Chapter 9
Unifying Micro-level and Macro-level Analyses of Segregation

Casting segregation indices in the difference of means framework provides a valuable option previously not available to researchers. It enables them to seamlessly connect macro-level segregation – as measured by the index score for a city – to micro-level processes of residential attainment. At the simplest level the value of any index placed in the difference of means framework can be obtained by performing an individual-level attainment analysis that predicts index-relevant residential outcomes (y, scored from area group proportion p) for individuals with a dummy variable (0,1) for racial group membership. The regression coefficient for race will exactly equal the index score obtained by standard computing formulas. This introduces a new interpretation of segregation index scores; their values reflect the effect of race on the attainment of residential outcomes that determine the segregation index score for the city.

Establishing the equivalence of between macro-level measures of segregation and the effect of race on residential attainments in a bivariate individual-level regression model paves the way for at least three important new options for segregation analysis. The first is to give researchers the ability to extend and elaborate bivariate models to investigate segregation in more detail using multivariate analyses. These models make it possible for researchers to address fundamental questions that previously could not be directly investigated. For example, researchers can assess whether or not the impact of race on segregation-determining residential outcomes seen in the bivariate analysis continues to persist when controls are introduced for other relevant individual- and household-level social characteristics (e.g., age, education, income, marital status, household composition, nativity, etc.) that may exert independent influence on residential outcomes.

A second new option for segregation analysis is to give researchers the opportunity to quantitatively dissect the underpinnings of segregation in more detail than has previously been possible. Specifically, researchers can use familiar tools of standardization and decomposition analysis to assess how the index score for a city is

quantitatively linked to group differences in the resources each group brings to the residential attainment process and to group differences in the parameters of the attainment process where resources (inputs) are converted to residential outcomes. Thus, one can develop improved answers to questions such as "Does segregation arise primarily because groups differ on income and other resources that affect residential contact with the reference group?" Or, "Does segregation arise primarily because groups differ with respect to their ability to convert income and other resources into residential contact with the reference group?" Or, "Do both factors play an important role in creating segregation?" Questions of this sort have been raised for many decades. But answers have been unsatisfactory because the available options for addressing the question have been crude and difficult to implement. The difference of means formulation provides new and superior options for developing answers to these long-standing questions.

A third new option for segregation analysis is for researchers to investigate cross-area and over-time variation in segregation in more detail using multi-level specifications of bivariate and multivariate segregation attainment models. Segregation attainment models are individual-level attainment models that predict the residential outcomes that exactly determine the level of segregation in a city. Multi-level specifications of the basic bivariate segregation attainment model enable researchers to investigate ecological variation in segregation by assessing how segregation – equated in this approach to the effect of race on segregation-determining residential outcomes – varies over time and across different cities depending on the time period and characteristics of the metropolitan area such as its size, rate of growth, industrial and occupational structure, unemployment rate, military presence, etc.

Multi-level specifications of individual-level, multivariate segregation attainment models make it possible to investigate these patterns in more detail and sophistication than ever before. Importantly, these models provide a superior approach for taking account of the role of non-racial social characteristics in shaping variation in segregation over time and across areas. Researchers routinely hypothesize that group differences on income, nativity, and other social characteristics may play a role in explaining cross area variation in segregation. Currently these hypotheses are assessed with aggregate-level models in which measures such as group income ratios, or percent foreign born for Latinos are used to predict segregation index scores for cities. The difference of means framework and the associated new option of analyzing segregation via attainment models make it clear that this long-standing practice is fundamentally flawed and should be discontinued.

Current practice carries risks of erroneous inference associated with the so-called "ecological fallacy" – the fallacy of using aggregate indicators to assess or control for the effects of variables that operate at the micro level. Researchers have relied on the aggregate-level approach to address these important questions because until now they did not have better options for analysis. Multi-level implementations of multivariate segregation attainment models now allow researchers to properly take account of variables that affect segregation-determining outcomes at the micro level

(e.g., income, nativity, English language ability, etc.) when investigating cross-area and cross-time variation in segregation.

The difference of means framework makes these three new options for segregation analysis possible. I discuss the first two in more detail in the remainder of this chapter. I provide a detailed discussion of the third option in Chap. 10.

9.1   New Ways to Work with Detailed Summary File Tabulations

To begin I illustrate how the difference of means formulation makes it possible for researchers to investigate segregation in new ways by revisiting and expanding on the analysis of White-Minority segregation in Houston, Texas reported earlier in Chap. 5 (Tables 5.1, 5.2 and 5.3). The summary file tabulations underpinning these analyses provide more than just simple counts of families by race for census block groups. The tabulations also provide counts of families by poverty status, family type, and presence of related children separately by race.[1] The analysis of segregation reported in Tables 5.1, 5.2 and 5.3 was simple and conventional. It assessed segregation in terms for race differences in residential outcomes without consideration for the role of the other social and economic characteristics available in the tabulation. There was no need to do so because index scores for the *overall* level of segregation between groups can be calculated using just group counts by race over areas. Accordingly, the scores reported in Tables 5.2 and 5.3 were obtained by collapsing the original detailed tabulations to obtain just the marginals for race.

The difference of means formulation of segregation indices makes it possible to draw on the detailed information in the full tabulation to gain a deeper understanding of how overall segregation is related to group differences in distribution across poverty status and family type. It has always been recognized, at least implicitly, that segregation arises out of group differences in distribution on individual residential attainments. And it also is widely recognized that residential attainments may vary, not only by race, but also with social characteristics such as age, gender, education, income, family status, and so on. Accordingly, researchers extending back at least to Duncan and Duncan (1955) have always wished for the option to take account of the possible role of social characteristics other than race when investigating racial segregation. They have been frustrated in this goal, however, because until now the macro-level outcome of segregation could not be directly linked to individual-level residential outcomes in a way that would allow researchers to undertake the kinds of quantitative analyses needed to explore the issues with greater detail and sophistication.

---

[1] Specifically, I draw on Tabulations P160 A-I of Census Summary File 3 of the 2000 Census.

Table 9.1 Descriptive statistics for poverty status and distribution of poverty status by family type for Whites, Black, Latinos, and Asians in Houston, Texas, 2000

Variable	Whites	Blacks	Latinos	Asians
Distribution of families by social characteristics				
Percent families not in poverty	95.8	80.9	80.2	90.4
Percent families in poverty	4.2	19.2	19.8	9.6
Percent families married couple	84.0	51.2	74.1	84.9
Percent families with children	47.3	62.0	69.2	59.3
Detailed distribution of families by poverty status and family type				
Families not in poverty by family type				
Married couple, no children	43.2	19.1	16.0	28.6
Married couple, children	38.7	27.8	46.0	49.5
Female headed, no children	3.9	7.9	2.9	3.4
Female headed, children	6.1	19.9	7.4	3.7
Other family type	4.0	6.2	7.9	5.3
Families in poverty by family type				
Married couple, no children	1.1	1.6	1.6	2.2
Married couple, children	1.1	2.7	10.6	4.6
Female headed, no children	0.3	1.7	0.5	0.5
Female headed, children	1.4	11.6	5.2	1.6
Other family type	0.4	1.6	1.9	0.7
	100.2	100.0	100.0	100.1
Sample N	627,613	195,928	294,931	55,746

Source: US Census 2000, Summary File 3

The difference of means framework provides a solution to this problem. Casting segregation index scores as a group difference of means on residential outcomes for individuals opens the door for researchers to apply a standard toolkit of methods that are currently used to investigate race differences on education, income, poverty status, and other socioeconomic outcomes. Specifically, researchers now can analyze segregation by combining individual-level attainment analysis with demographic techniques of standardization and components analysis to better assess the roles that race and other social characteristics play in determining segregation.

9.2   Some Preliminaries

Tables 9.1 and 9.2 present the relevant descriptive data for the case of Houston, Texas. Table 9.1 documents that Whites, Blacks, Latinos, and Asians differ in their distribution across categories of family type and poverty status. Table 9.2 documents how averages on the residential outcomes (y) that determine the separation index (S) vary across families grouped by family type, poverty status, and race. Table 9.3 similarly documents how averages on the residential outcomes (y) that

Table 9.2 Means on pairwise contact with Whites (y) scored for the separation index (S) by poverty status and family type for White-Minority comparisons, Houston, Texas, 2000

Family type	Whites		Minority group	
	Non-poverty	Poverty	Non-poverty	Poverty
White-Black comparison				
Married couple, no children	89.9	87.4	34.2	19.7
Married couple, children	91.3	86.0	41.3	29.4
Female headed, no children	86.1	87.3	24.9	19.7
Female headed, children	87.7	83.7	31.1	23.0
Other family type	87.2	83.9	30.6	20.1
All families	89.9		32.5	
Value of separation index (S)		57.4		
White-Latino comparison				
Married couple, no children	81.2	72.5	46.9	30.7
Married couple, children	83.9	74.0	42.6	30.2
Female headed, no children	73.3	70.7	38.4	31.3
Female headed, children	77.6	71.9	41.6	31.6
Other family type	75.7	66.9	36.2	28.2
All families	81.1		40.2	
Value of separation index (S)		40.9		
White-Asian comparison				
Married couple, no children	93.9	94.5	71.2	61.8
Married couple, children	93.9	94.0	71.7	62.3
Female headed, no children	93.2	94.6	65.6	63.6
Female headed, children	92.7	94.5	70.4	58.2
Other family type	93.3	92.9	64.8	59.4
All families	93.8		69.9	
Value of separation index (S)		23.9		

Source: US Census 2000, Summary File 3

determine the dissimilarity index (D) vary across families grouped by family type, poverty status and race. In the difference of means framework the patterns in these three tables carry clear and direct implications for segregation. The overall segregation index score for the group comparison is determined by the group difference of means on residential outcomes (y) and the mean for each racial group is in turn determined by the weighted average of the subgroup means for that racial group.

From that vantage point the data presented in Tables 9.2 and 9.3 can be understood as providing a simple "ANOVA-style" micro-level attainment analysis of residential segregation as measured by the separation index (S) and the dissimilarity index (D), respectively. The essence of the analysis is that individual families are cross-classified by the "independent variables" of race, family type, and poverty status and means on the "dependent variable" of scaled contact with Whites (y) are reported for the subgroups that are broken out in the cross tabulation. The overall group means reported in Table 9.2 in the rows labeled "All Families" reflect the

Table 9.3 Means on scaled pairwise contact with Whites (y) scored for the dissimilarity index (D) by poverty status and family type for White-Minority comparisons, Houston, Texas, 2000

Family type	Whites		Minority group	
	Non-poverty	Poverty	Non-poverty	Poverty
White-Black comparison				
Married couple, no children	87.5	82.2	20.2	8.8
Married couple, children	90.6	80.6	24.2	14.5
Female headed, no children	80.8	83.4	9.9	7.1
Female headed, children	84.6	77.8	13.5	8.8
Other family type	82.5	74.9	14.5	7.1
All families	87.7		16.8	
Dissimilarity index (D)		70.9		
White-Latino comparison				
Married couple, no children	81.5	67.6	32.9	13.3
Married couple, children	86.2	68.9	24.9	12.5
Female headed, no children	68.4	69.1	22.3	17.2
Female headed, children	76.6	65.8	23.6	13.5
Other family type	72.5	56.9	18.5	11.1
All families	81.5		23.1	
Dissimilarity index (D)		58.4		
White-Asian comparison				
Married couple, no children	75.3	78.8	19.2	13.0
Married couple, children	75.6	78.8	17.1	13.7
Female headed, no children	73.8	84.7	16.4	26.9
Female headed, children	70.7	78.6	15.7	12.1
Other family type	73.7	77.6	10.3	8.7
All families	75.2		16.9	
Dissimilarity index (D)		58.3		

Source: US Census 2000, Summary File 3

weighted sum of the subgroup means by family type and poverty status based on the relative frequencies reported in Table 9.1. The difference between the two "overall" group means yields the index score for the comparison. Thus, the score for the separation index (S) for the White-Black comparison is 57.4 based on the difference between Whites having mean (pairwise) contact with Whites of 89.9 compared to 32.5 for Blacks. Similarly, the score for the dissimilarity index (D) for the White-Black comparison is 70.9 based on the difference between Whites having a mean of 87.7 on (scaled pairwise) contact with Whites compared to a mean of 16.8 for Blacks.

It is not standard practice to analyze overall segregation index scores as arising from group differences in the distribution of individual families across subgroups with different average levels on residential outcomes (y) of scaled contact with Whites. In light of this I briefly review how the analysis presented in Tables 9.1, 9.2, and 9.3 can be performed using census summary tables. To begin, the data con-

tained in the block group-level census summary tabulation must be reconstituted as a micro-level data set for families. The first step is to recognize that the count for each "interior" cell in the full summary file tabulation represents a set of micro-level "cases" – families in this example – that have a particular configuration of social characteristics. The poverty status by family type summary file tabulation in question has eighteen (18) interior cells (note that tabulation marginals are excluded). The tabulation is repeated for all four racial groups yielding 72 separate "cases" (i.e., cells) for each block group. The final data set thus has one "record" for each interior cell in the summary file tabulation; that is a total of 72 separate records for each of the block groups in Houston. Each record has a unique combination for the characteristics of race, family type, and poverty status. The cell frequency indicates how many families with this unique combination of characteristics are found in each block group in the metropolitan area.

Next a set of variables is coded for each of the records. The first variable is area of residence (i.e., the block group code). The second is "nfamilies" which is set to the value of the cell frequency for this case (i.e., the count of families in that cell of the tabulation). This will later be used as the frequency weight for the record when performing statistical calculations.[2] Next a series of additional variables are coded to represent the social characteristics of each family – namely, their race, family type, poverty status, etc. – in the table. Each characteristic is coded as a separate variable and assigned values as appropriate for the needs of the analysis. Each record in the resulting data set represents a set of families that reside in a particular block group and hold a specific combination of social characteristics.

The variables that register social characteristics will serve as "independent" variables in micro-level residential attainment analyses. They may be coded a variety of equivalent ways. I created dummy (0,1) variables for race to select records for Whites, Blacks, Latinos, or Asians as relevant. I also created a dummy variable for "poverty" and I similarly created a set of dummy variables to represent the five categories of family type. Finally, I also created additional dummy variables to capture the possible interaction of poverty status and family type. Viewed from the perspective of analysis of variance (ANOVA) the set of dummy variables includes all combinations needed to estimate a "saturated" ANOVA model which includes all main effects and all possible interactions.

The next step is to prepare a separate block group data set. The cases in this data set are block groups. The first variable for the case is the block group code which will be used for merging with the first micro-level data set. In addition, a set of variables are coded for the total counts of families by race; specifically, separate variables for the count of White, Black, Latino, and Asian families. Next compute a set of variables

---

[2] Alternatively, one could create an individual-level data base by generating the relevant number of individual records for the families represented in each cell of the cross tabulation and assigning relevant codes for the social characteristics of the families as appropriate based. Of course, it is mathematically equivalent and computationally more efficient to use cells as cases and weight by cell frequency when performing analyses. However, the alternative approach can be used when statistical software cannot apply frequency weights for cases.

with the values of pairwise proportion White (p) for each of the three possible White-minority comparisons. These provide the basis for computing variables that score residential outcomes (y) from area (pairwise) proportion White (p) as relevant for different segregation indices. For example, in the case of the separation index (S), the relevant residential outcome (y) is the value of p. In the case of the dissimilarity index (D) the relevant residential outcome is the value of either 1 or 0 depending on whether area proportion White (p) is greater than proportion White for the city (P) or not. The resulting block group-level data set will then contain variables that will serve as dependent variables in micro-level segregation attainment analyses.

The final analysis data set is created by merging the second data with the first data set based on the common block group code. The resulting data set can then be used to perform micro-level statistical analyses to analyze residential segregation.

I followed the procedures just described to prepare a data set I used to perform the analyses establishing how means on the residential outcome of scaled contact with Whites (y) varies across subgroups and groups as reported in Tables 9.2 and 9.3. The results in these tables were obtained by via tabulation routines that calculate means on the relevant dependent variables (y) across the categories of a cross classification table based on micro-level variables measuring the social characteristics of race, family type, and poverty status. In the analysis the records in the family-level data set were weighted by the variable "nfamilies" which has the number of families that have the specific combination of social characteristics and reside in the block group in question. The same family-level data set can be used to perform micro-level statistical analyses such as analysis of variance (ANOVA) and multiple regression analysis predicting the dependent variable of individual residential attainments using the independent variables of race and other social characteristics.,[3,4] I report regression results obtained in this way later in the chapter.

9.3   Substantive Findings

I now discuss the analysis results in more detail. Table 9.2 shows that in all three White-Minority comparisons scaled (pairwise) contact with Whites varies across categories of poverty status and family type as well as by race. Group means on this residential outcome determine the value of the separation index (S). Two clear patterns warrant mention even on cursory inspection of the table. The first is that minority contact with Whites is consistently lower for poverty families compared with non-poverty families. The second is that, within non-poverty families, married

---

[3] Weighting cases by the cell counts from the summary file tabulation makes this an individual-level regression because the cell count registers the number of families that reside in the block group in question and have the exact combination of race, family type, and poverty status coded for the case.

[4] These can be termed "saturated" models because they include all possible effects of poverty status and family type (including all interactions).

couple families have higher levels of contact with Whites. Table 9.1 also documents that overall and within categories of family type minority families are consistently more likely to be in poverty than are White families but with Asians being substantially less disadvantaged than Blacks and Latinos. Table 9.1 also shows that the overall percentage of families that are married couples and non-poverty is much higher for Whites (81.9 %) and Asians (78.1 %) than for Blacks (46.9 %) and Latinos (62.0 %). The combination of these two patterns suggests it is plausible to hypothesize that group differences on poverty and family composition may play a role in making White-Black and White-Latino segregation more pronounced than White-Asian segregation.

Closer inspection of the patterns in Table 9.2 lends additional credibility to this conjecture. In the White-Black comparison pairwise contact with Whites (p) varies within a narrow interval of 7.6 points for White families ranging from a low of 83.7 % for female-headed families with children and in poverty to a high of 91.3 % for non-poverty married couples with children. For Black families contact with Whites is generally much lower than that observed for Whites in every category of family type. This suggests that race is a crucial factor in shaping the value of S (the group difference of means on p). However, it also is the case that Black contact with Whites varies by 21.6 points over categories of poverty status and family type for Blacks. The lowest level of 19.7 % is seen for married couple families without children and in poverty and this level also is seen for female-headed families without children and in poverty. The highest level of 41.3 % is seen for non-poverty married couples with children. The contrast is dramatic; the level of contact seen for the latter group is 21.6 points higher and more than double the level see for the first two groups. This suggests that, in addition to the important role of race alone, group differences in family type and poverty status also might impact the value of S for the White-Black comparison.

Similar patterns are evident in the results for the White-Latino comparison and the White-Asian comparison. In the White-Latino comparison Latino contact with Whites (p) is lower than that observed for Whites for every combination of family type and poverty status suggesting a clear "across the board" race effect. But it also is clear that contact with Whites varies across categories of family type and poverty status; by 18.7 points for Latinos and by 13.2 points for Whites. Combining this information with the knowledge that Latinos are disproportionately concentrated in categories of family type and poverty status that experience lower levels of contact with Whites suggests that group differences in distribution by poverty and family type may impact the level of White-Latino segregation.

In the White-Asian comparison Asian contact with Whites (p) is lower than that observed for Whites across all categories of family type and poverty status again suggesting an across the board" race effect but Asian contact with Whites varies much more (by 13.5 points) across categories of family type and poverty status than is observed for Whites (only 1.9 points) thus lending plausibility to the hypothesis that group differences in distribution by poverty and family type may impact the level of White-Asian segregation.

In sum, the patterns documented in Tables 9.1 and 9.2 lend plausibility to the hypothesis that group differences in social characteristics might play a non-trivial role independent of race in contributing to overall segregation. Without going into the same level of detail, I note that similar conclusions can be drawn based on reviewing the data on residential outcomes that determine the value of the dissimilarity index (D) presented in Table 9.3. The key finding is that the subgroup means that determine D vary across poverty and family type within race. This raises the possibility that group differences in distribution across these social categories may be a factor contributing to segregation as measured by D.

9.4 Opportunities to Perform Standardization and Components Analysis

The micro-level data set used to prepare Tables 9.1, 9.2, and 9.3 also can be used to apply the workhorse demographic techniques of standardization and components analysis (e.g., Kitagawa 1955; Winsborough and Dickinson 1971; Althauser and Wigler 1972; Iams and Thornton 1975; Jones and Kelley 1984) to gain insights into what factors give rise to segregation. The technique of standardization involves adopting a "standard" relative frequency distribution for poverty status and family type and using it, not the "observed" distributions given in Table 9.1, to weight the group-specific means on residential outcomes over poverty status and family type to calculate "expected" group means on residential outcomes. The resulting "standardized" group means can be interpreted as the group averages on segregation-relevant residential outcomes (y) that would result if both groups had the same "standard" distribution" on social characteristics while continuing to experience their "observed" residential outcomes documented in Table 9.2. The difference between the two group means in the standardized comparison can be interpreted as the level of segregation that remains when group differences in distribution by family type and poverty status have been "taken into account" by statistically setting them to be equal.

Table 9.4 reports results of standardization analyses of the type just outlined. In conducting this analysis I adopted the observed distribution of _all_ families (both White and minority group combined) over the categories of poverty status by family type as the relevant "standard" for the distribution of social characteristics. The top panel of the table reports results for the average levels on residential outcomes (y) that determine the value of the separation index (S) that would obtain for Whites and minorities if they had the same "standard" distribution for social characteristics. In the White-Black comparison the standardized mean for Whites is 89.46. This is about 0.40 points lower than the observed mean for Whites of 89.86. The standardized mean for Blacks is 35.07. This is about 2.59 points higher than the observed mean for Blacks of 32.48. The difference of the standardized group means can be interpreted as the value of the separation index (S) standardized to the condition of Whites and Blacks having identical distributions across family type and poverty

Table 9.4 Observed and standardized White-Minority segregation comparisons, Houston, Texas, 2000

	White-Black	White-Latino	White-Asian
Separation index (S)			
Observed group means on scaled contact with Whites (y) observed			
White mean (y)	89.86	81.12	93.79
Minority mean (y)	32.48	40.17	69.91
Difference	*57.38*	*40.95*	*23.88*
Means standardized on overall distribution for family type and poverty status			
White mean (y)	89.46	80.35	93.78
Minority mean (y)	35.07	42.06	69.59
Difference	*54.38*	*38.29*	*24.19*
Index of dissimilarity (D)			
Observed group means on scaled contact with Whites (y)			
White mean (y)	87.73	81.49	75.15
Minority mean (y)	16.75	23.12	16.93
Difference	*70.98*	*58.37*	*58.22*
Means standardized on overall distribution for family type and poverty status			
White mean (y)	87.04	80.26	75.28
Minority mean (y)	19.40	25.58	16.83
Difference	*67.64*	*54.68*	*58.45*

Source: US Census Summary File 3

status. The initial observed value of S was 57.38 points. The standardized value of S is 54.38 points. Thus, "standardizing" the comparison to a common distribution on poverty status and family type reduces the value of S by 3.00 points. This result provides a statistically sound basis for concluding that White-Black differences in the social characteristics considered here play only a small role in determining the overall level of White-Black segregation; simply put, "controlling" for group differences on social characteristics using sound methods of statistical analysis produces only a modest reduction in segregation.

This result also can be interpreted as indicating that the level of segregation as assessed by the observed value of S traces primarily to the effect of race. That is, group separation as measured by S traces to group differences in contact with Whites that arise independent of poverty status and family type. A more thorough decomposition analysis (per Kitagawa 1955; Althauser and Wigler 1972; Iams and Thornton 1975; Jones and Kelly 1984) could quantify this in a more careful way. Of course, like all standardization and decomposition exercises, thoughtful interpretations must consider the theoretical relevance of the "control" variables and the adequacy of the micro-level analysis that seeks to capture the relationship between non-racial social characteristics and segregation-relevant residential attainments.

Table 9.4 also reports results of standardization analyses for the separation index (S) for the White-Latino and White-Asian comparisons. These analyses also indicate that differences in group distribution over family type and poverty status do not play a major role in determining the overall level of segregation between the groups. In

the case of the White-Latino comparison, standardizing on poverty status and family type reduces S by 2.66 points lowering it from 40.95 to 38.29. In the case of the White-Asian comparison, standardizing on poverty status and family type increases S by 0.31 points raising it from 23.88 to 24.19. This suggests that group differences in family type and poverty status serve to obscure the impact of race on overall White-Asian segregation.

The lower panel of Table 9.4 reports results of a set of parallel analyses focusing on segregation measured using the index of dissimilarity (D). To perform this parallel analysis, I made only one change; I used a new dependent variable; namely, y as scored for D (reported in Table 9.3) instead of y as scored for S (reported in Table 9.2). Recall that in this case y is now scored 1 if $p \geq P$ and 0 otherwise. The impact of standardizing the White-Minority comparison to a common distribution on poverty status and family type here is very similar to that seen for the analysis for S. In the case of the White-Black comparison, standardizing on poverty status and family type reduces D by 3.44 points from 70.98 to 67.64. For the White-Latino comparison, the standardization exercise reduces D by 3.69 points from 58.37 to 54.68. For the White-Asian comparison, standardizing on poverty status and family type increases D by 0.23 points from 58.22 to 58.45. Thus, as seen in the analysis for S, the level of segregation measured using D changes little when one uses appropriate statistical methods to take account of the possible impact of group differences in family type and poverty status.

I also performed similar standardization exercises for other segregation indices – specifically, G, R, and H. However, I do not report the details here as the basic finding is the same in all cases.[5] That is, when analyzing group differences in residential outcomes that determine segregation as measured by G, R, and H, standardizing the White-Minority comparisons to a common group distribution on poverty status and family type reduces segregation by only modest amounts.

9.5 Comparison with Previous Approaches to "Taking Account" of Non-racial Social Characteristics

The ability to conduct the standardization exercises just reviewed is a completely new option made possible by the difference of means framework for measuring segregation. Several considerations make this approach superior to current practices for assessing or controlling for the role of non-racial social and economic characteristics of individuals on segregation. First, the approach can be easily extended to directly "control for" the role of many social characteristics in a single analysis where previously this has not been feasible. Second, the approach can draw on a broader range of information and a larger number of cases than is typical in current approaches to taking account of non-racial social characteristics and as a

---

[5] This is not surprising as the discussions in Chaps. 5, 6, and 7 note that when D and S give similar results all popular indices of uneven distribution will give similar results.

result yields results that are more appropriate and statistically reliable. Third, the results of the approach are much less susceptible to problems of distortion resulting from index bias and ecological fallacies than are results of current practices. I now briefly comment on each of these points.

The prevailing approach for taking account of the impact that factors other than group membership may have on segregation involves calculating segregation scores for subsets of individuals from the two groups that are matched on social characteristics. In the present context, that would involve calculating as many as 10 different White-Black segregation scores, one each based on the just the families found in the 10 categories of family type and poverty status. Or, for simplicity, the analysis might be limited to calculating the index score for one carefully chosen subgroup comparison such as non-poverty, married couple families with children, the family type with the largest number of families across all four racial groups. When the obtained index scores is lower than the score for the overall segregation comparison, the result is interpreted as indicating that segregation is lower when social characteristics are "controlled" and thus supports the conclusion that the impact of group differences on social characteristics on segregation is important. Alternatively, when the scores obtained is not lower than the score for overall segregation, the result is interpreted as indicating that the impact of group differences on social characteristics on segregation is modest or unimportant.

Unfortunately, basing the analysis on segregation scores calculated for matched comparisons involving small subgroup numbers often introduces non-trivial complications and concerns. One problem is that the approach subtly changes the substantive and quantitative relevance of the analysis. Note that the standardized segregation index scores reported in Table 9.4 are based on the full group distributions over many combinations of social characteristics and thus register the full spectrum of patterns of segregation for racial comparisons between and across all combinations of the 10 categories of family type and poverty status.

Anchoring the scores on the full range of data for both groups carries statistical and substantive benefits. Using the full group makes the comparison more statistically reliable; thus, for example, the standardized group means that determine the standardized values of S and D have smaller standard errors than group means computed for narrow subgroups. Substantively, using the full group data is attractive because it assesses segregation patterns between and across all combinations of social characteristics not just for a narrowly specified comparison that could potentially be idiosyncratic. Arguably this protects against getting unusual results for a particular narrowly defined comparison. Importantly, the approach also does not exclude the cross-category comparisons which quantitatively make large contributions to determining overall segregation but are completely ignored when comparisons are restricted to only one-to-one matches on social characteristics.

Another more technical problem is that scores based on narrowly defined subgroups are prone to being distorted by index bias. The problem of index bias is well-known and potentially vexing. Accordingly I give it extended attention in Chaps. 14, 15, and 16. Concern about index bias is especially relevant when group counts in spatial units are small and group ratios are imbalanced (Winship 1977).

This problem is likely to be salient when subgroup comparisons are based on small subsets of cases that exactly match on non-racial social characteristics. For example if one matches White and Black families on poverty status and family type, the counts families in each area will drop substantially. Furthermore, the underlying problem is likely to be even worse than it appears on first consideration. The reason for this is that the census tabulations that include other social characteristics in addition to race are based on samples instead of full counts. The summary file tabulations report "estimated" full counts. In fact, the analysis rests on a much smaller number of underlying cases. In the present example using data for 2000, the data are based on an approximate 1-in-6 (16.7 %) sample. Using more recent five-year summary files from the American Community Survey, the data would be based on a 1-in-20 (5 %) sample. Analysis of segregation between "matched" subsets of cases thus is likely to rest on a small set of cases in each block group.

Another problem is that, even under the best of conditions, it is usually infeasible to extend this conventional approach to take account of more than one or two non-racial characteristics at a time. Restricting the comparison to White and minority families matched on several characteristics at once will almost always result in basing the analysis on an unacceptably small number of micro-level cases. In contrast, the standardization approach applied in this chapter draws on the full population in each group and can in principle include many more social characteristics. The "ANOVA-style" reliance on categories instead of continuous predictors in the examples considered here can run into problems when means for some subgroups are less reliable due to being based on a small number of cases. However, the problem is less troublesome than the usual approach used in the literature. Moreover, it can be mitigated by using continuous measures in place of categories and adopting refined regression modeling strategies such as using multi-level specifications (discussed in Chap. 10) to improve estimation of effects. Thus, the difference of means framework provides clear advantages when researchers wish to take account of several non-racial characteristics at once.

9.6   Aggregate-Level Controls for Micro-level Determinants of Residential Outcomes

Segregation studies sometimes "take account" of group differences on social characteristics that play a role in residential outcomes in a fundamentally different way; namely, by estimating aggregate-level regressions where measures of group disparity on a relevant social characteristic (e.g., income or poverty status) is used to predict cross-city variation in segregation index scores. This strategy raises concerns about the risk of flawed inference associated with the "ecological" or "aggregate" fallacy.

It is fair to say that this concern does not seem to be widely recognized because the practice is routine in empirical studies and apparently not subject to strong criti-

cism.[6] Two factors may help explain why the prevailing practice is seen as non-controversial instead of seriously flawed. One is that traditional formulations of segregation indices encourage the view that the index score is an aggregate-level characteristic of cities that is not directly a product of individual-level attainment processes in way that would raise strong concerns about the undesirable consequences of the aggregate fallacy. The second is that, while studies in the location attainment tradition could potentially promote the view that segregation should be understood as arising out of micro-level residential attainment processes, they ultimately do not do so because until now micro-models could not be used to directly investigate segregation as measured by the dissimilarity index (D) and other popular aggregate-level indices.

The findings in this chapter show that analysis of segregation using popular aggregate-level measures can be joined seamlessly with analyses of micro-level residential attainment processes. The difference of means formulation of standard segregation indices makes this possible by establishing that segregation can be understood as a difference of group means on individual-level residential outcomes that in a given city are determined by a micro-level attainment process where many individual-level characteristics can impact segregation. The data and analyses presented in Tables 9.2, 9.3, and 9.4 clarify how the individual-level characteristics of race, poverty status, and family type affect residential outcomes (y) that then aggregate in a simple additive way to determine the level segregation in the city. This example establishes that the parallel with analyses of group differences other socio-economic attainment outcomes (e.g., education, occupation, income, home ownership, etc.) is exact. This then highlights a lack of correspondence on another point; namely, the failure of segregation researchers to show appropriate concern for the aggregate fallacy in aggregate-level segregation studies.

Researchers analyzing group differences in income understand that the aggregate-level outcome of inter-group income inequality in a particular city emerges as a product of an underlying micro-level process of income attainment for that city. As a result, it is easier for these researchers to recognize that the ideal way to obtain a sound assessment of the role that non-racial social characteristics play in producing group income inequality in a city is to draw on detailed micro data for that city. It also is easier for these researchers to recognize that attempts to take account of the role of non-racial social characteristics in producing inter-group income inequality using only aggregate data carries a high risk of mistaken inference due to the aggregate fallacy. I reviewed these issues more than two decades ago in an article that outlined the nature of the problem in detail and provided an empirical demonstration of how aggregate-level analysis leads to errors of inference and mistaken conclusions about the role of group differences in social characteristics for cross-area variation in group income inequality (Fossett 1988). Researchers interested in this topic appear to have adapted and moved forward. In recent decades there has been a fundamental change in the research literature. Aggregate-level analyses of cross-

---

[6] For example, in my experience journal reviewers not only do not object to this practice, they often *request* that it be incorporated into the analysis.

city variation in group income inequality were common in earlier decades and they routinely included aggregate-level measures to "control" for the impact of group differences on individual-level characteristics that predict income (e.g., income).[7] Such studies are no longer accepted as most researchers now understand that one must use disaggregated data to properly investigate these issues.

A similar reckoning is looming for the literature investigating cross-city variation in residential segregation. Concern about the aggregate fallacy currently is minimal because segregation researchers are not in the habit of viewing city-level segregation scores as mapping directly onto micro-level residential outcomes. Accordingly, segregation researchers do not automatically think in terms of using micro data to take account of the role of non-racial social characteristics in shaping residential segregation. This creates a "blind spot" for the possibility that key findings from studies that investigate cross-city variation in segregation may be suspect because the studies use research designs that incorporate the aggregate fallacy.

The data and analyses presented in Tables 9.2, 9.3, and 9.4 provide examples of how the differences of means approach makes it possible to "take account of" the impact of group differences on social characteristics on segregation in a way that is superior and offers a better chance to make correct inferences in comparison to past approaches. The data in these tables cast segregation as a group difference of means on residential outcomes (y) that emerge from a micro-level attainment process where race, poverty status, and family type all play a role in influencing residential outcomes. Once segregation is conceptualized in this way, it is clear that the proper statistical approach for taking account of group differences on poverty and family type is to perform city-specific standardization analyses using relevant attainment data disaggregated at the micro level for the city in question.

The limitations of the prevailing practice are revealed by the standardization analyses reported in Table 9.4. The results from the analyses directly answer the question of whether racial segregation arises due to group differences in poverty status and family type for the city in question. In each group comparison, the answer obtained is conceptually and statistically sound. The answer developed from analyses reported in Tables 9.1, 9.2, 9.3, and 9.4 also is definitive and complete. Group differences on social characteristics do not play a significant role in accounting for the observed level of White-Black segregation. This conclusion is anchored in a direct examination of the micro-level relationships between White and Black residential attainments in Houston in 2000. It cannot be improved by examining aggregate-level data for other White-Minority comparisons in the same city or even hundreds of such comparisons across other cities.

Moreover, analysis using only aggregate-level measures can easily lead to mistaken conclusions. For example, the analyses show that White-Asian segregation is lower than White-Black segregation and they also show that White-Asian differences in poverty are smaller than White-Black differences in poverty. The logic of

---

[7] Examples include Becker (1971 [1957]), Bahr and Gibbs (1966), Jiobu and Marshall (1971), Roof (1972), LaGory and Magnani (1979), and Elgie (1980) among many others as the practice was routine.

aggregate-level analysis would infer from this pattern that segregation is more pronounced when group differences in poverty are large. But analysis of the relationship using relevant micro-level data establishes that the impact of group differences in poverty is minimal.

Similarly, the result for the answer to the question cannot be improved by examining aggregate-level data for White-Black segregation in other cities. For example, if one examined a large sample of metropolitan areas and found a strong positive aggregate correlation between White-Black segregation and White-Black differences in poverty or income, the conclusion about the impact of White-Black differences in poverty on White-Black segregation in Houston based on the standardization analysis in Table 9.4 is not challenged and will stand unchanged. The aggregate-level findings are "trumped" by the direct analysis of relevant micro-level data for White-Black segregation in Houston.

In reviewing the general issues in detail in an earlier study (Fossett 1988) I noted that, while it is certainly plausible that group differences on social and economic characteristics could give rise to group differences on relevant attainment outcome, aggregate-level correlation is not a sound way to assess this possible effect. The sound way to assess the impact in a given city and group comparison is by working with relevant disaggregated data to examine the relationship at the micro level. Resorting to aggregate-level controls is tempting, but there are compelling reasons to discontinue this practice. One such reason can be summarized as follows.

> *Urban-ecological theories of cross-area variation in racial stratification provide a strong basis for expecting group differences on inputs to attainment processes to be spuriously correlated with group differences on outcomes of attainment processes at the aggregate level* (Fossett 1988).

The fundamental premise of urban-ecological theories of racial stratification is that some community-level factors shape group relations "across the board." If so, a general climate of minority disadvantage, tracing for example to comprehensive Jim Crow laws or high levels of White prejudice and discrimination in socioeconomic attainment processes, can lead to both high levels of White-Black differences in poverty and White-Black segregation. However, the resulting correlation of segregation with group differences in poverty and income produced by this social dynamic can easily be spurious, not causal. Thus, for example, if discrimination in housing severely constrains the residential opportunities of non-poverty Black households reducing their segregation-relevant contact with Whites, eliminating group differences in poverty will have no impact segregation.

It is not possible to sort out whether the aggregate relationship is spurious or causal with aggregate-level data. One must ultimately examine relevant micro data to directly assess whether reducing group differences in poverty or income would in fact reduce segregation in a given city. In Chap. 10 I present empirical analyses that illustrate how to perform such analyses. These analyses document and affirm that the empirical findings and central conclusions I reported in Fossett (1988) also apply to analyses of residential segregation. Specifically, the analyses document two parallels with the earlier study. The first is that aggregate-correlations and regres-

sions suggest that group differences in income play a major role in accounting for cross-city variation in segregation. The second is that this conclusion is shown to be incorrect when one uses micro-level data to properly take account of the impact of income differences on segregation. Based on this I caution segregation researchers to take seriously the concern that the practice of using aggregate-level regressions to assess the role of factors that operate at the micro-level is unsound and can yield misleading results.

9.7   New Interpretations of Index Scores Based on Bivariate Regression Analysis

Investigation of segregation using the technique of standardization analysis joins aggregate-level analysis with residential attainment analysis by clarifying how segregation index scores for a city arise from micro-level residential attainment processes shaped by racial and non-racial social characteristics. This point can be highlighted by noting that the data presented in Tables 9.2 and 9.3 correspond to predictions of mean residential attainments derived from individual-level models of residential attainment. More precisely, the subgroup means on residential outcomes correspond to predictions from individual-level analysis of variance (ANOVA) models or, alternatively, individual-level regression models. The tables reports means for residential attainments (y = scaled pairwise contact with Whites as relevant for S or D) for individual families grouped by category of race, poverty status, and family type. This corresponds to an individual-level ANOVA or regression analysis predicting residential attainments based on three categorical independent variables: family type (five categories), poverty-status (two categories), and race (two categories). Thus, the subgroup means reported in Tables 9.2 and 9.3 correspond to predictions from a "fully saturated" model which estimates all possible additive and non-additive effects for race, poverty status, and family type on residential attainments. The standardization analyses reported in Table 9.4 implicitly rest on these attainment models. It is a natural next step to explicitly focus on the results of the attainment model to assess more specifically how the effects of the independent variables shape residential segregation.

   The difference of difference of means framework yields a set of new and potentially attractive interpretations for segregation index scores. It is that the values of scores for indices such as S and D now can be described as reflecting the effect or impact of race on residential outcomes that determine segregation. This interpretation is straightforward in a bivariate model of individual-level residential attainment where race is the only predictor. When introducing the difference of means formulations, I offered computing formulas for obtaining index scores as a difference of group means. I now note that the index scores also can be obtained via an individual-level bivariate regression analysis in which a dummy variable for race (i.e., group membership) is used to predict the residential outcomes (y) that are relevant for a particular index.

Table 9.5 Bivariate segregation attainment regressions predicting residential outcomes (y) that additively determine White-Minority segregation for selected indices, Houston, Texas, 2000

Independent variable	G*	G/2*	D/2*	D	R	H	S
White-Black comparison (N=811,924)							
White (0,1 = White)	87.07	43.54	35.48	70.97	47.02	53.59	57.39
Constant	33.64	16.82	22.96	16.75	25.96	28.98	32.48
R Square	0.412	0.412	0.442	0.442	0.495	0.557	0.574
Implied index score	87.07	87.08	70.96	70.97	47.02	53.59	57.39
White-Latino comparison (N=911,060)							
White (0,1 = White)	74.19	37.09	29.19	58.37	28.11	35.46	40.96
Constant	49.53	24.76	30.14	23.12	35.26	37.50	40.17
R Square	0.359	0.359	0.317	0.317	0.385	0.404	0.410
Implied index score	74.19	74.18	58.38	58.37	28.11	35.46	40.96
White-Asian comparison (N=672, 968)							
White (0,1 = White)	76.28	38.14	29.11	58.22	34.96	31.31	23.88
Constant	29.94	14.97	23.27	16.93	35.74	52.16	69.91
R Square	0.131	0.131	0.122	0.122	0.135	0.198	0.239
Implied index score	76.28	76.28	58.22	58.22	34.96	31.31	23.88

Source: US Census 2000, Summary File 3
Notes: G* is given by $(Y_1 - Y_2)$ when y is scored 0–200 and G/2* and D/2* are given by $2 \cdot (Y_1 - Y_2)$ when y is scored 0–100. All regression coefficients are statistically significant at 0.001 or better

In the case of White-Black segregation as measured by the separation index (S), the regression would include a dummy variable for "White" coded 1 if White and 0 if Black to predict the residential outcome of pairwise contact with Whites (y = p). The value of the estimated regression intercept (b_0) will indicate the average contact with Whites for Blacks (i.e., the baseline group coded 0 on race). The value of the unstandardized regression coefficient for White (b_1) will indicate the extent to which the White mean for contact with Whites (i.e., the group coded 1 on race) deviates from the Black mean for contact with Whites. Accordingly, *the value of the regression coefficient also will exactly equal the value of the segregation index score; that is,* $b_1 = S$. (And, for the sake of completeness, mean contact with Whites for Whites will be given by $b_0 + b_1$.) At one level, this is not surprising as most readers will already be aware that bivariate dummy variable regression is mathematically equivalent to a difference of means comparison. But it is a new development in segregation measurement theory to interpret a segregation index score as the effect of race in a micro-level process of residential attainment. Thinking in this way opens up new avenues for exploring and interpreting segregation.

Table 9.5 reports results for a series of bivariate regressions of the type just described estimated using the micro-level data set for Houston, Texas introduced earlier in this chapter. Recall that this data set reconstitutes the block group-level summary tabulations so the information in the tabulation is organized in a data set appropriate for performing individual-level attainment analysis. Cells in the tabulation are treated as cases and are coded on independent variables – race, poverty status, and family type – to suit the needs of the analysis. Dependent variables relat-

ing to index-specific scores based on block-group race counts are assigned to block groups and then merged with the individual level data based on block group codes. The resulting data set can be used to estimate regression analyses in the conventional way with the proviso that a variable representing cell counts be used as a case-level frequency weight.[8]

The independent variable used in the bivariate regressions reported in Table 9.5 is a "dummy" (0, 1) variable for "White" coded 1 for White and 0 for minority depending on the race of the family's householder. The dependent variables are residential outcome scores (y) scaled from pairwise area proportion White (p) as appropriate for the segregation index of interest and the relevant group comparison. Table 9.5 reports results for separate regression analyses for five segregation indices – namely, G, D, R, H, and S.[9]

An important finding is evident in the results. In each regression analysis, the unstandardized regression coefficient for the dummy variable for race (here coded 1 if White and 0 otherwise) yields the value of the relevant index score (previously reported in Table 5.2). In the case of G, individual residential outcomes (y) are scored two ways; one to yield G and one to yield G/2. In the latter coding, the value of the coefficient for race must be doubled to obtain the value of G. The table also reports results for D taken as a crude version of G and thus scores residential outcomes (y) in relation to D/2. In this regression the coefficient for race must be doubled to obtain the value of D. The table also reports results based on the alternative formulation of D where residential outcomes (y) are coded as either 0 or 1.

Of course, relatively little is gained if we stop with the simple bivariate regression analysis. It merely recasts the difference of means comparison reviewed earlier in Table 5.2 in the regression (or ANOVA) framework. The most important descriptive findings to be gleaned from the analysis – namely, the group means and the group difference of means – are exactly the same as those previously reported earlier in Tables 5.2 and 5.3. So no new information is gained.

Regression analysis does potentially provide a useful framework for hypothesis testing regarding the level of segregation. But this has minimal practical value at the bivariate level of analysis as statistical significance is typically not a central concern in segregation analysis. Sample sizes and race effects both tend to be large in analyses of the overall level of segregation and thus statistical tests tend to be significant at levels far beyond conventional standards (i.e., 0.05 and 0.01). For example, the

---

[8] When estimation routines in statistical programs have this capability, the data set can be stored an efficient, compact form. Alternatively, one may create a separate record for each family included in the summary file tabulation. The resulting data set will be much larger. Some might find the less compact form of the data set more familiar for conducting individual level regression analysis. But regression results will be identical either way.

[9] The regressions were estimated using OLS regression. This is satisfactory for present purposes and is a convenient choice because it simplifies the presentation and discussion of results. In other situations, it may be necessary to use more technically appropriate regression procedures such as fractional logit regression (Papke and Wooldridge 1996; Wooldridge 2002) to deal with the problem that OLS assumptions are not valid when modeling bounded variables and OLS can yield predictions outside of the 0-1 bounds of segregation indices.

t-ratio for the effect coefficient of race in the bivariate regression of pairwise contact with Whites on a dummy variable for race for the White-Black comparison is over 1000 and the probability of chance deviation from 0.0 is zero out to many decimal places. In such circumstances, the usual concerns about statistical significance and technical regression assumptions fade into the background.

The more significant potential benefit of regression analysis is that it provides an opportunity to put segregation research on a new path for gaining a better, more direct understanding how segregation arises. Specifically, analyzing segregation from the difference of means framework sets segregation researchers on the path of investigating segregation using the methods and modeling strategies that status attainment researchers routinely use to investigate racial disparities and inequality on education, occupation, income, health, and other socioeconomic and life chance outcomes. These methods and modeling strategies previously have not been available to segregation researchers because the link between micro-level attainment and aggregate-level segregation (city segregation scores) was not established. The difference of means formulation of segregation indices thus allows researchers to move away from focusing simply on the calculation of descriptive index scores that summarize the state of segregation at the aggregate level. It instead allows researchers to move toward investigating segregation through the more analytically flexible method of performing multivariate analyses to assess the zero-order and net effects of race (group membership) on individual-level residential outcomes that directly determine the level of segregation.

I discuss the extension to multivariate analysis of segregation-relevant residential outcomes in more detail below. But first it is useful to point out that different indices register residential outcomes in different ways – based on index-specific functions $y = f(p)$ – and that these differences carry implications for interpreting the effect of race on residential outcomes in individual-level attainment analyses. Here it is useful to recall Fig. 5.1 which clarifies how these five segregation indices differ in registering residential outcomes (y) scored from area group proportion (p). In the case of G, D and R, y is scored as a nonlinear transformation of p that in these group comparisons tends to exaggerate group differences at high levels in p and minimize group differences over the middle ranges of p. H also involves a similar nonlinear transformation, but it is much less dramatic. In contrast, S scores y simply on the basis of p and does not subject p to a transformation. This makes the regression results for the separation index (S) especially easy to interpret and a good place to begin.

Table 9.5 reports the results for the bivariate regression $y = b_0 + b_1$ (race) relevant for investigating White-Black segregation as measured by S as $y = 32.5 + 57.4$ (race) where race is coded 1 for White and 0 for Black. In this example, the value of the regression constant (b_0) is 32.5 and reflects Blacks' average contact with Whites (Y_B). The value of the unstandardized regression coefficient for race (b_1) is 57.4. It reflects the impact that race has on average contact with Whites; namely, to raise contact with Whites by 57.4 points above the level of contact that Blacks experience. The sum of b_0 and b_1 gives the predicted value of 89.9 for Whites' average contact with Whites (Y_W). These values map exactly onto the terms reported in

Table 5.3 which showed how index scores can be obtained as differences of group means on residential outcomes. Thus, the value of S for White-Black segregation overall is 57.4 resulting because White families live in neighborhoods where pairwise percent White averages 89.9 while Black families live in neighborhoods where pairwise percent White averages 32.5.

This highlights the new interpretation available for S as indicating that race – specifically, being White instead of Black – "matters" for residential outcomes and in this case has the impact of increasing contact with Whites by 57.4 points in comparison with the reference group of Blacks. The magnitude of the effect makes it clear that race differences in residential attainment produce substantial residential separation between Whites and Blacks as Whites are predicted to reside in predominantly White areas and Blacks are predicted to live in predominantly Black areas.

It is instructive to compare the effect of race in the White-Black comparison with the effects of race in the bivariate segregation attainment analyses for the White-Latino and White-Asian comparisons. The race effect of 41.0 points in the White-Latino regression is approximately 16 points lower than that in the White-Black regression. Thus, we can conclude that race "matters less" in promoting residential separation of Whites from Latinos than it does in promoting residential separation of Whites from Blacks. However, the effect is still large and has the consequence of on average placing Whites in predominantly White areas while Latinos are in predominantly Latino areas. The race effect of 23.9 points in the White-Asian regression is approximately 34 points lower than in the White-Black regression. Based on this we can conclude that, while the effect of race is not trivial, race matters much less in promoting residential separation of Whites from Asians than it does in promoting residential separation of Whites from Blacks. One clear indication of this is that the effect of race on average leaves both Whites and Asians being predicted to reside in predominantly White areas.

The bivariate results for D suggest a somewhat different story. I focus on the results for D based on scoring residential outcomes as 0 or 1 based on whether the family attains parity on contact with Whites based on whether area proportion White equals or exceeds the level for the city as a whole (i.e., 1 if $p \geq P$, 0 otherwise). For this residential outcome, race matters a great deal in all three group comparisons. The unstandardized regression coefficients for race take high values in each analysis reaching approximately 71.0 in the White-Black analysis, 58.4 in the White-Latino analysis, and 58.2 in the White-Asian analysis. In substantive terms, we can interpret these effects as indicating that, in each comparison, race – that is being White in contrast to being minority – has a large impact on the probability of residing in an area where proportion White attains parity with city-wide proportion White.

This information is not without value. But it also is important to be aware of what is not revealed when modeling micro-level outcomes that determine the value of D. Namely, this analysis fails to provide a basis for assessing the quantitative differences in the racial composition of the neighborhoods the groups live in. If one does not bear this in mind, one can come away with an incomplete and potentially misleading impression of the nature of segregation in these three comparisons. This

is particularly true in the case of the White-Latino and White-Asian analyses. The comparison on the effect of race in these analyses shows that it is essentially the same in both two regression equations. This indicates that the White advantage in the probability of attaining parity on area proportion White is the same in relation to Latinos and Asians. In addition, comparison of the regression constants indicates that, overall, Asians are less likely than Latinos to attain parity on area proportion White. The combination suggests that White-Latino segregation and White-Asian segregation are very similar.

But it is important to bear in mind that D is sensitive to group differences in attaining "parity" on neighborhood proportion White *where "parity" is assessed in relation to the citywide pairwise racial proportions*. As a result, the effect of race in the models for D does not support inferences and interpretations relating to group differences in the actual level of pairwise contact with Whites or to group differences in "fixed" outcomes such as probabilities of residing in neighborhoods that are majority (50 %) White, two-thirds (67 %) White, or predominantly (e.g., 80 %) White. For example, in the case of Houston, Texas, Latinos are a much larger group than Asians. Accordingly, the "cut point" for scoring of residential outcomes as attaining "parity" on area proportion White for the White-Latino comparison is much different – specifically, much lower – than the "cut point" for scoring of residential outcomes as attaining "parity" on area proportion White for the White-Asian comparison. Consequently, a naive interpretation of the race effect in the attainment analysis for D might suggest the conclusion that Latinos and Asians fare similarly in comparison to Whites but with Asians being less likely than Latinos to live in areas that are disproportionately White. But the analysis for S shows that Asians live in areas that on average are 69.9 % White, a full 29.7 points higher than Latinos who on average live in majority Latino areas. This suggests that the substantive value of scoring residential "disadvantage" based on "parity" is open to reconsideration. In particular, I pose the question, "What are the substantive and sociological implications of Asians experiencing near-identical disadvantage as Latinos on attaining "parity" when the two groups differ greatly in terms of their residential separation from Whites?"

9.8   Multivariate Segregation Attainment Analysis (SAA)

The bivariate regression examples just discussed are interesting and useful in their own right. They illustrate some of the benefits of directly modeling the individual-level residential outcomes that give rise to segregation index scores. Specifically, the approach enables and encourages more thoughtful and careful interpretation of race effects on residential outcomes across group comparisons and different indices. In the long run, however, the bivariate regressions are just a useful preliminary step toward investigating how the impact of race on segregation compares with the impacts of other social characteristics. This can be done by investigating micro-level analyses segregation-relevant residential outcomes using multivariate

attainment models in the manner that is already universal in other literatures investigating racial disparities.

I term this new approach "segregation attainment analysis" (SAA). The justification for the label is that the effect of race in bivariate models corresponds directly to the aggregate level of segregation in the city and its effect in multivariate models yields insights into how the impact of race should be assessed and interpreted when taking account of the role of non-racial factors that also impact residential outcomes.

This can be accomplished by extending the micro-level attainment regressions to include additional independent variables beyond race. In this case, I used the tabulation of race by family type by poverty status to fashion the following independent variables: poverty status (0,1), married with spouse present (0,1), and presence of children under age 18 (0,1). Table 9.1 previously presented descriptive statistics based on this data set. It documents that the four groups in the analysis vary greatly on these variables. Non-poverty status runs from a high of 95.8 % for Whites to a low of 80.2 % for Latinos. Percent of families that are married couple families runs from a high of 84.9 % for Asians to a low of 51.2 % for Blacks. And percent of families with children under age 18 runs from a high of 69.2 % for Latinos to a low of 47.3 % for Whites. Given these group differences in distribution across social characteristics an obvious questions arise: "What role to these characteristics play in shaping residential outcomes that determining segregation?", "How does their role compare with the role of race?", and "How does the estimated effect of race change when other characteristics are controlled?"

Tables 9.6, 9.7, and 9.8 report results of bivariate and multivariate regression analyses that can be used to address these and related questions. Each table has five panels. Each of the five panels presents results from regression analyses predicting dependent variables that additively determine segregation indices. The regression analyses are estimated and reported separately by racial group for ease of discussion and presentation. For hypothesis testing and for cross-time and cross-city comparisons it may be more appropriate to estimate single-equation specifications which incorporate additive and non-additive race effects. White-Black comparisons are reported in Table 9.6, White-Latino comparisons in Table 9.7, and White-Asian comparisons in Table 9.8.

Analyses of this sort can be used to gain a richer understanding of the residential attainment process that gives rise to segregation by permitting direct examination and comparison of the separate effects of racial *and* non-racial social characteristics on residential outcomes. Table 9.6 presents results relevant for the analysis of White-Black segregation. Results are presented separately for five indices. I begin by discussing the results for the separation index (S) reported in the fifth panel in the table. The first and second columns report separate regressions for Whites and Blacks with no other independent variables included in the model. The constants in these equations of course equal the group means for scaled contact with Whites (y). In the case of the separation index (S) y is given by the pairwise proportion White (p) in the block group and difference between the two group means yields the value of the separation index. This is reported as "White Advantage (S)" which has the

Table 9.6 Group-specific attainment regressions for White-Black segregation

Variable	Simple comparison		Comparison with controls		Net impact
	Whites	Blacks	Whites	Blacks	
Residential outcome (y) scored for gini index (G)					
Non-poverty Family	–	–	5.66	8.37	−2.71
Married couple family	–	–	8.62	7.83	0.79
Children present	–	–	0.79	4.58	−3.79
Constant	120.71	33.64	107.67	20.03	87.64
White advantage (G)	87.07		87.64		81.93
Residential outcome (y) scored for index of dissimilarity (D)					
Non-poverty family	–	–	6.72	6.45	0.27
Married couple family	–	–	5.76	8.95	−3.19
Children present	–	–	2.89	3.07	−0.18
Constant	87.73	16.75	75.09	5.05	70.04
White advantage (D)	70.98		70.04		66.94
Residential outcome (y) scored for hutchens index (R)					
Non-poverty family	–	–	3.72	6.70	−3.98
Married couple family	–	–	3.67	5.29	−1.62
Children present	–	–	0.33	3.98	−3.65
Constant	72.98	25.96	67.14	15.38	60.42
White advantage (R)	47.02		60.42		42.52
Residential outcome (y) scored for theil index (H)					
Non-poverty family	–	–	3.37	8.19	−4.82
Married couple family	–	–	3.75	6.86	−3.11
Children present	–	–	0.93	4.83	−3.90
Constant	82.57	28.98	75.75	15.85	59.90
White advantage (H)	53.59		59.90		48.07
Residential outcome (y) scored for separation index (S)					
Non-poverty family	–	–	3.58	9.59	−6.01
Married couple family	–	–	3.34	8.19	−4.85
Children present	–	–	1.20	5.67	−4.47
Constant	89.86	32.48	83.06	17.01	66.05
White advantage (S)	57.38		66.05		50.72

Source: Summary File 3, Houston, Texas, 2000. Sample N: 627,613 for Whites and 195,928 for Blacks. All coefficients are statistically significant at 0.001 or better

value of 57.38. This value of S was reported previously in Tables 5.2, 5.3 and 9.5 and thus confirms the equivalence of the different approaches to assessing segregation.

The third and fourth columns report multivariate regressions separately for Whites and Blacks. Each equation has three independent variables – non-poverty status, married couple family, and presence of children – which have been coded as dummy (i.e., 0,1) variables. In this specification, the intercept of the equation can be interpreted as the expected group mean on scaled contact with Whites for families that are in poverty, are not married couple families, and do not have children residing with them.

Table 9.7 Group-specific attainment regressions for White-Latino segregation

Variable	Simple comparison		Comparison with controls		Net impact
	Whites	Latinos	Whites	Latinos	
Residential outcome (y) scored for gini index (G)					
Non-poverty family	–	–	14.94	15.33	−0.39
Married couple family	–	–	16.56	6.33	10.23
Children present	–	–	5.28	−2.70	7.98
Constant	123.72	49.53	93.01	34.41	58.60
White advantage (G)	74.19		58.60		76.42
Residential outcome (y) scored for index of dissimilarity (D)					
Non-poverty family	–	–	12.92	11.46	1.46
Married couple family	–	–	9.94	5.32	4.62
Children present	–	–	4.90	−3.76	8.66
Constant	81.49	23.12	58.46	12.58	45.88
White advantage (D)	58.37		45.88		60.62
Residential outcome (y) scored for hutchens index (R)					
Non-poverty family	–	–	5.13	7.18	−2.05
Married couple family	–	–	5.10	2.56	2.54
Children present	–	–	1.69	−0.73	2.42
Constant	63.37	35.26	53.38	28.10	25.28
White advantage (H)	28.11		25.28		28.19
Residential outcome (y) scored for theil index (H)					
Non-poverty family	–	–	6.60	9.33	−2.73
Married couple family	–	–	6.01	3.34	2.67
Children present	–	–	2.91	−1.01	3.92
Constant	72.95	37.50	60.49	28.24	32.25
White advantage (H)	35.45		32.25		36.11
Residential outcome (y) scored for separation index (S)					
Non-poverty family	–	–	7.65	11.30	−3.65
Married couple family	–	–	6.23	3.97	2.26
Children present	–	–	2.72	−1.24	3.96
Constant	80.12	40.17	67.28	29.01	38.27
White advantage (S)	40.95		38.27		40.84

Source: Summary File 3, Houston, Texas, 2000. Sample N is 627,613 for Whites and 294,931 for Latinos. All coefficients are statistically significant at 0.001 or better

The difference between the intercepts of the two equations can be interpreted as a White-Black segregation comparison that has been "standardized" to control for group differences in distributions on social characteristics. That is, the comparison reflects group differences on model predicted means on segregation-determining residential outcomes for White and Black families that are matched on social characteristics. For both Whites and Black the level of average contact with Whites for the subgroup reflected at the intercept is lower than the group's overall mean. The value for Whites is 83.06 which is 6.80 points lower than the value of 89.86 reported in the "constant only" equation for Whites. The value for Blacks is 17.01

Table 9.8 Group-specific attainment regressions for White-Asian segregation

Variable	Simple comparison		Comparison with controls		Net impact
	Whites	Asians	Whites	Asians	
Residential outcome (y) scored for gini index (G)					
Non-poverty family	–	–	−11.69	5.97	−17.66
Married couple family	–	–	0.91	4.11	−3.20
Children present	–	–	−2.67	*−0.11*	−2.56
Constant	106.22	29.94	117.91	21.12	96.79
White advantage (G)	76.28		96.79		73.37
Residential outcome (y) scored for index of dissimilarity (D)					
Non-poverty family	–	–	−4.98	3.09	−8.07
Married couple family	–	–	2.72	3.95	−1.23
Children present	–	–	*−0.17*	−1.29	1.12
Constant	75.14	16.93	77.72	11.55	66.17
White advantage (D)	58.21		66.17		57.99
Residential outcome (y) scored for hutchens index (R)					
Non-poverty family	–	–	−5.68	4.38	−10.06
Married couple family	–	–	*0.17*	2.42	−2.25
Children present	–	–	−1.57	*0.27*	−1.84
Constant	70.69	35.74	76.74	29.56	47.18
White advantage (H)	34.95		47.18		33.03
Residential outcome (y) scored for theil index (H)					
Non-poverty family	–	–	−2.91	6.71	−9.62
Married couple family	–	–	0.74	3.48	−2.74
Children present	–	–	−0.59	0.54	−1.13
Constant	83.47	52.16	85.91	42.82	43.09
White advantage (H)	31.31		43.09		29.60
Residential outcome (y) scored for separation index (S)					
Non-poverty family	–	–	−0.72	8.86	−9.58
Married couple family	–	–	0.81	4.18	−3.37
Children present	–	–	*0.00*	0.97	−0.97
Constant	93.79	69.91	93.80	57.78	36.02
White advantage (S)	23.88		36.02		22.10

Source: Summary File 3, Houston, Texas 2000. Sample N is 627,613 for Whites and 55,746 for Asians. Regression coefficients not significant at 0.01 are in gray italics

which is 15.47 points lower than the value of 32.48 reported in the "constant-only" equation for Blacks.

The White-Black difference of 66.05 at the intercept (83.06 minus 17.01) is reported in the third column of the "White Advantage" row. (As discussed below, it also is reported as a "net impact" in the fifth column.) This value can be understood as the impact of race on expected scaled contact with Whites for White and Black families that have the specific configuration of social characteristics associated with the intercept of the multivariate equation. Thus, it is the White-Black difference on average scaled contact with Whites predicted under the model for families coded

zero on all three independent variables (i.e., for non-married couple families with no children present and in poverty). In the bivariate regressions the impact of race represents the level of segregation in the city because it is exactly equal to the segregation index score. In the multivariate specification the impact of race can be interpreted as the expected level of segregation between Whites and Blacks when group differences in distribution on other social characteristics is controlled.

The group-specific regression coefficients reported in columns 3 and 4 give insights into how the three social characteristics included in the regression impact the residential attainments that additively determine segregation as measured by S. The regression coefficients for this analysis indicate that all three variables – non-poverty status, married couple status, and presence of children – have positive effects on family attainments of the residential outcome of scaled contact with Whites. This pattern is generally consistent across the multivariate regression analyses reported for all three White-Minority comparisons and for all five segregation indices considered. The effect of non-poverty status is positive and statistically significant in all equations. The effect of married couple status is positive and statistically significant in almost all equations. The effect of presence of children is less consistent. In the analyses for White-Black segregation it is positive and statistically significant in all of the equations but is small in size for Whites in analyses for some measures of segregation. In the analyses for White-Latino segregation the effect is positive for Whites and negative for Latinos. In the analyses for White-Asian segregation it is mixed in terms of direction but consistently small (absolute value under 1.0 in 7 of 10 possible cases).

The question of how these social characteristics impact segregation is answered by examining whether their effects ultimately reduce White-Minority differences on segregation-determining residential outcomes. For example, in the analyses for White-Black segregation moving from poverty to non-poverty status increases Black contact with Whites by 9.59 points. The comparable effect for Whites is 3.58. The "net impact" (i.e., White-Black effect difference) is –6.01 points and is reported in column five. This has direct implications for segregation. Specifically, it indicates that if one starts with White and Black families in poverty that are matched on other social characteristics and then move these families from poverty to not in poverty it would reduce segregation by 6.01 points. As a quick methodological aside, this "net impact" interpretation is based on using a linear, additive regression specification. Moving to a nonlinear and/or non-additive model for estimating effects of non-racial characteristics would require a more nuanced approach to assessing effects.[10]

In the analysis of White-Black segregation as measured by S the "net impact" (i.e., White-Black effect difference) is negative for all three social characteristics considered. Thus, in the same sense that the "net impact" indicates that moving from "poverty" to "non-poverty" reduces segregation by 6.01 points, moving from "non-married couple" to "married couple" on family type reduces segregation by

---

[10] Specifically, in nonlinear and/or non-additive models, the impact of a change from poverty to non-poverty would have to be assessed separately for each the initial configuration of the other social characteristics.

4.85 points and moving from "without children" to "with children" reduces segregation by 4.47 points. In the context of the linear, additive model used here, implementing all three "net impact" effects simultaneously would reduce the expected "White Advantage" in contact with Whites by 15.33 points; it would move the "White Advantage" from 66.05 at the intercept – that is, for the White-Black comparison standardized to non-married couple families without children and in poverty – to 50.72 for the White-Black comparison standardized to non-poverty, married couple families with children. This is reported in column five of the "White Advantage" row in the results.

These results help clarify how the impacts of racial group membership and non-racial social characteristics on segregation can be investigated in a more careful and nuanced way. The "net impact" reported in column five provides insight into the proximate impact of group differences on social characteristics on segregation. The regression coefficients reported in columns three and four clarify how the "net impact" comes about. Including the group-specific regression constants in the discussion provides a basis for comparing how the additive and non-additive effects of race compare with the effects of other factors in shaping segregation.

In the multivariate framework a wide range of logical possibilities can be imagined. At one extreme all block groups could have identical values on pairwise proportion White. In this possible but unlikely scenario of exactly zero race segregation the regression coefficients for non-racial social characteristics would be zero in both group equations and the intercepts of both equations would be identical. Another possibility is that race segregation is present and is due only to simple additive effects of race. In one scenario for this pattern, the regression coefficients for non-racial social characteristics are zero in both group equations but the intercept is higher in the equation for Whites and lower in the equation for Blacks. In a more complex scenario, the group equations differ at the intercept as just described and the regression coefficients for other social characteristics are not zero but are identical for both groups and both groups have identical distributions on the social characteristics.

A more plausible scenario is that race segregation is present and is produced by a complex combination of contributing factors including the following: additive race effects (i.e., differences at the intercepts of the attainment equation), non-additive race effects (i.e., race differences in the effects of non-racial characteristics), race differences in distribution on non-racial social characteristics, and the "interaction" of the last two factors. The results in Table 9.6 provide evidence that additive race effects are the most important factor contributing to segregation. The "White Advantage" of 66.05 reported in column three is one estimate of the quantitative contribution. This value can be described as the impact of race on segregation-determining residential outcomes for non-married couple families without children who are in poverty. That is, it is the value that would be estimated for the effect of being White (coded 1 if White and 0 if Black) if the regression analyses reported in columns 3 and 4 were replicated in a single equation specification using the combined samples.

The value of the intercept enters into all predictions and in this model specification the intercept corresponds to a set of families with a specific profile on social characteristics. So it is fair to describe the observed race difference at the intercept as applying "across the board" since reflects the expected level of segregation when social characteristics are fixed. Of course, the specific value of the intercept can vary depending on how variables are coded. So it is reasonable to ask whether the value of 66.05 is a fair or representative choice among all of the possible estimates of expected segregation for White-Black comparisons matched on social characteristics. The model predictions provide one answer to that question. Since all net impact calculations in column 5 are negative, 66.05 is the *maximum* race difference the attainment models will predict for White and Black families matched on all social characteristics. In contrast, the race difference of 50.72 predicted for the White-Black comparison for non-poverty, married-couple families with children present is *minimum* difference the attainment analysis will predict for White and Black families matched on all social characteristics.

This is useful information to consider. One can also apply predictions from the model to a hypothetical "standard" distribution of social characteristics to obtain expected White-Black differences on segregation-determining residential outcomes for "matched distributions." Results for this kind of standardization analysis were reported in Table 9.4 based on adopting the combined group distributions as the "standard" for matching Whites and Blacks on social characteristics. The White-Black difference obtained under this calculation was 54.38, which necessarily falls between the minimum and maximum predicted differences of 50.72 and 66.05.

The question at hand here is how the effect of race on segregation compares to the effect of other social characteristics. A range of estimates of the impact of race are on the table. The "net impact" estimates in column 5 provide one basis for assessing the impact of other social characteristics on segregation. The separate net impact estimates range from –4.47 to –6.01 and are small compared to the race effect. The impact of non-poverty status is the largest of the three values and its magnitude is less than 12 % of the lower-bound estimate of the additive impact of race. If one combines the impacts of all social characteristics to obtain the maximum possible combined effect on reducing segregation, the result is 15.33 which is about 30 % of the lower-bound estimate of the additive impact of race. On this basis, one can argue that race is clearly the dominant factor impacting residential outcomes that determined segregation. Poverty status, family type, and presence of children do impact segregation. But their effects on segregation are small compared to the broad effect of race.

Standardization and decomposition analysis can provide additional perspective on the role non-racial social characteristics play in shaping segregation. For example, Table 9.1 reported that 80.9 % of Black families were in not in poverty compared to only 95.8 % of White families. If the non-poverty rate for Black families was increased to match the rate of observed for White families, the model indicates segregation would be reduced by 1.43 points. This is less than 3 % of the lower-bound estimate of the effect of race. From many different vantage points, the analysis consistently indicates that White-Black differences in distribution on social

characteristics are not the major factor in determining segregation; the vast majority
of segregation is due to expected mean differences on segregation-determining out-
comes – in the case of S, pairwise contact with White – between Whites and Blacks
matched on social characteristics.

Similar findings emerge in the analyses of residential outcomes relevant for
determining the separation index (S) for White-Latino segregation (reported in
Table 9.7) and for White-Asian segregation (reported in Table 9.8). In both cases,
the net impact calculation for race based on the multivariate analysis of segregation-
determining residential attainments (i.e., the value of White advantage reported in
column 5) is much larger than the net impact calculations for the other social char-
acteristics included in the analysis. The same general finding holds up across all
three White-Minority segregation comparisons in analyses focusing on residential
attainments relevant for G, D, R, and H. That is, the net impact of race on index-
specific, segregation-determining residential outcomes is consistently much larger
than the net impact estimates for the other social characteristics considered in these
analyses.

These general conclusions are appropriate. But close inspection of the detailed
results reveals interesting differences across White-Minority comparisons and
across analyses focusing on different segregation indices. For example, in the case
of White-Black segregation, the net impact calculation for non-poverty status varies
across indices. Its absolute and relative magnitude is largest in the analysis focusing
on the separation index (S) and is small and modest in the analyses focusing on the
gini index (G) and the dissimilarity index (D). This is also true in the case of White-
Latino segregation. But the pattern is different in the analysis results for White-
Asian segregation. Here the net impact calculation for non-poverty status is sizeable
for all indices and largest of all in the results for the gini index (G).

I conclude this section by noting that other interesting results can be discovered
by making comparisons across groups. For example, the combined net impact cal-
culations for married couple status and children present serve to *reduce* White-
Black segregation across analyses for all segregation indices. A very similar pattern
is also found in the results for the analyses of White-Asian segregation. But a much
different pattern is seen in the results for the analyses of White-Latino segregation.
The combined net impact calculations for married couple status and children pres-
ent serves to increase segregation in the analyses for all segregation indices with the
magnitude of the combined impact being especially large in the case of the gini
index (G) and the dissimilarity index (D). These intriguing results and highlight
how the new approach opens the door for pursuing more careful exploration of the
social processes that produce White-Minority segregation. Future research may pro-
vide insight into why family structure appears to play a different role in White-
Latino segregation in comparison with White-Black and White-Asian segregation.
These and other possibilities for future analysis highlight the advantages of adopting
the difference of means framework and embracing its capabilities for exploring seg-
regation patterns in more detail.

9.9   Unifying Aggregate Segregation Studies and Studies of Individual-Level Residential Attainment

For many decades, dating back at least to the late 1960s, studies of segregation have followed one path while studies of racial and ethnic inequality and disparity on socioeconomic outcomes such as education, occupation, income, wealth, and home ownership have followed a different path. In the broader literature on racial socio-economic inequality and disparity it is conventional to see racial disparities on socioeconomic attainment outcomes (e.g., education, income, etc.) as emerging from micro-level processes of attainment. Accordingly, research focusing on inter-group inequality and disparity on most socioeconomic outcomes draws on micro-level attainment models to understand and analyze group differences on socioeconomic attainments.

This has not been the case in the study of residential segregation. To be fair, researchers understand that at some level residential segregation arises from micro-level processes wherein individuals and groups seek, compete for, and attain (or fail to attain) particular residential outcomes. But past statements on segregation measurement have focused almost exclusively on the task of aggregate-level description. Relatively little attention has been given to developing connections between index scores for uneven distribution and residential outcomes for individuals and families that are considered in studies of residential attainment.

I noted earlier that Duncan and Duncan lamented this fact observing that the literature on segregation indices provided no "suggestion about how to use them to study the *process* of segregation" (1955:216, emphasis in original). Unfortunately, the negative assessment they offered more than five decades ago applies with equal force today. Research clarifying how micro-level attainment dynamics give rise to aggregate segregation as measured by popular indices of uneven distribution is not well-developed. In my view, the point of concern that Duncan and Duncan raised has taken on much greater importance in the five decades that have passed since their study. In general, research on racial and ethnic differences in socioeconomic outcomes has advanced considerably based on steady, cumulative improvements in our understanding of how group differences in aggregate attainments arise from micro-level attainment dynamics. But this has not been the case in the subfield of segregation research. Until now there has been little progress in developing a better understanding of how aggregate level segregation (as measured by indices of uneven distribution) is linked with individual-level residential outcomes and the micro-level processes that shape them.

Of course, there is a large and vital literature that investigates micro-level dynamics of residential attainment. Studies using individual-level data to focusing on spatial assimilation and spatial attainment first appeared in the 1980s (e.g., Massey and Mullan 1984; Massey and Denton 1985) and then with increasing frequency in the 1990s and beyond (e.g., Alba and Logan 1993; Alba et al. 1999; Bayer et al. 2004; Crowder and South 2005; Crowder et al. 2006; Logan et al. 1996; South and Crowder 1997, 1998; South et al. 2005a, b; South et al. 2008). But, as valuable as

this literature has been, it has remained fundamentally disconnected from the literature investigating segregation at the aggregate level. The reason for this is simple; the literature on segregation measurement has never provided a simple, direct strategy for connecting segregation at the aggregate level (i.e., for a city) to individual residential attainments.

Casting indices of uneven distribution as group differences in means on individual residential outcomes addresses this gap in segregation studies. It establishes a simple, direct connection between individual residential outcomes and segregation index scores and in doing so creates the possibility of unifying studies of aggregate segregation and studies of residential attainment in a common overarching framework. Specifically, this approach allows researchers to *simultaneously* investigate both individual residential attainments *and* aggregate segregation in a single analysis. I noted earlier in this chapter that aggregate segregation now can be understood as the effect of group membership (coded 0–1) on the relevant residential outcome in a simple bivariate regression model of individual residential attainment.[11] But this is only a starting point for analysis, not an end point. The approach can be readily extended in a variety of ways that move the investigation of segregation beyond simply assessing aggregate-level uneven distribution.

Casting segregation as a difference of means on individual residential outcomes puts the investigation of segregation on the same methodological footing as the investigation of inter-group inequality and disparity on other important socioeconomic outcomes such as education and income. The key to this is that group disparity is conceived and modeled as emerging directly from an individual-level attainment process. This fundamental change in conceptualization opens up important new options for research. For example, it makes it possible to assess the role of social characteristics such as income using fine-grained measurement such as continuous measurement of income instead of crude category distinctions as used in current practice. Even more importantly, it makes it possible to take account of multiple social characteristics in analyses investigating group segregation; something that is difficult if not impossible to implement using standard methodological approaches to investigating segregation.

These new options become possible because multivariate modeling of individual residential outcomes provides a superior – specifically, a statistically more efficient – framework for taking account of the role of multiple social characteristics (including both race and non-racial characteristics). In this context, implications for aggregate-level segregation can be assessed using methods that are widely used in

---

[11] The analysis can be conducted using conventional OLS regression or analysis of variance. Statistical tests that rest on the assumption of normality and equal variances for the error term across groups may be questionable on technical grounds in some cases. But the typically large sample sizes used in segregation studies will minimize concerns about these issues. In any event, many good statistical alternatives are available. Boot-strapping or other methods may be used to perform statistical tests that do not rely on assumptions regarding normality and equal variances. Alternatively, the effect of group membership can be assessed using more technically appropriate modeling frameworks such as fractional regression (Papke and Wooldridge 1996) or beta regression (Smithson and Verkuilen 2006; Buis 2006; Buis et al. 2006).

the study of racial inequality and disparity in other socioeconomic outcomes. For example, regression standardization methods can be used to examine differences in residential outcomes for groups that are statistically matched on relevant social characteristics (i.e., other than group membership). Similarly, components analysis can be used to assess the contributions to aggregate segregation of group differences in attainment resources and group differences in ability to convert resources into attainments. These and related methods provide valuable new options for gaining a better understanding of the factors that produce segregation and new options for exploring the potential of different policies to impact aggregate segregation.

Regression-based analysis carries advantages on all these points. In general, the advantages derive from the fact that multivariate regression analysis is a more statistically efficient method with which to account for the effects of multiple social characteristics when comparing groups on average attainments on residential outcomes. Specifically, the statistical efficiencies of the regression standardization approach make it feasible to: (a) incorporate multiple non-racial social characteristics in the analyses and obtain reliable estimates of their separate effects on relevant residential attainments, (b) model the role of continuous social characteristics (e.g., income) in as much detail as the tabulations (or, as will be discussed below, microdata) will permit, (c) perform comparisons in cities where the small relative size of the minority population makes application of previous approaches problematic, and (d) perform significance tests of the role of race (i.e., group membership) on residential outcomes with social characteristics controlled.[12]

The empirical examples reviewed here provide preliminary illustrations of how the new methods can be used to good effect. But the next section shows that the examples introduced above only hint at what is possible. The new methods used in these examples permit one to imagine new options for analysis using micro data that can go far beyond what might be accomplished using traditional approaches for incorporating non-racial social characteristics into segregation analyses.

9.10 New Possibilities for Investigating Segregation Using Restricted Data

The methods introduced in this chapter permit researchers to investigate segregation in more detail than was previously possible. But the potential benefits of the new methods are relatively modest when segregation is investigated using publicly distributed census summary file tabulations. Summary file tabulations have been the "life blood" of segregation research to date. They have sustained traditional approaches to investigating residential segregation and, at least to some degree, they also can sustain analyses of individual residential attainments of the kind just reviewed. But public summary file tabulations have major limitations. For example,

---

[12] As discussed in an earlier note, this is based on performing pooled regression analyses to test additive and non-additive effects of race with non-racial social characteristics controlled.

tabulations rarely include more than a few non-racial social characteristics at one time, tabulations often provide only limited detail on non-racial social characteristics, and researchers have no control over the sample universe for the tabulations.[13]

The new methods outlined here can help researchers get more out of these traditional sources of data for segregation analysis. But the potential benefits of the new methods can be realized more fully and to greater effect if one draws on a new source of data for performing segregation analysis. The new source is restricted census datasets that contain individual-level data with detailed information about both individual social characteristics and also geographic information needed to pursue analyses of the residential attainment processes that produce segregation.[14] Working with restricted access census files is difficult, time consuming, and expensive. But it also affords great opportunities. For example, it is conceivable that one could use the most recent files from the American Community Survey (ACS) or the American Housing Survey (AHS) to investigate segregation without having to rely on summary file tabulations. This is possible because the difference of group means formulation of segregation indices allows segregation scores to be estimated by the effect of race in city-specific individual-level models of residential attainment. So, if one has access to detailed micro data, one has tremendous flexibility to investigate segregation in a wide range of new ways.

Additionally, because this approach allows for more efficient multivariate analysis, it expands the possibilities for investigating segregation reliably with smaller samples.[15] This is not only relevant for using smaller samples such as are found in the ACS and AHS. It also raises the possibility of investigating segregation using *non-census surveys*.[16] This is intriguing because non-census surveys can permit investigators to expand residential attainment analyses to consider variables such as individual racial attitudes, residential preferences, and other relevant measures that are not available in census datasets whether micro-data files or summary tabulations.

---

[13] For example, the attainment analyses I reported in the previous section found weak income effects based on a crude poverty/non-poverty distinction. Stronger income effects can be discerned using detailed income tabulations, but these tabulations do not include the other social characteristics in the analysis.

[14] A study by Bayer et al. (2004) takes a step in this direction by using restricted access census data to conduct refined individual-level analyses of residential contact. The framework I set forth here makes it possible to implement this kind of study to investigate uneven distribution.

[15] Significance tests and confidence intervals for the effect of race on residential attainments provides clear information about the reliability of segregation estimates obtained from analyses using smaller samples.

[16] Non-census surveys such as the Multi-City Study of Urban Inequality (MCSUI) can be used to study refined models of segregation so long as the households in the study are coded for area of residence at census geographies relevant for studying segregation (e.g., tract, block group, or block). Residential outcomes scored from census data can then be merged with the survey data to permit refined micro-level analyses of aggregated segregation.

9.11   An Example Analysis Using Restricted Microdata

A series of recently completed studies by Amber Fox Crowell provides insight into what the future of research on residential segregation is going to look like.[17] The primary focus of her research is on the factors determining White-Latino segregation. Her dissertation research (Fox 2014) presents detailed analyses investigating White-Latino in six major metropolitan areas. The analyses draw on restricted micro-data files of the 2000 decennial census and the restricted micro-data files of the 2008–2012 American Community Survey. Crowell applies the methods discussed in this work to the full potential that can be achieved with extant data. She measures residential outcomes at the level of census blocks and performs sophisticated quantitative analyses using the method of fractional regression to assess the impact of social and economic characteristics on White and Latino residential attainments. She then performs standardization and components analysis to assess the role of group differences in social and economic characteristics in explaining White-Latino residential segregation. Her studies present detailed results for analyses pertaining to segregation measured both using the separation index (S) and the dissimilarity index (D). I limit the presentation here to selected results from her analyses focusing on group separation (S) but note that the results for the dissimilarity index are similar in overall pattern.

The most striking contribution of her research is her ability to investigate how a comprehensive set of social and economic characteristics shape residential outcomes for Latino households. The list of micro-level predictors and the estimated coefficients indicating their impact on the residential attainments of Whites and Latinos in Houston, Texas in 2000 and in 2010 is presented in Table 9.9. Results for other cities are not presented to conserve space, but the results for Houston give the full flavor of the analyses Crowell is able to conduct. Her attainment equations include a wide range of relevant predictors including age, level of education, household income, military service, nativity and citizenship, year of immigration, English ability, marital/family status, and recent immigration experiences. No previous study has ever been able to take all of these factors into account simultaneously to quantitatively assess their impact on overall (city-level) residential segregation.

The results reported in Table 9.9 show that all of the micro-level variables have statistically significant effects in both the equation for Whites and the equation for Latinos. The "centered" constant reported in the table is the expected value of contact with Whites when independent variables are set at reference categories (for categorical variables) or values (for interval variables). The coefficients reported are fractional effects. These are additive effects on the logit value of the mean for contact with Whites. Positive effects are seen for education, income, and English language ability, produce greater average contact with Whites. Negative effects are

---

Table 9.9 Coefficients from fractional regressions predicting residential outcomes (y) determining the separation index (S) for White-Latino segregation in Houston, Texas in 2000 and 2010

Variable	Whites		Latinos	
	2000	2010	2000	2010
Degree (0–5)	0.1907[a]	0.1395[a]	0.2794[a]	0.2293[a]
Income (Ln)	0.0998[a]	0.0666[a]	0.0771[a]	0.0535[a]
Military	−0.0990[a]	−0.1103[a]	0.1088[a]	0.0828
U.S.-born citizen (ref)	–	–	–	–
Non-U.S. citizen	−0.0981[b]	−0.0873	−0.2506[a]	−0.2318[a]
Nat. U.S. citizen	−0.0991[a]	−0.1312[a]	−0.0315	−0.0280
Recent immigrant	−0.1836[a]	−0.0877	−0.1874[a]	−0.0042
English ability	0.2923[a]	0.3097[a]	0.1679[a]	0.2526[a]
Age 30–59 (ref)	–	–	–	–
Age 15–29	−0.1713[a]	−0.1673[a]	−0.1902[a]	−0.1950[a]
Age 60+	0.1579[a]	0.1386[a]	−0.0025	0.0843[a]
Married couple (ref)	–	–	–	–
Single mother	−0.2871[a]	−0.3010[a]	−0.1655[a]	−0.2784[a]
Other family	−0.3715[a]	−0.3325[a]	−0.0489[a]	−0.1283[a]
Recent mover	0.0940[a]	−0.0814[a]	0.2334[a]	0.0530[b]
Constant	−0.6652[a]	−0.6285[a]	−1.7437[a]	−1.8607[a]
Constant (centered)	1.5908[a]	1.2448[a]	0.0903[a]	−0.1099[a]

Notes: [a]denotes $p < 0.01$ and [b]denotes $p < 0.05$
Source: Restricted microdata files from the 2000 decennial census and the 2008–2012 American Community Survey

seen for foreign born status, non-citizen status, and recent immigration which all produce lower average contact with Whites. All of the effects are consistent with expectations from spatial assimilation theory. Group differences in the efficacy of the social and economic characteristics reflect the impact of minority status on contact with Whites. Altogether the results provide a wealth of information about the role of social and economic characteristics in shaping White and Latino residential outcomes and ultimately White-Latino segregation.

The implications of the results for White-Latino segregation in Houston are summarized in Table 9.10 which also presents results for the other cities included in Crowell's analyses. The results document that White-Latino differences in mean contact with Whites – the residential outcome that determines the value of the separation index (S) – vary across substantively relevant standardization scenarios. The scenario labeled "Latino group means & Latino rates of return" yields the predicted level of contact with Whites for Latinos in the Houston given their observed distribution on the social and economic characteristics in the attainment equations. Similarly, the scenario labeled "White group means & White rates of return" yields the predicted level of contact with Whites for Whites in the Houston given their observed distribution on the social and economic characteristics in the attainment equations. The difference between these two means yields the observed value of the

Table 9.10 Standardization analyses for White–Latino differences in residential outcomes (y) determining the separation index (S)

Standardization comparison	Atlanta Mean	Atlanta S*	Chicago Mean	Chicago S*	Houston Mean	Houston S*	Los Angeles Mean	Los Angeles S*	San Diego Mean	San Diego S*	Seattle Mean	Seattle S*
2000												
Latino group means & Latino rates of return	72.2	23.9	52.2	40.4	43.2	42.1	30.5	51.7	52.3	34.4	88.0	8.4
White group means & Latino rates of return	86.9	9.3	66.9	25.7	62.9	22.4	54.1	28.1	69.6	17.1	91.7	4.7
Latino group means & White rates of return	91.4	4.7	87.0	5.6	74.2	11.2	71.7	10.5	81.3	5.3	95.0	1.4
White group means & White r ates of return	96.1	–	92.6	–	85.3	–	82.2	–	86.6	–	96.4	–
2010												
Latino group means & Latino rates of return	63.1	29.6	49.5	39.8	36.8	42.3	27.4	50.4	47.8	34.2	82.5	11.3
White group means & Latino rates of return	80.9	11.8	63.1	26.3	54.5	24.5	46.4	31.4	61.9	20.1	88.7	5.1
Latino group means & White rates of return	85.9	6.8	84.2	5.2	67.4	11.7	72.5	5.3	76.7	5.3	90.8	2.9
White group means & White rates of return	92.7	–	89.4	–	79.1	–	77.8	–	82.0	–	93.7	–

Notes: "Mean" denotes predicted mean contact with Whites based on the standardization scenario. S* denotes the value of the separation index (S) (i.e., the White–Latino mean difference in contact with Whites) under the standardization scenario. S* under the "Latino group means & Latino rates of return is the observed value of the separation index. Source: Restricted micro-data files from the 2000 decennial census and the 2008–2012 American Community Survey

separation index (S) for White-Latino segregation in Houston. That is, the value of
42.1 in 2000 reflects the difference between the mean of 85.3 for Whites and the
mean of 43.2 for Latinos and is reported in the column labeled "S*" under Houston
on the first row of the panel reporting results for 2000.

Scanning the values reported on this row of the table reveals that White-Latino
separation varies greatly across the six cities in Crowell's analysis. The separation
index (S) is highest in Los Angeles (51.7) and only slightly lower in Houston (42.1)
and Chicago (40.4). It is somewhat lower in Atlanta (23.9) and very low in Seattle
(8.4). Drawing on methods reviewed earlier in this chapter, Crowell performed stan-
dardization analyses to explore address the question of whether White-Latino seg-
regation is due to group differences in resources for residential attainment or the
impact of group status itself in the residential attainment process. In the interests of
space group distributions on predictors are not shown but they are reported in
Crowell's studies. Not surprisingly, Latinos tend to have deficits on predictors that
have positive effects on contact with Whites (e.g., income) and surpluses on predic-
tors that reduce contact with Whites (e.g., non-U.S. citizen).

The role of group differences in resources is documented in the row labeled
"White group means & Latino rates of return." The values reported here indicate
how Latino residential outcomes would change if Latinos had the White "profile"
on social and economic characteristics. The implications for S* show that the role
of group differences assessed in this manner is always positive and substantively
important. Equalizing Latino inputs to residential attainment process reduces the
value of S by between 34 and 61 %.

The role of minority status is documented in the row labeled "Latino group
means & White rates of return" which indicates how Latino residential outcomes
would change if Latinos experienced White rates of converting inputs to the attain-
ment process into contact with Whites. The implications for S* show that the role
of this factor also is always positive and substantively important. Indeed, equalizing
Latino rates of return in the attainment process would reduce the value of S by
between 74 and 89 %.

I close this chapter by noting that the results presented in Tables 9.9 and 9.10
provide a wealth of information warranting additional discussion. Unfortunately, a
more detailed review is beyond the scope of the present discussion so I encourage
interested readers to seek out Crowell's research for more in-depth discussion of her
findings. The central point I stress here is this. Crowell's research shows that com-
bining the new methods outlined in this monograph with the restricted census
micro-data files opens the door to exciting new options for segregation analysis.
Crowell's research provides the best example to date of how segregation can be
analyzed in great detail in a single-city analysis. In the next chapter I outline how
this approach can be expanded to cover a larger sample of cities and explore the
impact of city-level characteristics on residential segregation via estimation of
multi-level models of residential attainments.

References

Alba, R. D., & Logan, J. R. (1993, May). Minority proximity to Whites in suburbs: An individual-level analysis of segregation. *American Journal of Sociology*, *98*, 1388–1427.

Alba, R. D., Logan, J. R., Stults, B., Marzan, G., & Zhang, W. (1999, June). Immigrant groups and suburbs: A reexamination of suburbanization and spatial assimilation. *American Sociological Review*, *64*, 446–460.

Althauser, R. P., & Wigler, M. (1972). Standardization and component analysis. *Sociological Methods and Research, 1*, 97–135.

Bahr, H. M., & Gibbs, J. P. (1966). Racial differentiation in American metropolitan areas. *Social Forces, 45*, 521–532.

Bayer, P., McMillan, R., & Reuben, K. S. (2004). What drives racial segregation: Evidence from census microdata. *Journal of Urban Economics, 56*, 514–535.

Becker, G. S. (1971). *The economics of discrimination*. Chicago: University of Chicago Press (First edition 1957).

Buis, M. L. (2006, September 12). *Proportions as dependent variables*. Paper presented at the 12th UK Stata Users Group Meeting, London, England.

Buis, M. L., Cox, N., & Jenkins, S. (2006). DIRIFIT: A Stata Module to fit a Dirichlet distribution. Posted at *EconPapers*. http://EconPapers.repec.org/software/bocbocode/s456725.htm. Accessed Aug 2008.

Crowder, K., & South, S. J. (2005). Race, class, and changing migration patterns between poor and nonpoor neighborhoods. *American Journal of Sociology, 110*(6), 1715–1763.

Crowder, K., South, S. J., & Chavez, E. (2006). Wealth, race, and inter-neighborhood migration. *American Sociological Review, 71*(1), 72–94.

Crowell, A. F., & Fossett, M. (2015, August). *White-Latino residential attainments in six cities: Assessing the role of micro-level factors*. Paper presented at the Annual Meetings of the American Sociological Association, Chicago, IL.

Duncan, O. D., & Duncan, B. (1955). A methodological analysis of segregation indices. *American Sociological Review, 20*, 210–217.

Elgie, R. A. (1980). Industrialization and racial inequality with the American south, 1950–1970. *Social Science Quarterly, 61*, 458–472.

Fossett, M. (1988). Community-level analyses of racial socioeconomic inequality: A cautionary note. *Sociological Methods and Research, 16*, 454–491.

Fossett, M. (2008, March 12–14). *Toward a unified framework for studying segregation: Studying individual-level residential attainments and city-level segregation within a single model*. Paper presented at the Session on Racial Residential Segregation at the Annual Meetings of the Southwestern Sociological Association, Las Vegas, NV.

Fox, A. R. (2014). *Latino residential segregation in the United States: New methods, new findings*. Doctoral dissertation, Texas A&M University, College Station, TX.

Fox, A. R., & Fossett, M. (2014a, May). *Residential attainments and segregation in Los Angeles: Applying new methods to connect the micro and the macro*. Poster presented at the Annual Meetings of the Population Association of America, Boston, MA.

Fox, A. R., & Fossett, M. (2014b, October). *Investigating White-Latino segregation patterns with micro-level controls*. Paper presented at the Annual Meetings of the Southern Demographic Association, Memphis, TN.

Iams, H. M., & Thornton, A. (1975). Decomposition of differences: A cautionary note. *Sociological Methods and Research, 3*, 341–352.

Jiobu, R., & Marshall, H. H., Jr. (1971). Urban structure and the differentiation between Blacks and Whites. *American Sociological Review, 36*, 638–649.

Jones, F. L., & Kelley, J. (1984). Decomposing differences between groups: A cautionary note. *Sociological Methods and Research, 12*, 323–343.

Kitagawa, E. M. (1955). Components of a difference between two rates. *Journal of the American Statistical Association, 50*, 1168–1194.

LaGory, M., & Magnani, R. J. (1979). Structural correlates of Black-White occupational differentiation: Will U.S. regional differences remain. *Social Problems, 27*, 157–169.

Logan, J. R., Alba, R. D., McNulty, T., & Fisher, B. (1996). Making a place in the metropolis: Locational attainment in cities and suburbs. *Demography, 33*, 443–453.

Massey, D. S., & Denton, N. A. (1985). Spatial assimilation as a socioeconomic process. *American Sociological Review, 50*, 94–106.

Massey, D. S., & Mullan, B. P. (1984). Processes of Hispanic and Black spatial assimilation. *American Journal of Sociology, 89*, 836–873.

Papke, L. E., & Wooldridge, J. (1996). Econometric methods for fractional response variables with an application to 401(k) plan participation rates. *Journal of Applied Econometrics, 11*, 619–632.

Roof, W. C. (1972). Residential segregation of Blacks and racial inequality in southern cities: Toward a causal model. *Social Problems, 19*, 393–407.

Smithson, M., & Verkuilen, J. (2006). A better lemon squeezer? Maximum likelihood regression with beta-distributed dependent variables. *Psychological Methods, 11*, 54–71.

South, S. J., & Crowder, K. D. (1997, November). Residential mobility between cities and suburbs: Race, suburbanization, and back-to-the-city moves. *Demography*, *34*, 525–538.

South, S. J., & Crowder, K. D. (1998, February). Leaving the 'hood: Residential mobility between Black, White, and integrated neighborhoods. *American Sociological Review*, *63*, 17–26.

South, S. J., Crowder, K., & Chavez, E. (2005a). Exiting and entering poor neighborhoods: Latinos, Blacks, and Anglos compared. *Social Forces, 84*(2), 873–900.

South, S. J., Crowder, K., & Chavez, E. (2005b). Migration and spatial assimilation among U.S. Latinos: Classical versus segmented trajectories. *Demography, 42*(3), 497–521.

South, S. J., Crowder, K., & Pais, J. (2008, September). Inter-neighborhood migration and spatial assimilation in a multi-ethnic world: Comparing Latinos, Blacks and Anglos. *Social Forces*, *87*, 415–443.

Winsborough, H. H., & Dickinson, P. (1971). Components of Negro-White income differences. *Proceedings of the American Statistical Association, Social Statistics Section* 6–8.

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

Wooldridge, J. M. (2002). *Econometric analysis of cross section and panel data*. Cambridge: MIT Press.