International Journal of Business and Economics, 2003, Vol. 2, No. 1, 27-37

Revisiting Residential Segregation by Income: A Monte Carlo Test

Junfu Zhang*
Research Fellow, *Public Policy Institute of California*, *U.S.A*.

Abstract

A long-standing hypothesis states that racial housing segregation in the U.S. results from the income inequalities between blacks and whites. This paper reexamines this hypothesis with a new methodology. We present a Monte Carlo study to show that segregation by income explains only a small proportion of the high level of segregation.

Key words: residential segregation; Monte Carlo test; dissimilarity index
JEL classification: C12; C15

1. Introduction

Residential segregation between blacks and whites has been a salient characteristic of American urban landscapes for many decades. The formation of black ghettos in the U.S. started in the early 20th century, following the massive black migration from the rural South to the urban North. By 1940, the silhouette of modern black ghettos was already in place in most northern cities. And by 1970, the average urban black lived in a census tract that was 68% black. The Civil Rights Movement in the 1950s and 1960s brought about legislative attempts to reduce residential segregation. In particular, the Fair Housing Act of 1968 was intended to eliminate racial discrimination in the housing market. While the level of residential segregation has dropped slightly since 1970, its persistence is beyond what most Americans had expected. By 1990, the average urban black still lived in a census tract that was 56% black (Cutler et al., 1999; Massey and Denton, 1993).

Residential segregation is believed to cause social problems, such as concentration of poverty, as well as incur social and economic costs. Social scientists across disciplines have been trying to understand the root causes of segregation and seek possible ways to promote racial integration. Some scholars argue that individuals prefer to live with like-colored neighbors, which leads to segregation (e.g., Schelling, 1971); some others believe that segregation persists because racial discrimination still exists in the housing and mortgage market (e.g., Yinger, 1995). Another

Received April 17, 2002, revised December 19, 2002, accepted April 8, 2003.
*Correspondence to: Public Policy Institute of California, 500 Washington Street, San Francisco, CA 94111, U.S.A. E-mail: zhang@ppic.org. I would like to thank an anonymous referee and the Managing Editor for their constructive comments. I also appreciate Nikesh Patel's assistance.

long-standing explanation of segregation, which traces back to sociologist Robert Park (1926), emphasizes the economic inequalities between blacks and whites. Park postulates that the physical distance between racial groups reflects their socio-economic distance. It is often argued that expensive housing tends to be located in separate neighborhoods from inexpensive housing; persons with high incomes and upper-level jobs tend to live apart from persons with low incomes and more menial jobs. Thus, because blacks have lower average incomes, they cannot afford to live in the same neighborhoods as whites. This argument implies that the observed racial housing segregation is, in fact, segregation by income. Clearly, various hypotheses regarding the causes of segregation are not mutually exclusive; yet, singling out the most important factor is crucial because different causes prescribe different desegregation policies.

Previous studies of residential segregation by income have employed various methods and produced mixed results. In this short paper, we argue that the Monte Carlo method is particularly useful for testing the income inequality hypothesis. Using this technique, we reexamine how much of residential segregation is attributable to income inequalities between blacks and whites. We find that segregation by income is able to account for only a small proportion of racial housing segregation.

The remainder of this paper is organized as follows. Section 2 introduces the commonly used dissimilarity index to measure residential segregation. Section 3 reviews previous studies and their limitations. Section 4 presents our Monte Carlo test. Section 5 concludes with some remarks.

2. The Dissimilarity Index

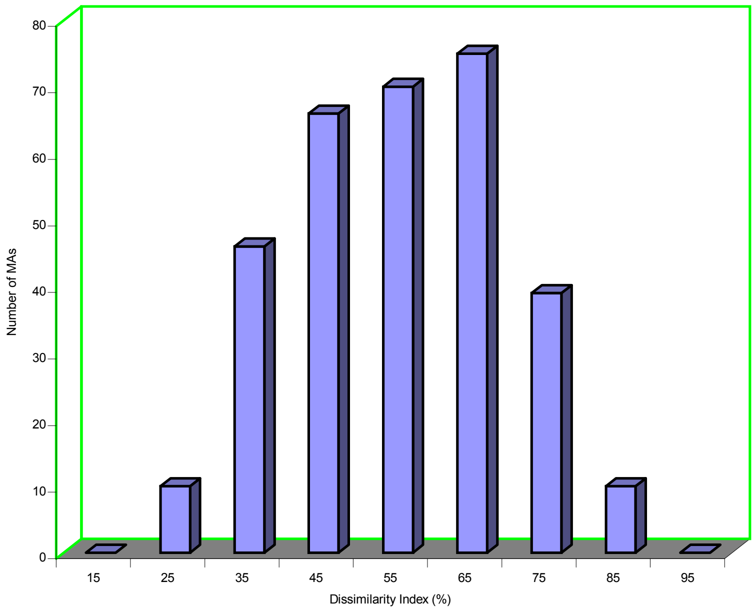
Although many indexes have been devised to measure residential segregation, one in particular has emerged as the most popular among scholars: the dissimilarity index (*DI*). It is defined as follows:

$$DI = \frac{1}{2}\sum_i \left| \frac{b_i}{B} - \frac{w_i}{W} \right|, \tag{1}$$

where b_i and w_i are the number of blacks and whites (measured in persons or households), respectively, living in neighborhood i and B and W are the total number of blacks and whites, respectively, living in the metropolitan area (MA). *DI* indicates the fraction of the black population that would have to change residential location, if whites do not move, in order to achieve an even distribution. It attains the value 0 if blacks and whites are evenly mixed together and attains the value 1 if blacks and whites never live in the same neighborhood. In general, a dissimilarity index above 60% is considered high, an index between 30% and 60% is considered moderate, and an index below 30% is considered low (Massey and Denton, 1993).

In empirical studies, neighborhoods are usually approximated by census tracts. Generally, each census tract has between 3000 and 8000 residents. Boundaries of census tracts reflect visible features such as major streets, highways, or rivers. Figure 1 describes the level of residential segregation calculated using the 1990 Census data. Among a total of 316 MAs in the U.S., 124 had segregation indices above 60%, and 49 had indices higher than 70%. The top 10 most segregated areas, including big metros such as Buffalo, Chicago, Cleveland, Detroit, and Milwaukee, all had segregation indices greater than 80%. That is, in these areas, more than 80% of the black population would need to be relocated in order to reach a perfect mixture of blacks and whites. It is this situation that Massey and Denton (1993) referred to as "hypersegregation" and "American apartheid."

Figure 1. Distribution of MAs over Dissimilarity Index, 1990



3. Limitations of Existing Literature

The majority of empirical works of residential segregation employed the dissimilarity index. Previous studies have taken three different methodological approaches to test the income inequality hypothesis: the cross-sectional approach, the direct standardization, and the indirect standardization (Massey, 1981).

The cross-sectional approach correlates degree of residential segregation with socioeconomic inequalities between racial groups. Using the technique of path analysis, Marshall and Jiobu (1975) studied residential segregation in 149 large urban areas. They used the dissimilarity index to measure residential segregation, income differentiation, and occupational differentiation for each area. They also calculated percentage of nonwhites, number of nonwhites, and white-nonwhite growth differential. Intercorrelations of those 6 variables were tabulated to make implications. Marshall and Jiobu found that income inequalities have the largest effect on residential segregation; occupational status differentials are also major determinants of segregation, independent of any indirect impact that occupation has through income differentiation. Other variables demonstrated lower correlations with the degree of residential segregation. Using multivariate analysis, Schnare (1977) tried to explain the variation of segregation in 112 metropolitan areas in both 1960 and 1970. Socioeconomic variables such as median housing costs as a percentage of median family income, black-white ratio of white-collar workers, and black-white median income ratio were all found to be positively and significantly correlated with black exposure to whites in both years.

The support of the hypothesis of socioeconomic inequalities is inferred from the strong correlation between residential segregation levels and socioeconomic indicators. However, it is well known that correlation does not imply causality. It remains unclear whether segregation causes inequality, or the other way around, or they both result from some other factors such as racial discrimination. In addition, the cross-sectional approach suffers from small sample biases. It takes a whole metro area as an observational unit, which does not yield many observations in the whole country, and therefore a few extreme cases could drive the empirical results.

Instead of looking at variations across metro areas, the direct standardization approach calculates the dissimilarity index within each income level in a single metro area. That is, it calculates the following index:

$$DI_j = \frac{1}{2}\sum_i \left| \frac{b_{ij}}{B_j} - \frac{w_{ij}}{W_j} \right|, \tag{2}$$

where i is the index of neighborhood and j the index of income category. Then the changes in segregation across different income levels are examined to give implications for the importance of income inequalities. This approach, more than the cross-sectional approach, works with even smaller samples since one cannot break the population into too many income brackets. Using this method, Massey (1979, 1981) found that Hispanic-white segregation falls off steeply as family income increases, which is interpreted as a support to the social class explanation of residential segregation. Yet between blacks and whites, the direct standardization method usually finds equally severe segregation in different income categories (Farley, 1977; Massey, 1979; Massey and Denton, 1993). That is, poor blacks are separated from poor whites, and rich blacks are separated from rich whites. Difficulties arise in this situation, because uniformly high *DI* values across income groups very well suggest that some non-income factors could be important. It is nevertheless inappropriate to conclude that income inequalities are not important based on the observation that blacks and whites are equally segregated across income levels.

Consider the example given in Table 1. There are two cities and each city has only two income groups: the poor and the rich. In each city, there are 4 poor neighborhoods and 4 rich neighborhoods. The racial composition of each neighborhood is given in the table. In both cities, blacks and whites are completely segregated and never share a neighborhood. The only difference between the two cities is in racial income distribution. In city 1, income is evenly distributed between blacks and whites: exactly 50% of the poor are blacks and 50% of the rich are blacks; in city 2, 75% of the poor are blacks and only 25% of the rich are blacks. The direct standardization detects no differences between these two cities. The dissimilarity index is 1 in both income categories in both cities. However, that does not imply income inequality is insignificant.

Suppose the color line is breached and people move freely without facing any racial barriers, but income inequalities still prevent the poor living side by side with the rich. In expectation, the two cities will end up with the situation described in Table 2. Now, in both cities, segregation is eliminated within each income category. However, the overall segregation index reduces all the way to 0 in city 1 but only declines to 0.5 in city 2. The income barriers in city 2 are able to keep segregation at a high level. This example shows that one cannot tell whether income inequality is an important factor in segregation given that people in different income groups are equally segregated. To see the role of income inequality, one must take into account the correlation between income distribution and race.

Table 1. Two Segregated Cities

City 1 $DI_{overall} = 1$		City 2 $DI_{overall} = 1$	
Poor Neighborhoods $DI_{poor} = 1$	Rich Neighborhoods $DI_{rich} = 1$	Poor Neighborhoods $DI_{poor} = 1$	Rich Neighborhoods $DI_{rich} = 1$
100 blacks, 0 whites	100 blacks, 0 whites	100 blacks, 0 whites	100 blacks, 0 whites
100 blacks, 0 whites	100 blacks, 0 whites	100 blacks, 0 whites	0 blacks, 100 whites
0 blacks, 100 whites	0 blacks, 100 whites	100 blacks, 0 whites	0 blacks, 100 whites
0 blacks, 100 whites	0 blacks, 100 whites	0 blacks, 100 whites	0 blacks, 100 whites

Table 2. Dissimilarity after Eliminating Racial Segregation

City 1 $DI_{overall} = 0$		City 2 $DI_{overall} = 0.5$	
Poor Neighborhoods $DI_{poor} = 0$	Rich Neighborhoods $DI_{rich} = 0$	Poor Neighborhoods $DI_{poor} = 0$	Rich Neighborhoods $DI_{rich} = 0$
50 blacks, 50 whites	50 blacks, 50 whites	75 blacks, 25 whites	25 blacks, 75 whites
50 blacks, 50 whites	50 blacks, 50 whites	75 blacks, 25 whites	25 blacks, 75 whites
50 blacks, 50 whites	50 blacks, 50 whites	75 blacks, 25 whites	25 blacks, 75 whites
50 blacks, 50 whites	50 blacks, 50 whites	75 blacks, 25 whites	25 blacks, 75 whites

The third approach, called indirect standardization, resolves the problem we just saw in the direct standardization method. Proposed by Duncan et al. (1961), this approach assumes an even distribution of blacks and whites within social classes and allows segregation between classes. One can then calculate an expected level of segregation and compare this value to the level of segregation actually observed. In practice, researchers usually proceed with the assumption that the black-white ratio in each income category in any neighborhood is equal to the black-white ratio at the same income level in the whole metro area. In that case, the indirect standardization essentially amounts to calculating the following index, which will be coined as the index of "income effect" (*IE*):

$$IE = \frac{1}{2}\sum_i \left| \frac{\sum_j (b_{ij} + w_{ij}) q_j}{B} - \frac{\sum_j (b_{ij} + w_{ij})(1 - q_j)}{W} \right|, \tag{3}$$

where i is the index of neighborhood and j the index of income category and q_j is the fraction of blacks in the total population that belongs to income category j over the whole metropolitan area. Note that *IE* is obviously defined relative to *DI*. The difference is that *IE* uses the "expected" number of blacks and whites in each neighborhood whereas actual numbers are used in *DI*. If income is the only factor that influences individuals' choice of residential location, *IE* seems to be the dissimilarity index that one would ask for.

Taeuber (1968) used the indirect standardization method to study the effect of income distribution on segregation in Cleveland and found that income inequalities have little explanatory power for the overall level of segregation. Bleda (1979) argued that indirect standardization is a method better for assessing the role of social classes in explaining racial housing segregation and got findings similar to Taeuber's. Using the same technique and data from the Toronto Area, Darroch and Marston (1971) showed that income, education, and occupation differentials by themselves could only account for a small amount of residential segregation.

In this paper, we propose to explicitly formulate the income inequality interpretation of residential segregation as a test hypothesis in the statistical sense. We simulate the data generating process under the assumption that individuals are segregated along income but not race. Like the indirect standardization approach, our approach takes into account the joint distribution of population over income and race and hence avoids the problem with the direct standardization method.

In addition, the Monte Carlo method has the following advantages. First, it gets around the small sample problem since a researcher can obtain as much data as desired by running the simulation. Second, it enables us to accommodate segregation resulting from random moves. Even in the case where people are equally rich and their residential choices are absolutely race neutral, we should still expect some degree of segregation that could emerge by chance. The Monte Carlo method is able to tell how much segregation is attributable to such random factors. Third, it allows us to state at what confidence level our conclusion could hold. This is a generic advantage of any explicit statistical test. And finally, the Monte Carlo method is so powerful and versatile that it can be used to test much more complicated hypotheses and many variations of such hypotheses. In particular, it is able to test multiple hypotheses simultaneously.

4. A Monte Carlo Test

The Monte Carlo method has become a standard technique in statistics, especially now that modern technology has significantly lowered its computing costs. The general idea behind a Monte Carlo study is to model the data-generating process, create several sets of artificial data with a computer, and then use these data to perform a test or, more commonly in statistics, to study properties of estimators. While the Monte Carlo technique is routinely employed by economists and statisticians, social scientists working in the area of residential segregation do not seem to take advantage of this powerful tool.

Originally proposed by Barnard (1963), a one-tailed Monte Carlo test can be carried out by following these steps below:

1. Specify a significance level α;
2. Choose integer values M and N such that $M/N = \alpha$;
3. Calculate T_1, the data-based value of the test statistic;
4. Simulate $N-1$ datasets under the null hypothesis and calculate $N-1$ test statistics $T_2, T_3, \ldots, T_N$ corresponding to the simulated datasets;
5. Reject the null hypothesis if T_1 is among the M largest values of the N test statistics.

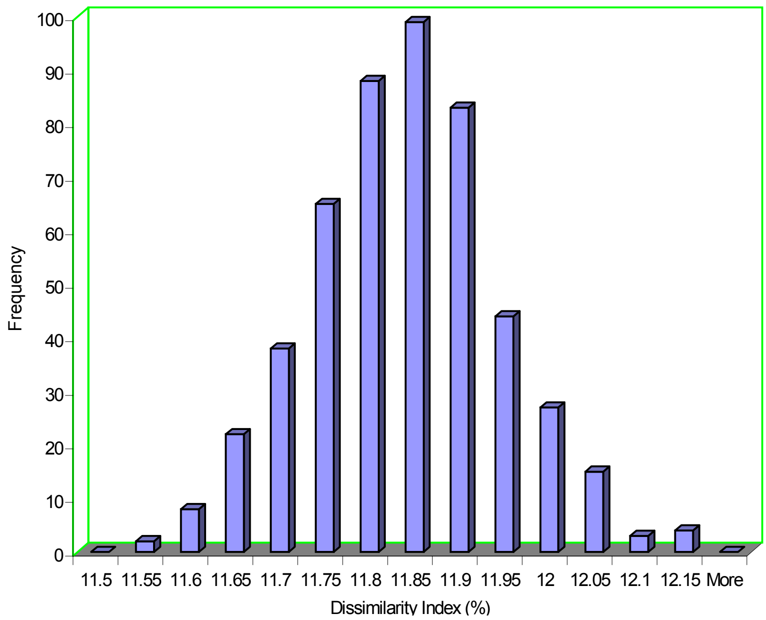
In our case, the null hypothesis is formulated as follows: conditional on income level, any two residential locations in a metro area are equally likely to be occupied by black (white) households. That is, household income alone affects where a family lives. We choose $\alpha = 0.01$, $M = 5$, and $N = 500$. Given the size of the test α, a larger M means a larger N (i.e., more repetitions of simulation). Since we will conduct the test based on an empirical distribution of the test statistic, the more replications we have, the more precise the critical region will be. Besag and Diggle (1977) argued that $M = 5$ suffices for most purposes.

We use the dissimilarity index as our test statistic. Given any metro area under examination, for example Baltimore, we first calculate its *DI* from real data (1990 Census data). We then randomly reshuffle families at each income level to simulate the data generating process under the null hypothesis. It is carried out this way: starting from the poorest income level, we count the total number of families over the whole metro area at that income level, of which we know how many are black and how many are white. We then randomly assign those families back to each neighborhood. If neighborhood i used to accommodate 20 of the poorest families, we will allocate 20 of the poorest families to it. Among those, whether a family is black or white is randomly determined. The Census data categorizes households into 9 different income groups, so we repeat this random assignment 9 times. In the end, we know the total number of black (white) families assigned to each neighborhood. This simulated dataset allows us to calculate a *DI*. The exercise is repeated 499 times, which gives us 499 simulated *DI*s. We put the *DI* from the real data and the simulated *DI*s together and sort them in ascending order. The 495th value (DI_{495}) is the critical value of our test. We shall reject the null hypothesis if the real *DI* exceeds

DI_{495}.

Figure 2 shows the distribution of the 499 simulated test statistics for Baltimore. The *DI* from the real data is 69.9%, which lies far off to the right of the distribution. We thus reject the null hypothesis. We also performed this Monte Carlo test for some other metro areas. The first two columns of Table 3 compare the real *DI* with the critical value of the test. The null hypothesis is uniformly rejected at the 1% significance level for all the areas under examination. In fact, the *DI* from the real data exceeds the critical value by so much in those areas that the *p*-value of the test must be far below 1%. As a result, it appears to be extremely unlikely that income disparity can fully account for segregation.

Figure 2. Simulated Values of the Dissimilarity Index for Baltimore, 1990



Consider Chicago, one of the most segregated metro areas in the U.S. throughout the twentieth century. Our simulation suggests that if income alone determines where a family is located, we should expect a dissimilarity index around 13%. In contrast, we observe an 82% index in Chicago.

A close look at the data reveals two features of the real world that help us better understand why income inequalities play a limited role in explaining segregation. First, housing quality is heterogeneous within any neighborhood, and therefore, a neighborhood always accommodates families from several different income levels. Second, although blacks on average are poorer than whites, distributions of the two races over income overlap a lot. For example, in Chicago, 5% of the rich families that have annual income higher than $100,000 are blacks; 30% of the poorest families, which earn less than $5,000 a year, are whites.

Table 3. Real Dissimilarity Indexes, Income Effect Indexes, and Simulated *DI*s (%)

	DI from Real Data	Critical Value, DI_{495}	Income Effect, *IE*	Simulated *DI*s		
				Minimum	Median	Maximum
Baltimore	69.90	12.09	11.57	11.52	11.81	12.14
Chicago	82.36	13.12	12.67	12.72	12.94	13.16
Cleveland	80.03	12.80	12.28	12.21	12.55	12.85
Detroit	86.76	13.91	13.47	13.52	13.72	13.97
Washington	65.67	12.24	11.86	11.79	12.04	12.31

Therefore, we should expect to see some blacks even in the richest neighborhoods and some whites even in the poorest neighborhoods. However, Chicago's 82% dissimilarity index suggests that most neighborhoods are either predominantly black or predominantly white, which is far beyond the explanatory power of income inequalities.

Figure 2 and Table 3 also show that random residential choices do cause some variation in the level of segregation. However, that variation is relatively small. In the case of Baltimore, random residential choices in 499 simulations only cause the level of segregation to vary within a less-than-one-percent range: between 11.52 and 12.14%. Interestingly, although the five metro areas differ a lot in terms of the dissimilarity index, their simulated *DI*s are quite similar. It suggests that if household income alone determines where people live, we should expect those metro areas to be more or less equally segregated.

We also calculate the *IE* index by making the black-white ratio in each income group in any neighborhood equal to the ratio in that income group in the whole metro area (Table 3, third column). Not surprisingly, the *IE* index tends to be in the lower range of the distribution of the simulated dissimilarity index. Since the whole range of random variation is small, however, it is very unlikely that the dissimilarity index under the null hypothesis will be 1% higher than the *IE* index. This implies that it is not a very stringent assumption to equate the black-white ratio in each income group in any neighborhood to the ratio in that income group in the whole metro area as the indirect standardization approach does.

5. Concluding Remarks

To improve upon the existing literature, we have proposed that we use the Monte Carlo technique to reinvestigate to what extent residential segregation is attributable to income inequalities between blacks and whites. Our Monte Carlo test rejects the hypothesis that income inequalities alone are able to account for the high levels of segregation in American urban areas. Although income inequality has a non-negligible effect on racial housing segregation, it in itself can only explain a small proportion of the overall segregation index. Our test suggests income redistribution to blacks will not be a very effective desegregation policy. And therefore, studies of segregation should focus on non-income factors such as residential preferences or discriminating behaviors.

The results of our Monte Carlo study are similar to previous findings using the indirect standardization method that completely ignores random factors. Thus, our study suggests that random residential choices do not cause much segregation. A possible explanation of this finding is that a neighborhood (census tract) in our sample is quite large so that the law of large numbers is at work. That is, the black-white ratio in each neighborhood is converging to the black-white ratio in the whole metro area. The finding raises new questions. For example, what happens if residential choices are not purely random but affected by the distance to school or availability of jobs? Does segregation increase a lot after adding such constraints on random moves? Further research in those directions using the same Monte Carlo method might be fruitful.

We have demonstrated that the Monte Carlo method is a useful tool for investigating hypotheses regarding residential segregation. The power of the method is not fully revealed here because we have focused on a fairly simple hypothesis. But the real world segregation process can be much more complicated. For example, it is quite possible that income inequalities and racial residential preferences reinforce each other to create severe segregation over time. However, researchers rarely deal with the residential preference hypothesis and the income inequality hypothesis simultaneously. Analytical tools commend that the two have to be studied one by one because a simultaneous treatment will make the analysis extremely difficult, if not completely intractable. The Monte Carlo technique makes many such complicated studies less formidable and hence opens up new areas for future research.

References

Barnard, G. A., (1963), "Discussion on Professor Bartlett's Paper," *Journal of the Royal Statistical Society*, Series B, 25, 294.

Besag, J. and P. J. Diggle, (1977), "Simple Monte Carlo Tests for Spatial Pattern," *Applied Statistics*, 26, 327-333.

Bleda, S. E., (1979), "Socioeconomic, Demographic, and Cultural Bases of Ethnic Residential Segregation," *Ethnicity*, 6, 147-167.

Cutler, D., E. Glaeser, and J. Vigdor, (1999), "The Rise and Decline of the American Ghettos," *Journal of Political Economy*, 107, 455-506.

Darroch, A. G. and W. G. Marston, (1971), "The Social Class Basis of Ethnic Residential Segregation: The Canadian Case," *American Journal of Sociology*, 77, 491-510.

Duncan, O. D., R. P. Cuzzort, and B. Duncan, (1961), *Statistical Geography*: *Problems of Analyzing Areal Data*, New York: Glencoe.

Farley, R., (1977), "Residential Segregation in Urbanized Areas of the United States in 1970: An Analysis of Social Class and Racial Differences," *Demography*, 14, 497-518.

Marshall, H. and R. Jiobu, (1975), "Residential Segregation in U.S. Cities: A Causal Analysis," *Social Forces*, 53, 449-460.

Massey, D. S., (1979), "Effects of Socioeconomic Factors on the Residential Segregation of Blacks and Spanish Americans in U.S. Urbanized Areas," *American Sociological Review*, 44, 1015-1022.

Massey, D. S., (1981), "Social Class and Ethnic Segregation: A Reconsideration of Methods and Conclusions," *American Sociological Review*, 46, 641-650.

Massey, D. S. and N. A. Denton, (1993), *American Apartheid: Segregation and the Making of the Underclass*, Cambridge: Harvard University Press.

Park, R. E., (1926), *The Urban Community as a Spatial Pattern and a Moral Order*, in *The Urban Community*, E. W. Burgess, ed., Chicago: University of Chicago Press.

Schelling, T. C., (1971), "Dynamic Models of Segregation," *Journal of Mathematical Sociology*, 1, 143-186.

Schnare, A., (1977), *Residential Segregation by Race in U.S. Metropolitan Areas*, Washington, D.C.: The Urban Institute.

Taeuber, K. E., (1968), "The Effect of Income Redistribution on Racial Residential Segregation," *Urban Affairs Quarterly*, 4, 5-14.

Yinger, J., (1995), *Closed Doors, Opportunities Lost*: *The Continuing Costs of Housing Discrimination*, New York: Russell Sage Foundation.