Microsegregation:

Analyzing Residential Patterns with Multi-level Data

Michael J. White

Brown University

Jennifer Glick

Arizona State University

Ann H Kim

Brown University

Background Paper for the Eastern Sociological Society Presentation

"Revisiting the Old Neighborhood"

Invited Session

24 February 2006

Contact: Michael J. White, Department of Sociology Box 1916, Brown University, Providence RI 02912

email: Michael_White@brown.edu.

Introduction

Residential patterns offer a crucial window on the social relations of groups. The people one has as neighbors reveal a great deal about patterns of tolerance and structural relations in society. From the viewpoint of social policy, housing integration is perhaps one of the most intimate areas of social life in which the government intervenes directly. Residential segregation statistics have become so well established that they are now cited regularly in policy documents and even in the more popular news media. This represents a great success for social statistics: Robert Park's adage that spatial distance reflects social distance has been taken to heart by social scientists who study urban social structure.

At the same time, however, the ethnic diversity of the United States has grown, challenging traditional approaches to segregation studies. The conventional descriptive approach is further strained by the movement of social science toward micro-level and multilevel data, and by the desire to assess outcomes in a multivariate way. These circumstances call for improved measurement of residential clustering. In this paper we argue that the current state of segregation measurement and analysis is inadequate in regard to methodology and demographic realities. The most popular segregation measures developed (and reinvented) from the 1950s through the 1980s, although quite adequate for the original task, need to be evaluated against a more stringent set of statistical criteria and a broader set of objectives.

Segregation reflects who one's neighbors are. Thus at the simplest level, comparing a trait (ethnic group, income category) of a reference person to that of a neighbor helps to clarify the amount and nature of segregation present in a region. If African Americans have predominantly African American neighbors, or if blue-collar workers have no professionals as neighbors, then one has evidence of segregation. Virtually all segregation measures try to capture this pattern of neighboring in some way; typically they work from data that have been tabulated geographically, such as for census tracts.

Urban residential segregation has been one of the phenomena measured most extensively by social scientists. Discussions of the importance of residential segregation permeate twentieth-century writing in urban sociology about the United States and other countries (Timms 1971). Efforts to develop a summary statistic for urban residential patterns extend back a half-century (Cowgill and Cowgill 1951; Jahn, Schmid, and Schrag 1947) and have overlapped with efforts to measure population distribution and population diversity more generally (Duncan and Duncan 1955; Lieberson 1969). Most of these measures were designed to capture black-white residential segregation in large U.S. cities. The application of segregation studies has been most extensive in regard to ethnic groups; some analysis has been based on socioeconomic status; life cycle segregation has been studied comparatively less.

Even though it is acknowledged that segregation indices summarize the outcome of several related processes, segregation measurement provides key insights into urban residential patterns from readily available observational data. Theoretical and empirical studies of segregation remain a major window on ethnic and racial relations (Bean and Tienda 1987; Farley and Allen 1987; Massey 1985; NAS 1997; Peach 1975). More recently, researchers have striven to tie residential segregation to its causes, either through the window of audit studies (Yinger

1991) or through opinion interviews with urban residents (Bobo 1986, 1989; Farley et al. 1994). Studies of urban neighborhood dynamics, now themselves a large portion of the literature, help to complete the picture (e.g., Alba and Logan 1991, 1993; Clark 1992; Krivo, et al., 1998; Massey and Denton 1985; Rosenbaum et al. 1991; Schelling 1972; Zhou and Logan 1991).

Most analysts still resort to statistical summary measures for dichotomies, and models must be tested at the aggregate level; this approach raises long-standing questions about the validity of ecological inference. Although this was an excellent extension of current work and a very reasonable response to the presence of competing indices, we still do not know whether other dimensions of segregation are untapped, nor the degree to which indices are comparable across metropolitan areas of differing size and ethnic composition.

Finally, segregation studies have been particularly challenged by growing ethnic diversity. Much work now recognizes this diversity. A substantive line of work examines residential patterns for “multi-ethnic metros” (Denton and Massey, 1991; Iceland, 2004). A related line of methodological work considers ways to summarize the segregation patterns when a dichotomous description is clearly no longer adequate (White, Kim and Glick, 2005).

Conceptual Approach, Data and Methods

Develop adjustments to segregation indices using multilevel data.

A simple model of segregation would reflect an individual's neighboring pattern, whether that pattern originated directly by self-selection or discrimination, or indirectly through composition. Until now, most segregation analysis has incorporated other traits (e.g., the effect of SES on racial segregation) either through ecological analysis or through an occasional cross-tabulation in census materials. Ecological analysis presents its own challenges (Langebein and Lichtman 1978). For a limited geographic area, it has been possible to conduct multilevel analysis by painstakingly constructing correlation matrices from census materials (Alba and Logan 1993). Two-way classification has been quite successful in showing the persistence of racial residential segregation across education or income categories (Denton and Massey 1988; Farley 1977; Farley and Allen 1987). Still, we would expect a variety of characteristics to influence residential patterns; extant empirical evidence suggests that it does so (White, Biddlecom, and Guo 1993; White, Dymowski, and Wang 1994). In this portion of the project we will investigate a micro-level model of residential segregation, relying on a unique design for the American Housing Survey.

Model. Consider a micro-level model of neighboring:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \gamma Z_i + \varepsilon_i \qquad (1)$$

The quantity Y_i measures the segregation outcome, typically the probability that the "neighbor" is of the same (ethnic, educational, age) category as the *i*th individual. We could also model the probability that a neighbor is a member of a specified reference group, or we could model the composition of the immediate neighborhood (e.g., percent Anglo; number of neighbors in poverty status). Z_i measures this trait for the index individual.

To conduct this analysis we face several tasks. First we must model with and without controls. We will employ the micro-level model, as indicated in equation (1). *Z* and *Y* capture the unconditional segregation pattern, much as one might obtain from a tabulation. For example, $Z = Y = 1$ if both the index individual and the neighbor describe themselves as of Mexican origin. By introducing the *X*s into the models after the *Z*s we can determine how the unconditional level of segregation is modified by the influence of other traits. For instance, we can examine how ethnic residential segregation is modified by socioeconomic status.

Second, we can correct for household aggregation, a nettlesome problem that permeates almost all segregation measurement. Intrahousehold diversity becomes reflected in the segregation statistic. If poor persons have larger families, tabulations of segregation by poverty status will reflect this influence. Native-born children (of second-generation parents) depress the measured amount of "ethnic" segregation in historical studies. The micro-level model will allow us to remove this effect and, equally important, to measure its relative importance.

Third, and building on these results, we will be able to calculate indices (probabilities of neighboring) adjusted for composition. Thus the field will have access to poverty-nonpoverty segregation or Irish-Mexican segregation, both unadjusted and adjusted for family size. One still must demonstrate how these adjusted and unadjusted indices are reexpressed most effectively in terms of traditional segregation indexes.

Data. Using the *American Housing Survey,* we will estimate a version of equation (1) with AHS data. The AHS is unique in that it is a representative sample of U.S. households, taken in alternate years (U.S. Department of HUD 1993). As the AHS documentation indicates, clusters of neighbors are identified in the machine-readable file. For several hundred of the sampled AHS housing units in urban areas, the 10 closest units were chosen and interviewed. One can therefore analyze income,, income mix, family composition, types of houses, etc. in each cluster of neighbors. (U.S. Department of HUD 1993).

This approach allows one to produce quantitative measures showing how much the calculated segregation is reduced (or increased) when individuals' background traits are controlled. In turn, the approach helps one interpret some of the origins of the intergroup differences (such as black-Anglo versus Asian-Anglo) that exist in the current literature.

In this work we begin with simple cross-tabulations of an index@ person's race/ethnicity by that of a neighbor's. (In this case the persons are the identified householders in the AHS.) The index-by-neighbor tabulation gives a simple measure of neighboring or clustering. This neighboring pattern reflects segregation in the city as well as overall ethnic composition. We then move on to adjusting the index for composition, calculating Yule's Q and other measures of association for cross-tabulation. We conclude with multivariate analysis, predicting neighbor's race on the basis of a host of characteristics B SES, life cycle, race/ethnicity of the index householder and his/her household.

Results Table 1 presents results from 1993 tabulations of the Index Householder Race by

Neighbor's race. This is done for all 6,138 persons in metropolitan areas sampled under this scheme. One observes for instance that for Anglo householders, some 86.8 percent of neighbors are also Anglo, whereas only 4.4 percent are black, 3.3 percent Asian, and so on. African-Americans householders have 64 percent of neighbors who are African American. The diagonal cells of this neighboring matrix reflect in-group neighboring. To a great degree this statistic is parallel to the Aisolation index@ (within-group-segregation) represented in several studies in recent years.

This neighboring matrix is inclusive of all US metropolitan areas and tells us about the overall pattern of neighboring. Already it is an improvement over most existing segregation approaches, because it compares all groups to all groups. Most existing indices compare within-group segregation (group vs. all others) or use a reference group (blacks vs. Anglos; Asians vs. Anglos). While we have the full spectrum, we all have a distribution across the 5x5 table that reflects the relative sizes of the five groups in US metropolitan territory as a whole.

To adjust for we calculate standardized values [(Observed-Expected)/(Expected) for each cell. Positive numbers mean an over-representation of that ethnic group compared to its proportions in US metropolitan areas; negative numbers reflect under-representation. Values larger in magnitude point to greater deviations from expectation. All groups are over-represented with neighbors of their own ethnic group, and this is no surprise. What is more interesting is to compare the magnitudes of these standardized values and also compare them across cells. Of the four identifiable ethnic groups, blacks and Asians clearly are more over-represented with neighbors of their own ethnic background. Note also that whites are less likely (in standard units) to have black neighbors than Asian or Hispanic neighbors. Latinos are more likely to have Asian and "Other Ethnicity" neighbors, but these numbers represent small samples. Finally, the table reveals that Black-White and Black-Asian are some of the largest standardized magnitudes of those cells with substantial counts, suggesting that these are pairwise segregation levels that are appreciable.

Table 2 repeats the basic distribution for seven large metropolitan areas. We can see some detect some notable differences across the metropolitan areas. Consider, for instance, New York and Los Angeles. Both have Anglo concentrations (see bottom row of table) of about 55 percent, yet Anglo-to-Anglo neighboring differs appreciably in the two cities. Los Angeles Anglos have 80% Anglo neighbors, whereas New York Anglos have 91% Anglo neighbors. This would indicate a higher degree of intra-group segregation, even if one were to control for city-level ethnic composition.

Table 3 turns to multivariate results. We use multinomial logit models to predict for each observed householder the ethnicity of the neighbor. Covariates included in the model include the index householder's ethnicity B black, Asian, Hispanic, with Anglo and other as the reference category B sex, marital status, and age. We also include household income and household composition (number of children and elderly) as predictors. For ease of exposition we include a basic set of covariates that covers key dimensions of urban differentiation: life cycle, socioeconomic status, and ethnicity. These results are pooled across all metropolitan areas.

The results are very telling. The effects of ethnicity swamp other traits. All ethnic covariates are significant and collectively, they absorb a substantial fraction of explained variance. We can compare directly the scale of the ethnic-to-ethnic coefficients. We see that black-to-black neighboring is most strongly indicated, even after controlling for other characteristics of the household. Note that this coefficient 3.79 exceeds the value of the two other intra-group neighboring coefficients.

We also observe that Latinos are somewhat more likely to be found in out-group neighboring patterns. Latinos are more likely to live next to Asians than are Blacks, and Latinos are more likely to live next to blacks than are Asians. Those who identify themselves as “other” in the race/ethnicity portion of the AHS do exhibit a very different neighboring pattern. In fact, such householders are quite likely to live next to blacks, Asians, and Hispanics.

What also emerges from these results is how unimportant life cycle characteristics are. Sex and age of household head are never statistically significant. The age composition of other household members has only modest predictive value. We find that those with more children are somewhat more likely to live near any one of the four minority groups, but are especially more likely to live near Latinos. Households with elderly member are not particularly likely to be spatially differentiated, save that such households are less likely to be found next to those of other race.

Income does matter, and the effects are quite important across the four types of minority neighbors. For each $10,000 of household income we find that the log-odds of living next to an African American family declines by 0.10, although this is not statistically significant. Similar negative relationships -0.11 and -0.13, respectively, exist for having Latino and AOther@ neighbors, and both of these coefficients are statistically significant. Finally, it is noteworthy y that household with higher income are more likely to have Asian-origin neighbors. The log odds here is about 0.06, and it is also statistically significant. Thus, these results are consistent with a preference ordering or discriminatory pattern across neighboring.

Conclusion

Our contributions in this paper are in two veins. First, we seek to make a methodological improvement in our analysis and understanding of residential segregation. Rather than being limited to aggregated data for block-groups, tracts, or other geographic aggregates with limited socioeconomic characteristics, we show how to exploit multi-level data to control for a variety of background traits while still measure inter-ethnic neighboring patterns. This approach allows the researcher to examine multi-ethnic settings more readily. Whereas much of earlier segregation analysis was based on pairwise use of the index of dissimilarity or some other index, out approach uses multi-level data to look simultaneously at ethnic neighboring across a wide array of available national origin or racial groups.

Our second line of contribution is substantive. Although our results are still preliminary and rely on one wave of the American Housing Survey, the findings are still quite suggestive of the factors associated with residential sorting in American metropolitan areas. We find that ethnicity itself is an extremely powerful predictor of neighboring. Controlling for household life cycle and income, while important, only removes a small fraction of the inter-ethnic differences in neighboring patterns. We find an important differentiation across ethnic groups for neighboring. As householders have more income we find they are less likely to live nearby blacks and Latinos, but more likely to live next to Asians, once other background characteristics are controlled.

Since previous segregation analyses have been often limited to pairwise comparisons and rarely control for socioeconomic and life cycle characteristics, we hope that our analysis begins to point the way to a new, more comprehensive style of analysis of residential intermingling in an increasingly diverse US metropolitan population.

References

Cowgill, D. O., and M. S. Cowgill (1951): “An Index of Segregation Based on Block Statistics,” *American Sociological Review*, 16(6), 825–831.

Denton, Nancy A. and Douglas S. Massey 1991. “Patterns of Neighborhood Transition in a Multiethnic World: U.S. Metropolitan Areas, 1970-1980” *Demography,* 28: 41-63.

Duncan, OD and B Duncan. 1955. “A methodological analysis of segregation indexes” *American Sociological Review*.

Fischer, Mary J. 2003. "The relative importance of income and race in determining residential outcomes in U.S. urban areas, 1970-2000." *Urban Affairs Review* 38:669-696.

Fong, Eric. 1994. "Residential proximity among racial groups in U.S. and Canadian neighborhoods." *Urban Affairs Quarterly* 30:285-297.

Frey, William H., and Reynolds Farley. 1996. "Latino, Asian and Black segregation in US metropolitan areas: Are multiethnic metros different?" *Demography* 33:35-50.

Iceland, John. 2004. “Beyond Black and White: Residential Segregation in Multiethnic America.” *Social Science Research* 33:248-271.

Jahn, J., C. F. Schmid, and C. Schrag (1947): “The Measurement of Ecological Segregation,” American Sociological Review, 12(3), 293–303.

Lieberson, Stanley. 1969. “Measuring Population Diversity” *American Sociological Review,* 34: 850-862.

Massey, D. S., and N. A. Denton (1988): “The Dimensions of Residential Segregation,” Social Forces, 67(2), 281–315.

Reardon, Sean F, and Glenn Firebaugh. 2002. "Measures of multigroup segregation." *Sociological Methodology* 32:33-67.

Taeuber, K. E., and A. F. Taeuber (1965): *Negroes in Cities*. Aldine Publishing Company.

Timms, D. 1971. *The urban mosaic: towards a theory of residential differentiation*. Cambridge U Press.

White, MJ, J. Glick and A. Kim. 2005. “Mapping Social Segregation” *Sociological Methods and Research* 34:173-203.

Table 1

Neighboring Patterns 1993 American Housing Survey

	Neighbor's Race					
	Anglo	N.H. Black	Asian	Hispanic	Other	Number of Householders
Householder's Race						
Anglo	86.8	4.4	3.3	5.1	0.5	4,183
N	3630.4	182.8	136.8	211.2	21.3	4,183
Expected	2879.7	660.0	169.0	401.9	71.9	4,183
Standardized	0.3	-0.7	-0.2	-0.5	-0.7	
N.H. Black	26.3	64.0	2.0	7.7	0.0	1,006
N	264.4	643.5	20.5	77.6	0.0	1,006
Expected	692.6	158.7	40.6	96.7	17.3	1,006
Standardized	-0.6	3.1	-0.5	-0.2	-1.0	
Asian	62.2	5.9	25.7	6.3	0.0	229
N	142.3	13.4	58.9	14.4	0.0	229
Expected	157.7	36.1	9.3	22.0	3.9	229
Standardized	-0.1	-0.6	5.4	-0.3	-1.0	
Hispanic	36.8	6.0	5.6	44.1	7.5	607
N	223.4	36.2	34.1	58.3	10.4	362
Expected	249.5	57.2	14.6	34.8	6.2	362
Standardized	-0.1	-0.4	1.3	0.7	0.7	
Other	49.0	24.8	0.0	0.0	26.2	113
N	55.4	28.0	0.0	0.0	29.6	113
Expected	77.8	17.8	4.6	10.9	1.9	113
Standardized	-0.3	0.6	-1.0	-1.0	14.2	
TOTAL	68.9	15.8	4.0	9.6	1.7	6,138
	4226	969	248	590	106	6,138

Table 2

Neighboring Patterns for Seven Large Metropolitan Areas

	SMSA						
	Los Angeles	New York	Atlanta	Chicago	Dallas	San Francisco	Houston
Householder Race							
Anglo							
% Black	0.9%	5.8%	21.3%	5.9%	6.2%	13.3%	4.6%
% Asian	8.8%	1.7%	0.0%	4.7%	0.0%	18.7%	1.5%
% Anglo	80.1%	90.5%	76.6%	86.1%	82.1%	63.4%	82.9%
%Hispanic	9.4%	2.1%	0.0%	2.5%	11.7%	3.5%	8.8%
Black							
% Black	52.8%	88.4%	87.3%	92.6%	40.6%	-	83.0%
% Asian	4.0%	0.0%	0.0%	0.0%	0.0%	-	0.0%
% Anglo	33.9%	5.0%	12.7%	3.7%	38.8%	-	13.5%
% Hispanic	9.2%	5.1%	0.0%	2.5%	20.6%	-	3.5%
Asian							
% Black	0.0%	11.1%	-	0.0%	-	4.9%	-
% Asian	30.1%	33.3%	-	33.3%	-	56.0%	-
% Anglo	24.6%	55.6%	-	66.7%	-	34.1%	-
% Hispanic	34.3%	0.0%	-	0.0%	-	4.9%	-
Hispanic							
% Black	2.1%	13.4%	-	15.4%	14.9%	6.7%	44.4%
% Asian	7.4%	9.4%	-	7.6%	0.0%	6.7%	0.0%
% Anglo	27.8%	34.4%	-	29.7%	24.5%	33.4%	33.3%
% Hispanic	57.3%	34.5%	-	31.5%	60.7%	47.4%	22.2%
All							
% Black	9.8%	25.7%	38.9%	42.6%	16.7%	10.9%	25.8%
% Asian	8.6%	3.9%	0.0%	3.7%	0.0%	23.8%	1.7%
% Anglo	55.1%	54.4%	59.5%	46.9%	57.7%	54.2%	63.2%
%Hispanic	23.6%	11.4%	0.0%	4.6%	25.5%	9.6%	8.0%

Table 3

Multinomial Regression Results for Household and Neighboring Patterns,

1993 American Housing Survey

	Race/Ethnicity of Neighbor							
	Black		Asian		Hispanic		Other	
Head's Race/Ethnicity								
Black	3.79	*	0.63	*	1.03	*	1.33	*
Asian	0.94	*	2.83	*	1.02	*	1.10	*
Hispanic	1.27	*	1.55	*	2.41	*	1.52	*
Head's Sex (female)	0.00		-0.24		0.01		0.09	
Head's Marital Status (vs. currently married)								
Div., widow, sep.	0.18		-0.11		0.12		-0.05	
Never Married	0.29	*	0.19		-0.12		-0.11	
Head's Age	0.01		0.01		0.00		0.00	
Household income (in $10,000)	-0.10		0.06	*	-0.11	*	-0.13	*
Household Composition								
Number of Kids < 18	0.09		0.08		0.17	*	0.03	
Number Adults 65+	-0.05		-0.02		0.02		-0.54	(a)
Constant	-2.94		-4.03		-2.12		-3.33	