DEPARTMENT OF ECONOMICS AND FINANCE WORKING PAPER SERIES • *February 2004*

Spatial and Cultural Autocorrelation in International Datasets

E. Anthon Eff *
Middle Tennessee State University, Murfreesboro, TN

Abstract

Positive autocorrelation implies that proximate observations take on similar values. "Proximate" can be defined in many different dimensions. In a cross-section of nations, it can be defined using physical distance, cultural similarity, ecological similarity, or using frequency and intensity of interaction, such as trade relationships or enemy and ally relationships. Autocorrelation of regression residuals presents well-known problems in least-squares estimation, but autocorrelation also provides useful information for exploratory data analysis and model specification. The paper shows that autocorrelation is widespread in international datasets. The paper demonstrates the usefulness of autocorrelation in uncovering stylized facts about international relations, and in specifying a least-squares model.

Key words: Spatial Autocorrelation; Culture; Religion

JEL category: F00, C49, Z10, Z12

*E. Anthon Eff, Associate Professor Department of Economics and Finance, Middle Tennessee State University, Murfreesboro, TN 37132, phone: 615-898-2387, fax: 615-898-5045, email: eaeff@mtsu.edu.
The weight matrices described in this paper are available by request from the author.

SPATIAL AND CULTURAL AUTOCORRELATION IN INTERNATIONAL DATASETS

Abstract: Positive autocorrelation implies that proximate observations take on similar values. "Proximate" can be defined in many different dimensions. In a cross-section of nations, it can be defined using physical distance, cultural similarity, ecological similarity, or using frequency and intensity of interaction, such as trade relationships or enemy and ally relationships. Autocorrelation of regression residuals presents well-known problems in least-squares estimation, but autocorrelation also provides useful information for exploratory data analysis and model specification. The paper shows that autocorrelation is widespread in international datasets. The paper demonstrates the usefulness of autocorrelation in uncovering stylized facts about international relations, and in specifying a least-squares model.

1. INTRODUCTION

In 1889, Edward Tylor presented what was to become the seminal paper in statistical cross-cultural analysis, before a panel at the Royal Anthropological Institute. Sitting on the panel was Sir Francis Galton, the statistician and cousin of Charles Darwin. Tylor had compiled information on institutions of marriage and descent for 350 cultures and examined the correlations among these institutions. The results showed that certain institutions were associated with each other far more often than chance would imply. Tylor interpreted these results as indications of a general evolutionary sequence, in which institutions changed focus from the maternal line to the paternal line. Galton disagreed, pointing out that similarity between cultures could be due to borrowing, could be due to common descent, or could be due to evolutionary development. He maintained that without controlling for borrowing and common descent one cannot make valid inferences regarding evolutionary development. In the literature, Galton's critique has become the eponymous "Galton's Problem." (Stocking 1968: 175).

At its most general, Galton's Problem is simply the observation that univariate correlations are unreliable since they do not control for other sources of influence. But it also contains a specific warning: statistical studies using cross-cultural data sets to examine structural-functional relationships among social phenomena must consider that common descent or cultural borrowing are two of the most important sources of outside influence that one must control.

Economic theory and econometrics work with models that are analogues of structural-functional perspectives: some social metric such as the average wage responds to changes in some other social metric such as the literacy rate, in much the same way that—in Tylor's study—descent rules respond to the complexity of social organization. And like Tylor, empirical work will often use a cross-section of social entities (typically nations) as the data for testing hypotheses. And again like Tylor, statistical results which at first glance suggest functional relationships might in fact be due to particular patterns of inheritance or borrowing.

Galton's Problem motivated at least two important research efforts in anthropology (Naroll 1965; Mace and Pagel 1994). The first approach, conducted by George Peter Murdock and his colleagues, attempted to tackle the problem by developing a sample of cultures relatively independent from each other—i.e., with relatively weak relationships of borrowing and common descent (Murdock, 1957; Murdock and White, 1969). The second approach sought to use statistical techniques to control for relationships of borrowing and descent. Working with ideas originally from geography, these anthropologists—some of whom were Murdock's students—used spatial statistics and spatially lagged models to study the relationships among cultures, and incorporated those relationships in regression models (Loftin 1972; Loftin and Ward 1983; White, Burton, and Dow 1981; Dow, Burton, and White 1982; Dow, White, and Burton 1982; Dow, Burton, Reitz, and White 1984). In this work there is the clear recognition that the distance between societies could be either spatial (based on physical distance) or cultural (based on language phylogenies).

This approach never quite caught on except among some mathematically oriented anthropologists, who in fact have further refined the methods (Mace and Pagel 1994). Outside of anthropology, there has been no use of these techniques with the exception of a recent paper in political science, which employs some of the statistical innovations of Dow, Burton, Reitz, and White (1984) without apparently being aware of their pioneering work (Beck and Gleditsch 2003). Despite a upsurge of interest in the cultural determinants of economic growth (Harrison and Huntington 2000; Landes 1998), the work of Dow, Burton, Reitz, and White (1984) has been ignored in economics.

Ordinary least squares requires that the residuals in the estimated model not be correlated with each other. Violation of this property causes the estimated standard errors of the coefficients to be biased, so that one cannot trust the t-statistics, and one therefore cannot make hypothesis tests regarding the estimated coefficients (Kennedy 1998). Even worse, the presence of autocorrelation is often a sign of omitted variables, so that even the estimated coefficients may be biased. For example, in a hedonic housing model a spatial autocorrelation test compares the residual for a particular house to the residuals of houses in the same neighborhood. Typically, these models show significant autocorrelation of the residuals, since the factors leading a house to have a particularly high (low) residual are often causing neighboring houses to have a particularly high (low) residual. These "neighborhood effects" are a sign of omitted variables, a problem that can usually most easily be cured by incorporating a spatially lagged dependent variable in the model.

In regressions on cross-national datasets, proximity of nations could be either spatial or cultural, and autocorrelation among spatially or culturally proximate residuals would indicate “neighborhood effects” in which variables similarly affecting related nations have been omitted. The sources of autocorrelation are exactly the sources mentioned by Sir Francis Galton: one has failed to control for relations of borrowing and descent.

The paper is organized as follows. The following section details the construction of 12 different weight matrices, each of which defines “proximity” between nations in a different way. The next section then compares the 12 matrices, using matrix correlation, to get a sense of how they covary. The paper then presents a dataset of 72 variables for 152 nations, drawn from a wide variety of sources. The variables are examined for autocorrelation and the results both show the very high incidence of autocorrelation in international data, and allow stylized facts to be produced of the form: “nations that are spatially (culturally, etc.) proximate, tend to have similar values of variable *y*.” The final section then shows how OLS regression residuals can be examined for autocorrelation, and how the autocorrelation statistics can be used to gain a sense of whether a model has an omitted variable problem.

2. Construction of Weight Matrices

The relationships among nations can be described in many different ways. International relationships are here operationalized by constructing 12 different weight matrices, each modeling a different dimension of international relationship. Each weight matrix **W** contains elements w_{ij} giving the proximity between nation i and nation j, where higher values of w_{ij} correspond to greater closeness between the pair of nations.

Physical Proximity

Physical distance forms the basis for a conventional spatial weights matrix. A common approach is to model contiguity, rather than proximity, so that countries with shared borders are proximate, and other countries have no connection. A contiguity matrix **W** will have elements w_{ij} equal to one when the nations are contiguous, and zero otherwise. Contiguity, however, has several problems. First, it is difficult to model the relationships of island nations or nations separated only by sea (e.g., Sweden and Denmark). Second, sparsely populated and inaccessible borders (e.g., that between Afghanistan and China) will bear the same weight as long and densely populated borders. And third, there may be no contiguity between otherwise spatially proximate nations (e.g., Botswana and Zambia). Proximity measures based on physical distance, rather than simple contiguity, are of two types: those that calculate distances between national centroids, and those that calculate distance between the closest points in two countries. The latter approach is similar to the contiguity approach in that a shared border gives the highest proximity, and has a similar disadvantage—that an inaccessible, sparsely populated border will count as much as a long and densely populated border. The former approach requires the calculation of national centroids, and these could either be physical area centroids or population centroids. Physical centroids can give misleading results: for example, Russia will appear to be very far from everywhere else, and the crescent shape of Croatia may situate its physical centroid between the prongs of the crescent—in Bosnia. Population centroids may be the best alternative, since the *perception* of physical proximity between countries is highest where population concentrations lie close to each other. Even population centroids, though, can be misleading when a nation has several widely dispersed centers of population, and the centroid is placed in a sparsely inhabited zone in the middle: for example, the United States and Canada have concentrations of population on each coast, and the population centroid for these countries is situated in its less densely inhabited midsection. Considering that each country's largest city is likely to contain its most important linkages to the outside world, it makes more sense to employ the centroid for the largest city as the national centroid. Thus Russia is centered on Moscow, so that its orientation is toward Europe; the U.S.'s centroid is New York, and is oriented toward the North Atlantic; and China's centroid is Shanghai, arguably its most internationally oriented big city. India's centroid however, is Calcutta, shifting its focus toward Bangladesh, rather than Pakistan.

The great circle distance in kilometers between each pair of centroids is calculated as follows:

(1) $$d_{ij}=6371.1*\arccos[\sin(y_i)*\sin(y_j)+\cos(y_i)*\cos(y_j)*\cos(x_i-x_j)]$$

where y_i is the latitude in radians for country *i*, x_i is the longitude in radians for country *i*, and the subscript *j* refers to similar measures for country *j*. Distance is converted to proximity using the following formula:

(2) $$w_{ij} = (1+0.001*d_{ij})^{-2} \quad \text{and} \quad w_{ii} = 0$$

The diagonal of the proximity matrix **W** is set to zero, and the matrix is then standardized by dividing each element w_{ij} by the largest of the elements w_{ij} so that each w_{ij} ranges from zero to one.

Cultural Proximity I (Linguistic Proximity)

Recent scholarship in economics has reintroduced the notion that culture is an important determinant of economic development (Harrison and Huntington 2000; Landes 1998). A pair of countries may have similar levels of economic development, not because they are physical neighbors, but because they are culturally closely related.

Language is the primary vehicle of inherited culture. Since language is usually transmitted from parent to child, geneticists have noted that the biometric distance between any pair of populations matches fairly closely the linguistic distance between those populations (Cavalli-Sforza, Menozzi, and Piazza 1994), so that linguistic similarity is a reasonable proxy for degree of biological relatedness—i.e., is a proxy for common descent. In cases where a native language is replaced by a new language, one generally sees “the adoption of most or all of a whole range of elements” from the culture represented by the new language (Mace and Pagel 1994: 552), so that linguistic similarity remains a good proxy for overall cultural similarity. Linguists have devoted considerable attention to unraveling the phylogenetic relationships among languages. Figure 1 gives an example of a language phylogeny, showing the relationships among 12 living languages in the Indo-European family. Language phylogenies present a best guess about genetic relationships among families, and are not simply based on lexical similarity. Thus, for example, English vocabulary contains a very large number of cognates from French, yet is as distant from French in the graph as it is distant from Hindi. English’s nearest neighbors are the languages that, like it, descended from early Germanic. Language phylogenies thus ignore relationships of borrowing, and focus on relationships of descent.

Intuitively, one can imagine each node in Figure 1 as a separate language. Thus, one can imagine that an ancient people spoke a single language (“Indo-European”), which split into daughter languages, again and again, to form the many nodes and branches of the graph. Figure 1 is a directed graph, where the lines point from a daughter language toward an ancestral language. Thus all lines point in toward Indo-European, and the nodes along the periphery of the graph with no in-pointing lines are living languages.

Historical linguistics can often be a contentious science, particularly regarding distant relationships among the major language families, though there tends to be more widespread agreement on the relationships among languages within each major language family. Thus, while some philologists would lump Indo-European, Uralic, Altaic, Dravidian, Afro-Asiatic, and a few other language families together in a super-family called Nostratic, most regard the evidence as too thin, and language taxonomy is therefore characterized by multiple families, each containing many languages, but without any connection among the families. Similarity measures can therefore be calculated for languages that are members of the same family, but languages in different families are assumed to be completely unrelated.

Mathematical anthropologists (White, Burton, and Dow 1981; Dow, Burton, Reitz, and White 1984) have employed phylogenetic language graphs to calculate distance measures between pairs of languages. In that work, distance is defined as it traditionally is in graph theory: the length (measured in number of links) of the shortest path in the undirected graph connecting the two languages (Scott 2000: 68). Similarity is then calculated as the inverse of distance. Examining Figure 1, one can find a few instances where the White-Burton-Dow (WBD) procedure leads to nonsensical results. For example, the number of links between English and German is seven, while the number of links between English and Armenian is six—which would imply, following the WBD procedure, that English is more closely related to Armenian than to German. The WBD procedure would work well only when all branches in the phylogenetic graph have the same number of links. Since branches in fact vary in the number of links they contain, one must look further for a good similarity measure.

One should keep in mind that each of the living languages lies an equal distance in time from the ancestral Indo-European language. Thus, even though Armenian's graph distance from Indo-European is two, and French's graph distance is ten, the two languages are equidistant in time from Indo-European. Likewise, even though the graph distance between English and West Germanic (its common ancestor with German) is two, and the graph distance between German and West Germanic is five, the two living languages are temporally equidistant from their common ancestor. One might approach the problem by standardizing distances such that each language connecting to a particular common ancestor is equally distant from that ancestor.

One way in which to standardize distances would be to determine for each node on the graph the longest path that can reach it. In Figure 1, the lengths of the longest paths are shown next to each node that serves as a common ancestor for a pair of living languages. Thus, the longest path that reaches West Germanic is five links long—the path originating at Standard German. One can use this longest path as the standardized distance to a common ancestor. Thus, the distance between English and West Germanic is now five, rather than two, and the distance between Armenian and Indo-European is now ten, rather than two.

The following formula provides a similarity measure between any pair of languages:

(3) $$S_{rk} = \frac{\partial_x - \partial_{rk} + 1}{\partial_x + 1}$$

where S_{rk} is the similarity between language *r* and language *k*, ∂_x is the length of the longest path in the language family (i.e., the length of the longest path to the common ancestor of the entire family), and ∂_{rk} is the length of the longest path to the nearest common ancestor of language *r* and language *k*. Thus, the similarity between English and German is (10-5+1)/(10+1)=0.5455, the similarity between English and Armenian is (10-10+1)/(10+1)=0.0909, and English's self-similarity is (10-0+1)/(10+1)=1.0. Proceeding in this way, one can calculate a proximity measure between each pair of languages within the same major language family. Since there are no links among the major language families, each language has a similarity of zero with languages outside its own family. The similarity between each language will thus always lie between zero and one. It will equal zero only when the two languages lie in different families, and will equal one only when a language is compared with itself.

The Summer Institute of Linguistics has produced a database called *Ethnologue*, that catalogs information on all of the world's languages. The database presents both taxonomic relationships and the number of speakers in each country (Grimes 2000). A few modifications were made to the *Ethnologue* taxonomy. The most important was to eliminate "Creole" as a separate family, and instead to treat each creole language as a sister language to the language from which it derives most of its vocabulary. In addition, unclassified languages, deaf sign languages, and languages with fewer than 1,000 speakers were eliminated. The linguistic proximity w_{ij} between countries *i* and *j* is calculated as follows:

(4) $$w_{ij} = \sum_r \sum_k p_{ik} p_{jr} S_{rk}$$

where p_{ik} is the percentage of the population in country *i* speaking language *k*, p_{jr} is the percentage of the population in country *j* speaking language *r*, and S_{rk} is the similarity measure between language *r* and language *k*. Thus, every language in country *i* is compared to every language in country *j*. Intuitively, the measure gives the expected similarity of the languages spoken by two persons, one drawn at random from each country. High values of w_{ij} will occur only when both countries have a high percent of their population in similar languages.[1]

---

[1] Note that a country's self-similarity w_{ii} provides a measure related to Ethno-Linguistic Fractioning (Easterly and Levine 1997). That measure is based on the Herfindahl Index ($ELF = 1 - \sum_k p_k^2$) and is the probability that two persons drawn at random from the population speak *different* languages. The diagonal of the language phylogeny matrix, however, gives the *expected similarity* of the languages spoken by two persons drawn at random from the population.

The diagonal of the proximity matrix **W** is set to zero, and the matrix is then standardized by dividing each element w_{ij} by the largest of the elements w_{ij} so that each w_{ij} ranges from zero to one.

Cultural Proximity II (Religious Proximity)

Samuel P. Huntington has made an influential argument that national cultures can be grouped into a taxonomy of perhaps eight or nine "civilizations" (Huntington 1997), and that these civilizations are primarily centered around religion. The most dangerous conflicts among nations will occur along the fault lines between civilizations, since members of the same civilization usually see each other as natural allies. Huntington's thesis suggests that a pair of nations may have similar levels of economic development because of cultural similarity—a cultural similarity that may not be reflected in linguistic similarity, but in religious similarity. Thus Afghanistan and Yemen, though speaking unrelated languages (Afghan languages are either Altaic or Indo-European, Yemeni languages are Afro-Asiatic) do have a degree of cultural similarity because of their similar religion.

Huntington's approach, followed by numerous other scholars, has been to assign each nation to a single civilization.[2] Thus a nation such as Lebanon, which contains adherents of religions forming the core of three Huntington civilizations (Western Christian, Orthodox Christian, Islam) is placed in the Islamic civilization. A more nuanced view of religious proximity, however, would not ignore the potential cultural affinities arising from minority religions.

The CIA World Factbook (United States, Central Intelligence Agency 2003) provides estimates of the percentage of each nation's population that adheres to the major world religions. The figures were modified in a few ways. Most importantly, percentages for "Atheist" and "None" were dropped, and percentages recalculated so that the sum would equal 100 percent. This was done since religion is typically an ancient cultural trait in each nation, while atheism is recent; focusing solely on religion therefore provides a better sense of deep cultural affiliations among nations. In addition, Protestant and Roman Catholic percentages were combined into Western Christian, and Buddhist and Taoist percentages were combined into Buddhist. The final religious categories are as follows:

1. Western Christian
2. Hindu
3. Buddhist

[2] Huntington does treat some nations as mixed, such as the Philippines, Guyana, Surinam, India, Sri Lanka, and portions of East and West Africa (Huntington 1997: 26-27).

4. Orthodox Christian
5. Muslim
6. Jewish
7. Indigenous

The religious proximity w_{ij} between countries i and j is calculated as follows:

(5) $$d_{ij} = \sqrt{\sum_{k}^{7} \left(p_{ik} - p_{jk}\right)^2}$$

(6) $$w_{ij} = \left(1 + d_{ij}\right)^{-2} \quad \text{and} \quad w_{ii} = 0$$

where p_{ik} is the percentage of country i's population adhering to religion k, and p_{jk} is the percentage of country j's population adhering to religion k. The distance measure d_{ij} is thus a seven-dimensional Euclidian distance between country i and country j. The squared inverse of d_{ij} is used to create the proximity matrix **W**, whose diagonal is set to zero, and the matrix is then standardized by dividing each element w_{ij} by the largest of the elements w_{ij} so that each w_{ij} ranges from zero to one.

Cultural Proximity III (Huntington Civilizations)

Samuel P. Huntington's division of nations into civilizations can also be directly employed in a proximity matrix, and in fact was used by Beck and Gleditsch (2003) in their paper employing spatial statistics to address autocorrelation of regression residuals in cross-national data sets. Huntington presented his classification in the form of a map, rather than a detailed country-by-country listing based on specific criteria (Huntington 1997: 26-27). A number of studies have employed Huntington's categories in statistical work, and have therefore drawn up lists specifically assigning states to civilizations (Henderson and Tucker 2001; Russett, Oneal, and Cox 2000; Beck and Gleditsch 2003). These efforts at classification all differ somewhat, but the classification of Russett, Oneal, and Cox (2000) will be used as a basis here, since these authors deviate only slightly from Huntington's map when assigning nations to civilizations.[3] This list is further modified, most importantly to assign states omitted from the list to their appropriate civilizations: these omitted states are Fiji, Mauritius, the former Soviet states, and the former states of Yugoslavia. In addition, Haiti is reassigned from the category of "Lone" state to the

[3] In addition, Huntington, in his published response to the Russett, *et al.* article (source), does not object to their classification scheme.

category of "African" state.[4] The relationship between a pair of states is modeled simply as a binary variable:

(7) $w_{ij} = 1$ if i and j are in the same civilization, $w_{ij} = 0$ otherwise.

The diagonal of proximity matrix **W** is set to zero, and the matrix is symmetric, with elements w_{ij} either equal to zero or to one.

Cultural Proximity IV (Colonization and Imperialism)

Many nations may share cultural characteristics since they share a common history of domination by a particular colonial or imperial power. Thus, for example, the legal codes of India and Nigeria are similar because English colonialists developed the basic legal institutions for each country, and Taiwan and Korea have similar educational systems since both were set up by Japanese colonialists. Colonial ties may also evolve into more durable cultural and economic ties, as they have in Francophone Africa, where Paris serves as the font of cultural innovation and French businesses constitute the most important source of foreign direct investment.

Distinguishing the most important colonial and imperial relationships is somewhat subjective. German colonialism, for example, was short-lived, coming to an end during the first World War. Nevertheless, the identity of a state such as Cameroon is based on German boundaries, and the tenure of Germans there was almost as long (36 years) as was the tenure of their French and British successors (44 years). The Ottoman, Russian, and Austrian empires dominated subject nations, often for hundreds of years, and when these subject nations became states, they shared a cultural legacy from that time of domination. Thus, the relationship of Bulgaria to Turkey is similar to the relationship of Slovakia to Austria. But is Norway related to Denmark or Sweden? Is Finland related to Sweden or Russia? If one doesn't simply pick the most recent colonial or imperial relationship, then one must determine the time horizon in which the significant relationships lie: should it be the previous hundred years, or thousand years? A reasonable answer may be about 300 years, which includes most of the relationships of European world expansion[5], as well as the relationships of the Soviet "empire." The time horizon excludes, however, early

---

[4] Since Surinam is classified as an African state, and Haiti seems more "African" than Surinam (Surinam has a large Asian population), it seems reasonable to reclassify Haiti as African. There are a number of other questionable assignments in the Russett, *et al.* classification: Belize is "Western," while the Bahamas is "Latin American," and much of the southwestern Pacific (Melanesia, Micronesia, and the Philippines) is classified as "Western." Nevertheless, these assignments are allowed to stand.

[5] The most notable exception would be Portuguese territories in southeast Asia and southern Africa, most of which were lost to the Dutch during the 17th Century.

empires which had powerful influences on the culture of subject nations, such as the Roman domination of Western Europe, and the Caliphate's domination of North Africa, the Near East, and Persia.

The relationship between a pair of states is modeled simply as a binary variable:

(8) w_{ij} =1 if i and j were dominated by the same power, w_{ij} =0 otherwise.

The diagonal of proximity matrix **W** is set to zero, and the matrix is symmetric, with elements w_{ij} either equal to zero or to one.

Ecological Proximity

An old strand of thought in the social sciences (Friedrich Ratzel 1882; Semple 1911) holds that humans at higher latitudes face greater challenges from the environment, and must therefore devote more effort to developing their material culture. Materialist perspectives, whether Marxist or otherwise, similarly maintain that the ecological environment in which a people are situated will determine many features of their material culture (Harris 1979, Seward 1968). One might reasonably expect, then, that countries at a similar level of economic development might be similar not because they are physically or culturally proximate, but because they are ecologically similar.

The World Wildlife Fund has produced a detailed map of the earth's ecological regions (Olson, et al. 2001).[6] The map divides the planet's land area into 867 unique ecoregions, and each of these are classified into one of 14 biomes. Thus the planet's surface can be classified into the following 16 categories:

[6] The website http://www.worldwildlife.org/ecoregions/dbaserequest.htm distributes the map as an ArcView Shapefile.

1. Boreal forests/taigas
2. Deserts and xeric shrublands
3. Flooded grasslands and savannas
4. Mangroves
5. Mediterranean scrub
6. Montane grasslands and shrublands
7. Temperate broadleaf and mixed forests
8. Temperate coniferous forests
9. Temperate grasslands, savannas, and shrublands
10. Tropical and subtropical coniferous forests
11. Tropical and subtropical dry broadleaf forests
12. Tropical and subtropical moist broadleaf forests
13. Tropical and subtropical grasslands, savannas, and shrublands
14. Tundra
15. Snow, ice, glaciers, and rock
16. Water

Using a GIS, the World Wildlife Fund biome theme is overlain with a theme of national borders. One can then calculate the percent of each country's area for each of the above 16 categories. The ecological proximity w_{ij} between countries i and j is calculated as follows:

(9) $$d_{ij} = \sqrt{\sum_{k}^{16} \left(p_{ik} - p_{jk}\right)^2}$$

(10) $$w_{ij} = \left(1 + d_{ij}\right)^{-2} \quad \text{and} \quad w_{ii} = 0$$

where p_{ik} is the percentage of country i's area in biome type k, and p_{jk} is the percentage of country j's area in biome type k. The distance measure d_{ij} is thus a 16-dimensional Euclidian distance between country i and country j. The squared inverse of d_{ij} is used to create the proximity matrix **W**, whose diagonal is set to zero, and the matrix is then standardized by dividing each element w_{ij} by the largest of the elements w_{ij} so that each w_{ij} ranges from zero to one.

Ally Relationships

Political scientists have long sought to understand the causes of international conflict. In recent years, empirical work has been facilitated by the compilation of large data sets. One of the most productive of these research efforts has been the Correlates of War (COW) Project, begun in 1963 by J. David Singer and Melvin Small. The data sets compiled by the COW project include the Militarized Interstate Dispute collection, which catalogs the events between 1816 and 2001 in which a state either used or threatened to use force against another state (Ghosn and Palmer 2003).

Two nations that frequently ally with each other during conflicts may already have particularly close economic and cultural ties, as for example is the case with the United Kingdom and the United States, or they may be compelled—because of their alliance—to coordinate economic policy, introduce similar industrial standards, or have similar views on the role of the state and human rights. Thus, for example, the long-standing alliance between Cuba and Russia was accompanied by Cuban adoption of Soviet methods in economic policy. One might believe, then, that two nations might have similar economies because they are allies.

Ally relationships can be of two kinds: either nations sign a formal agreement, or they can be on the same side during the same interstate militarized dispute. Here we look at the second type of ally relationship. The Militarized Interstate Dispute collection allows one to identify all nations that stood on the same side of a militarized conflict. The database presents the level of hostility of the conflict, the duration of the conflict, and the date of the conflict. The first step in forming an alliance score is determining the importance of each conflict to each participant. The importance of conflict *k* to participant *i* is found as follows:

(11) $$x_{ik} = e^{h_k} \ln(d_k + 1)\, e^{\left(\frac{\ln(0.5)}{66}(2001 - y_k)\right)}$$

where h_k is the level of hostility of conflict *k*, d_k is the duration—in days—of conflict *k*, and y_k is the year in which conflict *k* occurred. The level of hostility h_k takes on integer values from one to five: (1) no militarized action, (2) threat of force, (3) display of force, (4) use of force, and (5) war. As Figure 4 shows, using h_k as the exponent gives relatively more weight to actual use of force. Duration of hostility, on the other hand, is logged, since an additional day in a short conflict should weigh more heavily than an additional day in a long conflict (see Figure 3). The last term in Equation 11 is the age-decay component (Figure 2). A conflict a hundred years ago is of less salience than a conflict within the past decade, and should be weighted less. The formula gives a half-life of 66 years, reasoning that in two generations the conflict would be half as important as it was when actually occurring.

The importance as an ally of a particular country *j* to country *i* is found by taking the geometric mean of x_{ik} and x_{jk} for each conflict *k* in which the two countries were on the same side, and then summing all conflicts:

(12) $$w_{ij} = \sum_k \left(x_{ik} x_{jk}\right)^{\frac{1}{2}}$$

The diagonal of the proximity matrix **W** is of some interest since it serves as a measure of the involvement of each country in militarized disputes. However, for purposes of creating a proximity index, the diagonal is irrelevant, and is set to zero. The matrix is then standardized by dividing each element w_{ij} by the largest of the elements w_{ij} so that each w_{ij} ranges from zero to one.

Enemy Relationships

Nations often emulate their more powerful enemies, as for example, Peter the Great's administrative reforms (which were efforts to imitate Sweden), the Meiji reforms of Japan, and the Kemalist reforms of Turkey (Finer 1999: 1414). Enemies are usually neighbors, as well, and one might reasonably think that enemies have very similar economies. Of interest here, as in quantifying ally relationships above, are *long-term* enemy relationships, based not on the events of the past decade, but on the experience of two centuries. Thus, despite the amity of the last half century, much of Europe continues to be wary of Germany, and much of East Asia wary of Japan. The powerful neighbor, who historically has been an enemy, inspires respect and receives a great deal of attention, and is thus often emulated.

Equations 11 and 12 above provide the basis for calculating the strength of enmity relationships between each pair of countries. The only modification would be to Equation 12, where rather than summing over all conflicts in which countries i and j are on the same side, one sums over all conflicts in which they are on *opposite* sides. The diagonal of the proximity matrix **W** is set to zero, and the matrix is then standardized by dividing each element w_{ij} by the largest of the elements w_{ij} so that each w_{ij} ranges from zero to one.

Formal Alliance Relationships

The COW project has also compiled a collection of data on treaty relationships among nations. The Formal Alliance collection catalogs formal interstate agreements between 1816 and 2000. The agreements are of three types: defense pacts, nonaggression pacts, and ententes. The first binds signatories to mutual defense, the second binds signatories to never use force against each other, and the last simply requires signatories to consult each other during periods of crisis. (Gibler and Sarkees Forthcoming).

The data contain one record for each year in which a pair of countries are linked by a particular formal agreement. One can readily calculate the degree to which a pair of nations are attached by formal alliances. Equation 13 calculates the importance of agreement k in year y to country i:

(13) $$x_{iky} = e^{(4-h_k)} e^{\left(\frac{\ln(0.5)}{66}(2001-y)\right)}$$

where h_k is the level of military commitment of agreement *k*, and *y* is a year in which agreement *k* is in force. The level of military commitment h_k takes on integer values from one to three: (1) defense pact, (2) nonaggression pact, and (3) entente. Using $-h_k$ in the exponent gives relatively more weight to mutual defense treaties. The second term in Equation 13 is the age-decay component, identical to that found in Equation 11 and portrayed in Figure 2.

The importance as an ally of a particular country *j* to country *i* is found by taking the geometric mean of x_{iky} and x_{jky} for each agreement *k* in each year *y* in which the two countries were bound by the agreement, and then summing all agreements and years:

(14) $$w_{ij} = \sum_k \sum_y \left(x_{iky} x_{jky}\right)^{\frac{1}{2}}$$

The diagonal of the proximity matrix **W** is set to zero, and the matrix is then standardized by dividing each element w_{ij} by the largest of the elements w_{ij} so that each w_{ij} ranges from zero to one.

Trade Relationships

Nations that trade heavily with each other may have similar economies, due to the transmission of economic fluctuations through trade and the need to adapt to the standards of trading partners. Trade relations may be especially important in understanding autocorrelation in economic models employing international datasets. Beck and Gleditsch (2003) employ a trade matrix as one of their three spatial weight matrices.

The trade data employed here come from the International Monetary Fund's Direction of Trade CD-ROM, and were subsequently modified by Andrew Rose, who generously makes his data available on the web.[7] Rose deflates each trade flow (in nominal U.S. dollars) by the U.S. CPI, and then averages all four flows between each pair of countries (country A's exports, country B's exports country A's imports, country B's imports) to create a single trade value for each pair of countries. This single trade value is then logged. Rose's data are annual; to create the proximity matrix **W**, the average trade value for each pair of countries is averaged for the years 1990 through 1999. The resulting matrix is symmetric, with a diagonal of zero values. The matrix is

[7] The data are available at http://faculty.haas.berkeley.edu/arose/RecRes.htm, and are described in Rose (Forthcoming).

then standardized by dividing each element w_{ij} by the largest of the elements w_{ij} so that each w_{ij} ranges from zero to one.

Similarity in Level of Development

Nations at a similar level of development may be similar in a host of economic and social metrics. This view is little more than a conventional assumption implicit in many cross-national economic studies. Just as Herbert Spencer (1897) posited that social evolution is best summarized as increasing complexity, and that increasing complexity leads to regular and predictable changes in social structure, so do economists often posit that increasing economic development is accompanied by regular and predictable changes in economic institutions.

The United Nations' usual indicator of level of development—the Human Development Index—has three components: per capita GDP, life expectancy, and an education measure that combines the literacy rate with educational spending. The Human Development Index measures a nation's achievements on three important dimensions of development: life, knowledge, and prosperity (United Nations Development Program 2003). However, educational spending can be reasonably classified as instrumental, rather than as a measure of a desired outcome. Therefore, the dimensions of development are modified slightly to include the three components: life, liberty, and prosperity. The first is operationalized as average life expectancy, the second as the geometric mean of standardized measures of liberty from Freedom House and the Heritage Foundation, and the last as per capita GDP. Life expectancy is taken from the World Bank World Tables,[8] employing the average value for the years 1991 through 2000. Three measures of liberty are taken from Freedom House: Civil Liberties, Political Rights, and Press Freedom. The scores for each of these are based on a survey (Freedom House 2004). The average score between 1995 and 2001 is used for each of these. One measure of liberty is taken from the Heritage Foundation: Economic Liberty, which is a composite index based on ten measures of economic freedom (Miles *et al.* 2004). The average score for the years 1995 through 1999 is used for Economic Liberty. Each of the four scores is standardized and the geometric mean calculated, to produce an overall liberty score. Finally, per capita GDP (in 1995 US dollars) is taken from the World Bank World Tables,[9] employing the average value for the years 1991 through 2000.

Each of the three measures is standardized with a mean of 100 and a standard deviation of 15. Any country therefore has a specific location given by its scores for life, liberty, and prosperity

---

[8] The variable name in the World Tables is SP.DYN.LE00.IN.

[9] The variable name in the World Tables is NY.GDP.PCAP.KD.

within a three-dimensional space. The distance in level of development between countries i and j is calculated as follows:

(15) $$d_{ij} = \sqrt{\sum_{k}^{3} \left(x_{ik} - x_{jk}\right)^2}$$

where x_{ik} is the level of development of country i in dimension k, and x_{jk} is the level of development of country j in dimension k. The distance measure d_{ij} is thus a three-dimensional Euclidian distance between country i and country j. The proximity of country i and j is then calculated:

(16) $$w_{ij} = \left(1 + d_{ij}\right)^{-2} \quad \text{and} \quad w_{ii} = 0$$

The squared inverse of d_{ij} is used to create the proximity matrix **W**, whose diagonal is set to zero, and the matrix is then standardized by dividing each element w_{ij} by the largest of the elements w_{ij} so that each w_{ij} ranges from zero to one.

Dyadic Event Frequency

The interaction of nations is reported daily in the press. Using computational techniques, very large volumes of stories from news agencies such as Reuters can be coded to create databases of dyadic interactions among nations. Advocates of these methods maintain that the detail and daily frequency of these databases make them superior to aggregated, annual frequency databases such as those used by the Correlates of War project (King and Lowe 2003: 618). The data employed here are drawn from Reuters Business Briefings, and processed by Virtual Research Associates, Inc. The processed dyadic data contain over 2.8 million records, spanning from January 1991 through December 2000, and are generously made available by Gary King.[10]

These data are very rich and allow categorization of each interaction event as business, artistic, inter-governmental, etc. Thus, one could construct multiple interaction matrices (one for business interactions, another for artistic interactions, etc), and since each interaction event contains a date, one could also use an age-decay component, as was done earlier with the enemy and ally matrices. Nevertheless, here only the total interaction frequency is considered, without weighting for the age of the interaction event. The event frequency of country i and j is calculated as follows:

(17) w_{ij} = number of events where country i interacts with country j; w_{ii}=0

---

[10] The data are found at http://gking.harvard.edu/data.shtml .

The diagonal of the proximity matrix **W** is set to zero, and the matrix is then standardized by dividing each element w_{ij} by the largest of the elements w_{ij} so that each w_{ij} ranges from zero to one.

3. COMPARISON OF WEIGHT MATRICES

Weights could be found for 152 countries for each of the 12 measures. There are 27 countries in the World Bank World Tables that are not represented in the weight matrices. Together, these 27 excluded countries make up about three percent of the earth's population, and the included 152 countries make up about 97 percent.

Table 2 presents the matrix correlation between each pair of the 12 matrices described above. The significance level (from a permutation test) is given by the asterisks next to the correlation coefficient. Strikingly, with only three exceptions, the matrices are positively correlated with each other at the .90 level of significance. The three exceptions: physical distance, and allies are not correlated with the colonial-imperial matrix, and event frequency is not correlated with level of development.

Thus, overall, the weight matrices are positively correlated, but there is variation in the degree to which they are positively correlated. One can examine this variation to produce stylized facts regarding the ways in which these relationships overlap. For example, three of the culture matrices: language phylogeny, religion, and Huntington civilizations, are highly correlated with each other, while the fourth culture matrix—the colonial-imperial matrix—is more weakly correlated. The result suggests that a dominating nation often finds its colonies and subjects among nations that are not linguistically related and that are not coreligionists.

Some of the strong correlations are little more than tautologies. Event frequency is strongly correlated with trade and with allies, and also has fairly strong correlations with the enemies and formal treaty matrices. This result is predictable and not very interesting, since trade, allies, enemies, and formal treaty are the matrices that model interactions among nations, while the remaining matrices model similarities. In a similar vein, religion is highly correlated with Huntington Civilization—a trivial result since the Huntington classification is primarily based on religion.

Ignoring these trivial results, a few of the correlations are suggestive. Formal alliances are highly correlated with the cultural matrices, particularly with language phylogeny, implying that related nations are more likely to bind each other with formal agreements. Allies are highly correlated with trade, suggesting that states find their closest wartime allies among their most important trading partners. On the other hand, enemies are more highly correlated with allies than with any other matrix, suggesting that each state draws both enemies and allies from the same set of sister states, such that the relationship between a pair of states is likely to cycle between enmity and amity.

States at a similar level of development are likely to have similar ecologies and to have similar cultures, and though these correlations are not extremely strong, they are nevertheless stronger than the correlation between level of development and physical distance. Culture is more highly correlated with ecology than with physical distance. And so on—matrix correlation produces stylized facts that may suggest new lines of inquiry.

4. AUTOCORRELATION FOR EXPLORATORY DATA ANALYSIS

Table 3 presents 72 variables drawn from an array of international datasets. The data include psychometric variables, health and demographic variables, variables measuring diet and income inequality and a variety of other characteristics of nations. These data can be examined for autocorrelation, using the 12 weight matrices described above. By examining data representing many different categories of social life one might gain some sense of how autocorrelation might vary across these categories. In addition, by testing for autocorrelation across 12 different weight matrices, one might gain some sense of how the different weight matrices perform.

The autocorrelation statistic used here is Moran's I (Odland 1988; Anselin 1988):

$$I = \frac{n}{\sum_i \sum_j w_{ij}} \cdot \frac{\sum_i \sum_j w_{ij}\left(x_i - \bar{x}\right)\left(x_j - \bar{x}\right)}{\sum_i \left(x_i - \bar{x}\right)} \qquad (1)$$

Where w_{ij} is a weight representing the degree of relatedness between location i and j (greater relatedness implies a higher weight); n is the number of locations; x_i is the value of a variable at location i and x_j is the value of the same variable at location j. Intuitively, Moran's I differs from the usual correlation coefficient in that a correlation coefficient compares the values of two variables at each location, while Moran's I compares the value of a single variable for each pair of locations arrayed according to degree of relatedness. One can calculate a variance for a Moran's I, and then calculate a z-score. Alternatively, one can use simulation methods and calculate the Moran's I for random permutations of the variable vector, finding the distribution of the statistic. One can then reject or maintain the null hypothesis that there is no autocorrelation (Odland 1988; Anselin 1988).

Table 4 reports Moran's I for each of the 72 variables, using each of the 12 weight matrices. The immediately striking result is that most of the variables are highly autocorrelated most of the time. Table 5 summarizes Table 4, showing that 76.0 percent of the 864 autocorrelation tests (72 variables and 12 weight matrices) resulted in a p-value below 0.01; 88.7 percent of the 864 tests resulted in a p-value below 0.10.

Among the few variables showing little autocorrelation are a pair of financial variables: the interest rate spread between deposit and lending rates (prsprd) and FDI as a percentage of GDP (fdigdp). Two is too few, of course, to make any generalizations, but one could look at this with more examples to see if financial variables tend to be weakly autocorrelated.

The psychometric variables, with the exception of IQ, also show very weak autocorrelation. IQ is strongly autocorrelated, and this almost certainly is because Lynn and Vanhanen (2002) estimated IQ for a number of countries by simply taking averages of the IQ of countries that were geographically proximate and ethnically similar. Thus, the autocorrelation of IQ must be in part spurious. Hofstede's (2003) four psychometric measures are not highly autocorrelated using the four cultural weight matrices. This result suggests that Hofstede's measures present relatively ephemeral values rather than values that are deeply rooted in national cultures, a point made by some of Hofstede's critics (McSweeney 2002). Thus, one can see how the methodology can be used to examine the validity of the various questionnaire-based attempts at producing "dimensions of culture" (Hofstede 2003).

While psychometric measures are relatively weakly autocorrelated, life expectancy, the sex ratio, and the sexual division of labor are highly autocorrelated. Educational figures tend to be weakly correlated, while the fertility rate and indicators of child health are highly autocorrelated. The durable, deeply rooted bits of culture may in fact be habits regulating the body and regulating the relations between the sexes.

Moran's I can be used to test hypotheses of the form "Culturally (physically, etc.) proximate nations tend not to have similar ___," where the blank could be filled with: "GDP per capita," "expenditure on schools," "propensity to wage war," etc. For example, on the last row of Table 4, one can interpret the Moran's I figures as the test statistics for the null hypothesis: "Proximate nations tend not to have a similar record for engaging in war." One can then see that the null hypothesis is not rejected for two of the cultural matrices (language phylogeny and religion) and is only weakly rejected for the other two cultural matrices (Huntington civilization and colonial-imperial). The result could be used in *support* of an argument that the love of war does not vary systematically along religious and cultural lines.

The presence of autocorrelation in international datasets is clearly too widespread to ignore. For certain types of hypotheses, an autocorrelation test can directly be used as a test statistic, and can generate stylized facts. The next section shows how autocorrelation becomes an issue when specifying a regression model.

5. SPECIFYING A REGRESSION MODEL

The problems posed by autocorrelation of residuals in an OLS model are well known: the estimated coefficients are unbiased, though inefficient, but the standard errors are biased. Thus one cannot trust the t-statistics. In addition, the presence of autocorrelation can often signal the more serious problem of omitted variables. A pair of proximate observations may have similar residuals if omitted variables have similar values for the two observations. As one can see from Tables 4 and 5, most international variables are autocorrelated, suggesting that omitted variables are likely to be autocorrelated. Thus, autocorrelation in the residuals of regressions performed on international data, are likely to signal omitted variable problems. One should therefore test for autocorrelation of regression residuals, and if autocorrelation is present, one should consider that the problem may be due to omitted variables and attempt to remove the autocorrelation by respecifying the model.

As an example, this section examines the determinants of the fertility rate. A large body of evidence shows that humans tend to resemble their parents in such traits as the age at first pregnancy, inter-birth spacing, number of offspring, and parenting styles. The proximate determinants of fertility are therefore vertically transmitted.

Interspecies differences in number of offspring can be explained by a branch of evolutionary theory called life-history theory, which maintains that differences in mortality rates among species lead to differences in how reproductive effort is displayed by those species. Reproductive effort consists of mating effort and parenting effort. In general, the higher the mortality rate, the more reproductive effort consists of mating effort. Where mortality rates are low, reproductive effort emphasizes parenting effort. Body size, and the number of offspring vary together with the amount of parental investment: species with large bodies and few offspring have low mortality rates and emphasize parenting effort.

Humans are a species that have very high parental investment. Nevertheless, there is variation within the species, and a number of researchers have applied evolutionary perspectives to explain that variation. Two of the most influential papers in this field have been Belsky et al. (1991), who argue that a stressful early environment—typified by harsh parenting—predisposes children toward early and frequent reproduction, and Chisholm (1993) who maintains that the major source of environmental stress is high ambient mortality. These perspectives maintain that the natal environment provides the cues humans use to choose—albeit unconsciously—their reproductive strategy, where the choice is between emphasizing mating effort or emphasizing parenting effort. Thus, in examining the determinants of fertility rates, one should consider the natal environment of adult women, paying particular attention to the then-current mortality rates.

In addition, one should consider the costs of having children. Where females have the possibility of employment outside the home, the opportunity cost of having a child includes the foregone earnings. Where some of those costs can be transferred from the parents to the child—through child labor—one would expect higher fertility. Additional determinants of fertility would therefore include the percentage of women in the labor force and the percentage of children in the labor force.

Table 6 presents the results of an OLS regression on the average 1995-2000 fertility rates of 135 nations. At first glance, the model looks satisfactory: the coefficients are significant and have the expected sign, there is no problem with multicollinearity or heteroskedasticity, and the R^2 seems reasonably high. The potential endogeneity of two variables has been handled by creating instruments, and the other two variables are temporally lagged values and cannot be endogenous. Nevertheless, the Moran's I tests show that with 10 of the 12 weight matrices there is significant autocorrelation, and the RESET test suggests that the model is not correctly specified.

The finding of autocorrelation indicates that there are likely to be important omitted variables. The most common way to handle this problem is to create a spatially lagged dependent variable $\hat{\mathbf{y}} = \mathbf{W}\mathbf{y}$, where $\mathbf{y}$ is an $n \times 1$ vector of the dependent variable, and $\mathbf{W}$ is an $n \times n$ spatial weight matrix, with the rows standardized to sum to one. The spatially lagged variable $\hat{\mathbf{y}}$ will, however, be endogenous since if observation *i*'s value depends on the values of proximate nations, so do the values of proximate nations depend on the value of observation *i*. The simplest way around this problem is to substitute an instrument for $\mathbf{y}$.

Introducing $\hat{\mathbf{y}}$ as an independent variable in the model usually eliminates autocorrelation. If one were to use, for example, the spatial weight matrix for language phylogeny as $\mathbf{W}$, then the coefficient of $\hat{\mathbf{y}}$ can be interpreted as the effect on fertility of inherited culture. This interpretation, however, masks the specific inherited cultural traits which determine fertility—it tells us that inherited culture is a determinant, but doesn't tell us what about that inherited culture determines fertility (Mace and Pagel 1994). In addition, the use of multiple weight matrices brings forth a problem not encountered in spatial statistics, where only one weight matrix (physical proximity) is used —the problem of *which* spatially lagged independent variable to introduce. It is therefore preferable to use the spatially lagged variables sparingly, after first attempting to introduce other independent variables and trying various functional forms.

Table 7 presents the results after respecification. All terms are now logged, the adult mortality rate is now separated into a male and female mortality rate, and an interaction term is added between infant and adult male mortality. Two new variables are introduced: the temporally lagged average fertility rate, and the temporally lagged sex ratio for the population under the age of 14. The first of these variables can be interpreted as capturing additional, unknown features of the natal environment—apart from mortality rates—that conditioned adult women to adopt their reproductive strategies. The second variable provides some measure of the bargaining position of women when choosing mates (Posner 1992). If the sex ratio is high, then women have wider mate choices and would be likely to secure more assistance from mates in rearing their offspring. Thus, high sex ratios are likely to be associated with higher parental investment and lower fertility.

The respecified model performs better on the diagnostic tests: the R^2 is improved, heteroskedasticity is even less of a problem, the RESET test suggests the model is correctly specified, and there is no autocorrelation on any of the 12 dimensions. Multicollinearity is extremely high for the mortality figures, but the coefficients are significant, so the multicollinearity is not a problem.

The results show that as female mortality rates rise, fertility falls, do doubt due to the fact that fertility (the number of children the average woman will have in her lifetime) will fall if women tend to die before menopause. An additional factor may be that post-menopausal women often care for the offspring of their younger kinswomen, freeing these younger women to have more offspring. Thus, when post-menopausal women die young, reproductively active women will have longer inter-birth intervals.

When one considers the interaction term between adult male mortality and infant mortality, the results are both reasonable and suggestive (Figure 5). The highest fertility rates occur when both infant and adult male mortality is high, in agreement with Chisholm's (1993) thesis. When infant mortality is low, increases in adult mortality lead to only slight increases in fertility. When adult male mortality is low, infant mortality is negatively related to fertility: an indication, perhaps, that parents react to the high infant mortality of their childhood by increasing parental investment.[11] This result suggests that parents respond to the mortality they can do something about (infant mortality) by increasing parental investment, respond moderately to mortality impacting solely adult males, and respond to generalized mortality by increasing the number of offspring and reducing parental investment.

The example illustrates how the use of autocorrelation statistics can be used much like a RESET test to hunt for the best specification. But while a RESET test is only used for functional form (Wooldridge 2002:125), autocorrelation statistics can be used to guide a search for omitted variables. Here, though the initial model confirmed our expectations, it did not give the data much scope to modify our expectations. The respecification provided a more detailed picture of how mortality rates affect fertility, confirmed that features of the natal environment other than mortality rates affect fertility, and showed how fertility responds to the negotiating power of females.

5. Summary and Conclusions

The paper shows how autocorrelation in international datasets can be defined spatially, culturally, ecologically, and through the frequency and intensity of interactions such as trade, war, or alliance. Twelve autocorrelation weight matrices were developed. Each type of autocorrelation models similarity among nations, but similarity due to different processes: spatial autocorrelation models trait similarity due to diffusion across space; cultural autocorrelation models trait similarity due to common descent; ecological autocorrelation models trait similarity due to adaptation to similar environments, etc. Nevertheless, the autocorrelation weight matrices are overwhelmingly highly correlated with each other. Examining the correlation among the weight matrices led to a few interesting stylized facts, such as, for example, the fact that inter-state formal alliances are highly correlated with linguistic similarity.

Previous research has already recognized that autocorrelation can be defined spatially, culturally, or through trade interactions. In a series of four papers, Douglas R. White, Michael L. Burton, and Malcolm M. Dow present a very clear exposition of the cultural autocorrelation problem and show how spatial statistics can resolve that problem (White, Burton, and Dow 1981; Dow, Burton, and White 1982; Dow, White, and Burton 1982; Dow, Burton, Reitz, and White 1984).

[11] Since the observations are states, not parents, it would be more accurate to say that states react, rather than that parents react. Nevertheless, fertility decisions are in most states primarily in the domain of parents.

Like the present paper, these authors use a language phylogeny to model cultural similarity, though their focus was on comparing distinct homogeneous cultures, not on comparing contemporary states. Their method is refined in this paper to account for phylogenies with uneven numbers of nodes in each branch, and then applied to create inter-state measures of cultural similarity. In the only research that applies these ideas to contemporary states, Beck and Gleditsch (2003) examine autocorrelation defined by physical distance, cultural similarity, and trade flows. To model cultural similarity, these political scientists use Samuel P. Huntington's (1997) ideas to classify each state into about eight *civilizations*, based primarily on religious traditions. The weight matrix thus contains elements that are either zero (if the states are in different civilizations) or one (if the states are in the same civilization). The approach of Beck and Gleditsch (2003) to measuring cultural similarity is also employed in this paper, as is their approach of using trade flows.

Two cultural similarity matrices are developed here that are not found in the previous literature. The first is religious similarity, and the second is similarity based on similar experiences with colonial or imperial powers. Four similarity matrices are also developed that provide—in addition to trade—a view of the intensity and frequency of interaction among states. One of these models formal alliance frequency, another models *de facto* alliance in war, a third models enemy relationships in war, and the last models frequency of events recorded in the Reuter's news service. Finally, two additional similarity matrices were developed: ecological similarity and level of development similarity.

To assess the prevalence of autocorrelation in international data, a sample of 72 variables were drawn from a wide variety of sources. Moran's I was used to test these 72 variables, using the 12 weight matrices. Autocorrelation existed at the .95 level of significance about 86 percent of the time—a demonstration that autocorrelation is more likely than not in international data.

Autocorrelation tests such as Moran's I can be of use for the creation of stylized facts of the form "Culturally (physically, etc.) proximate nations tend to have similar ___," where the blank could be filled with: "GDP per capita," "expenditure on schools," "propensity to wage war," etc. Nevertheless, like a correlation coefficient, a Moran's I doesn't provide a rich and accurate view of the forces influencing "GDP per capita," "expenditure on schools," "propensity to wage war," etc. In most cases, the stylized facts are most useful during the phase of exploratory data analysis, prior to the development of a regression model.

The last section of the paper gave an example of how autocorrelation statistics can aid in the development of a regression model using international data. The model examined the determinants of fertility. A simple model was specified, containing variables suggested by theory. While most diagnostics suggested that the model had no serious problems, the residuals exhibited considerable autocorrelation. Since autocorrelation can be caused by autocorrelated omitted variables, and since our previous results suggest that omitted variables are very likely to be autocorrelated, the results were used as an informal signal of omitted variables. Guided by the RESET test and Moran's I, a new model was built with additional variables and a different functional form. While a spatially lagged dependent variable would have served well in the respecification, a spatially lagged dependent variable is not as informative as variables that provide information on the specific processes determining fertility, and hence should be used only after other options have been exhausted. In the present case, spatially lagged dependent variables were not necessary.

Autocorrelation measures provide information that can be used for model building, for exploratory data analysis, and for the production of stylized facts. With the recent upsurge of interest in the cultural determinants of economic growth (Harrison and Huntington 2000; Landes 1998), it has become imperative to develop methods that assist in identifying cultural relatedness among nations. Autocorrelation measures provide methods that are easily integrated into existing econometric techniques.

REFERENCES

Anselin, Luc. 1988. *Spatial Econometrics: Methods and Models.* Dordrecht: Kluwer Academic Publishers.

Beck, Nathaniel, and Kristian Gleditsch. 2003. "Space Is More than Geography." Paper presented at the Joint Meetings of the European Consortium for Political Research, Edinburgh, Scotland. March-April, 2003. http://polmeth.wustl.edu/papers/03/beck03.pdf

Belsky, J., L. Steinberg, and P. Draper. 1991. "Childhood Experience, Interpersonal Development, and Reproductive Strategy: An Evolutionary Theory of Socialization." *Child Development.* 62: 647-670.

Cavalli-Sforza, L. Luca, Paolo Menozzi, Alberto Piazza. 1994. *The History and Geography of Human Genes*. Princeton: Princeton University Press.

Chiang, Alpha C. 1984. *Fundamental Methods of Mathematical Economics*. New York: McGraw-Hill, Inc.

Chisholm, James S. 1993. "Death, Hope, and Sex: Life-History Theory and the Development of Reproductive Strategies." *Current Anthropology.* 34(1): 1-24.

Cliffe, A.D., and J.K. Ord. 1973. *Spatial Autocorrelation.* London: Pion Limited.

Diener, Ed and Eunkook Mark Suh. 1999. "National Differences in Subjective Well-Being." in Kahneman, Daniel, Ed Diener, and Norbert Schwarz, editors. *Well-Being: The Foundations of Hedonic Psychology*. New York: Russell Sage Foundation.

Dow, Malcolm M., Michael L. Burton, and Douglas R. White. 1982. "A Simulation Study of a Foundational Problem in Regression and Survey Research." *Social Networks*. 4:169-200.

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

Dow, Malcolm M., Douglas R. White, and Michael Burton. 1982. "Multivariate Modeling with Interdependent Network Data." *Behavior Science Research*. 17(2):216-245.

Easterly, William and Ross Levine. 1997. "Africa's Growth Tragedy: Policies and Ethnic Divisions." *The Quarterly Journal of Economics.* 112(4): 1203-1250.

Finer, Samuel E. 1999. *History of Government*. Oxford: Oxford University Press.

Freedom House. 2004. http://www.freedomhouse.org/research/freeworld/2000/methodology.htm Accessed February 2, 2004.

Ghosn, Faten, and Glenn Palmer. 2003. "Codebook for the Militarized Interstate Dispute Data, Version 3.0." Online: http://cow2.la.psu.edu.

Gibler, Douglas M., and Meredith Sarkees. (Forthcoming.) "Measuring Alliances: the Correlates of War Formal Interstate Alliance Data Set, 1816-2000." *Journal of Peace Research*.

GlobalBeauties.com. 2001. Miss World Ranking (1951-2001). Accessed February 2004. http://www.globalbeauties.com/gb/world/w_ranking.htm.

GlobalBeauties.com. 2002. Miss International Ranking (1960-2002). Accessed February 2004. www.globalbeauties.com/gb/international/i_ranking.htm .

Grimes, Barbara (editor). 2000. *Ethnologue: Languages of the World* (Fourteenth edition). Dallas: SIL International.

Harris, Marvin. 1979. *Cultural Materialism: The Struggle for a Science of Culture*. New York: Random House.

Harrison, Lawrence E., and Samuel P. Huntington (editors). 2000. *Culture Matters: How Values Shape Human Progress.* New York: Basic Books.

Hofstede, Geert. 2003. *Culture's Consequences : Comparing Values, Behaviors, Institutions and Organizations Across Nations* (second edition). London: Sage Publications, Ltd.

Hoy, Andrew R. 1994. "The Relationship Between Male Dominance and Militarism: Quantitative Tests of Several Theories."*World Cultures*8(2): 40-57

Huntington, Samuel P. 1997. *The Clash of Civilizations and the Remaking of World Order.* New York: Touchstone.

International Mathematical Olympiad. 2003. "Unofficial Country Rankings – IMO." http://imo.math.ca/results/CRBY.html . Accessed February 2004.

Kennedy, Peter. 1998. *A Guide to Econometrics* (4th edition). Cambridge: MIT Press.

King, Gary and Will Lowe. 2003. "An Automated Information Extraction Tool for International Conflict Data with Performance as Good as Human Coders: A Rare Events Evaluation Design." *International Organization.* 57(Summer): 617-642.

Landes, David. 1998. *The Wealth and Poverty of Nations.* New York: Norton.

Loftin, C. 1972. "Galton's problem as spatial autocorrelation: Comments on Ember's empirical test" *Ethnology* 11: 425-35.

Loftin, C. and S. Ward. 1983. "A spatial autocorrelation model of the effects of population density on fertility" *American Sociological Review* 48: 121-128.

Lynn, Richard, and Tatu Vanhanen. 2002. *IQ and the Wealth of Nations*. New York: Praeger Publishers.

Mace, Ruth and Mark Pagel. 1994. "The Comparative Method in Anthropology." *Current Anthropology*. 35(5): 549-564.

McSweeney, Brendan. 2002. "Hofstede's Model of National Cultural Differences and Their Consequences: A Triumph of Faith - A Failure of Analysis." *Human Relations*. 55(1):89-118.

Miles, Marc A., Edwin J. Feulner, Jr., Mary Anastasia O'Grady, and Ana I. Eiras. 2004. *Index of Economic Freedom, 2004*. Washington D.C.: Heritage Foundation.

Murdock, George Peter and Douglas R. White. 1969. "A Standard Cross-Cultural Sample." Ethnology 9:329-369.

Naroll, R. 1965. "Galton's problem: The logic of cross cultural research." *Social Research* 32: 428-51.

Odland, John. 1988. *Spatial Autocorrelation.* Newbury Park, California: Sage Publications.

Olson, David M. Eric Dinerstein, Eric D. Wikramanayake, Neil D. Burgess, George V.N. Powell, Emma C. Underwood, Jennifer A. D'Amico, Illanga Itoua, Holly E. Strand, John C. Morrison, Colby J. Loucks, Thomas F. Allnutt, Taylor H. Ricketts, Ymiko Kura, John F. Lamoreaux, Wesley W. Wettengel, Prashant Hedao, and Kenneth R. Kassem. 2001. "Terrestrial Ecoregions of the World: A New Map of Life on Earth." *BioScience* 51(11): 933-938.

Posner, Richard A. 1992. *Sex and reason*. Cambridge: Harvard University Press.

Ratzel, Friedrich. 1882. *Anthropogeographie*. Stuttgart: J. Engelhorn.

Rose, Andrew. Forthcoming. “Which International Institutions Promote International Trade?” *Review of International Economics*.

Russett, Bruce, John R. Oneal, and Michaelene Cox. 2000. “Clash of Civilizations, or Realism and Liberalism Déjà vu? Some Evidence.” *Journal of Peace Research.* 37(5): 538-608.

Scott, John. 2000. *Social Network Analysis: A Handbook* (second edition). London: Sage Publications, Ltd.

Semple, Ellen Churchill. 1911. *Influences of Geographic Environment, on the Basis of Ratzel's System of Anthropo-Geography*. New York: H. Holt and co.

Spencer, Herbert. 1897. *The Principles of Sociology* (third edition). New York: D. Appleton.

Steward, Julian H. 1968. “Cultural Ecology.” *International Encyclopedia of the Social Sciences*. David L. Sills, editor, New York, Macmillan Company: v.4, pp.337-344.

Stocking, George W. Jr. 1968. “Edward Burnett Tylor.” *International Encyclopedia of the Social Sciences*. David L. Sills, editor, New York, Macmillan Company: v.16, pp. 170-177.

Transparency International. 2004. *Corruption Perceptions Index.* Accessed February 2004. http://www.transparency.org/pressreleases_archive/2003/2003.10.07.cpi.en.html

United Nations Development Program. 2003. *Human Development Report 2003. Millennium Development Goals: A Compact Among Nations to End Human Poverty.* Accessed February 2004. http://www.undp.org/hdr2003/pdf/hdr03_complete.pdf .

United States, Central Intelligence Agency. 2003. *CIA World Factbook 2003*. Accessed September 2003. http://www.cia.gov/cia/publications/factbook/

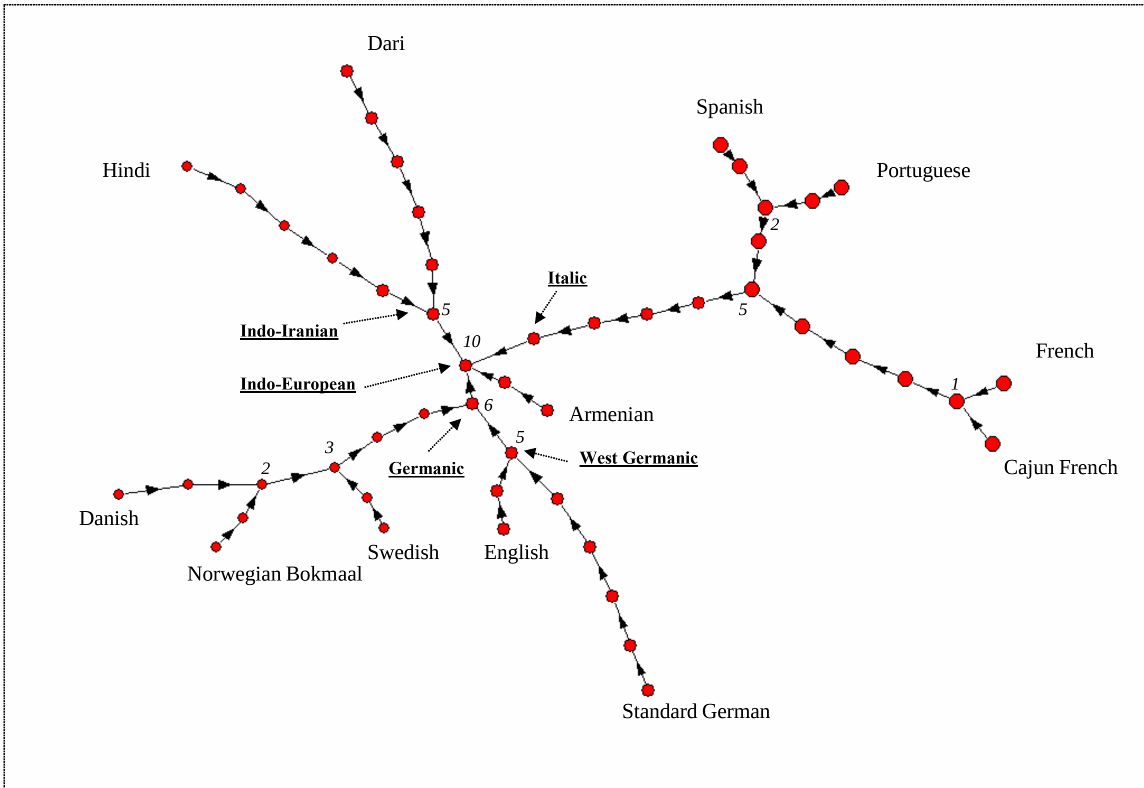
White, Douglas R., and Michael L. Burton, Malcolm M. Dow. 1981. “Sexual Division of Labor in African Agriculture: A Network Autocorrelation Analysis.” *American Anthropologist.* 83:824-849.

Wooldridge, Jeffrey W. 2002. *Econometric Analysis of Cross Section and Panel Data.* Cambridge: The MIT Press.

World Bank, International Economics Dept. 2002. *World Tables of Economic and Social Indicators, 1960-2000*. CD-ROM. Washington, DC: World Bank.

World Football Elo Ratings. 2004. http://www.eloratings.net/system.html. Accessed February 2004.

FIGURE 1: THE RELATIONSHIPS AMONG 12 LANGUAGES IN THE INDO-EUROPEAN FAMILY



Notes: The 12 selected languages are on the periphery of the digraph. Links point toward higher taxonomic levels, with all nodes ultimately connected to the node labeled Indo-European. The numbers indicate for selected nodes the maximum path length leading to that node. The taxonomy is from Grimes 2000.

FIGURE 2: THE AGE-DECAY COMPONENT OF ALLY AND ENEMY RELATIONSHIPS

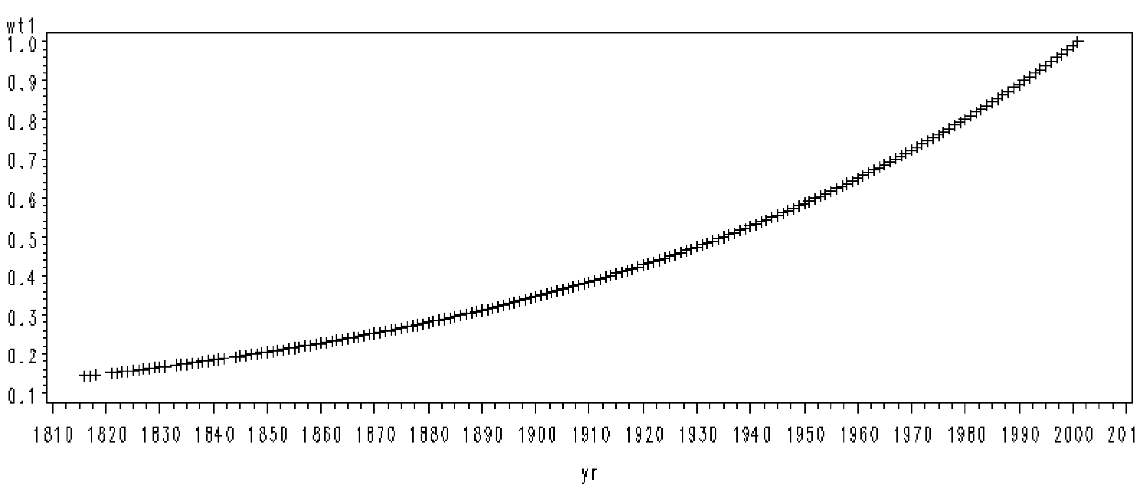


FIGURE 3: THE DURATION COMPONENT OF ALLY AND ENEMY RELATIONSHIPS

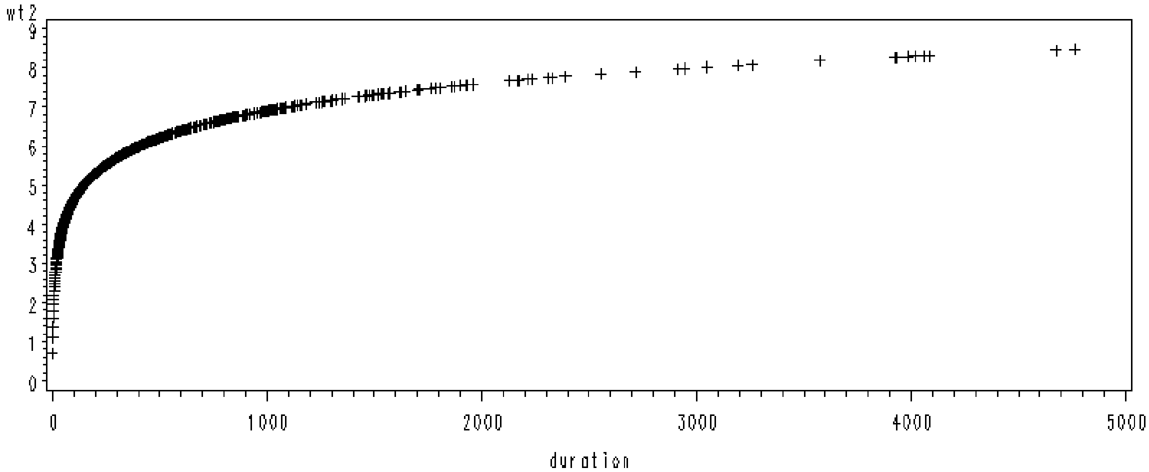


FIGURE 4: THE LEVEL OF HOSTILITY COMPONENT OF ALLY AND ENEMY RELATIONSHIPS

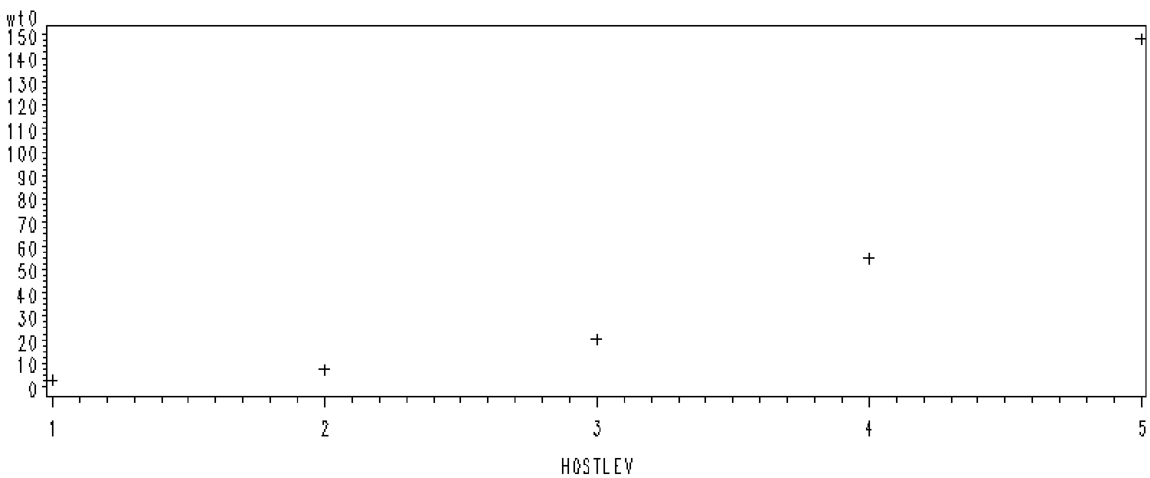


FIGURE 5: THE EFFECT OF MORTALITY RATES ON FERTILITY (MODEL 2)

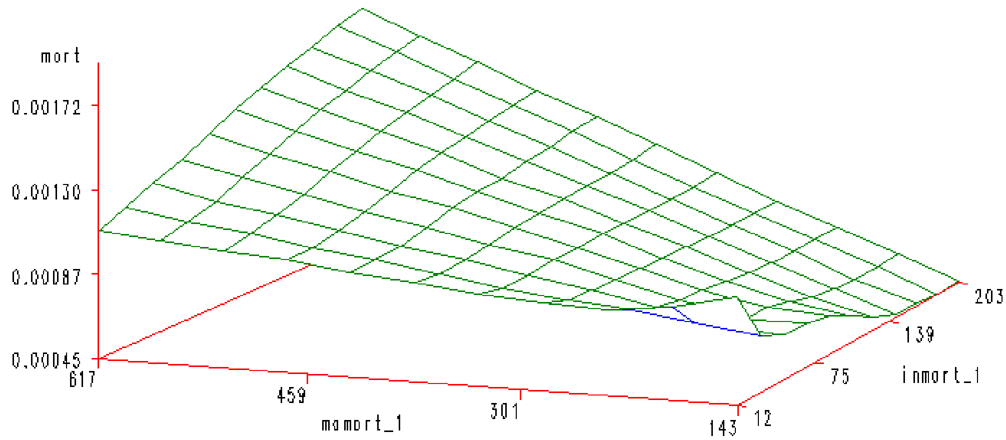


TABLE 1: TOP FIVE RELATED NATIONS FOR JAPAN, TURKEY, AND HAITI, FOR EACH OF THE 12 WEIGHT MATRICES

Level of Development	Trade	Ecology	Physical Distance	Language Phylogeny	Religion	Huntington Civilization	Colonial-Imperial	Formal Treaty	Allies	Enemies	Event Frequency
Japan											
Switzerland	USA	France	Korea	Korea	Cambodia		Philippines	USA	Italy	Russia	USA
Luxembourg	China	Macedonia	China	China	Mongolia		Germany	UK	France	China	China
Norway	Korea	Ireland	Philippines		Thailand		Korea	Thailand	USA	France	UK
Denmark	Germany	Poland	Mongolia		China		USA	Russia	UK	USA	Russia
Germany	Singapore	UK	Laos		Vietnam			Germany	Romania	Bulgaria	Korea
Turkey											
Ukraine	Germany	Chile	Romania	Turkmenistan	Bahrain	Albania	Syria	Greece	USA	Greece	USA
Russia	USA	Croatia	Bulgaria	Azerbaijan	Kuwait	Syria	Cyprus	UK	UK	Russia	Greece
Colombia	Italy	USA	Greece	Uzbekistan	Saudi Arabia	Turkmenistan	Romania	France	France	Italy	Iran
Fiji	France	New Zealand	Macedonia	Kyrgyzstan	Kyrgyzstan	Azerbaijan	Bulgaria	USA	Greece	France	Germany
Jordan	UK	Romania	Moldova	Kazakhstan	Iran	Iran	Greece	Italy	Canada	Cyprus	Russia
Haiti											
Gambia	USA	Dom. Rep.	Dom. Rep.	France	Venezuela	Cote d'Ivoire	Congo Rep.	Venezuela		USA	USA
Cambodia	Malaysia	Philippines	Jamaica	Mauritius	Colombia	Gabon	Cameroon	Colombia		Dom. Rep.	Dom. Rep.
Yemen	Dom. Rep.	Nicaragua	Bahamas	Dom. Rep.	Argentina	Togo	Cambodia	Argentina		UK	Colombia
Cameroon	France	Jamaica	Venezuela	Uruguay	France	Congo Rep.	France	Dom. Rep.		Canada	France
Congo Rep.	Japan	Laos	Panama	El Salvador	Dom. Rep.	Cameroon	Canada	USA		France	Canada

Notes: Each list gives the five most related countries, in order of descending relatedness. For matrices where top countries are tied, countries are sorted by average relatedness. Ties are most prevalent for Huntington Civilization, Religion, and Colonial-Imperial. For matrices with more than five tied top countries, only the first five (sorted by average relatedness) are shown. For matrices with fewer than five countries with non-zero weights, only those with non-zero weights are shown.

TABLE 2: MATRIX CORRELATIONS FOR 12 WEIGHT MATRICES

Matrix	Allies	Huntington Civilization	Colonial-Imperial	Ecology	Physical Distance	Event Frequency	Language Phylogeny	Formal Treaty	Level of Development	Religion	Trade	Enemies
Allies		0.1343***	0.0287	0.1161***	0.0167**	0.4965***	0.1314***	0.3768***	0.0346***	0.0924***	0.4268***	0.3301***
Huntington Civilization	0.1343***		0.1503***	0.2407***	0.0493***	0.0628***	0.4454***	0.3145***	0.1499***	0.5475***	0.0808***	0.0662***
Colonial-Imperial	0.0287	0.1503***		0.0945***	0.0177	0.0611***	0.2396***	0.1709***	0.0655***	0.0673***	0.0249*	0.0514***
Ecology	0.1161***	0.2407***	0.0945***		0.0493***	0.0579***	0.2371***	0.2299***	0.1212***	0.1365***	0.0746***	0.1045***
Physical Distance	0.0167**	0.0493***	0.0177	0.0493***		0.0098**	0.0667***	0.0292**	0.0147**	0.0195**	0.0087**	0.0520***
Event Frequency	0.4965***	0.0628***	0.0611***	0.0579***	0.0098**		0.0932***	0.1858***	0.0133	0.0491**	0.6899***	0.3017***
Language Phylogeny	0.1314***	0.4454***	0.2396***	0.2371***	0.0667***	0.0932***		0.4903***	0.1290***	0.4163***	0.0853***	0.1030***
Formal Treaty	0.3768***	0.3145***	0.1709***	0.2299***	0.0292**	0.1858***	0.4903***		0.1158***	0.3256***	0.2539***	0.1780***
Level of Development	0.0346***	0.1499***	0.0655***	0.1212***	0.0147**	0.0133	0.1290***	0.1158***		0.0931***	0.0293**	0.0179**
Religion	0.0924***	0.5475***	0.0673***	0.1365***	0.0195**	0.0491**	0.4163***	0.3256***	0.0931***		0.0780***	0.0454***
Trade	0.4268***	0.0808***	0.0249*	0.0746***	0.0087**	0.6899***	0.0853***	0.2539***	0.0293**	0.0780***		0.2123***
Enemies	0.3301***	0.0662***	0.0514***	0.1045***	0.0520***	0.3017***	0.1030***	0.1780***	0.0179**	0.0454***	0.2123***	

Notes: Matrix correlation coefficient (Wasserman and Faust 1994: 686). Significance levels reported for results from permutation test (10,000 permutations): '***'=p-value below 0.01; '**'=p-value below 0.05; '*'=p-value below 0.10.

TABLE 3: DESCRIPTIVE STATISTICS FOR SAMPLE OF VARIABLES

Category	Variable	Description	N	Maximum	Minimum	Mean	Std Dev
Children & Education	CHLF(i)	Pct Children 10-14 in labor force avg 1995-2000	149	52.49	0	10.33	14.05
Children & Education	PGR5F(i)	Persistence to grade 5, female (% cohort) avg 1995-2000	47	100.00	46.34	83.33	14.94
Children & Education	PGR5M(i)	Persistence to grade 5, male (% cohort) avg 1995-2000	47	100.00	51.14	82.29	14.27
Children & Education	PGR5T(i)	Persistence to grade 5, total (% cohort) avg 1995-2000	52	100	49	83	15
Children & Education	XSP(i)	Expenditure per student, primary (% GDP per capita) avg 1995-2000	67	45.79	2.08	14.83	8.58
Children & Education	XSS(i)	Expenditure per student, secondary (% GDP per capita) avg 1995-2000	77	84.85	4.67	22.73	16.05
Children & Education	XST(i)	Expenditure per student, total (% GDP per capita) avg 1995-2000	85	1,491.99	5.51	103.01	221.15
Competition	ELO2004(h)	Soccer Elo rating February 2004	152	2,022	706	1,465	278
Competition	ELORANK(h)	Soccer Elo rank February 2004	152	213	1	89	57
Competition	IMO9503(g)	Avg score International Math Olympiad 1995-2003	152	82.00	0	22.58	27.12
Competition	MISSINTERN(f)	Avg score Miss International 1960-2002	152	323.00	0.00	32.88	63.44
Competition	MISSWORLD(e)	Avg score Miss World 1951-2001	152	418	0	37	72
Culture	CPI2003(j)	Corruption Perceptions Index 2003	122	9.70000	1	4.27213	2.30634
Culture	LINGSIML(t)	Expected language self-similarity	152	1.0	0.17847	0.76048	0.22604
Culture	PATAPR(i)	Patent applications by residents avg 1995-2000	113	0.00	0.00	0.00	0.00
Diet	ALCOHOL909(d)	Litres per capita acohol cons avg 1990-99	30	15.80	1.60	9.89	3.18
Diet	CAL7097(d)	Pct change in calories per capita 1970 to 1997	125	85.20	-22.50	10.68	17.18
Diet	CALORIE70(d)	Calories per capita 1970	125	3,480	1,628	2,456	472
Diet	CALORIE97(d)	Calories per capita 1997	146	3,699	1,685	2,690	510
Diet	CIGARETTE(d)	Cigarettes per capita per year avg 1993-1997	116	4,075	70	1,216	867
Diet	FAT7097(d)	Pct change in fat grams per capita 1970 to 1997	125	229.50	-28.60	32.36	40.63
Diet	FAT97(d)	Fat grams per capita 1997	146	164	11	76	36
Diet	MALNH(i)	Pct under 5 malnourished height for age avg 1995-2000	82	54	0	27	15
Diet	MALNW(i)	Pct under 5 malnourished weight for age avg 1995-2000	85	59	0	20	14
Diet	PROT7097(d)	Pct change grams protein per capita 1970-1997	125	79	-31	11	21
Diet	PROTEIN97(d)	protein grams per capita 1997	146	115	28	74	21
Economy	FOODXP(i)	Food exports as pct exports avg 1995-2000	129	95.45	0.18	24.48	24.94
Economy	GCGDP(i)	Govt consumption as pct GDP avg 1995-2000	147	32.05	4.57	15.42	5.83
Economy	PCGDP(i)	Per capita GDP avg 1995-2000	150	39,510	491	7,824	8,203
Environment	C02GDP(i)	C02 emissions per dollar gdp avg 1995-2000	149	2.31942	0.00107	0.56460	0.51985
Environment	CERYLD(i)	Cereal yield (kg/ha) avg 1995-2000	146	7,559.84	219.82	2,550.84	1,705.44
Environment	GDPEU(i)	GDP per unit energy used (PPP) avg 1995-2000	118	10.24	1.06	4.45	2.13
Environment	PCTLANDPRO(i)	Pct national land protected 1999	51	32	1	9	8
Finance	FDIGDP(i)	FDI/GDP avg 1995-2000	138	111.30	0	5.76	10.29
Finance	PRSPRD(i)	(lending rate-deposit rate)/deposit rate avg 1995-2000	124	23.14	0.12	1.38	2.34
Health	HIVRATE(d)	HIV rate 1997	143	25.84	0.01	2.53	4.97
Health	NURSE(d)	Nurses per capita 1992-1995	118	2,184.00	3.20	317.06	392.04
Health	NURSE_DOCT(d)	Nurse/Doctor ratio 1992-1995	118	26.27	0.15	3.02	3.24
Health	PHYSPP(i)	Physicians per person avg 1995-2000	140	5.68	0.03	1.58	1.32
Health	TBRATE(d)	TB rate per capita 1997	152	588	2	72	95
Inequality	GINI(i)	Gini coefficient avg 1995-2000	85	60.70	21.70	40.19	9.11
Inequality	HI10(i)	Pct income held by highest income population decile	85	48.80	20.00	31.65	6.96
Inequality	LO10(i)	Pct income held by lowest income population decile	85	5.10	0.50	2.55	1.01
Life Expectancy	FAMORT(i)	Female adult mortality avg 1995-2000	152	679.66667	47.75000	203.25200	159.90215
Life Expectancy	FLEX(i)	Female life expectancy avg 1995-2000	152	83.70	38.84	67.30	12.57
Life Expectancy	INMORT(i)	Infant Mortality	152	165.20	3.57	42.62	38.40
Life Expectancy	MAMORT(i)	Male adult mortality avg 1995-2000	152	726.00	94.17	271.75	151.38
Life Expectancy	MFLEX(i)	Ratio of male life expect. to female avg 1995-2000	152	1.01	0.91	0.97	0.02
Life Expectancy	MLEX(i)	Male life expectancy avg 1995-2000	152	80	38	65	12
Life Expectancy	MORT05(i)	Mortality, ages 0-5 avg 1995-2000	149	276	4	64	66
Psychometric	HEDBAL(b)	Hedonic Balance (plaf-unpl): World Value Survey	39	2.90000	0.29000	1.27846	0.59723
Psychometric	IDV(a)	Individuality: Hofstede	51	91.00	6.00	45.45098	25.58462
Psychometric	IQ(c)	National avg IQ --Richard Lynn	152	106.00	59.00	84.93421	11.48202
Psychometric	LIFSAT(b)	Life Satisfaction: World Value Survey	40	8.39	5.03	6.93	0.86

Category	Variable	Description	N	Maximum	Minimum	Mean	Std Dev
Psychometric	MAS(a)	Masculinity: Hofstede	51	95.00	5.00	50.27451	18.78199
Psychometric	PDI(a)	Power Distance: Hofstede	51	104.00	11.00	54.98	22.02
Psychometric	PLAF(b)	Pleasant Affect: World Value Survey	39	3.63	1.12	2.44	0.60
Psychometric	UAV(a)	Uncertainty avoidance: Hofstede	51	112.00	8.00	66.08	24.37
Psychometric	UNPL(b)	Unpleasant Affect: World Value Survey	39	2.50	0.24	1.16	0.37
Sex Ratio & Fertility	FERT(i)	Fertility rate avg 1995-2000	152	7.36333	1.15500	3.31667	1.69967
Sex Ratio & Fertility	MF014(i)	Sex ratio: males/females, ages 0-14 avg 1995-2000	152	1.11385	0.99048	1.03670	0.02176
Sex Ratio & Fertility	MF1564(i)	Sex ratio: males/females, ages 15-64 avg 1995-2000	152	2.57028	0.79221	1.01074	0.15672
Sex Ratio & Fertility	MF65UP(i)	Sex ratio: males/females, ages 65 up avg 1995-2000	152	1.87116	0.44371	0.77692	0.16515
Sex Ratio & Fertility	MFTOT(i)	Sex ratio: all males/all females avg 1995-2000	152	1.97	0.86	0.99	0.10
Sexual Division Labor	FEMGOV(d)	Female as pct govt employees 1998	149	35.60	0	10.16	7.08
Sexual Division Labor	FEMGOVM(d)	Female as pct govt ministers 1998	149	43.50	0	8.62	8.03
Sexual Division Labor	FPCTLF(i)	Females as pct laborforce avg 1995-2000	148	51.95	14	39.68	7.97
Sexual Division Labor	PCTADMFEM(d)	Pct adminstrative female 2000	73	54.40	2.90	26.53	10.97
Sexual Division Labor	PCTMPFEM(d)	Percent Memb Parl female 1998	139	42.70	0.00	12.79	8.64
Sexual Division Labor	PCTPROFFEM(d)	Percent professional-technical female 2000	75	70.30	8.00	46.73	13.43
War	MILLF(i)	Military as pct labor force avg 1995-2000	149	7.86057	0	1.23178	1.31446
War	WARLIKE(t)	COW weighted militarized interstate contacts 1816-2001 (% total contacts)	152	0.05	0	0.01	0.01

Notes: (a) Hofstede (2003); (b) Diener and Suh (1999); (c) Lynn and Vanhanen (2002); (d) United Nations Development Program (2003); (e) GlobalBeauties.com (2001); (f) GlobalBeauties.com (2002); (g) International Mathematical Olympiad (2003); (h) World Football Elo Ratings (2004); (i) World Bank (2002)—all figures average of 1995-2000; (j) Transparency International (2004); (t) produced here as weight matrix diagonal.

TABLE 4: MORAN-I RESULTS FOR SAMPLE OF VARIABLES

Variable	Description	Allies	Huntington Civilization	Colonial-Imperial	Ecology	Physical Distance	Event Frequency	Language Phylogeny	Formal Treaty	Level of Development	Religion	Trade	Enemies
Children & Education													
CHLF	Pct Children 10-14 in labor force avg 1995-2000	0.3621***	0.4162***	0.2299***	0.1166***	0.6038***	0.3045***	0.4917***	0.6095***	0.6081***	0.0444***	0.1016***	0.6581***
PGR5F	Persistence to grade 5, female (% cohort) avg 1995-2000	-0.074	0.2531***	0.0943**	0.0063**	0.2383**	0.1534**	0.1716***	-0.101	0.3534***	-0.018	0.0049	0.1765*
PGR5M	Persistence to grade 5, male (% cohort) avg 1995-2000	-0.051	0.2986***	0.0432	0.0105**	0.2214**	0.1241**	0.1676**	-0.126*	0.3626***	-0.025	0.0006	0.0997
PGR5T	Persistence to grade 5, total (% cohort) avg 1995-2000	-0.024	0.2726***	0.1242**	0.0143**	0.2505***	0.1688**	0.1548**	-0.106*	0.3544***	-0.024	0.0499	0.1114
XSP	Expenditure per student, primary (% GDP per capita) avg 1995-2000	0.0811*	0.2316***	0.2967***	0.0391***	0.3301***	0.1327***	0.1604***	0.3529***	0.0143	0.0017*	0.1282***	0.1984**
XSS	Expenditure per student, secondary (% GDP per capita) avg 1995-2000	0.1107**	0.2112***	0.2009***	0.0241***	0.2835***	0.2300***	0.2155***	0.3091***	0.2421***	0.0311***	0.0895***	0.2777***
XST	Expenditure per student, total (% GDP per capita) avg 1995-2000	0.0292*	0.3291***	0.0389*	0.0306***	0.2236***	0.0224	0.4368***	0.1008**	0.2809***	0	0.0036	0.0187
Competition													
ELO2004	Soccer Elo rating February 2004	0.1623***	0.2155***	0.0266	0.0390***	0.2567***	0.1324***	0.1759***	0.1403***	0.1726***	0.0212***	0.1161***	0.3221***
ELORANK	Soccer Elo rank February 2004	0.1603***	0.2113***	0.0451**	0.0394***	0.2499***	0.1271***	0.1765***	0.1389***	0.1702***	0.0195***	0.1008***	0.3182***
IMO9503	Avg score International Math Olympiad 1995-2003	0.2303***	0.2309***	0.1480***	0.0239***	0.3091***	0.0976***	0.2168***	0.1984***	0.2093***	0.0276***	0.023	0.2836***
MISSINTERN	Avg score Miss International 1960-2002	0.3138***	0.2213***	0.1845***	0.0147***	0.1119**	0.2481***	0.2142***	0.4720***	0.2545***	0.0630***	0.1593***	0.2201***
MISSWORLD	Avg score Miss World 1951-2001	0.3442***	0.2193***	0.0455**	0.0080***	0.1086**	0.1747***	0.1812***	0.3162***	0.2712***	0.0512***	0.1650***	0.1436**
Culture													
CPI2003	Corruption Perceptions Index 2003	0.2187***	0.4534***	0.1390***	0.0448***	0.3830***	0.1703***	0.3343***	0.4432***	0.6700***	0.0716***	0.1899***	0.3059***
LINGSIML	Expected language self-similarity	0.1457***	0.1695***	0.1501***	0.0503***	0.2939***	0.1931***	0.1805***	0.2715***	0.1706***	0.0501***	0.0933***	0.3630***
PATAPR	Patent applications by residents avg 1995-2000	0.0810**	0.1251***	0.2507***	0.0295***	0.1825***	0.1005***	0.5732***	0.2201***	0.2583***	0.0043**	0.0523**	0.1574**
Diet													
ALCOHOL909	Litres per capita acohol cons avg 1990-99	-0.001	-0.008	0.0397	0.0280***	0.1874**	0.0914***	-0.02	0.0065	0.0329	-0.028	0.1377**	-0.039
CAL7097	Pct change in calories per capita 1970 to 1997	0.0733**	0.0987**	0.0691**	0	0.0840*	0.0551**	0.1830***	0.1433***	0.2140***	0.0248***	-0.003	0.018
CALORIE70	Calories per capita 1970	0.4035***	0.5256***	0.1578***	0.0910***	0.5103***	0.1892***	0.2952***	0.4641***	0.4919***	0.0612***	0.1520***	0.4398***
CALORIE97	Calories per capita 1997	0.4427***	0.4688***	0.0693***	0.1042***	0.5312***	0.2516***	0.3503***	0.5381***	0.4918***	0.0429***	0.1059***	0.5547***
CIGARETTE	Cigarettes per capita per year avg 1993-1997	0.3649***	0.3163***	0.1957***	0.0953***	0.4443***	0.1730***	0.2827***	0.4022***	0.3294***	0.0165***	0.0510**	0.4513***
FAT7097	Pct change in fat grams per capita 1970 to 1997	-0.007	0.3573***	0.0063	-0.004	0.2320***	0.0540**	0.1756***	0.0586*	0.1245***	0.0085**	0.0948***	0.1317**
FAT97	Fat grams per capita 1997	0.4759***	0.5428***	0.0552**	0.0844***	0.5789***	0.2574***	0.3959***	0.5717***	0.5763***	0.0810***	0.1892***	0.5249***
MALNH	Pct under 5 malnourished height for age avg 1995-2000	0.2269***	0.1963***	0.0498*	0.0272***	0.3260***	0.2548***	0.2789***	0.2610***	0.3892***	0.0011*	0.1154***	0.2456***
MALNW	Pct under 5 malnourished weight for age avg 1995-2000	0.1905***	0.1816***	0.1107**	0.0250***	0.4764***	0.3442***	0.2811***	0.3129***	0.3229***	0.0211***	0.1871***	0.4109***
PROT7097	Pct change grams protein per capita 1970-1997	-0.021	0.1069**	0.0426*	-0.005	0.0849*	0.0256	0.1592***	0.0445	0.1719***	0.0233***	-0.006	0.0375
PROTEIN97	protein grams per capita 1997	0.4809***	0.5159***	0.1372***	0.1242***	0.5746***	0.2536***	0.3866***	0.5521***	0.5444***	0.0482***	0.1090***	0.5858***
Economy													
FOODXP	Food exports as pct exports avg 1995-2000	0.1270***	0.1956***	0.0965***	0.0446***	0.3772***	0.1599***	0.1241***	0.2341***	0.1010***	0.0048*	0.0649***	0.3530***

Variable	Description	Allies	Huntington Civilization	Colonial-Imperial	Ecology	Physical Distance	Event Frequency	Language Phylogeny	Formal Treaty	Level of Development	Religion	Trade	Enemies
GCGDP	Govt consumption as pct GDP avg 1995-2000	0.1365***	0.2026***	0.1063***	0.0480***	0.3015***	0.1280***	0.1864***	0.1788***	0.1071***	0.0065**	0.0977***	0.1407***
PCGDP	Per capita GDP avg 1995-2000	0.3585***	0.4896***	0.1380***	0.0744***	0.4844***	0.2213***	0.3755***	0.5848***	0.7102***	0.0814***	0.1806***	0.4083***
Environment													
C02GDP	C02 emissions per dollar gdp avg 1995-2000	0.1374***	0.1268***	0.2682***	0.0480***	0.2497***	0.1612***	0.2928***	0.3263***	0.1438***	0.0331***	0.1603***	0.2471***
CERYLD	Cereal yield (kg/ha) avg 1995-2000	0.3359***	0.4210***	0.1337***	0.0953***	0.5081***	0.2584***	0.3527***	0.5051***	0.4386***	0.0685***	0.1388***	0.4538***
GDPEU	GDP per unit energy used (PPP) avg 1995-2000	0.1161**	0.1462***	0.1919***	0.0160***	0.2771***	0.1418***	0.1800***	0.2576***	0.1048***	0.0333***	0.1335***	0.2496***
PCTLANDPRO	Pct national land protected 1999	0.0780**	0.2714***	0.0503	0.0026**	0.2544***	0.1002***	0.1544**	0.1635**	0.1962***	0.1053***	0.1617***	0.3753***
	Finance												
FDIGDP	FDI/GDP avg 1995-2000	0.0037	0.0048	0.0027	0.0006**	-0.019	0.0128	0.0024	0.0310*	0.0257**	0.0146***	0.0112	0.0586*
PRSPRD	(lending rate-deposit rate)/deposit rate avg 1995-2000	0.0065	-0.004	-0.013	0.0000**	0.0152	-0.025	-0.036**	-0.006	-0.007	-0.01	-0.064**	0.0081
Health													
HIVRATE	HIV rate 1997	0.4600***	0.5045***	0.1252***	0.0604***	0.6411***	0.3778***	0.5483***	0.3603***	0.3305***	0.0436***	0.1871***	0.6167***
NURSE	Nurses per capita 1992-1995	0.2400***	0.3695***	0.5377***	0.0729***	0.5159***	0.1944***	0.3440***	0.5136***	0.2583***	0.0105**	0.1830***	0.4028***
NURSE_DOCT	Nurse/Doctor ratio 1992-1995	0.0457*	0.1297***	0.1435***	0.0137***	0.1187**	0.0626**	0.1307***	0.1072**	0.0417**	0.0135**	0.0077	0.1714**
PHYSPP	Physicians per person avg 1995-2000	0.4084***	0.5663***	0.3920***	0.1195***	0.6257***	0.2931***	0.4221***	0.6353***	0.3867***	0.0656***	0.2252***	0.5995***
TBRATE	TB rate per capita 1997	0.3121***	0.1815***	0.005	0.0116***	0.3975***	0.2417***	0.2437***	0.2174***	0.1755***	0.0075**	0.1005***	0.3633***
Inequality													
GINI	Gini coefficient avg 1995-2000	0.3115***	0.5359***	0.4238***	0.0871***	0.5916***	0.2019***	0.3149***	0.4787***	0.1046**	0.0050*	0.1945***	0.3712***
HI10	Pct income held by highest income population decile	0.2825***	0.5145***	0.4031***	0.0898***	0.5503***	0.1683***	0.2946***	0.4373***	0.1328***	0.0024*	0.1631***	0.3090***
LO10	Pct income held by lowest income population decile	0.2728***	0.4290***	0.3600***	0.0467***	0.4817***	0.2079***	0.3079***	0.4255***	0.0253	0.0226***	0.1945***	0.3462***
Life Expectancy													
FAMORT	Female adult mortality avg 1995-2000	0.5035***	0.6709***	0.2031***	0.1175***	0.7262***	0.4585***	0.7290***	0.6259***	0.6928***	0.0750***	0.1606***	0.7154***
FLEX	Female life expectancy avg 1995-2000	0.4959***	0.6723***	0.2113***	0.1257***	0.7108***	0.4282***	0.6874***	0.6615***	0.7390***	0.0844***	0.1504***	0.7066***
INMORT	Infant Mortality	0.3823***	0.5817***	0.2343***	0.1094***	0.6283***	0.3463***	0.5444***	0.6389***	0.6785***	0.0780***	0.1240***	0.5759***
MAMORT	Male adult mortality avg 1995-2000	0.4453***	0.6071***	0.1608***	0.0974***	0.6635***	0.4436***	0.6651***	0.5780***	0.6427***	0.0652***	0.1637***	0.6634***
MFLEX	Ratio of male life expect. to female avg 1995-2000	0.3077***	0.2391***	0.3621***	0.0607***	0.4466***	0.3371***	0.3521***	0.4828***	0.1778***	0.0441***	0.2777***	0.6110***
MLEX	Male life expectancy avg 1995-2000	0.4816***	0.6576***	0.1977***	0.1210***	0.6977***	0.4250***	0.6758***	0.6516***	0.7379***	0.0803***	0.1474***	0.6903***
MORT05	Mortality, ages 0-5 avg 1995-2000	0.4001***	0.5622***	0.2473***	0.1147***	0.6513***	0.3684***	0.5816***	0.6877***	0.6862***	0.0732***	0.1203***	0.6462***
Psychometric													
HEDBAL	Hedonic Balance (plaf-unpl): World Value Survey	0.0446*	0.1445**	0.1809**	-0.025	0.2399***	0.0623**	0.2530***	0.3114***	0.4458***	0.0324***	0.1501***	-0.005
IDV	Individuality: Hofstede	0.1785***	0.6620***	0.3335***	0.1192***	0.6645***	0.1021***	0.4468***	0.2995***	0.5549***	0.0100**	0.1041**	0.2869***
IQ	National avg IQ --Richard Lynn	0.4806***	0.6564***	0.2677***	0.1425***	0.7537***	0.4430***	0.6193***	0.6979***	0.5816***	0.0826***	0.1573***	0.7475***
LIFSAT	Life Satisfaction: World Value Survey	0.0719**	0.4006***	0.3427***	-0.01	0.2836***	0.1926***	0.3174***	0.4836***	0.4548***	0.0995***	0.2101***	0.1327**
MAS	Masculinity: Hofstede	0.0257	-0.04	0.3635***	-0.002	0.1585**	0.0745***	-0.063	0.0067	0.0967**	-0.027	0.0814**	-0.064
PDI	Power Distance: Hofstede	0.0791*	0.3669***	0.2293***	0.0720***	0.3813***	0.1216***	0.4606***	0.0847*	0.3728***	-0.004	0.1183**	0.2453***
PLAF	Pleasant Affect: World Value Survey	0.0551*	0.0484	-0.061	-0.022	0.1967**	0.0176	0.1625***	0.2315***	0.2385***	-0.001	0.0696**	-0.076
UAV	Uncertainty avoidance: Hofstede	0.0945**	0.1701***	0.4674***	-0.013	0.3499***	0.1316***	0.4512***	0.1127**	0.0551	-0.015	0.0982**	0.1446**
UNPL	Unpleasant Affect: World Value Survey	-0.023	0.1509**	0.0344	-0.018	0.1401**	0.011	-0.024	0.0027	0.1962***	-0.008	-0.015	-0.160**
Sexual Division Labor													
FEMGOV	Female as pct govt employees 1998	0.2097***	0.2694***	0.0842***	0.0133***	0.3059***	0.2363***	0.2837***	0.3361***	0.1547***	0.1178***	0.2022***	0.0663
FEMGOVM	Female as pct govt ministers 1998	0.1359***	0.2133***	0.0712**	0.0164***	0.2208***	0.1551***	0.2627***	0.2207***	0.1897***	0.0669***	0.1229***	0.0702
FPCTLF	Females as pct laborforce avg 1995-2000	0.2927***	0.2624***	0.2922***	0.0586***	0.3917***	0.2381***	0.3912***	0.5553***	0.0764***	0.0416***	0.1170***	0.5033***

Variable	Description	Allies	Huntington Civilization	Colonial-Imperial	Ecology	Physical Distance	Event Frequency	Language Phylogeny	Formal Treaty	Level of Development	Religion	Trade	Enemies
PCTADMFEM	Pct adminstrative female 2000	0.1341***	0.4467***	0.1220***	0.0042**	0.3153***	0.0978***	0.3373***	0.2119***	0.0553**	0.1561***	0.0938***	0.3302***
PCTMPFEM	Percent Memb Parl female 1998	0.1098***	0.2979***	0.2387***	0.0134***	0.2600***	0.1883***	0.2885***	0.2344***	0.3651***	0.1020***	0.1760***	0.2562***
PCTPROFFEM	Percent professional-technical female 2000	0.1380***	0.3815***	0.2633***	0.0118**	0.2758***	0.1628***	0.2371***	0.3147***	0.0809**	0.0894***	0.1427***	0.3014***
Sex Ratio & Fertility													
FERT	Fertility rate avg 1995-2000	0.4665***	0.6102***	0.2685***	0.1346***	0.6666***	0.3659***	0.5808***	0.7023***	0.5973***	0.0840***	0.1350***	0.7120***
MF014	Sex ratio: males/females, ages 0-14 avg 1995-2000	0.3567***	0.5614***	0.1693***	0.0907***	0.6059***	0.3318***	0.5333***	0.4831***	0.3976***	0.0402***	0.1146***	0.6029***
MF1564	Sex ratio: males/females, ages 15-64 avg 1995-2000	0.0984***	0.0659***	0.0508**	0.0273***	0.2690***	0.1028***	0.0528**	0.1455***	0.0818***	0.0238***	0.0508***	0.0781**
MF65UP	Sex ratio: males/females, ages 65 up avg 1995-2000	0.2057***	0.1779***	0.2617***	0.0556***	0.3869***	0.2227***	0.2019***	0.4275***	0.0808***	0.0586***	0.1830***	0.3609***
MFTOT	Sex ratio: all males/all females avg 1995-2000	0.1166**	0.1250***	0.0963***	0.0369***	0.3062***	0.1404***	0.0959***	0.2211***	0.0665***	0.0488***	0.0968***	0.1650***
War													
MILLF	Military as pct labor force avg 1995-2000	0.2936***	0.1343***	0.1313***	0.0502***	0.6049***	0.3130***	0.2476***	0.3070***	0.1120***	0.0349***	0.0758***	0.6646***
WARLIKE	COW weighted militarized interstate contacts 1816-2001	0.1821***	0.0289*	0.0267*	0.0074***	0.0974**	0.0866***	0.0155	0.3054***	0.0552**	-0.003	-0.053**	0.2802***

Notes: Moran's-I (Odland 1988). Significance levels from permutation tests (2,000 permutations): '***'=p-value below 0.01; '**'=p-value below 0.05; '*'=p-value below 0.10.

TABLE 5: PERCENTAGE SIGNIFICANT MORAN-I SCORES, BY MATRIX AND CATEGORY

Category	Allies	Huntington Civilization	Colonial-Imperial	Ecology	Physical Distance	Event Frequency	Language Phylogeny	Formal Treaty	Level of Development	Religion	Trade	Enemies	All 12 Matrices
Percentage with p-value<=.10													
Culture	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Economy	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Life Expectancy	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Sex Ratio & Fertility	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Environment	100.0	100.0	75.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	97.9
Inequality	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	66.7	100.0	100.0	100.0	97.2
Sexual Division Labor	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	66.7	97.2
Competition	100.0	100.0	80.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	80.0	100.0	96.7
Health	100.0	100.0	80.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	80.0	100.0	96.7
War	100.0	100.0	100.0	100.0	100.0	100.0	50.0	100.0	100.0	50.0	50.0	100.0	87.5
Diet	72.7	90.9	81.8	72.7	100.0	90.9	90.9	81.8	90.9	90.9	81.8	72.7	84.8
Children & Education	57.1	100.0	85.7	100.0	100.0	85.7	100.0	57.1	85.7	42.9	42.9	57.1	76.2
Psychometric	77.8	77.8	77.8	33.3	100.0	77.8	77.8	77.8	88.9	44.4	88.9	55.6	73.1
Finance	0.0	0.0	0.0	100.0	0.0	0.0	0.0	50.0	50.0	50.0	0.0	50.0	25.0
AllCategory	86.1	93.1	86.1	87.5	97.2	91.7	91.7	88.9	93.1	83.3	83.3	81.9	88.7
Percentage with p-value<=.05													
Culture	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Life Expectancy	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Sex Ratio & Fertility	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Environment	100.0	100.0	75.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	97.9
Economy	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	66.7	100.0	100.0	97.2
Sexual Division Labor	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	66.7	97.2
Competition	100.0	100.0	80.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	80.0	100.0	96.7
Health	80.0	100.0	80.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	80.0	100.0	95.0
Inequality	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	66.7	33.3	100.0	100.0	91.7
Diet	72.7	90.9	63.6	72.7	81.8	90.9	90.9	72.7	90.9	81.8	81.8	72.7	80.3
War	100.0	50.0	50.0	100.0	100.0	100.0	50.0	100.0	100.0	50.0	50.0	100.0	79.2
Children & Education	28.6	100.0	71.4	100.0	100.0	85.7	100.0	57.1	85.7	28.6	42.9	42.9	70.2
Psychometric	44.4	77.8	77.8	33.3	100.0	77.8	77.8	66.7	88.9	44.4	88.9	55.6	69.4
Finance	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	50.0	50.0	0.0	0.0	16.7
AllCategory	77.8	91.7	80.6	87.5	94.4	91.7	91.7	84.7	93.1	76.4	83.3	79.2	86.0
Percentage with p-value<=.01													
Life Expectancy	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Sex Ratio & Fertility	80.0	100.0	80.0	100.0	100.0	100.0	80.0	100.0	100.0	100.0	100.0	80.0	93.3
Economy	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	33.3	100.0	66.7	91.7
Sexual Division Labor	100.0	100.0	83.3	66.7	100.0	100.0	100.0	100.0	66.7	100.0	100.0	66.7	90.3
Culture	66.7	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	66.7	66.7	66.7	88.9
Inequality	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	33.3	33.3	100.0	100.0	88.9
Competition	100.0	100.0	40.0	100.0	60.0	100.0	100.0	100.0	100.0	100.0	80.0	80.0	88.3
Environment	50.0	100.0	75.0	75.0	100.0	100.0	75.0	75.0	100.0	100.0	100.0	100.0	87.5
Health	80.0	100.0	80.0	100.0	80.0	80.0	100.0	80.0	80.0	40.0	80.0	80.0	81.7
War	100.0	50.0	50.0	100.0	50.0	100.0	50.0	100.0	50.0	50.0	50.0	100.0	70.8

Category	Allies	Huntington Civilization	Colonial-Imperial	Ecology	Physical Distance	Event Frequency	Language Phylogeny	Formal Treaty	Level of Development	Religion	Trade	Enemies	All 12 Matrices
Diet	63.6	72.7	36.4	72.7	72.7	72.7	90.9	72.7	90.9	72.7	63.6	63.6	70.5
Children & Education	14.3	100.0	42.9	57.1	71.4	42.9	71.4	42.9	85.7	28.6	42.9	28.6	52.4
Psychometric	22.2	55.6	66.7	33.3	66.7	66.7	77.8	55.6	77.8	33.3	33.3	33.3	51.9
Finance	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	50.0	0.0	0.0	4.2
AllCategory	66.7	86.1	66.7	76.4	80.6	81.9	86.1	79.2	83.3	66.7	72.2	66.7	76.0

Notes: Summarizes Table 4. Matrix Rank based on Overall Average, such that (1) is the matrix which shows autocorrelation for the highest percent of sample variables and (12) is the matrix showing autocorrelation for the lowest percent of sample variables.

TABLE 6: DETERMINANTS OF FERTILITY, MODEL 1

Variable	Description	Coefficient	Std Error	P-value	VIF
Intercept	Intercept	1.5237	0.3812	0.0001	0.00
Pchlf	Instrument for % children 10-14 in labor force	3.0235	0.5793	<.0001	2.67
Pfpctlf	Instrument for % women in labor force	-1.8564	0.8473	0.0302	1.31
amort_1	Adult mortality avg 1960-1980	0.0038	0.0009	<.0001	5.22
inmort_1	Infant mortality avg 1960-1980	0.0116	0.0025	<.0001	5.56

Notes: Nobs=135; R^2= 0.8469; White's Test (H0: errors are homoskedastic) p-value= 0.2313; RESET test (H0: model is correctly specified) p-value= 0.001; Moran's I (H0: errors not autocorrelated): Allies: p-value=0.00007; Huntington Civilization: p-value=0; Colonial/Imperial: p-value=0.06389; Ecology: p-value=0.00504; Physical Distance: p-value=0.00001; Event Frequency: p-value=0.00474; Language Phylogeny: p-value=0.00066; Formal Treaty: p-value=0.00184; Level of Development: p-value=0.01875; Religion: p-value=0; Trade: p-value=0.89667; Enemies: p-value=0.0061.

TABLE 7: DETERMINANTS OF LOGGED FERTILITY, MODEL 2

Variable	Description	Coefficient	Std Error	P-value	VIF
Intercept	Intercept	9.503	1.525	<.0001	0.00
log(pchlf)	Instrument for % children 10-14 in labor force	0.1004	0.0278	0.0004	11.93
log(pfpctlf)	Instrument for % women in labor force	-0.2925	0.0832	0.0006	1.91
log(famort_1)	Adult female mortality avg 1960-1980	-0.2505	0.1322	0.0604	33.05
log(mamort_1)	Adult male mortality avg 1960-1980	-1.4327	0.3104	<.0001	96.66
log(inmort_1)	Infant mortality avg 1960-1980	-2.3776	0.3800	<.0001	508.96
inmamort	log(inmort_1)*log(mamort_1)	0.4246	0.0698	<.0001	1,010.00
log(fert_1)	Fertility rate avg 1960-1980	0.5990	0.0872	<.0001	8.45
log(mf014_1)	Sex Ratio: Males/Females ages 0-14. avg 1960-1980	-3.212	0.836	0.0002	2.32

Notes: Nobs=135; R^2= 0.9185; Model 2's R^2 on dependent variable of Model 1=0.9049; White's Test (H0: errors are homoskedastic) p-value= 0.9517; RESET test (H0: model is correctly specified) p-value= 0.4887; Moran's I (H0: errors not autocorrelated): Allies: p-value=0.8332; Huntington Civilization: p-value=0.14579; Colonial/Imperial: p-value=0.48297; Ecology: p-value=0.30988; Physical Distance: p-value=0.12387; Event Frequency: p-value=0.7853; Language Phylogeny: p-value=0.33126; Formal Treaty: p-value=0.76725; Level of Development: p-value=0.93906; Religion: p-value=0.30527; Trade: p-value=0.99988; Enemies: p-value=0.29362.