Accepted for publication in 2003 in **Annales d'Économie et de Statistique**

University of Connecticut

Department of Economics Working Paper Series

Segregation and Racial Preferences: New Theoretical and Empirical Approaches

Stephen L. Ross
University of Connecticut

Working Paper 2002-04R

April 2002, revised April 2003

341 Mansfield Road, Unit 1063
Storrs, CT 06269–1063
Phone: (860) 486–3022
Fax: (860) 486–4463
http://www.econ.uconn.edu/

Abstract

This paper investigates the role of preferences for social interactions or outcomes in determining observed patterns of racial segregation. In the theoretical section, consumers maximize utility by allocating time between personal time and social interactions within their neighborhood, and the dual of this problem is used to investigate the bidding and sorting of households over racial composition. The models suggests that African-American households may outbid white households to reside in white neighborhoods, and unlike previous models of segregation this model is consistent with either racial segregation or integration. In the empirical analysis, proxy variables are developed for unobservable attributes that enter household preferences based on measures of household outcomes and satisfaction, and then specifies an econometric model of residential location choice using those attributes. The paper finds evidence that racial differences in preferences for education can explain a substantial portion, but not all, of the racial segregation observed in 1985 Philadelphia using data from the American Housing Survey.

Journal of Economic Literature Classification: D1, D4, J7, R2

Segregation and Racial Preferences:

New Theoretical and Empirical Approaches

Racial segregation in housing has been a feature of major American Metropolitan areas ever since the first major rural to urban migration in the late 1800's. African-Americans are more segregated than other minority groups, and only a small fraction of this segregation can be explained by socio-economic factors. While segregation has declined over the last few decades, the levels are still quite high in many urban areas (Massey and Denton, 1993). Moreover, these declines have been accompanied by a narrowing of the black-white wage gap (Daly and Couch, 2000) and as a result socioeconomic factors still cannot explain very much of the remaining level of segregation (DeRango, 1999). Naturally, researchers have searched for alternative explanations for the high level of segregation in American metropolitan areas. Much of this debate has focused on housing market discrimination and racial differences in preferences as possible explanations.

This paper proposes and implements new theoretical and empirical approaches for examining the underlying causes of racial segregation in housing. Specifically, the theoretical and empirical models in this paper assume that households do not have preferences for neighborhood racial composition, nor for other easily observable location attributes like percent poverty or average education, but rather assume that apparent preferences for observed location attributes are derived demands that arise from a household's underlying preference for social interaction or for positive outcomes that arise from those social interactions.[1] Prejudice, cultural affinity, and other factors related to race and ethnicity may influence outcomes, but they do so by affecting the quality of across group social interactions. The paper sets up a simple consumer utility maximization problem where households allocate time between personal time and social interactions. The paper also specifies an econometric model of residential location choice in which the unobservable attributes that actually enter household preferences are replaced with proxy variables. These proxy variables are developed in a first stage based on measures of household outcomes and satisfaction. The model is estimated using the 1985 metropolitan sample of the American Housing Survey for Philadelphia.

On the theoretical side, traditional analyses have simply assumed that households have a preferences concerning the race of their neighbors (Schelling, 1972; Bailey, 1966; Schnare and MacRae, 1978) or their proximity to locations in which another group resides (Yinger, 1976, and Kern, 1981). The typical implication of such models is that complete segregation is the only stable equilibrium, as long as African-American households do not have a stronger preference to reside among white households than white households themselves. An assumption that on the surface seems quite reasonable.[2] Yinger (1986) suggests that housing discrimination may be practiced to prevent neighborhood racial transition when the minority population within a metropolitan area is growing, and Courant (1978) proposes a search model in which possibility of discrimination may raise minority search costs in predominantly white neighborhoods. While Yinger and Courant both illustrate that discrimination can in principle increase racial segregation, neither paper undermines the strong racial sorting implications of the racial preference literature.

More recent studies have focused on human capital accumulation as a motivation behind residential location choice. Benabou (1993), Borjas (1998), and Cutler and Glaeser (1997) examine models in which households sort across neighborhoods or communities based on peer group effects in education or human capital. Benabou shows that suboptimal equilibria exist in which low skill agents are segregated and drop entirely out of the labor market. Alternatively, Borjas focuses on integrated equilibria showing that income differences between the groups lead to increased segregation, but that education differences can decrease segregation if low skill households outbid high skill households for access to locations dominated by high skill residents. Cutler and Glaeser (1997) examine a model in which there are two ethnic and racial groups plus heterogeneity in human capital within the groups. In their model, household sort over skill level and high skill individuals outbid low skill individuals. If racial/ethnic discrimination is large enough, low skill members of the minority group may actually benefit from discrimination relative to an outcome in which all high skill individuals are segregated from all low skill individuals. In this style of model, racial/ethnic integration is possible without the minority group having a very strong preference to live with the majority group, but only because sorting is now driven by preferences over the distribution of skills in a location. Specifically, in Borjas's model the strong assumption concerning minority preferences is simply replaced with a strong assumption concerning the preference of low skill individuals.[3]

The theoretical model in this paper is closest to the model provided by Borjas. In Borjas's model, household utility is a function of the human capital externalities offered in its residential location, and social or cultural factors create a barrier to the transmission of human capital between racial and/or ethnic groups. Borjas's model is extended by recognizing that social interactions are costly and adding a trade-off between leisure and social interactions subject to a time constraint. In this way, the model is closer in approach to traditional discrimination models in labor economics where race influences an individual's decision to invest in human captial, see Lundberg and Startz (1983) and Lang (1986). When households choose a level of investment in social interactions and the consumption of leisure and social interaction are complementary, a member of the minority group may outbid an equally skilled member of the majority group for access to a predominantly majority neighborhood, and stable, integrated equilibria may exist.

On the empirical side, a major approach in the study of racial segregation has been the analysis of racial differences in housing prices. A finding that African-Americans pay a higher price for housing than whites suggests that they face substantial constraints in their residential location choices potentially due to housing discrimination. Studies from the 1960's tend to find evidence that African-Americans pay more for equivalent housing (King and Mieszkowski, 1973, Yinger, 1978), while studies from the 1970's (Schnare, 1976, Follain and Malpezzie, 1981) tend not to find evidence of a housing price premium. Cutler, Glaeser, and Vigdor (1999) confirm this pattern finding that the African-American rent premium fell dramatically between 1940 and 1970 and had reversed entirely by 1990. They argue that today segregation in America is enforced by a "decentralized racism" where whites outbid African-Americans for houses in white neighborhoods, and therefore pay more for housing than African-Americans.[4] Alternatively, Schafer (1979), Chambers (1992), and Kiel and Zabel (1996) argue that earlier work failed to find an African-American price premium because that work did not control for neighborhood quality and did not account for housing submarkets within metropolitan areas.[5]

This paper develops an alternative approach by drawing from and building on three separate literatures. First, many studies (e.g. Bayer, McMillan, and Rueben, 2002; Deng, Ross, and Wachter, 2001; Gabriel and Rosenthal, 1989; and Waddell, 1992) examine racial differences in residential location choices using a multinomial choice framework. These studies consistently find that African-American households are much more likely to reside in locations with high concentrations of African-Americans and with unfavorable attributes, such as high poverty rates or low rates of homeownership, even after controlling for differences in endowments and family structure. These studies simply examine conditional frequencies and cannot distinguish between preferences for a type of neighborhood and constraints that force a group into those neighborhoods.[6] Alternatively, Ihlanfeldt and Scafidi (In Press) use a unique data set in which individuals rank their preferences for neighborhoods with different levels of segregation. They conclude that if all African-Americans had a preference for perfect integration observed segregation would fall by 16.9, 9.2, and 10.3 percentage points in metropolitan Atlanta, Detroit, and Los Angeles, respectively. The two main concerns with their analysis are the endogeneity of the individual's revealed preference and their single dimension approach to location preferences. Finally, Borjas (1998) examines the determinants of ethnic residential segregation using the National Longitudinal Survey of Youths. He finds that both family and group human capital have an important effect on segregation, but as with Ihlanfeldt and Scafidi, his analysis looks at segregation in isolation from other factors that affect residential location.

This paper incorporates information from self-reported satisfaction and outcome measures into a standard multinomial model of the residential location choice. This information is used to isolate the influence of preferences on residential location outcomes. The responses to neighborhood satisfaction measures and to an employment status question are assumed to provide information about the amenities and opportunities offered by a location. Proxy variables for location quality on each dimension are created by estimating a model for satisfaction or outcomes based on household characteristics and neighborhood attributes and by calculating predicted values for each location based on the neighborhood attributes. The resulting proxy variables along with variables to control for racial composition and price level are included in the residential location choice model. If racial differences in location remain after controlling for these proxy variables, the residual differences cannot be explained by systematic racial differences in preferences for the attributes considered. Any unexplained segregation may be the result of racial prejudice or constraints that are operating in the housing market, such as housing market discrimination or racial steering. The paper finds evidence that racial differences in preferences concerning education can explain a substantial portion of the racial segregation observed in the sample, but that a sizable, unexplained component of segregation remains.

The paper is organized as follows. The next section presents a simple theoretical model in which household preferences depend upon social interactions which are costly and vary in value and cost by race. The third and fourth sections present an econometric model of residential location choice and describe the data that will be used to estimate such a model, respectively. The fifth section presents the results of the estimations, and the sixth section summarizes those results and suggests directions for future research.

Theoretical Framework for Studying Segregation and Preferences

This model is based on the premise that people care about the social interactions offered by a location rather than racial composition itself. Households exhibit a preference for racial composition because composition influences the opportunity for and quality of social interactions. Racial prejudice and other sociological factors may enter through the social interaction quality experienced by one group when that group interacts with a second group. Social interactions may be intrinsically valued or may be valuable because they lead to peer group effects as in DeBartolome (1990), also see Lundberg and Startz (1998) who examine how individuals search for productive interactions when race plays a factor in the return to those interactions.

In this model, there are two types of households (B,W). The utility of household type j depends upon the level of social interaction (I), leisure (L), and aggregate consumption. In order to focus our attention on the trade-off between social interaction and leisure, income (Y), the time endowment (T), and land consumption are assumed to be exogenously determined, and utility is assumed to follow a consumer surplus specification.

$$Max_{I,t,k} \quad U_j = V(I, T-t) + Y - P_k \tag{1}$$

$$I = \theta_{jk}\mu_{jjk}t + (1 - \theta_{jk})\mu_{jik}t \tag{2}$$

where t is the time allocated to social interaction, k is the community chosen, P_k is the price or rent for the standard consumption level of land in community k, θ_{jk} is the share of the population in community k made up of type j households, μ_{jik} is the average quality level of social interactions of group *j's* interactions with group *i* in community *k*. Note that leisure (*L)* is simply *T-t*. This specification arises from the assumption that households cannot control who they encounter in the search for social interactions and therefore their encounters are described by the distribution of the population within their chosen community.[7]

The first order condition for this problem is

$$\theta_{jk}\,\mu_{jjk} \;+\; (1-\theta_{jk})\,\mu_{jik} \;=\; U_L / U_I \;=\; MRS_{L,I} \qquad (3)$$

where the left hand side represents the marginal return from an increased allocation of time to social interactions and the right hand side is the marginal rate of substitution between leisure and social interactions. In addition to continuity and differentiability, some basic assumptions are imposed upon *V*. First, *I* and *L* are both assumed to be normal goods, but may be either substitutes or complements, and other things equal the marginal utility of social interactions is inelastic with respect to time investments.

$$\frac{dL}{dT},\frac{dI}{dT} > 0,\;\; -\frac{t}{V_I}\,\frac{dV_I}{dt} < 1 \qquad (4)$$

In order to examine the sorting problem, we abstract away from the community choice problem and write the household bid problem, which is simply the dual of utility maximization, as a function of racial composition

$$Max_t \;\; P_j \;=\; V\left(\theta_j\,\mu_{jj}\,t \;+\; (1-\theta_j)\,\mu_{ji}\,t\,,\; T-t\right) \;+\; Y \qquad (5)$$

Group populations are assumed to be homogenous and therefore the social interaction return parameters do not vary with racial composition, i.e. there is no within group sorting based on the quality of social interactions.

Finally, the model assumes that W offers higher quality social interactions than B, own group interaction is more productive than across group interactions controlling for the quality of the interaction, and the penalty or loss associated with across group social interaction is symmetric across groups.[8] Specifically,

$$\mu_{WW} > \mu_{BB} \text{ and } \mu_{WW} - \mu_{BW} = \mu_{BB} - \mu_{WB} > 0 \tag{6}$$

A segregated equilibrium can be expected to arise when the slope of R_B is always greater than slope of R_W over θ_B. This condition assures that household bids satisfy a standard single crossing condition over θ_B; namely, that the bids only cross once assuring perfect sorting of W households into W communities and B households into B communities. If this condition is violated, this model may be consistent with integrated equilibria and discriminatory barriers may be required to enforce segregated outcomes. The assumption that W offers higher quality social interactions, however, creates the possibility that the bid of group B over the fraction B in the neighborhood is flatter than the bid of group W, or group B outbids group W for predominantly W neighborhoods.

Theorem: For the above specification, segregation is a stable equilibrium if group B members benefit more from interacting with group B than with group W or if leisure and social interactions are substitutes.

Proof: The slope of the household bid-rent function can be found using the envelop theorem.

$$\frac{dP_B}{d\theta_B} = V_I \; t_B^* \left(\mu_{BB} - \mu_{BW}\right) <> 0 \tag{8}$$

$$\frac{dP_W}{d\theta_B} = V_I \; t_W^* \left(\mu_{WB} - \mu_{WW}\right) < 0 \tag{9}$$

where t_j^* is the optimal time investment in social interactions for group j.

Segregation is a stable equilibrium if W outbids B in predominantly W neighborhoods, which occurs if the slope of W's bid is less than the slope of B's bid over the share of B for all shares of B less than 0.5. If group B members benefit more from interacting with group B than with group W, the expression in equation (8) is positive and group B bids always have a greater slope.

On the other hand, if group B members prefer to interact with group W members or equivalently the racial difference in interaction quality exceeds the loss from across group interaction, the form of equilibrium depends on the relative slopes.

$$\begin{aligned} \frac{dP_W}{d\theta_B} - \frac{dP_B}{d\theta_B} &= V_I(t_W^*) \; t_W^* \left(\mu_{WB} - \mu_{WW}\right) - V_I(t_B^*) \; t_B^* \left(\mu_{BB} - \mu_{BW}\right) \\ &= \left(V_I(t_W^*) \; t_W^* - V_I(t_B^*) \; t_B^*\right)\left(\mu_{WB} - \mu_{WW}\right) \\ &\qquad - V_I(t_B^*) \; t_B^* \left(\left(\mu_{BB} - \mu_{BW}\right) - \left(\mu_{WB} - \mu_{WW}\right)\right) \end{aligned} \tag{10}$$

The second term in equation (10) is unambiguously negative based on the conditions in equation (6). Also based on equation (6), the first term will be negative if and only if

$$V_I(t_W^*)\ t_W^* - V_I(t_B^*)\ t_B^* > 0 \tag{11}$$

Next, based on the conditions in equation (4), equation (11) holds if and only if

$$t_W^* > t_B^* \tag{12}$$

since an increase in *t* increases the product of *t* and the marginal utility of social interactions. Finally, when the share B is below 0.5, the return to time spent in social interaction (I_j) is unambiguously higher for group W. Specifically,

$$\begin{aligned} \frac{dI_W}{dt} - \frac{dI_B}{dt} &= \left(\theta_B \mu_{WB} + (1-\theta_B)\mu_{WW}\right) - \left(\theta_B \mu_{BB} + (1-\theta_B)\mu_{BW}\right) \\ &= \left(\mu_{WW} - \mu_{BW}\right) - \theta_B\left(\mu_{BB} - \mu_{WB}\right) - \theta_B\left(\mu_{WW} - \mu_{BW}\right) \\ &= \left(\mu_{WW} - \mu_{BW}\right)\left(1 - 2\theta_B\right) > 0 \end{aligned} \tag{13}$$

The first term is positive and the second two terms are negative. All terms represent the penalty associated with across group social interaction, and the term is positive for neighborhoods that are predominantly W. Specifically, group W faces a lower price for social interactions in terms of lost leisure and as a result chooses a higher level of time investment or lower leisure consumption. Therefore, when *I* and *L* are substitutes, equations (12) and accordingly (11) are satisfied and the expression in equation (10) is unambiguously negative. #

On the other hand, if leisure and social interactions of complements and group W offers higher quality social interactions, the standard single crossing property may not be satisfied and segregation may not be an equilibrium. Group B may outbid group W in predominantly W neighborhoods because the higher price of social interactions leads to greater time investment and an accordingly greater return from interacting with group W as compared to the return received by the members of group W. This result reverses the strong sorting conclusions of the earlier literature, Schelling (1972), Bailey (1966), Schnare and MacRae (1978), and Kern (1981) This finding arises without a strong, exogenous assumption that low skill groups outbid high skill groups for predominantly high skill locations, and accordingly, raises questions about whether individual preferences for racial segregation can maintain the level of residential segregation observed in American society.

It is notable that this finding is also consistent with the previous literature. African-American households outbid whites in white locations whenever their return from the social interactions in these locations is higher than the return to white households; in other words, when African-American households have a stronger desire to reside in white neighborhoods. The above model simply shows that the assumption that "minorities do not outbid whites to reside in white neighborhoods" may be a much stronger assumption than previously thought. This assumption may be violated whenever the social interaction quality difference between groups is larger than the quality loss resulting from across group communication and leisure time and social interactions are complements.

Econometric Framework for Studying Segregation and Preferences

The utility level achieved by an individual *i* of race *j* in a given location *k* is

$$U_{ik} = \beta_Z X_i Z_{kj} + \beta_R X_i R_k + \beta_P X_i P_k + \epsilon_{ik} \tag{14}$$

where X is a vector of individual characteristics, Z is a vector of location attributes, R is the racial composition, P is the price level of housing, ϵ is a random error term, and the β's are behavioral parameters. The location attributes that actually enter preferences, Z, are allowed to vary by race. While demographic attributes, such as poverty rate, or performance measures, such as standardized school tests scores, are the same for all residents regardless of race, the actual benefits arising from residence in a neighborhood, such as employment or educational opportunities, may vary across racial groups.

The vector Z is actually unobserved; rather, information is available on easily recordable attributes of the location (W), such as percentage of households in poverty or average education of location residents, and potentially on resident satisfaction or outcomes (S) on the self-report of location residents. Assume that an individual's satisfaction or outcome depends upon individual characteristics and the location attributes that enter preferences

$$S_{ik} = \delta_{Xj} X_i + \delta_{Zj} Z_{kj} + \delta_{Rj} R_k + \mu_{ik} \tag{15}$$

and these location attributes can be predicted by observable location attributes

$$Z_{kj} = \gamma_{Wj} W_k R_k + \nu_{kj} \tag{16}$$

where the δ's and γ's are parameters that vary by race of the individual, j, and μ and ν are random error terms.

Substituting equation (16) into equation (15) yields the following reduced form equation

$$S_{ik} = \delta_{Xj} X_i + \delta_{Wj} W_k R_k + \tilde{\mu}_{ik} \tag{17}$$

Equation (17) can be consistently estimated by ordinary least squares if μ and ν are uncorrelated with the ϵ's associated with the location choices.

The estimated coefficients from equation (17) are used to create proxy variables for Z

$$\hat{Z}_{kj} = \hat{\delta}_{Wj} W_k R_k \tag{18}$$

which can be used to estimate equation (14)

$$U_{ik} = \beta_Z X_i \hat{Z}_{kj} + \beta_R X_i R_k + \beta_P X_i P_k + \tilde{\epsilon}_{ik} \tag{19}$$

If the error term in equation (19) is distributed as extreme value, the second stage of the problem can be modeled as a multinomial logit.

The assumption that μ and ν are uncorrelated with the ϵ's is a serious limitation of this analysis that needs to be addressed in future research. This assumption implies that the household residential location is exogenous to the dissatisfaction and outcome measures. Unobservable household variables, however, may influence both residential location choice and satisfaction with a given residential location, and under those circumstances the observable residential location attributes are likely to be endogenous to the satisfaction response or outcome variables. In other words, the residential sorting process may create a Tiebout-like bias in the first stage model estimates.

Sample and Data Description

The sample for this analysis is drawn from the 1985 Metropolitan Area (Metro) sample of the American Housing Survey (AHS) for Philadelphia. The Metro sample of the AHS contains detailed housing characteristics and the location of the housing unit down to a census tract identifier, which identifies all housing units that belong to the same tract, but does not actually identify the tract itself. The location of the housing unit is also described by its placement into one of 35 zones with population of approximately 100,000 each. The city of Philadelphia is divided into 13 zones, and the rest of the metropolitan area is divided into 22 zones. The survey also contains information on family structure and family member demographics, such as age and education level.

The 1985 survey included both a commuting and a neighborhood supplement. The commuting supplement collected limited information on the labor market outcomes of each family member including the employment location at the zone level for all family members who are currently employed and work at a fixed location. The neighborhood supplement collected information concerning the environment immediately surrounding the housing unit (within 300 feet), as well as whether the household respondent was dissatisfied with the neighborhood in general and for any specific reasons.[9]

The base sample includes all housing units in the sample that satisfy the following criteria:

1. The housing unit is occupied,
2. The housing unit is located in a tract containing 5 or more occupied units,
3. The household occupying the unit has annual family income exceeding $1000.
4. The head of household's race is white or African-American

This criteria leads to a sample of 3,971 occupied housing units in 495 census tracts. Each housing unit contains a single household whose responses form the basis for analysis. A second sample of prime age adults is created by selecting all household members whose age falls between 25 and 60 years from the base sample of households. Individual households may contain more than one prime age adult, and the resulting sample contains 4,963 prime age adults..

Three binary dissatisfaction/negative outcome measures are created. The first two measures are dissatisfied with public schools and/or children not attending public schools (education) and bothered by people in neighborhood (neighbors), which are based on responses at the household level. Table 1 contains the means and standard deviations of household variables by race. White households in public schools are somewhat less likely to report dissatisfaction with the public schools, but substantially more likely to opt out of the public school system leading to whites having a seven percentage point higher incidence of dissatisfaction.[10] On the other hand, African-Americans are nine percentage points more likely to report being bothered by people in the neighborhood.[11] The third measure is whether a prime age adult is not working and is based on the adult sample, see Table 2. African-American prime age adults are fourteen percentage points less likely to be working in this sample.

Table 1 also describes the socio-economic characteristics of households, which are used to control for household variation in the likelihood of dissatisfaction and in tastes for the attributes of residential locations. In this sample, white households are more likely than African-American households to involve married couples, to have higher education levels, to have higher family income, and to have smaller families. A similar set of control variables are created for the sample of prime age adults. The means for these variables are described in Table 2, and the pattern of racial differences in the adult sample is quite similar to the pattern in the household sample.

A detailed set of variables is created to describe each census tract represented in the sample. First, three socio-economic variables, mean years of education, mean family income, percent African-American, are created based on averages from either the household or prime age adult samples. Mean years of education is created from the sample of prime age adults, and mean family income and percent African-American are created from the sample of households. Both mean years of education and family income are means based on truncated tails. If the number of observations in a tract is between five and nine, the highest and lowest values are dropped from the calculation of the average. The two highest and two lowest values are dropped if the number is between ten and nineteen, and the four top and bottom values are dropped if there are twenty or more. Finally, when the means of these variables are assigned to a given household, this household's influence on the value of those variables is eliminated from the mean. Specifically, the respondent race, household income, and education level of family members are eliminated from the calculation of the averages, and each household is considered to evaluate the tract in which they reside as if no one resided in their housing unit. The residential zone in which the housing unit is located is also used to create a variable that is one if the unit is located in the central city and zero otherwise.

Second, the neighborhood and commute time supplements are used to create tract level variables. The neighborhood supplement is used to calculate the percentage of households in a tract reporting that their housing unit is located near commercial or industrial property or that their housing unit is located near green space. The commute time supplement and the sample of employed prime age adults is used to estimate a gravity model in which the flow of commuters between each tract and work zone depends upon the number of prime age adults in the tract, the number of employed adults working in the zone, and the mean travel time between the tract and the zone. An employment access measure is created as a weighted average of zone employment totals, where the parameter estimates and the commute times between a tract and each zone are used to create the weighting scheme.[12] In addition, a labor market earnings model is estimated using the adult sample based on adult characteristics and work location dummy variables. The mean unobserved, labor market quality of adult workers is calculated by averaging the work zone fixed effects for each employed adult in a tract.[13] As with the tract demographic variables, the work locations of household members of any given household do not affect the mean labor market quality for that household.

Third, a variable to represent the price level of housing is created from a housing price hedonic. Specifically, a model for the logarithm of house value is specified as a linear function of structural housing characteristics and census tract fixed effects. The estimated fixed effects are used to represent the price level.[14] This price level captures the premium that households are willing to pay to reside in specific locations and might be interpreted as the overall quality of the residential location. The other location attributes can explain approximately 50 percent of the variation in price level in the sample of census tracts. The unexplained variation may represent random error or market imperfections caused by factors like moving costs or racial discrimination, but this variation may also arise because the other eight location attributes do not capture all relevant features for describing the desirability of a location.

The means and standard errors of these variables are shown by racial composition in Table 3. Predominantly African-American tracts tends to be located in the central city, near commercial property, and away from green space. These tracts also have lower average income, lower education levels, worse job access, but actually better unobserved labor market quality of residents (after controlling for education and family structure). As a result of these negative location attributes, the price level of housing is lower in African-American locations.

Estimation and Simulation Results

Dissatisfaction/Negative Outcome Models

The first stage of the specification (equation 17) requires the estimation of linear models for dissatisfied with public schools/opted out of public schools and for bothered by people in the neighborhood using the household sample and for not working using the prime age adult sample. These estimations are conducted using a logit specification. A series of models are considered for each dependent variable, and the resulting loglikelihood values are shown in Table 4. The first row represents a baseline model that controls only for household or individual characteristics and interacts those characteristics with the race of the household head or adult. The next two models or rows include the location attributes and include those attributes plus the interaction between percent African-American and the other location attributes, respectively. The fourth model allows for the influence of the percent African-American interactions to vary by the race of the head of household or adult, and as such allows the effect of racial composition on unobserved location attributes to vary by race. The final model interacts all location attributes with race, or in other words adds the interaction of race with the non-interacted location attribute variables to the previous specification.

The data strong support the specification in which the influence of percent African-American on the return to other location attributes varies by race. In all but one case (model 3 for education), the earlier model is rejected for the expanded model until we get to the last row. In addition, the racial composition interaction model (model 3) is rejected in favor of a model that includes race interactions (model 4) for all three dependent variables. A model that only interacts race with the percent African-American interactions (model 4), not the attributes alone, cannot be rejected for a model that interacts all location variables with race (model 5).[15]

It should be noted that the interaction of race only with the percent African-American interactions implies an assumption concerning at what percent African-American the effect of location attributes are equal across the two races. A simple interaction action of percent African-American implies the effect of location does not vary by race when percent African-American equals zero, but this assumption only maximizes the log-likelihood function for the employment variable. The log-likelihood function for education and neighbors are maximized by specifications that assume that the effect of location does not vary by race when percent African-American equals 0.50 and 1.00. The likelihood values shown in Table 4 are based on these higher likelihood values, but this generalization does not change the main empirical findings of the paper.

The estimates for the fourth specification are shown in Table 5. The employment equation includes standard individual level variables as well as interactions involving the adult's gender and family structure. The dissatisfaction specifications are quite similar except for the interactions. For the employment variable, marital status is interacted with the head of household's age and education level. Race is interacted with all household/adult level variables, except for the marital status interaction terms which are omitted for parsimony.[16] As discussed above, tract level variables are included plus their interactions with racial composition and the effect of the racial composition interactions vary by race.

Focusing on the tract variables, which determine the instruments for the second stage, the coefficient estimates describing the direct effect of an attribute are shown in the nine rows that immediately follow the individual or household variables. The racial composition interactions and the interaction of those variables by race are in the last nine rows of the table under the Level and Race columns, respectively. Employment opportunities are worse for central city residents (a positive coefficient estimate for the direct effect). As tract education rises, employment opportunities improve for whites residing in locations with a high percentage of African-American residents (negative racial composition interaction estimate in the level column) . This education effect does not exist for African-Americans (coefficients in level and race column have the same magnitude and opposite sign). Increases in tract average family income and unobserved labor market quality, however, lead to improved employment opportunities for African-Americans in predominantly African-American neighborhoods (negative racial composition interaction estimates in race column).

The likelihood of dissatisfaction with public schools decreases in the central city and for not being near green space (negative coefficients for the direct effect). This result may arise from the assumption that private school attendance, which is concentrated among suburban residents, is evidence of dissatisfaction. It also must be noted, however, that these findings may arise from a Tiebout-like bias where more easily dissatisfied households sort into the neighborhoods with the best schools. The green space effect is exacerbated for whites in predominantly African-American neighborhoods and eliminated for predominantly white neighborhoods (positive coefficient in level column for interaction with percent African-American minus 0.5, which has no effect when percent African-American equals 50%). The effect of near green space does not vary by racial composition for African-Americans (race column for interaction with percent African-American). These results may be driven by a relationship between neighborhood racial composition and the attractiveness of public parks to white households.

Finally, labor market quality (negative direct effect coefficient) and suburban location (positive coefficient on central city) increases satisfaction with neighbors. These effects increase in magnitude as a neighborhood moves from being predominantly African-American to an integrated neighborhood and finally to a predominantly white neighborhood (opposite signs in level column for interaction with percent African-American-1.0, which has no effect when percent African-American is 100%). The increase in magnitude as percent African-American decreases occurs for both racial groups, but is strongest for whites.

A Model of Residential Location Choice

Following equations 17 and 18, race specific proxy variables are created to describe each census tract in terms of the employment, educational, and personal opportunities offered within that tract. The proxy variables are constructed using only the estimated coefficients on the tract and tract interaction variables from the three dissatisfaction/negative outcome models. Finally, the three proxy variables are scaled to have a mean of zero and a standard error of one by racial group.[17] These variables are constructed from only a portion of the estimated model and the magnitude of the predicted values is not a meaningful indication of the relative level of satisfaction between racial groups. Rather than affect the level of satisfaction, the inclusion of the tract interaction variables create the possibility that neighborhoods offer different opportunities to white and African-American households and that the opportunities offered vary based on the racial composition of the neighborhood. The resulting proxies should only represent neighborhood variation and should not contain information arising from across household variation in the dissatisfaction measures because the first stage estimations included the observed household or prime age adult socio-economic attributes as control variables.

Following equation (19), a multinomial logit model of household residential location choice across census tracts is estimated using the following location attributes: the three preference proxies, percent African-American, price level, whether the household resides in the central city, and total number of households in each tract. Percent African-American is important to capture the effect of discrimination, steering and other imperfections in the housing market that are related to race. Naturally, price level is expected to affect the likelihood of residing in a location. Central city is included in part to provide a reduced-form proxy for variation across local governments since the data set does not allow the identification of individual towns or school districts. Total number of households in tract is included as a control for option size in order to minimize the effect of aggregation bias.

The variable means are shown in Table 6 for the entire sample in the first column as well as by race in the last two columns. These means are conducted over all residential locations and therefore represent the average neighborhood attribute values over the available housing stock. For the lack of opportunity proxies, the means by race have been normalized to zero, and a positive value for both the entire sample and the other race sample implies that the racial group prefers their current distribution of locations relative to the opportunities offered by the metropolitan area overall. The proxy variables predict that whites are sorted into preferred locations based on employment opportunities, but less preferred locations based on the quality of neighbors. Alternatively, African-Americans are sorted into less favorable locations on employment, but more favorable on opportunities for education and neighbors. African-Americans are heavily segregated by race, are more likely to reside in the center city, and live in lower quality locations that have a low price level. All of these variables except central city are scaled by the sample mean and standard deviation prior to using in the residential location choice model.

These location attributes are interacted with the following household characteristics: whether the head of household is married, whether the head is female, whether there are children in the family, whether the head is a high school graduate, whether the head is a college graduate, and the logarithm of family income. Separate models are estimated for white and African-American households. The logarithm of family income is also scaled by the mean and standard deviation of the sample.

The estimation results are shown in Table 7. The rows of Table 7 contain the coefficient estimates for all household variables interacted with a given location attribute. The columns contain the coefficient estimates for specific household variables where the column labeled level represents the affect of the attribute for individual representing the omitted category on all binary variables, i.e. single, childless, male without a high school education and earning average income. For white households, a higher percent African-American reduces the likelihood of residing in a location even after controlling for racial differences in preferences over employment opportunities, educational opportunities, and opportunities for interactions with neighbors, but this effect is much smaller for educated white households (negative level and positive high school and college coefficients for percent African-American). Poor employment opportunities reduce the likelihood of choosing a location for educated whites (insignificant level and negative high school and college coefficients). Poor educational opportunities reduces the likelihood of choosing a location especially for middle to lower income households and for households without children, which is counter-intuitive again raising the possibility that the first stage estimates suffer from a Tiebout-like bias (negative level and positive children and log-income coefficients). Poor quality neighbors reduce the likelihood of choosing a location for single non-high school graduates with low incomes (negative level and positive married, high school, college, and log income coefficients). White households are less-likely to reside in the central city overall, and this effect eliminated or even reversed for college educated, high income households (negative level and positive college and log income coefficients). White households are more likely to locate in higher priced locations, but this effect increases with education and income, which suggests that the higher price level captures unobserved amenities (positive level, high school, college, and log income coefficients).

For African-American households, percent African-American increases the likelihood of choosing a location even after controlling for racial differences in preferences (positive level coefficient). Poor employment opportunities decrease the likelihood that higher income African-American households reside in a location (insignificant level and negative log income coefficients), and poor quality in terms of neighbors reduces the likelihood of residing in a location for all African-American households (negative level coefficient). African-Americans without at least a high school education are more likely to reside in the central city (positive level and negative high school and college coefficients). On the price level variable, the sign pattern of the estimated coefficients for African-Americans is almost identical to the pattern for white households, but only the result for income is statistically significant. High income African-Americans are more likely to choose high price level locations (insignificant level and positive log income coefficients).

Simulation Analysis

Finally, the estimates from this model are used to conduct a simple simulation analysis in order to decompose the determinants of segregation. Following Ihlanfeldt and Scafidi (In Press), the exposure to African-American's for a household in the sample is calculated as a weighted average of the neighborhood percent African-American where the predicted likelihood of the household residing in that neighborhood is used as the weight, and the exposure for the sample is calculated as simply the average exposure over all households. These predicted exposure rates are shown in Table 8. The first column represents the sample average of predicted exposure based on the estimated model for white households (coefficient estimates in Panel A of Table 7). The second column represents the sample average based initially on the African-American model (Panel B of Table 7) and later on modifications to the African-American model so that African-American preferences over a given attribute are set equal to estimated values for white preferences. These modifications are simply accomplished by changing the values of African-American parameters for a given location attribute to the estimated values from the white model.[18] This simulation is based on the notion that the multinomial logit produces estimates that describe the preferences of households and accordingly equal parameter values represent similar preferences over an attribute. It is important to note that the rows for these simulations are cumulative so that each row is based on a change relative to the previous row.

The final column presents simulation results to represent a situation where the marginal effect of an attribute on white and African-American choices is set to the same value. The predictions in the second column treat the estimated parameters as estimates of actual household preferences as opposed to a simple description of the pattern of residential outcomes. The approach in the third column avoids the strong assumption that the estimated parameters represent preferences, and simply restricts the marginal effect of an attribute on location choice to be the same across the two races. The marginal effect of a change in the value of an attribute is

$$\frac{\partial Prob[k|j,i]}{\partial Z_{kj}} = \beta_{Zj} X_i Prob[k|j,i] \left(1 - Prob[k|j,i]\right) \qquad (20)$$

where the probability is calculated at the household's actual residential location *(k)* for each racial classification (*j*), Z_k is a the vector of location attributes, and X_i are the characteristics of household *i*, as used earlier in equations 14 through 19. Alternative values for the minority coefficient are calculated for each observation so that

$$\hat{\beta}_{ZB} X_i Prob[k|B,i] \left(1 - Prob[k|B,i]\right) = \hat{\beta}_{ZW} X_i Prob[k|W,i] \left(1 - Prob[k|W,i]\right) \qquad (19)$$

where $\hat{\beta}_{ZB}$ is the simulation value for the African-American parameter, and $\hat{\beta}_{ZW}$ is the estimated parameter value for white households. These simulations are also cumulative where except for the parameter being modified the simulation parameters at each stage are based on the parameters from the previous stage. In addition, in the marginal effect analysis, the probabilities for equating marginal effects are based on the estimated probabilities from the previous model or row in Table 8. As a result, these simulations may be path dependent, and a simulation based entirely on white preferences may not equal the probabilities predicted directly from the white model.[19]

The first row shows the predictions based on the estimated model for each sample separately, which equals the actual sample means. The racial difference in exposure is almost 60 percentage points. The second row predicts the mean exposure by race using the entire sample for each prediction in order to eliminate household endowment and demographic differences. This exercise reduces the racial difference by seventeen percentage points, or over 25 percent, mostly by reducing the exposure of African-Americans to predominantly African-American neighborhoods. The elimination of racial differences in preferences for employment opportunities, third row, further reduces the difference by another four percentage points for the preference simulation in column 2, but does not reduce differences for the marginal effects simulation in column 3. Controlling for preferences differences over educational opportunities, the fourth row, has a large effect on the segregation level of African-Americans reducing exposure by 12 and 27 percentage points for the preference and marginal effect simulations, respectively. No other simulations reduced racial differences in exposure except for the elimination of racial differences in the effect of racial composition on location. Note that the marginal effect simulation for racial composition generates a predicted exposure that is very close to white exposure levels, which suggests that path dependancy is not a major issue for evaluating the analysis that is based on equal marginal effects.[20]

Summary and Conclusions

This paper suggests new theoretical and empirical approaches for studying the causes of racial segregation. A model is proposed in which households value social interactions, rather than racial composition or other easily observable location attributes, such as percent poverty or average education level. Household bids over different locations are examined in a simple partial equilibrium theoretical framework. The model does not in general satisfy a single crossing property over racial composition. Therefore, unlike previous theoretical research on segregation, this model suggests that segregation may not be the dominant outcome when households sort across location based on racial composition.

The paper next specifies and estimates a multinomial choice model using data from the 1985 metropolitan sample of the American Housing Survey. This model is estimated in two stages. The first stage estimates models for unemployment among prime age adults and for household satisfaction with the education provided by local public schools and with the household's neighbors. The model demonstrates that the relationship between observed location variables and satisfaction or outcomes varies by the racial composition of the neighborhood and that this relationship varies by race. The estimated coefficients from these three models are then included in a multinomial logit model of residential location choice. Simulations based on these estimates indicate that endowment differences can explain 17 percentage points of the observed 60 point difference in exposure.

Additional simulations indicate that racial differences in preferences concerning education can explain between 12 and 27 additional percentage points of the racial differences depending on the simulation approach. For one simulation approach, differences in preferences concerning employment opportunities can explain an additional 4 percentage points, but this variable has little effect in the other simulation. For both simulation approaches, substantial racial differences in exposure to minority neighborhoods remain that cannot be explained by racial differences in preferences concerning employment, education, and neighbors. These remaining differences can only be explained by the direct effect of neighborhood racial composition. This analysis, however, cannot determine whether the effect of racial composition arises due to preferences concerning racial composition, constraints on the housing market choices of African-American households, or some combination of the two.

These results should be interpreted with care. First, the first stage estimations require the strong assumption that household residential location is exogenous to the dissatisfaction and outcome measures. Specifically, households only sort over the observed location attributes based the observable household characteristics that are included in the dissatisfaction models. Second, Philadelphia may not be representative. The city has a very high level of segregation and a history of racial tension. Finally, while these findings are suggestive of the potential link between discrimination and segregation, the analysis does not actually contain any information on discrimination. If this methodology was extended to include multiple cities and existing measures of housing discrimination, it might be possible to tell whether large residual racial differences in exposure arise predominantly in cities with high levels of adverse treatment against African-Americans.

References

Bailey, M. J. 1966. Effects of race and other demographic factors on the values of single-family homes. *Land Economics*, 42, 215-220.

Bayer, McMillan, and Rueben. 2002. The causes and consequences of residential segregation: An equilibrium analysis of racial sorting. Unpublished manuscript.

Benabou, Roland. 1993. Workings of a city: Location, education, and production. *Quarterly Journal of Economics*, 108, 619-652.

Borjas, George J. 1998. To ghetto or not to ghetto: Ethnicity and residential segregation. *Journal of Urban Economics*, 44, 228-253.

Brock, William A. and Steven N. Durlauf. In press. Interaction-based models, in *Handbook of Econometrics Vol. 5* (J. Heckman and E. Leamer, Eds) Amsterdam: North-Holland.

Chambers, Daniel N. 1992. The racial housing price differential and racially transitional neighborhoods. *Journal of Urban Economics*, 32, 214-232.

Courant, Paul N. 1978. Racial prejudice in a search model of the urban housing market. *Journal of Urban Economics*, 5, 329-345.

Cutler, David M. and Edward L. Glaeser. 1997. Are ghettos good or bad? *Quarterly Journal of Economics*, 112, 827-872.

Cutler, David M., Edward L. Glaeser, and Jacob L. Vigdor. 1999. The rise and decline of the American Ghetto. *Journal of Political Economy*, 107, 455-506.

DeBartolome, Charles A.M. 1990. Equilibrium and Inefficiency in a community with peer group effects. *Journal of Political Economy*, 98, 110-133.

Deng, Yongheng, Stephen L. Ross, and Susan M. Wachter. 2001. Racial differences in homeownership: The effect of residential location. Unpublished manuscript.

DeRango, Kelly. 1999. The effect of inter-racial economic inequality on racial segregation re-examined. Unpublished manuscript.

Follain, James R. and Stephen Malpezzi. 1981. Another look at racial differences in housing prices. *Urban Studies*, 18, 195-203.

Gabriel, Stuart and Stuart S. Rosenthal. 1989. Household location and race: Estimates of a multinomial logit model. *Review of Economics and Statistics*, 71, 240-249.

Ihlanfeldt, Keith R. and Benjamin P. Scafidi. 2002. Black Self-Segregation as a Cause of Neighborhood Racial Segregation: Evidence from the Multi-City Study of Urban Inequality. *Journal of Urban Economics*, 51, 366-390.

Kern, Clifford R. 1981. Racial prejudice and residential segregation: The Yinger model revisited. *Journal of Urban Economics*, 10, 164-172.

Kiel, Katherine A. and Jeffrey E. Zabel. 1996. House price differentials in U.S. cities: Household and neighborhood racial effects. *Journal of Housing Economics*, 5, 143-165.

King, A. Thomas and Peter Mieszkowski. 1973. Racial discrimination, segregation, and the price of housing. *Journal of Political Economy*, 81, 590-606.

Lang, Kevin. 1986. A language theory of discrimination. Quarterly Journal of Economics, 101, 363-382.

Lundberg, Shelly and Richard Startz. 1998. Race, Information, and Segregation. Unpublished manuscript.

Manski, Charles F. In press. Economic analysis of social interactions. *Journal of Economic Perspectives*.

Massey, Douglass S. and Nancy A. Denton. 1993. *American Apartheid: Segregation and the Making of the Underclass*. Cambridge, MA: Harvard University Press

Schafer, Robert. 1979. Racial discrimination in the Boston housing market. *Journal of Urban Economics*, 6, 1176-1196.

Schnare, Ann B. 1976. Racial and ethnic price differentials in an urban housing market. *Urban Studies*, 13, 107-120.

Schnare, Ann B. and C. Duncan MacRae. 1978. The dynamics of neighborhood change. *Urban Studies*, 15, 327-301.

Schelling, Thomas C. 1972. A process of residential segregation: Neighborhood tipping, in *Racial Discrimination in Economic Life* (A. H. Pascal Ed.). Lexington, MA: Lexington Books, D.C. Heath & Co.

Paul Waddell. 1992. A multinomial logit model of race and urban structure. *Urban Geography*, 13, 120-141.

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

Yinger, John. 1986. Measuring racial discrimination with fair housing audits. *American Economic Review*, 76, 881-893.

Yinger, John. 1978. The Black-White price differential in housing: Some further evidence. *Land Economics*, 54, 187-206.

Yinger, John. 1976. Racial prejudice and racial residential segregation in an urban model. *Journal of Urban Economics*, 3, 383-396.

Zenou, Yves and Nicolas Boccard. 2000. Racial discrimination and redlining in cities. *Journal of Urban Economics*, 48, 260-285.

Table 1: Means of Household Sample			
Variable Names	Variable Descriptions	White[1]	African-American
Problem with Public Schools	One if dissatisfied with public schools or children do not attend public schools, zero otherwise	0.392 (0.488)	0.318 (0.466)
Problem with Neighbors	One if bothered by people in neighborhood, zero otherwise	0.151 (0.358)	0.241 (0.428)
Head Married	One if head of household married, zero single	0.554 (0.497)	0.308 (0.462)
Head female	One if head female, zero if male	0.315 (0.464)	0.541 (0.498)
Head Years in Labor Force	Age of head minus the sum of six and the years of education	30.340 (9.092)	29.470 (7.671)
High School Graduate	One if head has between 12 and 15 years of education, zero otherwise	0.503 (0.500)	0.541 (0.498)
College Graduate	One if head has 16 or more years of education, zero otherwise.	0.267 (0.442)	0.100 (0.300)
Family Income	Logarithm of family income	9.994 (0.840)	9.430 (0.870)
Retired	One if household contains individuals who are older than 65 years of age, zero otherwise	0.239 (0.426)	0.167 (0.373)
Children	One if the household contains children or other dependents of the respondent, zero otherwise	0.349 (0.476)	0.445 (0.497)
Kids Less Than Six	One if the household contains children younger than six years old, zero otherwise	0.156 (0.363)	0.201 (0.401)
Number of Adults	Number of adults in the household	0.935 (0.904)	0.819 (0.918)
Sample Size		3149	788

1. Standard errors are shown in parentheses for all tables in the paper.

Table 2: Means of Prime Age Adult Sample			
Variable Names	Variable Descriptions	White	African-American
Employment Status	One if the adult is not working, zero otherwise	0.302 (0.459)	0.441 (0.497)
Marital Status	One if adult married, zero single	0.682 (0.465)	0.394 (0.488)
Female	One if adult female, zero if male	0.513 (0.499)	0.581 (0.493)
Years in Labor Force	Age of adult minus the sum of six and the years of education	20.501 (11.457)	20.750 (11.424)
High School Graduate	One if adult has between 12 and 15 years of education, zero otherwise	0.584 (0.492)	0.631 (0.482)
College Graduate	One if adult has 16 or more years of education, zero otherwise.	0.282 (0.450)	0.134 (0.341)
Retired	One if household contains individuals who are older than 65 years of age, zero otherwise	0.079 (0.317)	0.087 (0.322)
Children	One if the household contains children or other dependents of the respondent, zero otherwise	0.934 (1.153)	1.091 (1.354)
Kids Less Than Six	One if the household contains children younger than six years old, zero otherwise	0.299 (0.628)	0.360 (0.711)
Number of Adults	Number of adults in the household	2.245 (1.035)	2.080 (1.025)
Sample Size		3969	994

Table 3: Means of Census Tract Sample				
Variable Name	Variable Description	Percent African-American		
		0.00	0.00-0.25	0.26-1.00
Percent African-American	Percent of AHS households in tract with African-American head	0.000 (0.000)	0.146 (0.047)	0.720 (0.263)
Average Income	Mean family income ($1000's) of AHS households in tract	29.218 (13.570)	23.306 (11.325)	15.893 (7.731)
Average Education	Mean education level of adults in AHS	12.531 (1.341)	12.636 (1.519)	11.592 (1.206)
Central City	One if tract in central city, zero otherwise	0.275 (0.447)	0.347 (0.478)	0.758 (0.429)
Percent Near Commercial	Percent AHS households reporting near commercial/industrial property	0.232 (0.252)	0.302 (0.289)	0.396 (0.265)
Percent Near Green Space	Percent AHS households reporting near green space	0.236 (0.280)	0.219 (0.202)	0.161 (0.189)
Employment Access	Predicted access for tract based on estimated gravity model	21.452 (7.425)	22.349 (9.924)	18.726 (6.675)
Labor Market Quality	Average wage fixed effects based on employment location of adults	8.911 (0.212)	8.940 (0.223)	8.999 (0.184)
Price Level	Tract fixed effect from hedonic price regression	9.368 (0.367)	9.253 (0.439)	8.706 (0.528)
Number of Households	Total number of households in tract	7.802 (2.856)	9.228 (3.241)	8.050 (2.598)
Sample Size		283	92	120

Table 4: Loglikelihood Values for Outcome and Satisfaction Models						
Model	Employment		Education		Neighbors	
	Loglikelihood	#[1]	Loglikelihood	#	Loglikelihood	#
1. Baseline	-2,702.4	27	-778.3	26	-1,736.2	28
2. Tract Attributes	-2,681.7	36	-728.5	35	-1705.5	37
Difference	20.7	9	49.8	9	30.7	9
3. Interact Racial Comp	-2,654.0	44	-723.5	43	-1,695.1	45
Difference	27.7	8	5.0	8	10.4	8
4. Interact Comp & Race	-2,645.1	52	-715.2	51	-1,685.8	53
Difference	8.9	8	8.3	8	9.3	8
5. Complete Interaction	-2,641.8	61	714.5	60	-1,682.4	62
Difference	3.3	9	0.7	9	3.4	9

1. The number sign represents the number of parameters. When the difference in parameters is eight, Chi-square statistics of 15.5 and 20.1 are statistically significant at the 0.05 and 0.01 level, respectively. The corresponding statistics for nine parameters are 16.9 and 21.7. These thresholds are based on the critical statistic from a Chi-square distribution with eight and nine degrees of freedom, respectively.

Table 5: Parameter Estimates for Outcome and Satisfaction Models[1]

Variable Names	Employment		Education		Neighbors	
	Level	Race	Level	Race	Level	Race
Intercept	-1.124 (1.174)	0.049 (0.283)	4.336 (3.439)	-0.483 (1.456)	13.358 (4.090)	-0.258 (0.870)
Marital Status	-0.361 (0.083)	-0.591 (0.117)	-0.032 (0.458)	0.105 (0.342)	0.046 (0.224)	0.084 (0.188)
Gender	0.054 (0.075)	-0.147 (0.106)	-0.031 (0.176)	0.219 (0.314)	- 0.115 (0.079)	-0.286 (0.158)
Years in Labor Force divided by 10	-0.147 (0.090)	-0.013 (0.019)	0.196 (0.270)	0.054 (0.355)	-0.052 (0.079)	-0.048 (0.162)
Years Squared	0.050 (0.019)	0.000 (0.000)	-0.004 (0.044)	-0.018 (0.064)	0.015 (0.011)	0.015 (0.026)
Years * Marital Status			0.131 (0.306)		0.084 (0.119)	
Years Squared * Marital Status			-0.040 (0.055)		-0.024 (0.018)	
High School Graduate	-0.406 (0.069)	-0.118 (0.138)	-0.199 (0.219)	0.081 (0.291)	0.153 (0.112)	-0.212 (0.169)
College Graduate	-0.640 (0.085)	-0.193 (0.197)	-0.190 (0.315)	-0.805 (0.45)	-0.059 (0.141)	-0.259 (0.268)
High School Graduate * Marital Status			-0.240 (0.262)		-0.171 (0.156)	
College Graduate * Marital Status			-0.474 (0.344)		-0.200 (0.185)	
Family Income			-0.090 (0.082)	0.061 (0.156)	0.012 (0.046)	0.053 (0.092)
Retired	0.122 (0.067)	0.433 (0.152)	-0.237 (0.233)	0.369 (0.604)	0.172 (0.090)	-0.290 (0.199)
Children	0.105 (0.048)	0.170 (0.057)			-0.140 (0.077)	-0.002 (0.152)
Kids Less Than Six	0.242 (0.102)	-0.217 (0.105)	-0.083 (0.064)	-0.143 (0.117)	0.060 (0.063)	-0.104 (0.111)
Number of Adults	0.016 (0.022)	0.173 (0.051)	-0.093 (0.067)	-0.057 (0.128)	-0.008 (0.039)	-0.003 (0.077)
Children * Marital Status	-0.110 (0.059)					
Kids Less Than Six * Marital Status	-0.213 (0.115)					
Female * Marital Status	0.727 (0.102)					
Children * Female * Marital Status	0.104 (0.050)					
Kids Less Than Six * Marital Status * Female	0.287 (0.087)					

Table 5 (cont.): Parameter Estimates for Outcome and Satisfaction Models

Variable Names	Employment		Education		Neighbors	
	Level	Race	Level	Race	Level	Race
Percent African-American[2]	6.273 (4.135)		-2.045 (7.621)		14.749 (4.622)	
Average Income	0.003 (0.002)		0.009 (0.011)		0.014 (0.011)	
Average Education	0.041 (0.024)		-0.011 (0.087)		0.014 (0.066)	
Central City	0.164 (0.063)		-0.581 (0.217)		0.514 (0.215)	
Percent Near Commercial	-0.042 (0.114)		0.538 (0.382)		-0.236 (0.278)	
Percent Near Green Space	-0.112 (0.093)		1.120 (0.461)		0.161 (0.345)	
Employment Access	-0.011 (0.002)		-0.003 (0.012)		-0.008 (0.010)	
Labor Market Quality	0.040 (0.120)		-0.418 (0.389)		-1.559 (0.439)	
Price Level	0.013 (0.076)		0.077 (0.175)		0.034 (0.162)	
Average Income*Percent African-American	-0.000 (0.022)	-0.069 (0.025)	0.021 (0.025)	-0.022 (0.045)	0.009 (0.012)	0.042 (0.017)
Average Education*Percent African-American	-0.315 (0.148)	0.286 (0.156)	-0.150 (0.207)	0.290 (0.388)	0.006 (0.075)	-0.025 (0.115)
Central City*Percent African-American	0.193 (0.369)	-0.492 (0.408)	0.725 (0.524)	0.497 (0.927)	0.724 (0.242)	-0.361 (0.303)
Percent Near Commercial* Percent African-American	0.498 (0.705)	-0.771 (0.722)	2.060 (0.931)	-4.043 (1.737)	-0.082 (0.328)	1.278 (0.555)
Percent Near Green Space* Percent African-American	0.427 (0.900)	-0.046 (0.949)	2.320 (1.027)	-2.435 (1.866)	0.184 (0.378)	-0.624 (0.738)
Employment Access* Percent African-American	0.002 (0.021)	-0.004 (0.022)	-0.016 (0.027)	0.044 (0.052)	-0.009 (0.011)	-0.002 (0.016)
Labor Market Quality* Percent African-American	0.121 (0.531)	-0.790 (0.341)	-0.138 (0.865)	0.577 (0.775)	-1.760 (0.493)	0.916 (0.288)
Price Level*Percent African-American	-0.425 (0.330)	0.614 (0.339)	0.472 (0.484)	-0.937 (0.815)	0.002 (0.202)	-0.984 (0.312)

1. Standard errors are shown in parentheses.
2. The percent minority variable takes a value of zero when actual percent minority is zero, fourty-five, and one hundred percent for the employment, location, and neighbor specifications, respectively.

Table 6: Mean Location Attributes for Household/Housing Unit Sample			
Variable Name	Total Sample	White Sample	African-American Sample
Employment: White Pref.	0.494 (1.823)	0.000 (1.000)	2.489 (2.786)
Education: White Pref.	0.007 (1.026)	0.000 (1.000)	0.078 (1.107)
Neighbor: White Pref	-0.416 (1.385)	0.000 (1.000)	-2.099 (1.448)
Employment: Black Pref.	-0.356 (0.613)	-0.444 (0.427)	0.000 (1.000)
Education: Black Pref.	1.259 (1.457)	1.507 (1.382)	0.000 (1.000)
Neighbor: Black Pref	0.976 (1.326)	1.217 (1.285)	0.000 (1.000)
Percent African-American	0.198 (0.321)	0.080 (0.167)	0.674 (0.351)
Central City Location	0.419 (0.493)	0.330 (0.470)	0.777 (0.415)
Price Level	9.172 (3.256)	9.243 (3.345)	8.885 (2.851)
Number of Households	9.177 (0.496)	9.296 (0.401)	8.695 (0.551)

Table 7: Parameter Estimates from Residential Location Choice Model

Panel A: White Parameter Estimates

Variable Names	Level	Married	Female	High School	College	Children	Log Income
Employment Opportunities Lacking	0.026 (0.067)	-0.027 (0.069)	-0.006 (0.067)	-0.175 (0.064)	-0.125 (0.076)	-0.068 (0.054)	-0.015 (0.031)
Poor Quality Public Education	-0.313 (0.145)	0.015 (0.154)	0.087 (0.151)	0.031 (0.139)	0.218 (0.167)	0.343 (0.120)	0.346 (0.069)
Poor Quality Neighbors	-0.464 (0.126)	0.354 (0.134)	-0.207 (0.131)	0.271 (0.121)	0.355 (0.145)	0.131 (0.102)	0.292 (0.060)
Percent African-American	-0.924 (0.178)	0.299 (0.187)	-0.361 (0.182)	0.471 (0.170)	0.649 (0.203)	-0.020 (0.144)	0.004 (0.081)
Central City	-0.529 (0.285)	0.193 (0.299)	-0.020 (0.295)	0.067 (0.271)	0.612 (0.326)	0.407 (0.236)	0.682 (0.137)
Price Level	0.159 (0.075)	-0.178 (0.082)	-0.094 (0.081)	0.141 (0.069)	0.632 (0.090)	-0.394 (0.061)	0.090 (0.036)
Number of Households	0.429 (0.052)	-0.145 (0.053)	0.009 (0.053)	-0.111 (0.049)	-0.066 (0.058)	0.000 (0.041)	0.022 (0.025)

Panel B: African-American Parameter Estimates

Variable Names	Level	Married	Female	High School	College	Children	Log Income
Employment Opportunities Lacking	-0.178 (0.140)	-0.014 (0.153)	-0.151 (0.132)	0.040 (0.123)	-0.224 (0.216)	0.092 (0.110)	-0.353 (0.070)
Poor Quality Public Education	-0.124 (0.184)	0.003 (0.208)	-0.218 (0.184)	-0.287 (0.168)	-0.231 (0.266)	0.084 (0.155)	0.021 (0.096)
Poor Quality Neighbors	-0.355 (0.140)	0.074 (0.158)	0.062 (0.142)	0.045 (0.132)	0.245 (0.197)	0.026 (0.120)	-0.043 (0.070)
Percent African-American	0.773 (0.146)	0.222 (0.163)	-0.242 (0.144)	0.100 (0.132)	-0.295 (0.213)	-0.054 (0.123)	-0.001 (0.075)
Central City	0.970 (0.430)	-0.755 (0.485)	-0.326 (0.429)	-1.065 (0.387)	-0.977 (0.639)	-0.118 (0.359)	-0.063 (0.226)
Price Level	0.244 (0.146)	-0.143 (0.164)	-0.171 (0.148)	0.182 (0.129)	0.322 (0.242)	-0.169 (0.124)	0.252 (0.076)
Number of Households	0.491 (0.116)	-0.036 (0.129)	-0.079 (0.118)	-0.205 (0.103)	-0.085 (0.159)	0.024 (0.097)	0.001 (0.060)

Table 8: Neighborhood Exposure to African-Americans			
Simulations	White Preferences	African-American Preferences	African-American Marginal Effects
Race Specific Sample	0.080	0.675	0.675
Same Observables[2]	0.092	0.508	0.508
Same Employment Model[3]	0.092	0.468	0.498
Same School Model	0.092	0.345	0.227
Same Neighbors Model	0.092	0.360	0.235
Same Central City Model	0.092	0.389	0.288
Same Price Level Model	0.092	0.407	0.296
Same Racial Composition Model[4]	0.092	0.092	0.071

1. The first row shows the predicted exposure rates for the white and African-American samples using the appropriate for each race.
2. The second row shows the predictions for the entire sample given estimated preferences for whites and African-Americans.
3. The next six rows eliminate racial differences in preferences over each location choice variable. The values in the third column eliminate racial differences in the marginal effect of the choice variable. White preferences are held fixed.
4. This simulation also eliminates any difference arising from number of households in tract, which as stated previously was included only to mitigate aggregation bias.

Endnotes

1. A growing literature examines the microeconomic and econometric foundations of social interactions, see Manski (In press) and Brock and Durlauf (In press)

2. Yinger (1976) claims to show that complete racial segregation is never a stable equilibrium in a non-discriminatory, competitive market, but Kern (1981) proves that segregation is a stable equilibrium for Yinger's model.

3. See Zenou and Boccard (2000) for a model that links segregation more directly to outcomes in the labor market.

4. Cutler et. al. also finds that the white price premium in the 1990's is highest in the most segregated cities. This additional finding is consistent with their decentralized racism hypothesis that whites pay a premium to live in segregated neighborhoods, but it is also consistent with African-Americans being segregated and steered into the worst neighborhoods in the metropolitan areas that are most highly segregated.

5. Racial differences in housing prices track very closely with the influx of southern, rural blacks into major U.S. cities, and the decline of the African-American housing price premium followed the conclusion of this great migration (Cutler et. al.). The observed housing price premium are as much a feature of this period as they are an indication of discrimination. As Cutler et. al. discuss, ghetto's can operate as a mechanism to help groups assimilate into new environments. If housing markets adjust slowly relative to the speed of migration, price spikes will arise in these ghettos whether or not housing discrimination is practiced. Similarly, as migration slows, discrimination is likely to force the new minority group to expand into specific regions of a city as opposed to preventing expansion, which is required for the price premium to persist. For example, Yinger (1995, p. 123) discusses the role played by real estate agents in the creation of new predominantly African-American neighborhoods in the Mattapan neighborhood of Boston, MA.

6. Formally, the estimates from the multinomial logit describe the preferences of households, but the estimates are biased whenever some options within the choice set are actually not available to some households (Maddala, 1984). If housing discrimination or racial steering effectively removes some neighborhoods from the choice set of African-American households, the estimates of preferences will be biased.

7. In this model, a household will chose to interact with every household encountered because the time search costs are sunk. If social interactions involve both interaction time and search time costs, an individual might choose not to interact with either their own or the other group. Moreover, if they can observe quality or a signal for quality, they may choose to interact with some members of their own group, but not all. In an imperfect information model, this selection process may create adverse selection problems, see Lundberg and Startz (1998).

8. An added complexity arises if one considers the possibility that an individual's quality in terms of social interaction depends on the quality of social interactions experienced. In this case, group quality is endogenous and depends upon the interaction between the decisions of both groups, as well as the equilibrium distribution of groups across communities. Multiple equilibria are likely to exist in such a model.

9. The commuting time supplement has not been administered as part of the metropolitan sample of the AHS since 1985, and therefore no information on work location, commuting, or mode choice is available for the AHS after that year. The information collected in the neighborhood supplement also has been dramatically reduced since 1988.

10. The education questions are only asked if the households report having school age children leading to a sample of 1223 respondents. The satisfaction questions do not refer explicitly to public schools and the structure implies that households with children attending private school most likely respond concerning their level of satisfaction with the schools that their children actually attend. Therefore, for households with children in private school, dissatisfaction must be inferred by the decision to opt out of the public school system.

11. The survey also asked households about dissatisfaction with police services and being bothered by crime. Neither of these variables provide a substantial amount of additional information over and above the information provided by the education and neighbors variables.

12. The approach used in this paper differs from the standard gravity model because the data on flows in the 1985 metro AHS is quite thin when considered at the tract level. In the standard gravity model, the sample is based on residential and work locations for which flows between these locations are observed because commute time is unobserved when there are no commuters traveling between the locations. In this paper, all possible residential and work locations are included in the sample and the commute time between the residential and employment zones is used as a proxy for the tract to employment zone commute time for routes that are not traveled by commuters in the AHS sample. Note that the final job access measure is based on a log-log specification in which the logarithm of one plus the number of flows is taken, but alternative flow models based on an ordered probit or a poisson regression yield very similar results.

13. The wage equation is corrected for sample selection into employment using a standard two-stage approach. A probit model of employment is estimated controlling for the standard individual characteristics plus residential location fixed effects at the census tract level. The selection correction model is identified by the assumption that residential location does not directly influence earnings after controlling for observed individual characteristics and employment location.

14. The housing price regression is corrected for sample selection using the results of a first stage probit that estimates the likelihood that a household will reside in either rental or owner-occupied housing. This model is identified by the standard assumption that household demographics influence tenure choice, but do not directly influence the equilibrium price of a housing unit after controlling for the attributes of the housing. If a tract does not contain any owner-occupied housing or any rental housing, the tract fixed effect is imputed using the fixed effect from a hedonic price equation for annual rent or house value, respectively.

15. An alternative approach is to estimate an interaction between the basic location attributes and race prior to interacting race with the interactions between racial composition and the basic location attributes. This alternative model would then replace model four. When this specification is used, model three cannot be rejected in favor of model four. Moreover, a non-nested hypothesis test supports the current model four over this alternative.

16. The estimated coefficients on the interaction of race with the gender or marital status interaction terms are statistically insignificant and very close to zero in models that incorporate these additional interaction variables. Also note that the children variable is not identified for

the education dissatisfaction measure because households only respond to the education questions if they have school aged children.

17. All later results in the paper (Tables 7 and 8) are replicated when these proxies are simply scaled by mean and standard deviation of the entire sample rather than using the subsample values.

18. The simulations concerning preferences for employment, education, and quality neighbor opportunities are conducted by eliminating racial differences in the coefficients that predict residential location, as well as the coefficients that are used to generate the opportunity proxies.

19. The denominator of the simulation coefficient contains the African-American probability associated with the household's residential location. In practice, this denominator can be quite small resulting in very large marginal effects for some observations and causing the procedure to break down after the simulation procedure has been performed for a couple of location attributes. In order to avoid the impact of outliers, the predicted probabilities were restricted to fall between 0.01 and 0.99 and were set to these limits when they fell outside of those ranges. These restrictions were not imposed when predicting the exposure levels presented, and moreover the simulations for the first two location attributes provide very similar results when this restriction is not imposed.

20. The simulation labeled Same Racial Composition Model also eliminates any difference arising from number of households in tract, which as stated previously was include only to mitigate aggregation bias.