

# **Analysis of Policy Issues Concerning Social Integration**

**BRIAN STIPAK**

*Institute of Public Administration, Pennsylvania State University, University Park, PA., U.S.A.*

---

## **ABSTRACT**

Empirical policy analysis can often contribute information relevant to public policy-making that affects different kinds of social integration. Policy analyses concerning social integration in a variety of settings face common analytical issues and potential statistical problems. As an example of this type of empirical analysis, this paper uses survey and census data from the Detroit metropolitan area to analyze the effect of neighborhood racial composition on residents' satisfaction with their neighborhoods.

---

## **Introduction**

Either by design or by accident, many types of public policies affect the degree of integration of different socio-demographic groups in different settings. Probably the most salient examples in the United States today concern racial integration of schools and housing. Some policy proposals seek to directly use governmental authority to manipulate the social context. For example, proposals for inter-school busing of school children aim at changing the racial composition of schools. These complex policy problems involve a plethora of considerations not amenable to empirical analysis. Nonetheless, empirical policy analysis can often contribute relevant information to the policy debate. Although obfuscated by the emotion and controversy that frequently surround such policies, policies affecting social context pose common analytical issues for empirical analysis. This paper first discusses these underlying analytical issues, then turns to the potential statistical problems analysts may face, and finally presents an example of this type of empirical analysis.

## **Policy Analysis of Effects of Social Context on Individual-Level Outcome Variables**

One general goal of public policy can be thought of as influencing characteristics of individual citizens that affect their quality of life. Such characteristics can span a diversity of objective and subjective variables, such as educational achievement,

**TABLE 1**  
**Examples of Individual-Level Outcome Variables and Individual-Level and Contextual Independent Variables**

Individual-level outcome variable	Individual-level independent variable(s)	Contextual independent variable(s)
(1) Educational achievement	Socio-demographic characteristics	Socio-demographic composition of school, classroom
(2) Propensity for a student to be truant from school	Past truancy record	Truancy record of classmates
(3) Probability of completing high school	Socio-demographic characteristics	Socio-demographic composition of school
(4) Probability of recidivism of a felony offender after parole	Criminal background	Criminal background of fellow prisoners
(5) Satisfaction with neighborhood	Socio-demographic characteristics	Socio-demographic composition of neighborhood

personal health, satisfaction with different aspects of one's life, and so on. Improvements on individual-level variables therefore constitute one set of the ultimate outputs of governmental intervention in the social system. To understand how public policy can effect improvements in these outcome variables, the policy analyst must ask what are the determinants of these variables, and which of the determinants governmental policy can influence.

The determinants that governmental policy can influence are often contextual variables, as opposed to other individual-level variables. Whereas an individual-level variable is a characteristic of an individual person, a contextual variable is a characteristic of an individual's environment. For example, a person's race is an individual-level variable, but the racial composition of his neighborhood, or of the school he attends, is a contextual variable. Governmental policy cannot easily affect most individual-level socio-demographic variables, such as a person's race, but can affect their contextual analogues, such as the social context in which individuals live or attend school.

Table 1 provides, for several individual-level outcome variables, examples of possible individual-level independent variables and their policy-relevant contextual analogues. In each example, any effects of the individual-level independent variable(s) are given from the point of view of policy makers [1], but effects of the contextual independent variables are potentially manipulable. To the school administrator, effects of students' socio-demographic backgrounds on educational achievement or likelihood of completing high school are unmanipulable constraints [2]; in contrast, contextual effects offer potential avenues for influencing the outcome variable, since the administrator can change the socio-demographic composition of schools or classrooms. Similarly, the prison administrator must view as given whatever effect

the criminal background of individual prisoners has on the probability of recidivism, but effects due to the type of fellow prisoners a prisoner interacts with are manipulable by prison policy. In short, public policy is often better able to influence contextual, rather than individual-level, determinants of individual-level outcome variables.

The policy importance of social context therefore stems from the greater efficacy of public policy to manipulate contextual variables, compared to individual-level variables, in order to improve an individual-level outcome variable. However, if social context does not affect the outcome variable in question, the ability to influence contextual variables offers little solace to the policy-maker. Therefore, an important task for policy analysis is to enlighten the policy debate about what effects contextual variables do exert, and which variables offer potentially efficacious means of affecting outcomes.

Policy research must investigate not only which contextual variables affect the outcome variables and the strength of those effects, but also the functional form of those effects. The functional form of the relationship between an outcome variable and a contextual determinant limits the degree that policy can have an overall beneficial effect by manipulating the contextual determinant. For example, assume that students' achievement levels decrease monotonically with the percent non-white enrollment in the students' schools, *ceteris paribus*. That is, for purposes of illustration, assume that regardless of a student's own race, the student will perform more poorly if he attends a school with a higher percentage of minority enrollment. Implementing a public policy for achieving greater racial integration would therefore tend to decrease the educational achievement of students originally attending low-minority schools, and increase the educational achievement of students originally attending high-minority schools. However, the degree that greater integration would decrease achievement of some students, compared to increasing achievement for

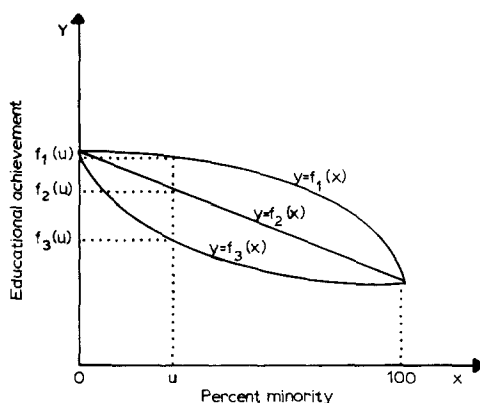


Fig. 1. Alternative assumptions about the functional form of the regression of educational achievement on percent minority in the school, other factors constant.

other students, depends on the functional form of the relationship between achievement and minority enrollment.

Figure 1 uses the above example to illustrate the policy importance of the functional form of a contextual variable's independent effect on an individual-level outcome variable. Three different hypothetical forms are shown for the relationship between the contextual variable (percent minority) and the outcome variable (educational achievement):  $f_1$  has a negative second derivative;  $f_2$  is linear; and  $f_3$  has a positive second derivative. Each of these functions conforms to the assumption of a decreasing monotonic relationship, but their policy implications differ. Assume for simplicity that before a new integration policy is implemented complete segregation exists, and that after implementation complete integration exists in the sense that all schools have a percent minority enrollment equal to the overall percent for all schools,  $\mu$ . Before integration, expected educational achievement is either  $f(0)$  or  $f(100)$ , and after integration, expected achievement is  $f_1(\mu)$ ,  $f_2(\mu)$ , or  $f_3(\mu)$ , depending on whether  $f_1$ ,  $f_2$ , or  $f_3$  is the true function. If  $f_1$  is the true function, integration lowers achievement of students previously in all-white schools much less than if  $f_3$  is true. Conversely, if  $f_3$  is the true function, integration raises achievement of students previously in all-minority schools much less than if  $f_1$  is true. The effects of  $f_2$  are between  $f_1$  and  $f_3$ . The functional form of the relationship between an individual-level outcome variable and a contextual determinant clearly affects the relative beneficial and harmful consequences of particular policies.

One convenient, although arbitrary, criterion for comparing the relative beneficial and harmful consequences is the overall mean of the individual-level outcome variable. In the above example, expected achievement before integration for all schools is simply a weighted average of  $f(0)$  and  $f(100)$ :

$$E(Y) = \frac{f(100)\mu}{100} + \frac{f(0)(100-\mu)}{100}$$

This is exactly  $f_2(\mu)$ ; thus, if the true independent impact of the contextual variable is linear, manipulation of the social context through racial integration cannot change mean achievement for all students, although it will change the expected performance of students from different schools. On the other hand, under  $f_1$  integration improves mean achievement, since  $f_1(\mu)$  is greater than  $f_2(\mu)$ , and under  $f_3$  integration decreases mean achievement. Obviously, estimated effects on overall mean achievement should not be the final arbiter for policies concerning complex social issues like school integration, but any information the policy analyst can supply about the relative harmful versus beneficial consequences of policy proposals increases the informational basis for a policy decision.

Policy research must also consider whether the strength and functional form of a contextual variable's effect are the same for individuals from different socio-demographic groups. Interactions between contextual variables and the individual's socio-demographic characteristics can have great policy importance. For example,

the U.S. Commission on Civil Rights (1967) concluded that school integration increases black students' achievement, but does not decrease achievement for white students attending classes having black compositions as high as about 50%. If that is true, school desegregation can potentially increase achievement for blacks without decreasing achievement for whites, yielding a Pareto improvement. In short, to provide useful information about the possible consequences of social integration for an outcome variable, policy research must investigate (1) which contextual variables affect the outcome variable, (2) the strength of those effects, (3) the functional form of those effects, and (4) whether those effects interact with the individual's socio-demographic characteristics.

### **Statistical Approaches to Estimating the Effects of Social Context**

The ideal data for estimating the effects of social context on an individual-level outcome variable would be experimental data derived from a classic experimental design. The policy analyst would randomly assign subjects to different "treatments" consisting of different types of social contexts. For example, a policy analyst interested in understanding how school racial composition affects achievement of black school children would randomly assign black students to schools with different racial compositions [3]. In reality, policy analysts can seldom implement experimental designs, especially concerning sensitive issues like social integration. Quasi-experimental longitudinal data may occasionally be available – for example, data on student achievement scores before and after integration. However, policy analysts will often have only non-experimental, cross-sectional data available for analyzing the effects of social context. Contextual analyses using these data face a number of common considerations and potential problems.

Estimating the separate effects of individual-level and contextual variables using non-experimental, cross-sectional data requires that analysts include the individual-level and contextual variables explicitly in their statistical analyses. The analyses must be done using the individual as the case for analysis, appending the necessary contextual data to the individual-level data. For example, a student-level dataset might include data for the student's test score and demographic characteristics, as well as appended data on the demographic composition of the student's school or class, and perhaps variables concerning the educational approach used in the student's school or class. With such a dataset the analyst can attempt to estimate the independent effects of individual-level and contextual variables on the outcome variable. In contrast, an analysis of educational achievement that uses school-level or classroom-level data confounds effects of (1) the individual student's own social characteristics, and (2) the social characteristics of the school or class the student attends.

Estimating models that contain contextual variables can potentially encounter a number of statistical problems. These problems are not unique to contextual

analysis, but rather are typical problems in multivariate analysis, especially in non-experimental social research. These statistical problems are treated at length in the econometrics literature and are well understood. Those critics (e.g. Hauser, 1970, 1974) who have made general critiques of contextual analysis have over-emphasized the uniqueness and severity of the statistical obstacles. The potential contextual researcher should consult Boyd and Iversen (1979) or Hensler and Stipak (1979) for a more detailed discussion than presented below of the problems of statistical inference in contextual analysis, and for suggestions about how to avoid or minimize those problems.

Probably the most serious statistical problems that can plague contextual analyses are specification errors and measurement errors. The objective of contextual analysis is to estimate the independent effects attributable to contextual variables or settings. The analytical task involves estimating the parameters of models of the form

$$Y = f(\mathbf{I}, \mathbf{C})$$

where:  $Y$  = a dependent variable describing the individual;

$\mathbf{I}$  = vector of independent variables describing the individual;

$\mathbf{C}$  = vector of independent variables describing or identifying the context in which the individual is a member.

The susceptibility of contextual analysis to statistical bias from specification or measurement errors stems from the strong correlations that typically exist between individual-level ( $\mathbf{I}$ ) and contextual ( $\mathbf{C}$ ) predictors. That is, in most applications in non-experimental social research the individuals are not randomly distributed across the contexts, but rather different contexts have members who differ on many individual characteristics. Indeed, the importance of social context results from this non-randomness, since if individuals of different socio-demographic characteristics were randomly distributed across contexts, no differences in social context would exist.

Because of the correlation that typically exists between individual-level and contextual variables, misspecifying the model by omitting relevant individual-level variables having an effect on the dependent variable will in general bias parameter estimates for the correlated contextual variables. Measurement error in the individual-level variables can similarly lead to biased parameter estimates for the contextual variables. For example, if an analyst wants to estimate the independent effect on a dependent variable attributable to the economic level of a person's neighborhood, the analyst must include independent variables for both the economic level of the neighborhood and the person's own economic level. Omission of the individual-level economic variable, or measurement error in that variable, can bias the coefficient estimate for the neighborhood variable [4]. For these reasons, contextual analysts should exercise care in specifying the individual-level component of a contextual model, and in minimizing measurement error.

Another potential statistical problem is simultaneity between the dependent variable and a contextual predictor. Hauser (1974, pp. 373–374) refers to this as selection on the dependent variable, and offers the example of low-income parents of gifted children choosing their residence to allow their children to attend the presumed better schools serving higher-income families. Armor (1972, p. 112) similarly warns about the difficulty of causal inference when self-selection is present – such as families of higher-achieving students choosing to live in integrated neighborhoods – and argues that self-selection may account for the Coleman study's finding (Coleman et al., 1966, pp. 29, 331) that integrated black students had higher achievement than segregated students. This type of simultaneity can bias contextual parameter estimates. To illustrate using Hauser's example, estimating a single-equation model that regresses educational achievement on family income, a school-level income measure, and other predictors, could spuriously attribute effects to the contextual income measure that in reality result from the selection process.

To avoid misleading conclusions, contextual analysts must question how likely serious biases might result from simultaneity, specification errors, or measurement errors. To what extent might the outcome variable influence selection to contexts? How strong is the justification for the specification of the model? How reliable are the measures? Realistically, these statistical problems will always be present to some degree in non-experimental social research; therefore, if the analyst finds strong effects attributable to individual-level variables and only weak effects attributable to correlated contextual variables, he should view the estimated contextual effects with suspicion.

A practical question the contextual analyst must consider is how to represent the contextual variables in the models to be estimated. One approach is to append data for the values of the contextual variables to the individual-level data, and to include these quantitative measures explicitly as contextual variables in the model. A second approach is to estimate the net contextual effect for each context by including dummy variables representing each context [5]. The first approach offers the advantage of yielding estimates of the independent effects of each of the contextual variables, whereas the second approach confounds the contextual effects in single estimates of the net effect for each context. If there is more than one contextual variable on which the contexts differ, then the dummy variable model is not identified, since there is more than one potential explanation for the estimates of net inter-context differences.

Although it confounds effects of different contextual variables, the dummy variable method does offer the advantage of feasibility when using quantitative contextual variables is not possible because (1) data on the contextual variables are unavailable, (2) the appropriate contextual variables are unknown, or (3) there are too few contexts [6]. Implementing the dummy variable approach requires only the relevant individual-level data, and knowledge of the context in which the individual is a member. Therefore, the dummy variable approach can be a useful exploratory

tool when limitations of data or theory prevent directly estimating independent effects for specific contextual variables.

### **An Empirical Example: Neighborhood Satisfaction**

As an example of policy-relevant contextual analysis, Detroit data will be used to investigate the effect of neighborhood racial composition on the satisfaction of residents with their neighborhoods. Different population sub-groups having different preferences for the racial composition of their neighborhood will be identified and analyzed separately. Since the data are cross-sectional and non-experimental, the analysis cannot explore possible dynamic effects over time, such as changes in neighborhood racial preferences that might result from changes in neighborhood racial composition. Rather, the analysis statically treats existing preferences as given, and attempts to estimate the effect neighborhood racial composition has on neighborhood satisfaction, independent of the economic level of the neighborhood or the socio-demographic characteristics of the individual.

#### **Data**

The analysis uses a dataset of merged survey and census data from the Detroit metropolitan area. The University of Michigan's Institute for Social Research collected the survey data from a probability sample of 1194 respondents who were interviewed during October 1974 through February 1975. Census data from the 1970 Census of Population and Housing were merged with these survey data. Thus, for each case (respondent) the dataset contains both survey data about the respondent, and census data about the respondent's census tract.

#### **Dependent Variable**

Five survey items were used to construct a neighborhood satisfaction scale. Four of the items required the respondents to rate their neighborhoods on seven-point scales according to different characteristics: (1) friendly people–unfriendly people, (2) very good place to live–very poor place to live, (3) pleasant–unpleasant, and (4) good neighbors–bad neighbors [7]. The fifth item is a seven-point rating scale of the respondents' general satisfaction with their neighborhoods as places to live: "All things considered, how satisfied are you with this neighborhood as a place to live?" These five items were chosen on the basis of face validity as measures of respondents' general satisfaction with their neighborhoods and neighbors. Satisfaction items with more specific referents – e.g. neighborhood noise, trees, traffic, safety, vandalism – were not used, to avoid contaminating general feelings of neighborhood satisfaction with more specific satisfactions.

The product-moment correlations among the satisfaction items are all moderately



high and positive, ranging from 0.46 to 0.73. No variable has markedly higher inter-correlations than the others. A principal components factor analysis yields a first factor which explains 67% of the total variance, whereas the second factor explains less than 14% of the total variance. All five variables have loadings on the first factor of about the same magnitude (0.77–0.87). In short, the correlation and factor analysis results support the interpretation on the basis of face validity that these items tap the same general dimension of neighborhood satisfaction.

Because all five items load about the same on the underlying dimension, there is little need to differentially weight the items. A simple summated scale was therefore constructed. The estimated scale reliability (Cronbach's alpha) is 0.87, indicating that only about 13% of the scale variance results from random error. The summated scale was divided by five to return it to the original seven-point rating scale units of its components items.

### **Independent Variables**

The model estimates the effect of racial composition on neighborhood satisfaction, taking into account the economic level of the neighborhood and the socio-demographic characteristics of the individual. Since blacks and whites compose all but a small fraction of the area's population, the percent of the census tract population that is black serves as the racial composition variable [8]. The median family income of the census tract was used to control for the independent effects of the economic level of the neighborhood. The individual-level socio-demographic variables are age, sex, and family income [9].

Several variables were hypothesized to have possible interactive effects with the other variables in the model. First, the individuals' race may affect the impact of neighborhood racial composition, and possibly other independent variables, on neighborhood satisfaction. Also, the effect of neighborhood composition may depend on the individuals' preferences for the racial composition of their neighborhoods. Fortunately, preference data were obtained by the following interview questions:

(whites) Would you personally prefer to live in a neighborhood with all whites, mostly whites, mostly blacks, or a neighborhood that's mixed half and half?

(blacks) Would you personally prefer to live in a neighborhood with all blacks, mostly blacks, mostly whites, or a neighborhood that's mixed half and half?

Interaction of the individuals' race and racial preferences with the independent variables was allowed for by estimating the model separately for respondents in the different race and preference categories [10]. This approach obviates entering individual-level race or racial preference variables explicitly in the model.

Possible effects of other variables were tested by entering those variables in the

model, and estimating the model separately for black and white sub-samples. These variables included respondent education, median school years for people 25 and older in the census tract, housing unit vacancy rate in the census tract, percent single family houses in the census tract, percent occupied housing units sub-standard in the census tract, respondent homeowner versus renter, and respondent length of residence in present dwelling unit. Possible interactive effects of homeownership and length of residence were investigated by separately estimating equations for homeowners versus renters, and for old versus new residents. No evidence of strong effects was found for any of these variables. Coefficient estimates were small and not statistically significant. The coefficient estimate for the neighborhood racial composition variable, the variable of primary interest, changed little when these other variables were entered. Therefore, these variables were not included in the model, for the sake of parsimony and to avoid problems of high multicollinearity with some of the other predictors.

#### Specification of the Functional Form

The percent of the neighborhood that is black might be expected to non-monotonically affect neighborhood satisfaction of respondents preferring neighborhoods of mixed racial composition. To investigate possible non-monotonicity, the percent black variable was divided into a number of dummy variables corresponding to different ranges of percent black [11]. Using the dummy variables instead of the original variable, the model was estimated for the different race and racial preference categories [12]. The results showed little evidence of any non-monotonicity, even for respondents voicing a mixed racial preference. Therefore, the analysis assumes that any effects of the percent black variable are monotonic for all race and racial preference categories.

Although the dummy variable results showed monotonic effects of the percent black variable, the effects are not necessarily linear. That is, the marginal effect of a certain percentage change in the neighborhood black population may vary according to the initial percentage that is black. To explore such non-linearities, power function models were estimated:

$$S = \alpha + \gamma B^\lambda + \beta'X$$

where:  $S$  = neighborhood satisfaction scale;

$B$  = percent black in census tract;

$X$  = vector of other independent variables;

$\alpha$ ,  $\gamma$ ,  $\lambda$ ,  $\beta$  = parameters to be estimated.  $\beta$  is a coefficient vector corresponding to the  $X$  variables.

Allowing  $\lambda$  to vary allows the marginal effect of the percent black variable to vary. If  $\lambda = 1$ , then the marginal effect of a certain percentage change in the neighborhood black composition is constant, regardless of the initial percentage that is black. If  $\lambda < 1$ , then there is a diminishing marginal effect as black composition increases, whereas for  $\lambda > 1$ , a percentage change has an increasing effect. If a non-linear least squares program is conveniently available, the parameters of this type of model can be estimated directly; otherwise, an approximate least squares solution can be obtained using ordinary least squares. For this research, an approximate solution was obtained by estimating equations for power transformations of  $B$  with  $\lambda$  set at values between 0.1 and 2.0, in 0.1 increments, and choosing the solution that minimized the sum of squared deviations, i.e. maximized  $R^2$ .

The individual and contextual income variables were entered in the equations as logarithmic transformations. Base two logarithms were used, so that the coefficient estimates indicate the change in the neighborhood satisfaction scale resulting from a doubling of income. A logarithmic specification was chosen over a linear specification because of the expectation that income has a decreasing rather than a constant marginal effect – specifically, that approximately equal satisfaction changes will result from equal proportional changes in income. An empirical test of this expectation was made by estimating the model separately for logged and unlogged income variables, for each race and racial preference category. The logged versions resulted in higher predictive power.

Respondent sex was entered as a dummy variable, coded as zero for women and one for men. Age was broken into five categories, and entered as four dummy variables. The youngest age category, 18–29 years, is used as the reference. This dummy variable representation of age allows for possible non-linearity and non-monotonicity due to cohort or life-cycle effects.

## Results

Table 2 presents the coefficient estimates, with the estimated standard errors in parentheses [13], for equations for each of the race and racial preference sub-groups. Four groups are omitted, because of the small number of respondents in each: whites, mostly blacks preference (0 cases); blacks, all blacks preference (11 cases); blacks, mostly blacks preference (15 cases); and blacks, mostly whites preference (14 cases). Since these omitted cases constitute only 3% of the sample, ignoring them for the remainder of the analysis will not greatly affect the conclusions. Table 2 does include results for whites and blacks who answered that racial composition makes no difference to them. Thus, the total sample was broken down into four white sub-samples and two black sub-samples, each of which was analyzed separately.

For only three of the groups did the estimation results show strong effects of neighborhood racial composition: (1) whites, all whites preference, (2) whites, mostly whites preference, and (3) blacks, mixed racial preference. Table 2 presents the

Table 2  
 Estimation Results For White and Black Respondents Having Different Preferences For Neighborhood Racial Composition

Independent variables	Whites, all whites racial preference	Whites, mostly whites racial preference	Whites, mixed racial preference	Whites, no difference racial preference	Blacks, mixed racial preference	Blacks, no difference racial preference
	$\lambda = 1.0$	$\lambda = 0.4$				
(Percent Black) <sup><math>\lambda</math></sup>	-0.055 (0.009)	-0.198 (0.080)	-0.004 (0.006)*	-0.003 (0.006)*	$\lambda = 0.2$ -0.923 (0.269)	-0.003 (0.007)*
Median tract income, base 2 logarithm	0.743 (0.205)	0.380 (0.286)	0.424 (0.347)	0.934 (0.305)	0.702 (0.314)	1.063 (0.464)
Family income, base 2 logarithm	0.054 (0.066)	0.299 (0.089)	-0.085 (0.117)	0.184 (0.100)	0.025 (0.100)	-0.006 (0.155)
Sex	-0.086 (0.120)	-0.193 (0.143)	-0.073 (0.254)	0.300 (0.204)	0.246 (0.241)	-0.559 (0.295)
Age:						
60+	0.368 (0.193)	0.988 (0.254)	0.513 (0.398)	0.507 (0.318)	1.890 (0.374)	1.447 (0.547)
50-59	0.446 (0.182)	0.237 (0.233)	0.736 (0.393)	-0.206 (0.313)	1.901 (0.368)	1.307 (0.496)
40-49	0.618 (0.190)	0.060 (0.220)	0.436 (0.381)	-0.476 (0.328)	1.165 (0.348)	1.403 (0.509)
30-39	0.196 (0.195)	-0.014 (0.197)	0.457 (0.311)	-0.616 (0.311)	0.420 (0.358)	0.616 (0.476)
<i>R</i>	0.46	0.39	0.28	0.50	0.58	0.63
<i>N</i>	354	218	98	122	147	67
<i>N</i> /(total <i>N</i> )	0.352	0.217	0.097	0.121	0.146	0.067

\* Results arbitrarily presented for untransformed percent black variable ( $\lambda = 1.0$ ). No transformation yielded a statistically significant coefficient estimate.

Note: Table entries are unstandardized partial regression coefficients, and numbers in parentheses are estimated standard errors.

estimation results for these three groups for the power function that was the best least squares fit. The best fit was obtained for  $\lambda = 1.0, 0.4, \text{ and } 0.2$ , respectively. The coefficient estimates for the racial composition variable for these three groups are highly statistically significant, with respective  $t$  statistics of  $-6.21, -2.46, \text{ and } -3.43$ . In contrast, the results for the other three groups – (1) whites, mixed preference, (2) whites, no difference, and (3) blacks, no difference – show no strong effect of neighborhood racial composition. For these three sub-samples no power function transformation resulted in even a statistically significant (0.05 level) coefficient estimate for the racial composition variable. This is not due merely to a difference in sub-sample size, but also to the magnitude of the effects: if the coefficient estimates for the untransformed version of the percent black variable ( $\lambda = 1.0$ ) are compared, the estimates for the first three sub-samples are  $-0.055, -0.014, \text{ and } -0.012$ , respectively, whereas for the second three sub-samples the estimates are  $-0.004, -0.003, \text{ and } -0.003$ , respectively. Table 2 does present estimation results for the three groups for which there was no discernible effect on racial composition in order that effects of other variables can be compared across all six groups. The results presented are for the untransformed version of the percent black variable.

The signs of the coefficient estimates for the racial composition variable are negative in all cases, suggesting a tendency for a higher percentage of blacks in the neighborhood population to lower neighborhood satisfaction [14]. For the three groups showing no statistically significant effect on racial composition, the tendency is negligible, but for the other three groups it is substantial. For example, in the case of white respondents with all whites preferences, a change in the neighborhood racial composition from all whites to 50% whites decreases neighborhood satisfaction an estimated 2.75 points on the seven-point neighborhood satisfaction rating scales [15]. Since the best least squares fit was obtained for a  $\lambda$  of 1 for white respondents with all whites preferences, but for a  $\lambda$  of  $< 1$  for white respondents with mostly whites preferences and for black respondents with mixed racial preferences, the negative marginal effect of percent black in the neighborhood does not appear to decrease for whites with all whites preferences, but does for the other two groups.

The results for the contextual income variable indicate that greater neighborhood affluence increases neighborhood satisfaction. Doubling tract income levels raises expressed satisfaction about 0.4–1.0 unit on the seven-point rating scales. This is consistent with Little's (1976) finding that neighborhood income level has an important independent effect on preferences for housing. On the other hand, the effect of individual-level income appears much weaker. Similarly, respondent sex shows no large or statistically significant effects. The coefficient estimates for the respondent age dummies indicate a tendency for satisfaction to increase with age, consistent with past research [16]. Age appears to interact with race, as the effect of age is greater for blacks than whites.

### Analysis

As noted above, the only sub-samples showing evidence of strong effects of neighborhood racial composition are whites, all whites and mostly whites preferences, and blacks, mixed racial preference. Thus, of the four white sub-samples examined, the two with the strongest preferences for white neighborhoods manifest the strongest negative effect on neighborhood satisfaction of percent black residents in the neighborhood. Also, as noted above, the effect for whites voicing an all whites preference appears stronger than for whites stating a mostly whites preference. The marginal effect of percent black apparently diminishes for the mostly whites preference group ( $\lambda = 0.4$ ), but not for the all whites preference group ( $\lambda = 1.0$ ). The results for whites therefore consistently reveal greater negative effects of percent black for whites having stronger preferences for white neighbors.

For blacks voicing a no difference racial preference, the results show no strong effect of percent black, just as for whites voicing a no difference racial preference. Blacks having a mixed racial preference do show a strong negative effect of percent black, in contrast to whites having a mixed racial preference. The results therefore suggest that the mixed racial preference category for whites indicates close to indifference, whereas for blacks it indicates a preference for neighborhoods of greater white concentrations. In light of this interpretation, it is interesting that all but 9% of black respondents eschewed the all blacks or mostly blacks racial preference categories. Other Detroit survey data (Farley et al., 1979, p. 104), as well as national survey results (Pettigrew, 1973, p. 44), have also found that most blacks prefer integrated rather than all black neighborhoods.

An important analytical question concerns how changes in racial integration would affect overall levels of neighborhood satisfaction. Considering all independent variables except percent black as constant, the model predicts satisfaction ( $S_1$ ) for white respondents, all whites preference as

$$S_1 = \alpha_1 - 0.055B_1 \quad (1)$$

where:  $\alpha_1$  = some constant, not necessarily the same as for the complete equation;  
 $B_1$  = percent neighborhood population black.

Similarly, for whites, mostly whites preference ( $S_2$ ), and for blacks, mixed preference ( $S_3$ ), the model predicts

$$S_2 = \alpha_2 - 0.198(B_2)^{0.4} \quad (2)$$

$$S_3 = \alpha_3 - 0.923(B_3)^{0.2} \quad (3)$$

Since the results (Table 2) show no statistically discernible effects for the other three groups, assume such effects are minimal and can be ignored. If we further assume the

percent neighborhood population black is about the same for most people in each race and preference group, we can estimate the predicted overall mean satisfaction,  $\bar{S}$ , by merely taking a weighted average of  $S_1$ ,  $S_2$ , and  $S_3$ , weighting each group's satisfaction by that group's relative frequency:

$$\bar{S} = 0.352S_1 + 0.217S_2 + 0.146S_3 + 0.285S_4 \quad (4)$$

$$= 0.352 [\alpha_1 - 0.055B_1] + 0.217 [\alpha_2 - 0.198(B_2)^{0.4}] + 0.146 [\alpha_3 - 0.923(B_3)^{0.2}] + 0.285S_4 \quad (5)$$

where:  $\bar{S}_4$  = mean satisfaction for whites, mixed and no difference preferences, and blacks, no difference preference.

Since  $S_4$  was assumed constant with respect to neighborhood racial composition, we can simplify eqn. (5):

$$\begin{aligned} \bar{S} &= \alpha^* + (0.352)(-0.055)B_1 + (0.217)(-0.198)(B_2)^{0.4} \\ &\quad + (0.146)(-0.923)(B_3)^{0.2} \\ &= \alpha^* - 0.019B_1 - 0.043(B_2)^{0.4} - 0.135(B_3)^{0.2} \end{aligned} \quad (6)$$

The analyst can use eqn. (6) to compare the predicted effect of different patterns of racial integration on the overall mean satisfaction of citizens with their neighborhood. Since the objective is to compare relative mean satisfaction for different patterns of integration, we can simplify eqn. (6) yet further by discarding the constant term:

$$S^* = -0.19B_1 - 0.043(B_2)^{0.4} - 0.135(B_3)^{0.2} \quad (7)$$

$S^*$  will be used to compare relative mean neighborhood satisfaction predicted to result from three different patterns of integration.

(1) *Complete Segregation.* Under a pattern of complete segregation, the percent neighborhood population black for all whites is 0%, and for all blacks 100%:

$$B_1 = B_2 = 0$$

$$B_3 = 100$$

$$S^* = -0.019(0) - 0.043(0)^{0.4} - 0.135(100)^{0.2} = -0.34$$

(2) *Complete Integration.* Under a pattern of complete integration, the percent neighborhood population black is the same for all residents. For the Detroit sample, the percent black is 24.9%:

$$B_1 = B_2 = B_3 = 24.9$$

$$S^* = -0.019(24.9) - 0.043(24.9)^{0.4} - 0.135(24.9)^{0.2} = -0.89$$

(3) *Integration of Blacks and Whites of Mixed and No Difference Racial Preference.* Under this integration pattern, the percent neighborhood population black for whites with all whites and mostly whites preferences is 0%. The percentage that blacks constitute of the remaining four groups combined is 49.3%. This pattern assumes that integration of these groups is complete in that all integrated neighborhoods have the same percent black

$$B_1 = B_2 = 0$$

$$B_3 = 49.3$$

$$S^* = -0.019(0) - 0.043(0)^{0.4} - 0.135(49.3)^{0.2} = -0.29$$

Comparing the results for the three integration patterns shows that complete integration lowers mean neighborhood satisfaction considerably compared to complete segregation, but that integration of only blacks and whites with no difference and mixed racial preferences slightly raises mean neighborhood satisfaction, compared to complete segregation. Predicted mean neighborhood satisfaction is over 0.5 seven-point rating scale unit less under complete integration compared to complete segregation, and 0.05 seven-point rating scale unit higher under the partially integrated pattern. Complete integration lowers overall mean satisfaction, compared to complete segregation, because the increased satisfaction for blacks with mixed racial preferences does not compensate for the decreased satisfaction for whites with all whites and mostly whites racial preferences. On the other hand, integration of blacks and whites with no difference and mixed racial preferences slightly increases overall mean satisfaction, compared to perfect segregation, by increasing satisfaction for blacks with mixed racial preferences without decreasing satisfaction for any other group.

The result that integration of blacks and whites with no difference and mixed racial preferences slightly raises mean neighborhood satisfaction, compared to complete segregation, is based on eqn. (7). Equation (7) does not include terms for the three groups showing no statistically discernible effect of neighborhood racial composition. We can test the sensitivity of the findings for  $S^*$  to the omission of these three groups using the coefficient estimates in Table 2 and assuming the effects for the previously omitted groups are approximately linear:

(1) *Complete Segregation*

$$\begin{aligned} S^{**} &= S^* + (0.097) (-0.004) (0) + (0.121) (-0.003) (0) \\ &\quad + (0.067) (-0.003) (100) \\ &= S^* - 0.02 = -0.34 - 0.02 = -0.36 \end{aligned}$$



*(2) Complete Integration*

$$\begin{aligned}
 S^{**} &= S^* + (0.097) (-0.004) (24.9) + (0.121) (-0.003) (24.9) \\
 &\quad + (0.067) (-0.003) (24.9) \\
 &= S^* - 0.02 = -0.89 - 0.02 = -0.91
 \end{aligned}$$

*(3) Integration of Blacks and Whites of Mixed and No Difference Racial Preference*

$$\begin{aligned}
 S^{**} &= S^* + (0.097) (-0.004) (49.3) + (0.121) (-0.003) (49.3) \\
 &\quad + (0.067) (-0.003) (49.3) \\
 &= S^* - 0.05 = -0.29 - 0.05 = -0.34
 \end{aligned}$$

The conclusions about the relative effects of these three integration patterns remain unchanged.

### Summary and Interpretation of the Results

The results of the analysis show two contrasting possible effects of racial integration on neighborhood satisfaction:

(1) Complete integration of all groups, including those strongly preferring to live in a racially homogeneous neighborhood, decreases overall mean neighborhood satisfaction.

(2) Integration among only those groups not strongly preferring to live in racially homogeneous neighborhoods does not decrease, and may even slightly increase, overall mean neighborhood satisfaction.

If not prevented by racial discrimination or other barriers, some type (2) integration may occur through personal location choices [17]. Thus, to the extent that the private housing market allows citizens to satisfy their preferences for the racial composition of their neighborhood, overall mean satisfaction may tend toward a maximum [18]. To the extent that barriers to individual choice have preserved a perfect segregation pattern, intervention through governmental policy can promote greater integration while maintaining and possibly slightly increasing overall satisfaction, providing that policy intervention results in integrating those groups preferring mixed neighborhoods or without strong preferences about the racial composition of their neighborhoods. However, if the effect of governmental policy is to integrate those groups strongly preferring racially homogeneous neighborhoods, overall mean neighborhood satisfaction will decline.

### Limitations of the Analysis

When presenting analytical results such as these, the policy analyst should carefully point out their limitations. The analysis statically assumes that the effect of neighborhood racial composition on neighborhood satisfaction does not change with experience living in integrated versus segregated settings. In predicting long-term effects, this static assumption may seriously mislead; Pettigrew (1973), for example, concludes on the basis of survey data available for recent decades that racially integrated housing may itself erode opposition to housing integration. Therefore, any predictive ability of the analysis presented here may be limited to the short-term. Moreover, using overall mean neighborhood satisfaction as a criterion for comparison is an arbitrary, *ad hoc* procedure. Although computationally convenient, mean satisfaction is not theoretically defensible as a social welfare function [19]. Nonetheless, this analysis does provide information about possible effects of racial integration on neighborhood satisfaction for different types of individuals, as well as information on the relative magnitude of the estimated effects for different patterns of racial integration. By providing some empirically-derived information about one of the outcomes that policy-makers may want to consider, the policy analyst increases the informational basis for policy decisions.

### Conclusion

Policy issues concerning social integration involve many complex issues not amenable to empirical analysis. However, when measurable individual-level outcome variables exist, empirical analysis can potentially contribute policy-relevant information on (1) the consequences of policy proposals for the outcome variables, and (2) the relative beneficial versus harmful consequences of different policies. Policy analysts should summarize and interpret the analytical results in a manner understandable to policy-makers not trained in statistics, at the same time pointing out the limitations, assumptions, and possible flaws in the analysis. In this way, empirical studies of the effect of social context on individual-level outcome variables can contribute useful information to policy-making.

### Acknowledgments

The author gratefully acknowledges his debt to Carl Hensler, incurred during their many conversations together, for developing ideas presented in this paper about statistical inference in contextual analysis and the relevance of contextual analysis to public policy. Robert Marans and Willard Rodgers generously allowed the author to use data from their Detroit Quality of Life Study, and provided useful comments on a draft of this manuscript. Robert Mowitz and James McDavid also provided helpful comments.

### Notes

1. Individual-level socio-demographic variables such as race or sex are clearly constants vis-a-vis governmental policy. Others, such as education or income, can potentially be influenced by governmental policy, but only in the long-term.
2. Individual-level socio-demographic variables can usually be considered proxy variables. For example, a student's family income or race may be proxies for cultural differences in upbringing that influence later educational performance. Similarly, a prisoner's criminal background may be a proxy for the extent the prisoner has learned criminal attitudes and behavior.
3. Assigning a large number of new students of one race to a school would itself change the "treatment" variable, the school's racial composition. Therefore, to preserve the analogy to the classic experimental design the number assigned would have to be small compared to the total student population.
4. Bias will result only if (1) the individual-level variable has an effect on the dependent variable, and (2) the individual-level variable and neighborhood variable are statistically associated.
5. For  $n$  contexts, the analyst includes  $n-1$  dummy variables, allowing one context to be the reference.
6. The number of contexts limits the number of contextual variables that can be included in the model. Given  $n$  contexts, the model can include no more than  $n-1$  contextual variables, regardless of the number of individuals from each context, because  $n$  or more contextual variables will be linearly dependent.
7. Some researchers (e.g. Lansing et al., 1970; Marans and Wellman, 1977) constructing neighborhood satisfaction scales from these items have omitted items (1) and (4), to keep satisfaction with one's neighbors separate from general neighborhood satisfaction. This study includes those items, since that component of general neighborhood satisfaction concerning satisfaction with neighborhood residents is relevant to the study.
8. Alternatively, the percent housing units that are black occupied could be used. However, these variables correlate so highly for this sample ( $r = 0.996$ ) that the choice is inconsequential.
9. Individual-level and contextual education variables were tested in the equations but their coefficient estimates were not statistically significant. They were dropped from the equations because of high multicollinearity with other independent variables.
10. Since dividing the sample decreases the sample size used to estimate each equation, this approach has the disadvantage of increasing the sampling variances of the estimators. By using interaction terms, the analyst can often avoid dividing the sample, although in this example the number of possible interactions and the specification of the functional form made that approach impractical.
11. Entering a squared term into the equation was not practical because of the high correlation (0.95 for whites, 0.97 for blacks) between the percent black and the percent black squared variables.
12. The dummy racial composition variables were defined differently for equations estimated for black and for white respondents, since the distribution of the percent black variable differs greatly for these two groups.
13. The estimated standard errors are only approximations, since they assume simple random sampling. However, judging from Frankel's (1971) findings about design effects for partial regression coefficients, these approximations are probably reasonable.
14. This tendency may result from what Taylor (1979, p. 35) refers to as social disorganization found in black neighborhoods. For white respondents, racial prejudice may add to the effect attributable to social disorganization.
15. Note that  $\lambda = 1$  for white respondents, all white preferences, and that the dependent variable is in the original seven-point rating scale units.
16. Marans and Rodgers (1975, pp. 314-315) found that community satisfaction tends to increase with age. Campbell et al. (1976, p. 152) found positive relationships between age and satisfaction in a variety of domains.
17. The Detroit data do show some relationship between preferences and actual neighborhood racial composition. Of those whites voicing an all whites or mostly whites preference, 85% and 88%, respectively, live in areas with less than 1% black population, and 60% and 62%, respectively, live in completely white areas. In contrast, of those whites having mixed or no difference preferences, 70% (both groups) live in areas with less than 1% black population, and 42% and 41%, respectively, live in completely white areas. Of those blacks with mixed racial preferences, 40% live in areas with less

- than 75% black population and 72% live in areas with less than 90% black population. In contrast, of those blacks voicing a no difference preference, 20% live in areas with less than 75% black population, and 54% live in areas with less than 90% black population.
18. Urban economists have investigated the related question of under what conditions stable patterns of integrated housing are possible (e.g., Courant and Yinger, 1977; Yinger, 1976). Yinger (1976) concludes that market competition without discrimination cannot generate stable integrated housing patterns. However, Yinger's analysis assumes that all whites prefer not to live in integrated neighborhoods.
  19. The impossibility of constructing a social welfare function stems from the inability to make interpersonal comparisons in a theoretically sound way. Computing a mean satisfaction value aggregates the measured satisfactions of individuals, and therefore implicitly makes interpersonal comparisons.

### References

- Armor, David J. (1972). "The Evidence on Busing," *Public Interest*, 28: 90-126.
- Boyd, Lawrence H. Jr. and Iversen, Gudmund R. (1979). *Contextual Analysis: Concepts and Statistical Techniques*. Belmont, CA.: Wadsworth.
- Campbell, Angus, Converse, Philip E. and Rodgers, Willard L. (1976). *The Quality of American Life*. New York: Russell Sage.
- Coleman, James S., Campbell, Ernest Q., Hobson, Carol J., McPartland, James, Mood, Alexander M., Weinfeld, Frederic D. and York, Robert L. (1966). *Equality of Educational Opportunity*. Washington: U.S. Government Printing Office.
- Courant, Paul N. and Yinger John (1977). "On Models of Racial Prejudice and Urban Residential Structure," *Journal of Urban Economics*, 4: 272-291.
- Farley, Reynolds, Bianchi, Suzanne and Colasanto, Diane (1979). "Barriers to the Racial Integration of Neighborhoods: The Detroit Case," *The Annals of the American Academy of Political and Social Science*, 441: 97-113.
- Frankel, Martin R. (1971). *Inference from Survey Samples: An Empirical Investigation*. Ann Arbor, Mich.: Institute for Social Research, University of Michigan.
- Hauser, Robert M. (1970). "Context and Consex: A Cautionary Tale," *American Journal of Sociology*, 75:645-664.
- Hauser, Robert M. (1974). "Contextual Analysis Revisited," *Sociological Methods and Research*, 2: 365-375.
- Hensler, Carl and Stipak, Brian (1979). "Contextual Analysis: Problems of Statistical inference and their Solutions." Unpublished manuscript available from Brian Stipak, Institute of Public Administration, Pennsylvania State University, University Park, PA 16802.
- Lansing, John B., Marans, Robert W. and Zehner, Robert B. (1970). *Planned Residential Environments*. Ann Arbor, Mich.: Institute for Social Research, University of Michigan.
- Little, James T. (1976). "Residential Preferences, Neighborhood Filtering, and Neighborhood Change," *Journal of Urban Economics*, 3:68-81.
- Marans, Robert W. and Rodgers, Willard (1975). "Toward an Understanding of Community Satisfaction," in Amos H. Hawley and Vincent P. Rock, (eds.), *Metropolitan America in Contemporary Perspective*. New York: Wiley.
- Marans, Robert W. and Wellman, John D. (1977). *The Quality of Non-Metropolitan Living: Evaluations, Behaviors, and Expectations of Northern Michigan Residents*. Ann Arbor, Mich.: Institute for Social Research, University of Michigan.
- Pettigrew, Thomas F. (1973). "Attitudes on Race and Housing: A Social-Psychological View," in Amos H. Hawley and Vincent P. Rock (eds.), *Segregation in Residential Areas*. Washington: National Academy of Sciences.
- Taylor, D. Garth (1979). "Housing, Neighborhoods, and Race Relations: Recent Survey Evidence," *The Annals of the American Academy of Political and Social Science*, 441: 26-40.
- U.S. Commission on Civil Rights (1967). *Racial Integration in the Public Schools*. Washington: U.S. Government Printing Office.
- Yinger, John (1976). "Racial Prejudice and Racial Segregation in an Urban Model," *Journal of Urban Economics*, 3: 383-396.