal, and while the OLS estimator remains unbiased, it is no longer efficient. Furthermore, the basic standard errors associated with $\hat{\boldsymbol{\beta}}$ are biased, resulting in erroneous statistical inference.

A regression analysis that accounts for phylogenetic dependence starts with a more general assumption that $\mathbb{V}[\boldsymbol{\varepsilon}|\mathbf{X}] = \mathbf{C}$, where the off-diagonal elements of $\mathbf{C}$ reflect the expected similarity between societies, given our knowledge of their evolutionary history. Economists working with cross-cultural data typically address this issue by calculating clustered standard errors, where clusters are defined based on spatial or language family classifications. While the OLS estimator remains inefficient in this case, hypothesis testing regarding $\boldsymbol{\beta}$ is valid if the assumed clustering structure accurately captures the true covariance matrix $\mathbf{C}$. Other limitations of this approach have been discussed in section 2.

Evolutionary biologists, who face a similar issue in the context of cross-species analyses, have developed an alternative approach that relies on specifying the $\mathbf{C}$ matrix based on the estimated evolutionary relationships between species, captured by phylogenetic trees and an assumed process of evolution. If $\mathbf{C}$ is correctly specified, the generalized least squares (GLS) estimator $\tilde{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{C}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{C}^{-1}\boldsymbol{y}$ is efficient, with the covariance matrix equal to $(\mathbf{X}'\mathbf{C}^{-1}\mathbf{X})^{-1}$. Since $\mathbf{C}$ is intended to represent phylogenetic dependence between observations, the corresponding linear model is called phylogenetic regression, and $\tilde{\boldsymbol{\beta}}$ is referred to as the phylogenetic generalized least squares (PGLS) estimator (Grafen, 1989; Martins and Hansen, 1997; Revell, 2010). Intuitively, whereas in a standard OLS regression each observation contributes equally to the estimation process, PGLS weights observations based on their phylogenetic relatedness (Symonds and Blomberg, 2014). Closely related species are downweighted as "pseudoreplicates," while those that are more phylogenetically distinct are upweighted because they provide more independent bits of information.

The "canonical" version of the phylogenetic regression model assumes that the error terms result from a BM process with zero mean and an evolutionary rate σ^2 on a given phylogenetic tree. Under this assumption, $\boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{C})$, where $\mathbf{C} = \sigma^2 \cdot \mathbf{V}$, and the phylogenetic covariance matrix $\mathbf{V}$ has a well-known structure. Specifically, as described in section 4, each off-diagonal element of $\mathbf{V}$ is equal to the length of the evolutionary path shared by respective pair of societies (the height of their MRCA), and for an ultrametric tree, each diagonal element is simply equal to the total tree height.[10] Thus, $\mathbf{V}$ is easily constructed from the given phylogeny, and, because $\mathbf{C} = \sigma^2 \cdot \mathbf{V}$, the PGLS estimator is expressed as $\tilde{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\boldsymbol{y}$, with $\mathbb{V}[\tilde{\boldsymbol{\beta}}|\mathbf{X}] = \sigma^2 \cdot (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}$. An unbiased

---

[10] In this case, the standard phylogenetic regression has homoskedastic error terms.

estimate for σ^2 can be obtained in the standard way using regression residuals: $\tilde{\sigma}^2 = (\boldsymbol{y} - \mathbf{X}\tilde{\boldsymbol{\beta}})'\mathbf{V}^{-1}(\boldsymbol{y} - \mathbf{X}\tilde{\boldsymbol{\beta}})/(n-k)$.

Although the PGLS estimator is unbiased for any choice of $\mathbf{C}$, its relative efficiency and the validity of hypothesis testing rely on the accurate specification of the covariance matrix. One potential concern is errors in the phylogeny. Reassuringly, simulations and analytic studies have shown that phylogenetic regression, and PCMs more generally, are fairly robust to local perturbations in tree specifications, including errors in both branch lengths and tree topology (Symonds, 2002; Martins and Housworth, 2002; Stone, 2011). More importantly, in our analysis, we use the entire sample of likely language phylogenies from Bouckaert et al. (2022) and directly explore the sensitivity of our results to tree estimates.

Another possible source of misspecification is the choice of the underlying evolutionary process. While BM is a versatile model consistent with various evolutionary scenarios, it is unlikely to precisely capture the past dynamics of $\boldsymbol{\varepsilon}$. However, if a strong case can be made for a specific evolutionary process, the corresponding $\mathbf{C}$ matrix can then be derived and used in the PGLS estimation framework. Indeed, implementations of phylogenetic regression under popular alternative models of evolution, including time-varying BM or the Ornstein-Uhlenbeck process, are readily available (Symonds and Blomberg, 2014; Revell and Harmon, 2022).

More commonly, instead of taking a strong stance on a specific evolutionary model, researchers employ a flexible setup that allows them to estimate one or more parameters of the evolutionary process from the available data. The most popular example of such an approach is to adjust the covariance matrix in the canonical BM case using Pagel's λ, which was introduced in section 4. Specifically, the off-diagonal elements of $\mathbf{V}$ are multiplied by a factor $\lambda \in [0, 1]$, which serves as an additional parameter of the model that captures the strength of phylogenetic signal in the error term. The coefficients of interest, $\boldsymbol{\beta}$, and Pagel's λ are then jointly estimated, and the appropriate level of phylogenetic correction is dictated by the data (Revell, 2010). This approach offers substantial flexibility by accommodating the extreme cases of complete phylogenetic independence ($\lambda = 0$) and canonical BM ($\lambda = 1$), as well as intermediate levels of phylogenetic signal.[11] Note that for $\lambda = 0$, the PGLS estimator coincides with basic homoskedastic OLS.

Simulation studies have shown that simple BM-based phylogenetic regression performs reasonably well, in terms of type I error rates, under a variety of alternative evolutionary

---

[11] It is the degree of phylogenetic signal in the error term, rather than in the dependent or independent variables, that ultimately justifies the application of phylogenetic regression (Revell, 2010).

models (Díaz-Uriarte and Garland, 1996; Martins et al., 2002). Even when deviations from the BM are severe, such as in complex heterogeneous models of evolution, simple PGLS with Pagel's λ significantly outperforms the standard OLS approach. Intuitively, while BM may not correspond precisely to the evolutionary process for $\boldsymbol{\varepsilon}$, it still captures the fundamental idea that societies that diverged from one another relatively recently are expected to be more similar, a property that is consistent with many alternative evolutionary models. Thus, even imperfect accounting for phylogenetic dependence via a BM-based model is better than none.

5.2 OLS vs. PGLS: A simulation study on language supertrees

The advantages of phylogenetic regression over standard OLS estimation methods in the presence of phylogenetic dependence have been well-established in the evolutionary biology literature (Nunn, 2011). We complement earlier studies by conducting a simulation exercise specific to cultural evolution on language supertrees for a sample of EA societies and comparing the performance of phylogenetic regression to common strategies in the economics literature.

Our setting is based on the simulation study by Revell (2010) that explores the implications of ignoring phylogenetic structure in regression analysis. The author considers a simple bivariate regression model and several types of data-generating processes that vary based on the presence of phylogenetic signal in the independent variable and the error term. To produce phylogenetically dependent data, he simulates a simple BM process on 1,000 stochastic pure-birth phylogenetic trees, all of the same height and each containing 100 tips. Consistent with basic theory, simulations show that, in scenarios with phylogenetic signal present in the error term, PGLS is more efficient than OLS and produces estimates that are closer to the true value of the slope coefficient in the vast majority of runs. Furthermore, OLS yields highly inflated type I error rates.

We focus on a model where both $\mathbf{X}$ and $\boldsymbol{\varepsilon}$ exhibit phylogenetic signal. Specifically, these variables are generated by simulating two independent BM processes with zero mean and $\sigma^2 = 1$ occurring on the global phylogenies of languages from Bouckaert et al. (2022). We use the entire sample of 902 trees and perform 100 simulations per tree, resulting in a total of 90,200 runs for each model specification. The trees are pruned to our baseline sample of 1,184 languages corresponding to EA societies, as described in section 3. The outcome variable is assumed to be independent of $\mathbf{X}$ and is generated as $\boldsymbol{y} = \mathbf{X} \cdot \mathbf{0} + \boldsymbol{\varepsilon}$. For each simulated dataset, we compare the performance of OLS and PGLS (with Pagel's

λ adjustment) in estimating the slope coefficient and testing the basic hypothesis that it is equal to zero at the 5% significance level.[12]

Unlike Revell (2010), we evaluate the properties of phylogenetic regression in relation not only to basic homoskedastic OLS but also to estimation strategies that incorporate clustered standard errors and various fixed effects. In both cases, we consider four types of groupings used in the literature to capture, directly or indirectly, the role of common ancestry. The first two categories are continent and country. EA societies in our sample fall into 148 modern countries across six continents (Africa, Asia, Europe, North America, Oceania, and South America). The other two categories reflect cultural affinity based on primary language families. In addition to the modern *Glottolog* classification, described in section 3, we also consider the language groupings available from the EA. Despite being outdated, this classification is commonly used by researchers due to its immediate availability in the EA. According to the original groupings in the EA, languages in our baseline sample correspond to 67 primary families and 41 isolates.[13] This is in contrast to 126 primary language families and 23 isolates according to *Glottolog*. In what follows, we refer to these two alternative groupings as MLF and GLF, which stand for *Glottolog* and Murdock language family classifications, respectively.

Table 2 summarizes the main simulation results. The first two columns of the table show the averages and variances of the coefficient estimates across simulations for different model specifications. As expected, although the distributions of OLS (panel A) and PGLS (panel B) estimates are both centered around the true value of 0, the variance is much smaller under PGLS. In the baseline bivariate case, the variance of the OLS estimates is more than 200 times greater than that of PGLS. Including fixed effects in the OLS estimation improves precision, with the lowest variance achieved in a specification that accounts for *Glottolog* language families. But even in this case it remains over 16 times greater than for the bivariate PGLS specification. As shown in panel B, fixed effects have only a minor impact on the dispersion of PGLS estimates.

Figure 6 illustrates the dramatic improvement in precision achieved by phylogenetic regression compared to standard OLS. Panel (a) shows that the distribution of estimates is the tightest under bivariate PGLS, even though the variance of OLS estimates decreases with the incorporation of fixed effects. Panel (b) shows that the simplest bivariate PGLS

---

[12] We used the following R packages in this analysis: `fixest` for OLS estimation, `phytools` for BM simulation, and `phylolm` for PGLS estimation. Following Revell (2010), we also assessed statistical power by simulating the model with the true slope coefficient set to 1. Both OLS and PGLS showed good power across scenarios, and we omit a detailed presentation of these results.

[13] The original language family coding in the EA contains over 60 missing values. We use up-to-date information to assign respective languages to one of the EA families or designate them as isolates.

Table 2: OLS vs. PGLS: Summary of simulation results

	Coefficient estimates		Type I error rate by clustering category				
Model specification	Mean	Variance	None	Continent	Country	MLF	GLF
A. OLS							
Bivariate	1.30×10^{-3}	1.80×10^{-1}	0.89	0.37	0.69	0.44	0.44
Continent FE	1.48×10^{-3}	1.18×10^{-1}	0.87	0.32	0.63	0.41	0.40
Country FE	1.27×10^{-3}	8.89×10^{-2}	0.83	0.47	0.28	0.39	0.41
MLF FE	1.20×10^{-3}	7.43×10^{-2}	0.82	0.42	0.45	0.32	0.31
GLF FE	1.82×10^{-4}	1.41×10^{-2}	0.60	0.24	0.30	0.21	0.23
B. PGLS							
Bivariate	-1.07×10^{-6}	8.51×10^{-4}	0.05				
Continent FE	2.43×10^{-6}	8.55×10^{-4}	0.05				
Country FE	2.30×10^{-4}	9.59×10^{-4}	0.05				
MLF FE	1.16×10^{-4}	9.25×10^{-4}	0.05				
GLF FE	1.22×10^{-4}	9.63×10^{-4}	0.05				

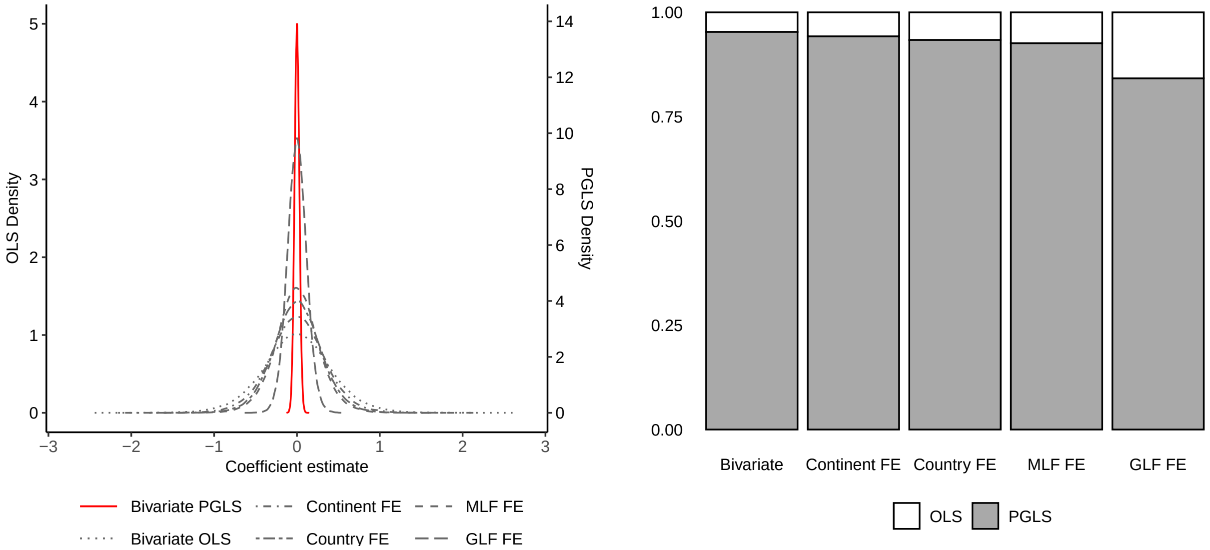
Notes: Panels A and B present results from OLS and PGLS estimation frameworks, respectively. MLF and GLF stand for Murdock and *Glottolog* language family classifications, respectively. Each reported statistic is based on 90,200 simulations of the corresponding model.

regression yields coefficient estimates that are closer to the true value of 0 in the vast majority of simulations. Specifically, PGLS is more precise than bivariate OLS in over 95% of simulations, while the best-performing OLS specification (with GLF fixed effects) is more precise in only about 16% of runs.

The right side of table 2 presents type I error rates associated with the null hypothesis that the slope coefficient is equal to zero. As expected, simple homoskedastic OLS yields highly inflated type I error rates, with nearly 90% of false positives in the bivariate regression case. While clustering standard errors helps mitigate this problem, the rates of false positives remain unacceptably high, with the lowest at 37% when clustering by continent. Adding most fixed effects to the homoskedastic OLS specification only marginally reduces the type I error rate, although accounting for *Glottolog* language families brings it down to 60%.[14] The table also presents the rates for various combinations of fixed effects and clustering categories. Overall, employing both strategies simultaneously further reduces false positives, with the best performance observed when clustering is used alongside GLF fixed effects. Still, even a 21% type I error rate is significantly higher than the nominal 5% level maintained across PGLS specifications. Figure C.1 in appendix C shows the distributions of p-values across simulations for a subset of estimation approaches.

---

[14] This result parallels the finding that, generally, regional fixed effects cannot properly account for spatial dependence, except under very special circumstances (Anselin and Arribas-Bel, 2013).



(a) Distributions of estimates

(b) Shares of runs with more precise estimates

Figure 6: Evaluating the relative precision of PGLS estimates.

Notes: Panel (a) displays kernel density plots for the distributions of coefficient estimates, based on 90,200 simulations per model specification. Density values for PGLS estimates are measured on a separate vertical axis for visual clarity. Panel (b) shows the fractions of simulations where the estimated coefficient was closer to its true value of zero under PGLS and OLS. Here, PGLS always refers to the bivariate specification, while column labels refer to respective OLS specifications.

Given that phylogenetic and spatial distance tend to be positively correlated, we also computed standard errors adjusted for spatial autocorrelation using the method of Conley (1999) for different distance cutoff values between 200 and 1,500 kilometers. Table C.1 in appendix C shows that while the Conley adjustment reduces the type I error rate – especially with larger distance cutoff values – it remains severely elevated compared to the nominal 5% rate. The specification with GLF fixed effects continues to have the best performance, yielding 23-24% of false positives under the largest cutoff values.

6 PGLS in action: Revisiting published results

6.1 Study sample and approach to reanalysis

Informed by the insights from our simulation exercise, we now turn to the application of phylogenetic regression in a reanalysis of cross-cultural patterns in real data. We focus on 10 papers, published in leading economics journals between 2013 and 2024, that use EA data to estimate society-level relationships. The research question is different in each of these papers and we do not intend to reevaluate any particular theory. Rather, our main goal is to see how the estimation results change in a phylogenetic regression framework and pinpoint any common patterns that emerge when accounting for non-independence of observations due to shared ancestry.

Importantly, we only look at a subset of estimates presented in each paper, namely those coming from OLS estimation of society-level regressions, which typically vary based on included fixed effects, other control variables, and approaches to computing standard errors. In some papers, these regressions represent the core empirical analysis, while in others they are merely "motivational" or represent the first line of evidence before turning to more complex analyses aimed at establishing causality. Our reanalysis is thus not comprehensive in many cases and should only be viewed with reference to basic OLS results.[15] We do not have strong priors regarding the impact of our reanalysis, and PGLS may in theory reveal patterns that are weaker, stronger, or unchanged relative to baseline OLS.

Overall, our analysis covers 28 regression models, 12 of which are based on global samples of EA societies, while the remaining 16 focus solely on Africa. In each case, we rely on publicly available replication data and ensure the consistency of original findings by replicating baseline OLS estimates before comparing them to PGLS results. Table 3 provides a summary of each model specification in our reanalysis. For easy identification of the original results, we reference the corresponding table and column numbers from published papers. We next list the dependent and main independent variables and provide the number of controls used, not counting any spatial or language family fixed effects.[16] The number and type of such controls differ significantly across models. In a substantial number of cases, dependent variables are binary in nature. We follow the original studies and estimate linear probability models for consistency and ease of comparison.[17] It should be noted, however, that specialized methods designed for modeling binary outcomes, such as phylogenetic logistic regression, are also readily available (Ives and Garland, 2010; Gershman and Mumladze, 2024).[18]

In a separate column, we report categories of fixed effects employed by the authors to account for unobserved heterogeneity. Many studies employ basic spatial units, such

---

[15] Incorporating phylogenetic adjustment in more sophisticated settings such as instrumental-variables regressions is a fruitful direction for future research.

[16] See appendix A for detailed definitions of all variables in table 3.

[17] One exception is the second model from Fenske (2014) which is originally estimated using ordered probit. We proceed with the OLS estimation after confirming that it yields a qualitatively similar result.

[18] Simulations have shown that PGLS performs reasonably well with binary response variables in large samples (Mundry, 2014).

as continent and country, though precise classifications vary slightly across papers. To capture cultural and linguistic affinity, some studies incorporate fixed effects representing language groupings from the EA. These include, in order or increasing granularity, language continents, families, subfamilies, and cultural provinces. We also report the authors' approaches to computing standard errors, which include heteroskedasticity-robust estimation, clustering by geographic or language group, and the Conley (1999) correction for spatial correlation.

The remaining columns of table 3 describe the baseline result, in terms of sign and statistical significance of the main coefficient, and provide the sizes of both the original sample and the subsample used in our reanalysis. The loss of observations results from an imperfect matching of EA societies to language trees, as explained in section 3. We typically lose 5% to 10% of observations in global samples and fewer than 5% in specifications focusing on Africa. In each case, we verify that the baseline OLS result remains intact in the reduced sample, that is, the coefficient sign and statistical (in)significance are preserved, and its magnitude is within 20% of the original estimate. Baseline OLS results are all statistically significant at least at the 10% level, with 3 exceptions.

Similar to our simulation exercise, we examine a range of model specifications and clustering strategies for each regression. We start with a simple bivariate case followed by the "baseline" specification incorporating all control variables (not counting fixed effects), which differ by study and sometimes also by model within a study. Next, we introduce spatial fixed effects represented by continents for global samples and countries for Africa. Finally, we estimate specifications with language family fixed effects using the MLF classification.[19] Our definitions for all fixed effects are identical to those used in section 5.

Each model specification is estimated using OLS and PGLS. In the OLS case, we present point estimates and statistical significance levels based on three types of standard errors: heteroskedasticity-robust, clustered by spatial unit and by language family. Appendix D additionally reports significance levels for standard errors adjusted for spatial correlation using different distance cutoffs. For each model from table 3, at least one set of reported estimation results is identical or very close to those in the original study.[20] When applying phylogenetic regression, we produce PGLS estimates for each of the available 902 global language trees. We then report the interquartile range (IQR) of PGLS estimates

---

[19] Some studies using the language family classification from the EA treat all language isolates as belonging to one category. Instead, we treat them as singleton groups, a more conceptually justified approach. Estimation results with GLF fixed effects are similar to those obtained using MLF classification.

[20] For example, while we do not report significance levels associated with clustering by cultural province (Alsan, 2015), using MLF as clustering category yields very similar results.

Table 3: Summary of models used in the reanalysis

Model	Table (column)	Dependent variable	Independ. variable	Num. of controls	Fixed effects	Standard errors	OLS estimate	Sample size	Subsample size
Global samples									
Alesina et al. (2013)									
1	1(1)	Female part. in agricult.	Plow use	5	None	Conley	Positive 1%	660	625
Becker (2024)									
1	1(2)	Patrilocal residence	Depend. on pastoralism	0	Continent	Robust	Positive 1%	1167	1072
2	1(4)	Insist. on virginity	—"—	0	—"—	—"—	Positive 5%	592	538
3	A1(1)	Women as sexual	—"—	2	—"—	—"—	Negative > 10%	1124	1034
Bentzen and Gokmen (2023)									
1	1A(6)	High gods	Class stratif.	4	Language family	Language family	Positive 1%	666	600
2	1B(6)	High gods	Irrigation potential	4	—"—	—"—	Positive 10%	560	529
Enke (2019)									
1	3-3	Kinship tightness	Malaria ecology	2	Continent	Language subfamily	Positive 5%	1216	1107
2	4(4-5)	High gods	Kinship tightness	3	Continent/ country	—"—	Negative 1%	770	694
3	4(9)	Purity	—"—	2	Continent	—"—	Positive 5%	371	335
4	4(14-15)	Jurisd. hierarchy	—"—	2	Continent/ country	—"—	Positive 1%	1141	1039
Fenske (2013)									
1	2(8)	Log pop. density	Land quality	11	Language continent	Conley	Positive 1%	1205	1105
Mayshar et al. (2022)									
1	1(1)	Jurisd. hierarchy	Reliance on cereals	0	Continent	Country/ Conley	Positive 1%	952	870
African samples									
Alsan (2015)									
1	1(4)	Large animals	TseTse suitability	11	None	Cultural province	Negative 1%	484	477
2	1(4)	Intensive agriculture	—"—	11	—"—	—"—	Negative 1%	485	478
3	1(4)	Log pop. density	—"—	11	—"—	—"—	Negative 1%	398	390
4	1(4)	Indigenous slavery	—"—	11	—"—	—"—	Positive 5%	446	440
5	1(4)	State central.	—"—	11	—"—	—"—	Negative 5%	467	460

Model	Table (column)	Dependent variable	Independ. variable	Num. of controls	Fixed effects	Standard errors	OLS estimate	Sample size	Subsample size
Depetris-Chauvin and Weil (2018)									
1	2(6)	Log pop. density	Malaria burden	9	Cultural province	Robust	Positive 1%	398	390
2	3(3)	State central.	—"—	9	—"—	—"—	Positive > 10%	520	512
3	3(4)	Intensive agriculture	—"—	9	—"—	—"—	Negative > 10%	520	512
Enke (2019)									
1	3(10)	Kinship tightness	Malaria ecology	3	None	Language subfamily	Positive 5%	500	494
2	3(10)	Kinship tightness	TseTse suitability	3	—"—	—"—	Positive 1%	500	494
Fenske (2013)									
1	2(6)	Indigenous slavery	Land quality	11	Language continent	Conley	Positive 1%	416	409
2	2(9)	Log pop. density	—"—	11	—"—	—"—	Positive 10%	486	477
Fenske (2014)									
1	3(1)	Jurisd. hierarchy	Ecological diversity	16	None	UN region	Positive 5%	440	432
2	5(2)	Class stratif.	—"—	16	—"—	—"—	Positive 1%	364	358
Moscona et al. (2020)									
1	2A(3)	Conflict incidents	Segmentary lineage	11	Country	Robust	Positive 1%	141	137
2	A4A(3)	Conflict incidents	—"—	11	—"—	—"—	Positive 5%	458	433

Notes: The first column numbers the models within each paper. This numbering is used in figures 7, 8, 10, and 11. In the column reporting the type of standard errors in original studies, "Robust" refers to heteroskedasticity-robust estimates, "Conley" refers to estimates adjusted for spatial correlation (Conley, 1999), and other items correspond to clustering categories (country, language family, language subfamily, cultural province, UN region). The last two columns report the sample size in the original study and in the reanalysis, respectively. The difference in sample sizes is due to the loss of observations in the process of matching EA societies to global language trees. The ditto mark (—"—) means repetition of the previous value in the same column.

for each specification, directly accounting for phylogenetic uncertainty in our analysis. We also provide the corresponding IQRs for estimated strength of phylogenetic signal in the error terms (Pagel's λ). All coefficient estimates are standardized for visual clarity and comparability across models.

6.2 Summary of results

Figure 7 presents a summary of point estimates from models using global samples of EA societies. For each specification, the OLS estimate is marked with an × symbol, the IQR of PGLS estimates is represented by a black box plot, and the IQR for λ is depicted as a solid red segment. Figure 8 provides a summary of corresponding statistical significance levels, which are categorized into four groups: 1%, 5%, 10%, and greater than 10%. For OLS estimates, significance levels are reported for three versions of standard errors described above using markers of different shapes. For phylogenetic regressions, we show the histograms of p-values corresponding to PGLS estimates within the reported IQRs and based on the four significance-level bins.

A common finding across models is the presence of substantial phylogenetic signal in regression residuals. The estimates of λ are generally higher in the bivariate and baseline specifications, but often remain substantial even after including fixed effects. Accounting for language family reduces residual phylogenetic signal the most and sometimes, but not always, eliminates it altogether.

PGLS typically produces more conservative coefficient estimates that are smaller in magnitude relative to OLS. As illustrated in panel (a) of figure 9, the differences in coefficient magnitudes is positively correlated with the strength of estimated phylogenetic signal. In these scatterplots, the vertical axis represents the percentage change in the OLS coefficient relative to the median PGLS estimate, the horizontal axis shows the corresponding values of λ, and each point corresponds to one model from table 3. In bivariate and baseline specifications, characterized by the highest values of λ, PGLS coefficient estimates are typically 60-80% smaller than their OLS counterparts, and sometimes the difference is even larger.

The gap between OLS and PGLS estimates is reduced with the introduction of fixed effects. In about half of the models with continent fixed effects, PGLS coefficients are 30-50% smaller than OLS estimates, although reductions of 80% or higher are still observed in a substantial number of cases. In one model, we identify no phylogenetic signal, and both estimation approaches yield the same coefficient estimate. When language family fixed effects are controlled for, phylogenetic signal weakens notably across the board and is eliminated in 5 out of 12 models, leading to convergence between OLS and PGLS. In the remaining models, we keep observing moderate to substantial change in the coefficient estimates.

An important observation is that across different specifications for the same model, PGLS coefficient estimates remain very stable. In contrast, OLS estimates are much more

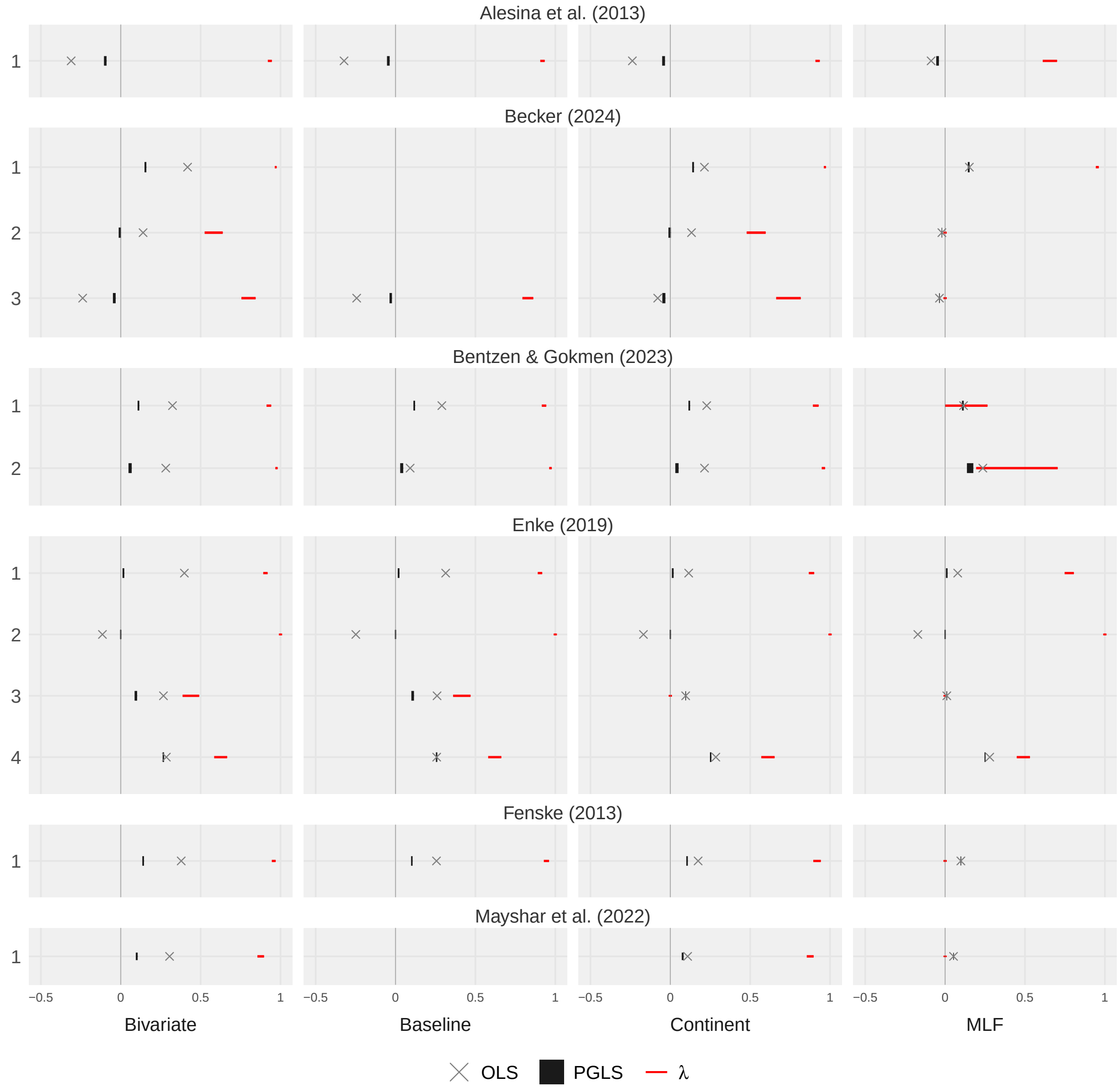


Figure 7: OLS and PGLS point estimates and phylogenetic signal in global samples.

Notes: Model numbers on the left correspond to those in column 1 of table 3. OLS point estimates are marked with an × symbol, the IQR of PGLS estimates is represented by a black box plot, and the IQR for λ estimates is depicted as a solid red segment. Baseline specification includes all controls from the original studies (excluding fixed effects). Continent and MLF specifications add respective fixed effects to the baseline specification. All coefficient estimates are standardized. In three cases, namely models 1 and 2 from Becker (2024) and model 1 from Mayshar et al. (2022), baseline specifications are identical to bivariate cases and are omitted.

sensitive to the set of included controls. Furthermore, PGLS estimates are quite robust to phylogenetic uncertainty, as evident from consistently narrow interquartile ranges of coefficients. This remains true even in models with notable variability in estimated residual

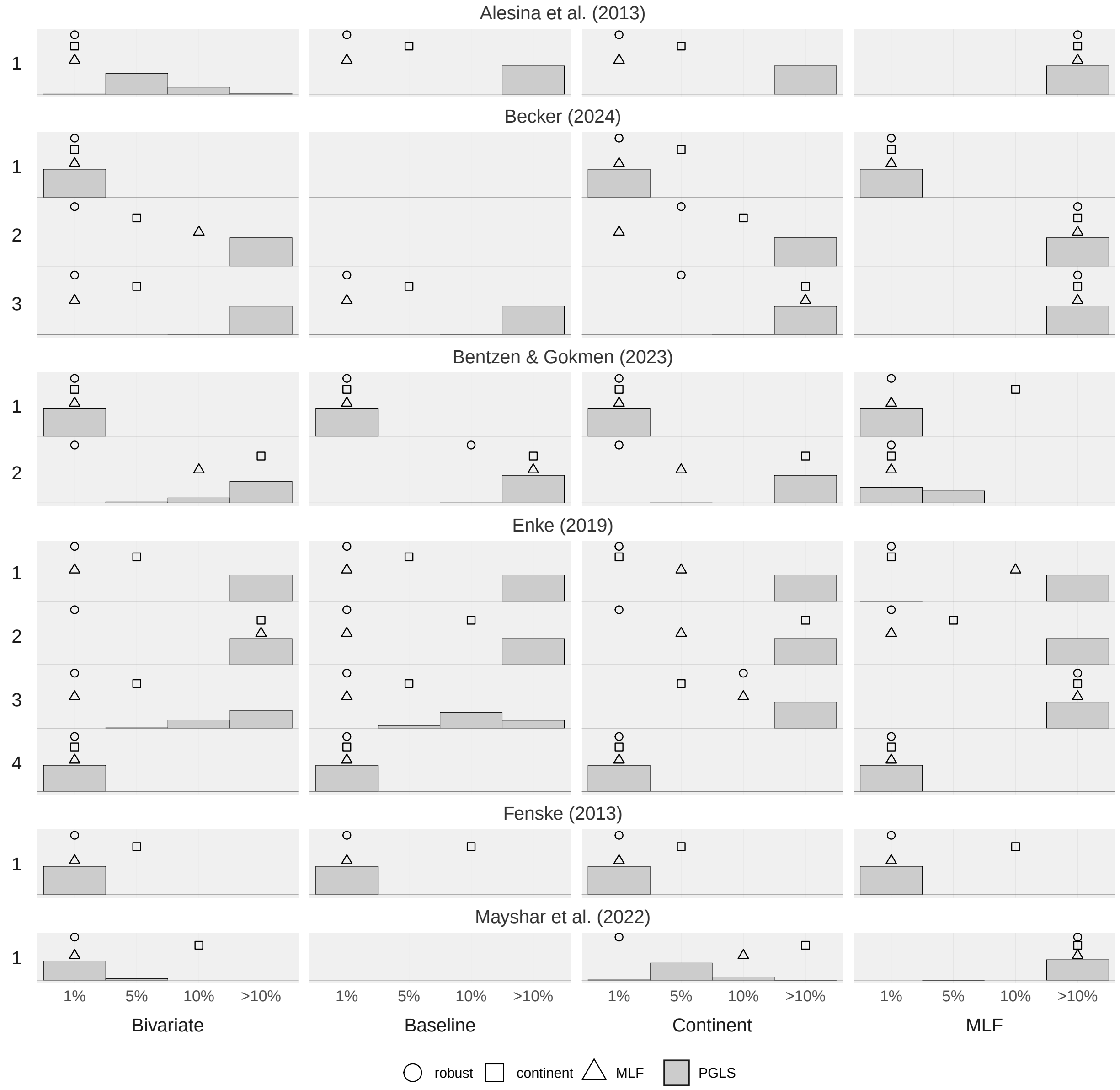
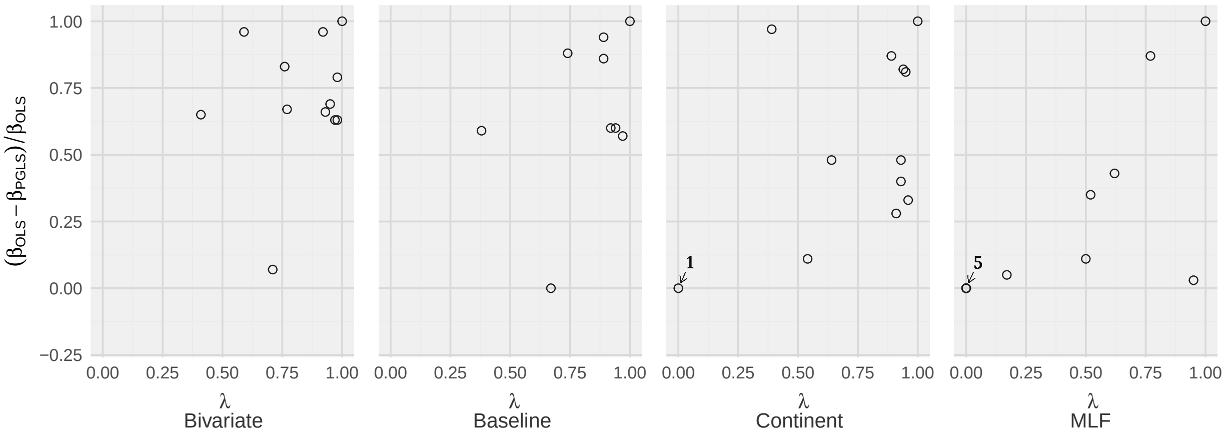


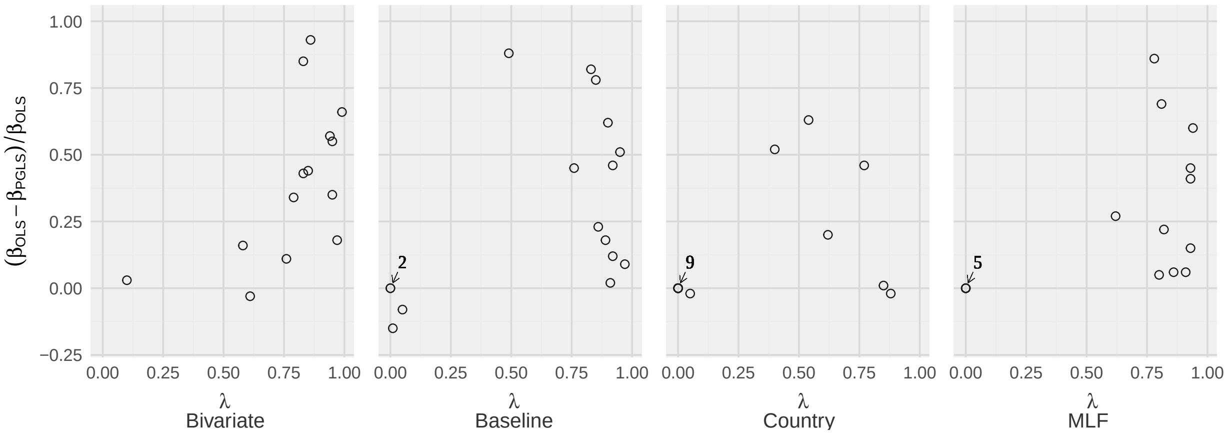
Figure 8: Statistical significance of OLS and PGLS estimates in global samples.

Notes: Levels of statistical significance are grouped into four categories: 1%, 5%, 10%, and greater than 10%. For OLS estimates, significance levels are shown based on three types of standard errors: heteroskedasticity-robust (circles), clustered by continent (squares) and MLF (triangles). For PGLS, the figure shows histograms of *p*-values corresponding to estimates within the IQRs and based on the four significance-level bins. Baseline specification includes all controls from the original studies (excluding fixed effects). Continent and MLF specifications add respective fixed effects to the baseline specification. In three cases, namely models 1 and 2 from Becker (2024) and model 1 from Mayshar et al. (2022), baseline specifications are identical to bivariate cases and are omitted.

phylogenetic signal. In other words, the differences in global language trees do not convert into substantial variation in PGLS coefficient estimates.



(a) Global samples



(b) African samples

Figure 9: Residual phylogenetic signal and the change in coefficient magnitudes.

Notes: Each panel plots the percentage difference between the OLS coefficient and the median PGLS estimate against the corresponding value of residual phylogenetic signal. Each point corresponds to one model specification. Negative values indicate cases where PGLS coefficients are larger in magnitude than OLS. Two outlier cases, in which OLS and PGLS coefficients have different signs, are omitted. Labeled points reflect the number of models in which OLS and PGLS estimates are identical.

Figure 8 shows that OLS estimates tend to be more statistically significant than their PGLS counterparts, regardless of the approach to standard errors. Clustering and inclusion of fixed effects typically reduce the significance of OLS estimates. It appears to be particularly sensitive to clustering standard errors by continent and controlling for language family categories. Similar to the case of point estimates, statistical significance of PGLS coefficients is lower and shows more stability across models compared to OLS. Already in the bivariate and baseline specifications, about a half of the estimates become statistically insignificant under PGLS estimation. The OLS and PGLS results are most closely, but not perfectly aligned in specifications with language family fixed effects. Additionally, the rather consistent distribution of p-values across significance-level bins shows once again the robustness of PGLS estimation to phylogenetic uncertainty.

Figure D.1 in appendix D compares the statistical significance of OLS estimates when standard errors are adjusted for spatial correlation using different distance cutoff values to PGLS results. We find that this approach yields p-values that are qualitatively similar to clustering strategies and frequently different from PGLS-based inference.

Next we revisit the studies focusing on African societies. The setting and visualization approach is the same as for the global-sample reanalysis above, but following the literature, we use country rather than continent as a clustering and fixed effect category. The results are summarized in figures 10, 11, and panel (b) of figure 9.[21]

Once again, we detect substantial phylogenetic signal in most models, which is persistent across specifications. The differences in OLS and PGLS coefficients are typically smaller than in the global sample. In roughly one third of bivariate and baseline specifications, the reduction in coefficient magnitude is between 30% and 60%. Moderate decreases of 10%-20% are also common, while in some cases the drop in magnitude reaches 80% or higher.

As in the global sample, inclusion of fixed effects reduces both the measured residual phylogenetic signal and the gap in coefficient estimates, with country fixed effects being especially influential. After controlling for language family, the change in the coefficient magnitude is less than 50% in the vast majority of cases, with roughly half of all models showing little to no change. In specifications with country fixed effects, the difference between OLS and PGLS estimates is very small in most cases and non-existent in about half of all models. Consistent with earlier findings, PGLS coefficients remain stable across specifications, with narrow interquartile ranges.

The patterns of statistical significance also resemble earlier findings, although the differences between OLS and PGLS are somewhat less pronounced. As before, clustering and fixed effects reduce the statistical significance of OLS coefficients, but the results still do not fully align with PGLS. Figure D.2 in appendix D shows that adjusting standard errors for spatial correlation yields conclusions that are qualitatively similar to those obtained from clustering approaches.

Overall, our reanalysis demonstrates that accounting for non-independence of observations in a phylogenetic regression setting often leads to a substantial reduction in both

---

[21] In several specifications, the PGLS estimation routine did not converge for a small subset of trees not exceeding 6% of the entire sample. In these cases, all reported results are based on the estimates obtained from the remaining set of trees.

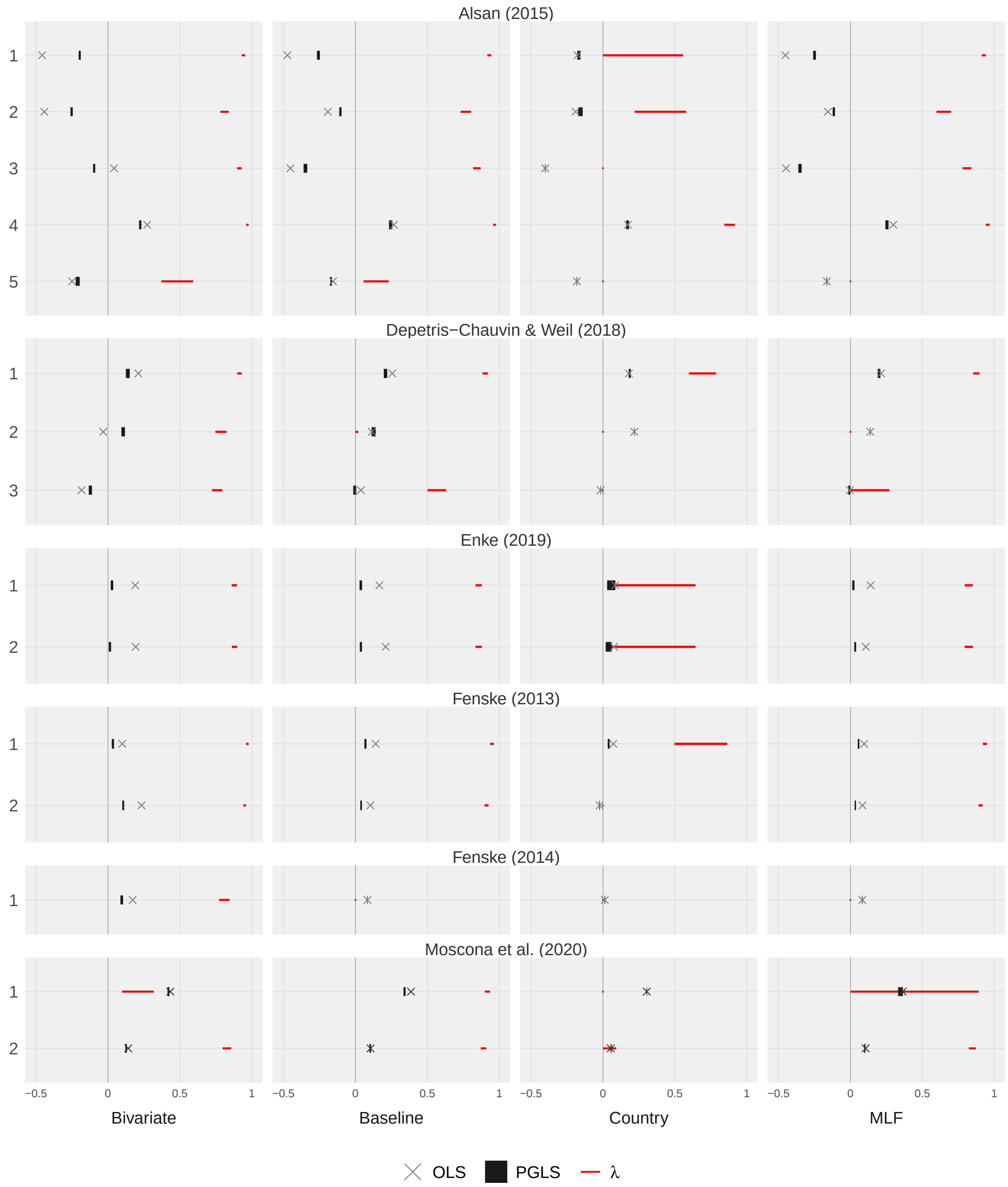


Figure 10: OLS and PGLS point estimates and phylogenetic signal in African samples.

Notes: Model numbers on the left correspond to those in column 1 of table 3. OLS point estimates are marked with an × symbol, the IQR of PGLS estimates is represented by a black box plot, and the IQR for λ estimates is depicted as a solid red segment. Baseline specification includes all controls from the original studies (excluding fixed effects). Country and MLF specifications add respective fixed effects to the baseline specification. All coefficient estimates are standardized.

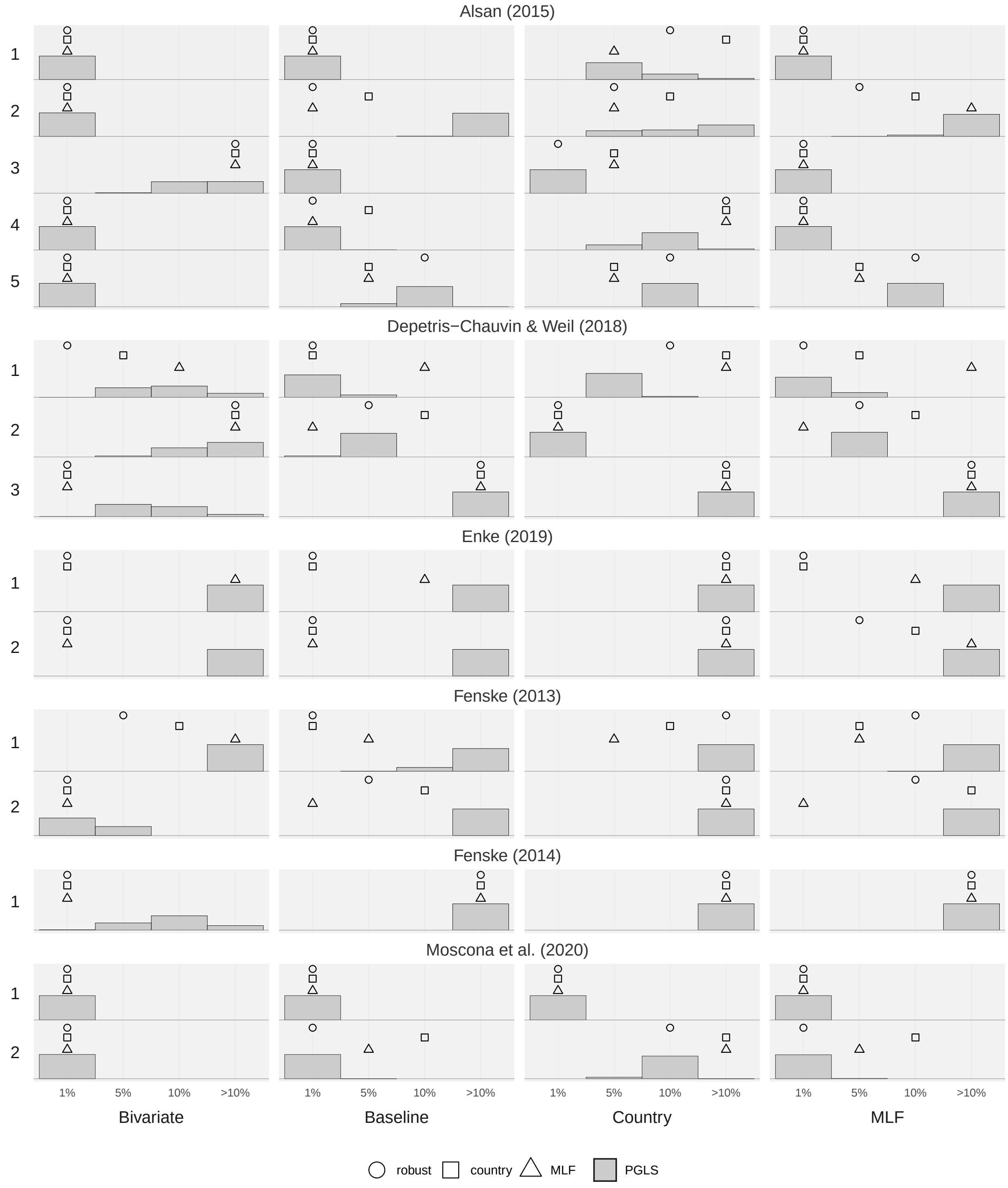


Figure 11: Statistical significance of OLS and PGLS estimates in African samples.

Notes: Levels of statistical significance are grouped into four categories: 1%, 5%, 10%, and greater than 10%. For OLS estimates, significance levels are shown based on three types of standard errors: heteroskedasticity-robust (circles), clustered by continent (squares) and MLF (triangles). For PGLS, the figure shows histograms of p-values corresponding to estimates within the IQRs and based on the four significance-level bins. Baseline specification includes all controls from the original studies (excluding fixed effects). Country and MLF specifications add respective fixed effects to the baseline specification.

estimated coefficient magnitudes and their statistical significance. Furthermore, PGLS estimates are stable across model specifications and robust to phylogenetic uncertainty.

7 Concluding remarks

This paper argues that the growing empirical literature in economics, which uses ethnic-group-level datasets to examine culture, institutions, and long-term development, has not adequately addressed the issue of non-independence of observations due to shared ancestry. We propose the adoption of phylogenetic comparative methods, a suite of statistical techniques designed to deal with this problem. We demonstrate how newly available global phylogenies of languages facilitate the application of PCMs in large cross-cultural datasets, with a focus on the widely used *Ethnographic Atlas*.

Our study reports three key sets of findings. First, phylogenetic signal tests reveal that various economic, institutional, and cultural characteristics are more similar among societies with closer ancestral ties, highlighting the non-independence of observations in the EA. Second, our simulation exercise shows that conventional approaches such as using clustered standard errors or introducing fixed effects as additional control variables fall short of fully addressing the problem of phylogenetic correlation. Third, a reanalysis of previously published results in a phylogenetic regression setting finds that the resulting estimates often have lower magnitude and weaker statistical significance.

Our matched dataset of EA societies and existing statistical software make phylogenetic regression straightforward to apply, and we encourage researchers to include this robustness check in any cross-cultural analysis. More broadly, adjustment for phylogenetic non-independence should become standard practice, similar to how spatial correlation is now routinely accounted for.

This research opens several avenues for future work incorporating PCMs in empirical economic analysis. In an ongoing project, we explore how phylogenetic regression can be adapted for different units of observation – such as individuals, ethnic homeland partitions, or grid cells – commonly used in studies leveraging ethnicity-level variation. Another important research direction involves developing methods that account for both phylogenetic and spatial dependence, helping to disentangle their distinct effects. Finally, economists could benefit from other PCMs, including techniques for modeling trait co-evolution on phylogenetic trees.

Appendices

A Definitions of variables

This section defines the variables analyzed throughout the paper. We first list the variables from sections 3 and 4, followed by those from section 6. Unless otherwise specified, all variables come from the version of the *Ethnographic Atlas* available through D-PLACE (Kirby et al., 2016) at `https://d-place.org/contributions/EA`. Original D-PLACE variables are mentioned below with an EA prefix.

Basic economy and social organization

Intensive agriculture. An indicator variable equal to 1, if a society practices: 1) intensive agriculture on permanent fields, utilizing fertilization, crop rotation, or other techniques so that fallowing is either unnecessary or is confined to relatively short periods, or 2) intensive cultivation largely dependent upon irrigation. Corresponds to the original variable EA028 taking values 5 or 6.

Plow use. An indicator variable equal to 1, if animals are used in plow cultivation, whether this practice is aboriginal or not. Corresponds to the original variable EA039 taking values 2 or 3.

Class stratification. An indicator variable equal to 1 for any type of significant wealth distinctions and class differentiation. Corresponds to the original variable EA066 taking values above 1. Fenske (2014) treats it as an ordinal variable with 5 levels of stratification, which we treat as continuous in our reanalysis of his paper.

Reliance on agriculture. Dependence on agriculture for subsistence. Derived from an ordinal variable EA005, ranging from 0 (0-5% share) to 9 (86-100% share). We compute midpoints for each interval and treat the resulting variable as continuous.

Reliance on gathering. Dependence on gathering wild plants and small land fauna for subsistence. Derived from an ordinal variable EA001, ranging from 0 (0-5% share) to 9 (86-100% share). We compute midpoints for each interval and treat the resulting variable as continuous.

Settlement complexity. The prevailing type of settlement pattern. The original variable EA030 is ordinal: 1) fully migratory or nomadic bands, 2) semi-nomadic communities, 3) semi-sedentary communities, 4) compact but impermanent settlements, 5) neighborhoods of dispersed family homesteads, 6) separated hamlets, 7) compact and relatively permanent settlements, 8) complex settlements. We treat this variable as continuous.

Institutions

Indigenous slavery. An indicator variable equal to 1, if a society had any form of slavery. Corresponds to the original variable EA070 taking values above 1.

Local democracy. An indicator variable equal to 1, if a society is a traditional local democracy, that is, political succession (applying to the office of local headman or a close equivalent) occurs through election or formal/informal consensus. Corresponds to the original variable EA072 taking values 6 or 7.

Partible land inheritance. An indicator variable equal to 1, if a society practices partible land inheritance, where land is distributed equally or relatively equally among all members of the same category. Corresponds to the original variable EA075 taking the value 1.

Jurisdictional hierarchy beyond local community. The number of jurisdictional levels beyond the local community. The original variable EA033 is ordinal: 1) no political authority beyond the local community (e.g., autonomous bands or villages), 2) one level (e.g., petty chiefdoms, 3) two levels (e.g., larger chiefdoms), 4) three levels (e.g., states), 5) four levels (e.g., large, complex states). We treat this variable as continuous.

Culture

Bride price. An indicator variable equal to 1, if bride-price or bride-wealth was the prevailing type of transfer at marriage. Corresponds to the original variable EA06 taking the value 1.

Polygamy. An indicator variable equal to 1, if polygyny or polyandry were common or general in a society. Corresponds to the original variable EA009 taking values above 2.

High gods. An indicator variable equal to 1, if a high god is present and active in human affairs. Corresponds to the original variable EA034 taking values above 2.

Matrilineal descent. An indicator variable equal to 1, if matrilineal descent was a major mode of descent. Corresponds to the original variable EA043 taking the value 3.

Kinship tightness. An index describing the extent to which people are interconnected in tightly structured, extended family systems. Constructed based on several EA variables describing family structures and descent systems. *Source*: Enke (2019).

Additional variables used in section 6

Conflict incidents. Log transformation of the number of deadly conflict incidents between 1997 and 2014, based on ACLED data. *Source*: Moscona et al. (2020).

Dependence on pastoralism. Historical reliance on pastoralism (involving herded animals) for subsistence. *Source*: Becker (2024).

Ecological diversity. An index of ecological fractionalization, based on vegetation zones. *Source*: Fenske (2014).

Female participation in agriculture. An ordinal variable with 5 categories: 1) males only, 2) males appreciably more, 3) equal participation by males and females, 4) females appreciably more, 5) females only. *Source*: Alesina et al. (2013).

Insistence on virginity. An indicator variable equal to 1, if premarital sex relations were not allowed for unmarried women or allowed but sanctioned if pregnancy results. *Source*: Becker (2024).

Irrigation potential. The land area where agriculture is impossible without irrigation as a share of total arable land. *Source*: Bentzen and Gokmen (2023).

Land quality. Index of land quality based on climate, soil and terrain slope constraints on rain-fed agriculture. *Source*: Fenske (2013).

Large domesticated animals. An indicator variable equal to 1, if the predominant type of domesticated animals is equines, deer, camelids, or bovines. *Source*: Alsan (2015).

Log population density. Log of population density based on the EA data. *Source*: Alsan (2015), Fenske (2013).

Malaria burden. A measure of malaria burden based on the prevalence of the mutation that causes sickle cell disease. *Source*: Depetris-Chauvin and Weil (2018).

Malaria ecology. Average malaria stability index. *Source*: Enke (2019).

Patrilocal residence. An indicator variable equal to 1, if marital residence is with or near the male patrilineal kinsmen of the husband. *Source*: Becker (2024).

Purity. The duration of sexual abstinence following childbirth. An ordinal variable with 6 categories ranging from none to more than two years. *Source*: Enke (2019).

Reliance on cereals. An indicator variable equal to 1, if cereal grains are the principle type of cultivated crops. *Source*: Mayshar et al. (2022).

Segmentary lineage. Specification in table 2A, column 3, uses a binary index of presence or absence of segmentary lineage organization based on the *Ethnographic Survey of Africa*. Specification in the appendix table 4A, column 3, uses an alternative index of segmentary lineage organization based on the EA data. *Source*: Moscona et al. (2020).

State centralization. An indicator variable equal to 1, if there is one or more (Depetris-Chauvin and Weil, 2018) or two or more (Alsan, 2015) jurisdictional levels beyond the local community.

TseTse suitability. The average TseTse fly suitability index. *Source*: Alsan (2015).

Women as sexual. Log of the share of motifs that depict women as sexual among the total number of motifs in society's folklore. *Source*: Becker (2024).

B Glossary of terms related to phylogenetic trees

The terms explained in this section broadly apply to both phylogenetic and language trees, with obvious adaptations. We use an example tree in figure B.1 for illustration. See Nunn (2011) for more details.

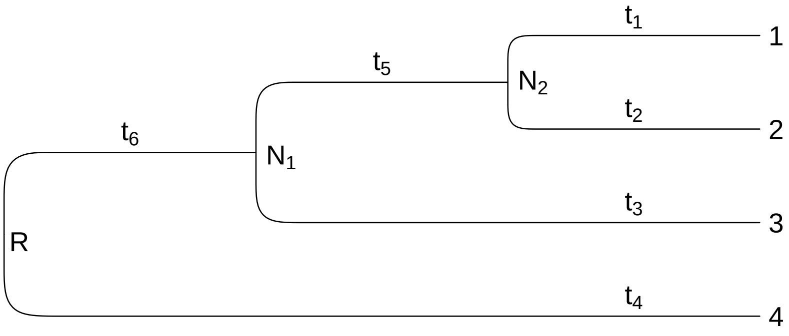


Figure B.1: Example phylogenetic tree.

Phylogeny. The branching evolutionary history of societies, represented as a tree-like structure. Often used synonymously with "phylogenetic tree."

Phylogenetic tree. A graphical representation of the branching evolutionary history of societies, arranged hierarchically to show nested groups that indicate when two or more societies last shared a common ancestor. A tree consists of nodes connected by branches and can represent either relative or absolute timing of evolutionary divergence events. The tree in figure B.1 illustrates ancestral relationships among four societies.

Node. A point on a phylogenetic tree corresponding to a lineage-splitting event. Represents a present-day society or its ancestor.

Root node (root). The base of a phylogenetic tree, representing the most recent common ancestor (MRCA) of all societies. In figure B.1, the root is marked with R.

Internal node. A node with at least one descendant. The nodes labeled N_1 and N_2, as well as the root R, are internal nodes in figure B.1.

Terminal nodes (tips, leaves). The outermost points of the tree representing current societies. In figure B.1, tips $1, 2, 3, 4$ represent extant societies.

Branch (edge). A line connecting two nodes in a phylogenetic tree. Branches connecting internal nodes are called **internal**. Branches connecting internal nodes to tips are called

terminal. In figure B.1, branches labeled with t_1, t_2, t_3, and t_4 are terminal, and those labeled with t_5 and t_6 are internal. These labels correspond to branch lengths.

Clade. A part of the tree that includes a group of societies together with their common ancestor and all of its descendants. For example, in figure B.1, societies 1 and 2, together with corresponding terminal branches of lengths t_1 and t_2 and their MRCA N_2 form a clade.

Sister clades. Two societies or clades that diverge from the same tree node.

Node age. The time that has passed since the divergence event that a node represents. Typically measured relative to the present.

Crown age (tree height). The distance from the root to the most distant tip, representing the total evolutionary change or time that has passed from the root to the present. In figure B.1, the crown age of the tree is equal to t_4.

Ultrametric tree. A phylogenetic tree where all tips are equidistant from the root. The tree in figure B.1 is ultrametric.

Time calibrated tree. A phylogenetic tree where branch lengths capture the passage of time.

Bifurcating tree. A phylogenetic tree where each internal node has one ancestral branch and two descendant branches. The tree in figure B.1 is bifurcating.

Polytomy. An internal node in a phylogenetic tree from which more than two branches descend simultaneously. An example of a polytomy is shown in the right panel of figure 2.

Rooted tree. A phylogenetic tree with a single, defined starting point (root). The tree in B.1 is rooted.

Star phylogeny. A tree in which all lineages diverge simultaneously from the root. In this tree, there are no internal nodes or branches, indicating that all societies evolved completely independently.

Pruned tree. A phylogenetic tree with certain branches and nodes removed. Pruning deletes the tree elements that are not relevant to a particular analysis because of missing data or sample restrictions, while retaining the original tree structure.

C Additional simulation results

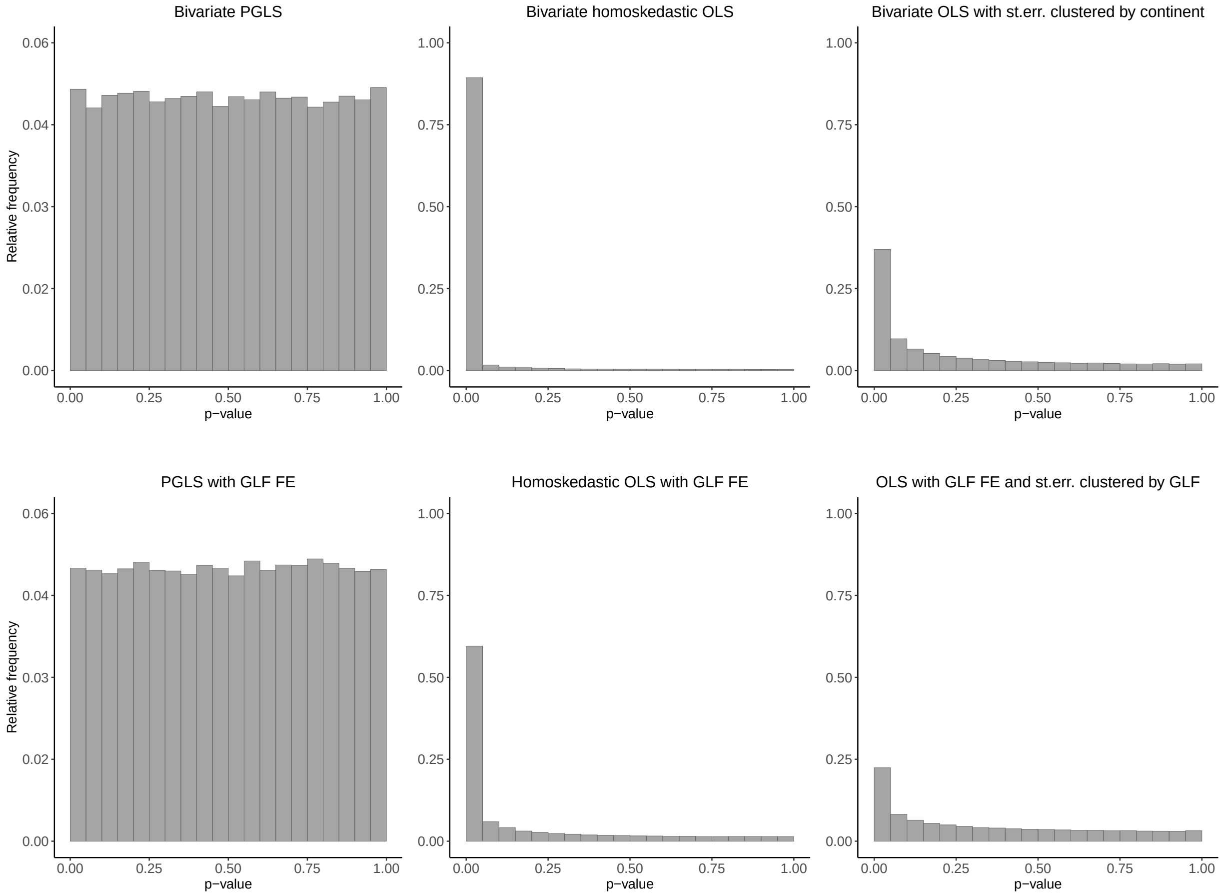


Figure C.1: Distributions of p-values for selected estimation strategies.

Notes: The figure displays the distributions of p-values associated with testing the statistical significance of the slope coefficient in selected model specifications and estimation strategies described in section 5. Each histogram is based on 90,200 simulations.

Table C.1: Type I error rates in models estimated by OLS with Conley standard errors.

	Distance cutoff (in km)				
Model specification	200	500	800	1000	1500
Bivariate	0.83	0.73	0.67	0.63	0.58
Continent FE	0.76	0.64	0.58	0.55	0.51
Country FE	0.66	0.50	0.43	0.40	0.36
MLF FE	0.53	0.42	0.38	0.37	0.37
GLF FE	0.40	0.26	0.23	0.23	0.24

Notes: Each rate is based on 90,200 simulations. See section 5 for details of the modeling setup.

D Additional results from the reanalysis

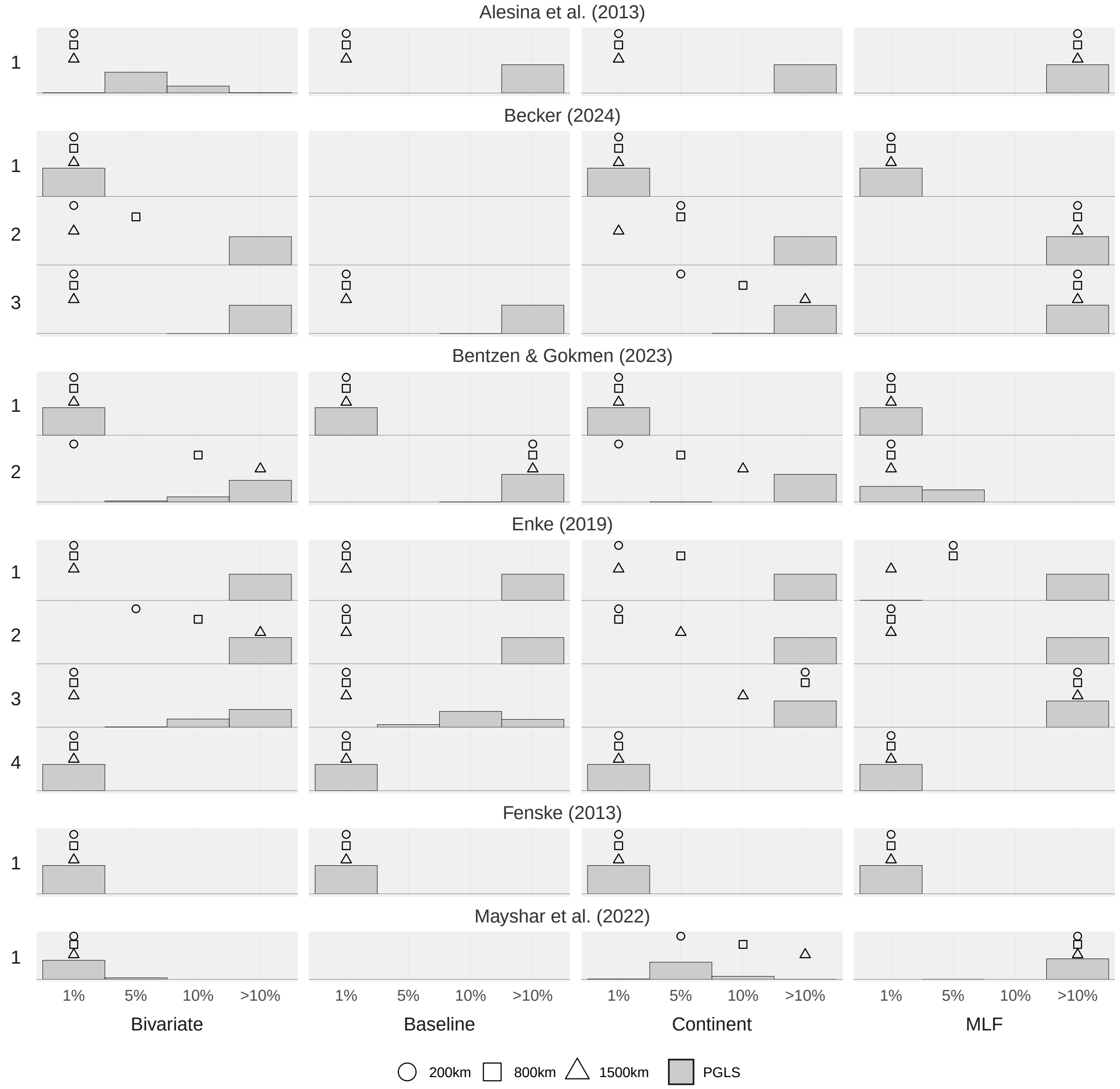


Figure D.1: OLS with spatially adjusted standard errors vs PGLS in global samples.

Notes: Levels of statistical significance are grouped into four categories: 1%, 5%, 10%, and greater than 10%. For OLS estimates, significance levels are based on standard errors adjusted for spatial correlation using the method of Conley (1999) for three distance cutoff values: 200 km (circles), 800 km (squares), and 1500 km (triangles). For PGLS, the figure shows histograms of *p*-values corresponding to estimates within the IQRs and based on the four significance-level bins. Baseline specification includes all controls from the original studies (excluding fixed effects). Continent and MLF specifications add respective fixed effects to the baseline specification. In three cases, namely models 1 and 2 from Becker (2024) and model 1 from Mayshar et al. (2022), baseline specifications are identical to bivariate cases and are omitted.

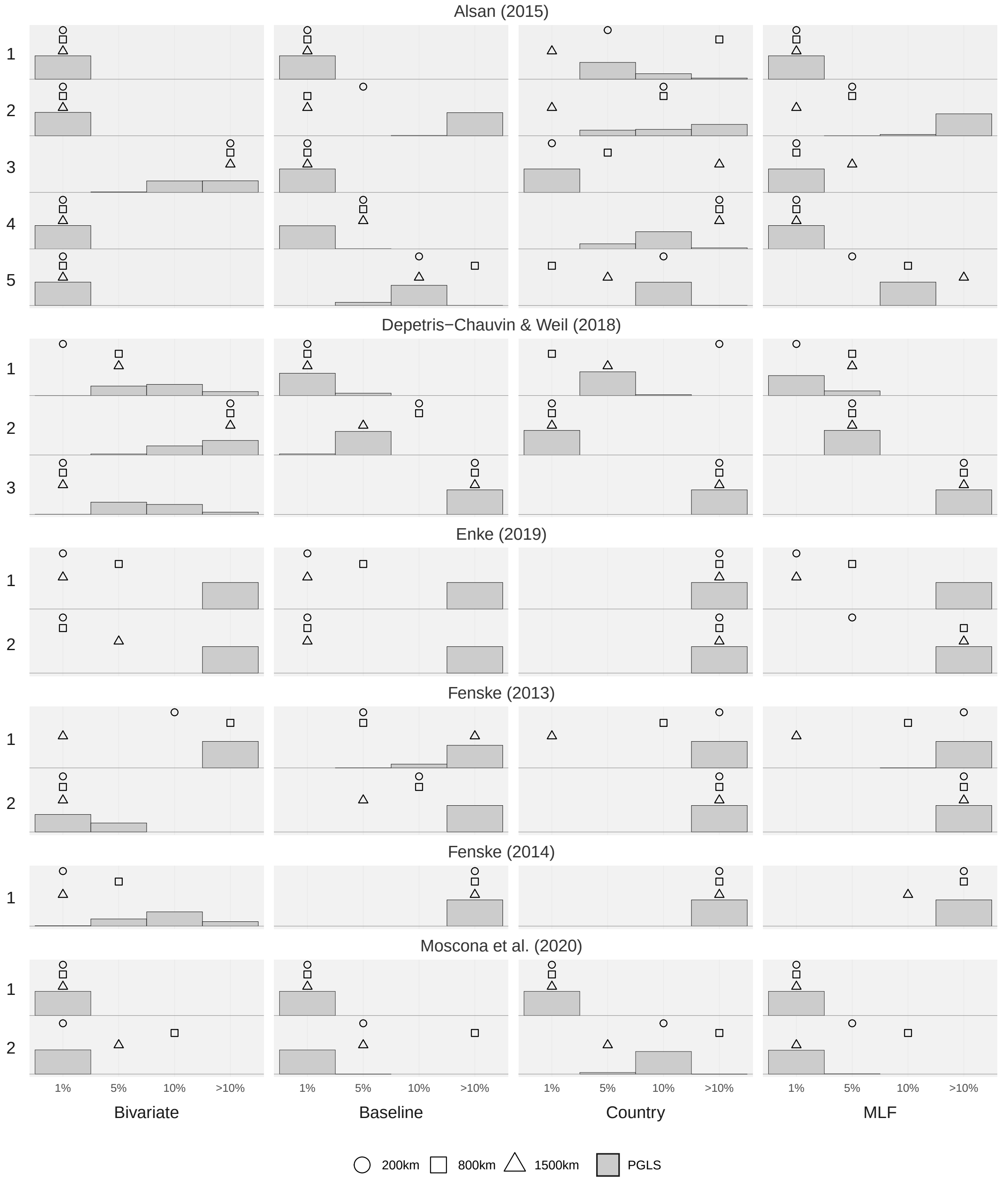


Figure D.2: OLS with spatially adjusted standard errors vs PGLS in African samples.

Notes: Notes: Levels of statistical significance are grouped into four categories: 1%, 5%, 10%, and greater than 10%. For OLS estimates, significance levels are based on standard errors adjusted for spatial correlation using the method of Conley (1999) for three distance cutoff values: 200 km (circles), 800 km (squares), and 1500 km (triangles). For PGLS, the figure shows histograms of p-values corresponding to estimates within the IQRs and based on the four significance-level bins. Baseline specification includes all controls from the original studies (excluding fixed effects). Continent and MLF specifications add respective fixed effects to the baseline specification.

References

Abouheif, Ehab, "A Method for Testing the Assumption of Phylogenetic Independence in Comparative Data," *Evolutionary Ecology Research*, 1999, *1*, 895–909.

Alesina, Alberto, Paola Giuliano, and Nathan Nunn, "On the Origins of Gender Roles: Women and the Plough," *Quarterly Journal of Economics*, May 2013, *128* (2), 469–530.

Alsan, Marcella, "The Effect of the TseTse Fly on African Development," *American Economic Review*, January 2015, *105* (1), 382–410.

Anselin, Luc and Daniel Arribas-Bel, "Spatial Fixed Effects and Spatial Dependence in a Single Cross-Section," *Papers in Regional Science*, 2013, *92* (1), 3–18.

Bahrami-Rad, Duman, Anke Becker, and Joseph Henrich, "Tabulated Nonsense? Testing the Validity of the Ethnographic Atlas," *Economics Letters*, 2021, *204*, 109880.

Bastide, Paul, Claudia Solís-Lemus, Ricardo Kriebel, K William Sparks, and Cécile Ané, "Phylogenetic Comparative Methods on Phylogenetic Networks with Reticulations," *Systematic Biology*, June 2018, *67* (5), 800–820.

Becker, Anke, "On the Economic Origins of Concerns Over Women's Chastity," *Review of Economic Studies*, 2024, *forthcoming.*

Bentzen, Jeanet and Gunes Gokmen, "The Power of Religion," *Journal of Economic Growth*, 2023, *28* (1).

Blomberg, Simon P., Theodore Garland, and Anthony R. Ives, "Testing for Phylogenetic Signal in Comparative Data: Behavioral Traits Are More Labile," *Evolution*, 2003, *57* (4), 717–745.

Bouckaert, Remco, David Redding, Oliver Sheehan, Thanos Kyritsis, Russell Gray, Kate Jones, and Quentin Atkinson, "Global Language Diversification Is Linked to Socio-Ecology and Threat Status," July 2022. SocArXiv working paper.

Cameron, A. Colin and Douglas L. Miller, "A Practitioner's Guide to Cluster-Robust Inference," *Journal of Human Resources*, 2015, *50* (2), 317–372.

Claessens, Scott, Thanos Kyritsis, and Quentin D. Atkinson, "Cross-National Analyses Require Additional Controls to Account for the Non-Independence of Nations," *Nature Communications*, September 2023, *14* (1), 5776.

Conley, Timothy G., "GMM Estimation with Cross Sectional Dependence," *Journal of Econometrics*, September 1999, *92* (1), 1–45.

_ , "Spatial Econometrics," in S.N. Durlauf and L.E. Blume, eds., *Microeconometrics*, London: Palgrave Macmillan, 2010, pp. 303–313.

_ **and Morgan Kelly**, "The Standard Errors of Persistence," *Journal of International Economics*, 2024, *forthcoming.*

Currie, Thomas E., Simon J. Greenhill, and Ruth Mace, "Is Horizontal Transmission Really a Problem for Phylogenetic Comparative Methods? A Simulation Study Using Continuous Cultural Traits," *Philosophical Transactions of the Royal Society B: Biological Sciences*, 2010, *365* (1559), 3903–3912.

Depetris-Chauvin, Emilio and David N. Weil, "Malaria and Early African Development: Evidence from the Sickle Cell Trait," *Economic Journal*, May 2018, *128* (610), 1207–1234.

Díaz-Uriarte, Ramón and Jr. Garland Theodore, "Testing Hypotheses of Correlated Evolution Using Phylogenetically Independent Contrasts: Sensitivity to Deviations from Brownian Motion," *Systematic Biology*, March 1996, *45* (1), 27–47.

Dow, Malcolm M., Michael L. Burton, Douglas R. White, and Karl P. Reitz, "Galton's Problem as Network Autocorrelation," *American Ethnologist*, 1984, *11* (4), 754–770.

Duda, Pavel and Jan Zrzavý, "Towards a Global Phylogeny of Human Populations Based on Genetic and Linguistic Data," in Yonatan Sahle, Hugo Reyes-Centeno, and Christian Bentz, eds., *Modern Human Origins and Dispersal*, Tübingen, Germany: Kerns Verlag, 2019, chapter 12, pp. 331–359.

Eff, Ellis Anthon, "Does Mr. Galton Still Have a Problem? Autocorrelation in the Standard Cross-Cultural Sample," *World Cultures*, 2004, *15* (2), 153–170.

Enke, Benjamin, "Kinship, Cooperation, and the Evolution of Moral Systems," *Quarterly Journal of Economics*, May 2019, *134* (2), 953–1019.

Felsenstein, Joseph, "Maximum-Likelihood Estimation of Evolutionary Trees from Continuous Characters," *American Journal of Human Genetics*, September 1973, *25* (5), 471–492.

_ , "Phylogenies and the Comparative Method," *The American Naturalist*, 1985, *125* (1), 1–15.

Fenske, James, "Does Land Abundance Explain African Institutions?," *Economic Journal*, December 2013, *123* (573), 1363–1390.

_ , "Ecology, Trade, and States in Pre-Colonial Africa," *Journal of the European Economic Association*, 2014, *12* (3), 612–640.

Freckleton, Robert P., Paul H. Harvey, and Mark Pagel, "Phylogenetic Analysis and Comparative Data: A Test and Review of Evidence," *American Naturalist*, 2002, *160* (6), 712–726.

Fritz, Susanne A. and Andy Purvis, "Selectivity in Mammalian Extinction Risk and Threat Types: A New Measure of Phylogenetic Signal Strength in Binary Traits," *Conservation Biology*, August 2010, *24* (4), 1042–1051.

Gershman, Boris, "The Economic Origins of the Evil Eye Belief," *Journal of Economic Behavior & Organization*, February 2015, *110*, 119–144.

_ , "Long-Run Development and the New Cultural Economics," in Matteo Cervelatti and Uwe Sunde, eds., *Demographic Change and Long-Run Development*, Cambridge: MIT Press, 2017, chapter 9, pp. 221–261.

_ **and Tinatin Mumladze**, "Headhunting and Warfare in Austronesia: A Phylogenetic Comparative Analysis," *Journal of Economic Behavior & Organization*, 2024, *220*, 768–791.

Ginsburgh, Victor and Shlomo Weber, "The Economics of Language," *Journal of Economic Literature*, June 2020, *58* (2), 348–404.

Gittleman, John L. and Mark Kot, "Adaptation: Statistics and a Null Model for Estimating Phylogenetic Effects," *Systematic Zoology*, September 1990, *39* (3), 227–241.

Giuliano, Paola and Nathan Nunn, "Ancestral Characteristics of Modern Populations," *Economic History of Developing Regions*, 2018, *33* (1), 1–17.

Grafen, Alan, "The Phylogenetic Regression," *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, December 1989, *326* (1233), 119–157.

Gray, Russell D. and Joseph Watts, "Cultural Macroevolution Matters," *Proceedings of the National Academy of Sciences*, 2017, *114* (30), 7846–7852.

_ , David Bryant, and Simon J. Greenhill, "On the Shape and Fabric of Human History," *Philosophical Transactions of the Royal Society B: Biological Sciences*, 2010, *365* (1559), 3923–3933.

Greenhill, Simon J., Thomas E. Currie, and Russell D. Gray, "Does Horizontal Transmission Invalidate Cultural Phylogenies?," *Proceedings of the Royal Society B: Biological Sciences*, 2009, *276* (1665), 2299–2306.

Guarnieri, Eleonora and Ana Tur-Prats, "Cultural Distance and Conflict-Related Sexual Violence," *Quarterly Journal of Economics*, 2023, *138* (3), 1817–1861.

Hammarström, Harald, Robert Forkel, and Martin Haspelmath, "Glottolog 4.0," 2019. Dataset available at `https://doi.org/10.5281/zenodo.10804357`.

Hansen, Thomas F. and Emília P. Martins, "Translating Between Microevolutionary Process and Macroevolutionary Patterns: The Correlation Structure of Interspecific Data," *Evolution*, August 1996, *50* (4), 1404–1417.

Harmon, Luke, *Phylogenetic Comparative Methods: Learning From Trees*, Moscow, ID: EcoEvoRxiv, 2019.

Harvey, Paul H. and Mark D. Pagel, *The Comparative Method in Evolutionary Biology*, Oxford: Oxford University Press, 1991.

Heled, Joseph and Remco R. Bouckaert, "Looking for Trees in the Forest: Summary Tree from Posterior Samples," *BMC Evolutionary Biology*, 2013, *13*, 221.

Ives, Anthony R. and Theodore Garland Jr., "Phylogenetic Logistic Regression for Binary Dependent Variables," *Systematic Biology*, January 2010, *59* (1), 9–26.

Jäger, Gerhard, "Global-Scale Phylogenetic Linguistic Inference From Lexical Resources," *Scientific Data*, October 2018, *5* (1), 180189.

Kirby, Kathryn R., Russell D. Gray, Simon J. Greenhill, Fiona M. Jordan, Stephanie Gomes-Ng, Hans-Jörg Bibiko, Damián E. Blasi, Carlos A. Botero, Claire Bowern, Carol R. Ember, Dan Leehr, Bobbi S. Low, Joe McCarter, William Divale, and Michael C. Gavin, "D-PLACE: A Global Database of Cultural, Linguistic and Environmental Diversity," *PLoS ONE*, July 2016, *11* (7), e0158391.

Lowes, Sara, "Ethnographic and Field Data in Historical Economics," in Alberto Bisin and Giovanni Federico, eds., *Handbook of Historical Economics*, London: Academic Press, January 2021, pp. 147–177.

_ , "Culture in Historical Political Economy," in Jeffery A. Jenkins and Jared Rubin, eds., *The Oxford Handbook of Historical Political Economy*, Oxford University Press, 2023.

Mace, Ruth and Clare J. Holden, "A Phylogenetic Approach to Cultural Evolution," *Trends in Ecology & Evolution*, March 2005, *20* (3), 116–121.

_ **and Hanzhi Zhang**, "Cross-Cultural Comparative Methods for Testing Evolutionary Hypotheses," in Jamshid J. Tehrani, Jeremy Kendal, and Rachel Kendal, eds., *The Oxford Handbook of Cultural Evolution*, Oxford Academic, 2023.

_ **and Mark Pagel**, "The Comparative Method in Anthropology," *Current Anthropology*, December 1994, *35*, 549–564.

MacKinnon, James G., Morten Ørregaard Nielsen, and Matthew D. Webb, "Cluster-Robust Inference: A Guide to Empirical Practice," *Journal of Econometrics*, February 2023, *232* (2), 272–299.

Martins, Emília P. and Elizabeth A. Housworth, "Phylogeny Shape and the Phylogenetic Comparative Method," *Systematic Biology*, 2002, *51* (6), 873–880.

_ **and Thomas F. Hansen**, "Phylogenies and the Comparative Method: A General Approach to Incorporating Phylogenetic Information into the Analysis of Interspecific Data," *American Naturalist*, 1997, *149* (4), 646–667.

_ **, José Alexandre F. Diniz-Filho, and Elizabeth A. Housworth**, "Adaptive Constraints and the Phylogenetic Comparative Method: A Computer Simulation Test," *Evolution*, 2002, *56* (1), 1–13.

Mayshar, Joram, Omer Moav, and Luigi Pascali, "The Origin of the State: Land Productivity or Appropriability?," *Journal of Political Economy*, April 2022, *130* (4), 1091–1144.

Michalopoulos, Stelios and Elias Papaioannou, "Spatial Patterns of Development: A Meso Approach," *Annual Review of Economics*, 2018, *10*, 383–410.

Moscona, Jacob, Nathan Nunn, and James A. Robinson, "Segmentary Lineage Organization and Conflict in Sub-Saharan Africa," *Econometrica*, 2020, *88* (5), 1999–2036.

Mundry, Roger, "Statistical Issues and Assumptions of Phylogenetic Generalized Least Squares," in László Zsolt Garamszegi, ed., *Modern Phylogenetic Comparative Methods and Their Application in Evolutionary Biology*, Heidelberg: Springer Berlin, 2014, chapter 6, pp. 131–153.

Münkemüller, Tamara, Sébastien Lavergne, Bruno Bzeznik, Stéphane Dray, Thibaut Jombart, Katja Schiffers, and Wilfried Thuiller, "How to Measure and Test Phylogenetic Signal," *Methods in Ecology and Evolution*, August 2012, *3* (4), 743–756.

Murdock, George P., *Ethnographic Atlas*, Pittsburgh, PA: University of Pittsburgh Press, 1967.

_ **and Douglas R. White**, "Standard Cross-Cultural Sample," *Ethnology*, October 1969, *8* (4), 329–369.

Nunn, Charles L., *The Comparative Approach in Evolutionary Anthropology and Biology*, Chicago, IL: University of Chicago Press, 2011.

_ **, Monique Borgerhoff Mulder, and Sasha Langley**, "Comparative Methods for Studying Cultural Trait Evolution: A Simulation Study," *Cross-Cultural Research*, 2006, *40* (2), 177–209.

Nunn, Nathan, "The Historical Roots of Economic Development," *Science*, 2020, *367* (6485), eaaz9986.

Pagel, M., "Inferring the Historical Patterns of Biological Evolution," *Nature*, October 1999, *401* (6756), 877–884.

Revell, Liam J., "Phylogenetic Signal and Linear Regression on Species Data," *Methods in Ecology and Evolution*, 2010, *1* (4), 319–329.

_ **and Luke J. Harmon**, *Phylogenetic Comparative Methods in R*, Princeton and Oxford: Princeton University Press, 2022.

Ringen, Erik J., Pavel Duda, and Adrian V. Jaeggi, "The Evolution of Daily Food Sharing: A Bayesian Phylogenetic Analysis," *Evolution and Human Behavior*, 2019, *40* (4), 375–384.

Rohlf, F. James, "Comparative Methods for the Analysis of Continuous Variables: Geometric Interpretations," *Evolution*, 2001, *55* (11), 2143–2160.

Šaffa, Gabriel, Jan Zrzavý, and Pavel Duda, "Global Phylogenetic Analysis Reveals Multiple Origins and Correlates of Genital Mutilation/Cutting," *Nature Human Behaviour*, May 2022, *6* (5), 635–645.

Schulz, Jonathan F., Duman Bahrami-Rad, Jonathan P. Beauchamp, and Joseph Henrich, "The Church, Intensive Kinship, and Global Psychological Variation," *Science*, November 2019, *366* (6466), eaau5141.

Spolaore, Enrico and Romain Wacziarg, "Ancestry, Language and Culture," in Victor Ginsburgh and Shlomo Weber, eds., *The Palgrave Handbook of Economics and Language*, London: Palgrave Macmillan, 2016, chapter 6, pp. 174–211.

Stone, Eric A., "Why the Phylogenetic Regression Appears Robust to Tree Misspecification," *Systematic Biology*, February 2011, *60* (3), 245–260.

Surowiec, Alexandra, Kate T. Snyder, and Nicole Creanza, "A Worldwide View of Matriliny: Using Cross-Cultural Analyses to Shed Light on Human Kinship Systems," *Philosophical Transactions of the Royal Society B: Biological Sciences*, 2019, *374* (1780), 20180077.

Symonds, Matthew and Simone Blomberg, "A Primer on Phylogenetic Generalised Least Squares," in László Zsolt Garamszegi, ed., *Modern Phylogenetic Comparative Methods and Their Application in Evolutionary Biology*, Heidelberg: Springer Berlin, 2014, chapter 5, pp. 105–130.

Symonds, Matthew R. E., "The Effects of Topological Inaccuracy in Evolutionary Trees on the Phylogenetic Comparative Method of Independent Contrasts," *Systematic Biology*, July 2002, *51* (4), 541–553.

Tylor, Edward B., "On a Method of Investigating the Development of Institutions; Applied to Laws of Marriage and Descent," *The Journal of the Anthropological Institute of Great Britain and Ireland*, October 1889, *18*, 245–272.