The Spatial Distribution of Changes in County-Level Racial/Ethnic Diversity in the United States, 1980-2000: Using ESDA, Spatial Regression, and GWR

Lauren A. Hughes, The Pennsylvania State University

BACKGROUND. In recent years the United States has experienced a national trend toward greater racial/ethnic diversity, however, little is known about the spatial dynamics of racial/ethnic diversity at the subnational level. Studying the spatial patterns of racial/ethnic diversity and changes in racial/ethnic diversity at the county level has potential implications for a wide range of topics. Specifically, better understanding the spatial patterns of change in racial/ethnic diversity could add to the immigration literature, the diversity literature, and to theories on migration and residential mobility.

SPECIFIC AIMS. This project will investigate the spatial aspects of changes in racial/ethnic diversity, focusing on U.S. county-level data from 1980 to 2000. It will consider common correlates of racial/ethnic diversity, such as poverty, immigration, and population size, as well as contextual variables such as region, county type and county functional specialization. I will utilize exploratory spatial data analysis (ESDA) to examine the spatial distribution of changes in racial/ethnic diversity and will employ OLS and spatial regression models to explore the relationship between changes in a county's racial/ethnic diversity and a range of social and contextual variables. Specifically, this study seeks to answer the following questions:

- Q1. How do change in percent poverty, change in percent foreign born, and change in population size in a county correspond to the change in racial/ethnic diversity of the county?
- Q2. How do the region, county type, and functional specialization of a county correlate with the change in racial/ethnic diversity of the county?
- Q3. In what ways is the change in racial/ethnic diversity of a county associated with the change in racial/ethnic diversity of neighboring counties?

The county unit allows an examination of changes in racial/ethnic diversity across the entire continental United States. This study will employ geographically weighted regression (GWR) to explore the possibility that the relationships between change in racial/ethnic diversity and the aforementioned variables might vary across the nation, which leads to the final research question:

- Q4. Are the previous associations constant across the U.S. or are they non-stationary?

DATA. The data for the study come from the U.S. Decennial Census for 1980, 1990 and 2000. The dependent variables are measures of the raw change (difference score) in racial/ethnic diversity within U.S. counties over three periods of time: 1980-1990, 1990-2000, and 1980-2000. Diversity is quantified using a standardized entropy score for the county, using five exhaustive and mutually exclusive racial/ethnic groups: Hispanics, non-Hispanic whites, non-Hispanic blacks, non-Hispanic Asians, and other. The study will also include three common correlates of racial/ethnic diversity: poverty, immigration, and population. Change in percent poverty is the raw change in the percentage of families within the county that fell below the poverty level during the given time period. Similarly, change in percent foreign born is the raw change, during a given time period, in the percentage of residents within the county that were born outside of the United States. It is also possible that the spatial variation in changing patterns of racial/ethnic diversity could be attributed to regional historic, cultural, or climatic differences throughout the country. For example, the historically large black population in Southern states is likely to influence the racial/ethnic diversity of Southern counties, as well as those counties' patterns of change. In order to test for regional effects, I will include a set of dummy variables that correspond to census regions in my analyses. In their study of changes in racial residential segregation, Farley and Frey outline six functional specialties of metropolitan areas that help explain the variation in segregation across the country: retirement communities, durable goods communities, nondurable goods communities, government communities, university communities, and military communities (1994). It is possible that these functional specialties are correlated with changes in racial/ethnic diversity as well. For example, increases in racial/ethnic diversity are likely to be seen in counties with large government, university, or military presences, due to these institutions' dedication to racial/ethnic equality and active minority recruitment. In order to evaluate whether these functional specialties can explain changes in racial/ethnic diversity at the county level, I will include the same six categories as dummy variables in my models. Finally, racial/ethnic diversity has been linked with county type, with higher levels of diversity occurring in metropolitan areas than in micropolitan or rural areas. In order to determine whether differences in county type can help explain changes in racial/ethnic diversity, I will include dummy variables for metropolitan and micropolitan statistical areas in the models. Counties are classified as metropolitan if they contain a metropolitan statistical area, micropolitan if they include a micropolitan, but not a metropolitan, statistical area, and rural if they contain neither a metropolitan nor a micropolitan statistical area.

METHODS AND PRELIMINARY FINDINGS. For this study I will utilize a variety of ESDA techniques to visually examine the spatial patterning of changes in racial/ethnic diversity. The next step in the analysis will be a comparison of a basic, non-spatial, OLS regression model to spatial lag and spatial error models. Finally, I will employ geographically weighted regression (GWR) to explore the possibility that the relationships that I've found using OLS and spatial regression vary across the country and, thus, should not be represented using a global model. The results of preliminary analyses, focusing on changes in racial/ethnic diversity from 1990 to 2000 are presented in tables 1-3. In these analyses I used a four-group measure of racial/ethnic diversity, which only includes the four largest racial/ethnic groups in the United States: whites, blacks, Hispanics, and Asians. While I will use the five-group measure of racial/ethnic diversity described above in future analyses, in order to capture the entire populations of each county, sensitivity testing suggests that I will find similar results.

The results of the OLS regression of change in diversity on change in percent poverty, change in percent foreign born, change in the log of population, region, functional specialty, and county type are presented under Model 1 in Table 1. Model 1 explains 39.6% of the variance in the change in racial/ethnic diversity within counties between 1990 and 2000, $F(15, 3089) = 144.49$, $p<.001$. Percent poverty, percent foreign born, and population have positive, statistically significant relationships with change in diversity. Change in percent foreign born's large b-coefficient, relative to its standard deviation, reveals that it has the largest effect on change in racial/ethnic diversity. This makes sense because as non-white immigrants enter majority white counties they add to the racial/ethnic diversity of the county. The region variables are also strongly statistically significant, indicating that region explains part of the variation in levels of change in racial/ethnic diversity and that the Northeast, Midwest, and West are significantly different from the South in their relationships with changes in diversity. Only three of the functional specialties are statistically significant at the $p<.05$ level: nondurable goods manufacturing, university, and military. Neither of the variables for county type is statistically significant in this model.

The statistically significant results of the Lagrange multiplier tests, included at the bottom of Table 1, indicate the presence of autocorrelation and that the OLS model is, thus, misspecified. Since the Robust form of the Lagrange multiplier error is statistically significant at a higher order of magnitude than the Robust form of the Lagrange multiplier lag, a spatial error model appears to be the best choice (Anselin 2005). The results of the spatial error model are included under Model 3 in Table 1. Model 3 explains 44.7% of the variance in the change in racial/ethnic diversity within counties, an improvement to Model 1. All of the variables that are statistically significant in the OLS model, except nondurable goods manufacturing, are also significant in the spatial error model and operate in the same direction. The spatial error term that has been added to the model is also significant and with the inclusion of this term the metropolitan dummy variable is now statistically significant and university has increased in significance. Comparing the Log Likelihood from Model 1 to Model 3, there is some improvement. The Log Likelihood increases from -8577.4 in Model 1 to -8476.70in Model 3, indicating a somewhat better fit. In addition, the AIC decreases from 17185 in Model 1 to 16983 in Model 3, adding further support that the spatial error model fits the data better than non-spatial OLS model. For the sake of comparison, the results of the spatial lag model are included under Model 2 in Table 1. The results confirm my decision to proceed with the spatial error model. Compared to Model 3, Model 2 has a smaller explained variance, a smaller Log Likelihood, and a larger AIC, all of which indicate a worse model fit.

GWR accounts for spatial nonstationarity by fitting a regression model at each observation location (Wheeler and Paez 1994). In order to estimate the models, a local weights matrix is constructed from a kernel function, which assigns greater weight to the observations that are located closer to the calibration location (Wheeler and Paez 1994, Mennis and Jordan 2005). Since my unit of analysis, the county, varies so much in size, a fixed bandwidth presents a potential problem. In order to combat this issue, I utilized an adaptive spatial kernel function, the corrected Akaike Information Criterion (AIC). The AIC minimization process converged at a local sample size of 160, accordingly, each estimation was influenced by the 160 closest observations and all other points were weighted zero. By estimating a series of local regression models, GWR allows for nonstationarity and the representation of varying relationships between attributes across space. Thus, the inclusion of region variables in the model would be redundant and problematic. In addition, binary variables in general can present problems in GWR. For example, most of the functional specialization variables only represent a small proportion of the counties, making it likely that there are local areas with little or no variation in a specific functional specialization. Without variation, the local models cannot be estimated. For this reason, my GWR model (Model 4) only includes the continuous variables, change in percent poverty, change in percent foreign born and change in the log of population.

The GWR estimation fits the model better than the global estimation. This is evident from its reduced AIC score as well as its increased adjusted r-square score. Table 3 includes a five number summary of the parameter estimates for Model 4 as well as the results from the Monte Carlo significance test for spatial variability. A significant result from the Monte Carlo test indicates that the parameter is non-stationary. All of the parameters, including the intercept, were statistically significant at the $p<.001$ level, indicating non-stationarity. Glancing at the summary of the parameter estimates supports this finding; the direction of the relationship with change in racial/ethnic diversity changes across local estimates for each parameter. The local r-square values are visually depicted in Figure 1. The map illustrates that locally this model ranges from explaining as little as four percent to as much as ninety percent of the variation in the change in racial/ethnic diversity.

These preliminary analyses provide evidence that there is a spatial element to the distribution of the change in racial/ethnic diversity scores within counties across the United States. In other words, across the country, the change in racial/ethnic diversity that a county experienced between 1990 and 2000 was positively related to the change in racial/ethnic diversity experienced by its neighbors. However, this relationship was particularly strong for certain clusters across the country. These differences were partially explained by the predictor variables. Change in percent poverty, change in percent foreign born, change in the log of population, and the region variables had statistically significant positive relationships with change in racial/ethnic diversity using spatial regression. However, by utilizing GWR and counties as units of analysis, this study was able to illustrate that the relationships between change in racial/ethnic diversity and change in percent poverty, change in percent foreign born, and change in the log of population are non-stationary. Figure 1 illustrates the wide range of explained variance at the local level and highlights the danger of limiting research on racial/ethnic diversity to global models. By providing new insights on the local nature of these spatial patterns, this study aims to add to the immigration literature, the diversity literature, and to theories on migration and residential mobility. These preliminary findings are examples of the results I expect to find in this study. However, in the next steps of this analysis I will explore a larger time frame, using a broad range of spatial tools to understand and illustrate the intricacies of the spatial distribution of changes in racial/ethnic diversity.

Figure 1: Local R-Square Estimates (Model 4)

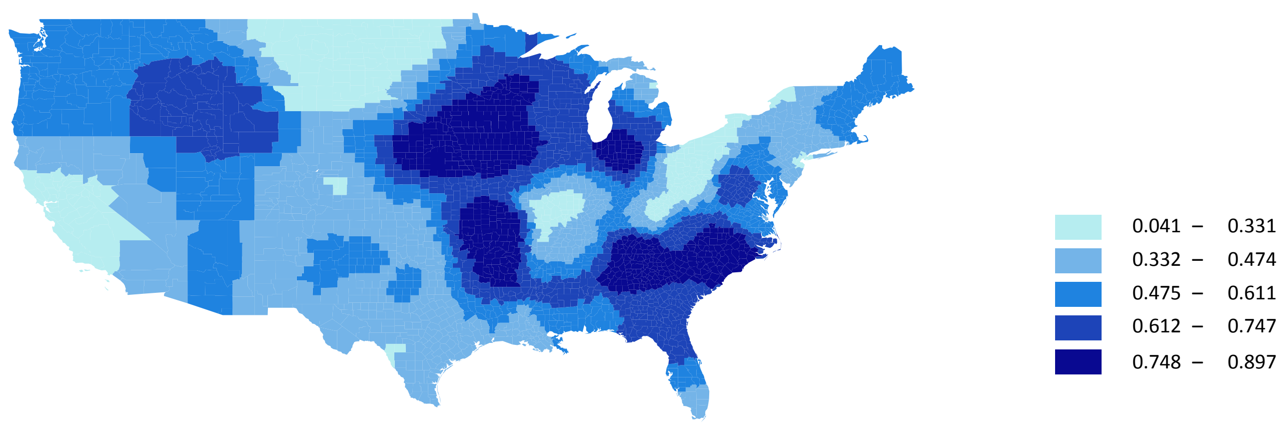

Table 1: Change in Racial/Ethnic Diversity from 1990-2000: OLS, Spatial Lag, and Spatial Error Models						
	Model 1 (OLS)		Model 2 (lag)		Model 3 (error)	
Independent Variable	Coeff.	Std. Err.	Coeff.	Std. Err.	Coeff.	Std. Err.
Constant	2.163***	.201	1.325***	.206	2.085***	.232
Change in % Poverty	10.747***	2.715	7.749**	2.641	6.943*	2.746
Change in % Foreign Born	140.187***	3.798	126.908***	3.940	141.008***	4.003
Change in Log Population	5.424***	.615	4.635***	.605	5.951***	.660
Region						
Northeast	1.001***	.301	.847**	.293	.980*	.424
Midwest	1.045***	.174	.956***	.170	1.042***	.240
West	-1.069***	.235	-1.025***	.228	-1.085***	.328
Functional Specialty						
Retirement	.302	.215	.095	.209	.065	.221
Durable Goods	-.108	.202	-.133	.196	-.213	.212
Nondurable Goods	.462*	.216	.335	.210	.224	.224
Government	.161	.217	.207	.210	.314	.216
University	.456*	.211	.440*	.205	.619**	.205
Military	-.977*	.383	-.914*	.372	-.786*	.371
Type						
Metropolitan	-.293	.190	-.432*	.185	-.426*	.195
Micropolitan	-.122	.192	-.124	.186	-.149	.186
Spatial Lag (Rho)			0.254***	0.022		
Spatial Error (Lambda)					0.362***	0.025
R^2	0.396		0.428		0.447	
F-statistic	144.493***					
Log Likelihood	-8577.4		-8510.8		-8476.7	
AIC	17185		17054		16983	
Lagrange Multiplier (error)	247.109***					
Robust LM (error)	92.005***					
Lagrange Multiplier (lag)	155.552***					
Robust LM (lag)	.449					

*p<0.05; **p<0.01; ***p<0.001

Table 2: Global and Local GWR Models (Model 4)		
	Global	GWR (local)
Residual Sum of Squares	47324.520	28584.031
Effective # of Parameters	4.000	191.829
Sigma	3.907	3.132
AIC	17275.132	16111.481
Coef. of Determination	.374	.622
Adjusted R-Square	.373	.597

Table 3: GWR Parameter Summary Results (Model 4)						
	Min	Lower Quartile	Median	Upper Quartile	Max	Monte Carlo
Intercept	-2.045	1.407	2.114	2.954	6.608	Non-stationary***
Change in % Poverty	-141.354	-3.944	6.885	22.227	94.546	Non-stationary***
Change in % Foreign Born	-41.390	116.269	223.071	271.960	409.507	Non-stationary***
Change in Log Population	-8.218	0.556	3.835	6.838	31.792	Non-stationary***

*p<0.05; **p<0.01; ***p<0.001

References

Anselin, Luc. 2005. “Exploring Spatial Data with GeoDa: A Workbook.” Retrieved March 2, 2011 (http://www2.geog.ucl.ac.uk/~jessletzbichler/GEOG2003/geodaworkbook.pdf).

Farley, Reynolds and William H. Frey. 1994. “Changes in the segregation of whites from blacks during the 1980s: small steps toward a more integrated society.” American Sociological Review 59:23-45.

Mennis, Jeremy L. and Lisa Jordan. 2005. “The Distribution of Environmental Equity: Exploring Spatial Nonstionarity in Multivariate Models of Air Toxic Releases.” Annals of the Association of American Geographers 95(2):249-268.

Wheeler, David C. and Antonio Paez. 2010. “Geographically Weighted Regression.” Pp. 462-486 in Handbook of Applied Spatial Analysis, edited by Manfred M. Fischer and Arthur Getis. New York: Springer.