APPENDIX A

Variables and Variable Definitions

All demographic county-level variables have been drawn directly from the 1970, 1980, and 1990 U.S. *Censuses of Population,* published by the U.S. Department of Commerce, Bureau of the Census. Variables for political party registration and party voting originate from the state agency in each state in charge of administering elections.

% 1980 Group Population	Immigrant group's percentage of total population residing in the county in 1980
% Unemployment 1980	Unemployment rate in 1980
Change in Real Median Income 1980–90	Dollar change in real median family income from 1980–90 expressed in thousands of 1992 dollars
Net Population Change	Percentage of population growth 1970–80, 1980–90
Population Density	Population/square mile of land area (in 1980, 1990)
% College Students	Percentage of county population enrolled in colleges within the county in 1980
Median Income	Median family income (in 1980 and 1990)
% College Educated	Percentage of population within county with four-year college degree (in 1980 and 1990)
% Born Out of State (or % Internal Migrants)	Percentage of population born outside the state where currently residing, 1980 and 1990, exclud-

	ing the foreign-born population and those born in U.S. territories such as Guam or Puerto Rico
% Pre-1965 Immigrants	Percentage of population that is foreign born arriving in the United States prior to 1965 (in 1980 and 1990)
Isolation of Minorities from Whites	Dissimilarity index calculated across tracts within counties for white-Asian, white-Hispanic, and white-black concentration and then summed (in 1980 and 1990) (*See Chapter 2, footnote 1*)
Asian Segregation	Dissimilarity index calculated across tracts within counties for white-Asian concentration (in 1980 and 1990) (*See Chapter 2, footnote 1*)
Hispanic Segregation	Dissimilarity index calculated across tracts within counties for white-Hispanic concentration (in 1980 and 1990)
% Post-1970 Immigrants	Percentage of population that is foreign born arriving in the United States after 1965 (in 1980 and 1990)
% of Foreign Born Naturalized	Percentage of the foreign-born population that is naturalized (in 1980 and 1990)
% Black	Percentage of the population that is African American (in 1980 and 1990)
% Turnout	Percentage of registered voters voting in either a gubernatorial or presidential election in year
Party Irregularity	Percentage of voters registered Republican in a year minus percentage of vote for the Republican

	candidate leading the ticket (either president or governor) in year, expressed in absolute terms
Change in % Born Out of State	Percentage born out of state (but in the United States) in 1990 minus percentage born out of state (but in the United States) in 1980 and the same for 1970 to 1980
% Foreign Born	Percentage foreign born (in 1980 and 1990)
% Hispanic	Percentage of Hispanic ancestry of any race, (in 1980 and 1990)
% Asian	Percentage of Asian ancestry (in 1980 and 1990)
Change in % Hispanic	Percentage Hispanic in 1990 minus percentage Hispanic in 1980 and the same for 1970 to 1980
Change in % Asian	Percentage Asian in 1990 minus percentage Asian in 1980 and the same for 1970 to 1980
% Republican Registrants	Republicans registered as a percentage of total registrants, 1980 and 1990
Change in % Republican Registrants	Republicans registered as a percentage of total registrants in 1990 minus the same figure for 1980 and the same for 1980 to 1970
Change in % Unemployment	Percentage unemployment in 1990 minus percentage unemployment in 1980 and the same for 1970 to 1980
% Over Age 65	Percentage of residents over age sixty-five (in 1980 and 1990)
Spatial Lag	Dependent variable spatially weighted to account for the influence of nearby observations. (*see appendix B*)

TABLE A2.1. Characteristics of California Natives, Immigrants, and Internal Migrants

Demographic	Native	Migrant	Immigrant
Mean age	37.8	49.4	40.0
Mean education	10.9	11.0	8.8
Mean wage and salary income	$22,687	$26,342	$18,624
Median wage and salary income	$10,000	$8,230	$6,213
Mean Social Security income	$2,061	$3,118	$1,840
% non-Hispanic white	70.7	80.0	20.0

Source: 1990 U.S. Census Public Use Microdata Sample for California, 1 percent sample, respondents over age eighteen. Figures are percentages unless otherwise indicated.

Note: These data reflect individual responses, not households.

TABLE A2.2. Influence of Spatial Segregation Internal to California Counties on Naturalization Rates, 1980 and 1990

Variable	1980	1990
Population density, 1980–90	.0008**	.0006**
	(.0003)	(.0003)
% foreign born	–.86**	–.41**
	(.17)	(.10)
Asian segregation	.33**	.10
	(.09)	(.09)
Hispanic segregation	–.41**	–.26**
	(.11)	(.12)
% college education	.47**	.53**
	(.18)	(.11)
Spatial lag	.44**	.14
	(.14)	(.20)
Constant	31.51	29.66
N	58	58
R^2_a	.85	.65

Note: Spatial autoregressive model, weighted for population; dependent variable = percentage of the immigrant population naturalized. See appendix A for a full description of variables.

$*p < .10$. $**p < .05$.

TABLE A2.3 Logistic Regression of the Influence of Nativity on Presidential Vote Choice, Controlling for Party Identification, Education, Income, Length of Residence, and Urban/Rural Location, by Ethnic Background

Variable	European and Canadian	Hispanic	Asian
Birthplace (native = 0,	.14	.31	.49
foreign born = 1)	(.21)	(.51)	(1.09)
Rural	.04	.55	–.13
	(.10)	(.58)	(1.24)
Urban	–.18	.48	1.01
	(.11)	(.41)	(1.13)
Income	.12**	.33**	–.32
	(.04)	(.19)	(.47)
Education	–.05*	.01	.63
	(.03)	(.13)	(.57)
Age	.0003	.004	.02
	(.003)	(.02)	(.05)
Length of residence	.003	–.016	–.05
	(.003)	(.015)	(.05)
Black	–.52	–.04	—
	(1.09)	(1.15)	
Party identification	1.69**	1.97***	2.35***
(1 = *D,* 2 = *I,* 3 = *R*)	(.05)	(.28)	(.63)
Constant	–3.09	–4.86	–7.52
N	4037	218	51
% correctly classified	81.1%	83.5%	84.3%
Null model	59.9%	61.0%	56.0%
Model χ^2	1,814.1	104.1	31.4
Significance	$p < .0001$	$p < .0001$	$p < .0001$

Source: ICPSR, *American National Election Studies Cumulative Datafile 1980–1994.*

Note: Dependent variable = presidential vote choice coded: 0 = Democrat, 1 = Republican.

*$p < .10$. **$p < .05$. ***$p < .01$.

TABLE A3.1. Characteristics of Colorado Natives, Immigrants, and Internal Migrants

Demographic	Native	Migrant	Immigrant
Mean age	41.0	44.4	42.3
Mean education	10.4	11.3	9.9
Mean wage and salary income	$12,365	$16,101	$11,651
Median wage and salary income	$5,600	$7,530	$6,000
Mean Social Security income	$2,218	$2,570	$2,101
% non-Hispanic white	77.1	88.6	51.5

Source: 1990 U.S. Census Public Use Microdata Sample for Colorado, 1 percent sample, respondents over age eighteen. Figures are percentages unless otherwise indicated.

Note: These data reflect individual responses, not households.

TABLE A3.2. Influence of Spatial Segregation Internal to Colorado Counties on Naturalization Rates, 1980 and 1990

Variable	1980	1990
Population density, 1980–90	.001	–.001
	(.001)	(.002)
% foreign born	.49	–1.99*
	(1.19)	(1.07)
Asian segregation	.26	–.17
	(.20)	(.20)
Hispanic segregation	–.43**	.08
	(.20)	(.25)
% college education	–.49**	.01
	(.23)	(.12)
Spatial lag	.34**	.40**
	(.13)	(.13)
Constant	43.07	40.82
N	63	63
R^2_a	.30	.23

Note: Spatial autoregressive model, weighted for population; dependent variable = percentage of the immigrant population naturalized.

$*p < .10$. $**p < .05$.

TABLE A4.1. Characteristics of Kansas Natives, Immigrants, and Internal Migrants

Demographic	Native	Migrant	Immigrant
Mean age	47.1	45.7	46.6
Mean education	10.3	10.7	9.7
Mean wage and salary income	$10,986	$13,993	$11,216
Median wage and salary income	$5,600	$7,530	$6,000
Mean Social Security income	$1,400	$1,166	$636
% non-Hispanic white	92.9	87.9	41.6

Source: 1990 U.S. Census Public Use Microdata Sample for Kansas, 1 percent sample, respondents over age eighteen. Figures are percentages unless otherwise indicated.

Note: These data reflect individual responses, not households.

TABLE A4.2. Influence of Spatial Segregation Internal to Kansas Counties on Naturalization Rates, 1980 and 1990

Variable	1980	1990
Population density, 1980–90	–.0007	.007
	(.006)	(.005)
% foreign born	–5.88**	–3.45**
	(1.34)	(.83)
Asian segregation	–.04	–.14
	(.08)	(.13)
Hispanic segregation	–.10	–.09
	(.18)	(.15)
% college education	–.10	–.35**
	(.39)	(.17)
Spatial lag	–.17	.17
	(.16)	(.13)
Constant	80.25	62.15
N	105	105
R^2_a	.17	.32

Note: Spatial autoregressive model, weighted for population; dependent variable = percentage of the immigrant population naturalized. See appendix A for a full description of variables.

*p < .10. **p < .05.

TABLE A5.1. Characteristics of Kentucky Natives, Immigrants, and Internal Migrants

Demographic	Native	Migrant	Immigrant
Mean age	45.4	42.2	41.6
Mean education	9.2	10.5	11.1
Mean wage and salary income	$10,250	$13,823	$14,045
Median wage and salary income	$3,671	$6,400	$3,725
Mean Social Security income	$1,113	$925	$641
% non-Hispanic white	92.4	90.0	57.1

Source: 1990 U.S. Census Public Use Microdata Sample for Kentucky, 1 percent sample, respondents over age eighteen. Figures are percentages unless otherwise indicated.

Note: These data reflect individual responses, not households.

TABLE A5.2. Influence of Spatial Segregation Internal to Kentucky Counties on Naturalization Rates, 1980 and 1990

Variable	1980	1990
Population density, 1980–90	.004	.008**
	(.004)	(.004)
% foreign born	–4.77*	.36
	(2.91)	(4.59)
Asian segregation	–.29*	–.32*
	(.17)	(.18)
Hispanic segregation	.08	–.01
	(.23)	(.08)
% college education	–2.99**	–1.35**
	(.81)	(.53)
Spatial lag	–.11	.02
	(.15)	(.14)
Constant	100.42	74.78
N	120	120
R^2_a	.19	.08

Note: Spatial autoregressive model, weighted for population; dependent variable = percentage of the immigrant population naturalized. See appendix A for a full description of variables.

*$p < .10$. **$p < .05$.

TABLE A6.1. Characteristics of Florida Natives, Immigrants, and Internal Migrants

Demographic	Native	Migrant	Immigrant
Mean age	38.8	50.5	47.7
Mean education	10.0	10.5	9.4
Mean wage and salary income	$12,163	$12,475	$10,662
Median wage and salary income	$7,800	$4,500	$5,000
Mean Social Security income	$2,061	$3,335	$2,672
% non-Hispanic white	67.7	89.0	28.4

Source: 1990 U.S. Census Public Use Microdata Sample for Florida, 1 percent sample, respondents over age eighteen. Figures are percentages unless otherwise indicated.

Note: These data reflect individual responses, not households.

TABLE A6.2. Influence of Spatial Segregation Internal to Florida Counties on Naturalization Rates, 1980 and 1990

Variable	1980	1990
Population density, 1980–90	.006**	–.0004
	(.002)	(.002)
% foreign born	–.69**	–.41**
	(.10)	(.14)
Asian segregation	–.54**	.23
	(.20)	(.28)
Hispanic segregation	.19	–.10
	(.12)	(.17)
% college education	–1.29**	–.63**
	(.47)	(.29)
Spatial lag	.04	.33
	(.18)	(.21)
Constant	85.62	43.88
N	67	67
R^2_a	.55	.31

Note: Spatial autoregressive model, weighted for population; dependent variable = percentage of the immigrant population naturalized. See appendix A for a full description of variables.

*$p < .10$. **$p < .05$.

TABLE A7.1. Characteristics of Pennsylvania Natives, Immigrants, and Internal Migrants

Demographic	Native	Migrant	Immigrant
Mean age	45.6	44.3	48.7
Mean education	9.9	10.8	9.5
Mean wage and salary income	$12,106	$15,208	$11,953
Median wage and salary income	$6,000	$6,899	$1,594
Mean Social Security income	$1,004	$943	$1,229
% non-Hispanic white	95.5	88.5	66.7

Source: 1990 U.S. Census Public Use Microdata Sample for Pennsylvania, 1 percent sample, respondents over age eighteen. Figures are percentages unless otherwise indicated.

Note: These data reflect individual responses, not households.

TABLE A7.2. Influence of Spatial Segregation Internal to Pennsylvania Counties on Naturalization Rates, 1980 and 1990

Variable	1980	1990
Population density, 1980–90	–.001	–.0004
	(.001)	(.002)
% foreign born	4.68**	–.41**
	(1.28)	(.14)
Asian segregation	–.37**	.23
	(.17)	(.28)
Hispanic segregation	–.19**	–.10
	(.07)	(.17)
% college education	–2.41**	–.63**
	(.45)	(.29)
Spatial lag	.47**	.33
	(.16)	(.21)
Constant	64.47	43.88
N	67	67
R^2_a	.55	.31

Note: Spatial autoregressive model, weighted for population; dependent variable = percentage of the immigrant population naturalized. See appendix A for a full description of variables.

*$p < .10$. **$p < .05$.

TABLE A8.1. Characteristics of New York Natives, Immigrants, and Internal Migrants

Demographic	Native	Migrant	Immigrant
Mean age	43.9	47.2	46.7
Mean education	10.6	10.9	9.2
Mean wage and salary income	$15,934	$17,118	$12,946
Median wage and salary income	$8,600	$7,596	$5,000
Mean Social Security income	$1,162	$1,352	$1,017
% non-Hispanic white	87.9	73.7	37.2

Source: 1990 U.S. Census Public Use Microdata Sample for New York, 1 percent sample, respondents over age eighteen. Figures are percentages unless otherwise indicated.

Note: These data reflect individual responses, not households.

TABLE A8.2. Influence of Spatial Segregation Internal to New York Counties on Naturalization Rates, 1980 and 1990

Variable	1980	1990
Population density, 1980–90	–.00002	–.0001*
	(.00005)	(.00006)
% foreign born	–.53**	–.38**
	(.12)	(.11)
Asian segregation	–.29**	–.17*
	(.10)	(.10)
Hispanic segregation	–.05	–.08
	(.05)	(.06)
% college education	–.28**	.04
	(.11)	(.07)
Spatial lag	.33**	.42**
	(.15)	(.16)
Constant	63.89	47.16
N	62	62
R^2_a	.85	.82

Note: Spatial autoregressive model, weighted for population; dependent variable = percentage of the immigrant population naturalized. See appendix A for a full description of variables.

$*p < .10$. $**p < .05$.

APPENDIX B

Basics of Spatial Regression Analysis

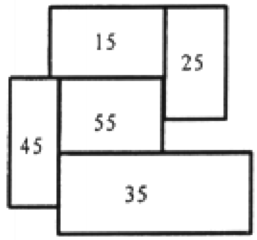
Boundaries around geographical units (counties, cities, census tracts, districts) are usually arbitrarily drawn. This gives rise to the problem of spatial dependence among observations that are spatially related. When the values of a variable across units of analysis are spatially related, the standard regression assumption of independence among the observations is violated. Fortunately, recent developments in statistics and econometrics have provided potential corrections or controls for the effects that the spatial arrangement may have on the values of given observations for a variable. When the values of an observation are closely related to the values of nearby observations, the condition is analogous to time-series autocorrelation when the values at a particular time for a given observation are related to the values of previous times for that observation:

Time series autocorrelation

Time:	$t-4$	$t-3$	$t-2$	$t-1$	t
Observation:	15	25	35	45	55

The analogy is imperfect, however, because spatial autocorrelation may occur in multiple "directions." In the time-series case, the dependency is only backward in time. Typically, the values at time t are modeled as a function of the values at $t-1$ and other explanatory variables. The values at time t can also be modeled as a function of the values at $t-2$, $t-3$, . . . $t-n$. But the directionality is always unidimensional, making temporally lagged variables very easy to compute.

In the case of *spatial autocorrelation,* the dependency can be in multiple directions, in two dimensions:



As in time-series autocorrelation, the dependency can extend to more than just the immediately adjacent observations to observations that are more distant.

Just as the goal in time-series analysis is to account for the influence of serial correlation on the dependent variable, the goal in spatial data analysis is to eliminate the influence of spatial dependence. According to Anselin, spatial dependence is the existence of a functional relationship between what happens at one point in space and what happens elsewhere (1988, 11). Since county, tract, and other jurisdictional boundaries are arbitrarily drawn, it is likely that observations from one point on a map are similar to those of nearby points, violating classical regression assumptions. Certainly for geographically arranged data, where the observations are determined by boundaries that do not truly separate observations from one another, it is safer to assume that observations are related to one another than to assume that all observations are independent.

One of the most difficult areas of spatial data analysis is to determine the precise nature of the relationship among observations arranged in space (Haining 1990, 341–42). In time-series analysis, it is straightforward to lag a variable by one time period or more, and econometric theory often provides clear direction for specification. But in spatial analysis some units in the geographic vicinity of a given observation may have more influence than others (Anselin 1988, 16–17). It is a safe assumption that closer neighbors make more of a difference than those that are far away, but a given county may be more closely related to its close neighbors to the north and west than to equally contiguous neighbors to the south and east. This could be the case, for instance, if highways ran to the north and west but not in the other two directions.

After considering the complications involved in attempting to specify the precise nature of spatial dependency for every observation across seven states, I decided on a simpler distance criteria for the analysis in this book.

Ideally one should run models using a variety of different specifications of spatial dependency. Space constraints, and the desire to write a book about the political consequences of population mobility, rather than one about methodology, prevented me from reporting models using other weighting schemes. The reader should note, however, that different weighting schemes will produce different regression results than the ones reported here.

Taking each state separately, I calculated the arc (great circle) distance between county centroids for all observations using the statistical software SPACESTAT. This produced an $N \times N$ matrix specifying the distance between each observation (county) in each state. I then converted the distance matrix to a binary contiguity matrix (1,0) based on the minimum distance necessary to link each observation to at least one other observation. For example, if the minimum distance necessary to establish that every California county has at least one neighboring county is 77.2 miles, then this was the distance criteria used to calculate the number of neighbors for each county in the state. An especially small California county could have many neighboring counties whose centroids were within 77.2 miles, and each of these counties are scored a "1," as neighbors. In adopting this method of defining the values of the spatial weights matrix, then, I assumed that no county in any of the states was totally isolated from, or independent of, all other counties in the state. Counties that are small in land area obviously have more neighbors than those that are large in land area. The conversion of the distance matrix to a binary contiguity matrix results in a $N \times N$ matrix, W, of ones and zeros, with ones indicating linkages to the most proximate observation(s), and zeros indicating no such linkage (see table B1.1).

The spatial weight matrix, once constructed, is then standardized such that the row elements sum to one. This facilitates later interpretation of the coefficients in spatial regression analysis (Anselin 1988, 23).

The goal of creating the spatial weight matrix in my analysis is to calculate spatially lagged dependent variables. Once the spatial weights matrix of ones and zeros is generated, it can be used with a program such as GAUSS or SPACESTAT to construct a spatial lag for any variable. Briefly, the spatial lag is a weighted sum of the observations adjacent to a given observation (those nearby observations given a value of 1) for a given variable. The terms of the sum are obtained by multiplying the dependent variable by the associated weight in the spatial weights matrix (Anselin 1988).

The resulting regression model, in matrix form, appears as follows:

$$y = \rho W_1 y + X + X\beta + \varepsilon$$
$$\varepsilon = \lambda W_2 \varepsilon + \upsilon$$

where y is an $N \times 1$ vector of observations on the dependent variable, $W_1 y$ is the spatially lagged dependent variable, and X symbolizes an $N \times K$ matrix of other explanatory variables. β is a vector of K regression coefficients, ρ is the spatial autoregressive coefficient, and ε is the random disturbance term. ε, in turn, is a function of the errors of adjacent observations ($W_2\varepsilon$) plus random error, υ.

Spatial autocorrelation of the type I focus on here is the condition where the dependent variable at each location is correlated with observations on the dependent variable at other locations:

$$E(y_i, y_j) \neq 0.$$

If there is no spatial autocorrelation, $\rho = 0$ and OLS estimation will be unbiased. If spatial autocorrelation is present, however, $\rho <$ or > 0 and OLS estimation will be biased in the same manner as if one had omitted an important explanatory variable.

By including the spatially lagged dependent variable as an explanatory variable, I can directly test the degree of spatial dependence while controlling for the effects of other explanatory variables. This appendix was designed to provide only an intuitive and general grasp of the method I used to specify the weights matrix from which I calculated the spatially lagged dependent variables in my regression models. Replications using alternative weighting schemes would be most welcome. For additional detail on spatial regression analysis, the reader may want to consult Anselin 1988, Haining 1990, or Cressie 1993.

Characteristics of the Weights Matrices for States

In addition to providing useful mathematical and data analytic functions for spatial analysis, SPACESTAT commands allow researchers to examine characteristics of the weights matrices used in computing spatial lags. The characteristics of the weights matrices used in the regression models for each state appear in table B1.1. One of the more helpful figures is the average number of links for counties in each state appearing in the last column. Based on the distances used to compute the weights used in my analysis, it is no surprise that California has the largest average number of links. California has some of the smallest and largest counties in terms of land area, so the average distance between county centroids will be sufficiently large that some counties will have many "influential neighbors" while others will have only one. Yolo County, for example, in northern California, had 19 observations to which it was linked, while much larger counties, including Inyo and Modoc, had only 1. New York is similar in that the state contains very small and large counties. The average New York county has 6.9 linked observations or "influential neighbors." Tompkins County, in upstate New York, was linked in my analysis to eleven nearby neighbors. Suffolk County, on the eastern end of Long Island, had only one link, Nassau.

Computation of Moran's I Statistic for Spatial Autocorrelation

For the state maps I present in the text for various variables, I have included a measure of spatial autocorrelation present in the observations. This measure of spatial autocorrelation is known as the "I statistic" or as "Moran's I" (Moran 1948; Cliff and Ord 1981; Anselin 1988). Moran's I can be applied to test both variables and regression residuals for the presence of spatial autocorrelation. Moran's I is defined as:

TABLE B1.1. Characteristics of Spatial Weights Matrices for Each State

State	Dimension of Matrix	Nonzero Links	% Nonzero Weights	Average Weight	Average Number of Links
California	58	464	14.0	.13	8.0
Colorado	63	204	5.2	.31	3.2
Kansas	105	388	3.6	.27	3.7
Kentucky	120	462	3.2	.26	3.9
Florida	67	212	4.8	.32	3.2
Pennsylvania	67	228	5.2	.29	3.4
New York	62	428	11.3	.14	6.9

Source: Figures based on weights matrices computed in SPACESTAT for spatial data analysis in chapters 2 through 8.

$$I = [N/S] \times \{[e'We]/e'e\}$$

where e is a variable or a vector of OLS residuals, W is a spatial weight matrix whose (i, j)th element is either 1 or 0, N is the number of observations and S is a standardization factor, equal to the sum of all elements in the weight matrix (Anselin 1988, 101). For a row standardized weight matrix, Moran's I can be expressed as:

$$I = e'We/e'e.$$

Because the mathematics of calculating Moran's I is similar to that of calculating the correlation coefficient (r), the values it takes on ranges from –1 to +1. Values of Moran's I that approach +1 indicate positive spatial autocorrelation—where similar values are clustering in spatially adjacent areas on a map. The generally rarer condition of negative autocorrelation is present when values approach –1. Here dissimilar values are clustering together. When $I = 0$, the values of the variable (or residuals) are randomly scattered, indicating no spatial autocorrelation.