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

Spatial, Cultural, and Ecological Autocorrelation in U.S. Regional Data

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 U.S. regions, it can be defined using physical distance, cultural similarity, ecological similarity, or using frequency and intensity of interaction, such as migration or commuting 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 U.S. regional data.

Key words: Spatial Autocorrelation; Culture; Religion

JEL category: R15, 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. A summer 2004 research grant from the Jennings Jones College of Business, Middle Tennessee State University, helped make this paper possible.

SPATIAL, CULTURAL, AND ECOLOGICAL AUTOCORRELATION IN U.S. DATA

Abstract: Positive autocorrelation implies that proximate observations take on similar values. "Proximate" can be defined in many different dimensions. In a cross-section of U.S. regions, it can be defined using physical distance, cultural similarity, ecological similarity, or using frequency and intensity of interaction, such as migration or commuting 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 U.S. regional data.

1. INTRODUCTION

Regression analysis using socioeconomic data typically tries to uncover functional relationships among different social phenomena. For example, one might regress the average wage rate on potential determinants of that wage rate, including the average level of educational attainment. One would then interpret the regression results as an expression of a functional relationship between wages and education, a functional relationship that would be elicited universally, in any social system. The interpretation of regression results as functional relationships occurs regardless of whether the observations in the socioeconomic data are nations, regions, cultures, or persons.

Nevertheless, the association between wages and education need not be a functional relationship. The two could be associated for other reasons. For example, suppose that at some time in the past all peoples had low wages and low educational attainment. But one ethnic group, entirely by chance, acquired both high wages and high educational attainment as features of its culture. Over time, as the high wage, high educational attainment ethnic group migrated and settled new lands, it carried with it its culture. And perhaps, for a variety of reasons, this ethnic group was admired, so that its culture was emulated. Over time, then, one would find that these two traits (high wages, high educational attainment) would spread, and that they would seldom or never be found alone, but any population that had one trait (say high wages), would have the other (high educational attainment).

Francis Galton was the first to point out that a significant correlation between two cultural traits need not signify a functional relationship, but could actually be the result of processes of cultural borrowing or cultural inheritance (Stocking 1968: 175). The eponymous "Galton's Problem" has become an important issue in the discipline of Cross-Cultural Anthropology, where a variety of methods have evolved to mitigate the problem (Naroll 1965; Mace and Pagel 1994; Murdock, 1957; Murdock and White, 1969). Among the most useful for economists and others conducting regression analyses are methods based on spatial autocorrelation statistics (White, Burton, and Dow 1981; Dow, Burton, and White 1982; Dow, White, and Burton 1982; Dow, Burton, Reitz, and White 1984; Loftin 1972; Loftin and Ward 1983). In the same way that one might construct a spatial weight matrix, to see if physically proximate observations have similar values, so might one construct a cultural weight matrix, to see if culturally proximate observations have similar values.

In the regression context, it's quite clear that Galton's Problem is an omitted variable problem. The spurious functional relationship appears because the regression failed to control for processes of inheritance and borrowing. It has long been recognized in housing price analysis (e.g., Can 1998) that the presence of spatially autocorrelated regression residuals indicates "neighborhood effects," where proximate homes have similar prices because they share features with each other—features that are not included in the set of independent variables. The usual remedy for spatially autocorrelated residuals is to create a spatially lagged dependent variable (taking appropriate steps to avoid endogeneity) and to add this to the model as the proxy for the omitted variables (Anselin 1988). One can, in a similar manner, create a culturally lagged dependent variable, and add this to a regression model to control for the association of cultural traits due to processes of inheritance (White, Burton, and Dow 1981; Dow, Burton, and White 1982; Dow, White, and Burton 1982; Dow, Burton, Reitz, and White 1984).

A previous paper (Eff 2004) examined the prevalence of autocorrelation in international datasets. Unlike regional economists, who have long recognized the importance of spatial autocorrelation, economists working with international data have ignored spatial autocorrelation. It was found in that paper that autocorrelation (in spatial, cultural, and other dimensions) is highly prevalent among a wide variety of international data series. In the regional context, no work has yet been done on the prevalence of autocorrelation in dimensions other than that of physical distance. The purpose of the present paper is to determine the extent of various types of autocorrelation in U.S. regional data. To that end, 35 different weight matrices are constructed, each measuring—in a different dimension—the proximity of the 394 U.S. Labor Market Areas. These 35 weight matrices are then used to produce autocorrelation statistics on a sample of 205 variables from a variety of sources. The results give an indication of the prevalence of the autocorrelation, in different dimensions, that might be found in U.S. regional data.

The paper is organized as follows. The following section details the construction of 35 different weight matrices, each of which defines "proximity" between regions in a different way. The next section then compares the 35 matrices, using matrix correlation, to get a sense of how they covary. The paper then presents a dataset of 205 variables for 394 regions (U.S. Labor Market

Areas),[1] 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 regional data, and allow stylized facts to be produced of the form: "regions that are spatially (culturally, etc.) proximate, tend to have similar values of variable *y*."

2. CONSTRUCTION OF WEIGHT MATRICES

The relationships among regions can be described in many different dimensions. In this section, the relationships among U.S. Labor Market Areas (LMAs) are operationalized by constructing 35 different weight matrices, each modeling a different dimension of inter-regional relationship. Each weight matrix **W** contains elements w_{ij} giving the proximity between LMA *i* and LMA *j*, where higher values of w_{ij} correspond to greater closeness between the pair of LMAs.

Physical Proximity

Two matrices are constructed to model physical proximity: a contiguity matrix, and a matrix based on the great circle distance between the population-weighted centroids of each LMA. The contiguity matrix **W** has elements w_{ij} equal to one when the LMAs are contiguous, and zero otherwise:

(1) $$w_{ij} = 1 \;\; if \;\; LMA_i \;\; is \;\; contiguous \;\; to \;\; LMA_j, \;\; w_{ij} = 0 \;\; otherwise$$

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

(2) $$d_{ij}=(0.62)*6371.1*\arccos[\sin(y_i)*\sin(y_j)+\cos(y_i)*\cos(y_j)*\cos(x_i-x_j)]$$

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

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

Table 1 summarizes the two physical proximity matrices.

Cultural Proximity I (Ancestry)

Economic theory has long contained speculations about the role of culture in shaping economic behavior. Edmund Burke's view that the inherited traditions of a people constitute laboriously worked-out solutions to recurring problems (Muller 2002) was to find favor with many other

[1] The Labor Market Area boundaries used are those of Killian and Tolbert (1993). Goetz (1999) provides a good review of the issues involved in defining Labor Market Areas.

economists who respected tradition, perhaps most notably Friedrich Hayek (Hayek 1979). Thorstein Veblen's economic analyses focused on "habits of thought" transmitted from the past, though Veblen differs from Burke and Hayek in that he deemphasizes the instrumental features of inherited tradition, instead viewing tradition as a ceremonial buttress for elite power and status (Veblen 1899). Recent scholarship in economics has reintroduced the notion that culture is an important determinant of economic development (Harrison and Huntington 2000; Landes 1998).

Culture should be distinguished both from universally observed human behavior and from individual learning. The former is part of the human genetic endowment, shaped by selective forces in the environment of our mammalian and hominid ancestors. The latter is not received from other humans, but simply acquired through trial and error. Since individual learning is costly, culture provides an efficient way for humans to acquire useful knowledge and behavior, particularly in environments experiencing little change. Culture can be transmitted vertically (from parents to offspring), horizontally (among members of the same generation), and obliquely (from unrelated adults to a younger generation) (Boyd and Richerson 1985).

In the United States, oblique transmission of culture is characteristic of the many institutions that assimilate Americans of many different ancestries into a national culture. Perhaps most important are public education and the news and entertainment media. Since age-segregation is common in the United States, horizontal transmission is also important—especially in public schools—and would usually work to assimilate different ancestry groups into a national culture. Vertical transmission, on the other hand, would typically work to preserve the traditions of ancestry groups. Ancestry groups that are highly endogamous would tend to conserve traditions more than highly exogamous ancestry groups. Ancestry groups that tend to be spatially concentrated, whether in urban ghettoes or homogeneous rural districts, would have an additional impetus toward tradition-conservation, since oblique and horizontal cultural transmission would more often be within the same ancestry group. Nevertheless, even dispersed settlement can be compensated by practices of association and communication, so that an ancestry group can remain cohesive (Zelinsky and Lee 1998).

A small literature describes how practices associated with a particular ethnic group tend to be transmitted over the generations, even when the group's technology and environment change. A well-known example from ethnography is that of the American Plains Indians. With the introduction of the horse, a number of peoples moved onto the Plains and developed a culture based on nomadic buffalo hunting. While the Plains Indians all possessed a common core of cultural traits (such as the horse, the travois, and the teepee), they differed from each other in features such as political organization, and these differences can be attributed to cultural inertia—the persistence of traits from the time before they migrated onto the Plains (Boyd and Richerson 1985: 57). Cultural inertia is well documented even among European-Americans. For example, Sonya Salamon (1984) describes how farming practices in the Midwest vary according to whether proprietors are of German, Irish, Yankee, or Swedish ancestry. Salamon's farmers are at least three or four generations removed from their European peasant ancestors, and yet the ethnic differences have persisted.

Delineation of U.S. regional cultures has been attempted by a number of scholars, primarily in the field of cultural geography. Perhaps the most comprehensive attempts are those of Raymond D. Gastil (1975) and Wilbur Zelinsky (1992). A well-known effort by a journalist is Joel Garreau's (1981) *Nine Nations of North America*.

Historians have often differed when describing the origins of U.S. regional or national culture. A tradition most often associated with Frederick Jackson Turner (1893) maintains that European immigrants adapted their culture to frontier conditions, so that U.S. culture—and particularly the culture of the American West—is less a product of cultural transmission from the peoples of Europe than it is a product of individual learning in the novel frontier environment. Other historians have chosen to emphasize the role of cultural transmission, and deemphasize the role of individual learning. The best known name here is certainly David Hackett Fischer (1991), who maintains that the regional cultures of the United States stem from four regional British cultures brought to those areas in colonial times. Thus, New England was settled by Puritans from East Anglia, Virginia was settled by "Cavaliers and Servants" from southern England, Pennsylvania was settled by Quakers from the English North Midlands, and the western frontier was populated by Borderers from the border between Scotland and England. Fischer argues that the first migrants develop a culture to which subsequent migrants must adapt. A similar argument has been advanced by George Foster (1960) for Latin America. Foster believes that the first Spanish immigrants to Latin America—predominantly from Seville—"crystallized" Latin American colonial culture, and that subsequent immigrants had little influence on Latin American regional cultures. The cultural geographer Wilbur Zelinsky (1992: 23) calls this the "doctrine of first effective settlement"—"the hypothesis that the first European or American white population that established the economic and social basis of an area had a decisive influence on later patterns" (Gastil 1975: 27).

From this brief discussion, it is evident that many different perspectives exist regarding the persistence of culture among U.S. ancestry groups. Some might maintain that cultural traditions were not particularly useful in the novel frontier environment, so that different ancestry groups assimilated into a new culture created through the individual learning of first generation immigrants. Others might maintain that immigrant groups assimilate to the preexisting culture, so that only the initial wave of British colonial immigrants conserve their cultural traditions. Yet others would maintain that ancestry groups retain many cultural traditions, particularly ancestry groups that are endogamous and spatially concentrated.

Census data from 1990 and 2000 provide information on the race, ancestry, and Hispanic status of county populations. These data can be used to calculate the percent of population in each LMA from each ancestry group. In the county-level data, the decennial Census presents "ancestry" only for those persons who are in the racial category "white" (the original peoples of Europe, North Africa, and Asia as far east as Afghanistan) or "black" (though very few people who are "black" will specify an ancestry). Native Americans and ancestry groups from Asia, the Pacific, or Latin America will have detailed ancestry presented in detailed tables for race and Hispanic status. Census presents ancestry in almost all cases as a nation-state, though there are a few Census ancestry groups that have no nation-state, such as the Rom, the Kurds, and the Assyrians. Detailed race tables for Asians and Pacific Islanders also typically assign race to a nation-state.

Many persons do not specify an ancestry, and many others will declare their ancestry to be "American," or some variant such as "Tennessean." Others might name a religion when declaring an ancestry—a response which is disqualified and classified as unclassifiable. Thus, Jewish-Americans are usually classified as members of the ancestry groups of Eastern and Central Europe.

The geographic distribution of ancestry groups shows that those who specify "American" (as well as those who do not specify an ancestry) are especially common in the area settled by Fischer's "Borderers" (often described as the "Scotch-Irish"): Tennessee, Kentucky, and then westward up to Missouri and down to Texas. Borderers arrived in large numbers in the mid 18^{th} century. A poor people, with a tradition of skirmish warfare, they squatted on land on the western Pennsylvania frontier, then migrated along the Appalachian valleys toward the south. In Virginia, they mingled with poor farmers from Fischer's Cavalier and Servant culture, then migrated through the Cumberland Gap into Kentucky and Tennessee, whence they made the first inroads of European settlement both to the northwest and southwest (Gastil 1975: 10-11). Unlike the people of New England, who continue to describe their ancestry as "English," these descendents of the Borderers have lost sight of their European roots, a phenomenon already apparent in the early 20^{th} century (Fischer 1991: 618). Others who describe themselves as "American" or who do not specify an ancestry may simply be a mixture of many different ancestry groups, and may think of themselves as “white” or some other racial identity.

The decennial Census allows each respondent to specify two ancestries. For persons who specify two ancestries, the number of ancestries is divided by half, and then added to the single ancestry figure. Ancestry codes for “American,” “Did not specify,” and “Unclassifiable” are then replaced by race and Hispanic categories as follows. First the number of black persons is reduced by the number of ancestries from Sub-Saharan Africa (excluding South Africa) and the non-Hispanic West Indies. Then black, detailed Asian, detailed Hispanic, and detailed Pacific Islander and American Indian are added to the ancestry figure. This figure is then subtracted from the total population, and the difference is then labeled “white.” Thus the 2000 Census provides 95 ancestry categories and the 1990 Census 71 ancestry categories. For each LMA, one can calculate the percentage of its population in each of these ancestry groups.

For each pair of LMAs, one can calculate the cultural proximity between them, based on the similarity of their ancestries. Four types of proximity indices are used. The first is the inverse of the Euclidean distance between the two LMAs:

$$w_{ij} = \left(\sum_{k=1}^{m} \left(p_{ik} - p_{jk} \right)^2 \right)^{-\frac{1}{2}} \tag{4}$$

Where p_{ik} is the percentage of the population in LMA i in ancestry group k, and p_{jk} is the percentage of the population in LMA j in ancestry group k. The second proximity index is a Herfindahl-type index:

$$w_{ij} = \sum_{k=1}^{m} \left(p_{ik} p_{jk} \right) \tag{5}$$

Where the notation is as above. The third proximity index is a modification of the Herfindahl-type index:

$$w_{ij} = \sum_{k=1}^{m} \sum_{p=1}^{m} \left(p_{ik} p_{jp} S_{kp} \right) \tag{6}$$

Where S_{kp} is a similarity index between ancestry group k and ancestry group p. Equation (6) can be interpreted as the expected similarity in ancestries between a person randomly chosen in LMA i and a person randomly chosen in LMA j. The similarity index S_{kp} can be derived from a variety of sources. In a previous paper, utilizing international data, language phylogenies served as the basis of S_{kp} (Eff 2004). Genomic similarity, as in the work of Cavalli-Sforza, Menozzi, and Piazza

(1994), would provide another feasible source of a similarity index. In the present paper, ancestry groups were judged to be more similar if there was a higher rate of intermarriage among them. Appendix A describes the data and methods used to produce a marriage matrix **S**, where each element S_{kp} is a scaled ratio, with the numerator the percentage of married persons in ancestry group k married to persons in ancestry group p, and the denominator the percentage of married persons in the married population who are in ancestry group p.

The final proximity index takes on values of one when the similarity between two LMAs lies above a threshold, and takes on the value of zero otherwise. The index is similar to the contiguity index, in that a relationship is modeled as a binary variable. In the present case, the principal ancestry group is extracted for each LMA, using ancestry data from both 1990 and 2000. A total of 13 ancestry groups serve as the principal ancestry group in at least one LMA (see Figure 1). If two LMAs share the same principal ancestry group, then they are judged to be similar; otherwise they are judged to be dissimilar.

$$(7) \qquad w_{ij} = 1 \;\; if \;\; A_i = A_j, \;\; w_{ij} = 0 \;\; otherwise$$

Where A_i is the principal ancestry group for LMA i, and A_j is the principal ancestry group for LMA j.

Table 1 provides a summary description of the six weight matrices created from ancestry data.

Cultural Proximity II (Religion)

The political scientist 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. In a previous study (Eff 2004), it was found that a weight matrix based on Huntington's classifications was very successful in eliciting autocorrelation for a variety of international data series. Arguments similar to those of Huntington have been made by the cross-cultural anthropologist Andrey Korotayev (2004), who maintains that the cultures of the "Old World Oikumene" can phenetically be divided into two groups—one Christian and one Moslem. Korotayev believes that the Christian proscription of polygyny gave rise to numerous changes in social structure in Christian cultures, changes that differentiated Christian cultures from Moslem cultures. Korotayev's findings lend some support to the view that religion may be one of the most important features of a culture—important because religion might in some sense determine many of the other features of social life. David Hackett Fischer (1991: 795) maintains that "of all the determinants which shaped the cultural character of British North America, the most powerful was religion."

Within the United States there is considerable regional variation in the membership of Christian denominations, though the numbers of non-Christian believers are small, and in fact little or no data are available for members of non-Christian religions. Even for Christian denominations, county-level data on membership is not plentiful. The Censuses of 1890, 1906, 1916, 1926, and 1936 provide reliable data, but for more recent data one must rely on private sources. The Glenmary Research Center, in Atlanta, provides a decennial census of Judeo-Christian membership and attendance by county, though "members" and "attendants" are not always defined in the same way by all Christian denominations (Bradley 1992). The 2000 Glenmary census reportedly provides figures for followers of Islam, though these data were not available to me. [2]

Some Christian denominations are associated with particular ancestry groups. For example, the area of Puritan settlement into the old Northwest can most readily be mapped by examining the westward spread of Congregational churches (Gastil 1975: 52). Each of Fischer's four regional cultures was associated with a Christian denomination: the Puritans were Congregationalists, the Borderers were Presbyterians, the Quakers were Quakers, and the Cavaliers and Servants were Episcopalians (Fischer 1991). Each of these cultures, though, experienced substantial changes in religion. Congregationalists become steadily more liberal—less puritanical—over time, eventually giving rise to theologically liberal denominations such as Unitarianism. Congregationalist migrants to upstate New York or the old Northwest might often convert to another strain of Calvinism, such as Presbyterianism. New England was also affected by the Revivalist movements of the early 19th century, in which many turned to Baptist and Methodist denominations (Gastil 1975: 52). The Quakers ceased to proselytize, becoming a "tribal" religion, and soon became an insignificant minority in Pennsylvania. However, they had invited many similar German Pietist and Anabaptist sects to settle in their colony, and many of these German groups had high rates of natural increase, so that a Pennsylvania style of religion continued and spread to the west. In the Chesapeake, the Episcopal Church never lost favor with the "Cavaliers," but was displaced by less liturgically oriented denominations such as the Baptists among the "Servants." The evangelization of African slaves eventually led to segregated churches for African-Americans, and one could argue that African-American culture is the largest branch descending from the culture of the Chesapeake (Butler 1990). The Borderers were affected by the great backcountry religious revivals of the early 19th century, when Presbyterianism lost ground

---

[2] The U.S. Bureau of the Census county-level religious membership data from the censuses of 1890, 1926, and 1936, as well as the county-level Glenmary Research Center data for 1990, were generously made available by the American Religion Data Archive (http://www.thearda.com/).

to more charismatic or enthusiastic denominations such as the Baptists and Methodists (Conkin 1990). These changes make it difficult for one to use Census data from 1890 or later to pick out the areas settled by each of Fischer's four cultures. Nevertheless an attempt was made. Figure 2 shows the results.

For each pair of LMAs, one can calculate the cultural proximity between them, based on the similarity of their religions. Four types of proximity indices are used. The first is the inverse of the Euclidean distance between the two LMAs:

$$w_{ij} = \left(\sum_{k=1}^{m} \left(p_{ik} - p_{jk} \right)^2 \right)^{-1/2} \tag{8}$$

Where p_{ik} is the percentage of the population in LMA i in religious denomination k, and p_{jk} is the percentage of the population in LMA j in religious denomination k. The second proximity index is a Herfindahl-type index:

$$w_{ij} = \sum_{k=1}^{m} \left(p_{ik} p_{jk} \right) \tag{9}$$

The third proximity index is a modification of the Herfindahl-type index:

$$w_{ij} = \sum_{k=1}^{m} \sum_{p=1}^{m} \left(p_{ik} p_{jp} S_{kp} \right) \tag{10}$$

Where S_{kp} is a similarity index between religious denomination k and religious denomination p. Equation (10) can be interpreted as the expected similarity in religious affiliation between a person randomly chosen in LMA i and a person randomly chosen in LMA j. The similarity index S_{kp} is here derived from a phenetic classification[3] presented by the Glenmary Research Center (Bradley 1992). Appendix Table A2 shows the phenetic classification, and the weights S_{kp} are created using the method presented in Appendix Equation (A1).

The fourth proximity index is binary: if two LMAs have the same principal religious denomination, then they are judged to be similar; otherwise they are judged to be dissimilar.

$$w_{ij} = 1 \;\; if \;\; A_i = A_j, \;\; w_{ij} = 0 \;\; otherwise \tag{11}$$

---

[3] *Phenetic* means that the denominations are classified according to observed similarity in doctrine and style of worship. A *genetic* classification would be preferable, since it would capture the actual historical relationships among fissioning and fusing denominations, but the compilation of the historical relationships among Christian denominations was too time-consuming a task for the purposes of this study.

Where A_i is the principal religious denomination for LMA i, and A_j is the principal religious denomination for LMA j. For the 1990 figures, the number of adherents are aggregated into four categories: Protestants, Catholic/Orthodox, Jewish, and non-adherents. For the 1890 figures, eight principal religious categories are used. Four of these are from David Hackett Fischer's four British cultures (Borderer, Quaker, Cavalier and Servant, and Puritan). For areas of the country with a colonial non-English culture (such as the Hudson Valley or New Mexico), or where immigrant groups played an important a part in crystallizing the local culture (such as Minnesota and the Dakotas), Fischer's four cultures can not reasonably be assigned. The four additional religious groups used for these cases are Dutch, Catholic, Lutheran, and Mormon.[4]

Table 1 provides a summary description of the seven weight matrices created from religion data.

Cultural Proximity III (Place Name Similarity)

Toponymy is widely used by historians to delineate areas settled by pre-modern people. For example, the suffix "*by*" means "town" in Danish, and those areas in Britain settled by Viking Age Danes will have many villages with the suffix. One might argue that U.S. places with the same place names would be similar—perhaps because settlers in the West would choose the place names common in their natal regions to the east, or perhaps because names such as "Madison" or "Jefferson" would have only a brief period of popularity, and then only among people with similar political views, so that places sharing these names are likely to have been settled at the same time by people with similar cultural affinities.

The U.S. Bureau of the Census provides as part of its geocoded data a set of point files for U.S. places, containing 15,528 named places. About 80 percent of the place names are found only in one LMA; about 10 percent are found in two LMAs, and the remaining 10 percent are found in up to 27 different LMAs. Thus, about 3,000 place names are found in more than one LMA. Equation (12) presents the formula used to create a proximity measure between a pair of LMAs:

$$(12) \qquad w_{ij} = \frac{\left(\sum_{(k \in R_i) \cap (k \in R_j)} X_{ik} \sum_{(k \in R_i) \cap (k \in R_j)} X_{jk} \right)^{1/2}}{\left(\sum_{k \in R_i} X_{ik} \sum_{k \in R_j} X_{jk} \right)^{1/2}}$$

[4] One might, however, choose to regard Mormon adherents as simply part of the Puritan group, since the founding events of Mormonism occurred in upstate New York among people who were of New England stock (Gastil 1975: 12-13).

Where X_{ik} is the population of LMA i located in a place with place name k, X_{jk} is the population of LMA j located in a place with place name k, and R_i and R_j are the set of place names for LMA i and LMA j, respectively.

Table 1 provides a summary description of the weight matrix created from place name data.

Cultural Proximity IV (Presidential Election Similarity)

Salient differences in regional cultures are likely to give rise to regional differences in political affiliations. Political scientists have long been interested in regional and ethnic patterns in electoral behavior, with important contributions by Daniel Elazar (1972) and Kevin Phillips (1969). Elazar classified British colonial cultures and the cultures of post-colonial immigrant groups in the categories: Moralistic, Individualistic, and Traditionalistic. Puritans, for example, are Moralistic, as are Scandinavians, Jews, and Anglo-Canadians. Southerners of British stock, on the other hand, are Traditionalistic, as are post-colonial immigrants of Mediterranean, East European, and French-Canadian ancestry. Elazar uses his taxonomy to explain why Scandinavians and Yankees in Minnesota tend to have similar political views, but Yankees and French-Canadians in Massachusetts are often politically at loggerheads (Gastil 1975: 55-58).

Since voting behavior is an expression of values, and values are an important component of culture, one can use national elections as an opportunity to quantify cultural differences among regions. The Inter-university Consortium for Political and Social Research distributes county-level returns for national elections between 1840 and 1972 (Clubb, Flanigan, and Zingale 1986). Congressional elections can be used to see if voters are voting for the same party in different regions, but a party might stand for quite different principles in different parts of the country—as for example, the Democratic Party in the 1970s represented different principles for southern whites than it did for northern African-Americans. Presidential elections are the most useful to consider, since voters in all regions are—in most cases—choosing among the same slate of candidates. Data before 1912 suffer from the problem of missing counties, since a fairly large number of western counties were created in the first decade of the 20th Century.

For each pair of LMAs, one can calculate the cultural proximity between them, based on the similarity of their tallies in Presidential elections. Two types of proximity indices are used. The first is the inverse of the Euclidean distance between the two LMAs:

(13) $$w_{ij} = \left(\sum_{y} \sum_{k_y} \left(p_{ik_y y} - p_{jk_y y} \right)^2 \right)^{-1/2}$$

Where $p_{ik_y y}$ is the percentage of the tallied votes in LMA i in election year y for candidate k_y, and $p_{jk_y y}$ is the percentage of the tallied votes in LMA j in election year y for candidate k_y. The second proximity index is a Herfindahl-type index:

$$w_{ij} = \sum_{y} \sum_{k_y} \left(p_{ik_y y} p_{jk_y y} \right) \tag{14}$$

Where the notation is as in Equation (13). Table 1 provides a summary description of the two weight matrices created from Presidential election data.

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 also interpret the perspective of historians like Frederick Jackson Turner (1893) as suggesting that the local environment is a particularly important determinant of U.S. regional cultures.

Delineations of U.S. ecological regions have been produced by the United States Forest Service (2004). The data are in shapefile format, readable in ArcView. By overlaying the ecological regions theme with the LMA boundary theme, one can measure the percentage of an LMA that is in each ecological region.

The U.S. Forest Service ecological regions are hierarchically structured in four levels of aggregation. The grossest level is the "Domain," of which there are four in the United States: Humid Temperate, Humid Tropical, Dry, and Polar. Below the Domain level is a finer level, called the "Division." Each of the 24 Divisions is a member of exactly one Domain. Below the Division level is the "Province" level; each of the 51 Provinces a member of only one Division. The finest level is the "Section." Each of the 192 Sections is a member of only one Province.

The hierarchical structure permits one to construct a 192x192 similarity matrix **S** among ecological regions at the "Section" level. If Section k is identical to Section p, then S_{kp}=4; if Section k is in the same Province as Section p, then S_{kp}=3; if Section k is in the same Division as Section p, then S_{kp}=2; if Section k is in the same Domain as Section p, then S_{kp}=1; and if Section k is not in the same Domain as Section p, then S_{kp}=0. With this similarity matrix, one may then construct the following ecological similarity measure between a pair of LMAs:

$$w_{ij} = \sum_{k=1}^{m} \sum_{p=1}^{m} \left(p_{ik} p_{jp} S_{kp} \right) \tag{15}$$

Where p_{ik} is the percentage of the land area in LMA i in ecological Section k, and p_{jp} is the percentage of the land area in LMA j in ecological Section p. Equation (15) can be interpreted as the expected ecological similarity between a hectare of land drawn at random from LMA i and a hectare of land drawn at random from LMA j. Table 1 summarizes the weight matrix created from ecological data.

Labor Flow Relationships

Commuting flows have been extensively used to delineate the relationships among counties. The *Metropolitan Statistical Area* (United States, Office of Management and Budget 2000) and the Bureau of Economic Analysis *Economic Area* (Johnson 1995) are both defined by assigning to a large urban county the surrounding counties linked by strong commuting flows. One might also use commuting flows to delineate the linkages among large urban counties (Eff 2003). Flows among LMAs would represent phenomena such as intra-firm temporary assignments of workers to branch offices, or the provision of skilled services by highly specialized firms serving national or global markets. An LMA would be more closely related to those LMAs with which it has relatively strong commuting flows.

Commuting flows are detailed in the decennial census Journey to Work data. The data are at the county level, and can be aggregated up to the LMA level. Journey to Work data are available from the Censuses of 1970, 1980, 1990, and 2000. The Bureau of Economic Analysis includes these data on its annual Regional Economic Information System CD-ROM (United States, Bureau of Economic Analysis 2003). In the present study only the data from 1990 and 2000 are used.

Longer term labor flows among LMAs would be found in population migration data rather than commuting data. The STP-28 file from the 1990 census gives inter-county migration counts for persons by age, sex, race, educational attainment, nativity, and poverty status. Thus, working age migrants can be singled out, to identify labor flows. These data can be aggregated to identify migration patterns at the LMA level. One might expect working age population flows to be particularly useful in delineating longer distance flows, such as those between LMAs at the middle levels of the central place hierarchy and their higher order centers (Eff 2003).

Equation (16) represents how commuting and migration are employed in the weight matrix:

(16) $$w_{ij} = \left[\ln\left(1 + x_{ij}\right)\ln\left(1 + x_{ji}\right)\right]^{1/2}$$

where x_{ij} is the flow of commuters (or working age migrants) from LMA i to LMA j. Table 1 summarizes the three weight matrices **W** created from labor flow data.

Similarity in Level of Development

Regions 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-regional economic studies: 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). Measures for U.S. counties can be found for all three dimensions. The Bureau of Economic Analysis, on its annual Regional Economic Information System CD-ROM, provides a county-level measure of per capita income (United States, Bureau of Economic Analysis 2003). The average LMA per capita income between 1990 and 2000 is taken as the measure of prosperity. The 2000 Census, in the STF3 data, provides a count of the number of persons, age 25 or older, in each educational attainment category in each county. Summing up to the LMA level, one can then calculate the average educational attainment level in each LMA. This provides a measure of knowledge. Murray, Michaud, McKenna, and Marks (1998), of the Harvard Center for Population and Development Studies, provide a county-level measure of life expectancy. From this one can calculate the population-weighted average life expectancy for each LMA. This provides a measure of life.

Each of the three measures is standardized with a mean of 100 and a standard deviation of 15. Any LMA therefore has a specific location given by its scores for life, knowledge, and prosperity within a three-dimensional space. The proximity in level of development between LMAs i and j is calculated as the inverse Euclidean distance as follows:

(17) $$w_{ij} = \left(\sum_{k}^{3}\left(x_{ik} - x_{jk}\right)^2\right)^{-1/2}$$

where x_{ik} is the level of development of LMA i in dimension k, and x_{jk} is the level of development of LMA j in dimension k. Table 1 summarizes the weight matrix created in this section.

The geometric mean of the standardized measures for life, knowledge, and prosperity can be used as a measure of an LMA's level of development. Figure 3 maps this level of development measure.

Similarity in Economic Structure

Economists are often interested in the similarity of regions in economic dimensions. There are two kinds of data on the economic structure of regions. One categorizes employment by occupation, the other categorizes employment or establishments by industry. The location of industrial districts, and the understanding that a regional economy can be tied to a particular set of industries, underlies much of regional economics. Agglomeration economies can encourage industries to cluster together, and export-base analysis and input-output analysis are just some of the tools regional economists have used to understand the effects of industrial clustering (Richardson 1979). Beginning in the 1980s, regional economists began to acknowledge that the salient differences among regions might more easily be shown in the occupational structure than in the industrial structure. A multi-locational firm would tend to concentrate functions requiring unskilled labor in regions where labor was cheap but willing, while it would concentrate skilled functions in regions with highly educated workers and with business services that provided information. Thus, a region's comparative advantage might not be in a particular industry, but in a particular kind of worker.

Data on county occupational structure can be found in the Equal Employment Opportunity Files, produced by the Bureau of the Census (1992). For each county, the number of workers are given for each of 512 occupational categories. Data on county industrial structure can be found in County Business Patterns (United States, Bureau of the Census 2003).

Two types of proximity indices are used. The first is the inverse of the Euclidean distance between the two LMAs:

(18) $$w_{ij} = \left(\sum_k \left(p_{ik} - p_{jk} \right)^2 \right)^{-1/2}$$

Where p_{ik} is the percentage of the establishments (labor force) in LMA i in NAIC category (occupational category) k, and p_{jk} is the percentage of the establishments (labor force) in LMA j in NAIC category (occupational category) j. The second proximity index is a Herfindahl-type index:

$$w_{ij} = \sum_k \left(p_{ik} p_{jk} \right) \tag{19}$$

Where the notation is as in Equation (18). Table 1 provides a summary description of the four weight matrices designed to measure proximity among LMAs in economic structure.

Similarity in Regional Role

Spatial weight measures based on physical distance or contiguity assume that the most important relationships among regions are those between adjacent regions. Innovations in one region, for example, diffuse to neighbors, and then to others further away. A little reflection, however, shows that this may be too simple. Diffusion of innovations are likely to be unidirectional—from urban centers to rural areas, and not from rural areas to urban centers. Large urban centers are likely to receive innovations from even larger urban centers, in the hierarchical structure given by Central Place Theory (Christaler 1933). One can also imagine that similar processes might also affect all regions occupying a similar regional role, even though there is no necessary inter-regional transmission of innovations. For example, a national change in the mortgage interest rate might spur homebuilding in rural areas on the periphery of large urban centers. These peripheral rural areas exhibit very similar behavior, but it is behavior that differs from their neighbors (large urban centers on one side and isolated rural counties on the other). Relative to physically proximate neighbors, population growth and housing construction may even be negatively autocorrelated. Relative to other regions in similar central place roles, however, population growth and housing construction would be positively autocorrelated.

Four proximity measures are created to address some of the issues raised by similarity in regional role. The first measure employs the regional codes created by Calvin Beale, and made available by the USDA ERS. The code ranges from 0 to 9, where lower numbers indicate counties with larger populations closer to urban centers, and higher numbers indicate counties with smaller populations further away from urban centers. Equation (20) presents the proximity measure used for the Beale codes:

$$w_{ij} = \left(1 + \left| \bar{c}_i - \bar{c}_j \right| \right)^{-1} \tag{20}$$

where $\bar{c}_i$ is the population weighted mean of the Beale codes of counties in LMA *i*, and $\bar{c}_j$ is the population weighted mean of the county Beale codes in LMA *j*. This proximity matrix W should reflect that similar processes might occur simultaneously in regions at a similar level of the Central Place hierarchy, without the processes occurring in their immediate neighbors.

The other three measures attempt to capture how LMAs might acquire innovations from other LMAs at a higher level of the Central Place hierarchy. Using the methodology developed in Eff (2004a), flows among n regions can be depicted in an n x n matrix **F**, where the flow from region I to region j is given by each element f_{ij}. Flows can be converted to percentages, to dampen the effect of differential region sizes, in an n x n matrix **P**, where the flow from region I to region j is given by each element $p_{ij} = f_{ij} / \sum_j f_{ij}$. The element-wise geometric mean of **P** and its transpose creates a symmetric matrix **A**, where each element $a_{ij}=(p_{ij}p_{ji})^{½}$. Setting the diagonal equal to zero, one can use **A** to calculate a centrality vector **c**:

(21) $$\lambda c = Ac$$

where λ is an eigenvalue and **c** is an eigenvector of matrix **A** (Strang 1980: 181). The first principal eigenvector from Equation (21) has long been used by network analysts to calculate network centrality scores (Wasserman and Faust 1994: 207; Bonacich 1987). The centrality score for any region i is given below:

(22) $$c_i = \frac{1}{\lambda}\sum_{j=1}^{n} a_{ij}c_j \ , \qquad a_{ii} = 0 \ , \ \forall i$$

Tam (1989) describes this specification as *endogenous* centrality, since the centrality of a region i is a function not only of the magnitude of the flows a_{ij} between i and j, but also of the centrality of the regions j to which it is connected. A peripheral region will have flows directed primarily at regions with low centrality—it will lie low on the hierarchy, with upward connections primarily to centers that themselves are not particularly high on the hierarchy. A central region will have flows directed primarily toward regions with high centrality. One can base the proximity matrix **W** directly on Equation (22), as follows:

(23) $$w_{ij} = \frac{x_{ij}}{\sum_j x_{ij}}, \ \ where \ \ x_{ij} = \frac{a_{ij}}{\lambda}c_j$$

In effect, one weights the flows between LMA i and LMA j by the centrality of the destination LMA. Thus, *ceteris paribus,* an LMA is "more proximate" to a more central LMA than to a less central LMA. Equation (24) shows how one may select only the upward links:

(24) $$x_{ij} = \frac{a_{ij}}{\lambda}c_j \quad if \ \ c_j > c_i, \ \ x_{ij} = 0 \ \ otherwise$$

One may then create a proximity measure showing the relative importance to LMA i of its upward link to LMA j:

(25) $$w_{ij} = \frac{x_{ij}}{\sum_j x_{ij}}$$

Equation (24) can also be used as a basis for selecting each region's single most important upward link:

(26) $$w_{ij} = 1 \;\; if \;\; x_{ij} = \max_j (x_{ij}), \;\; w_{ij} = 0 \;\; otherwise \qquad \forall i$$

Note that the link to the selected region provides the single largest contribution to region i's own centrality score, and thus is the single greatest source of connectedness to the network of regions. Table 1 summarizes the four weight matrices **W** that depict some aspect of similarity in regional role.

Modifications to Proximity Matrices prior to Testing for Autocorrelation

The diagonal of the proximity matrix **W** is set to zero, and the matrix is made symmetric by replacing each element w_{ij} by the largest of the elements w_{ji} and w_{ij}. The matrix is then standardized, first by subtracting from each element w_{ij} the smallest of the off-diagonal elements w_{ij}, and then by dividing each element w_{ij} by the largest of the elements w_{ij}. Thus, each w_{ij} ranges from zero to one. When using **W** to test for autocorrelation, each element w_{ij} is divided by the row sum ($\Sigma_i w_{ij}$) so that the sum of the row elements equals one.

While only part of the standardization (setting the diagonal equal to zero and setting the row sums equal to one) is strictly necessary for an autocorrelation weight matrix, nevertheless converting **W** to a symmetric matrix does ensure that the proximity measures conform to the formal symmetry property of a distance metric (a proximity measure is essentially an inverted distance measure). The symmetry property is simply that the distance between LMA i and LMA j equals the distance between LMA j and LMA i (Rektorys 1969: 998). While proximity measures based on inverse Euclidean distance are always symmetric, the same is not true of the flow-based measures.

The most attractive feature of the Herfindahl-type measure is its intuitive meaning. In the case of Equation (5), for example, the proximity measure can be interpreted as the probability that a person drawn at random from LMA i is of the same ancestry as a person drawn from LMA j. Nevertheless, Herfindahl-type measures violate a formal property of distance measures, that the distance between an LMA i and itself equal zero, and the distance between an LMA i and all other LMAs be greater than zero (Rektorys 1969: 998). Paradoxically, a Herfindahl-type measure often returns the result that an LMA is closer to another LMA than to itself. For example, suppose there are 51 ancestry groups, and LMA *i* has 50 percent of its population in ancestry group "white," and the remaining 50 percent scattered evenly in 50 other ancestry groups. The proximity between LMA *i* and itself (as given in Equation (5)) would equal $.5^2+50*(.01)^2=0.255$. Suppose LMA *j* has 90 percent of its population in ancestry group "white," and the remaining 10 percent scattered evenly in 50 other ancestry groups. The Herfindahl-type proximity between LMA *i* and LMA *j* would equal .5*.9+50*(.01*.002)=0.45002. Thus LMA *i* is closer to LMA *j* than it is to itself.

The following section compares the weight matrices described in the present section. Using matrix correlation, one can produce a set of stylized facts regarding the relationships among regions.

3. COMPARISON OF WEIGHT MATRICES

Table 2a presents the matrix correlation (Wasserman and Faust 1992: 686) between each pair of the 35 matrices described above. To fit the table to the page, only the rounded t-statistic is given. A negative value indicates the correlation is negative. The standard error is determined from a permutation test as follows. For each pair of matrices, the elements of one matrix are randomly rearranged, and the matrix correlation between the pair is calculated. The procedure is repeated 1,000 times to give a sample of matrix correlation coefficients. The standard deviation from this sample is used as the standard error. If the correlation is insignificant (with one-sided p-value >0.05), then the rounded t-statistic is replaced in Table 2a by a zero.

Table 2b summarizes Table 2a by showing, for each weight matrix, the number of times the results of Table 2a are negatively correlated, uncorrelated, or positively correlated. Eighty-five percent of the time, the matrices correlate positively with each other. This result suggests that inter-regional relationships tend to be reflected in multiple dimensions—that, for example, a pair of regions strongly tied in a distance dimension tend also to be strongly tied in other dimensions such as religion, ancestry, and economic structure. One can readily see that *pnsiml* (place name similarity) is the most anomalous in the sense that it correlates positively the fewest times with other matrices—only 17 times out of 34, followed closely by *edcbc* (Inverted Absolute Difference Avg. Calvin Beale Codes 1993)—only 19 times out of 34. The matrices *pop* (working age migration 1985-1990), *hicbp* (Herfindahl-type County Business Patterns 2001), and *edpres* (inverse Euclidean Distance, presidential elections) all returned at least five significant negative correlations.

Table 2b also reports the probability that a weight matrix will have a higher t-statistic than other weight matrices in a given row. This figure gives some sense of which matrices are most strongly correlated with other weight matrices. One can see that *contig* (physical contiguity) is most likely to have the highest t-statistic among the 34 weight matrices in any given row (higher 88 percent of the time), while *pnsiml* (place name similarity) is least likely to have the highest t-statistic (higher 16 percent of the time).

Anomalous weight matrices—*i.e.,* ones that correlate poorly with other weight matrices—are perhaps especially interesting, since they express interregional relationships that are not expressed by other matrices. When the weight matrices are used to examine autocorrelation, it may be that the anomalous weight matrices work particularly well for a given variable—that anomalous matrices show high autocorrelation for this variable, whereas other matrices show little or none. Such results would suggest that the forces determining the values taken on by this variable are uniquely modeled by the weight matrix. On the other hand, the weight matrices highly correlated with the weight matrix *contig* (physical contiguity) can probably be proxied by *contig* in autocorrelation tests on variables. One might thus argue that the current practice—using exclusively physical proximity weight matrices—is entirely correct *unless* one finds that the anomalous matrices return higher autocorrelation for certain classes of variables.

Matrix correlations can be used to establish some stylized facts about inter-regional relationships. For example, the negative correlations for *pop* are suggestive: working age population flows are likely to be larger between regions with dissimilar ancestry (*elem, pumsanc*), with dissimilar economic structure (*hicbp, hiocc*), and with dissimilar Calvin Beale codes (*edcbc*). Migration flows between LMAs apparently work to dissipate ethnic concentrations, unlike the migration flows within LMAs which typically lead to increased racial segregation. And migration flows apparently typically take place either up or down the central place system (not laterally), and between areas with dissimilar economic bases.

The results differ in suggestive ways from matrix correlations of international weight matrices (Eff 2004). *Interregional* matrix correlations show that physical distance weight matrices are the most *highly* correlated with other weight matrices; *international* matrix correlations show that physical distance is the most *weakly* correlated with others. Thus, in international relationships, physical proximity is not as closely associated with cultural and other dimensions of measuring proximity as it is in interregional relationships.

4. AUTOCORRELATION FOR EXPLORATORY DATA ANALYSIS

Table 3 presents 205 variables drawn from an array of regional datasets. The data include crime variables, health and demographic variables, variables measuring income distribution and income levels and a variety of other characteristics of regions. These data can be examined for autocorrelation, using the 35 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 35 different weight matrices, one might gain some sense of how the different weight matrices perform. The weight matrices reporting the highest levels of autocorrelation would be those that most accurately represent inter-regional relationships. The autocorrelation statistic used here is Moran's I (Odland 1988; Anselin 1988):

$$I = \frac{n}{\sum_i \sum_j w_{ij}} \cdot \frac{\sum_i \sum_j w_{ij}\left(x_i - \bar{x}\right)\left(x_j - \bar{x}\right)}{\sum_i \left(x_i - \bar{x}\right)} \tag{27}$$

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).

With 205 variables and 35 weight matrices, a total of 7,175 autocorrelation statistics are calculated. Tables 4, 5, and 6 attempt to report the results in a condensed fashion. Table 4 reports, for each weight matrix, the percent of the 205 autocorrelations that were significant at the 0.10, 0.05, and 0.00000 size of test. The fourth column of figures reports the net probability that the weight matrix has a higher Moran-z than any other weight matrix, for any variable.[5] The net probability is calculated by comparing, for each variable, all pairs of Moran z-scores. The percent

[5] The net probability compares the Moran z-scores, rather than the p-value of the Moran-I. Since the p-values equal zero at five significant decimal places when the Moran-z is about 4.5, the p-value is effectively truncated. Comparing the Moran-z therefore allows a more complete comparison.

of time that a weight matrix has a lower Moran-z in a pairwise comparison is subtracted from the percent of time that the weight matrix has a higher Moran-z. A large positive net probability indicates that the weight matrix usually outperforms (has a higher Moran z-score than) other weight matrices. A negative number indicates that the weight matrix is usually outperformed by (has a lower Moran z-score than) other weight matrices. Thus one can see that the weight matrix *dist*, with a net probability of 69.0 percent, usually has a higher Moran z-score than other weight matrices, while the weight matrix *pnsiml*, with a net probability of -74.0 percent, usually has a lower Moran z-score.

The fifth column of figures calculates the net probability that a weight matrix will have a higher Moran z-score than other weight matrices *in its category*. Thus, for example, when compared to other weight matrices in the category *Ancestry,* the weight matrix *anhi2000* has a net probability of 71.3 percent. These net probability figures are used to select the best-performing weight matrix in each category.

Perhaps the most striking result of Table 4 is the very high degree of autocorrelation, for all weight matrices. Even the poorest performing weight matrix (*pnsiml*—place name similarity) produces significant autocorrelations (at the 0.05 size of test) 32.3 percent of the time. A z-score of 4.5 will return a p-value equal to zero at five decimal places. The best overall weight matrix (*dist*—inverse squared great circle distance) has a p-value equal to zero 82.4% of the time.

Distance performs very well, as do labor flows, and the regional structure matrices derived from commuting flows. One of the ancestry matrices (*anhi2000*) performs very well, though the modified Herfindahl-type similarity measures (*elem, pumsanc, rs1890, rs1926, rs1936, rs1990*)—so difficult to create and from which good results were expected—performed poorly. The only exception to poor performance of the similarity measures is *simleco*—the similarity in ecological regions. The disappointing performance of similarity measures echoes the results of international comparisons in Eff (2004), where a simple binary matrix based on Huntington's Civilizations outperformed (albeit only slightly) a sophisticated weight matrix based on language similarities. Similarity measures have not shown that they are worth the trouble.

Table 5 presents the net probability of having a higher Moran z-score for the nine weight matrices that performed best in their categories.[6] The results are broken down by variable category, to give

[6] There are 11 weight matrix categories, but the best-performing matrices in the Place Name and Industries categories performed so poorly that they were dropped. The categories, and the best performing matrices, are: Ancestry (*anhi2000*), Distance (*dist*), Elections (*edpres*), Regional Structure (*evnup*), Religion (*hi90*), Occupation (*hiocc*), Labor Flow (*n10*), Ecology (*simleco*), and Level of Development (*sliv*).

some sense of how weight matrices compare with each other in each variable category. Table 6 summarizes the results of Table 5, showing the overall net probability of having a higher Moran z-score for each weight matrix in each variable category. The number in parentheses is the rank net probability (1=highest, 9=lowest) for the row.

Unquestionably, *Distance* is the best performing weight matrix in Table 6. Never having a negative net probability, *Distance* is always in the top five and turns in the highest net probability for seven of the 17 variable categories. This is a reassuring result, since it suggests that current practice (using distance as the basis of weight matrices in analysis of regional data) is the best single option.

Level of Development appears to be the second best performing matrix, but since the matrix is based on the Euclidean distance between regions in *Education*, *Life Expectancy*, and *Income*, its top performance in those variable categories is nothing more than a tautology. The matrix's high performance on *Male/Female Occupation Difference*, however, is not tautological, and suggests that interregional variations in the sexual division of labor may have more to do with variation in level of development than with variations in ancestry and religion, or diffusion from neighboring regions.

Ancestry performs quite well, ranking first in four out of 17 variable categories—five out of 17 if one considers that its number two rank for *Life Expectancy* is only because of the tautological assignment of first rank to *Level of Development*. *Ancestry* performs particularly well for some of the variables representing household structure: *Percent Children in Arrangement*, *Percent Households by Household Type*, and *Marital Status*. Both *Sex Ratio* and *Life Expectancy* vary systematically across ethnic groups,[7] and it is therefore not surprising that *Ancestry* performs best among the weight matrices for these two variable categories.

Religion performs especially well for *Education*, *Life Expectancy*, and *Level of Development* variables. The two cultural weight matrices—*Ancestry* and *Religion*—appear especially important in explaining variation in variables that measure social structure and parental investment.

Regional Structure performed much better than other weight matrices for *Politics*, a variable category consisting of only one variable: the relative conservativeness of the voting record of the region's congress members. This surprising result suggests that ancestry, religion, and the level of

---

[7] The sex ratio at birth varies across populations, and is unusually low for African-Americans (Draper and Harpending 1988; Draper 1989; Posner 1992: 137-138).

development are less important than the situation of a region in the regional structure (the variety and intensity of its connection to higher order centers) in determining its political orientation.

Labor Flows performs somewhat better than *Regional Structure*, but both perform more poorly than *Distance*. The relations among regions apparently are better modeled through distance than through more specific measures. Note that the amount of work that went into developing the weight matrix is inversely related to its performance: *Distance* is the easiest to produce, followed by *Labor Flows*, followed by *Regional Structure*.

Ecology performed poorly, with modest success only for two variable categories: *Environment* and *Population Growth. Elections* also performed poorly, even for the variable category *Politics. Occupations* was another poor performer, though it ranked second in *Male/Female Occupation Difference* (perhaps tautological) and in *Income Growth 1969-2001*.

Moran's I can be used to test hypotheses of the form "Culturally (physically, etc.) proximate regions tend *not* to have similar ___," where the blank could be filled with: "per capita income," "expenditure on schools," "Congressional voting records," etc. For example, for the variable *rating* (*National Journal* Congressional Representative Conservative Rating 2004), one can interpret the Moran's I figures as the test statistics for the null hypothesis: "Proximate regions tend *not* to have Congressional Representatives with a similar voting record." One can then see that the null hypothesis is not rejected for only two of the nine best weight matrices (*Ancestry* and *Occupation*). Thus, while variations in ancestry apparently do not account for variations in the conservativeness of Congressional Representatives, variations in religion do. The result could be used in support of an argument that the salient dividing line between conservative and liberal Americans is *not* based on ethnic affiliation, but rather on other factors such as religion, or the conservativeness of neighboring regions.

To a certain extent, there is a bit of ambiguity about whether *Distance* is really the best performing weight matrix. *Level of Development*, *Regional Structure*, and *Labor Flows*, after all, return a higher percent of significantly autocorrelated variables at the 0.05 size of test (Table 4). In addition, there is the question perhaps most important for the user of weight matrices: Are all of the weight matrices picking up the same patterns of variation, or are there important differences among them? This question has been partly answered by Table 2. One can see that the weight matrices are not always significantly correlated with each other, and are occasionally even negatively correlated. From Tables 5 and 6 one can also infer that for certain categories of data—perhaps especially data on family structure—autocorrelation on cultural dimensions would prove more significant than autocorrelation on the dimension of physical distance.

5. SUMMARY AND CONCLUSIONS

The paper shows how autocorrelation in regional datasets can be defined spatially, culturally, ecologically, and through the intensity of labor flows. Thirty-five weight matrices were constructed, each showing interregional relationships in a different dimension. The 35 weight matrices were then compared, using matrix correlation, to evaluate the degree to which they differed in describing the strength of ties among regions (Tables 2a and 2b). Then, to assess the prevalence of autocorrelation in regional data, a sample of 205 variables were drawn from a wide variety of sources. Moran's I was used to test these 205 variables, using the 35 weight matrices. Autocorrelation existed at the .95 level of significance about 77 percent of the time, with results for individual weight matrices ranging from as high as 91 percent to as low as 32 percent (Table 4). The results demonstrate that autocorrelation is more likely than not in regional data.

Current practice in regional economics is to use physical proximity matrices—especially physical contiguity matrices—when calculating autocorrelation statistics. The results in this paper suggest that weight matrices based on physical distance are the best single option for uncovering patterns of autocorrelation. The two physical distance-based matrices—and especially the contiguity matrix—exhibited the highest degree of correlation with other matrices, a result suggesting that interregional patterns of cultural proximity, ecological proximity, labor flow proximity, and level of development proximity are often such that adjacent regions happen to be the most closely related. Thus, autocorrelation in a variety of dimensions may be proxied by using a physical proximity weight matrix. One might even go a step further and say that physical proximity provides a more accurate picture of the relations among regions than any single one of these other weight matrices. If similarity among regions is due to multiple processes (cultural, ecological, etc.), then no single one of these will tell the whole story, but physical proximity may provide a useful composite measure. Evidence in favor of this view can be seen in the autocorrelations on the sample of 205 variables. Here, the overall best-performing weight matrix was physical proximity, a result that suggests that physical proximity provides a more accurate view of the similarity among regions than does any other single weight matrix.

Nevertheless, among some classes of variables, cultural autocorrelation appears to be stronger than physical distance autocorrelation. Table 6 shows that variables measuring aspects of family structure, the level of development, education, life expectancy, and the sex ratio were all likely to have their highest autocorrelation with cultural weight matrices—*viz* ancestry or religion. One might argue that for these classes of variables, variation among regions might be due to variations in borrowing or inheritance of cultural traits. Cultural traits such as patterns of marital status, for example, may be similar in regions that are similar in their ethnic composition, or similar in religious affiliation. Since physically neighboring regions are likely to be similar in ethnic composition or religion, physical proximity weight matrices will return significant autocorrelation statistics, but the autocorrelation will be stronger with weight matrices describing the actual salient source of similarity—ancestry and religion.

Compared with previous work on autocorrelation in international datasets (Eff 2004), U.S. regional data exhibit a few differences. The most important of these may be that physical proximity weight matrices are poorly correlated with cultural weight matrices in international data, but very highly correlated with cultural weight matrices in regional data. Thus, when looking at the relationships among nations, physical proximity does not provide a good proxy for cultural proximity, whereas it does when examining the relationships among U.S. regions. The implication is that the current practice of regional economists—using exclusively physical proximity weight matrices—is probably good enough to capture cultural proximity, whereas economists using international datasets should probably use cultural proximity weight matrices in addition to physical proximity. Currently, economists using international datasets do not normally conduct autocorrelation tests—whether cultural or spatial—though the amount of autocorrelation present in international data (at the .95 significance level, 86 percent of the trials using 72 sample variables with 12 weight matrices) is no less than the amount of autocorrelation in regional data (at the .95 significance level, 77 percent of the trials using 205 sample variables with 35 weight matrices).

Four different methods of specifying weight matrices from attribute data are used: inverse Euclidean distance, Herfindahl-type, binary, and similarity-weighted Herfindahl-type. The similarity-weighted Herfindahl-type matrices all perform relatively poorly (Table 4), and since they are the most difficult to construct, it seems that this type of weight matrix is a poor choice. The performance of inverse Euclidean distance and Herfindahl-type weight matrices is hard to distinguish—sometimes one performs better, sometimes another. The binary weight matrices (*contig, samec8, sameancest*) are relatively easy to construct, they produce sparse weight matrices (*i.e.*, with many zero elements—providing computational advantages), and they perform relatively well. One might also make the generalization that weight matrices based on more recent data outperform matrices based on older data—a pattern most evident with the commuting flow weight matrices.

Autocorrelation measures are useful for the production of stylized facts through tests on null hypotheses such as “regions similar in religion tend not to have similar murder rates.” The weight matrices can also be directly used to produce stylized facts, by employing matrix correlation to test null hypotheses such as "regions with stronger commuting ties tend not to be more similar in religion." Autocorrelation measures are also useful for exploratory data analysis, since comparing the autocorrelation results from several different weight matrices can give some sense of whether the processes causing interregional variation in a variable are more likely to be rooted in cultural, economic, or level of development factors. For example, the variable measuring sex differences in occupational status returns higher autocorrelation with level of development than with ancestry, religion, or distance (Table 6), suggesting that variations in female occupational status may have more to do with level of development than with other factors. Economists, however, because of their emphasis on regression analysis, are most likely to find autocorrelation statistics useful for examining regression residuals. The presence of autocorrelated residuals is likely to indicate an omitted variable problem. One can then use the appropriate weight matrix to create a "spatially-lagged" variable that serves as an instrument for the omitted variables.

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

Bonacich, Phillip. 1987. "Power and Centrality: A Family of Measures." *American Journal of Sociology*. 92(5): 1170-1182.

Bradley, Martin B. 1992. *Churches and Church Membership in the United States, 1990: An Enumeration by Region, State, and County, Based on Data Reported for 133 Church Groupings*. Atlanta, Georgia: Glenmary Research Center.

Butler, Jon. 1990. *Awash in a Sea of Faith: Christianizing the American People*. Cambridge: Harvard University Press.

Can, Ayşe. 1998. "GIS and Spatial Analysis of Housing and Mortgage Markets." *Journal of Housing Research* 9(1): 61-86.

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

Christaller, Walter. 1933. *Die zentralen Orte in Süddeutschland*. Jena: Gustav Fischer. (Translated by Carlisle W. Baskin, as *Central Places in Southern Germany*. Englewood Cliffs, NJ: Prentice Hall 1966).

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

Clubb, Jerome M., William H. Flanigan, and Nancy H. Zingale. 1986. *Electoral Data for Counties in the United States: Presidential and Congressional Races, 1840-1972*. Ann Arbor, MI: Inter-university Consortium for Political and Social Research.

Conkin, Paul Keith. 1990. *Cane Ridge, America's Pentecost*. Madison: University of Wisconsin Press.

Daly, Chris, and George Taylor. 2000. *United States Average Annual Precipitation, 1961-1990.* [Computer file]. Corvallis, OR: Spatial Climate Analysis Service, Oregon State University. http://nationalatlas.gov/atlasftp.html

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.

Draper, P. 1989. "African Marriage Systems: Perspectives from Evolutionary Ecology", *Ethology and Sociobiology.* 10: 145-69.

Draper, P. and Harpending, H. 1988. "A Sociobiological Perspective on the Development of Human Reproductive Strategies", in K.B. MacDonald (ed) *Sociobiological Perspectives on Human Development*. pp. 340-372. New York: Springer Verlag.

Eff, E. Anthon. 2004. "Spatial and Cultural Autocorrelation in International Datasets." *MTSU Department of Economics and Finance Working Papers*. February 2004. http://www.mtsu.edu/%7Eberc/working/spatial%20autocorrelation.pdf

Eff, Ellis. 2004a. "Eliciting a Hierarchy of Regions: Network Analysis and Interregional Flows." presented at The 100th Association of American Geographers' meeting, Philadelphia, Pennsylvania, March 18, 2004.

Eff, E. Anthon. 2003. "On the Relationships of Regions: Network Analysis and Interregional Flows." presented at The 42nd Southern Regional Science Association meeting, Louisville, Kentucky, April 11, 2003.

Elazar, Daniel. 1972. *American Federalism: A View from the States.* Second Edition. New York: Thomas Y. Crowell.

Fischer, David Hackett. 1991. *Albion's Seed: Four British Folkways in America*. Oxford: Oxford University Press.

Foster, George. 1960. *Culture and Conquest: America's Spanish Heritage*. Chicago: Quadrangle Books.

Garreau, Joel. 1981. *The Nine Nations of North America*. New York: Avon.

Gastil, Raymond D. 1975. *Cultural Regions of the United States*. Seattle: University of Washington Press.

Goetz, Stephan J. 1999. "Migration and Local Labor Markets." in *The Web Book of Regional Science* (www.rri.wvu.edu/regscweb.htm), ed., Scott Loveridge. Morgantown, WV: Regional Research Institute, West Virginia University.

Granovetter, Mark S. 1973. "The Strength of Weak Ties." *American Journal of Sociology*. 78: 1360-1380.

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.

Hayek, Friedrich A. 1979. *Law, Legislation, and Liberty.* Chicago: University of Chicago Press.

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

Isard, Walter. 1960. *Methods of Regional Analysis: An Introduction to Regional Science*. Cambridge, Massachusetts: MIT Press.

Johnson, Kenneth P. 1995. "Redefinition of the BEA Economic Areas." *Survey of Current Business*. February, 1995: 75-80.

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

Killian, Molly S. and Charles M. Tolbert. 1993. "Mapping Social and Economic Space: The Methodology of Defining Local Labor Markets." Pp. 69-79 in Singelmann, Joachim and F.A. Deseran (eds.). *Inequality in Local Labor Markets*. Boulder: Westview.

Korotayev, Andrey. 2004. *World Religions & Social Evolution Of The Old World Oikumene Civilizations: A Cross-cultural Perspective*. Edwin Mellen Press.

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.

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

Mizruchi, Mark S., Peter Mariolis, Michael Schwartz, and Beth Mintz. 1986. "Techniques for Disaggregating Centrality Scores in Social Networks." In Tuma, Nancy Brandon (editor) *Sociological Methodology* 1986, pp. 26-48. Washington, D.C.: American Sociological Association.

Moody, James. 2000. "SPAN: SAS Programs for Analyzing Networks" The Ohio State University, Columbus, OH.

Muller, Jerry Z. 2002. *The Mind and the Market: Capitalism in Modern European Thought*. New York: Alfred A. Knopf.

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

Murray, C.J.L., C.M. Michaud, M.T. McKenna, and J.S. Marks. 1998. "U.S. Patterns of Mortality by County and Race: 1965-1994." *Harvard Center for Population and Development Studies*. http://www.hsph.harvard.edu/organizations/bdu/Onlinedata_files/county.htm

National Journal. 2004. *2003 Vote Ratings*. Friday, Feb. 27, 2004. Accessed September 2, 2004. http://nationaljournal.com/members/news/2004/02/0227nj1.htm

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.

Odum, Howard, and Harry Moore. 1938. *American Regionalism: A Cultural-Historical Approach to National Integration.* New York: Henry Holt.

Phillips, Kevin P. 1969. *The Emerging Republican Majority*. New Rochelle, NY: Arlington House.

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

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

Rektorys, Karel. 1969. *Survey of Applicable Mathematics.* Cambridge, Massachusetts: MIT Press.

Richardson, Harry W. 1979. *Regional Economics*. Urbana: University of Illinois Press.

Ross, Sheldon M. 1980. *Introduction to Probability Models* (second edition). New York: Academic Press.

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.

Tam, Tony. 1989. “Demarcating the Boundaries between Self and the Social: The Anatomy of Centrality in Social Networks.” *Social Networks*. 11: 387-401.

Turner, Frederick Jackson. 1893. "The Significance of the Frontier in American History." Address at the American Historical Association. Chicago. 12 July 1893.

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, Bureau of Economic Analysis. 2003. *Regional Economic Information System 1969-2001.* [Computer file]. Washington, DC: Bureau of Economic Analysis.

United States, Bureau of the Census. 2004a. *Summary File 3, 2000 Census of Population and Housing: Technical Documentation.* June 2004. Accessed July 2004. http://www.census.gov/prod/cen2000/doc/sf3.pdf

United States, Bureau of the Census. 2004b. *Public Use Microdata Sample, 2000 Census of Population and Housing: Technical Documentation.* July 2004. Accessed July 2004. http://www.census.gov/prod/cen2000/doc/pums.pdf

United States, Bureau of the Census. 2003. *Federal, State, and Local Governments: 2001 Public Elementary-Secondary Education Finance Data.* http://www.census.gov/govs/www/school01.html

United States, Bureau of the Census. 2002. *County Business Patterns: 2000.* Accessed July 2004. http://www.census.gov/epcd/cbp/download/dwncbp00.html

United States, Bureau of the Census. 1992. *Census of Population and Housing, 1990: Equal Employment Opportunity (EEO) File*. [Computer file].Washington, DC: Bureau of the Census.

United States, Department of Agriculture, Forest Service. 2004. *Ecoregions and Subregions of the United States, Puerto Rico, and the U.S. Virgin Islands*. Reston, VA: U.S. Geological Survey. http://nationalatlas.gov/atlasftp.html

United States, Department of Justice, Federal Bureau of Investigation. 2004. *Uniform Crime Reporting Program Data: County-Level Detailed Arrest and Offense Data, 2002*. [Computer file]. ICPSR ed. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [producer and distributor], 2004.

United States, Geological Survey. 2001. *Urban Areas of the United States*. Reston, Virginia: U.S. Geological Survey. http://nationalatlas.gov/atlasftp.html

United States, Office of Management and Budget. 2000. "Standards for Defining Metropolitan and Micropolitan Statistical Areas." *Federal Register* 65(249): 82228-82238. http://www.census.gov/population/www/estimates/00-32997.pdf

Veblen, Thorstein Bunde. 1899. *Theory of the Leisure Class.* New York: MacMillan.

Wasserman, Stanley and Katherine Faust. 1994. *Social Network Analysis: Methods and Applications*. Cambridge: Cambridge University Press.

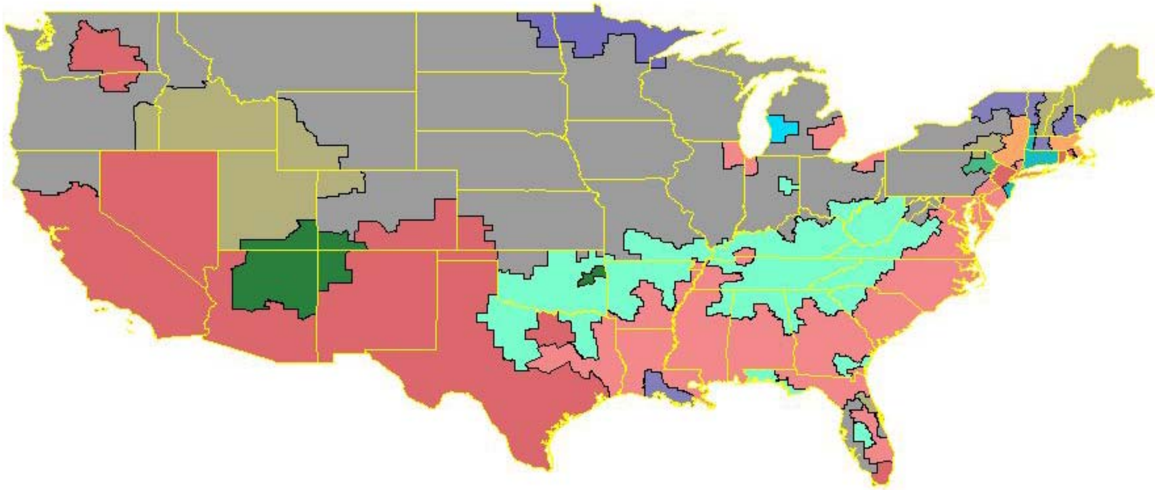
White, Douglas R., and Michael L. Burton, Malcolm M. Dow. 1981. "Sexual Division of Labor in African Agriculture: A Network Autocorrelation Analysis." *American Anthropologist.* 83:824-849.

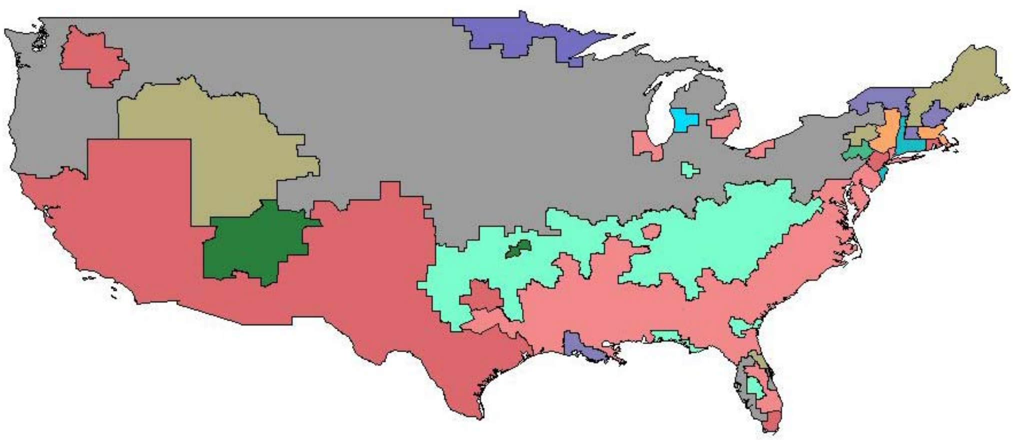
Wooldridge, Jeffrey W. 2002. *Econometric Analysis of Cross Section and Panel Data.* Cambridge: The MIT Press.

Zelinsky, Wilbur. 1992. *The Cultural Geography of the United States.* Second Edition. Englewood Cliffs, NJ: Prentice-Hall.

Zelinsky, Wilbur, and Barrett A. Lee. 1998. "Heterolocalism: An Alternative Model of the Sociospatial Behavior of Immigrant Ethnic Communities." *International Journal of Population Geography* 4:281-298.

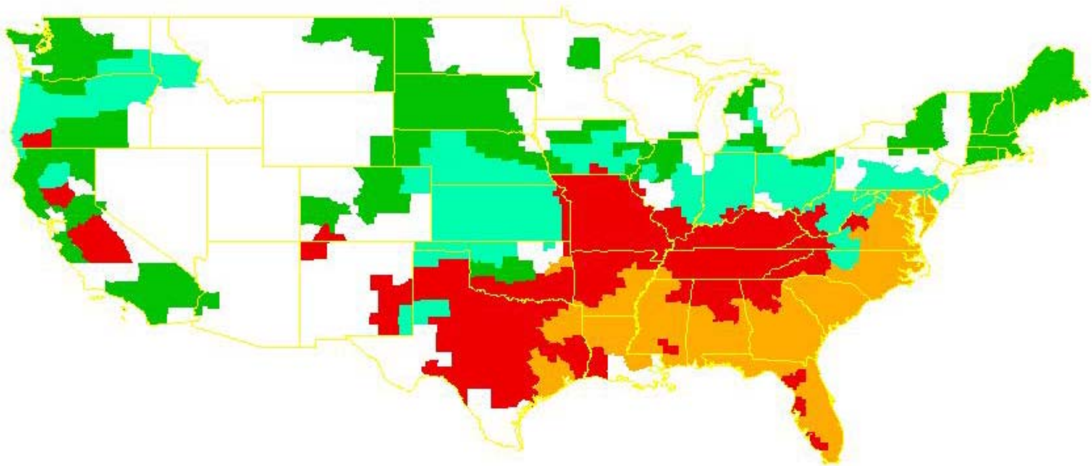
FIGURE 1: PRINCIPAL ANCESTRY REGIONS, 1990-2000

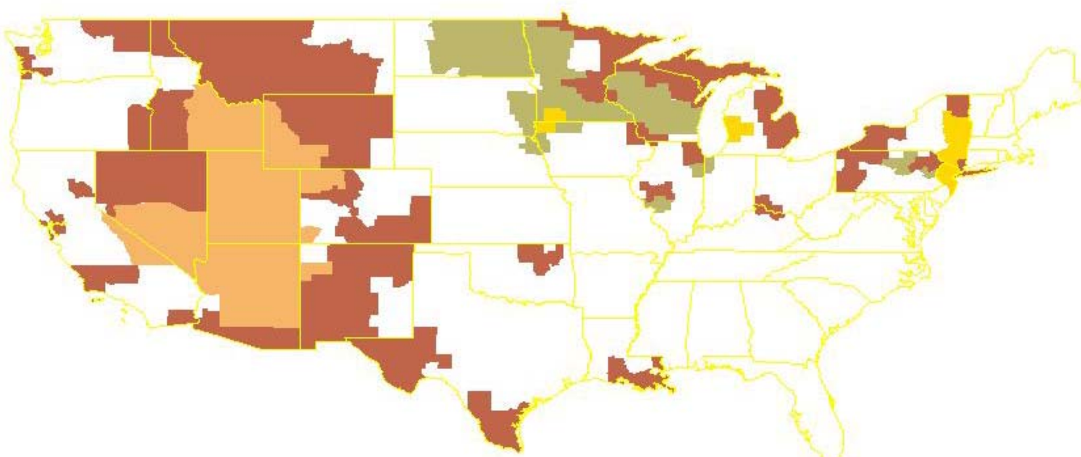




Notes: These two maps show the principal ancestry for each LMA. The top map differs from the bottom only in that it shows state lines, to better understand the location of the regions. Ancestry-based regions tend to be contiguous. The largest region is that of German ancestry, which stretches across the northern tier of states. Borderer ancestry stretches across the upper South, from West Virginia to north Texas. African ancestry covers the lower South, extending up the Mississippi river and the East Coast, and includes large northern cities such as Chicago, Detroit, Cleveland, and New York. The Latino ancestry region stretches from the Texas Gulf to California, and also includes much of eastern Washington, south Florida, northern New Jersey, and Rhode Island. American Indian covers the Four Corners area, a section of eastern Oklahoma, and also includes Alaska (not shown). British covers northern New England, a piece of upstate New York, and the area of Mormon settlement. Southern Louisiana and portions of New England and northern New York are French. Smaller areas are assigned to Irish (eastern Massachusetts and the Hudson Valley), Italian (Connecticut and portions of Massachusetts, Vermont, and New Jersey), Dutch (western Michigan), Scandinavian (northern Minnesota and North Dakota), and Eastern European Catholic (eastern Pennsylvania). This latter contains ancestries stretching from Lithuania, through Poland, down to Croatia. Hawaii (not shown) is Asian.

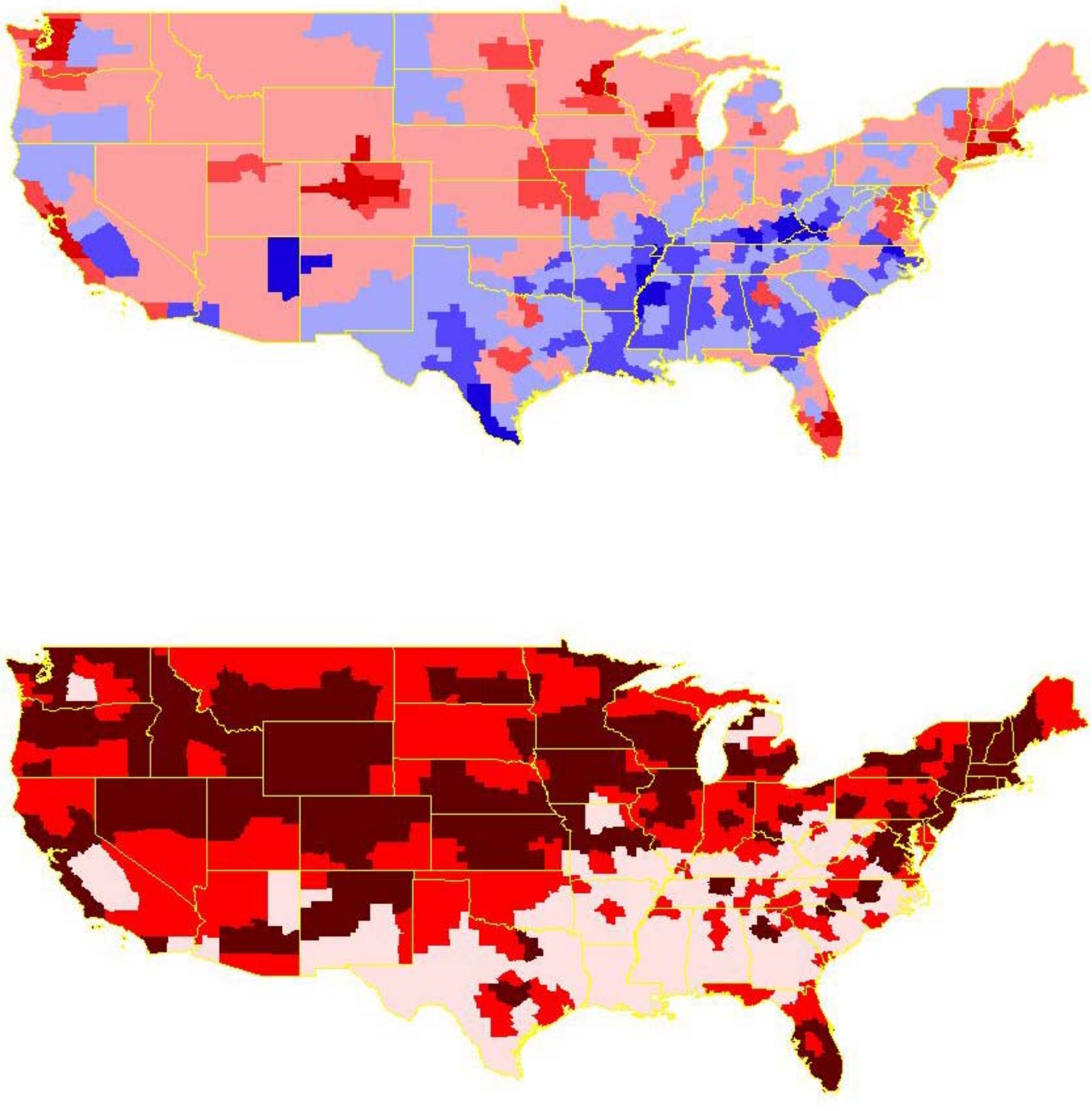
FIGURE 2: FISCHER-RELIGION REGIONS, CENSUSES OF 1890, 1926, 1936





Notes: These two maps show the principal religious affiliation for each LMA. The assignments are based on the 1890 Census, if available, and on the 1926 and 1936 Censuses otherwise. The top map shows the religious affiliations corresponding to the four colonial cultures of David Hackett Fischer. The red region across the upper South and into Texas is the Borderers. The Puritans, in dark green, cover New England, skip across the northernmost tier of states, and are dominant along the West Coast. The Pennsylvania culture covers the light green area stretching from Pennsylvania through Ohio and Indiana, and then to Kansas, with a presence on the West Coast and in far western Virginia and North Carolina. The Chesapeake culture (predominantly African-American denominations) occupies the orange region in the lowland South. The second map shows regions that could not reasonably be assigned to any of the four Fischer groups. The Dutch occupy the yellow area—primarily the Hudson valley and western Michigan. The dark brown area is Catholic, the greenish area concentrated in the Dakotas, Minnesota, and Wisconsin is Lutheran, and the light brown area is Mormon.

FIGURE 3: REGIONAL STANDARD OF LIVING: LIFE, KNOWLEDGE, PROSPERITY.



Notes: These two maps present LMA variation in standard of living, as measured by a regional analogue of the UN Human Development Indicator. The standard of living simply equals the geometric mean of standardized scores for per capita income, average life expectancy, and average educational attainment. The top map divides LMAs into standard deviation categories. The darker the blue, the lower the standard of living, the darker the red, the higher the standard of living. The second graph simply separates the 394 LMAs into terciles: the darker the color, the higher the standard of living. One can readily see that the traditional four poorest areas of the U.S. rank lowest: the Four Corners, Appalachian Kentucky, the Mississippi Delta, and South Texas. The highest standards of living are found in the San Francisco Bay Area and south to Monterey County, Puget Sound, Minneapolis, Madison, the eastern slope of the Colorado Rockies, Miami, Connecticut, and eastern Massachusetts.

TABLE 1: SUMMARY OF PROXIMITY MATRICES

Category	Name	Equation	Data Year	Description
Distance				
	contig	Eq. (1)	2004	Contiguous LMAs
	dist	Eq. (2)-(3)	2004	Inverted Squared Great Circle Distance
Culture				
Ancestry				
	aned2000	Eq. (4)	2000	Inverse Euclidean distance ancestry 2000
	anchi90	Eq. (5)	1990	Herfindahl-type ancestry 1990
	anhi2000	Eq. (5)	2000	Herfindahl-type ancestry 2000
	elem	Eq. (6)	1990	Similarity ancestry marriage matrix 1990 (Eq. A1)
	pumsanc	Eq. (6)	2000	Similarity ancestry marriage matrix 2000
	sameancest	Eq. (7)	1990, 2000	Binary Similarity Principal Ancestry 1990 2000 (Figure 1)
Religion				
	ed90	Eq. (8)	1990	Inverse Euclidean distance Religion 1990
	xed	Eq. (8)	1990	Inverse Euclidean distance Religion 1990, 4 categories
	hi90	Eq. (9)	1990	Herfindahl-type religion 1990
	rs1890	Eq. (10)	1890	Similarity religion, phenetic classification 1890
	rs1926	Eq. (10)	1926	Similarity religion, phenetic classification 1926
	rs1936	Eq. (10)	1936	Similarity religion, phenetic classification 1936
	rs1990	Eq. (10)	1990	Similarity religion, phenetic classification 1990
	xhi	Eq. (11)	1990	Herfindahl-type religion 1990, 4 categories
	xbn	Eq. (11)	1990	Religion 1990, 4 categories, binary same 2 main religions
	xn	Eq. (11)	1990	Religion 1990, 4 categories, binary same main religion
	samec8	Eq. (11)	1890-1936	Religion 1890, binary same main religion (Figure 2)
Place Names				
	pnsiml	Eq. (12)	2000	Place name similarity, population weighted
Elections				
	edpres	Eq. (13)	1912-1972	Inverse Euclidean Distance, Presidential Elections
	hipres	Eq. (14)	1912-1972	Herfindahl-type, Presidential Elections
Ecology				
	simleco	Eq. (15)	2004	Similarity ecoregions
Labor Flows				
	n9	Eq. (16)	1990	Commuting flows 1990
	n10	Eq. (16)	2000	Commuting flows 2000
	pop	Eq. (16)	1990	Migration flows 1990 of ages 25-64
Level of Development				
	sliv	Eq. (17)	1990-2000	Inverse Euclidean distance Life expect., PCI, Educ. attainment
Economic Structure				
Industries				
	edcbp	Eq. (18)	2001	Inverse Euclidean distance County Business Patterns 2001
	hicbp	Eq. (19)	2001	Herfindahl-type County Business Patterns 2001
Occupations				
	edocc	Eq. (18)	1990	Inverse Euclidean distance EEOF occupations 1990
	hiocc	Eq. (19)	1990	Herfindahl-type EEOF occupations 1990
Regional Structure				
	edcbc	Eq. (20)	1993	Inverted Absolute Difference Avg. Calvin Beale Codes 1993
	evnal	Eq. (23)	2000	Eigenvector weights, all links
	evnup	Eq. (25)	2000	Eigenvector weights, upward links
	evn10	Eq. (26)	2000	Eigenvector weights, single strongest upward link

Notes: S

TABLE 2A: MATRIX CORRELATIONS AMONG WEIGHT MATRICES: T-STATISTICS

column	Matrix	description	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35
1	contig	Contiguous LMAs	.	115	34	25	28	16	19	23	30	23	23	13	17	19	12	16	14	13	23	0	30	20	46	39	87	51	17	18	7	19	22	4	125	130	90
2	dist	Inverted Squared Great Circle Distance	115	.	16	18	25	9	12	20	20	17	16	6	9	9	7	10	12	16	28	3	8	10	28	19	36	16	12	10	8	8	10	4	84	94	69
3	aned200	Inverse Euclidean distance ancestry 2000	34	16	.	12	16	12	12	22	17	12	10	3	6	6	-3	6	8	13	19	3	10	7	8	7	10	8	15	7	0	8	0	3	17	21	15
4	anchi90	Herfindahl-type ancestry 1990	25	18	12	.	24	12	15	17	13	11	6	5	8	9	4	11	7	10	21	4	0	8	11	2	2	0	9	6	7	2	5	4	7	11	7
5	anhi200	Herfindahl-type ancestry 2000	28	25	16	24	.	11	16	25	19	13	13	6	10	10	7	11	10	13	28	5	2	11	17	5	7	2	12	7	8	4	9	4	12	16	11
6	elem	Similarity ancestry marriage matrix 1990 (Eq. A1)	16	9	12	12	11	.	14	13	5	5	0	3	5	5	0	4	2	4	6	4	4	3	6	0	0	-2	8	4	3	0	3	2	2	5	3
7	pumsanc	Similarity ancestry marriage matrix 2000	19	12	12	15	16	14	.	13	6	7	0	3	6	6	0	7	4	6	12	3	4	5	8	0	0	-3	9	4	5	0	3	3	3	6	3
8	sameanc	Binary Similarity Principal Ancestry 1990 2000 (Figure 1)	23	20	22	17	25	13	13	.	23	13	14	5	7	8	2	4	8	12	24	3	10	7	10	6	8	4	13	6	2	3	8	0	14	16	11
9	ed90	Inverse Euclidean distance Religion 1990	30	20	17	13	19	5	6	23	.	28	34	6	12	11	0	15	22	33	36	2	10	8	8	9	12	9	14	8	0	5	5	3	20	22	16
10	xed	Inverse Euclidean distance Religion 1990, 4 categories	23	17	12	11	13	5	7	13	28	.	16	5	8	8	2	22	34	48	18	2	5	6	6	6	8	5	9	8	4	5	5	4	14	17	12
11	hi90	Herfindahl-type religion 1990	23	16	10	6	13	0	0	14	34	16	.	5	9	9	9	6	16	23	31	0	6	8	6	6	12	9	9	4	0	4	4	2	16	19	13
12	rs1890	Similarity religion, phenetic classification 1890	13	6	3	5	6	3	3	5	6	5	5	.	9	9	0	3	3	4	10	2	-3	7	5	0	2	0	3	0	0	0	0	0	4	6	4
13	rs1926	Similarity religion, phenetic classification 1926	17	9	6	8	10	5	6	7	12	8	9	9	.	16	3	6	5	9	17	2	-5	14	6	0	2	0	4	2	2	0	3	0	6	8	6
14	rs1936	Similarity religion, phenetic classification 1936	19	9	6	9	10	5	6	8	11	8	9	9	16	.	4	7	6	8	16	0	-5	16	5	0	0	0	5	2	3	0	4	0	5	9	5
15	rs1990	Similarity religion, phenetic classification 1990	12	7	-3	4	7	0	0	2	0	2	9	0	3	4	.	2	5	6	7	0	-4	4	2	0	0	-2	0	2	3	0	7	0	3	4	3
16	xhi	Herfindahl-type religion 1990, 4 categories	16	10	6	11	11	4	7	4	15	22	6	3	6	7	2	.	26	34	15	0	0	4	4	2	0	0	5	4	6	2	2	5	5	8	5
17	xbn	Religion 1990, 4 categories, binary same 2 main religions	14	12	8	7	10	2	4	8	22	34	16	3	5	6	5	26	.	72	17	0	4	5	4	4	6	3	7	5	4	3	5	5	9	10	7
18	xn	Religion 1990, 4 categories, binary same main religion	13	16	13	10	13	4	6	12	33	48	23	4	9	8	6	34	72	.	21	0	8	7	4	5	7	4	11	8	5	5	7	6	10	10	7
19	samec8	Religion 1890, binary same main religion (Figure 2)	23	28	19	21	28	6	12	24	36	18	31	10	17	16	7	15	17	21	.	2	8	16	18	10	15	8	14	7	6	5	6	3	17	18	12
20	pnsiml	Place name similarity, population weighted	0	3	3	4	5	4	3	3	2	2	0	2	2	0	0	0	0	0	2	.	0	0	5	2	3	2	0	2	0	0	0	0	0	0	0
21	edpres	Inverse Euclidean Distance, Presidential Elections	30	8	10	0	2	4	4	10	10	5	6	-3	-5	-5	-4	0	4	8	8	0	.	0	0	4	6	5	7	3	-3	0	0	0	14	20	13
22	hipres	Herfindahl-type, Presidential Elections	20	10	7	8	11	3	5	7	8	6	8	7	14	16	4	4	5	7	16	0	0	.	7	3	5	2	5	2	2	2	3	0	8	11	8
23	simleco	Similarity ecoregions	46	28	8	11	17	6	8	10	8	6	6	5	6	5	2	4	4	4	18	5	0	7	.	7	12	5	5	6	6	5	2	0	24	29	21
24	n9	Commuting flows 1990	39	19	7	2	5	0	0	6	9	6	6	0	0	0	0	2	4	5	10	2	4	3	7	.	27	14	2	2	-3	5	0	0	29	30	21
25	n10	Commuting flows 2000	87	36	10	2	7	0	0	8	12	8	12	2	2	0	0	0	6	7	15	3	6	5	12	27	.	22	3	3	-5	8	2	0	66	72	53
26	pop	Migration flows 1990 of ages 25-64	51	16	8	0	2	-2	-3	4	9	5	9	0	0	0	-2	0	3	4	8	2	5	2	5	14	22	.	2	0	-6	7	0	-3	31	40	29
27	sliv	Inverse Euclidean distance Life expect., PCI, Educ. attainment	17	12	15	9	12	8	9	13	14	9	9	3	4	5	0	5	7	11	14	0	7	5	5	2	3	2	.	12	4	9	8	11	5	6	4
28	edcbp	Inverse Euclidean distance County Business Patterns 2001	18	10	7	6	7	4	4	6	8	8	4	0	2	2	2	4	5	8	7	2	3	2	6	2	3	0	12	.	8	7	6	12	6	7	5
29	hicbp	Herfindahl-type County Business Patterns 2001	7	8	0	7	8	3	5	2	0	4	0	0	2	3	3	6	4	5	6	0	-3	2	6	-3	-5	-6	4	8	.	0	5	6	-3	-3	-3
30	edocc	Inverse Euclidean distance EEOF occupations 1990	19	8	8	2	4	0	0	3	5	5	4	0	0	0	0	2	3	5	5	0	0	2	5	5	8	7	9	7	0	.	3	10	9	10	8
31	hiocc	Herfindahl-type EEOF occupations 1990	22	10	0	5	9	3	3	8	5	5	4	0	3	4	7	2	5	7	6	0	0	3	2	0	2	0	8	6	5	3	.	8	6	8	5
32	edcbc	Inverted Absolute Difference Avg. Calvin Beale Codes 1993	4	4	3	4	4	2	3	0	3	4	2	0	0	0	0	5	5	6	3	0	0	0	0	0	0	-3	11	12	6	10	8	.	0	0	0
33	evnup	Eigenvector weights, upward links	125	84	17	7	12	2	3	14	20	14	16	4	6	5	3	5	9	10	17	0	14	8	24	29	66	31	5	6	-3	9	6	0	.	176	149
34	evnal	Eigenvector weights, all links	130	94	21	11	16	5	6	16	22	17	19	6	8	9	4	8	10	10	18	0	20	11	29	30	72	40	6	7	-3	10	8	0	176	.	135
35	evn10	Eigenvector weights, single strongest upward link	90	69	15	7	11	3	3	11	16	12	13	4	6	5	3	5	7	7	12	0	13	8	21	21	53	29	4	5	-3	8	5	0	149	135	.

Notes: Matrix correlation formula given in Wasserman and Faust (1992: 686). Standard errors determined through permutation test (1,000 simulations).

TABLE 2B: NUMBER OF SIGNIFICANT CORRELATIONS, BY MATRIX

Category	Matrix	Description	Negative	Insignificant	Positive	P(t-stat>z)
Distance	contig	Contiguous LMAs	0	1	33	88%
Distance	dist	Inverted Squared Great Circle Distance	0	0	34	80%
Ancestry	aned2000	Inverse Euclidean distance ancestry 2000	1	2	31	62%
Ancestry	anchi90	Herfindahl-type ancestry 1990	0	2	32	59%
Ancestry	anhi2000	Herfindahl-type ancestry 2000	0	0	34	72%
Ancestry	elem	Similarity ancestry marriage matrix 1990 (Eq. A1)	1	5	28	33%
Ancestry	pumsanc	Similarity ancestry marriage matrix 2000	1	5	28	41%
Ancestry	sameancest	Binary Similarity Principal Ancestry 1990 2000 (Figure 1)	0	1	33	65%
Religion	ed90	Inverse Euclidean distance Religion 1990	0	2	32	72%
Religion	xed	Inverse Euclidean distance Religion 1990, 4 categories	0	0	34	64%
Religion	hi90	Herfindahl-type religion 1990	0	4	30	58%
Religion	rs1890	Similarity religion, phenetic classification 1890	1	8	25	26%
Religion	rs1926	Similarity religion, phenetic classification 1926	1	4	29	42%
Religion	rs1936	Similarity religion, phenetic classification 1936	1	6	27	43%
Religion	rs1990	Similarity religion, phenetic classification 1990	3	10	21	20%
Religion	xhi	Herfindahl-type religion 1990, 4 categories	0	4	30	42%
Religion	xbn	Religion 1990, 4 categories, binary same 2 main religions	0	1	33	49%
Religion	xn	Religion 1990, 4 categories, binary same main religion	0	1	33	62%
Religion	samec8	Religion 1890, binary same main religion (Figure 2)	0	0	34	78%
Place Names	pnsiml	Place name similarity, population weighted	0	17	17	16%
Elections	edpres	Inverse Euclidean Distance, Presidential Elections	5	8	21	31%
Elections	hipres	Herfindahl-type, Presidential Elections	0	3	31	43%
Ecology	simleco	Similarity ecoregions	0	2	32	53%
Labor Flows	n9	Commuting flows 1990	1	8	25	36%
Labor Flows	n10	Commuting flows 2000	1	6	27	45%
Labor Flows	pop	Migration flows 1990 of ages 25-64	5	7	22	30%
Level of Development	sliv	Inverse Euclidean distance Life expect., PCI, Educ. attainment	0	2	32	50%
Industries	edcbp	Inverse Euclidean distance County Business Patterns 2001	0	2	32	40%
Industries	hicbp	Herfindahl-type County Business Patterns 2001	7	6	21	23%
Occupations	edocc	Inverse Euclidean distance EEOF occupations 1990	0	9	25	31%
Occupations	hiocc	Herfindahl-type EEOF occupations 1990	0	6	28	32%
Regional Structure	edcbc	Inverted Absolute Difference Avg. Calvin Beale Codes 1993	1	14	19	24%
Regional Structure	evnup	Eigenvector weights, upward links	1	2	31	62%
Regional Structure	evnal	Eigenvector weights, all links	1	2	31	71%
Regional Structure	evn10	Eigenvector weights, single strongest upward link	1	2	31	57%

Notes: Summary of performance shown in Table 2a. Insignificant correlations are those with p-value greater than 0.05 (one-sided test). P(t-stat>z) gives the probability that the matrix will have a higher t-statistic than other matrices in a random row in Table 2a.

TABLE 3: 205 SAMPLE VARIABLES USED IN AUTOCORRELATIONS

Variable	Label	N	Maximum	Minimum	Mean	Std Dev
Income 1991-2001						
pfarmy2	farm income as pct personal income 1991-2001	394	0.2255	-0.0029	0.0185	0.0256
pnetyy2	net earnings as pct personal income 1991-2001	394	0.7691	0.4351	0.6404	0.0501
ptrany2	transfer income as pct personal income 1991-2001	394	0.3554	0.0753	0.1707	0.0432
pincmy2	income maintenance income as pct personal income 1991-2001	394	0.0731	0.0048	0.0177	0.0096
puecmy2	unemployment benefits as pct personal income 1991-2001	394	0.0178	0.0008	0.0045	0.0024
pretiy2	retirement income as pct personal income 1991-2001	394	0.2773	0.0670	0.1485	0.0357
pdiry2	dividends interest and rent income as pct personal income 1991-2001	394	0.4189	0.1137	0.1889	0.0346
pwagey2	wage and salary income as pct personal income 1991-2001	394	0.8263	0.5621	0.7607	0.0397
ppropy2	proprietor income as pct personal income 1991-2001	394	0.3648	0.0447	0.1290	0.0455
avgempy2	average POW earnings per job 1991-2001	394	45,178	17,024	24,560	3,975
avgwage2	average POW earnings per wage and salary job 1991-2001	394	40,613	17,191	23,070	3,537
pci2	per capita income 1991-2001	394	32,156	11,459	19,919	3,314
pfarmprop2	farm proprietors as pct of all proprietors 1991-2001	394	0.467	0.001	0.169	0.107
pprop2	proprietors as pct of POW jobs 1991-2001	394	0.3384	0.1022	0.1905	0.0471
nettoty2	Ratio of net earnings to total POW earnings 1991-2001	394	1.4080	0.8433	0.9662	0.0651
avgpropy2	average POW earnings per proprietor 1991-2001	394	49,059	8,472	16,669	5,134
Income growth 1969-2001						
dpfarmy	farm income as pct personal income	394	0.234	-1.274	-0.666	0.214
dpnetyy	net earnings as pct personal income	394	-0.0318	-0.2704	-0.1422	0.0417
dptrany	transfer income as pct personal income	394	1.340	-0.024	0.431	0.206
dpincmy	income maintenance income as pct personal income	394	2.664	-0.476	0.394	0.436
dpuecmy	unemployment benefits as pct personal income	394	5.167	-0.738	-0.206	0.478
dpretiy	retirement income as pct personal income	394	1.734	0.028	0.494	0.213
dpdiry	dividends interest and rent income as pct personal income	394	1.254	-0.008	0.439	0.176
dpwagey	wage and salary income as pct personal income	394	0.4319	-0.1052	0.0234	0.0775
dppropy	proprietor income as pct personal income	394	0.631	-0.666	-0.222	0.211
davgempy	average POW earnings per job	394	3.114	0.963	1.830	0.300
davgwage	average POW earnings per wage and salary job	394	2.891	1.171	1.903	0.245
dpci	per capita income	394	3.694	1.895	2.866	0.342
dpfarmprop	farm proprietors as pct of all proprietors	394	0.140	-0.790	-0.415	0.133
dpprop	proprietors as pct of POW jobs	394	0.820	-0.338	0.064	0.183
dnettoty	Ratio of net earnings to total POW earnings	394	0.1178	-0.1661	-0.0147	0.0297
davgpropy	average POW earnings per proprietor	394	3.323	-0.128	1.103	0.524
Income Distribution by Age						
nfdm15w25	Lieberson Net Difference, income higher for 25to34 than for und25 2000	394	0.5796	0.2501	0.3921	0.0565
nfdm25w35	Lieberson Net Difference, income higher for 35to44 than for 25to34 2000	394	0.3201	0.0686	0.1556	0.0355
nfdm35w45	Lieberson Net Difference, income higher for 45to54 than for 35to44 2000	394	0.1542	-0.0016	0.0849	0.0269
nfdm45w55	Lieberson Net Difference, income higher for 55to64 than for 45to54 2000	394	0.0148	-0.2128	-0.1216	0.0340
nfdm55w65	Lieberson Net Difference, income higher for 65to74 than for 55to64 2000	394	-0.0556	-0.3817	-0.2214	0.0524
nfdm65w75	Lieberson Net Difference, income higher for 75up than for 65to74 2000	394	-0.0415	-0.3343	-0.1992	0.0384
Education						
fea	female educational attainment 2000	394	9.876	6.189	8.603	0.549
mea	male educational attainment 2000	394	10.088	6.554	8.614	0.615
tea	total educational attainment 2000	394	9.873	6.358	8.608	0.576
xppp	operating expenditures per pupil	392	11.08	4.17	6.45	1.02

Variable	Label	N	Maximum	Minimum	Mean	Std Dev
nfdmedat	Lieberson Net Difference Female>Male on Educational Attainment 2000	394	0.1352	-0.1141	-0.0027	0.0325
Environment						
range	annual precipitation inches from shapefile	391	97.79	5.62	39.39	13.50
urb0	percent area rural from shapefile 2000	394	1.0000	0.6510	0.9834	0.0300
urb1	percent area urban from shapefile 2000	394	0.3490	0	0.0166	0.0300
Marital Status						
msf1	Pct Females over14, never married 2000	394	0.3615	0.1243	0.2154	0.0410
msf2	Pct Females over14, now married 2000	394	0.6539	0.4220	0.5643	0.0343
msf3	Pct Females over14, now married, husb present 2000	394	0.6045	0.3453	0.5188	0.0402
msf4	Pct Females over14, now married, husb absent 2000	394	0.1007	0.0211	0.0456	0.0123
msf5	Pct Females over14, now married, separated 2000	394	0.0494	0.0065	0.0211	0.0087
msf6	Pct Females over14, now married, other 2000	394	0.0675	0.0118	0.0245	0.0062
msf7	Pct Females over14, widowed 2000	394	0.1620	0.0491	0.1142	0.0200
msf8	Pct Females over14, divorced 2000	394	0.1457	0.0498	0.1061	0.0157
Life Expectancy						
tle	total life expectancy average 1965-1994	394	79.07	70.73	75.32	1.48
mle	male life expectancy average 1965-1994	393	75.73	66.10	71.68	1.81
fle	female life expectancy average 1965-1994	393	82.35	74.93	78.80	1.27
Sex Ratio						
sr0	Sex Ratio, Male/Female, und1 2000	394	1.5273	0.7919	1.0613	0.0979
sr1	Sex Ratio, Male/Female, 1 2000	394	1.3613	0.7616	1.0515	0.0885
sr2	Sex Ratio, Male/Female, 2 2000	394	1.4389	0.7178	1.0547	0.0969
sr3	Sex Ratio, Male/Female, 3 2000	394	1.3806	0.7539	1.0431	0.0919
sr4	Sex Ratio, Male/Female, 4 2000	394	1.3814	0.7491	1.0518	0.0890
sr5	Sex Ratio, Male/Female, 5 2000	394	1.4804	0.7929	1.0607	0.0854
sr6	Sex Ratio, Male/Female, 6 2000	394	1.4575	0.8271	1.0552	0.0870
sr7	Sex Ratio, Male/Female, 7 2000	394	1.3457	0.6875	1.0495	0.0826
sr8	Sex Ratio, Male/Female, 8 2000	394	1.4076	0.7765	1.0497	0.0862
sr9	Sex Ratio, Male/Female, 9 2000	394	1.3546	0.8428	1.0597	0.0843
sr10	Sex Ratio, Male/Female, 10 2000	394	1.4087	0.8001	1.0636	0.0842
sr11	Sex Ratio, Male/Female, 11 2000	394	1.4098	0.8088	1.0545	0.0858
sr12	Sex Ratio, Male/Female, 12 2000	394	1.5504	0.8565	1.0608	0.0902
sr13	Sex Ratio, Male/Female, 13 2000	394	1.3257	0.8018	1.0538	0.0780
sr14	Sex Ratio, Male/Female, 14 2000	394	1.3394	0.8395	1.0630	0.0796
sr15	Sex Ratio, Male/Female, 15 2000	394	1.4589	0.8387	1.0629	0.0775
sr16	Sex Ratio, Male/Female, 16 2000	394	1.3542	0.7280	1.0704	0.0888
sr17	Sex Ratio, Male/Female, 17 2000	394	1.3629	0.8067	1.0770	0.0870
sr18	Sex Ratio, Male/Female, 18 2000	394	1.467	0.717	1.065	0.121
sr19	Sex Ratio, Male/Female, 19 2000	394	2.159	0.607	1.069	0.165
sr20	Sex Ratio, Male/Female, 20 2000	394	2.173	0.383	1.075	0.173
sr21	Sex Ratio, Male/Female, 21 2000	394	2.130	0.704	1.069	0.159
sr22	Sex Ratio, Male/Female, 22to24 2000	394	1.595	0.836	1.073	0.112
sr23	Sex Ratio, Male/Female, 25to29 2000	394	1.4788	0.8054	1.0438	0.0836
sr24	Sex Ratio, Male/Female, 30to34 2000	394	1.6892	0.8695	1.0245	0.0782
sr25	Sex Ratio, Male/Female, 35to39 2000	394	1.4735	0.8563	0.9959	0.0615
sr26	Sex Ratio, Male/Female, 40to44 2000	394	1.3087	0.8715	0.9998	0.0548
sr27	Sex Ratio, Male/Female, 45to49 2000	394	1.2742	0.8631	0.9899	0.0499
sr28	Sex Ratio, Male/Female, 50to54 2000	394	1.2341	0.8428	0.9806	0.0497
sr29	Sex Ratio, Male/Female, 55to59 2000	394	1.1535	0.7929	0.9485	0.0500
sr30	Sex Ratio, Male/Female, 60to61 2000	394	1.3225	0.6654	0.9330	0.0794
sr31	Sex Ratio, Male/Female, 62to64 2000	394	1.1521	0.7318	0.9121	0.0596
sr32	Sex Ratio, Male/Female, 65to66 2000	394	1.2052	0.6701	0.8991	0.0793
sr33	Sex Ratio, Male/Female, 67to69 2000	394	1.1577	0.6336	0.8663	0.0679
sr34	Sex Ratio, Male/Female, 70to74 2000	394	1.0960	0.5837	0.8088	0.0662
sr35	Sex Ratio, Male/Female, 75to79 2000	394	0.9696	0.5110	0.6984	0.0695
sr36	Sex Ratio, Male/Female, 80to84 2000	394	0.9379	0.4054	0.5861	0.0790
sr37	Sex Ratio, Male/Female, 85up 2000	394	0.7675	0.2050	0.4084	0.0699
srupto15	Sex Ratio, Male/Female, 0to15 2000	394	1.0943	0.9825	1.0527	0.0156
sr18to29	Sex Ratio, Male/Female, 18to29 2000	394	1.4594	0.8949	1.0557	0.0854
sralt	Sex Ratio, Male(22to44)/Female(20to39) 2000	394	1.8076	1.0102	1.2089	0.0878
nfdmage	Lieberson Net Difference Female>Male on Age 2000	394	0.0925	0.0049	0.0598	0.0127
Male/Female Occupation Difference						
dif	Male/Female Occupation Difference Index 2000	394	1.1293	0.5194	0.8100	0.0985
Percent Population by Household Type						
ppht1	Pct population, in households 2000	394	0.9885	0.8802	0.9657	0.0164

Variable	Label	N	Maximum	Minimum	Mean	Std Dev
ppht2	Pct population, in family households 2000	394	0.9373	0.6987	0.8228	0.0321
ppht3	Pct population, in family households, householders 2000	394	0.2962	0.2174	0.2644	0.0131
ppht4	Pct population, in family households, male householder 2000	394	0.2467	0.1323	0.2042	0.0177
ppht5	Pct population, in family households, female householder 2000	394	0.1078	0.0290	0.0602	0.0116
ppht6	Pct population, in family households, spouse 2000	394	0.2495	0.1351	0.2087	0.0170
ppht7	Pct population, in family households, child 2000	394	0.4029	0.2039	0.2903	0.0250
ppht8	Pct population, in family households, child, natural-born 2000	394	0.3812	0.1861	0.2648	0.0246
ppht9	Pct population, in family households, child, adopted 2000	394	0.0134	0.0050	0.0074	0.0010
ppht10	Pct population, in family households, child, step 2000	394	0.0307	0.0078	0.0181	0.0039
ppht11	Pct population, in family households, grandchild 2000	394	0.0678	0.0047	0.0176	0.0087
ppht12	Pct population, in family households, sibling 2000	394	0.0193	0.0030	0.0070	0.0028
ppht13	Pct population, in family households, parent 2000	394	0.0145	0.0014	0.0054	0.0020
ppht14	Pct population, in family households, other relatives 2000	394	0.0405	0.0028	0.0119	0.0058
ppht15	Pct population, in family households, nonrelatives 2000	394	0.0360	0.0097	0.0175	0.0041
ppht16	Pct population, in nonfamily households 2000	394	0.2399	0.0481	0.1429	0.0246
ppht17	Pct population, in nonfamily households, male householder 2000	394	0.0903	0.0185	0.0530	0.0084
ppht18	Pct population, in nonfamily households, male householder, living alone 2000	394	0.0570	0.0119	0.0414	0.0056
ppht19	Pct population, in nonfamily households, male householder, NOT living alone 2000	394	0.0365	0.0027	0.0116	0.0041
ppht20	Pct population, in nonfamily households, female householder 2000	394	0.0921	0.0228	0.0644	0.0096
ppht21	Pct population, in nonfamily households, female householder, living alone 2000	394	0.0815	0.0192	0.0567	0.0087
ppht22	Pct population, in nonfamily households, female householder, NOT living alone 2000	394	0.0241	0.0014	0.0077	0.0034
ppht23	Pct population, in nonfamily households, nonrelatives 2000	394	0.0831	0.0051	0.0255	0.0112
ppht24	Pct population, in group quarters 2000	394	0.1198	0.0115	0.0343	0.0164
ppht25	Pct population, in group quarters, institutionalized 2000	394	0.1123	0.0057	0.0190	0.0120
ppht26	Pct population, in group quarters, NONinstitutionalized 2000	394	0.0624	0.0015	0.0152	0.0107
Percent Households by Household Type						
ht1	Pct Households, 1 person 2000	394	0.3008	0.1137	0.2558	0.0243
ht2	Pct Households, 1 male person 2000	394	0.1387	0.0436	0.1082	0.0111
ht3	Pct Households, 1 female person 2000	394	0.1846	0.0702	0.1476	0.0168
ht4	Pct Households, 2 or more persons 2000	394	0.8863	0.6992	0.7442	0.0243
ht5	Pct Households, Family 2000	394	0.8428	0.5997	0.6936	0.0332
ht6	Pct Households, Family, married couple 2000	394	0.7107	0.4096	0.5453	0.0384
ht7	Pct Households, Family, married couple, children und18 2000	394	0.4345	0.1435	0.2411	0.0316
ht8	Pct Households, Family, married couple, NO children und18 2000	394	0.4087	0.2235	0.3042	0.0303
ht9	Pct Households, Family, NOT married couple 2000	394	0.3173	0.0734	0.1483	0.0325
ht10	Pct Households, Family, male householder, no wife 2000	394	0.0595	0.0244	0.0385	0.0053
ht11	Pct Households, Family, male householder, no wife, children und18 2000	394	0.0357	0.0122	0.0213	0.0037
ht12	Pct Households, Family, male householder, no wife, NO children und18 2000	394	0.0290	0.0100	0.0172	0.0034
ht13	Pct Households, Family, female householder, no husb 2000	394	0.2666	0.0491	0.1098	0.0295
ht14	Pct Households, Family, female householder, no husb, children und18 2000	394	0.1553	0.0320	0.0667	0.0165
ht15	Pct Households, Family, female householder, no husb, NO children und18 2000	394	0.1114	0.0171	0.0431	0.0143
ht16	Pct Households, Non-Family 2000	394	0.1454	0.0149	0.0506	0.0192

Variable	Label	N	Maximum	Minimum	Mean	Std Dev
ht17	Pct Households, Non-Family, male householder 2000	394	0.0933	0.0099	0.0303	0.0106
ht18	Pct Households, Non-Family, female householder 2000	394	0.0666	0.0050	0.0203	0.0089
Percent Females: Child Status and Labor Force Status						
fesch1	Pct Females 16up w own chidren und18 2000	394	0.4203	0.1970	0.3081	0.0293
fesch2	Pct Females 16up w own chidren und6 2000	394	0.1133	0.0447	0.0724	0.0097
fesch3	Pct Females 16up w own chidren und6, in LF 2000	394	0.0677	0.0281	0.0475	0.0061
fesch4	Pct Females 16up w own chidren und6, in LF,employed 2000	394	0.0611	0.0264	0.0438	0.0061
fesch5	Pct Females 16up w own chidren und6, in LF,unemployed 2000	394	0.0094	0.0010	0.0037	0.0014
fesch6	Pct Females 16up w own chidren und6, NOT in LF 2000	394	0.0592	0.0118	0.0250	0.0069
fesch7	Pct Females 16up w own chidren und6 and 6to17 2000	394	0.1191	0.0389	0.0610	0.0109
fesch8	Pct Females 16up w own chidren und6 and 6to17, in LF 2000	394	0.0592	0.0218	0.0385	0.0062
fesch9	Pct Females 16up w own chidren und6 and 6to17, in LF,employed 2000	394	0.0535	0.0191	0.0360	0.0059
fesch10	Pct Females 16up w own chidren und6 and 6to17, in LF,unemployed 2000	394	0.0103	0.0007	0.0025	0.0014
fesch11	Pct Females 16up w own chidren und6 and 6to17, NOT in LF 2000	394	0.0668	0.0110	0.0225	0.0071
fesch12	Pct Females 16up w own chidren 6to17 2000	394	0.2190	0.1130	0.1747	0.0135
fesch13	Pct Females 16up w own chidren 6to17, in LF 2000	394	0.1673	0.0889	0.1338	0.0127
fesch14	Pct Females 16up w own chidren 6to17, in LF,employed 2000	394	0.1649	0.0866	0.1283	0.0132
fesch15	Pct Females 16up w own chidren 6to17, in LF,unemployed 2000	394	0.0158	0.0018	0.0055	0.0021
fesch16	Pct Females 16up w NO own chidren und18, NOT in LF 2000	394	0.0985	0.0194	0.0409	0.0115
fesch17	Pct Females 16up w NO own chidren und18 2000	394	0.8030	0.5797	0.6919	0.0293
fesch18	Pct Females 16up w NO own chidren und18, in LF 2000	394	0.4642	0.2154	0.3434	0.0397
fesch19	Pct Females 16up w NO own chidren und18, in LF,employed 2000	394	0.4478	0.1952	0.3223	0.0410
fesch20	Pct Females 16up w NO own chidren und18, in LF,unemployed 2000	394	0.0454	0.0093	0.0212	0.0049
fesch21	Pct Females 16up w NO own chidren und18, NOT in LF 2000	394	0.5056	0.2302	0.3485	0.0439
Percent Children in Arrangement						
ches1	Pct own children under6 2000	394	0.3762	0.2634	0.3127	0.0165
ches2	Pct own children under6 lw both parents 2000	394	0.3402	0.1200	0.2245	0.0236
ches3	Pct own children under6 lw both parents, both in LF 2000	394	0.1927	0.0709	0.1266	0.0205
ches4	Pct own children under6 lw both parents, only Far in LF 2000	394	0.1917	0.0309	0.0810	0.0174
ches5	Pct own children under6 lw both parents, only Mor in LF 2000	394	0.0141	0.0028	0.0075	0.0018
ches6	Pct own children under6 lw both parents, neither in LF 2000	394	0.0411	0.0032	0.0095	0.0049
ches7	Pct own children under6 lw one parent 2000	394	0.1850	0.0348	0.0882	0.0208
ches8	Pct own children under6 lw only Far 2000	394	0.0371	0.0091	0.0213	0.0042
ches9	Pct own children under6 lw only Far, Far in LF 2000	394	0.0290	0.0081	0.0176	0.0034
ches10	Pct own children under6 lw only Far, Far NOT in LF 2000	394	0.0156	0.0005	0.0037	0.0018
ches11	Pct own children under6 lw only Mor 2000	394	0.1498	0.0242	0.0669	0.0188
ches12	Pct own children under6 lw only Mor, Mor in LF 2000	394	0.0929	0.0187	0.0477	0.0120
ches13	Pct own children under6 lw only Mor, Mor NOT in LF 2000	394	0.0569	0.0044	0.0192	0.0082
ches14	Pct own children 6to17 2000	394	0.7366	0.6238	0.6873	0.0165
ches15	Pct own children 6to17 lw both parents 2000	394	0.6222	0.3116	0.5012	0.0439
ches16	Pct own children 6to17 lw both parents, both in LF 2000	394	0.5112	0.1945	0.3328	0.0572

Variable	Label	N	Maximum	Minimum	Mean	Std Dev
ches17	Pct own children 6to17 lw both parents, only Far in LF 2000	394	0.2307	0.0651	0.1239	0.0233
ches18	Pct own children 6to17 lw both parents, only Mor in LF 2000	394	0.0651	0.0108	0.0234	0.0063
ches19	Pct own children 6to17 lw both parents, neither in LF 2000	394	0.1075	0.0066	0.0211	0.0117
ches20	Pct own children 6to17 lw one parent 2000	394	0.3835	0.0781	0.1861	0.0397
ches21	Pct own children 6to17 lw only Far 2000	394	0.0654	0.0145	0.0408	0.0069
ches22	Pct own children 6to17 lw only Far, Far in LF 2000	394	0.0539	0.0125	0.0343	0.0058
ches23	Pct own children 6to17 lw only Far, Far NOT in LF 2000	394	0.0184	0.0011	0.0065	0.0027
ches24	Pct own children 6to17 lw only Mor 2000	394	0.3253	0.0637	0.1453	0.0373
ches25	Pct own children 6to17 lw only Mor, Mor in LF 2000	394	0.2266	0.0502	0.1124	0.0240
ches26	Pct own children 6to17 lw only Mor, Mor NOT in LF 2000	394	0.0987	0.0080	0.0330	0.0157
Population Growth						
grni	Population Growth Rate 1990-1998, Natural Increase Component	394	0.0233	-0.0045	0.0051	0.0038
grim	Population Growth Rate 1990-1998, Net International Migration Component	394	0.0151	-0.0001	0.0013	0.0020
grdm	Population Growth Rate 1990-1998, Net Domestic Migration Component	394	0.0427	-0.0170	0.0031	0.0078
popsamehus	Pct population living in same house 2000 and 1995 2000	394	0.7029	0.3614	0.5568	0.0580
Level of Development						
hhi	composite standard of living	394	130.19	64.57	99.64	12.37
Politics						
rating	National Journal Congressmember Conservative rating 2004	394	92.01	5.66	58.7072081	17.4956176
Crime						
pmurder	Uniform Crime Reports, reported murders per 1000 persons 2001	385	0.3714	0	0.0457	0.0365
prape	Uniform Crime Reports, reported rapes per 1000 persons 2001	385	0.903	0	0.324	0.144
probbery	Uniform Crime Reports, reported robberies per 1000 persons 2001	385	5.751	0	0.847	0.765
pagasslt	Uniform Crime Reports, reported aggrevated assaults per 1000 persons 2001	385	8.11	0.41	2.71	1.54
pburglry	Uniform Crime Reports, reported burglaries per 1000 persons 2001	385	20.56	2.31	7.57	3.10
plarceny	Uniform Crime Reports, reported larcenies per 1000 persons 2001	385	53.97	4.87	24.56	8.01
pmvtheft	Uniform Crime Reports, reported motor vehicle thefts per 1000 persons 2001	385	11.49	0.35	2.76	1.86
parson	Uniform Crime Reports, reported arsons per 1000 persons 2001	385	1.258	0	0.248	0.160

Sources: Crime (United States, Department of Justice, Federal Bureau of Investigation 2004); Education (United States, Bureau of the Census 2003; United States, Bureau of the Census 2004a); Environment (Daly and Taylor 2000; United States, Geological Survey 2001); Male/Female Occupation Difference (United States, Bureau of the Census 1992); Percent Children in Arrangement (United States, Bureau of the Census 2004a); Percent Females: Child Status and Labor Force Status (United States, Bureau of the Census 2004a); Percent Households by Household Type (United States, Bureau of the Census 2004a); Level of Development (United States, Bureau of the Census 2004a; Murray, Michaud, McKenna, and Marks 1998; United States, Bureau of Economic Analysis 2003); Percent Population by Household Type (United States, Bureau of the Census 2004a); Life Expectancy (Murray, Michaud, McKenna, and Marks 1998); Marital Status (United States, Bureau of the Census 2004a); Politics (National Journal 2004); Population Growth (United States, Bureau of the Census 2004a); Income 1991-2001 (United States, Bureau of Economic Analysis 2003); Income growth 1969-2001 (United States, Bureau of Economic Analysis 2003); Sex Ratio (United States, Bureau of the Census 2004a); Income Distribution by Age (United States, Bureau of the Census 2004a).

TABLE 4: PERFORMANCE OF WEIGHT MATRICES

Name	% of Moran-z with p-value below: 0.10	0.05	0.00000	net probability has higher Moran-z than all matrices	category	Description
Distance						
Contig	91.2	89.8	74.6	36.0	-40.2	Contiguous LMAs
Dist	90.7	89.3	82.4	69.0	40.2	Inverted Squared Great Circle Distance
Culture						
Ancestry						
anchi90	85.4	82.9	59.0	33.0	15.4	Herfindahl-type ancestry 1990
aned2000	88.3	85.9	62.4	31.0	11.5	Inverse Euclidean distance ancestry 2000
anhi2000	89.3	87.8	71.7	61.5	71.3	Herfindahl-type ancestry 2000
Elem	76.6	72.7	37.1	-18.9	-72.3	Similarity ancestry marriage matrix 1990 (Eq. A1)
pumsanc	82.0	80.0	47.3	4.2	-38.9	Similarity ancestry marriage matrix 2000
sameancest	86.8	85.9	61.5	33.9	13.1	Binary Similarity Principal Ancestry 1990 2000 (Figure 1)
Religion						
ed90	84.9	82.4	49.8	6.0	41.4	Inverse Euclidean distance Religion 1990
hi90	86.3	84.9	62.0	33.3	58.4	Herfindahl-type religion 1990
rs1890	53.2	45.9	3.4	-65.6	-44.3	Similarity religion, phenetic classification 1890
rs1926	53.2	46.8	1.0	-67.9	-50.8	Similarity religion, phenetic classification 1926
rs1936	57.6	51.2	3.4	-60.2	-30.1	Similarity religion, phenetic classification 1936
rs1990	65.9	59.5	5.9	-57.3	-29.3	Similarity religion, phenetic classification 1990
samec8	88.8	87.3	62.0	26.4	49.1	Religion 1890, binary same main religion (Figure 2)
Xbn	80.0	76.6	29.8	-18.0	12.7	Religion 1990, 4 categories, binary same 2 main religions
Xed	81.0	77.6	31.2	-14.8	22.4	Inverse Euclidean distance Religion 1990, 4 categories
Xhi	57.1	51.2	12.7	-52.8	-39.7	Herfindahl-type religion 1990, 4 categories
xn	69.3	66.8	36.1	-16.4	10.0	Religion 1990, 4 categories, binary same main religion
Place Names						
pnsiml	45.9	32.2	0.5	-74.0	0.0	Place name similarity, population weighted
Elections						
edpres	88.8	87.3	47.3	19.7	44.6	Inverse Euclidean Distance, Presidential Elections
hipres	60.5	54.6	13.7	-49.8	-44.6	Herfindahl-type, Presidential Elections
Ecology						
simleco	89.3	87.8	55.1	17.2	0.0	Similarity ecoregions
Labor Flows						
n10	91.2	91.2	78.5	59.2	54.6	Commuting flows 2000
n9	90.2	86.3	51.7	3.4	-46.2	Commuting flows 1990
pop	86.3	85.9	63.4	25.8	-8.5	Migration flows 1990 of ages 25-64
Level of Development						
sliv	91.7	91.2	75.6	60.7	0.0	Inverse Euclidean distance Life expect., PCI, Educ. attainment
Economic Structure						
Industries						
edcbp	86.3	82.4	24.4	-16.1	47.1	Inverse Euclidean distance County Business Patterns 2001
hicbp	68.8	64.4	5.9	-50.2	-47.1	Herfindahl-type County Business Patterns 2001
Occupations						
edocc	87.3	85.9	38.0	-9.3	-45.1	Inverse Euclidean distance EEOF occupations 1990
hiocc	89.8	88.3	61.0	24.8	45.1	Herfindahl-type EEOF occupations 1990
Regional Structure						
edcbc	75.1	73.2	42.4	2.8	-4.0	Inverted Absolute Difference Avg. Calvin Beale Codes 1993
evn10	86.8	84.9	35.1	-16.5	-47.9	Eigenvector weights, single strongest upward link
evnal	91.2	90.2	66.8	15.5	3.5	Eigenvector weights, all links
evnup	90.7	90.2	71.2	24.4	48.4	Eigenvector weights, upward links

Notes: Net probability is the probability that—for a given variable—the weight matrix has a higher Moran z-score than other weight matrices minus the probability that the weight matrix has a lower Moran z-score than other weight matrices. A negative net probability therefore indicates that a weight matrix tends to have a relatively low Moran-z; a high positive net probability indicates that it has a relatively high Moran-z.

TABLE 5: NET PROBABILITIES AMONG 10 WEIGHT MATRICES, BY VARIABLE CATEGORY.

Weight Matrix	Mean	Ancestry	Distance	Elections	Reg. Str.	Religion	Occupation	Labor Flow	Ecology	Lev. Dev.
Crime										
Ancestry	42	0	25	75	50	75	25	25	100	0
Distance	50	-25	0	50	100	75	50	75	100	25
Elections	8	-75	-50	0	25	50	25	25	100	-25
Reg. Str.	-8	-50	-100	-25	0	25	50	-50	100	-25
Religion	-39	-75	-75	-50	-25	0	-25	-50	0	-50
Occupation	-22	-25	-50	-25	-50	25	0	-50	50	-75
Labor Flow	11	-25	-75	-25	50	50	50	0	50	25
Ecology	-67	-100	-100	-100	-100	0	-50	-50	0	-100
Lev. Dev.	25	0	-25	25	25	50	75	-25	100	0
Education										
Ancestry	0	0	20	60	20	-100	-20	20	100	-100
Distance	0	-20	0	100	100	-100	-60	-60	100	-60
Elections	-40	-60	-100	0	60	-100	-60	-100	60	-60
Reg. Str.	-44	-20	-100	-60	0	-100	-60	-100	100	-60
Religion	62	100	100	100	100	0	60	60	100	-60
Occupation	18	20	60	60	60	-60	0	60	60	-100
Labor Flow	18	-20	60	100	100	-60	-60	0	100	-60
Ecology	-80	-100	-100	-60	-100	-100	-60	-100	0	-100
Lev. Dev.	67	100	60	60	60	60	100	60	100	0
Environment										
Ancestry	-15	0	-100	100	-33	-33	-33	100	-100	-33
Distance	67	100	0	100	100	100	100	100	33	-33
Elections	-67	-100	-100	0	-33	-100	-33	-100	-100	-33
Reg. Str.	0	33	-100	33	0	33	-33	33	33	-33
Religion	0	33	-100	100	-33	0	-33	33	33	-33
Occupation	0	33	-100	33	33	33	0	33	33	-100
Labor Flow	-37	-100	-100	100	-33	-33	-33	0	-100	-33
Ecology	15	100	-33	100	-33	-33	-33	100	0	-33
Lev. Dev.	37	33	33	33	33	33	100	33	33	0
Male/Female Occupation Difference										
Ancestry	-22	0	-100	100	100	-100	-100	-100	100	-100
Distance	22	100	0	100	100	100	-100	-100	100	-100
Elections	-89	-100	-100	0	-100	-100	-100	-100	-100	-100
Reg. Str.	-44	-100	-100	100	0	-100	-100	-100	100	-100
Religion	0	100	-100	100	100	0	-100	-100	100	-100
Occupation	67	100	100	100	100	100	0	100	100	-100
Labor Flow	44	100	100	100	100	100	-100	0	100	-100
Ecology	-67	-100	-100	100	-100	-100	-100	-100	0	-100
Lev. Dev.	89	100	100	100	100	100	100	100	100	0
Percent Children in Arrangement										
Ancestry	61	0	46	100	62	54	92	54	77	62
Distance	38	-46	0	8	100	23	100	85	100	-23
Elections	-4	-100	-8	0	23	23	31	-8	46	-46
Reg. Str.	-32	-62	-100	-23	0	-31	46	-100	23	-46
Religion	-3	-54	-23	-23	31	0	69	-8	38	-62
Occupation	-64	-92	-100	-31	-46	-69	0	-100	-38	-100
Labor Flow	15	-54	-85	8	100	8	100	0	100	-46
Ecology	-48	-77	-100	-46	-23	-38	38	-100	0	-85
Lev. Dev.	38	-62	23	46	46	62	100	46	85	0
Percent Females: Child Status and Labor Force Status										
Ancestry	38	0	-14	100	71	62	71	-5	33	24
Distance	60	14	0	90	100	52	100	90	71	24
Elections	-47	-100	-90	0	-14	-52	14	-90	-14	-71
Reg. Str.	-32	-71	-100	14	0	-14	14	-100	14	-43
Religion	-10	-62	-52	52	14	0	33	-52	5	-24
Occupation	-44	-71	-100	-14	-14	-33	0	-90	5	-81
Labor Flow	36	5	-90	90	100	52	90	0	62	14
Ecology	-23	-33	-71	14	-14	-5	-5	-62	0	-33
Lev. Dev.	21	-24	-24	71	43	24	81	-14	33	0
Percent Households by Household Type										
Ancestry	46	0	0	100	89	0	67	44	67	44
Distance	32	0	0	56	100	11	56	22	78	-33
Elections	-25	-100	-56	0	0	-11	0	-22	22	-56
Reg. Str.	-43	-89	-100	0	0	-56	0	-100	-11	-33
Religion	11	0	-11	11	56	0	56	0	33	-44

Weight Matrix	Mean	Ancestry	Distance	Elections	Reg. Str.	Religion	Occupation	Labor Flow	Ecology	Lev. Dev.
Occupation	-40	-67	-56	0	0	-56	0	-56	-33	-89
Labor Flow	16	-44	-22	22	100	0	56	0	67	-33
Ecology	-28	-67	-78	-22	11	-33	33	-67	0	-33
Lev. Dev.	31	-44	33	56	33	44	89	33	33	0
Level of Development										
Ancestry	44	0	100	100	100	-100	100	100	100	-100
Distance	22	-100	0	100	100	-100	100	100	100	-100
Elections	-22	-100	-100	0	100	-100	100	-100	100	-100
Reg. Str.	-67	-100	-100	-100	0	-100	-100	-100	100	-100
Religion	67	100	100	100	100	0	100	100	100	-100
Occupation	-44	-100	-100	-100	100	-100	0	-100	100	-100
Labor Flow	0	-100	-100	100	100	-100	100	0	100	-100
Ecology	-89	-100	-100	-100	-100	-100	-100	-100	0	-100
Lev. Dev.	89	100	100	100	100	100	100	100	100	0
Percent Population by Household Type										
Ancestry	27	0	-15	77	62	8	46	15	46	8
Distance	44	15	0	62	100	15	77	54	100	-23
Elections	-28	-77	-62	0	0	-23	0	-62	31	-62
Reg. Str.	-37	-62	-100	0	0	-31	-15	-92	23	-54
Religion	3	-8	-15	23	31	0	31	-15	23	-38
Occupation	-32	-46	-77	0	15	-31	0	-77	15	-85
Labor Flow	23	-15	-54	62	92	15	77	0	62	-31
Ecology	-38	-46	-100	-31	-23	-23	-15	-62	0	-38
Lev. Dev.	36	-8	23	62	54	38	85	31	38	0
Life Expectancy										
Ancestry	67	0	100	100	100	33	100	100	100	-33
Distance	7	-100	0	-33	100	-100	100	100	100	-100
Elections	7	-100	33	0	100	-100	100	33	100	-100
Reg. Str.	-67	-100	-100	-100	0	-100	100	-100	-100	-100
Religion	59	-33	100	100	100	0	100	100	100	-33
Occupation	-89	-100	-100	-100	-100	-100	0	-100	-100	-100
Labor Flow	-15	-100	-100	-33	100	-100	100	0	100	-100
Ecology	-44	-100	-100	-100	100	-100	100	-100	0	-100
Lev. Dev.	74	33	100	100	100	33	100	100	100	0
Marital Status										
Ancestry	50	0	25	100	75	50	50	50	75	25
Distance	42	-25	0	25	100	50	50	75	100	0
Elections	-3	-100	-25	0	25	25	0	0	75	-25
Reg. Str.	-44	-75	-100	-25	0	0	-25	-100	0	-75
Religion	-19	-50	-50	-25	0	0	25	-50	25	-50
Occupation	-28	-50	-50	0	25	-25	0	-50	0	-100
Labor Flow	19	-50	-75	0	100	50	50	0	100	0
Ecology	-50	-75	-100	-75	0	-25	0	-100	0	-75
Lev. Dev.	33	-25	0	25	75	50	100	0	75	0
Politics										
Ancestry	-67	0	-100	-100	-100	-100	100	-100	-100	-100
Distance	67	100	0	100	-100	100	100	100	100	100
Elections	-22	100	-100	0	-100	-100	100	-100	-100	100
Reg. Str.	89	100	100	100	0	100	100	100	100	100
Religion	22	100	-100	100	-100	0	100	-100	100	100
Occupation	-89	-100	-100	-100	-100	-100	0	-100	-100	-100
Labor Flow	44	100	-100	100	-100	100	100	0	100	100
Ecology	0	100	-100	100	-100	-100	100	-100	0	100
Lev. Dev.	-44	100	-100	-100	-100	-100	100	-100	-100	0
Population Growth										
Ancestry	6	0	0	0	0	50	0	0	-50	50
Distance	72	0	0	100	100	100	100	100	50	100
Elections	-56	0	-100	0	-100	-50	-50	-100	-50	-50
Reg. Str.	6	0	-100	100	0	50	100	-100	-50	50
Religion	-22	-50	-100	50	-50	0	50	-100	-50	50
Occupation	-44	0	-100	50	-100	-50	0	-100	-50	-50
Labor Flow	44	0	-100	100	100	100	100	0	0	100
Ecology	28	50	-50	50	50	50	50	0	0	50
Lev. Dev.	-33	-50	-100	50	-50	-50	50	-100	-50	0
Income 1991-2001										
Ancestry	1	0	-50	88	25	25	-50	25	38	-88
Distance	38	50	0	88	100	50	-50	63	75	-38
Elections	-60	-88	-88	0	-75	-25	-63	-75	-38	-88

Weight Matrix	Mean	Ancestry	Distance	Elections	Reg. Str.	Religion	Occupation	Labor Flow	Ecology	Lev. Dev.
Reg. Str.	-25	-25	-100	75	0	0	-75	-50	38	-88
Religion	-22	-25	-50	25	0	0	-75	-13	25	-88
Occupation	46	50	50	63	75	75	0	63	75	-38
Labor Flow	4	-25	-63	75	50	13	-63	0	88	-38
Ecology	-39	-38	-75	38	-38	-25	-75	-88	0	-50
Lev. Dev.	57	88	38	88	88	88	38	38	50	0
Income growth 1969-2001										
Ancestry	0	0	-63	63	0	38	0	-50	25	-13
Distance	50	63	0	88	75	75	50	0	63	38
Elections	-46	-63	-88	0	-25	-13	-38	-88	-38	-63
Reg. Str.	-21	0	-75	25	0	13	-13	-100	38	-75
Religion	-29	-38	-75	13	-13	0	-38	-63	13	-63
Occupation	1	0	-50	38	13	38	0	-38	38	-25
Labor Flow	47	50	0	88	100	63	38	0	75	13
Ecology	-29	-25	-63	38	-38	-13	-38	-75	0	-50
Lev. Dev.	26	13	-38	63	75	63	25	-13	50	0
Sex Ratio										
Ancestry	32	0	19	52	43	52	33	29	33	29
Distance	31	-19	0	52	52	48	43	24	57	19
Elections	-14	-52	-52	0	10	24	0	-29	10	-38
Reg. Str.	-26	-43	-52	-10	0	-19	5	-48	-24	-48
Religion	-20	-52	-48	-24	19	0	5	-48	-14	-14
Occupation	-19	-33	-43	0	-5	-5	0	-43	-5	-38
Labor Flow	16	-29	-24	29	48	48	43	0	38	-5
Ecology	-13	-33	-57	-10	24	14	5	-38	0	-19
Lev. Dev.	13	-29	-19	38	48	14	38	5	19	0
Income Distribution by Age										
Ancestry	22	0	-33	33	100	-33	100	0	67	-33
Distance	59	33	0	100	100	0	100	100	100	0
Elections	-19	-33	-100	0	67	-33	67	-100	33	-67
Reg. Str.	-48	-100	-100	-67	0	-33	33	-100	-33	-33
Religion	30	33	0	33	33	0	67	33	67	0
Occupation	-63	-100	-100	-67	-33	-67	0	-100	0	-100
Labor Flow	26	0	-100	100	100	-33	100	0	100	-33
Ecology	-41	-67	-100	-33	33	-67	0	-100	0	-33
Lev. Dev.	33	33	0	67	33	0	100	33	33	0
Total										
Ancestry	31	0	-1	77	51	31	40	22	47	9
Distance	41	1	0	60	87	36	57	51	80	-3
Elections	-26	-77	-60	0	0	-10	0	-47	14	-54
Reg. Str.	-30	-51	-87	0	0	-20	2	-80	12	-50
Religion	-7	-31	-36	10	20	0	19	-26	19	-40
Occupation	-26	-40	-57	0	-2	-19	0	-52	4	-70
Labor Flow	20	-22	-51	47	80	26	52	0	66	-16
Ecology	-32	-47	-80	-14	-12	-19	-4	-66	0	-45
Lev. Dev.	30	-9	3	54	50	40	70	16	45	0

Notes: Net probability is the probability that—for a given variable—the weight matrix has a higher Moran z-score than other weight matrices minus the probability that the weight matrix has a lower Moran z-score than other weight matrices. A negative net probability therefore indicates that a weight matrix tends to have a relatively low Moran-z; a high positive net probability indicates that it has a relatively high Moran-z.

TABLE 6: NET PROBABILITY OF HIGHER MORAN Z-SCORE, NINE WEIGHT MATRICES, BY VARIABLE CATEGORY

Variable Group	Ancestry	Distance	Elections	Regional Structure	Religion	Occupations	Labor Flows	Ecology	Level of Development
Crime	42 (2)	50 (1)	8 (5)	-8 (6)	-39 (8)	-22 (7)	11 (4)	-67 (9)	25 (3)
Education	0 (5)	0 (5)	-40 (7)	-44 (8)	62 (2)	18 (3)	18 (3)	-80 (9)	67 (1)
Environment	-15 (7)	67 (1)	-67 (9)	0 (4)	0 (4)	0 (4)	-37 (8)	15 (3)	37 (2)
Male/Female Occupation Difference	-22 (6)	22 (4)	-89 (9)	-44 (7)	0 (5)	67 (2)	44 (3)	-67 (8)	89 (1)
Percent Children in Arrangement	61 (1)	38 (2)	-4 (6)	-32 (7)	-3 (5)	-64 (9)	15 (4)	-48 (8)	38 (2)
Percent Females: Child Status and Labor Force Status	38 (2)	60 (1)	-47 (9)	-32 (7)	-10 (5)	-44 (8)	36 (3)	-23 (6)	21 (4)
Percent Households by Household Type	46 (1)	32 (2)	-25 (6)	-43 (9)	11 (5)	-40 (8)	16 (4)	-28 (7)	31 (3)
Level of Development	44 (3)	22 (4)	-22 (6)	-67 (8)	67 (2)	-44 (7)	0 (5)	-89 (9)	89 (1)
Percent Population by Household Type	27 (3)	44 (1)	-28 (6)	-37 (8)	3 (5)	-32 (7)	23 (4)	-38 (9)	36 (2)
Life Expectancy	67 (2)	7 (4)	7 (4)	-67 (8)	59 (3)	-89 (9)	-15 (6)	-44 (7)	74 (1)
Marital Status	50 (1)	42 (2)	-3 (5)	-44 (8)	-19 (6)	-28 (7)	19 (4)	-50 (9)	33 (3)
Politics	-67 (8)	67 (2)	-22 (6)	89 (1)	22 (4)	-89 (9)	44 (3)	0 (5)	-44 (7)
Population Growth	6 (4)	72 (1)	-56 (9)	6 (4)	-22 (6)	-44 (8)	44 (2)	28 (3)	-33 (7)
Income 1991-2001	1 (5)	38 (3)	-60 (9)	-25 (7)	-22 (6)	46 (2)	4 (4)	-39 (8)	57 (1)
Income growth 1969-2001	0 (5)	50 (1)	-46 (9)	-21 (6)	-29 (7)	1 (4)	47 (2)	-29 (7)	26 (3)
Sex Ratio	32 (1)	31 (2)	-14 (6)	-26 (9)	-20 (8)	-19 (7)	16 (3)	-13 (5)	13 (4)
Income Distribution by Age	22 (5)	59 (1)	-19 (6)	-48 (8)	30 (3)	-63 (9)	26 (4)	-41 (7)	33 (2)

Notes: Net probability is the probability that—for a given variable—the weight matrix has a higher Moran z-score than other weight matrices minus the probability that the weight matrix has a lower Moran z-score than other weight matrices. A negative net probability therefore indicates that a weight matrix tends to have a relatively low Moran-z; a high positive net probability indicates that it has a relatively high Moran-z.

Appendix A: Marriage Matrix

The Census Public Use Microdata Sample (PUMS) is structured such that each record is either a household or a person within a household, and each field is drawn from the decennial census long form. PUMS data are available for either one percent or five percent of households, and in this section the one percent sample from 2000 is used.

For each person, there are two fields for ancestry, one field for race, and one field for Hispanic status. The codes entered in these fields are much more detailed than the categories available in county-level data (for example, there are 569 ancestry codes in the 2000 PUMS data). For all responses in which only the first ancestry was specified, the second ancestry was set equal to the first. For all responses where ancestry was "American," "Unspecified," or "Unclassifiable," the ancestry was set to detailed Hispanic classification, if the respondent was Hispanic, and set to detailed racial classification, if the respondent was not Hispanic. To make the categories compatible with the county-level data, the codes were aggregated to the 95 categories in the county-level STF3 data. Thus, for each person there exist two fields giving one of 95 racial/ethnic codes. For persons of homogeneous descent the codes in the two fields are identical; for persons of mixed ancestry, the codes are different.

The roles of each member of a household are coded in a specific field. One code indicates the head of the household, and another code indicates the spouse. Sex is given in yet another field. One can thus extract each married couple in the PUMS data. In each couple, there are four ancestries, and the number of pairs of male ancestries with female ancestries is four {(male1,female1), (male1,female2), (male2,female1), (male2,female2)}. **M** is a 95x95 matrix, where each element m_{ij} is the number of times a male ancestry i is paired with a female ancestry j. From **M**, one can construct a second 95x95 matrix **W**=**M**+**M**\`, so that each element $w_{ij} = m_{ij} + m_{ji}$, giving the number of times a person from ancestry i is paired with a person from ancestry j. Finally, one can create a 95x95 matrix **Q**, where each element $q_{ij}=w_{ij}/\Sigma_j w_{ij}$, giving the percent of times a person from ancestry i is paired with a person from ancestry j. The diagonal of **Q** gives the rate of endogamy for each ancestry group.

One problem with matrix **Q** is that it gives no clue about how the pattern of marriage for an ancestry group deviates from chance. For example, a person from the ancestry group "Danish" is very likely to marry someone from the ancestry group "Black" based on chance, since each "Danish" person would have about a 10 percent chance of drawing a "Black" person at random from the population. To address this issue, one can express the elements of matrix **Q** in a form similar to a location quotient. Create a new matrix **T**, where each element $t_{ij}=\Sigma_i\, m_{ij}/\Sigma_i\, \Sigma_j m_{ij}$, giving the probability that a person from ancestry group i would draw a person from ancestry group j at random for the population. Each row in matrix **T** is, of course, identical. One can then create a new matrix **Z**, where each element z_{ij} is the element-wise division of matrix **Q** by matrix **T**. Dividing each element z_{ij} by the row maximum, and then replacing each element z_{ij} with the geometric mean of z_{ij} and z_{ji}, one obtains the matrix **S** used in Equation (7) as a similarity index between ancestry groups.

Table A1 summarizes some of the information from marriage matrix **Q**. The table is sorted, such that the ancestry groups that have the highest percent endogamous marriages are at the top, and those with the lowest rate of endogamy are at the bottom. The table also shows for each ancestry group the four ancestry groups most often married. Most ancestry groups have themselves as the principal marriage partner. A few groups, however, do not have themselves among the top four principal marriage partners.

Table A1: Principal Marriage Partners among Ancestry Groups

Ancestry	Endogamy %	Partner 1	%	Partner 2	%	Partner 3	%	Partner 4	%
Black	91	Black	91	Hispanic	1	White	1	African	1
Somalian	86	Somalian	86	Black	8	African	3	Other Race	1
Afghan	84	Afghan	84	Asian	3	White	2	Iranian	2
Hispanic	84	Hispanic	84	White	4	German	2	Irish	2
Sudanese	83	Sudanese	83	Black	5	Lebanese	2	Haitian	2
Haitian	81	Haitian	81	Black	8	Hispanic	2	Jamaican	1
Asian	80	Asian	80	White	4	German	3	Hispanic	2
Ethiopian	80	Ethiopian	80	Black	6	African	5	White	2
White	80	White	80	German	4	Irish	2	English	2
Ghanian	76	Ghanian	76	Black	10	White	3	African	3

Ancestry	Endogamy %	Partner 1	%	Partner 2	%	Partner 3	%	Partner 4	%
Senegalese	76	Senegalese	76	Black	14	Lithuanian	5	African	5
Nigerian	74	Nigerian	74	Black	16	African	2	German	1
Assyrian	74	Assyrian	74	White	4	German	3	Iraqi	2
Guyanese	71	Guyanese	71	Black	6	Asian	5	Jamaican	4
Iranian	70	Iranian	70	White	6	German	3	Hispanic	3
Egyptian	69	Egyptian	69	White	5	German	4	Hispanic	2
Albanian	67	Albanian	67	Irish	5	White	5	Italian	4
Kenyan	63	Kenyan	63	Black	15	Asian	5	African	5
Palestinian	63	Palestinian	63	Arab	5	White	5	Hispanic	3
Jamaican	62	Jamaican	62	Black	18	Hispanic	3	White	1
African	61	African	61	Black	21	Hispanic	3	White	2
Brazilian	61	Brazilian	61	White	6	Hispanic	5	Portuguese	5
Liberian	60	Liberian	60	Black	15	African	10	Hispanic	2
Iraqi	60	Iraqi	60	White	6	Hispanic	5	Assyrian	4
European	60	European	60	White	6	German	6	English	4
Arab	60	Arab	60	White	6	Hispanic	5	German	3
Jordanian	59	Jordanian	59	White	6	German	4	Hispanic	4
Armenian	58	Armenian	58	White	5	German	5	Italian	4
Cape Verdean	58	Cape Verdean	58	Black	10	Irish	5	English	4
Bulgarian	57	Bulgarian	57	White	7	German	7	Irish	5
Trinidadian	55	Trinidadian	55	Black	17	Jamaican	4	Hispanic	4
South African	51	South African	51	White	13	German	5	British	5
Eastern European	51	Eastern European	51	White	7	German	6	Irish	4
Northern European	51	Northern European	51	German	7	White	6	English	5
Macedonian	50	Macedonian	50	German	10	English	5	Irish	5
Pacific Islander	46	Pacific Islander	46	Asian	12	White	8	Hispanic	5
Israeli	45	Israeli	45	White	15	Russian	5	German	5
British West Indian	43	British West Indian	43	Black	15	Jamaican	6	Trinidadian	6
Barbadian	43	Barbadian	43	Black	23	British West Indian	9	Jamaican	5
Yugoslavian	41	Yugoslavian	41	German	10	Irish	8	White	6
Turkish	40	Turkish	40	White	10	German	6	Irish	6
Portuguese	40	Portuguese	40	Irish	8	White	7	German	7
Belizean	37	Belizean	37	Black	20	Hispanic	11	White	9
West Indian	36	West Indian	36	Black	24	Hispanic	6	Jamaican	3
Romanian	36	Romanian	36	German	9	White	7	Russian	6
Moroccan	36	Moroccan	36	White	10	Hispanic	6	Irish	5
Greek	35	Greek	35	German	10	White	9	Irish	9
German	35	German	35	Irish	12	English	10	White	8
Russian	33	Russian	33	German	11	Polish	8	White	7
English	33	English	33	German	17	Irish	11	White	7
Other Race	32	Other Race	32	White	28	Hispanic	5	Asian	5
Cypriot	30	Cypriot	30	Greek	16	German	12	Irish	8
Italian	30	Italian	30	Irish	13	German	13	White	8
Ukrainian	29	Ukrainian	29	German	12	Irish	8	Italian	7
Estonian	29	Estonian	29	Irish	11	White	9	German	7
US Virgin Islander	29	US Virgin Islander	29	Black	18	German	7	Hispanic	7
Serbian	27	Serbian	27	German	15	White	8	English	7
British	27	British	27	German	14	White	10	Irish	8
Carpatho-Rusyn	27	Carpatho-Rusyn	27	German	11	Slovak	11	Italian	6
Syrian	26	Syrian	26	German	9	Irish	8	Italian	8
American Indian	25	American Indian	25	White	17	German	14	Irish	11
Lebanese	25	Lebanese	25	White	11	German	10	Irish	9
Maltese	24	Maltese	24	German	14	Italian	13	Irish	9
French Canadian	24	French Canadian	24	German	12	Irish	11	White	10
Irish	23	Irish	23	German	19	English	11	White	8
Polish	22	Polish	22	German	17	Irish	11	Italian	8
Canadian	19	Canadian	19	White	18	German	11	English	8
Latvian	19	Latvian	19	German	12	White	8	English	7
Norwegian	18	German	25	Norwegian	18	English	9	Irish	9
Croatian	16	German	17	Croatian	16	Irish	11	English	7
Slovene	16	German	18	Slovene	16	Irish	9	English	8
Australian	15	Australian	15	White	15	English	11	German	11

Ancestry	Endogamy %	Partner 1	%	Partner 2	%	Partner 3	%	Partner 4	%
Finnish	15	German	17	Finnish	15	English	10	Irish	9
Slovak	15	German	17	Slovak	15	Irish	11	Italian	10
Dutch	15	German	21	Dutch	15	Irish	12	English	11
French	14	German	18	French	14	Irish	12	English	11
Scotch Irish	14	German	18	English	14	Scotch Irish	14	Irish	13
Basque	12	German	15	White	13	Basque	12	English	12
Bahamian	12	Black	42	Bahamian	12	White	6	Jamaican	6
Czech	12	German	23	Czech	12	Irish	11	English	9
Belgian	12	German	21	Belgian	12	Irish	10	White	8
Lithuanian	12	German	15	Lithuanian	12	Irish	11	Polish	9
Dutch West Indian	11	German	15	American Indian	14	White	14	Irish	14
Icelander	11	German	17	English	12	Icelander	11	White	11
Scandinavian	11	German	20	Scandinavian	11	English	11	Irish	9
Luxemburger	11	German	29	Irish	11	Luxemburger	11	English	7
Hungarian	11	German	17	Hungarian	11	Irish	11	White	9
New Zealander	10	White	13	German	12	English	12	New Zealander	10
Swiss	9	German	25	English	14	Irish	9	Swiss	9
Swedish	9	German	22	English	12	Irish	10	Swedish	9
Scottish	9	German	18	English	16	Irish	13	Scottish	9
Danish	7	German	22	English	15	Irish	9	White	8
Slavic	6	German	20	Irish	11	English	10	White	8
Austrian	6	German	18	Irish	11	English	10	White	7
Welsh	4	German	21	English	15	Irish	13	White	7

Figure A1 provides an additional perspective on the marriage relationships among ancestry groups. Using a technique known as block-modeling (Scott 2000: 131f; Wasserman and Faust 1994; Moody 2000), one can build a dendritic diagram that represents the similarity among ancestries. The similarity is not simply that of lumping together two ancestries that inter-marry with each other, but also considers that two ancestries may have similar patterns in marrying into other ancestries. While not all the dendritic relationships seem reasonable, the overall pattern resonates with conventional ideas of interethnic relations. All ancestries pointing toward node ‘5’ on the upper right of the graph are ultimately of African origin. Non-African ancestries that existed in large numbers during colonial times (American Indian, British ancestries, French, Dutch) share the branch pointing toward node ‘12’ on the lower right of the graph. At node ‘6,’ this branch is joined by another branch with Scandinavian, German, and other west European ancestries. This large branch pointing to node ‘6’ consists of ancestries with relatively low rates of endogamy, and one can presume that the differences among these groups are not particularly salient (assimilation has done its work.

The graph in Figure A1 provides a suggestion of a second way in which one might construct a similarity matrix **S** among ancestries. Following the methods first developed by White, Burton, and Dow (1981), and then modified slightly by Eff (2004), one may apply the following formula:

(A1) $$s_{ij} = \frac{\partial_{\max} - \partial_{ij} + 1}{\partial_{\max} + 1}$$

where s_{ij} is the similarity between ancestry group i and ancestry group j, d_{max} is the length of the longest path in the relationship graph (Figure A1), and d_{ij} is the length of the longest path to the nearest node connecting ancestor group i and ancestor group j. The similarity between each ancestry group will thus always be greater than zero and less than or equal to one.

Equation (A1) is also used to produce a similarity matrix among religions, based on the taxonomy presented in Table A2. The similarity matrix is then used in Equation (10).

TABLE A2: PHENETIC CLASSIFICATION OF DENOMINATIONS

level1	level2	level3	level4	level5	level6	level7
Judaeo-Christian	Christian	Western	Protestant	I. Mainline		
Judaeo-Christian	Christian	Western	Protestant	II. Evangelical Prot	A. Revivalist/Experiential	1. Holiness
Judaeo-Christian	Christian	Western	Protestant	II. Evangelical Prot	A. Revivalist/Experiential	2. Pentecostal
Judaeo-Christian	Christian	Western	Protestant	II. Evangelical Prot	B. Revivalist/Rational	1. Baptist
Judaeo-Christian	Christian	Western	Protestant	II. Evangelical Prot	B. Revivalist/Rational	2. Other
Judaeo-Christian	Christian	Western	Protestant	II. Evangelical Prot	C. Evan Prot Other	
Judaeo-Christian	Christian	Western	Protestant	III. Black Protestant	A. Black Mainline	
Judaeo-Christian	Christian	Western	Protestant	III. Black Protestant	B. Black Other	
Judaeo-Christian	Christian	Western	Catholic	IV. Catholic	A. Latin & B. Other Rites	
Judaeo-Christian	Christian	Western	Catholic	IV. Catholic	B. Other Rites	
Judaeo-Christian	Christian	Western	Protestant	V. Other Christian	B. Confessional	
Judaeo-Christian	Christian	Western	Protestant	V. Other Christian	C. Friends	
Judaeo-Christian	Christian	Western	Protestant	V. Other Christian	D. Latter Day Saints	
Judaeo-Christian	Christian	Eastern		V. Other Christian	E. Orthodox	
Judaeo-Christian	Christian	Western	Protestant	V. Other Christian	F. Mainline Tributaries	
Judaeo-Christian	Christian	Western	Protestant	V. Other Christian	G. Pietist	
Judaeo-Christian	Christian	Western	Protestant	V. Other Christian	H. Theologically Liberal	
Judaeo-Christian	Christian	Western	Protestant	V. Other Christian	I. Other	
Judaeo-Christian				VI. Non-Christian	D. Jewish	
Hindic					Buddhist	

Notes: Source is Bradley (1992).

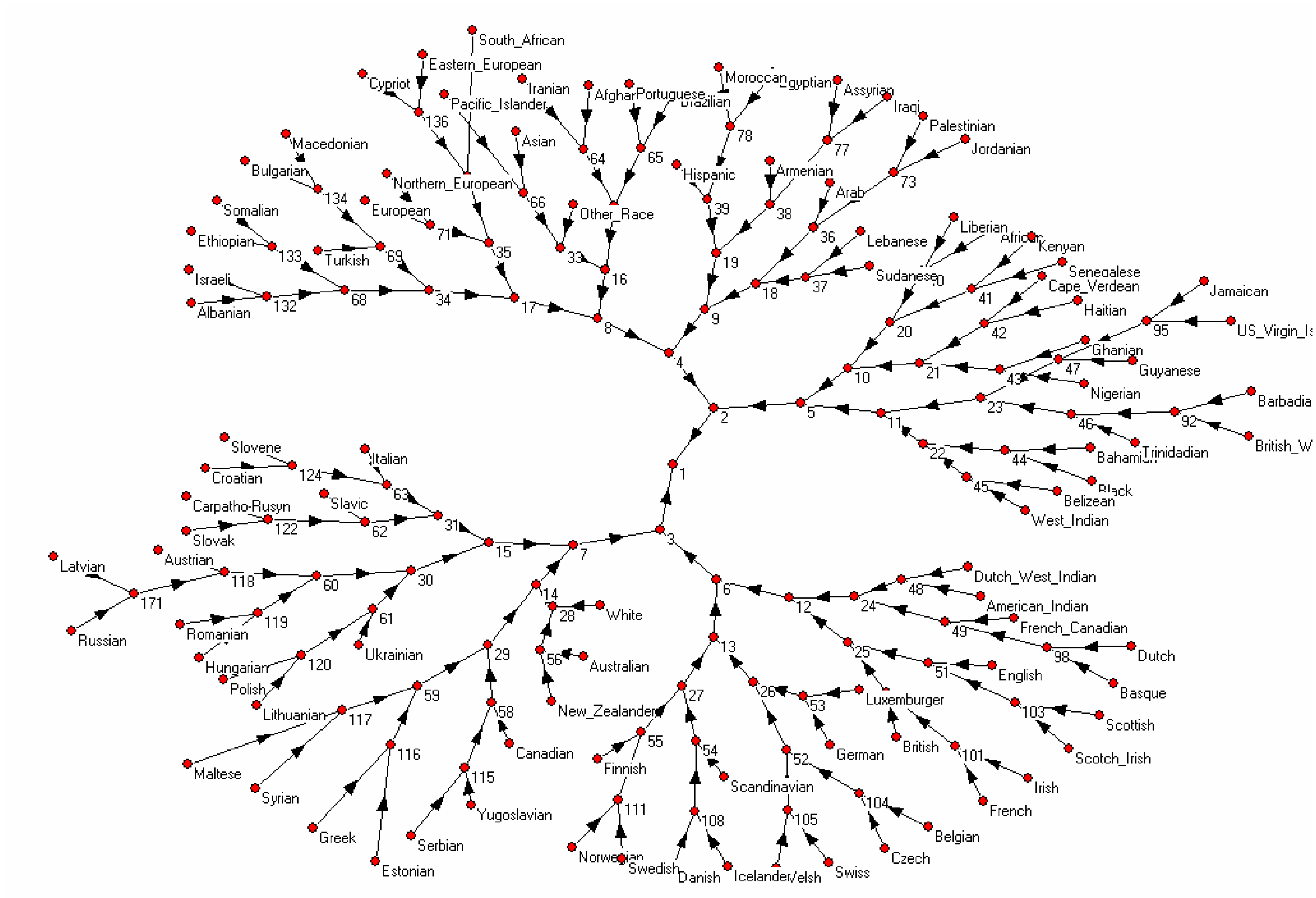


Figure A1: Block-Modeling the Marriage Matrix.