*A Revaluation of Indexes of Residential Segregation**

CHRISTOPHER WINSHIP, *Harvard University*

ABSTRACT

New criteria for indexes of residential segregation are developed. It is argued that a pattern of random segregation rather than complete desegregation should be used as a baseline for measuring segregation. It is shown that any index whose relationship to one baseline is independent of the proportion black in a city, necessarily has a dependent relationship with respect to the other baseline. The index of dissimilarity is adjusted to serve as a measure of deviation from random segregation. Eta-square, which was shown by Duncan and Duncan to depend on the proportion black, is shown to be independent of the proportion black when random segregation is used as a baseline. It is argued that segregation should be measured from a situation of complete desegregation when its effects are of concern, but that it should be measured from random segregation when its causes are being analyzed.

Over the years, sociologists have been continually interested in residential segregation. Much of this interest has been in the segregation of blacks from whites. Recently such inquiries have increased as sociologists have realized that residential segregation is the basis for other sorts of discrimination. In particular, residential segregation is the basis for much discrimination in both education and employment (Pettigrew).

From the years 1947 to 1955 a series of articles appeared in the literature, mostly in the *American Sociological Review*, concerned with developing various indexes of residential segregation. The standard dogma during this period was that there were many different indexes, each of which was appropriate for different theoretical reasons. The actual consequences of these articles were, however, quite the opposite. The outcome has been an overwhelming tendency to look at residential segregation from one perspective—that of the segregation curve and its deviation from complete desegregation—and the summary of this deviation in terms of the index of dissimilarity.

The purpose of the paper is to show that there are at least two different perspectives from which residential segregation can be examined. Segregation can be measured as it deviates from a situation of complete desegregation or in terms of a situation in which there is random segregation in the city.

The indexes that have been developed and those that will be examined in this paper assume that the segregation data are in a very specific form. It is assumed that the data are either the number of people or households that are black and that are white for either census tracts or blocks of a given city. In this paper it will be assumed that the data are reported in terms of households per block.

*I am in debt to the following people for help and suggestions: Thomas F. Pettigrew, Karl W. Taueber, Stanley S. Wasserman, Nancy Kreuger, Ann Young, and Louisa Amos. The research was supported in part by National Science Foundation Grant GS-2689 to Harvard University, and a National Science Foundation Graduate Fellowship.

INDEXES OF SEGREGATION

CRITERIA

From 1947 to 1955 a number of criteria were advanced for indexes of ecological segregation. Those suggested by Jahn et al. represent the criteria considered during that period. They are as follows:

> In addition to these two basic stipulations (i.e., of a maximum and minimum value), a satisfactory measure of ecological segregation should be (1) expressed as a single quantitative value so as to facilitate such statistical procedures as comparison, classification, and correlation; (2) be relatively easy to compute; (3) not be distorted by the size of the total population, the proportion Negroes, or the area of the city; (4) be generally applicable to all cities; and (5) differentiate degrees of segregation in such a way that the distribution of intermediate scores covers most of the range between the extremes of 0 and 100.

When the value of an index is "0," this represents the case where there is complete desegregation, i.e., each block has the same proportion black as the total population of the city. "100" or at times "1" represents the case where there is complete segregation, i.e., each block is either all black or all white.

There are two points made above that are important here. First is the position that segregation should be measured as a deviation from a situation in which there is complete desegregation. Second, there is the position that an index of segregation should not be biased by the proportion black in a city. Much of the debate during these years has been over the question whether a certain index was distorted by the proportion black; the essential problem was what was meant by the term "distortion." A solution to this problem was provided by Otis and Beverly Duncan in their article: "A Methodological Analysis of Segregation Indexes."

SEGREGATION CURVE

In their paper, Duncan and Duncan examined segregation indexes in terms of what they call a segregation curve. The segregation curve is constructed as follows: (1) the blocks are ordered in terms of the proportion black within each; (2) the percentage of the city's black population and white population within each block is computed; (3) the percentages are then cumulated starting with the block that has the greatest percentage black; (4) the cumulated percentages for blacks (X) and whites (Y) are then plotted to form the segregation curve. Table 1 and Figure 1 are a set of hypothetical data and their segregation curve. The diagonal in Figure 1 represents the situation in which there is complete desegregation. The X and Y lines represent the case of complete segregation.

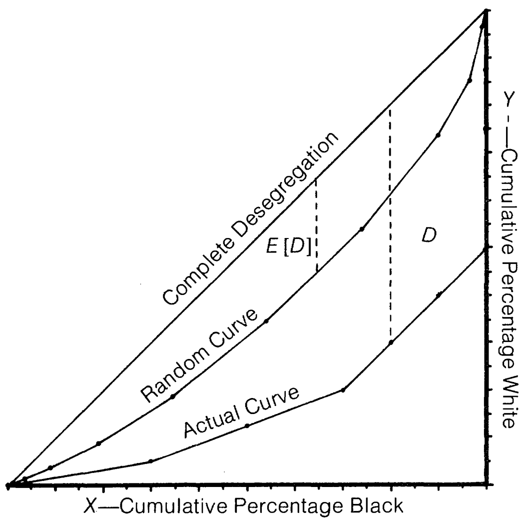
The significance of the Duncan and Duncan article is that they show that almost all the suggested indexes have a geometrical relationship to the segregation curve. What is important about this is that the geometrical relationship for some indexes is independent of the proportion black in the city, and for the others it is not. An index's geometrical relationship to the segregation curve provides a criterion for deciding whether it is biased by the proportion black in a city. The index of

Table 1. HYPOTHETICAL DATA FOR SEGREGATION CURVE

Block	Black Households	White Households	Percent of Blacks	Percent of Whites	Cumulative Black	Cumulative White
1	15	10	30	5	30	5
2	10	15	20	7.5	50	12.5
3	10	15	20	7.5	70	20
4	5	20	10	10	80	30
5	5	20	10	10	90	40
6	5	20	10	10	100	50
7	0	25	0	12.5	100	62.5
8	0	25	0	12.5	100	75
9	0	25	0	12.5	100	87.5
10	0	25	0	12.5	100	100

Total number of households	250	Percent black in population	20
Total number of black households	50	Index of dissimilarity	.50
Total number of white households	200	Expected value of index of dissimilarity	.20
Households per block	25	Adjusted index of dissimilarity	.30
Total number of blocks	10		

Figure 1. HYPOTHETICAL SEGREGATION CURVE



dissimilarity (index no. 4 Jahn et al.), the Gini index (index no. 3 Jahn et al.), and the generalized Cowgills' index (see Duncan and Duncan) are geometrically independent of the proportion black, whereas the "Nonwhite Ghetto Index" (index no. 1 Jahn et al.), the "Reproducibility Index" (Jahn), and eta-square (Bell) are dependent on it. The reader may refer to Duncan and Duncan for a more detailed explanation of these indexes and their relationship to the segregation curve.

INDEX OF DISSIMILARITY

Because of its relationship to the segregation curve and its interpretability, the index of dissimilarity has been very widely used. The index of dissimilarity is given by the formula $D = \sum Ti \mid Pi - P \mid / 2TP(1 - P)$ where Pi is the proportion black in block i, P is the proportion black in the city, Ti is the total number of households in block i, and T is the total number of households in the city. D is equal to the maximum distance between the diagonal and the segregation curve (see Figure 1).

The index of dissimilarity has been interpreted as the proportion of people who would have to move in order for the city to be completely desegregated. This is true only for a fairly peculiar interpretation of what we mean by moving. It is the proportion of black and/or white households that would have to move into the blocks that are disproportionately represented by the other group, assuming that no one in the other group moved out of these blocks. It seems a bit more reasonable to think about moving in terms of the total number of black and white households that would have to switch homes in order for a city to be completely desegregated. This number is given by $\sum Ti \mid Pi - P \mid$. Unfortunately, however, the number of households that have to move in a city is affected by the proportion black in the city. This can be seen by looking at the case of two completely segregated cities, one of which is 10 percent black and the other 50 percent black. In the first city no more than 20 percent of the households will have to move, whereas in the second city fully 50 percent of the households will have to move. The maximum number of households that would have to move for a city with a given proportion black is given by $2TP(1 - P)$. The index of dissimilarity is equal to the ratio betwen the number of households that actually have to move and the maximum number that would have to move for any distribution of housing. We have a new and more accurate interpretation of the index of dissimilarity. The index is equal to the proportion of households that still need to move in order to achieve complete desegregation.

RANDOM PATTERNS OF SEGREGATION

So far, segregation has been thought of as a deviation from complete desegregation. Alternatively we might look at desegregation as existing if the pattern of residential segregation is random. What do we mean by random? A residential housing pattern is random if households[1] choose their place of residence without regard to the racial composition of the neighborhood. In a city with a random housing pattern there may still be neighborhoods that are all black or all white. Only if people chose to live in neighborhoods that were completely desegregated, would we have a situation in which there was complete desegregation in the city as a whole. It should be noted, though, that even though blacks and whites may live in blocks that are totally segregated only by chance, this does not mean that this segregation will not have important effects on their lives.

A natural model for random segregation is the binomial distribution. If we assume households independently and randomly choose a house to live in within the city, the number of blacks in a block will be binomial variable with parameters Ti (the block size) and P (the proportion Negro in the city). If we assume constant block size then the curve representing random segregation can be parametrized in N as follows: $X = (1/TiP)\sum_{Wi=0}^{N}(Ti - Wi)b(Ti,\ 1-P)$ and $Y = [1/Ti(1-P)]\sum_{Wi=0}^{N} Wi\ b(Ti,\ 1-P)$, where N ranges in integer values from 0 to Ti, and Wi is the number of white households in a block. $b(Ti,\ 1-P)$ is the formula for the binomial distribution $\binom{Ti}{Wi}(1-P)^{Wi}Pi^{Ti-Wi}$. Different values of N give the inflexion points for the curve (see Figure 1). Inflexion points are connected by straight lines. A smooth curve can be gained by using the normal approximation to the binomial. The random curve for the case where $Ti = 25$ and $P = .2$ is shown in Figure 1.

One might argue that using the curve of random segregation as a baseline for measuring the degree of desegregation in a city instead of the diagonal of complete desegregation only amounts to a shift in the scale. For instance, if the index of dissimilarity is used as a measure then using the random curve as the base would only reduce all values by say .2. An index that was .5 when the diagonal was used would be equal to .3 if the curve of random segregation were used.

This is an important question. If true, it would mean that thinking about things in terms of random segregation instead of complete segregation only amounts to choosing a different zero point. We can answer this question by examining the expected value of D.

The expected value of D is $\sum Ti\ E[\,|Pi - P|\,]/2TP(1-P)$ where the expected value of $|Pi - P|$ is given by $\sum_{X=0}^{Ti}|(X/Ti) - P|\,b(Ti,\ P)$. For large Ti (>25) a normal approximation can be used. The expected value of $|Pi - P|$ is approximated by $2\sqrt{\{[Ti\,P(1-P)]/2\pi\}}$. Table 2 gives both the exact and approximate values of D for different blocksizes and proportion black when blocksize is assumed constant.

From Table 2 it is clear that the expected value of D varies by blocksize and the proportion black in a city. All indexes of segregation that have been suggested depend on block size. A city in which there is only one block is necessarily completely desegregated. A city which has only one person per block is necessarily completely segregated.

The fact that the expected value of D depends on the proportion black in a city is important. If we have two cities, each of which has a random pattern of housing, but which are different proportions black, then they will have different indexes of dissimilarity. Table 2 suggests that these differences can be fairly large. For instance if both cities had 25 households per block and one city was 10 percent black and the other was 50 percent black, then D would equal .272 for the first and .161 for the second.

Thinking about segregation in terms of random segregation is not just a matter of shifting the zero point. The expected value of D is equal to the maximum distance between the diagonal and the curve of random segregation. From Table 2 we know that this depends on the proportion black in a city. Thus if we measure segregation in terms of its deviation from the curve of random segregation, instead of the diagonal, we will get results that are essentially different.

Table 2. EXPECTED VALUES OF INDEX OF DISSIMILARITY WITH CONSTANT BLOCK SIZE *Ti*

Households per block		.1 − .9	.2 − .8	.3 − .7	.4 − .6	.5
10	exact	.387	.302	.267	.251	.246
	approximation	.421	.315	.272	.258	.252
25	exact	.272	.196	.176	.161	.161
	approximation	.266	.199	.174	.163	.160
50	exact	.185	.140	.122	.115	.112
	approximation	.188	.141	.123	.115	.113
100	exact	.131	.099	.087	.081	.080
	approximation	.133	.100	.087	.082	.080

$$\text{Exact} = 1/2Ti\ P(1-P)\ \sum_{B=0}^{Ti} |B - Ti\ P| \binom{Ti}{B} P^B (1-P)^{Ti-B}$$

$$\text{Approximation} = 1/\sqrt{[2\pi\ Ti\ P(1-P)]}$$

Another way to think about this is to recall the problem of whether an index is dependent on the proportion black in a city. Duncan and Duncan gave one answer to this problem which was discussed above. Our analysis above suggests another way to answer it. An index may be considered to be independent of the proportion black in a city if its relationship to the curve of random segregation is not dependent on the proportion black. A simpler criterion would be that its expected value not be dependent on the proportion black in a city.

The analysis of the expected value of D has shown that the relationship between the diagonal and the random curve depends on the proportion black in a city. This means that our two criteria for lack of bias with respect to the proportion black in a city are incompatible. Any index that is not biased with respect to one criterion is necessarily biased with respect to the other. One way to think of this is that we have two separate dimensions for residential segregation. One dimension is measured by comparing a city to a situation of complete desegregation, the other by comparing a city to random segregation.

ADJUSTING D

In order to make a comparison between a curve representing random segregation and a curve representing the actual segregation in a city, indexes need to be developed. The simplest suggestion is to adjust the index of dissimilarity by subtracting its expected value from it. This index will have a minimum of $-E[D]$ and a maximum of $1 - E[D]$. It will equal zero when the number of households that still need to be moved to accomplish complete desegregation is equal to the number of households that would need to be moved when there is random segregation. The index will be negative when the number of households that need to be moved is less than would be expected if segregation were random and will be positive when the number is greater. The index can be standardized to have a maximum of 1 by dividing by $1 - E[D]$. The formula for the case with constant blocksize is

$$D - E[D] / 1 - E[D] = \frac{\{[\sum Ti \mid Pi - P \mid / 2T\, P(1 - P)] - 1 / \sqrt{2\pi Ti\, P(1 - P)}\}}{[1 - 1 / \sqrt{2\pi Ti\, P(1 - P)}].}$$

Geometrically the index is just the difference between the maximum distance from the actual curve to the diagonal and the maximum distance from the random curve to the diagonal divided by one minus the maximum distance from the random curve to the diagonal.

OTHER INDEXES

Earlier we gave a list of indexes that were biased by the Duncan and Duncan criteria. Are these indexes also biased with respect to the random curve? One index will be examined here, eta-square.

Eta-square is the square of the correlation ratio of color on block and is identical with the square of the mean square contingency coefficient phi. By definition it is the variance between block proportions divided by the total variance of the population: $\text{Eta}^2 = \sum Ti(1 - Pi)^2 / TP(1 - P)^2 - (1 - P)/P$. It yields a value of 1 for complete segregation and a value of 0 for complete desegregation. Since eta-square is non-linear it has no simple geometrical relationship with the random curve. We can however evaluate its expected value. Its expected value is $1/Ti$ where Ti is the mean block size for the city. Eta-square's expected value does not depend on the proportion black. We may think of eta-square as being an index measuring the departure of a city from a pattern of random segregation except that it has a zero when there is complete desegregation and a value of $1/Ti$ when there is random segregation. If desired, eta-square can be adjusted in the same way that the index of dissimilarity was.

In his article "A Probability Model for the Measurement of Ecological Segregation," Bell shows that eta-square is just a standardized probability: $\text{eta}^2 = (H - 1 + P)/P$. H represents the probability that the next person from his block that a random black will meet is another black. In a weak sense, H characterizes the nature of the minority group's encounters.

IMPLICATIONS

We have argued that there are two dimensions underlying residential segregation. In reality how different are these two dimensions and how are we to use them? The adjusted index of dissimilarity has not been calculated for any data. The index of dissimilarity and eta-square have been calculated on the same data. Taueber and

Taueber report a correlation of .19 between the two indexes. Zoloth reports correlations of .87 and $-.24$. Choice of index does make a substantial difference in analyzing data.

We now have two different ways of measuring segregation, but how are we to use them? Adopting an unqualified position, it is appropriate to measure residential segregation as a deviation from complete desegregation if we are concerned with the *effects* of segregation. Alternatively it is appropriate to measure segregation as a deviation from random segregation if we are concerned with the *causes* of segregation.

If we are interested in the effects of residential segregation then it makes little difference whether segregation is random or nonrandom. Residential segregation may affect blacks in terms of the availability of jobs and schooling. What is important is the degree of their isolation and not the mechanisms underlying it. In this case the index of dissimilarity is an appropriate measure.

If we are interested in the causes of segregation then it is appropriate to use a random model as a baseline. Residential segregation may be determined by the reluctance of whites to live with blacks, discriminating real estate brokers and /or inequalities in income. If the effects of these variables are to be determined, then it is necessary to know how much segregation there would be if people chose where they wanted to live randomly. In this case the adjusted index of dissimilarity would be the appropriate measure.

In terms of policy both views are important. In designing policy it is important to gauge residential segregation in terms of its effects. In making comparisons between cities in which various programs have been implemented, we will want to know what effect this has had on blacks. At the same time, in implementing policies, especially those aimed at preventing whites from being discriminating in their choices, it is important to use a random model as a baseline. In any policy program it is important to know how much segregation there would be if households were not discriminating. From a policy point of view both the cause and effect side of residential segregation need to be examined.

CONCLUSIONS

Two different ways of measuring segregation have been examined—as a deviation from complete desegregation and as a deviation from random segregation. It was shown that these two viewpoints imply two different and contradictory ways of deciding whether an index of segregation is biased by the proportion black. If an index is not biased with respect to one baseline then it must be biased with respect to the other. It was then argued that segregation should be measured from the diagonal of complete desegregation if we are interested in the effects of segregation and from the random curve if we are interested in causes. In the former case the index of dissimilarity is an appropriate measure and in the latter, the adjusted index of dissimilarity.

NOTE

1. It makes little sense to assume that individuals independently choose houses to live in within a city. For this reason we have restricted our discussion to households.

REFERENCES

Bell, W. 1954. "A Probability Model for the Measurement of Ecological Segregation." *Social Forces* 32(May):357–64.

Duncan, O. D., and B. Duncan. 1955. "A Methodological Analysis of Segregation Indexes." *American Sociological Review* 20:210–17.

Jahn, A. 1950. "The Measurement of Ecological Segregation: Derivation of an Index Based on the Criterion of Reproducibility." *American Sociological Review* 15:100–4.

Jahn, A., F. Schmid, and C. Schrag. 1947. "The Measurement of Ecological Segregation." *American Sociological Review* 12:293–303.

Pettigrew, Thomas F. (ed.). 1975. *Racial Discrimination in America*. New York: McGraw-Hill.

Taueber, Karl E., and Alma F. Taueber. 1965. *Negroes in Cities*. Chicago: Aldine.

Zoloth, B. S. 1974. "An Investigation of Alternative Measures of School Segregation." Institute for Research on Poverty Discussions Papers. Madison: University of Wisconsin.