An Examination of the Spatial Distribution of Immigrant Residential Segregation

Weiwei Zhang
Department of Sociology
Brown University
Providence, RI 02904
Tel: (401) 863-6323
Email: weiwei_zhang@brown.edu

Prepared for presentation at the 2008 Annual Meeting of the Population Association of America

INTRODUCTION

The unprecedented migration into the United States started about four decades ago has been of increasing interest to scholars of contemporary social science. Research has focused on migration motivation, adaptation/assimilation, and consequences of migration on sending and receiving countries. To understand immigrants' adaptation and its consequential changes to U.S. metropolitan areas, studies on immigrant residential settlement patterns have also thrived. Previous literature has shown that immigrants are more likely drawn to ethnic neighborhoods because of shared social and economic resources that are unavailable elsewhere (Massey and Denton, 1987). Residential duration, language skill, and socioeconomic status are revealed to be important predictors of immigrant residential pattern (Iceland, 2004; Logan et al., 2004; Massey and Denton, 1987). Generally, recent immigrants are found to be more likely living in neighborhoods of the same race/ethnicity; immigrants of better language skills and higher socioeconomic status tend to have greater residential mobility than their counterparts.

At this point, researchers have focused on the segregation at the individual level. It is also reasonable to speculate that the spatial distribution of segregation at an upper-level may also present certain patterns. In other words, immigrant neighborhoods may cluster across space. However, there are few studies looking at the spatial distribution of immigrant residential segregation. With the utilization of the techniques of exploratory spatial data analysis and spatial regression models, the current study intends to examine the spatial distribution of immigrant residential segregation by answering the following research questions: 1) are immigrant neighborhoods randomly distributed across space? 2) if not, will the predictors for residential segregations of individual immigrants also predict the non-randomly spatial arrangement of the residential segregation clustering and; 3) how?

THEORETICAL FRAMEWORK AND SPATIAL MODELS

Spatial Statistical Analysis

The first step for a spatial analysis is to examine the spatial autocorrelation, which refers to the non-randomly spatial arrangement of a certain outcome. A single summary statistic, Global Moran's I, is used to measure the direction and magnitude of spatial autocorrelation. A significant positive value of Moran's I indicates that the value of the dependent variable for location *i* is positively related to the weighted average value of the dependent variable for all neighboring locations.

If spatial autocorrelation is detected, spatial analysis proceeds to examine local clustering of the outcome variable using the Moran's scatterplot map. Potential clusters of similar values can be observed from the map. In addition to local clusters, the map also demonstrates distinctive regions, if they exist, in which local clustering presents different patterns. In such cases, we need to consider spatial heterogeneity of the large regions before I move further to assess the local interactions using multivariate methods. Spatial effects from the heterogeneity between large regions could confound our analysis of a local process happening in each region.

The above two steps belong to the exploratory phase of spatial data analysis, which is then followed by an ordinary least-squares (OLS) regression model. At the multivariate stage, independent variables are incorporated. The equation is expressed as following:

$$Y_i = \Sigma_k X_{ki} \beta_k + \varepsilon_i ,$$

where *i* is the location and X is the *k*th independent variable. If the variation in the outcome variable is fully accounted for by the independent variables, there should be no residual spatial autocorrelation. The relationship between the outcome of location *i* and the outcome of neighboring locations is not significant. The examination of residual spatial autocorrelation is conducted based on the residuals of the OLS regression model.

Assuming spatial autocorrelation remains controlling for independent variables, the next step is to select a spatial dependence model with proper specifications. Generally, we can choose between a spatial lag model and a spatial error model. In a spatial lag model, a weighted average value of the dependent variable for the neighboring locations is introduced as an additional covariate:

$$Y = \rho WY + X\beta + u\ ,$$

where ρ is the spatial autoregressive parameter and *W* is the spatial weights matrix. This equation can be represented as followings as well:

$$Y(1-\rho W) = X\beta + u\ ,$$

$$Y = (1-\rho W)^{-1} X\beta + (1-\rho W)^{-1} u\ .$$

The equation above illustrates that the value of *Y* at each location is not only determined by *X*s at that location, but also by the *X*s at all other locations through the spatial multiplier $(1\text{-}\rho W)^{-1}$. Spatial dependence in a spatial lag model is suggestive of a possible diffusion process – events in a location increase the likelihood of similar events to occur in neighboring locations.

In a spatial error model, spatial dependence is incorporated in an autoregressive error term, which is indicative of omitted covariates:

$$\varepsilon = \lambda W\varepsilon + u\ ,$$

So, the full equation becomes

$$Y = X\beta + (1-\lambda W)^{-1} u,$$

where notations are the same as in the spatial lag model. It indicates that the value of the outcome variable for each location is affected by errors of all locations through the spatial multiplier $(1\text{-}\lambda W)^{-1}$.

In addition to spatial dependence, spatial heterogeneity is also considered in this paper. As noted, larger regions may present different spatial patterns. Hence, we need to disentangle spatial effects due to differences between regions and spatial effects that are specifically due to spatial dependence between nearby locations. Generally, the regions are defined as spatial regimes. In different regimes, the spatial pattern of the outcome variable may vary and the effects of the independent variables may also change. Statistical tests of the overall stability and the stability of individual regression coefficients are needed for the assessment of spatial regimes.

Residential Segregation Clustering

With regard to the clustering of immigrant residential segregation, we can think of two scenarios. First, a group of geographic locations may share some common external forces from an upper-level structure. For instance, some states may have favorable social and economic structures and policies that pull foreign-born populations to certain states. Therefore, the demonstration of residential segregation clustering is partially a result of some common exogenous structural factors that are not included in the county level data.

Second, the clusters may come from a process of population flows from a location to neighboring locations. Immigrants may initially be drawn to an "ideal" location. Gradually, due to accumulated growth, population expands from the first-stop location to the neighboring locations. This process of population flows indicates a "spill-over" effect of population growth, which is often seen in developing countries. However, unlike the case of developing countries in which only several cities serve as the centers, the "spill-over" effect may be limited in explaining immigration residential segregation in the US because more than a few locations act as "ideal" locations for immigrants.

Considering these two scenarios, I hypothesize that either a spatial lag model or a spatial error model has the explanatory power for the potential clustering of immigrant residential segregation. The probability of being a fitted model may be greater for a spatial lag model than a spatial error model in regions where high residential segregation is surrounded by high residential segregation. In addition to the speculation about spatial dependence models, I also derive several other hypotheses corresponding to each stage in the spatial analysis.

HYPOTHESES

1. Immigrant residential segregation will exhibit statistically significant and positive spatial autocorrelation, suggesting segregation clusters in space.
2. Assuming that spatial randomness is rejected, the traditional OLS regression model with a set of independent variables will be insufficient to explain the variation in the spatial clustering of immigrant residential segregation.
3. Assuming residual spatial autocorrelation is observed, a spatial lag model or a spatial error model will be used to explain the residual spatial autocorrelation of segregation accounting for the spatial heterogeneity.

DATA

Data for this analysis come from the 2000 Summary Tape Files 3A (U.S. Bureau of the Census 2000), which provides information of social, economic, and housing characteristics compiled from a sample of approximately 19 million housing units.[1] Advantages of using the Summary files 3 rely on the detailed information on geographic organizational units and subgroups by characteristics, such as nativity and language ability. The geographic unit of analysis is US County because: 1) it is a politically and economically meaningful unit in space; 2)

[1] Further descriptions of Summary File 3 can be found on the website of U.S. Census Bureau. Http://www.census.gov/press-release/www/2002/sumfile3.html

it is a common-used geographic unit between state and tract, another two meaningful units; 3) it is a good choice for studying population flows. From the Summary Tape Files, I extract data of both county and tract levels. County level data are used for structural variables of the county level. Data of both levels are used for the calculation of the dependent variable measuring immigrant residential segregation.

METHOD

Dependent Variable

I use the Entropy method to assess segregation (Duncan and Duncan, 1955). The basic logic of constructing the Entropy values is to assess the level of segregation using two levels of geographic units. Computing steps can be illustrated by the following equations:

$$H^* = (-1)\sum_{k=1}^{K} P_k \log(P_k)$$

$$\overline{H} = (-1)\sum_{i=1}^{I} \frac{w_i}{W} \sum_{k=1}^{K} P_{ik} \log(P_{ik})$$

$$H = \frac{H^* - \overline{H}}{H^*}$$

where we have *i* tracts and *k* groups, *P* denotes the proportion of a particular group in total county population, *W* denotes total population of the county. Hence, the Entropy value (*H*) can be regarded as the proportional reduction in error (PRE) for measuring diversities at the county level by using data on distribution patterns at the tract level. In other words, it is "the weighted average deviation of each category's diversity from the total diversity, standardized by the total diversity (White and Kim, 2004)." White (1986) explicitly describes the properties of the entropy index as followings:

The entropy index varies between 0, when each parcel has the same composition as the city, so knowledge of parcel sheds no light on population composition, and 1, when each tract contains one group only.

Independent Variables

Drawing on previous literature on immigrant residential segregation, I construct five structural predictors: county total population, percent of foreign born population, percent of recent foreign born population, percent of naturalized foreign born population, and percent of foreign born who speak only English at home. The first two variables are transformed into the log form due to the highly skewed distribution of the original values. In particular, percent of recent foreign born population and percent of foreign born who speak only English at home are expected to be positively associated with the segregation. Percent of foreign born population and percent of naturalized foreign born population are expected to be negatively associated with the segregation. In addition to these structural predictors, I also include a geographic variable indicating US divisions defined by Census Bureau. [2]

Weights Matrix

Neighboring counties may share broader structural factors that are missing in the county-level model and explain a partial force of the spatial effect. Immigrants are drawn to a region because of certain socio-economic characteristics of that place. These certain features may come from a broader geographic sphere and are shared by counties belonging to this sphere. This type of residential segregation clustering may be captured by weights matrixes that define neighbors based on the unit of analysis, i.e., county in this study. On the other hand, for high segregation surrounded by high segregation, the intrinsic expansiveness of heavily concentrated population at

[2] Census classifications of divisions: New England, Middle Atlantic, East North central, West North central, South Atlantic, East South central, West South central, Mountain, and Pacific.

a particular county may demonstrate a spill-over effect. This special type of clusters may be captured by weights matrixes that define neighbors based on both distance and shared structural factors, because the expansion of population is significantly determined by the effect of distance, which then could be mediated by transportation, for instance, the availability of inner and inter-county highways.

Based on these criteria, I first exclude the distance-based weight matrix due to its ineffectiveness of capturing the essence of shared broader structural factors. For example, suppose we have a scenario in which county A, one of the neighbors of a hypothetical central county, is so large that its centroid locates beyond the defined distance from the central county. By definition, the distance-based weight matrix will not count county A into the neighborhood of the hypothetical county regardless of the high likelihood that county A shares broader structural factors with the hypothetical county. I then rule out the *K*-nearest neighbor weight matrixes after examining the matrix record files. What I find in the files shows that neighborhoods are forced into creation with a certain number of counties depending on how many nearest neighbors I define earlier in the program. The serious problem is the lack of primary attention on shared structural factors and distance, which are the key determinants of spatial effects in my study on spatial distribution of the Entropy values.[3] Eventually, I decide to adopt the first-order queen weights matrix for further examination.

Analytical Procedures

The first step of my spatial analyses is an exploratory analysis of the spatial clustering of county-level Entropy values. It consists of the examination of the global and local patterns of

[3] For example, I choose the four-nearest neighbor weights matrix. The selected four neighboring counties for county 114 (the unique ID of counties in my study) are county 131 with a distance of 43,724, county 148 with a distance of 95,943, county 164 with a distance of 98,180, and county 178 with a distance of 119,338. Compared to neighbor 131, neighbor 178 is about three times further away from county 114. The distance here is certainly not the main concern using this matrix.

spatial autocorrelation in the Entropy values. Global autocorrelation is assessed by means of Moran's I statistic, with a positive and significant value indicating clustering in space of similar Entropy values. Local spatial autocorrelation is assessed by means of a local Moran statistic, which indicates to what extent the Entropy value each location is significantly correlated with the values at neighboring counties. Local clustering of high (a high Entropy value surrounded by high values) or low (a low Entropy value surrounded by low values) values reject the spatial randomness of the arrangement of Entropy values. Additionally, maps of local clustering are indicative of distinctive spatial regimes. The non-randomly spatial arrangements are presented in different formats across regimes.

The exploratory spatial data analysis is then followed by an OLS regression of county-level Entropy values on the independent variables, including total county population, percent of the foreign born population, percent of recent immigrants, percent of naturalized immigrants, and percent of immigrants who speak only English at home. In addition to the traditional statistics, the OLS results also report spatial diagnostics that can be used for examining residual spatial autocorrelation in Entropy values and specifying a spatial dependence model. However, these diagnostics are not independent from the confounding effects of spatial heterogeneity in a larger scale. If no distinctive spatial regimes are observed from the exploratory phase, I would move forward to spatial models with the initial diagnostic statistics. Otherwise, I need to consider disentangling the spatial effects from the heterogeneity beyond the level of local clustering.

I will conduct formal tests on whether the variations between different regimes are significant. The tests could be done by examining the overall stability of the regression coefficients and the stability of individual regression coefficients. From the latter one, individual variables that affect the Entropy values differently across regimes can be detected. In order to carrying out such tests on spatial regimes, I construct a geographic variable that will be used as a classification of spatial regimes for the tests. If the tests of spatial regimes are not significant, spatial diagnostics from the original OLS model will be used for the specification of spatial models. Otherwise, the original sample needs to be divided into two sub-samples classified by the regime variable, and separate OLS models are needed for each sub-sample. Similarly, the OLS models will produce two sets of spatial diagnostics for two regimes, respectively.

Spatial diagnostics contain three groups of information. First, it reports residual Moran's I statistic of the Entropy values from the OLS model. Second, test results of a spatial lag and a spatial error model are included for comparison. The model of a significant statistic will be the choice of the final spatial model. If both are significant, the third group of information on robust tests will be helpful. The model of a significant statistic from a robust test will then be the choice of the final spatial model. Either a spatial lag model or a spatial error model reports individual coefficients for each predictor. In addition, a spatial lag model reports the coefficient of a lagged dependent variable, which is the weighted average of Entropy values of neighboring counties; a spatial lag model reports the coefficient of a spatial autoregressive error term, which is the effect of a vector of unmeasured variables.

RESULTS

Exploratory Spatial Data Analysis

I begin by examining the Global Moran's I statistic of the county-level Entropy values, which is shown in Figure 1. It is a single correlation coefficient between the Entropy value of a county and the weighted average Entropy values of its neighboring counties that are defined by the weights matrix. The coefficient is .3528, which is statistically significant at the .001 level.[4]

[4] The level is assessed based on a permutation approach with a 999 random permutation.

Figure 2 is the Moran scatterplot map of the Entropy values. There are four categories in the legend. However, only the "High-High" and the "Low-Low" clusters contribute the positive spatial autocorrelation of the Entropy values. The clustering of high Entropy values is mostly in the East (as indicated by the red color and the label "High-High"). The clustering of low Entropy values is found through out the West Central and parts of the mountains. From this map, I conclude that two most important regimes that are the West and the East should be incorporated into the multivariate analyses. A formal test of the overall stability of regression coefficients is conducted later in SpaceStat.

Figure 1 Global Moran's I Statistic of the County-Level Entropy Values

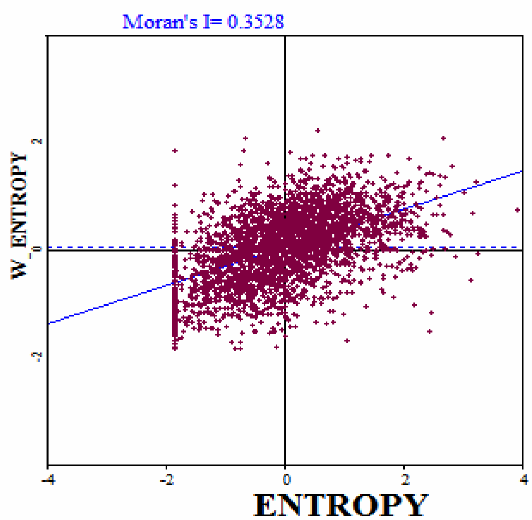
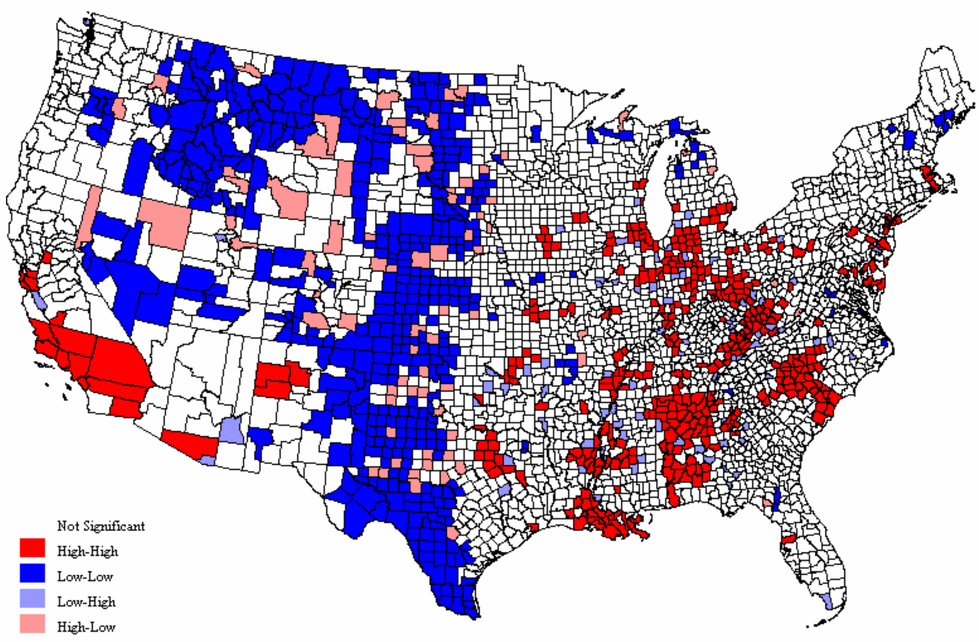


Figure 2 The Moran Scatterplot of the County-Level Entropy Values



The OLS Model[5]

Table 1 reports regression results from the OLS model. The total model fit is good, with a R^2 approximate of .51. All structural predictors are significant and relationships are as expected. The non-significant coefficient of the geographic variable "West" indicates that the average Entropy value of the West is not significantly different the average value of the East. However, it does not necessary mean that the patterns of local clustering in two regions are not significantly different from each other. The inclusion of interaction terms of the geographic variable and other structural predictors may be able to examine whether the effect of each individual predictor changes from West to East. Given a set of formal tests is available in

[5] Three measures on multicollinearity, non-normality, and heteroskedasticity are also reported in GeoDa following the results of an OLS regression. No serious multicollinearity was detected (multicollinearity condition number is 26.87). The Jarque-Bera test on normality of the errors is significant, indicating non-normality of the errors. For samples with comparatively large sizes, this may not be too serious a problem. Diagnostic tests on heteroskedasticity report significant results, suggesting non-constant variance in errors.

SpaceStat, I decide to address this concern of spatial heterogeneity later in a separate session.[6]

Table 1 Ordinary Least-Squares Regression of County Entropy Values

Variable	Coefficient	Std.Err.	t
(Constant)	-0.203	0.0071	-28.372 ***
Total population (logged)	0.023	0.0005	47.591 ***
% Foreign-born (logged)	-0.017	0.0007	-24.643 ***
% Recent foreign-born	0.032	0.0046	7.035 ***
% naturalized foreign-born	-0.037	0.0050	-7.336 ***
% Foreign-born who speaks only English at home	-0.018	0.0021	-8.814 ***
West	-0.002	0.0012	-1.331
R^2	0.51		

**p* <.05; ***p* < .01; ****p* < .001

Examination of Residuals from the OLS model

Using the ESDA techniques, I examined the residuals from the OLS model. The residual spatial autocorrelation is estimated as .1182, significant at the .001 level. It indicates that the OLS model partially explains the variation in the Entropy values.

[6] Actually, I used both methods and found that results are consistent, although the method using interaction terms produces coefficients less significant than the one generated by SpaceStat.

Figure 3 Residual Global Moran's I Statistics from the OLS Model

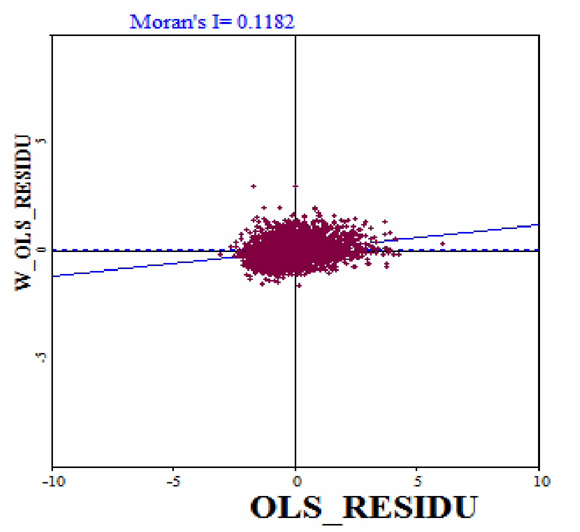


Figure 4 illustrates patterns of over- or under-prediction and the magnitude of the residuals from the OLS analysis. The three categories in blue tones are over-prediction and the other three in brown tones are under-prediction. The darkest blue and brown categories represent cases with the largest residuals and suggest a need to include other exogenous variables to explain the remaining variance in the Entropy values. Overall, the visualization of the standard deviational map presents obvious spatial patterns of over- or under-prediction. It also indicates potential spatial regimes that can be roughly generalized as the East and West (brown versus blue). The examination of the local clusters of residuals (Figure 5) reinforces the findings. Actually, three regions can be identified from the map: several clusters of low segregations (in blue) in West, some clusters of high segregations (in red) in the Mid-West region, and a cluster of low segregations in the North-East region. For this study, I only define two regimes: West and East.

Figure 4 Standard Deviation Map of Residuals from the OLS Model

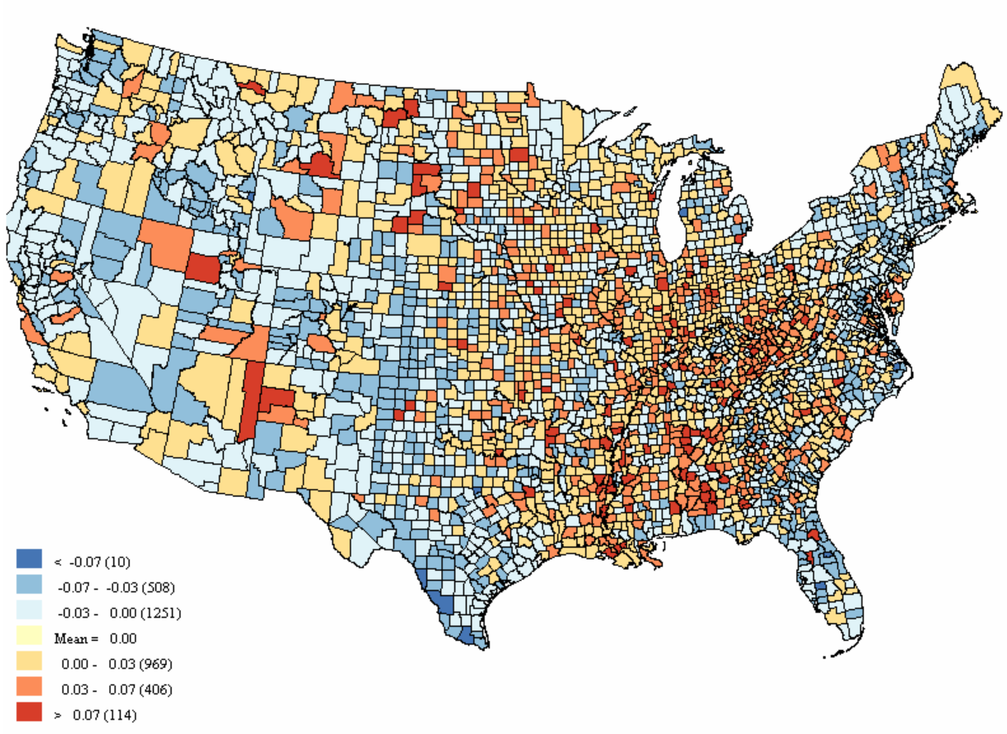
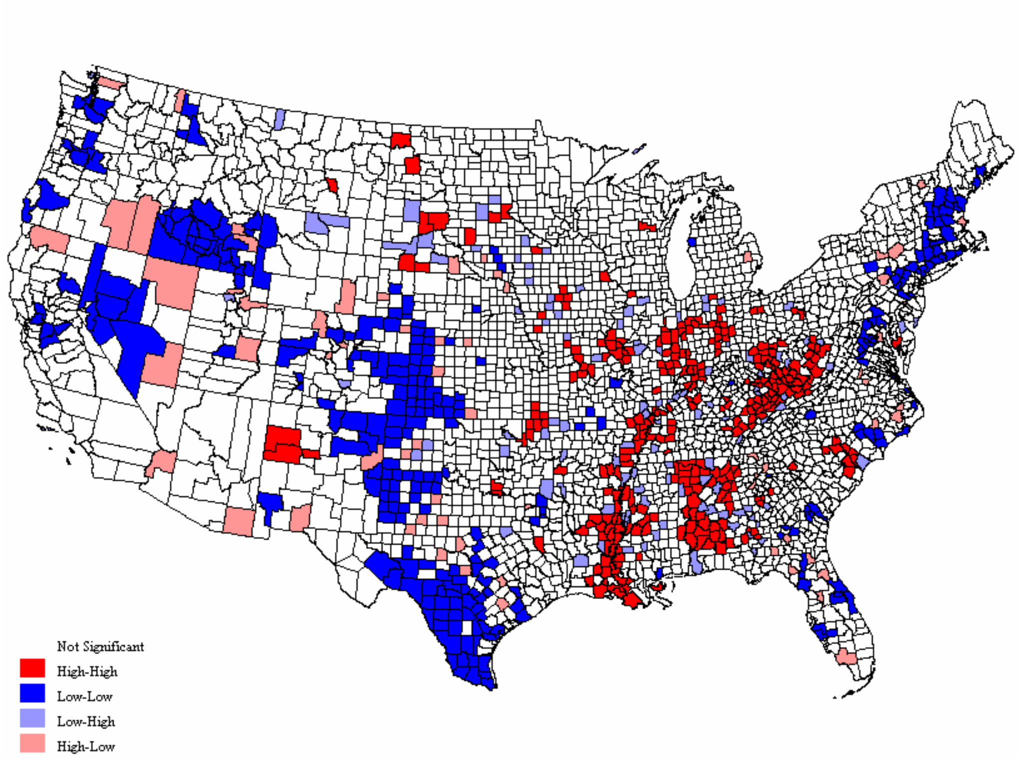


Figure 5 Local Cluster Map of Residuals from the OLS Model



Spatial Regimes

After defining spatial regimes based on the observation from the maps, I asses the null hypothesis of the overall stability of the regression model for different regimes. In Table 2, the Chow Wald test shows that the structural stability of the two regimes is rejected and the stability of two of the individual regression coefficients is also rejected (percent of foreign born and percent of naturalized foreign born). In this case, I divide the original sample into two sub-samples and construct multivariate analyses for two samples separately (Baller et al., 2001).

Table 2 Stability of Regression Coefficients by Spatial Regime (Groupswise Heteroskedastic Error Model)

Test on Structural Stability for 2 Regimes[a]		
Chow Wald	33.589	***
Test on Stability of Individual Coefficients [b]		
(Constant)	1.145	
Total population (logged)	0.455	
% Foreign-born (logged)	24.192	***
% Recent foreign-born	0.365	
% naturalized foreign-born	5.495	*
% Foreign-born who speaks only English at home	2.746	†

†P <.1; *p <.05; **p < .01; ***p < .001
a: distributed as χ^2 with 6 degrees of freedom
b: distributed as χ^2 with 1 degree of freedom

Spatial Models

I repeat the OLS regression for the two sub-samples. For each, a set of spatial diagnostic statistics are reported in Table 3. The basic logic for choosing between a spatial lag model and a spatial error model is to first compare the results of the standard Lagrange Multiplier tests (LML and LME). The one with significant results is the proper model. If the standard LM tests for a lag or an error model are both significant, I move to the Robust LM tests. In this way, I choose the spatial error model for the West (.0242 for robust LML vs. .0000 for robust LME) and the spatial lag model for the East (.0006 for robust LML vs. .0022 for robust LME).

In addition, Table 3 also reports Moran's I statistics controlling for independent variables and spatial heterogeneity. The Moran's I statistics for both regimes remain significant. It indicate that although the unevenness of the spatial distribution of the Entropy values has been explained partially by the set of independent variables, residual spatial autocorrelation exists. This finding is consistent with what is presented earlier using the full sample.

Table 3 Spatial Diagnostics

	West		East	
Test Statistics	**Value**	**Prob.**	**Value**	**Prob.**
Moran's I (error)	0.13	0.0000	0.10	0.0000
LM (lag)	57.27	0.0000	50.67	0.0000
Robust LM (lag)	5.08	0.0242	11.92	0.0006
LM(error)	75.71	0.0000	48.09	0.0000
Robust LM (error)	23.52	0.0003	9.34	0.0022
LM (SARMA)	80.79	0.0000	60.01	0.0000

LM: Lagrange Multiplier

Table 4 includes statistics from the spatial model for each regime, respectively. Generally, the coefficients of independent variables in spatial models are smaller than their counterparts in the OLS model, suggesting that the original coefficients were inflated. Overall, spatial models present better model fit than the OLS model by observing three measures of model fit (log-likelihood, Akaike Info Criterion, and Schwarz Criterion). Higher log-likelihood and lower Akaike Info Criterion and Schwarz Criterion statistics of the spatial error models reinforce the conclusion from comparing Lagrange Multiplier tests. The R^2 is less useful here because it is not directly comparable with the measure given for OLS results. Moreover, the examination of residuals from spatial models indicates no existence of significant residual spatial autocorrelation in the Entropy values (both Moran's I statistics -.0141 and .0246 are close to zero and not significant at the .05 level). Overall, the spatial models assess to what extent clustering of immigrant residential segregation can be accounted for by independent structural variables and spatial effects.

Table 4 Spatial Regression Models for West and East Regions

	West		East	
	OLS	Error Model	OLS	Lag Model
Variable	b	b	b	b
(Constant)	-0.200 ***	-0.174 ***	-0.216 ***	-0.209 ***
Total population (logged)	0.023 ***	0.022 ***	0.023 ***	0.022 ***
% Foreign-born (logged)	-0.014 ***	-0.010 ***	-0.021 ***	-0.018 ***
% Recent foreign-born	0.035 ***	0.028 ***	0.029 ***	0.027 ***
% Naturalized foreign-born	-0.023 ***	-0.022 ***	-0.047 ***	-0.045 ***
% Foreign-born Who speak only English at home	-0.014 ***	-0.014 ***	-0.021 ***	-0.018 ***
Lagged Entropy	---	---	---	0.167 ***
Lambda	---	0.353 ***	---	---
Log-likelihood	3130.05	3168.86	3509.62	3533.35
AIC[a]	-6249.79	-6325.72	-7007.24	-7052.70
SC[b]	-6217.71	-6293.64	-6974.58	-7014.60

**p* <.05; ***p* < .01; ****p* < .001

a: Akaike info criterion

b: Schwarz criterion

CONCLUSION

This analysis reveals several findings. First, immigrant residential segregation is not randomly distributed in space. The county-level Entropy values exhibit moderate positive spatial autocorrelation. Second, distinctive regional differences in presenting local clusters of the Entropy values are observed and statistically tested. It suggests future studies on spatial distribution of segregation should consider spatial heterogeneity across large regions.

Furthermore, important predictors for immigrant residential segregation at the individual level can be aggregated to structural variables of the corresponding level, and they are important predictors for the clustering of segregation at the county level. However, the Entropy values are not completely determined by these structural variables, because of significant residual spatial autocorrelation after the OLS regression is conducted.

For the two define large regions, different spatial models are applied. The spatial error model applies to the West where local clusters are due to low segregation; the spatial lag model applied to the East where local clusters are due to high segregation. This finding is consistent with the logic of each spatial model, respectively. For high-high clusters, population grows and expands, suggesting a close relationship between nearby locations. For low-low cluster, spatial effects are mainly from unmeasured characteristics of counties that are adverse conditions for immigrants' residence.

REFERENCES

Anselin, L. (2005) (Revised) *Exploring Spatial Data with GeoDa: A Workbook.* http://www.sal.uiuc.edu/stuff/stuff-sum/tutorials

Duncan, O. D., and Duncan, B. (1955). A methodological analysis of segregation indexes. *American Sociological Review, 20,* 210-217.

Iceland, J. (2004). Beyond Black and White: Residential Segregation in Multiethnic America. *Social Science Research, 33*, 248-271

Logan, J. R., Stults, B. J., and Farley, R. (2004). Segregation of minorities in the Metropolis: Two decades of change, *Demography, 41*, 1-22.

Massey, D. S. and Denton, N. A. (1987). Trends in the residential segregation of blacks, Hispanics and Asians: 1970-1980, *American Sociological Review, 52*, 805-825.

White, M. J. (1986) Segregation and diversity measures in population distribution. *Population Index, 52,* 198-221.

White, M. J., and Kim, A. H. (2004). Residential segregation. *Encyclopedia of Social Measurement, 00,* 1-7.

Baller, R. D., Anselin, L., Messner, S. F., Deane, G., and Hawkins, D. F. (2001). Structural Covariates of US County Homicide Rates: Incorporating Spatial Effects. *Criminology, 39(3)*, 561-590.