Chapter 16
Comparing Behavior of Unbiased and Standard Versions of Popular Indices

In the previous chapter I outlined the rationale for unbiased versions of indices of uneven distribution. Additionally, I presented results from analysis of expected group residential distributions under a binomial probability model to establish that the unbiased versions of popular indices have expected values of zero when residential distributions are random. In this chapter I report analyses of the behavior of standard and unbiased versions of indices of uneven distribution to document two things: the potential undesirable impact of bias on the scores of standard versions of indices and the attractive behavior of the scores of unbiased versions of the same indices.[1]

To document index behavior I conducted a series of simulation experiments to systematically "exercise" standard and unbiased versions of popular indices under a wide range of demographic contexts and neighborhood definitions. I performed the analyses using residential distributions generated by SimSeg, a computational model that simulates residential segregation dynamics. The SimSeg program has been described in more detail elsewhere (e.g., Fossett and Waren 2005; Fossett 2006, 2011a, b; Fossett and Dietrich 2009; Clark and Fossett 2008). Examining results generated by the SimSeg program is useful for the purposes of this chapter for two reasons. First, the program implements routines that calculate both standard and unbiased versions of G, D, R, H, and S. Second, the program can systematically generate residential distributions over a wide range of study designs that can reveal how the behavior of standard and unbiased versions of indices differ under varying circumstances.

Using SimSeg I designed and executed simulation experiments that implemented a two-group city in which segregation is assessed using bounded neighborhoods of uniform size. The two groups in the simulation are of course "virtual", but for convenience of discussion and consistency with examples discussed in earlier chapters

---

[1] The analyses I report in this chapter elaborate and extend analyses I conducted for an earlier study on this topic (Fossett and Zhang 2011).

I refer to them as "White" and "Black". I varied the conditions of the experiments to exercise index behavior by varying the racial mix of the city randomly from 2 to 98 % White separately in each experiment. I then ran 2,500 experiments separately for each of eight neighborhood sizes based on a square housing grid for the bounded area ranging from 3 to 10 houses on a side. The resulting neighborhood sizes were 9, 16, 25, 36, 49, 64, 81, 100, and 225. The simulation experiments conducted using these varying settings for neighborhood size and city racial composition were relatively simple.[2] The program first created the relevant virtual neighborhoods and housing units within them. Next it created the virtual population of households according to the racial demography setting. It then distributed households distributed randomly across housing units. Then, it calculated and recorded a battery of segregation index scores including scores for standard versions of all popular measures of uneven distribution G, D, A, R, H, and S and unbiased versions for G, D, R, H, and S.[3]

Tables 16.1 and 16.2 report the means and standard deviations for scores for standard versions of indices of uneven distribution under random distribution at the initialization of the city landscape over varying conditions of effective neighborhood size (ENS) and percent White for the city (P). For economy of presentation, results are given only for ENS settings of 9, 16, 25, 49, 100, and 225. Inspection of Table 16.1 shows that the level and pattern of index scores varies systematically by index and over different combinations of settings for ENS and P. Inspection of the results presented in Table 16.2 shows that index scores vary in a relatively narrow range around the mean for index scores under any particular combination of settings for ENS and P and the results also show that the degree of dispersion in index scores is generally similar in magnitude across different indices.

Figure 16.1 provides visual documentation of the patterns of index behavior summarized in Tables 16.1 and 16.2. The figure provides separate graphs for each index considered; namely, G, D, A, R, H, and S. Each graph plots the values of the relevant index score calculated from the random residential distribution at the beginning of the simulation experiment (i.e., cycle 0) against percent White in the city population (P). The graphs plot index scores for the simulations where effective neighborhood size (ENS) is set to value of 9, 16, 25, 49, and 100. In addition, each graph also plots a black line tracing the expected index score (e.g., E[D]) based on calculations using a binomial model (per Winship 1977). To reduce visual clutter and facilitate clarity of patterns, the graphs do not depict results for ENS settings of

---

[2] Other details of the simulations are uniform across all simulations and have no impact on results. For example, neighborhoods are arranged to form an approximately circular form for the overall city. The dimensions of the city were calibrated to yield between 6400 and 8500 virtual households depending on number of households per neighborhood and the number of neighborhoods in the simulated city. The resulting virtual cities are similar in form to those described in Fossett (2006, 2011a, b). I conducted additional simulations using larger cities with more neighborhoods and more virtual households. All relevant index behavior was fundamentally similar. So I used smaller cities to keep the computational burdens for generating the analysis data sets at reasonable levels.

[3] Results for an unbiased version of Atkinson's A are not shown because I have not been able to place this index in the difference of means framework. Hutchens R is a closely related measure.

Table 16.1 Means for standard versions of popular indices of uneven distribution computed for random residential distributions under varying combinations of relative group size (P) and neighborhood size

Index	P[a]	Neighborhood size					
		9	16	25	49	100	225
Gini (G)	≤5	81.4	70.5	60.1	45.2	33.0	22.0
	11–15	56.5	43.7	35.1	25.1	17.6	11.7
	36–50	40.2	30.1	24.0	17.1	12.0	8.0
Dissimilarity (D)	≤5	77.7	62.8	48.0	33.4	24.0	15.6
	11–15	40.8	31.6	25.2	17.9	12.5	8.3
	36–50	28.7	21.4	17.1	12.2	8.5	5.7
Atkinson[b] (A)	≤5	78.4	64.4	49.8	28.4	12.9	4.5
	11–15	41.9	23.3	13.0	5.5	2.6	1.1
	36–50	14.5	7.5	4.7	2.4	1.1	0.5
Hutchens (R)	≤5	54.0	40.8	29.6	15.6	6.7	2.3
	11–15	23.8	12.4	6.7	2.8	1.3	0.5
	36–50	7.6	3.8	2.4	1.2	0.6	0.3
Theil (H)	≤5	34.8	24.1	16.9	9.0	4.4	1.8
	11–15	18.8	10.5	6.4	3.1	1.5	0.6
	36–50	9.9	5.3	3.3	1.7	0.8	0.4
Separation (S)	≤5	12.3	6.9	4.4	2.2	1.1	0.5
	11–15	12.3	7.0	4.4	2.3	1.1	0.5
	36–50	12.3	6.9	4.4	2.3	1.1	0.5

[a]Here P denotes the city-wide group percentage for the smaller group in the comparison
[b]Atkinson index (A) is computed with δ set at 0.5, the value at which A is "symmetric"

36, 64, 81, and 225. However, Table 16.2 documents that the results for these settings are consistent with the results shown in the figure. For example, the means for index scores when ENS is set at 36, 64, and 81 fall between the scores for ENS settings immediately above and below the ENS setting in question.

The results presented in Fig. 16.1 document several clear patterns. First, all of the indices take values above zero in each and every simulation trial reflected in the 12,500 data points plotted in the figure. The gray points for individual simulation trials indicate that index scores calculated from the random residential distributions at initialization in individual simulation trials vary in relatively narrow ranges around their expected values based on binomial theory. The Black lines show that the expected values of the indices based on analytic calculations vary systematically with effective neighborhood size and percent White in the city population. As noted earlier, the nature of the systematic variation in index scores is simple in its main features. For all indices, scores for both the expected values under random assignment and the observed random segregation at initialization in the simulations are systematically higher when effective neighborhood size (ENS) is lower. Thus, the highest curve is for the set of simulations that use the lowest value of ENS (in this case 9) and the curves move systematically lower as ENS moves to successively higher values. Also, for all indices *except the separation index (S)*, both the expected

Table 16.2 Standard deviations for standard versions of popular indices of uneven distribution computed for random residential distributions under varying combinations of relative group size (P) and neighborhood size

Index	P[a]	Neighborhood size					
		9	16	25	49	100	225
Gini (G)	≤5	4.9	6.6	7.2	7.4	5.7	3.9
	11–15	2.4	2.4	2.4	2.1	1.4	1.0
	36–50	1.2	1.2	1.2	1.2	0.9	0.6
Dissimilarity (D)	≤5	6.4	9.6	10.3	5.9	4.2	2.8
	11–15	1.8	2.2	1.9	1.5	1.1	0.8
	36–50	0.9	0.9	0.9	0.9	0.6	0.4
Atkinson[b] (A)	≤5	6.1	8.9	10.4	10.7	6.3	2.0
	11–15	3.8	3.3	2.5	1.1	0.4	0.2
	36–50	1.2	0.6	0.5	0.3	0.2	0.1
Hutchens (R)	≤5	6.7	7.6	7.5	6.5	3.4	1.0
	11–15	2.5	1.9	1.3	0.6	0.2	0.1
	36–50	0.6	0.3	0.2	0.2	0.1	0.1
Theil (H)	≤5	3.9	3.9	3.5	2.7	1.5	0.6
	11–15	1.4	1.1	0.9	0.5	0.2	0.1
	36–50	0.6	0.4	0.3	0.2	0.1	0.1
Separation (S)	≤5	0.7	0.5	0.4	0.3	0.2	0.1
	11–15	0.6	0.5	0.4	0.3	0.2	0.1
	36–50	0.7	0.5	0.4	0.3	0.2	0.1

[a]P is the city-wide, pairwise group percentage for the smaller group in the comparison
[b]The Atkinson index (A) is computed with "tuning" value δ set at 0.5, the value at which A is symmetric

values and the observed random outcomes are systematically higher when proportion White for the city (P) departs from balance at 0.50 and the expected values and observed outcomes take especially high values when P falls below 0.10 or rises above 0.90.

The existing methodological literature has documented similar patterns of random variation for D many times before and also occasionally for G. But reports on patterns of variation for expected values of A, R, H, and S under random assignment are rare if they exist at all. To my knowledge, the results presented here are the first to systematically compare the bias behavior of all popular indices of uneven distribution.

Comparing the figures for each index reveals several noteworthy differences in their behavior under random assignment. One obvious pattern is that indices vary considerably in the magnitude of bias under random assignment. The highest expected values under random assignment are observed for G followed closely by D and then A. The lowest scores under random assignment are for S. H and R have the next lowest scores for expected values. The "takeaway" point here is that D, the most popular and widely used index of uneven distribution has higher expected values under random assignment than all other indices except G.
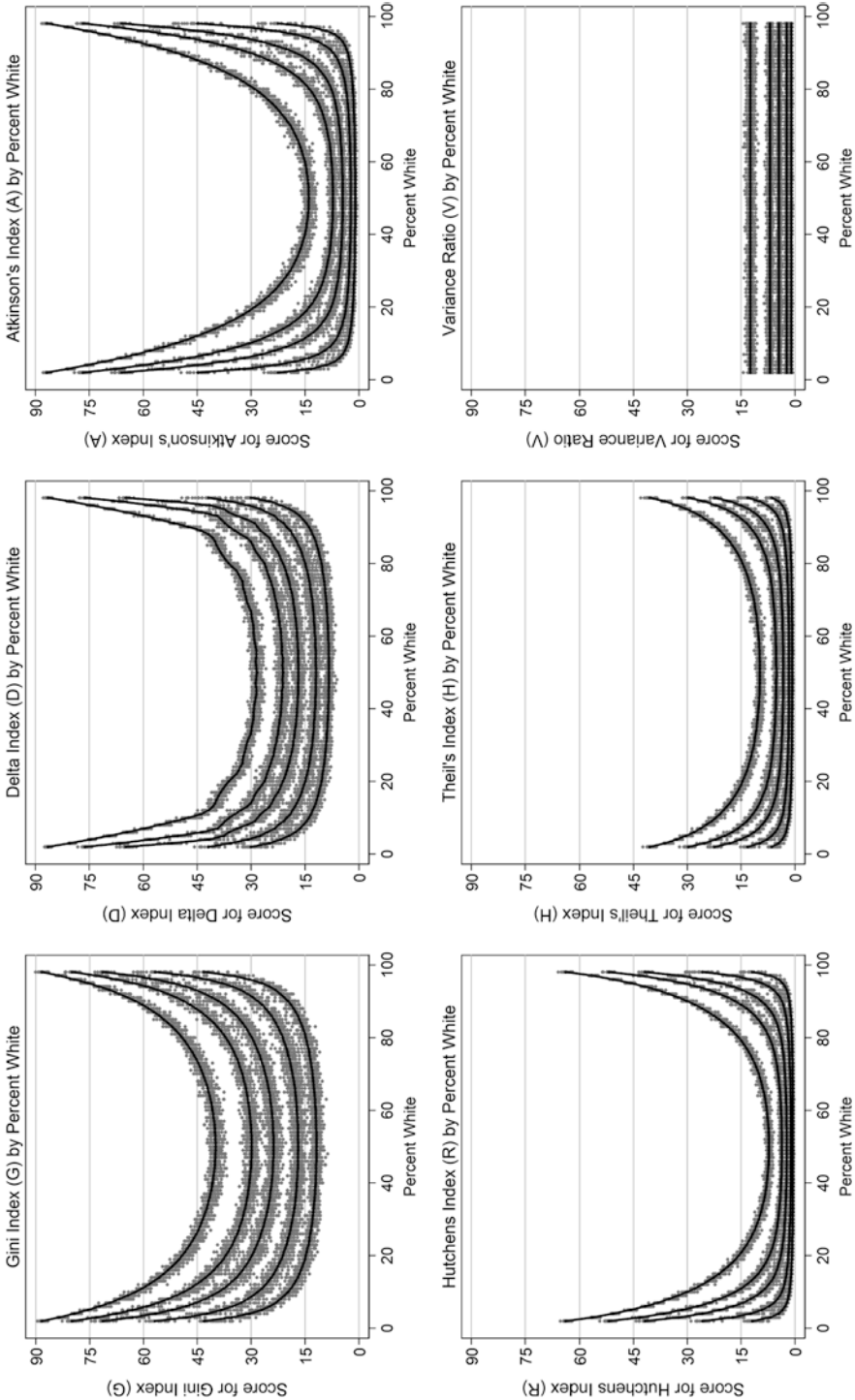
Fig. 16.1 Scores for "Standard" versions of indices of uneven distribution under random assignment by city percent White and neighborhood size (Note: Points plotted in *light gray* are values calculated from random residential distributions. Points plotted in *black* are expected values (e.g., E[D]). Values for effective neighborhood size (ENS) are 9, 16, 25, 49, and 100)

Another clear pattern is that expected values under random assignment are lower for every index when effective neighborhood size (ENS) is larger. One additional finding is that, for all indices except S, index bias is highest, often alarmingly so, when group size is imbalanced. These findings provide at least some justification for two crude rules-of-thumb for research designs used in many segregation studies. One practice is that most studies in recent decades examine segregation scores calculated using data for spatial units with larger population counts (e.g., use tracts over blocks). This tends to promote, but does not guarantee, higher levels of effective neighborhood size which, all else equal, serves to reduce bias. Another practice is that studies often avoid analysis of comparisons involving groups that are small in relative population size. All else equal, this tends to exclude comparisons where bias is likely to be larger.

Another common practice in the empirical literature is to avoid analysis of comparisons that involve groups that are small in absolute population size. The results presented here provide no support for this practice. Analytic formulas for bias (e.g., Winship 1977) identify a clear role of neighborhood size and relative group size but they do not identify a role for absolute group size. Empirically, absolute size may be correlated with relative group size but only relative size has a consequence for index bias. So if one is screening cases on relative group size there is no justification for additional screening on absolute size, at least not for the purpose of avoiding problematic bias.[4]

Similarly, there is no support in these results for the practice of "dealing with bias" by weighting cases in aggregate-level analyses by the size of the minority population. Absolute group size has no bearing on bias. Accordingly, weighting cases on minority size serves only to skew results toward findings for cities with larger minority populations.

Figure 16.1 also reveals a few findings that are not currently widely appreciated. One is that *for most indices, and especially for G and D, effective neighborhood size (ENS) and group ratio (GR) interact such that index bias is especially high when ENS is low and GR is highly imbalanced*. This has an important practical implication. It indicates that the standard rules-of-thumb commonly used in restricting analysis samples in empirical studies are crude and are not necessarily reliable for their intended purpose of identifying cases prone to high levels of bias. The standard rules of thumb are crude first because they applied using a "rough-and-ready" cut points when bias behavior varies continuously across ENS and GR and second because the rules are applied in a simple additive way and do not take account of the important interaction between ENS and GR that is so clear in these results. As a result, the prevailing practices can easily exclude cases where bias may low enough to be viewed as negligible (e.g., $E[\bullet] < 2 - 3$); particularly when using R, H, and

---

[4] In Chap. 8 I noted that screening on absolute group size may be relevant for other reasons. For example, if group size is insufficient to "fill" 3–5 of the spatial units used in measuring segregation, scores for indices that measure residential polarization will likely be biased downward because the spatial units being used may be too large to capture concentrated displacement for the group in question.

S. Conversely, they can sometimes include cases where bias is high and problematic; particularly when using G, D, and A.

In sum, not only do current practices for dealing with bias greatly restrict the scope of segregation studies, they also are likely to be less reliable and effective for their intended purpose than researchers may realize. If researchers apply these practices in future research, they should revise them to take account of the findings reported here.

16.1   Documenting the Attractive Behavior of Unbiased Versions of Indices of Uneven Distribution

I now review the behavior of the new *unbiased* versions of popular indices of uneven distribution under random distribution at the initialization of the city landscape. Tables 16.3 and 16.4 report the means and standard deviations, respectively, for the sampling distributions of scores for the unbiased versions of the indices over the simulations conducted over varying conditions of effective neighborhood size (ENS) and percent White in the city (P). Figure 16.2 documents these patterns visually with separate graphs for G′, D′, R′, H′, and S′.[5] As with Fig. 16.1, each graph plots the values of the relevant index score at the beginning of the simulation experiment (i.e., cycle 0) against percent White in the city population. Also as before the individual graphs plot observed segregation outcomes from simulations in which effective neighborhood size (ENS) is variously set to 9, 16, 25, 49, and 100. I should note two important differences from Fig. 16.1. One is that the expected values of the unbiased indices (e.g., E[D′]) all are zero under calculations using an "exact" binomial model (per Winship 1977). So the resulting plotted "curve" for the expected values for all of the indices is a horizontal straight line centered on zero on the vertical (y) axis of the figure. The other is that the vertical range of the "y" axis of the figures is covers a much smaller range of scores than in Fig. 16.1. This aids in making visual inspection of patterns in Fig. 16.2. But it is important to take account of the difference when making visual comparisons with Fig. 16.1. The range of variation is much smaller in Fig. 16.2 but this is not visually obvious.

The graphs in Fig. 16.2 show that the unbiased index scores based on the 12,500 random residential distributions vary in an approximately bell-shaped distribution around zero and thus take both negative and positive values. The vertical dispersion of unbiased index scores around the expected value of zero gives intuitive insight into the expected sampling distribution of the scores for the unbiased versions of the different indices. The dispersion depicts the range and pattern of index scores that occur when there is no statistical association between race and residential location; that is when residential distributions are random. Intuitively, this provides a basis

---

[5] There is no graph for Atkinson's A′ because I have not been able to place it in the difference of means framework. Hutchens R is a closely related index with an available difference of means formulation needed to develop and unbiased version of the index.

Table 16.3 Means for unbiased versions of popular indices of uneven distribution computed for random residential distributions under varying combinations of relative group size (P) and neighborhood size

Index	Pa	Neighborhood size					
		9	16	25	49	100	225
Gini (G)	≤5	−0.1	−0.5	−0.4	−0.7	−1.0	−1.5
	11–15	−0.2	0.2	−0.5	−0.6	−0.7	−1.1
	36–50	−0.2	−0.4	−0.5	−0.8	−0.8	−1.0
Dissimilarity (D)	≤5	−0.0	−0.4	−0.4	−0.8	−0.2	−0.6
	11–15	−0.0	0.1	−0.1	0.3	−0.1	0.1
	36–50	−0.2	−0.2	−0.1	−0.1	−0.1	0.1
Hutchens (R)	≤5	−0.1	−0.2	−0.1	−0.1	−0.1	−0.1
	11–15	−0.1	0.2	−0.0	0.0	0.0	−0.0
	36–50	−0.0	−0.0	−0.0	−0.1	−0.0	0.0
Theil (H)	≤5	−0.0	−0.1	−0.1	−0.1	−0.1	−0.1
	11–15	−0.0	0.1	−0.0	0.0	0.0	−0.0
	36–50	−0.0	−0.0	−0.0	−0.0	−0.0	−0.0
Separation (S)	≤5	−0.0	−0.0	−0.0	−0.0	−0.0	−0.0
	11–15	−0.0	0.0	−0.0	−0.0	0.0	−0.0
	36–50	−0.0	−0.0	−0.0	−0.0	−0.0	−0.0

[a]P is the city-wide, pairwise group percentage for the smaller group in the comparison

Table 16.4 Standard deviations for unbiased versions of popular indices of uneven distribution computed for random residential distributions under varying combinations of relative group size (P) and neighborhood size

Index	Pa	Neighborhood size					
		9	16	25	49	100	225
Gini (G)	≤5	4.9	5.3	5.8	6.7	5.2	3.2
	11–15	3.3	3.5	3.6	3.6	2.5	1.9
	36–50	2.4	2.4	2.4	2.4	1.8	1.2
Dissimilarity (D)	≤5	4.8	5.4	5.8	6.9	6.1	4.3
	11–15	3.3	3.6	3.9	4.6	3.5	3.3
	36–50	2.5	2.7	2.9	3.7	3.0	2.6
Hutchens (R)	≤5	3.4	3.3	3.2	3.0	1.6	0.5
	11–15	1.7	1.3	0.9	0.5	0.2	0.1
	36–50	0.5	0.3	0.2	0.2	0.1	0.0
Theil (H)	≤5	2.2	1.9	1.7	1.4	0.8	0.3
	11–15	1.2	0.9	0.7	0.4	0.2	0.1
	36–50	0.6	0.4	0.3	0.2	0.1	0.1
Separation (S)	≤5	0.8	0.5	0.4	0.3	0.2	0.1
	11–15	0.7	0.6	0.5	0.3	0.2	0.1
	36–50	0.8	0.5	0.4	0.3	0.2	0.1

[a]P is the city-wide, pairwise group percentage for the smaller group in the comparison
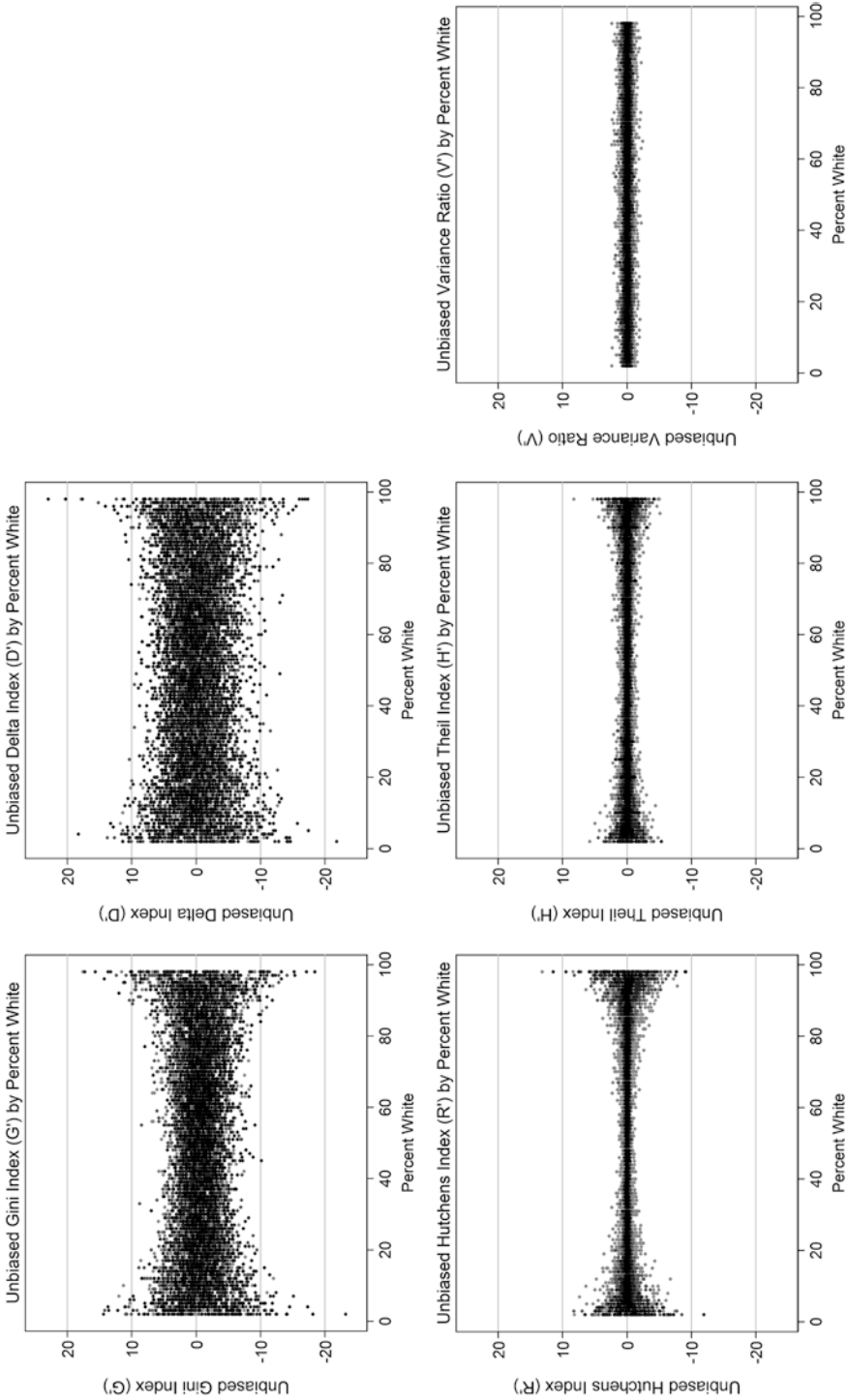
Fig. 16.2 Scores for unbiased versions of indices of uneven distribution under random assignment by percent White and neighborhood size (Note: Values for effective neighborhood size (ENS) are 9, 16, 25, 49, and 100. Cases for higher values of ENS are plotted in darker shades)

for evaluating observed scores for residential segregation. Observed scores that fall within the middle portion of the sampling distribution can easily occur by chance. But chance is a less plausible explanation for observed scores that fall in the low probability tails of the sampling distribution. Accordingly, scores in these regions are likely to reflect the impact of structured social processes that promote either greater or lesser segregation than would occur based on chance.

Because the expected value for an unbiased index under the null hypothesis of no association between race and residential location is zero and the sampling distribution is bell-shaped, one half of the values in the sampling distribution of an unbiased index will be negative. Some segregation researchers may not be initially comfortable with seeing negative scores for unbiased indices. But negative scores have a straightforward interpretation on both narrow statistical grounds and also on substantive grounds. On statistical grounds *negative scores indicate that scores for the standard version of the index take values that are lower than would be expected under random assignment*. Under the null hypothesis, negative values that fall in the middle region (e.g., in the middle 95 % region) of the sampling distribution for unbiased index scores can be set aside in the usual way; they can be attributed to chance and the observed departure from the expected value of zero can be viewed as not statistically significant. In contrast, negative scores that fall in the left tails of the sampling distribution can be viewed as statistically significant; they are unlikely to occur by chance and thus invite a substantive sociological explanation of how (scaled pairwise) contact with Whites among neighbors could come to be higher on average for Blacks than for Whites.

I note below that interesting sociological explanations are available. But I first pause to note that unbiased indices necessarily take negative values under exact even distribution. For example, consider the values of the standard and unbiased versions of the separation index for a city that is 90 % White and 10 % Black and has exactly 10 households per block. Under exact even distribution every block will have nine White households and one Black household. Proportion White among *neighbors* differs by race and will be 0.889 (i.e., 8/9) for every White household and 1.000 (i.e., 9/9) for every Black household. In contrast, proportion White for area population will be 0.900 (i.e., 9/10) for every White and every Black household. Accordingly, the standard version of S will be zero but the unbiased version S′ will be −0.111.

The comparison on D would be even more extreme. The value of the standard version of D would again be zero. But the value of the unbiased version D′ would be–1.000 because all White households are scored 0 on attaining parity (i.e., 0.90 or higher) on proportion White among neighbors while all Black households are scored 1.[6]

These negative values for unbiased indices under conditions of exact even distribution will be unfamiliar and perhaps also surprising to most readers, but they are

---

[6] The example serves to highlight a difference between D and S; namely, that, whether in standard or unbiased form, D responds much more strongly than S to quantitatively small deviations from parity on racial proportions.

fully expected and have a clear substantive interpretation. Negative values result because exact even distribution – the zero point for standard measures of uneven distribution – is a highly unexpected outcome under random distribution. The occurrence of such an unexpected residential distribution invites a sociological explanation identifying the structured social process that could bring about exact even distribution. Ready examples could include social dynamics such as quota systems in state policies governing assignments of households to housing units or institutional housing policies that structure housing assignments in dorms at colleges and universities, public housing, barracks in military bases, juvenile detention facilities, jails and prisons, orphanages, institutions for persons with disabilities, and the like. Thus, statistically significant negative values for unbiased indices are not only possible, they can and should obtain in certain empirical settings (albeit not ones that are commonly studied) where group distributions are highly structured to produce even distribution. Thus, negative scores for unbiased indices are valid and carry a clear sociological meaning.

Table 16.4 and Fig. 16.2 document patterns of dispersion in scores for unbiased indices under random distribution. The main differences across the five unbiased indices are seen in three areas. The first is the general level of volatility in the dispersion of scores around the expected value of zero. Holding simulation conditions constant, scores for G′ and D′ consistently exhibit greater variability under random assignment; scores for R′ and H′ exhibit less variability; and scores for S′ exhibit the lowest variability of all.

Another interesting pattern in the sampling distributions of the unbiased indices is how the dispersion of index scores under random distributions varies with effective neighborhood size (ENS). Table 16.4 documents that variability in the distribution of scores around zero is greater when effective neighborhood size (ENS) is small. This pattern is highlighted in visual form in Fig. 16.2 by plotting the points in successively darker shades of gray as ENS increases in size from 9–16 to 25–49 to 100 producing a concentration of the darkest points near the center of the distribution.

A third pattern in the sampling distributions of the unbiased indices is how the dispersion of index scores under random distributions varies with city racial proportion; in this case proportion White in the city (P). Here the unbiased separation index (S′) stands apart from the other indices. Other things equal, the dispersion in the scores for S′ is constant across levels of percent White in the city (P). In contrast, a much different pattern holds for G′, D′, R′, and H′; they all exhibit greater dispersion in index scores when percent White in the city (P) departs further from balance (i.e., 50). Figure 16.2 documents that the increase in the magnitude of the dispersion in index scores becomes especially pronounced when P begins to approach the bounds of 0 and 100.

I offer the following intuitive explanation for these patterns. The pattern of dispersion in values of the unbiased version of the separation index (S′) serves as a ready benchmark. Variation in dispersion is a simple function of effective neighborhood size. This is easy to understand; smaller samples of neighbors lead to greater volatility in residential outcomes. Dispersion in S′ is unaffected by relative group

size because values of unbiased contact (p′) map on segregation-determining scores for residential outcomes (y′) without change. For all other indices, the scaling functions mapping scores of p′ onto scores of y′ are nonlinear. This assures that random deviations of p′ from P will be exaggerated. Furthermore, because nonlinearity in the scaling functions is stronger when group size is imbalanced, the impact will be greater when group size is more imbalanced.

Finally, it is important to note that Fig. 16.2 documents that scores for unbiased indices are distributed symmetrically around zero at all levels of effective neighborhood size (ENS) and all levels of percent White for the city (P). So, while the magnitude of dispersion for scores for unbiased indices varies across indices and over study conditions, the expected value (zero) and shape of dispersion in scores (symmetrical and bell-shaped) remain constant for all of the indices.

16.1.1   Summary of Behavior of Unbiased Indices

In sum, under random distribution, dispersion in scores of unbiased indices varies in magnitude depending on the particular index, the value of effective neighborhood size (ENS), and, with the lone exception of S′, percent White in the city (P). These patterns indicate that one must be mindful of these distinctive sampling distributions for different indices when evaluating the statistical significance of particular index scores. Exact analytic solutions for standard errors of unbiased index scores under varying circumstances have not yet been established. For exploratory analysis "t" and "Z" tests for group differences of means on scaled contact with the reference group may perhaps serve as reasonable approximations. For more definitive assessments, researchers should use bootstrapping or other similar computation-intensive approaches that require less stringent assumptions regarding the nature of error distributions.

16.2   Documenting Additional Desirable Behavior of Unbiased Indices Based on the Difference of Means Formulation

I now review the behavior of standard and unbiased versions of popular indices of uneven distribution in multi-group situations. My purpose is to show that "norming" adjustments proposed by Winship (1977) and Carrington and Troske (1997) and discussed in Chap. 14 can be problematic in these situations while the unbiased indices that I introduce here behave in desirable ways.

The essence of the problem with norming adjustments is that the expected values of indices under random assignment are more complicated in multi-group situations than previous methodological discussions have acknowledged. The logic of per-

forming "norming" adjustments proposed previously in the literature rests on the crucial assumption that the expected value of standard indices under random distribution is invariant (is a constant) under a given combination of area size and (pairwise) group proportions. Unfortunately, this assumption is not correct. Instead, the expected value of standard indices is uncertain and can vary substantially even when area size and group proportions are known and simple in nature (e.g., all areas are constant size). The variation in index behavior traces to the presence of other groups in the population; the residential distributions for these groups can have nontrivial impacts on expected values of standard indices. This possibility ultimately undermines the potential effectiveness of previously proposed procedures for performing norming adjustments to deal with the impact of bias on the scores of standard indices.

I present results from simulation analyses conducted using the SimSeg simulation model to highlight the complex problems of bias in standard indices. The simulations all involve three groups; one large minority group, and two smaller minority groups. At the initialization of each simulation trial the households in the majority group are highly segregated from the households in the two minority groups but the households in the two minority groups are randomly distributed in relation to each other. This is depicted in the top panel in Fig. 16.3.[7] The simulation is then run for ten cycles (i.e., time periods). During each cycle, 25 % of households are chosen at random and are assigned randomly to a new residential location. Not surprisingly, systematic segregation between the majority group and the two minority groups quickly dissipates under this process of random movement resulting in majority households being randomly intermixed with minority households. This is depicted in the bottom panel in Fig. 16.3. At all times, starting at initialization and continuing to conclusion, the households in the two minority groups are randomly distributed in relation to each other.

The simulation experiments I used to generate the results for the analysis here follow the general design used in the simulations described earlier. The simulations here use the same neighborhood size (25) and the same city size and area configuration (i.e., 256 areas and 6,400 housing units). The racial composition of the city is set at 80-10-10. A total of 2,500 separate simulation experiments are run using this setting.

Index behavior is depicted in Fig. 16.4 which provides four graphs, two on the top row for the unbiased formulation of the dissimilarity index (D′) and two on the bottom row for the standard formulation of the dissimilarity index (D). The graphs in the left column depict majority-minority segregation; the graphs in the right column depict minority-minority segregation. The box plots in the top left graph show how D′ for the majority-minority comparison starts at very high levels and falls to zero as the ten cycles of random movement dissipate the initial segregation at the start of the simulation. The box plots in top right graph show that the distributions of D′ for minority-minority segregation are always centered on zero as

---

[7] Note that to facilitate visual inspection the example city depicted in the figure is smaller (about 1/3rd size) than the city size used in the simulations.

Fig. 16.3 Illustration of the transition from the initial state of minority-minority integration and high majority-minority segregation to the end state of all-way integration (Random distribution) (Note: Households from the majority group and two minority groups are depicted in shades of *gray* (light, medium, and dark gray, respectively). Vacant housing units are in *White*. *Grid lines* delimit areas. For easy visual review, the city here is 40% the size of the city in the simulations but faithfully depicts city shape and residential patterns)
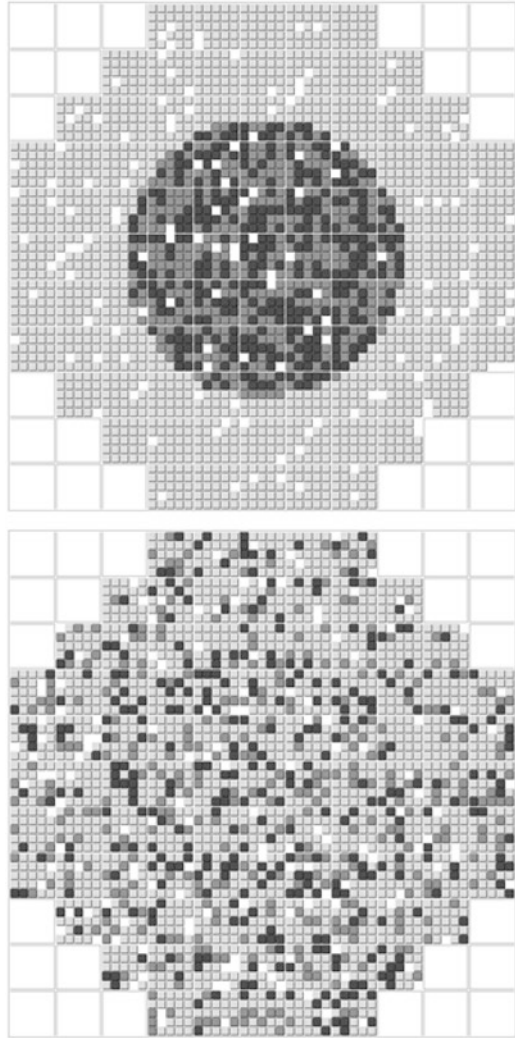
expected since households in the two minority groups are distributed randomly in relation to each other over the entire course of the simulation.

The box plots in the bottom left graph depict the distribution of scores for the standard version of the index of dissimilarity (D) for majority-minority segregation. This shows that D is very high at the beginning of the experiment and then falls sharply as households move randomly for ten cycles. But D does not fall to zero due to the intrinsic bias in D. Thus, the final level of D essentially reflects a "bootstrap" estimate of the expected value of D (E[D]) for majority-minority segregation under random assignment. The box plots in the bottom right graph depict the distributions of scores for D for minority-minority segregation. These reflect only random residential variation over the course of simulation. The surprising finding here is that D
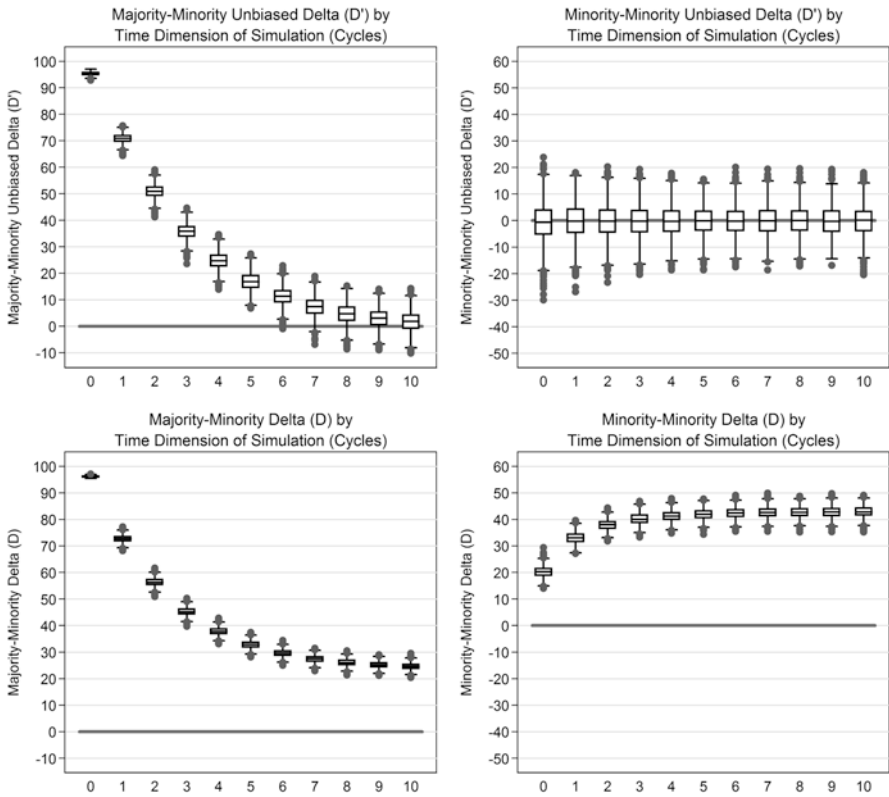
Fig. 16.4 Box plots depicting distributions of scores for unbiased and standard delta Index (D′ and D) for majority-minority segregation and minority-minority segregation over ten simulation cycles (Note: The graphs in the *top row* depict unbiased delta Index (D′) for majority-minority segregation on the *left* and minority-minority segregation on the *right*. The graphs on the *bottom row* depict values for standard delta (D) for the same comparisons. See text for details regarding the simulation designs)

increases over the course of the simulation. Why does this occur when the two minority groups are distributed randomly in relation to each other over the entire simulation? The answer traces to the complicated nature of effective neighborhood size in residential patterns for cities with three or more groups.

As illustrated in Fig. 16.4, the simulations begin with the two minority groups being highly segregated from the majority group. Under this pattern, effective neighborhood size (ENS) for the minority-minority segregation comparison is approximately 25 (i.e., the size of the neighborhoods) because households from the two minority groups live together in a small subset of the city's areas where majority households are absent. But the value of ENS for the minority-minority comparison changes over the course of the simulation. Under the final pattern of random distribution for all groups, effective neighborhood size (ENS) for minority-minority

segregation falls to approximately 5 (i.e., 20 % of the neighborhood size of 25).[8] The change in ENS has important implications for the expected value of D under random assignment (i.e., E[D]) because E[D] is a negative function of effective neighborhood size. Consequently, over the course of the simulation, ENS falls from 25 to 5 and the value of E[D] for the minority-minority segregation comparison increases.

Figure 16.5 graphically summarizes results from additional analyses that replicate the analysis just reviewed using additional multiple racial demographic distributions for the virtual city. These are for group distributions of 80-15-5 and 91-6-3. The findings closely parallel those presented in Fig. 16.5. The results document two key findings. The first is that the unbiased version of D that is set forth in this study behaves in a desirable way under a wide range of conditions. The second is that standard version of D behaves in an undesirable way under these same conditions.

These findings document that previous suggestions by Winship (1977) and Carrington and Troske (1997) for dealing with index bias face a serious obstacle. They suggest adjusting observed values of D in relation to D's expected value under random distribution based on the calculation $D* = (D - E[D])/(1 - E[D])$. The obstacle this approach faces is that the proposed adjustments can be effective only when the value of E[D], whether estimated by formula or by bootstrap methods, is accurate. Unfortunately, the results just reviewed show that the value of E[D] for the minority-minority segregation is not a simple constant. In the simulations under review here the two minority groups are distributed randomly in relation to each other. Accordingly, the value of D for this comparison reflects a bootstrap simulation estimate of E[D] for the minority-minority segregation comparison. The results from the simulations show that the value of E[D] is significantly impacted by an important factor that is not considered in previous discussions of potential solutions for dealing with index bias. Specifically, the value of E[D] is impacted by how the two groups in the comparison are distributed in relation to a third group – that is, the value of E[D] for the minority-minority comparison is impacted by how the two minority groups are distributed in relation to the majority group. In more general terms, the findings reviewed here indicate that E[D] for any two-group comparison is complicated in the multi-group situation and will be affected by: (a) the extent to which the two groups in the comparison are jointly segregated from other groups and (b) the relative size of other groups in the city population.

Space does not permit a detailed review of the issue, but in analyses not reported here, I have found that this finding applies to all standard indices of uneven distribution and that two broad conclusions hold in multi-group situations. One is that expected values of index scores under random assignment (i.e., E[•]) can potentially vary over wide ranges. The other is that adjustments of index scores in relation to expected values (E[•]) based on assumptions of simpler conditions can be inappropriate and perform poorly. *In the extreme the adjustments can generate assessments of segregation that are as problematic as the original unadjusted index scores.*

---

[8] I say approximately because a precise discussion of effective neighborhood size would take account of the city vacancy rate (which is 6 % in these simulations).
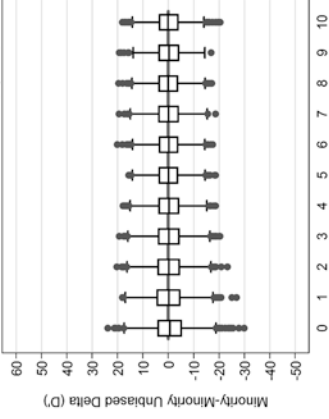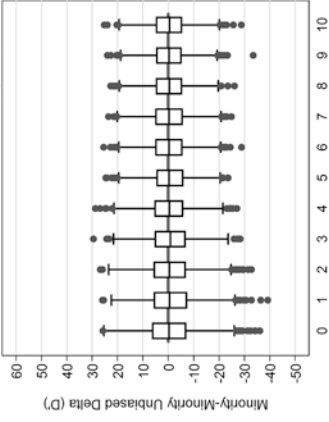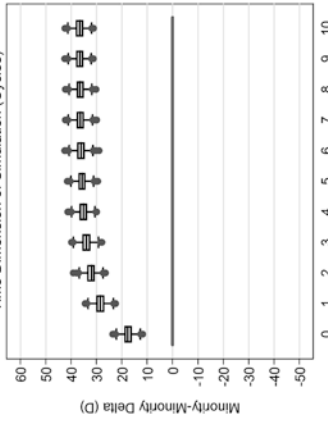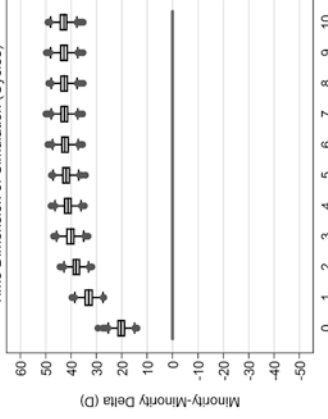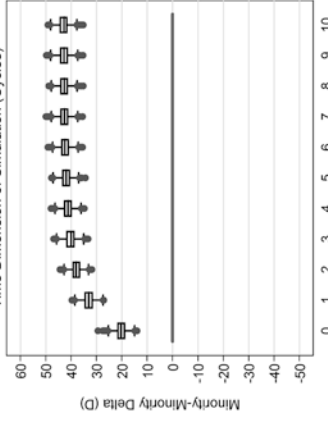
Fig. 16.5 Scores for unbiased and standard delta index (D' and D) for minority-minority segregation over time for three combinations of ethnic mix (Note: *Top row* depicts the unbiased delta index(D') for minority-minority segregation; *bottom row* is the standard delta index (D) for minority-minority segregation. Each simulation begins with an initial residential distribution in which the majority group is very highly segregated from two minority groups and the two minority groups are randomly distributed in relation to each other. Ten periods of random residential movement follow and the segregation pattern moves rapidly toward random distribution for *all* groups. Ethnic mix settings are 80/10/10 (column 1), 80/15/5 (column 2), and 91/6/3 (column 3). Neighborhood size is 25)

This may help explain why adjustment methods such as those proposed by Winship (1977) and Carrington and Troske (1997) are rarely used in empirical analyses. My own experience has been that the adjustment methods work quite well in methodological exercises where the underlying assumptions of the method are met (or closely approximated). However, when I apply the adjustments in the context of multi-group situations, they tend to "break down" and often yield unexpected results sometimes including results that are substantively implausible.

It is possible that the general approach of adjusting standard index scores could be "salvaged." This could be accomplished by using more sophisticated methods to develop refined estimates of expected index values under random assignment (i.e., E[•]) that take account of the complications associated with population groups not included in the segregation comparison. For example, I have found that bootstrap methods can be used to obtain serviceable situation-specific estimates of E[•]. One approach that appears to work well is to take the observed distribution across areas of the combined count of the two groups in the segregation comparison. Then perform bootstrap simulations wherein households from the two groups in the comparison are assigned randomly to areas until the observed area counts for the two groups combined are duplicated in each area. Performing a sufficiently large number of bootstrap simulations (e.g. 1,000 or more) will then establish the expected value of the index of interest under random assignment.

Alternatively, one could apply formula-based methods to obtain expected values of indices. But the formulas would have to be refined to take into account the observed distribution of effective neighborhood size across areas of the city. This makes implementing the formulas more complicated and also more computationally demanding.

Estimates of E[•] obtained in these ways are specific, not only to the nature of the multi-group residential pattern, but also to other potential complicating factors such as variation in area size. Unfortunately, most researchers are likely to view these technical refinements as exceedingly burdensome to implement. For example, in the simulation results just reviewed, the values of E[D] would have to be recalculated anew – using computation-intensive bootstrap methods or complex analytic computations – at least at the beginning of every time period of the simulation and perhaps even more frequently in the early stages of the simulations when the empirically assessed value of E[D] is changing rapidly. For this reason, reason it is unlikely that this approach will ever gain wide use.

The good news is that the unbiased indices I introduce in this monograph provide a superior alternative. The approach I propose is effective in both simple and complicated conditions, is conceptually appealing and easy to understand, and is much easier to implement in empirical analyses. The new unbiased indices I propose eliminate the source of bias at its root cause and do not rely on "after the fact" adjustments to purge unwanted consequences of index bias. Accordingly, the expected values of the unbiased indices are zero regardless of whether other groups are present in the population and, if so, regardless of the nature of the residential segregation pattern between the two groups of interest and other groups. Indeed, the only impact I have been able to discern so far is that the dispersion of the sampling

distribution of the unbiased indices is affected by the presence of other groups. More specifically, while the mean for unbiased indices is always approximately zero, the standard error of the mean varies inversely with ENS as basic sampling theory would lead one to expect. But this pattern holds for the expected distributions of scores of both standard and unbiased versions of indices of uneven distribution and so does not diminish the advantage of using unbiased versions of indices.

16.3   Conceptual and Practical Issues and Potential Impact on Research

When should researchers use the new unbiased versions of indices of uneven distribution I have introduced here? One simple and reasonable answer is that researchers can and should use the unbiased versions of the indices in most if not all situations. Unbiased versions of index scores are not burdensome to compute; they support familiar substantive interpretations; they also expand available substantive interpretations; they eliminate concerns that index bias may distort findings; and they give researchers the option to expand research designs to consider a wider range of situations where standard versions of index scores would be untrustworthy and misleading.

Significantly, *few, perhaps no, unwelcome consequences are associated with using unbiased indices*. If standard versions of indices of uneven distribution are non-problematic, the unbiased versions indices will closely replicate their scores. This is because scores of unbiased indices differ from scores of standard indices in meaningful ways only when the scores for the standard indices are problematic. When this is happens, the scores of the standard version of the index are called into question as untrustworthy for many research purposes and the scores of the unbiased version of the index provide a more trustworthy assessment of the nature of group differences in residential distribution.

Will using the unbiased versions of familiar indices lead to major changes in research findings? I answer this question in two parts. The first part of my answer begins by noting that studies conducted in recent decades have tended to use research designs that try to guard against index bias. I have characterized the strategies used as a patchwork of practices that can be criticized for being crude and in some cases weakly justified. But in general the strategies do tend to minimize the most egregious impacts of index. As a result, findings of many, perhaps most, previous studies using standard indices are not necessarily likely to be contradicted in dramatic ways if they are *exactly replicated* but using unbiased indices. I place emphasis on the phrase "exactly replicated" to stress that this means using exactly the same set of cases. Below I note that future studies may differ from past studies by being able to use a wider range of cases and more varied group comparisons instead of being limited to using the smaller, restricted set of cases and group comparisons used in past research.

The reason why the specific findings of many past studies are not likely to change when exactly replicated using unbiased indices is straightforward. To the extent that the practices researchers have incorporated into research designs have been conservative and excluded cases that are most seriously affected by problems with index bias, replications that use unbiased versions of indices for the same cases will not be likely to yield dramatically different results. This is because the unbiased versions of indices yield scores similar to standard versions when bias is low. Substantively meaningful differences might arise for marginal cases that were not effectively screened because the ad hoc screening practices were crude and imprecise. But in many, perhaps most, studies these cases should not dominate the findings and so results will likely remain similar when the analysis is replicated using unbiased indices.

Certain kinds of past studies would be most susceptible to changes in results if "exactly" replicated using unbiased versions of indices of uneven distribution instead of standard versions. These are studies where research designs were less stringent in screening out cases where index scores are most susceptible to bias. Examples would include: studies that use block-level data instead of tract data; studies that focus on segregation for groups that are imbalanced in size, studies that focus on subgroups that are small in combined size, and studies that are based on sample data instead of full count data.

Another kind of study result that might change when replicated using unbiased measures are studies where findings differ when cases are weighted by minority population size in comparison to when cases are weighted equally. Presumably findings do often differ. Otherwise the practice of weighting cases would not be so widely used. Instead, an early study would report the finding that it makes no difference and study designs would weight cases equally. The results reviewed here show that minority group size has no intrinsic relationship to bias. So the logical justification for weighting cases by minority group size to minimize the consequences of index bias can be questioned under all circumstances. The practice would clearly be unwarranted if studies are replicated using unbiased versions of indices. I suspect this might lead to some changes in findings. The current widespread practice of weighting by minority group size skews findings toward the cases in the sample that have larger minority populations. To the extent that this subset of cases has different segregation outcomes, from the remainder of the cases, findings would change when studies are replicated using unbiased versions of indices.

A broader interpretation to the notion of study replication would lead to a different answer. "Exact" replications of past studies involves excluding many cases that can be included when using unbiased versions of indices. Similarly, "exact" replications of past studies means foregoing many group comparisons that can be examined when using unbiased versions of indices. The availability of unbiased indices frees the literature from the need to accept these past compromises in study design. With this in mind I now offer the second part of my answer.

There are at least three ways that results for empirical studies are likely to change in welcome and potentially important ways when researchers adopt unbiased indices. One is that *using unbiased index scores will give researchers much greater*

ability to discuss and compare specific cases without concern for the distorting influence of bias. These discussions are more difficult when standard scores are used. Scores for individual cases are potentially subject to different levels of distortion by index bias. Researcher recognition of this concern motivates the widespread practice of weighting cases differentially in statistical analyses. Concern about case-to-case variation in the impact of bias on index scores complicates the interpretation of scores of individual cities and it also complicates the direct comparison of scores for any given city with the scores of any other cities. Such complications are eliminated when using unbiased scores. Scores for individual cases can be evaluated with ease. Similarly, scores for two cases and scores for the same case at two points in time can be compared without concern.

A second way results may change is that *the logic of case weighting as implemented in statistical analyses in current studies will no longer be justified when using scores for unbiased versions of indices*. The stated motivation for differentially weighting cases – that is, to minimize the distorting impacts biased cases may exert on findings – is of course negated entirely. The main implication of this is that results of statistical analyses will no longer be driven by segregation patterns for cities with large minority populations. It is unclear whether this will in fact lead to important changes in findings. But it is a distinct possibility that results of statistical analyses may differ because many cases which previously would have had little or no influence on results of statistical analyses will now carry equal weight.

The third way using unbiased indices will impact segregation studies is the most important. It is that *researchers will be free to greatly expand the scope of segregation studies*. Researchers will no longer need to limit analysis to the small subset of cities that survive sample restrictions and receive weights that give them disproportionate influence on results after prevailing practices exclude and discount potentially problematic cases to guard against index bias. Instead, future studies will be able to conduct expanded analyses that may investigate segregation in many situations that previously were not examined because conventions in restricting study designs foreclosed this possibility. Relatedly, using unbiased indices will allow researchers to consider many kinds of group comparisons that previously could not be considered. This includes, for example, comparisons involving small population groups and comparisons involving small subgroups within particular populations. In the past, such comparisons have gone unexamined because index scores are potentially subject to high levels of bias. These concerns can be set aside when unbiased versions of indices are used.

Eliminating the need to impose draconian restrictions on research designs of segregation studies can only be a good thing. It will allow researchers to expand samples and explore a broader range of research questions. The following is a brief list of research applications where the benefits of using unbiased indices are especially likely to be seen.

- studies assessing segregation at small spatial scales such as the census block and block group; or classrooms within schools; or the very small neighborhoods typically used in agent simulation analyses of segregation[9];
- studies assessing segregation when groups are imbalanced in size; for example, studies of segregation involving small population groups such as Asian and Latino populations in areas of new settlement; and
- studies assessing segregation for subgroups within broader populations which will result in small effective neighborhood size; for example, the segregation of Latino and Asian subgroups, and the segregation of high-income Whites and high-income African Americans.

I conclude by strongly encouraging researchers to take advantage of the new option to use unbiased versions of popular indices of uneven distribution. One is never worse off for examining the new unbiased versions of popular indices and there are many ways they may yield benefits. Accordingly, I argue that it will always make good sense to examine the scores of the unbiased versions of indices. As I said, one can never be worse off for doing so because findings will be unchanged if bias is not a problem and the positive confirmation on this point will provide an additional basis for placing confidence in one's findings. Moreover, there are many reasons to expect one would be better off, perhaps by a great deal, in comparison to following prevailing practices. Current "rule-of-thumb" practices that aim to minimize undesirable complications associated with index bias are crude and imprecise and can be "hit and miss" in effectiveness. Concerns on this point can be completely set aside by examining the unbiased versions of the indices even if one in the end elects to report results for standard versions of indices. However, it is likely that standard indices will be used as often as in the past because the availability of unbiased versions of indices of uneven distribution makes it possible for researchers to examine segregation in a wider range of situations than was previously possible. Once this occurs, scores for standard indices will be even less trustworthy than they currently are and researchers will increasingly need to rely on unbiased versions when attempting to answer the new questions these measures permit researchers to investigate.

---

[9] In fact, I began pursuing the development of unbiased measures of uneven distribution to cope with the problem of bias in measuring segregation in simulation studies. In that context, the unbiased measures allow researchers to explore a much wider range of combinations of neighborhood scale and population composition than can be considered using standard versions of segregation indices.

References

Carrington, W. J., & Troske, K. R. (1997). On measuring segregation in samples with small units. *Journal of Business and Economic Statistics, 15*, 402–409.

Clark, W. A. V., & Fossett, M. (2008, March 18). Understanding the social context of the Schelling model. *Proceedings of the National Academy of Sciences*, *105*(11), 4109–4114.

Fossett, M. (2006). Ethnic preferences, social distance dynamics, and residential segregation: Theoretical explorations using simulation analysis. *Journal of Mathematical Sociology, 30*, 185–274.

Fossett, M. (2011a). Generative models of segregation: Investigating model-generated patterns of residential segregation by ethnicity and socioeconomic status. *Journal of Mathematical Sociology, 35*, 114–135.

Fossett, M. (2011b). *Spatial implementations of segregation indices: New implementations for popular indices of uneven distribution.* Paper presented at the session on racial residential segregation at the annual meetings of the Southwestern Sociological Association, Las Vegas, Nevada, March 16–19.

Fossett, M., & Dietrich, D. R. (2009). Effects of city size, shape, and form, and neighborhood size and shape in agent-based models of residential segregation: Are Schelling-style preference effects robust? *Environment and Planning B: Planning and Design., 36*, 149–169.

Fossett, M., & Waren, W. (2005). Overlooked implications of ethnic preferences for residential segregation in agent-based models. *Urban Studies, 11*, 1893–1917.

Fossett, M., & Zhang, W. (2011). Unbiased indices of uneven distribution and exposure: New alternative for segregation analysis. Paper presented at the annual meetings of the population association of America, Washington, DC, March 31-April 2.

Winship, C. (1977). A re-evaluation of indices of residential segregation. *Social Forces, 55*, 1058–1066.