# Bridging simulation and inference methods in the study of residential segregation: a case study

Miruna Petrescu-Prahova*

April 2, 2008

**Abstract**

In this paper we conduct a two-stage analysis of residential segregation in the Yuma, AZ, metropolitan area. First, we estimate the contribution of effects such as homophily, xenophobia or push-pull between rent and income levels in determining the observed distribution of households across tracts. Second, we simulate a scenario with covariates from Census data, where we vary the parameter values to analyze the impact of various effects on segregation patterns, and obtain a range of possible Yuma residential configurations. Exploring the range of alternate residential pattern configurations as well as the observed configuration allows us to understand how even slight changes in some parameters could affect the overall assignment of households to tracts and therefore segregation levels in this area. This combination of simulation and inference gives us a powerful tool for understanding residential segregation processes, while at the same time suggesting implications for public policy aimed at reducing segregation.

## 1 Introduction

Ethnic residential segregation has been a visible and salient aspect of urban life in the U.S., especially after the country experienced massive waves of immigration during

*Department of Sociology, University of California, Irvine, 3151 Social Science Plaza A, University of California, Irvine, CA 92697; mpetresc@uci.edu

the 19th and early 20th century. Empirical studies conducted at the beginning of the 20th century noted the existence of ethnic neighborhoods in metropolitan areas with large immigrant populations, such as Chicago and New York (Thomas, 1921). Post-1965 immigrants, although hailing from different origins than their predecessors, have exhibited the same tendency to form ethnic communities in which institutions and services are tailored to the characteristic needs of the ethnic groups (Zhou, 1992; Foner, 2000; Waldinger, 2001). Apart from ethnic neighborhoods that formed as a result of immigration, cities in the US are home to a large African American population, which is, and has consistently been, residentially segregated from the native-born white population (Taeuber and Taeuber, 1965; Massey and Denton, 1993; Gottdiener and Hutchinson, 2000).

Previous residential segregation studies have sought to identify the factors that determine residential settlement patterns (Clark, 1992; Zubrinsky Charles, 2001; Wilson and Hammer, 2001; Alba and Nee, 2003). They suggest several factors, which can be classified into three main categories: physical characteristics of the urban environment, individual and aggregate socioeconomic characteristics, and individual preferences for neighborhood composition. Important strides have recently been made in the direction of studying the interactions among these factors by researchers using agent-based and cellular automata models (Epstein and Axtell, 1996; Mare and Bruch, 2003; Benenson, 2004; Zhang, 2004; Fossett, 2006), based on early work by Schelling (1969, 1971) and Sakoda (1971). However, these results have largely been based on simulations of "toy" worlds, and efforts to extend the analyses to real cases have been hampered by a lack of inferential tools to connect theoretical models with extant data.

Using new models which bridge this "inferential gap" (Butts, 2007; Petrescu-Prahova, 2007), we conduct a two-stage analysis of a real case, the Yuma, AZ, metropolitan area. In the first stage, we use real covariates from U.S. Census data to infer the value of the parameters in the Yuma, AZ, metropolitan area, thus estimating the relative contribution of effects such as homophily, xenophobia or push-pull between rent and income levels in determining the *observed* distribution of households across areal units. In the second stage, we simulate a simple scenario with real, fixed covariates and real geographic information about census tract placement and boundaries from U.S. Census data, in which we vary the parameter values to analyze the impact of various effects on residential patterns. In doing so, we obtain a range of *possible* Yuma residential pattern configurations.

The choice of the Yuma metropolitan area for this study is motivated by theoretical and methodological considerations. First, the population consists mainly of two ethnic groups, Hispanics and non-Hispanic whites (37% of households have a Hispanic householder, and 58% have a non-Hispanic White householder). The index of dissimilarity (D) for Yuma, calculated at the census tract level based on Census 2000 data is 0.47 [1]. Second, Yuma is a relatively small metropolitan area (it contains 53,848 households), and this reduces the computational challenges associated with our analyses. Also, the presence of mainly two ethnic groups in this area facilitates the comparison of the re-

[1] This is the value of D for households, since the unit of analysis in our study is the household; we also calculated D at the individual level, and obtained a value of 0.46.

sults of this analysis with previous simulation studies conducted on a toy scenario with two ethnic groups, using the same methods (Petrescu-Prahova, 2007).

## 2 Potential determinants of residential segregation in US metropolitan areas

Modern cities have certain man-made features, which are intrinsic to their structure and to some extent independent of their resident population, as well as natural features, all of which may be conducive to certain patterns of land use (McKenzie, 1924, Hawley, 1950). Fixed infrastructure (e.g., roads, factories), the spatial distribution of land available for residential use (as opposed to economic use), and the number of housing units, combined with natural barriers such as rivers or hills can influence settlement patterns, since locations which present spatially isolated clusters of housing units may be more prone to segregation than locations with minimal barriers between units.[2] In the same vein, Grannis (1998) shows that neighborhoods connected by tertiary streets (i.e., streets with one lane in each direction and no center divider) are more likely to be similar in terms of ethnic composition than spatially adjacent neighborhoods that are separated by non-tertiary streets. Foner (2000) notes that in the early years of the Jewish and Italian influx into New York, most immigrants settled in the downtown neighborhoods situated below Fourteenth street, which ensured that they were living close to the sources of jobs – docks, warehouses, factories, and business streets (p. 39). They were able to move out of these neighborhoods only after the infrastructure

[2]One of the expressions through which the urban vernacular has captured this situation is "the wrong side of the tracks", which reflects the fact that the borders of segregated neighborhoods are determined by such barriers as railroad tracks (Massey and Denton, 1993).

of public transportation, roads and bridges eased the access to new destinations such as Harlem, Brooklyn and Queens. However, even in the extremely densely populated area below Fourteenth street, Italians and Jews were rarely close neighbors. The grid structure of the streets provided the barriers, and "most blocks were heavily dominated, if not exclusively populated, by one or the other immigrant group" (Foner, 2000, p. 41).

Another set of factors are individual and aggregate socioeconomic characteristics, especially personal income and rent levels. The relationship between rent and personal income is a hard constraint on residential choice, especially for low-income households. As a consequence, households with comparable incomes seek locations with similar and affordable rent levels and consequently cluster together in certain parts of the metropolis (Hawley, 1950). If, in addition, we take into account the fact that poverty disproportionately affects members of minority ethnic groups, we have the premises of ethnic residential segregation through income levels alone (Clark, 1986; Gottdiener and Hutchinson, 2000). On the other hand, settlement patterns of ethnic groups in urban areas are determined partly by social networks of kinship, friendship, and co-ethnicity. To a large extent, these networks offer support to new immigrants, who are unfamiliar with American society and frequently lacking proficiency in English. This leads to geographic concentration of ethnic or even national origin groups (Thomas, 1921; MacDonald and MacDonald, 1970; Massey et al. 1998; Menjívar, 2000). Researchers have generally emphasized the positive aspects of social networks. However, in her study of Salvadoran immigrants in the Bay Area, Menjívar (2000) notes that networks are prone to breaking down and brings to attention several aspects that have been overlooked in the study of immigrant social networks. Immigrant networks are dynamic structures that are affected by larger social and economic forces, and so the social ties that immigrants use can weaken at the point of destination. On the other hand, the same networks that are supportive at some point can be sources of conflict at other points, especially in times of economic hardship, when support needed by recent immigrants can be experienced as a substantial burden by earlier immigrants. She also points out that common background is not enough for the development and maintenance of co-ethnic relationships. Groups may have internal divisions along social class or political ideology that prevent supportive interactions. This phenomenon is not restricted to immigrants, however; human geography studies suggest that internal migrants also make settlement decisions based on the geographic location of friends and relatives (Clark, 1986a).

One of the most influential theories for the interpretation of ethnic population distribution across metropolitan space is the spatial assimilation framework, developed by Massey (1985), on the basis of earlier work by Robert Park and Louis Wirth. According to this framework, which is related to the normative view of immigrant assimilation in the host societies (as presented by Gordon, 1964), immigrant groups initially settle in enclaves located in the inner city, mainly in economically disadvantaged areas. As their members experience social mobility and acculturation, they usually leave these areas and move to "better" neighborhoods, namely areas that do not have such a high concentration of ethnic minorities, leading to a reduction in residential segregation levels. This hypothesis is supported in the case of Hispanics, whose levels of segregation from non-Hispanic whites decline with generation spent in the U.S. (Massey, 1979).

The underlying assumptions of the framework are that neighborhood location and housing are largely determined by market processes and that individuals are motivated to improve their residential status once they have acculturated and made some socioeconomic gains. In this context, residential exposure to the majority group is hypothesized to improve as a result of gains in an ethnic family's socioeconomic standing, acculturation (as measured, for instance, by its members' proficiency in speaking English), and generational status or, in the case of first generation immigrants, length of residence in the country of destination. Residence in the suburbs is also taken into account in the model because it is seen as a sign of enhanced residential assimilation. A series of studies of spatial assimilation for some of the main metropolitan regions, summarized by Alba and Nee (2003), focus especially on the median household income of the census tract of residence and the percent of non-Hispanic whites, the majority group, among residents, as indicators of spatial assimilation. For Hispanics, they find that the most powerful determinant of living in a high income, high percent white neighborhood is their own socioeconomic position: the greater their income and the higher their educational status, the larger, for instance, the percentage of non-Hispanic whites in the population of the neighborhood where they reside (Alba and Nee, 2003).

The spatial assimilation framework does not apply, however, to African American communities and to immigrant groups that have mixed African ancestry (Haitians, West Indians), because of racial discrimination by the white population (Massey and Denton, 1993). Apart from this shortcoming, the spatial assimilation model, which was built primarily on the experience of the mainly Southern and Eastern European immigrant flows in the early 20th century, also does not fully account for the experience of new immigrant groups. The achievement of social mobility is no longer linked with the exit from the ethnic community – especially for those groups that have financial capital when they arrive in the U.S. – and remaining in the ethnic community represents a choice rather than a constraint for members of some national-origin groups such as Cubans (South, Crowder and Chavez, 2005).

This result is related to the final set of factors suggested by previous research as a potential determinant of ethnic residential segregation: individual preferences for neighborhood composition (Clark, 1992; Zubrinsky Charles, 2001), which can vary according to the reference combination of ethnic groups. One of the first factors emphasized is the preference for homogeneity, which can be understood either as a desire to be close to co-ethnics (homophily), or a desire to be apart from ethnic "others" (xenophobia). This type of preference is mostly exhibited by the non-Hispanic white population, who prefer neighborhoods that are 70% or more white, when viewed as combinations of non-Hispanic white and black households (Clark, 1992). In contrast, blacks appear to want a sizable population of coethnics and substantial integration at the same time, leading to a preference for 50/50 neighborhoods (Zubrinsky Charles, 2001). Hispanics tend to approximate the preferences of blacks, when the reference composition is Hispanic/non-Hispanic white, but approach a preference for neighborhoods that are 75% Hispanic when the potential neighbors are black. In turn, Asian respondents are much more open to integration with non-Hispanic whites than with other groups and find integration with blacks least appealing, while at the same time showing strong preferences for co-ethnic neighbors (Zubrinsky Charles, 2001).

Within the Hispanic population, researches have also observed differences based on race. Denton and Massey (1989) study Caribbean Hispanics, who have a mix of racial characteristics, and do not define themselves in terms of black and white. They find that Caribbean Hispanics display a low degree of segregation from white Hispanics and a high degree of segregation from both black Hispanics and non-Hispanic blacks. They are highly segregated from non-Hispanic whites, however, suggesting that they are accepted by white Hispanics on the base of ethnicity/language and shunned by non-Hispanic whites based on their mixed race, while inevitably recognizing the stigma associated with being black in America, and distancing themselves residentially from the black population. This characteristic is also observed among West Indian immigrants, who strive to maintain their ethnic identity in order to avoid being confused with native blacks, the most stigmatized and discriminated segment of U.S. population (Waters, 1999).

Apart from influencing personal residential choices, neighborhood composition preferences are important because they can lead to discrimination in the housing market, for instance through restrictive covenants signed by neighborhood associations, which limited the choices available to minority groups and led to the creation of segregated neighborhoods (Massey and Denton, 1993). Although some of these extreme, formally implemented measures are now illegal, personal discrimination by real estate agents is harder to identify and eradicate, and its global effects are not well known (Clark, 1992).

A slightly different approach to explaining residential segregation of Hispanics is taken by Betancur (1996), who uses a world systems theory framework in his analysis of residential segregation patterns in Chicago. He argues that Hispanic immigration reflects the economic exploitation and political domination of Latin America by the U.S., and that as a consequence the Hispanic experience in the U.S. has been characterized by domination and exclusion manifested through employer abuses, labor segmentation and immobility, and residential segregation.

# 3 Methods and Data

## 3.1 The Model

Despite the wealth of empirical studies that analyze the potential determinants of residential segregation, very few of them have attempted to explicitly identify the manner in which spatial residential patterns emerge from the interaction of these elements in field settings (Clark, 1986b). Important steps in this direction have been made by researchers using agent-based models, which provide useful theoretical insights into the dynamics of residential segregation, but are nonetheless limited in their application to real cases (Fossett, 2006).

Agent-based model research has shown that factors such as neighborhood composition preferences and socioeconomic characteristics influence spatial residential patterns (Schelling, 1969; Sakoda, 1971; Hegselmann and Flache, 1998; Mare and Bruch, 2003; Benenson, 2004; Zhang, 2004; Fossett 2006). In these studies, researchers simulate scenarios in which multiple agents, usually belonging to two racial/ethnic groups make residential location decisions based on their preferences and resources; the resulting spatial residential patterns are then examined for the presence of segregation. For instance, Fossett (2006) shows that

> "ethnic preferences and social distance dynamics can, when combined with status preferences, status dynamics, and demographic and urban-structural settings common in American cities, produce highly stable patterns of multi-group segregation and hyper-segregation (i.e., high levels of ethnic segregation on multiple dimensions) of minority populations" (p. 185)."

With regard to ethnic preferences, most studies in this tradition have used and extended Schelling's model in which agents react to the fraction of familiar agents within the neighborhood (i.e., a threshold). The alternative, proposed by Sakoda (1971) and replicated by Hegselmann and Flache (1998), is to define the attitude of an agent to agents of her own and the other type as attraction, neutrality, or avoidance. In the present study we employ the latter approach and examine the ways in which xenophobia (i.e., a preference not to reside in the same neighborhood as dissimilar alters) and homophily (i.e., a preference to reside in the same neighborhood as similar alters) combine with other factors to influence the spatial distribution of households to neighborhoods.

The assumption on which the approach employed in this study is built is that at any point in time, we can interpret the spatial residential pattern as an equilibrium state of a system of households and neighborhoods, with households located in neighborhoods[3].

[3] "Neighborhood" is a general term that may be replaced with any of the subdivisions of the

However, this system contains various kinds of dependencies: households are tied to one another by kin or friendship relations, and neighborhoods are related by virtue of being contiguous or being a certain distance apart from one another. As such, a traditional regression framework is not going to be very reliable in explaining outcomes, and it will fail to represent the complex dependencies within the system. One area of sociology that has seen tremendous advances toward developing stochastic models for social systems with complex dependence structures is social network analysis, where researchers have drawn on earlier results in other scientific fields such as spatial statistics and statistical physics (Robins and Pattison, 2005) to build models to account for such complex dependencies. Building further on these developments, Butts (2007) has proposed "a general framework for modeling and analysis of systems which can be specified in terms of the arrangement of a finite set of objects with respect to a finite set of locations" (p. 285). This framework can be used to study a range of social processes such as occupational segregation, stratification and settlement patterns.

In this framework, Butts (2007) defines a generalized location system model as a stochastic model for the equilibrium state of a generalized location system. The generalized location system consists of objects (households, organizations, etc.) and locations (places, jobs, etc.). A particular assignment of objects to locations represents a configuration of the system. Given the set of *all* possible configurations, $C$, the system will be found to occupy any *particular* configuration with some specified probability. The probability that at equilibrium the system will be found in a given configuration $l$

metropolitan area usually employed in segregation studies, such as block or census tract.

can be written as

$$\Pr(S = l) = I_C(l) \frac{\exp(\mathcal{P}(l))}{\sum_{l' \in C} \exp(\mathcal{P}(l'))} \tag{1}$$

where $S$ is the equilibrium state, $l$ is the given configuration (i.e., a certain assignment of households to census tracts), $I_C(l)$ is an indicator function that takes a value of 1 if $l \in C$ and 0 if $l \notin C$, and $\mathcal{P}$ is a quantity called the *social potential*.

The location system is more likely to be found in areas of high probability which, as can be observed from equation (1), are also areas of high potential. In other words, the system is more likely to be found to occupy configurations that maximize the social potential function. Since a location system configuration represents the assignment of certain objects to certain locations, the social potential function should include effects that reflect this assignment, and take into account object and location characteristics *at the same time*. Butts (2007) specifies such a functional form for the social potential, one that is based on object attributes and relations among objects, as well as on location attributes and relations among locations.

Table 1 (cf. Butts, 2007, p. 297) presents the classes of effects that can be included in the social potential function; each of them is based on both object and location features. We now consider these four classes of effects and some examples, as well as their functional form. We then turn to the social potential function, which is a linear combination of these effects, and discuss some more about its properties.

Since we apply the location system model in a residential settlement context, we deal

| | Location Attributes | Location Relations |
|---|---|---|
| Object Attributes | Attraction/Repulsion Effects | Object Homogeneity/ Heterogeneity Effects (through Locations) |
| Object Relations | Location Homogeneity/ Heterogeneity Effects (through Objects) | Alignment Effects |

Table 1: Elements in the Social Potential Function

with a particular type of objects and locations, namely households and neighborhoods, respectively. From this point on we use these terms for clarity. It is also important to note that by describing an "assignment" of households to neighborhoods we do not wish to convey the idea that some entity is responsible for assigning certain houses to certain neighborhoods; rather, the assignment of households to neighborhoods (i.e., the equilibrium configuration of a system composed of households and neighborhoods) is the result of underlying social mechanisms we wish to identify. As suggested by Hedstrm and Swedberg (1998), these mechanisms are expressed as functions transforming variables, and are therefore included as effects in the present model.

Attraction/repulsion (or "push/pull") effects are based on the attributes of households (such as income) and the attributes of neighborhoods where they reside (such as rent levels). Neighborhoods have attributes that may make them attractive (or undesirable) to households with particular attributes. For instance, high-rent neighborhoods attract households with high income, and at the same time repel households with low income. One important mechanism that can be modeled as an attraction/repulsion effect is discrimination. In this framework, discrimination may be understood as a conditional tendency for households with certain attributes to be found in (or denied access to) neighborhoods with certain attributes[4] .

We define the attraction potential, $\mathcal{P}_\alpha$, as

$$\mathcal{P}_\alpha(l) = \sum_{i=1}^{a} \alpha_i \sum_{j=1}^{n} \mathbf{Q}_{l_i j} \mathbf{X}_{ij} \tag{2}$$

where $\mathbf{X}$ is a matrix of household attributes, $\mathbf{Q}$ is a matrix of neighborhood attributes, and $\alpha$ a parameter vector.

The second class of effects covers household homogenity/heterogeneity based on neighborhood relations. In other words, this effect captures the tendency for associated (e.g., contiguous) neighborhoods to be occupied by households with similar (or different) attributes. A xenophobia effect can be understood in this framework as the tendency for households of the same race/ethnicity to reside in contiguous neighborhoods, thus reducing local heterogeneity. We define the object homogeneity potential, $\mathcal{P}_\beta$, as

$$\mathcal{P}_\beta(l) = \sum_{i=1}^{b} \beta_i \sum_{j=1}^{n} \sum_{k=1}^{n} \mathbf{B}_{il_j l_k} |\mathbf{Y}_{ij} - \mathbf{Y}_{ik}| \tag{3}$$

where $\mathbf{Y}$ is a matrix of household attributes, $\mathbf{B}$ is an array of adjacency matrices on the neighborhood set, and $\beta$ a parameter vector.

Effects of location homogeneity/heterogeneity through relations of objects capture the tendency for locations that are similar to be occupied by people who are associated in some way. An example of such effects is recruitment by entrepreneurs through

[4] This formulation does not include, however, any indication as to why things happen as they do, i.e., who is denying access, etc.

networks of immigrants. The result is that similar types of jobs (supermarket assistants, for instance) are occupied by people from the same family or community. It is slightly more difficult to interpret this type of effect when locations are geographical units. We define $\mathcal{P}_\gamma$ as

$$\mathcal{P}_\gamma(l) = \sum_{i=1}^{c} \gamma_i \sum_{j=1}^{n} \sum_{k=1}^{n} \mathbf{A}_{ijk} |\mathbf{R}_{l_j i} - \mathbf{R}_{l_k i}| \tag{4}$$

where $\mathbf{R}$ is a matrix of location attributes, $\mathbf{A}$ is an array of adjacency matrices on the object set, and $\gamma$ a parameter vector. Finally, alignment effects reflect the tendency for households that are related to occupy locations that are related in their turn. An example of such an effect is propinquity, the tendency for households that are linked through kinship or friendship ties to reside in neighborhoods that are contiguous (or close, in Euclidean distance terms). We define $\mathcal{P}_\delta$ as

$$\mathcal{P}_\delta(l) = \sum_{i=1}^{d} \delta_i \sum_{j=1}^{n} \sum_{k=1}^{n} \mathbf{W}_{ijk} \mathbf{D}_{i l_j l_k} \tag{5}$$

where $\mathbf{W}$ is an array of household relation adjacency matrices, $\mathbf{D}$ is an array of neighborhood relation adjacency matrices, and $\delta$ a parameter vector. The category of alignment effects is the most flexible, since its functional form uses matrices as inputs and many mechanisms can be expressed based on matrices built from household and neighborhood features [5].

$$\mathcal{P}(l) = \sum_{i=1}^{a} \alpha_i \sum_{j=1}^{n} \mathbf{Q}_{l_j i} \mathbf{X}_{ji} + \sum_{i=1}^{b} \beta_i \sum_{j=1}^{n} \sum_{k=1}^{n} \mathbf{B}_{i l_j l_k} |\mathbf{Y}_{ji} - \mathbf{Y}_{ki}|$$
$$+ \sum_{i=1}^{c} \gamma_i \sum_{j=1}^{n} \sum_{k=1}^{n} \mathbf{A}_{ijk} |\mathbf{R}_{l_j i} - \mathbf{R}_{l_k i}| + \sum_{i=1}^{d} \delta_i \sum_{j=1}^{n} \sum_{k=1}^{n} \mathbf{W}_{ijk} \mathbf{D}_{i l_j l_k} \qquad ()$$

where $\mathbf{X}$ and $\mathbf{Y}$ are matrices of household attributes, $\mathbf{Q}$ and $\mathbf{R}$ are matrices of neighborhood attributes, $\mathbf{B}$ and $\mathbf{D}$ are arrays of neighborhood relation adjacency matrices, and $\mathbf{A}$ and $\mathbf{W}$ are arrays of household relation adjacency matrices. Each of the terms in the social potential function thus involves an interaction between the attributes or relations of households and the attributes or relations of the neighborhoods in which they reside.

We can also express the social potential in terms of parameters and statistics as

$$\mathcal{P}(l) = \sum_{i=1}^{a} \alpha_i t_i^{\alpha}(l) + \sum_{i=1}^{b} \beta_i t_i^{\beta}(l) + \sum_{i=1}^{c} \gamma_i t_i^{\gamma}(l) + \sum_{i=1}^{d} \delta_i t_i^{\delta}(l) \qquad (6)$$

where $\alpha$, $\beta$, $\gamma$ and $\delta$ are the model parameter vectors, and $t^{\alpha}$, $t^{\beta}$, $t^{\gamma}$ and $t^{\delta}$ are vectors of sufficient statistics, i.e., deterministic functions of object and location attributes and relations. Using such a potential function leads to a regular exponential family on $C$, which has well known statistical properties.

To summarize, we propose here the use of the generalized location system models developed by Butts (2007) to study the factors that influence spatial residential patterns.

[5] In fact, all the four categories of effects presented so far could be written as products of matrices. This is not advisable, however, first because we need to be able to differentiate among the various types of effects, and second because of the high computational costs associated with using matrices.

Within this approach, we model the probability of observing a particular distribution of households across neighborhoods as resulting from the interaction of various factors such as availability of housing, wealth, and preferences for neighborhood composition. The advantages of this framework are that it can be readily simulated, allowing for the testing of simple scenarios, it is specifiable in terms of directly measurable properties, and supports likelihood-based inference (using Markov Chain Monte Carlo methods). Another set of characteristics that recommends the use of this framework for the study of residential settlement patterns is the ability to include as covariates a range of factors such as population density, inter-household ties, individual preferences and neighborhood characteristics, and examine the effect of their interactions in determining residential patterns. By specifying values of the parameters in a simulated scenario, we can obtain assignments of households to locations that illustrate what the spatial patterns would be if particular social mechanisms were at play. By estimating the parameters from real data we can obtain information about the relative influence of the effects on the observed assignment.

## 3.2 Data

As mentioned above, the choice of the Yuma metropolitan area for this study is motivated by its relatively small size and the fact that the population consists mainly of two ethnic groups, Hispanics and non-Hispanic whites. These characteristics are desirable for the analyses we conduct both because they make results easier to interpret and because they are less expensive computationally.

| Household Attributes | Census Tract Attributes | Census Tract Relations |
|---|---|---|
| Race of householder | Median rent | Contiguity |
| Hispanic origin | Number of housing units | |
| Income | | |

Table 2: Household-level and census tract-level variables

The Yuma, AZ, metropolitan area is situated on the U.S. border with Mexico and comprises the entire Yuma county. The area is divided into 33 census tracts and has a population of 53,848 households, out of which 37% have a Hispanic householder, and 58% have a non-Hispanic White householder. We use geographic information about census tract placement and boundaries from the U.S. Census to define the spatial layout for our analyses.

The choice of covariates for household and census tract characteristics is based on factors suggested by the literature as potential determinants of residential segregation, and the availability of data [6]. These covariates are listed in Table 2 below.

Although the variables we include are race and Hispanic origin, we use only the two categories of Hispanic and Non-Hispanic white. Based on these variables, we include the following effects in the present analysis:

1. Attraction: based on household income and census tract rent levels

2. Xenophobia: object heterogeneity effect based on household race/ethnicity and tract contiguity. Here, xenophobia is interpreted as stemming from a preference for being away from dissimilar alters (i.e., outgroup avoidance), without any pref-

[6] Unfortunately, data on links among households is not available for this case, which means that no effects based on household relations can be included in the analysis.

| Effect | Covariate | Form |
|---|---|---|
| Attraction | **X** | Vector of household income values |
| Attraction | **Q** | Vector of neighborhood rent values |
| Xenophobia | **Y** | Vector of household ethnicities |
| Xenophobia | **B** | Tract contiguity matrix |
| Single Homophily | **W** | Matrix of ethnic similarity for one group |
| or Double Homophily | | Matrix of ethnic similarity for both groups |
| Homophily (any type) | **D** | Tract contiguity matrix |

Table 3: Covariates and effects

erence for members of the same group. Negative values of $\beta$ are associated with xenophobia, while positive values of $\beta$ would be associated with xenophily, a preference to be close to dissimilar alters.

3. Homophily: alignment effect between similarity in race/ethnicity and tract contiguity, interpreted as stemming from a preference for being close to similar alters, without any preference toward members of the other group. In a scenario with two groups it can have two forms:

   (a) Differential homophily, where only the members of one of the groups prefer to be close to similar alters

   (b) Uniform homophily, where members of both groups prefer to be close to similar alters.

   Positive values of $\delta$ are associated with homophily, while negative values of $\delta$ would be associated with homophobia, or same-group avoidance.

The relationship between covariates and effects is presented in Table 3.

Our analysis will proceed as follows: first, we estimate the values of the parameters for the observed Yuma residential pattern configuration. Second, keeping the attraction parameter constant, we vary the values of the xenophobia and homophily parameters, and simulate alternate Yuma residential pattern configurations. This will allow us to understand how changes in the parameters could affect the overall assignment of households to neighborhoods and therefore segregation levels in this area.

## 4 Analysis

In the first instance, we used a reduced version of the data, in which each of the "objects" in our analysis is not one household, but (approximately) 100 households; we use the term "representative agents" to refer to these entities. We cross-tabulated the two variables we are using, race/ethnicity and income, to create a contingency table that contains the number of households that fall into each of the possible combinations between the categories of the two variables. Although the variables we include are race and Hispanic origin, we use only the two categories of Hispanic and Non-Hispanic white, and since there are 16 income categories reported by the Census, we have 32 possible race/ethnicity/income categories in our table. To obtain the representative agents data we then divided the population counts in each of the cells of the table by 100 and rounded to the nearest integer. The total number of representative agents obtained is 463, and because of the rounding procedure, each of them represents 50 or more households in the same race/ethnicity/income category. Each representative agent was assigned an income drawn randomly from a uniform distribution over the interval covered by the respective income category (for instance, a representative agent in the \$15,000-\$20,000 income category was assigned a value randomly drawn from the Uniform(15000, 20000) distribution).

Figure 1 shows the spatial distribution of representative agents in the Yuma metropolitan area, with red circles representing hundreds of Hispanic households, and black circles representing hundreds of non-Hispanic White households. The circles are drawn at the population centroid of each tract, and their positions are jittered so as to prevent overlap.

We note a few things by looking at Figure 1. Some of the census tracts are empty, which is a result, in this case, of the fact that there were less than 50 households in each of the 32 categories defined by our variables. The highest concentration of households is present in the census tracts around downtown Yuma. We also notice a concentration of red circles in the lower left corner, which is an indication of a high concentration of Hispanic households in the area. This is not entirely surprising, since the area is right on the border with Mexico, and a closer examination indicates the existence of a locality that spans the border

We use the spatial proximity index to characterize the spatial distribution of representative agents in the Yuma metropolitan area. Although they are based on proportions of minority/majority population in clearly defined neighborhoods, most residential segregation indices do not take into account the location of these spatial units of measurement relative to each other, thus ignoring important aspects of segregation such as the geographic distance between two group concentrations (White, 1983; Massey and

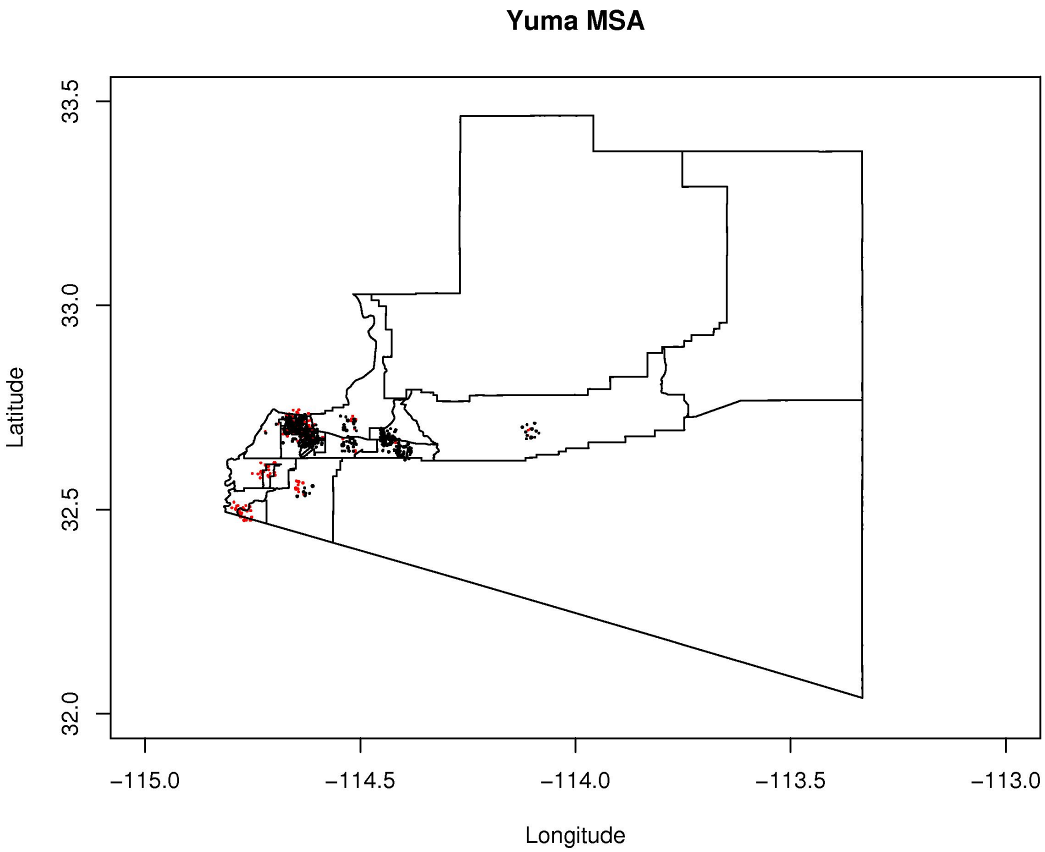


Figure 1: Spatial distribution of representative agents, Yuma MSA

Denton, 1988; Grannis, 2002). Clustering indices address this shortcoming and measure "the extent to which areal units inhabited by minority members adjoin one another, or cluster, in space" (Massey and Denton, 1988, p. 293). The spatial proximity index, ($SP$), is a clustering index proposed by White (1986), which calculates the average of intragroup proximities for the minority and majority populations, weighted by the proportions each group represents of the total population.

Spatial proximity equals 1 if there is no differential clustering between minority and majority group members. It is greater than 1 when members of each group live nearer to one another than to members of the other group, and is less than 1 if minority and majority members live nearer to members of the other group than to members of their own group. For the present case, using representative agents, $SP = 1.22$, suggesting low levels of segregation. A further use of the spatial proximity index is in the creation of "regime" plots, which will be presented shortly.

To fit the location system model to the data, we need to go through the following steps:

1. obtain initial parameter estimates via maximum pseudo-likelihood estimation

2. based on these parameters, simulate assignments

3. calculate mean statistics for the simulated assignments

4. calculate the Euclidean distance between the mean simulated statistics and the vector of observed statistics

5. refine parameters iteratively and repeat steps 2-4 until the distance is less than 0.01

Mean simulated statistics were calculated over a sample of 25,000 draws from the probability distribution of equation ( 1), uniformly thinned from a total sample of size 25,000,000. Parameter estimates obtained through the procedure outlined above are presented in Table 4, together with the deviance estimate (i.e., -2×loglikelihood) and the corresponding AIC scores [7]. We present estimates for models including the attraction effect, on the one hand, and attraction, xenophobia and homophily effects, on the other [8].

These results indicate that the best model is the one that includes only the attraction effect based on household income and rent, since it has smallest AIC score. Although the attraction, xenophobia, and differential homophily model is a better fit, the AIC is larger for this model because it includes more parameters. This suggests that the dominant mechanism within this residential system is attraction/repulsion based on household income and rent. However, it is useful to examine the estimates for the other models, especially because they can be linked with the next step in the analysis, the comparison with *alternate* Yuma residential system configurations.

To obtain such alternate configurations, we vary the xenophobia and homophily parameters while keeping the attraction parameter constant, and simulate assignments

[7] Akaike's Information Criterion (AIC), equals $2k - 2\ln(L)$, where $k$ is the number of parameters, and $L$ is the likelihood. Therefore, this measure not only rewards goodness of fit, but also includes a penalty that is an increasing function of the number of estimated parameters.

[8] At the time of writing we had not obtained reliable estimates for models that included only xenophobia or only homophily effects.

| Effect | $\alpha$ Only | All Effects with Differential Homophily | All Effects with Uniform Homophily |
|---|---|---|---|
| *Attraction/Repulsion Effects* | | | |
| Income × Rent | $6.752579\times10^{-8}$ | $6.788228\times10^{-8}$ | $7.161418\times10^{-8}$ |
| *Object Heterogeneity Effects* | | | |
| Xenophobia = Ethnicity × Tract Contiguity | | $9.260738\times10^{-7}$ | $3.857106\times10^{-7}$ |
| *Alignment Effects* | | | |
| Homophily = Ethnic Similarity × Tract Contiguity | | $1.337139\times10^{-6}$ | $-3.857106\times10^{-7}$ |
| Deviance | 392.908 | 389.3142 | 395.6693 |
| Model Degrees of Freedom | 1 | 3 | 3 |
| AIC | 394.908 | 395.3142 | 401.6693 |

Table 4: Maximum likelihood estimates for Yuma models

of households to census tracts based on these parameters. To examine these assignments with regard to the presence of residential segregation, we then calculate and plot the mean values of $SP$ for each pair of parameter values. Plotting the values of $SP$ that characterize the simulated residential configurations resulting from varying the values of the parameters may suggest the existence of various "regimes", areas of the parameter space where the system becomes locked into configurations with certain characteristics.

Figures 2 and 3 show the mean values of the spatial proximity index for 100 residential configurations simulated based on a model that includes attraction, xenophobia and differential homophily effects (Figure 2) and attraction, xenophobia and uniform homophily effects (Figure 3). The attraction parameter was kept constant at a value equal to the maximum likelihood estimate in Table 4, while the xenophobia and homophily parameters were left to vary between -0.75 and 0.75.

For both figures, low values of the spatial proximity index are represented by light blue areas (or light gray in a gray scale printout), while high values are represented by pink areas (or darker gray in a gray scale printout). The x-axis of the plot is labelled "Ethnic Heterogeneity" to provide a more intuitive understanding of the consequences of varying the xenophobia parameter, $\beta$: positive values increase heterogeneity (thus decreasing segregation), while negative values increase homogeneity, thus increasing segregation.

To interpret these plots, we first divide them into four quadrants, based on the sign of parameter values. For instance, the upper-left quadrant shows $SP$ scores for positive values of the homophily parameter, $\delta$, and negative values of the ethnic heterogeneity

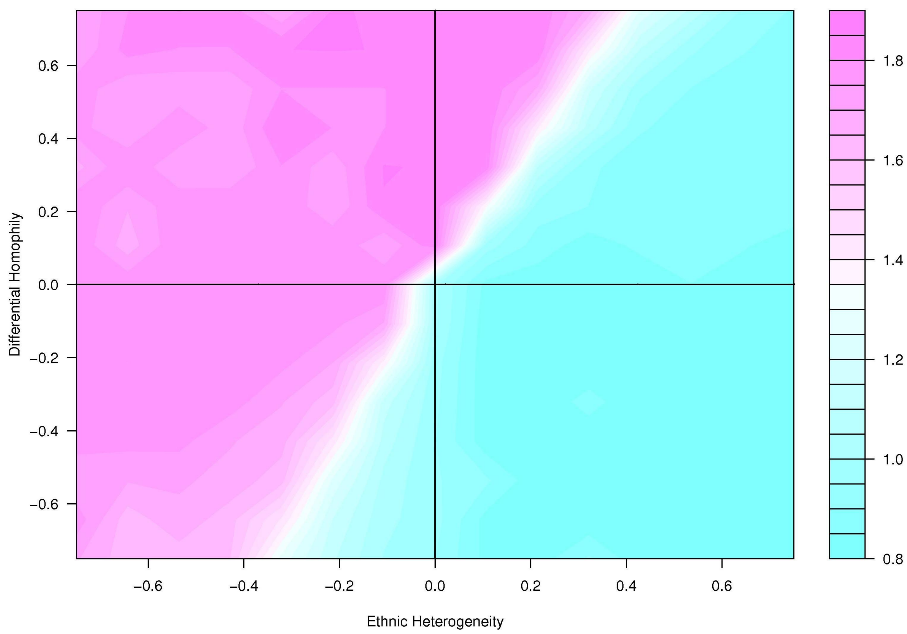


Figure 2: Spatial proximity index values for simulated Yuma configurations, attraction, xenophobia and differential homophily effects

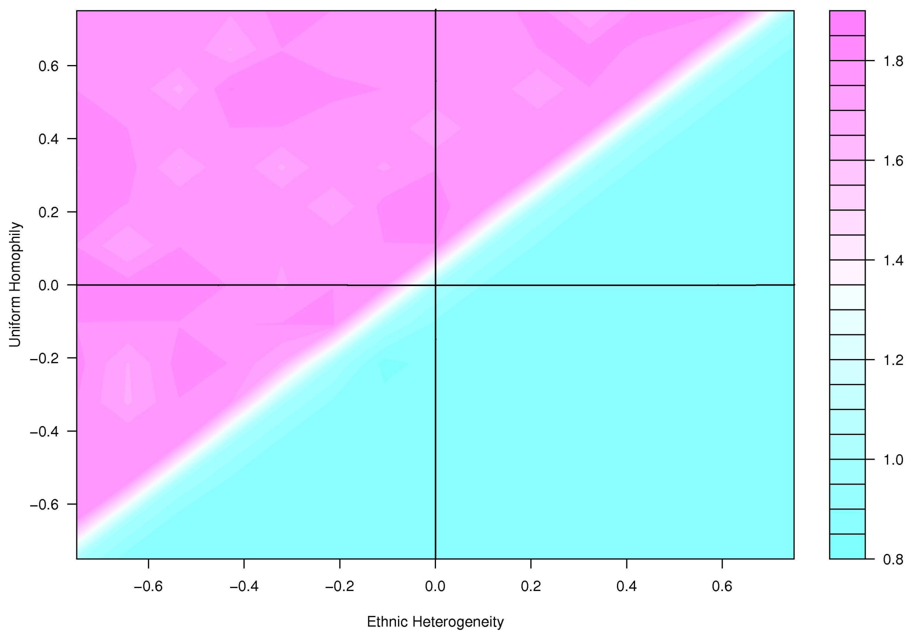


Figure 3: Spatial proximity index values for simulated Yuma configurations, attraction, xenophobia and uniform homophily effects

parameter, $\beta$. This corresponds to scenarios in which we have both homophily and xenophobia mechanisms, and the result would be segregation, as indicated by the values of the $SP$. In other words, if the distribution of households across neighborhoods in the Yuma residential system would be a result of attraction, xenophobia, and homophily mechanisms, we would expect to see values of the spatial proximity index in the neighborhood of 1.7, which would indicate high levels of segregation. As mentioned above, for the observed Yuma residential system configuration, $SP$=1.22, so according to these plots, the observed Yuma configuration should be characterized by a combination of parameters that places this case in the light blue areas of the regime plot. This is indeed the case, for both models including all three effects. The $\beta$ and $\delta$ parameter estimates for the attraction, xenophobia and differential homophily model are both positive, which means that our case should be located in the upper-right quadrant. As we can see, $SP$ values around zero in this quadrant are around 1.1-1.2, very close to the value obtained from our data. Similarly, the positive $\beta$ and negative $\delta$ estimates of the attraction, xenophobia and uniform homophily model would suggest that our case should be located in the lower-right quadrant, and, indeed, simulated assignments found in this quadrant have $SP$ values close to 1.22.

# 5 Conclusion

We have presented here a newly developed statistical framework for the modeling of residential systems of households and neighborhoods. We have applied this model in the case of the Yuma metropolitan area, not necessarily because we think that

Yuma is representative of metropolitan areas that have a Hispanic/non-Hispanic white population mixture, but mostly because this case represented, due to its population composition properties, as simple an example as possible. Such an example was needed since this is the first attempt to fit the model to real data. The novelty of the approach is also the ground for trying to keep the number of covariates and effects to a minimum. Until the behavior of this new family of models is well understood, it is inadvisable to include every effect that has ever been suggested by the literature.

The framework allows the researcher to conduct both simulation and parameter estimation based on real data, and thus represents a step forward from previous studies that employed either agent based simulations or classical statistical techniques alone. One of the most important implications of being able to compare the real distribution (or assignment) of households to neighborhoods in the Yuma case with simulated assignments is the ability to understand what are some likely consequences of changes in parameter values for segregation levels. For example, we can see that, for the same levels of the $\beta$ parameter (which, in both models, actually indicates the presence of "xenophily"), an increase in the homophily parameter would lead to more segregated configurations, since it would move the assignment in the area of the regime plot characterized by high levels of $SP$. Similarly, for the same levels of the $\delta$ parameter, a decrease in the xenophobia parameter would lead to more segregated configurations. One could imagine such scenarios in which some event would determine one of the groups to seek the proximity of members of the same group and which would make member of the other group feel threatened by the dissimilar alters. Most importantly, however, such analyses may prove informative for policy makers who are interested in reducing segregation levels.

# 6 References

Alba, Richard and Victor Nee. 2003. *Remaking the American Mainstream: Assimilation and Contemporay America* . Cambridge: Harvard University Press.

Benenson, Itzhak. 2004. "Agent-Based Modeling: From Individual Residential Choice to Urban Residential Dynamics." In Michael F. Goodchild and Donald G. Janelle (eds.), *Spatially Integrated Social Science*. New York: Oxford University Press

Betancur, John J. 1996. "The Settlement Experience of Latinos in Chicago: Segregation, Speculation, and the Ecology Model". *Social Forces*. 74(4): 1299-1324

Butts, Carter T. 2007. "Models for Generalized Location Systems". *Sociological Methodology* 37:283-348

Clark, W.A.V. 1986. "Residential Segregation in American Cities: A Review and Interpretation". *Population Research and Policy Review*, 5:95-127

Clark W.A.V. 1992. "Residential Preferences and Residential Choices in a Multiethnic Context". *Demography*, 29:451-466

Denton, Nancy A., and Douglas S. Massey. 1989. "Racial Identity Among Caribbean Hispanics: The Effect of Double Minority Status on Residential Segregation". *American Sociological Review*. 54(5):790-808.

Epstein, Joshua M. and Robert Axtell. 1996. *Growing Artificial Societies: Social Science from the Bottom Up* Washington, D.C.: The Brookings Institution Press

Foner, Nancy. 2000. *From Ellis Island to JFK: New Yorks Two Great Waves of Immigration*. Yale University Press.

Fossett, Mark. 2006. "Ethnic preferences, social distance dynamics, and residential segregation: Theoretical explorations using simulation analysis". *Journal of Mathematical Sociology* 30:185-274

Frey, William H. and Reynolds Farley. 1996. "Latino, Asian, and Black Segregation in U.S. Metropolitan Areas: Are Multi-ethnic Metros Different?". *Demography*. 33(1):35-50

Gottdiener, Mark, and Ray Hutchinson. 2000. *The New Urban Sociology*. Boston: McGraw-Hill

Grannis R (1998) The Importance of Trivial Streets: Residential Streets and Residential Segregation. American Journal of Sociology 103:1530-1564

Grannis R (2002) Discussion: Segregation indices and their functional inputs. Sociological Methodology 32:69-84.

Hawley, Amos H. 1950. *Human Ecology: A Theory of Community Structure*. New York: Ronald

Hedstrm P, Swedberg R (1998) Social mechanisms: An analytical approach to social theory. Cambridge University Press, New York

Hegselmann R, Flache A (1998) Understanding complex social dynamics: A plea for cellular automata based modelling. Journal of Artificial Societies and Social Simulation 1(3):1, http://www.soc.surrey.ac.uk/JASSS/1/3/1.html

Iceland, J., Weinberg, D., and Steinmetz, E. 2002. *Racial and Ethnic Residential Segregation in the United States: 1980-2000*. Washington, DC: U.S. Census Bureau

Logan, John R. and Wenquan Zhang. 2004. “Identifying Ethnic Neighborhoods with Census Data: Group Concentration and Clustering.” In Michael F. Goodchild and Donald G. Janelle (eds.), *Spatially Integrated Social Science*. New York: Oxford University Press

MacDonald J.S. and L. MacDonald. 1974. ”Chain Migration, Ethnic Neighborhood Formation, and Social Networks,” in *An Urban World*, edited by C. Tilly, Boston: Little, Brown

Mare, Robert and Elizabeth Bruch. 2003. “Spatial inequality, neighborhood mobility, and residential segregation”. California Center for Population Research Working Paper No. 003-03. University of California, Los Angeles

Massey, Douglas S. 1985. “Ethnic Residential Segregation: A Theoretical Synthesis and Empirical Review.” *Sociology and Social Research*, 69:315-350

Massey, Douglas S., Joaquin Arango, Graeme Hugo, Ali Kouaouci, Adela Pellegrino and J. Edward Taylor. 1998. *Worlds in Motion: Understanding International Migration at the End of the Millennium*. Oxford and New York: Oxford University Press.

Massey, Douglas S. and Nancy A. Denton. 1985. “Spatial Assimilation as a Socioeconomic Outcome”. *American Sociological Review*, 50(1):94-106

Massey, Douglas S. and Nancy A. Denton. 1987. “Trends in the Residential Segregation of Blacks, Hispanics, and Asians: 19701980.” *American Sociological Review*, 94:802825

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

Massey, Douglas S., White, Michael J. and Phua, Voon-Chin. 1996. “The Dimensions of Segregation Revisited”, *Sociological Methods and Research*, 25(2):172206

McKenzie, Roderick. 1923. “The Ecological Approach to the Study of Urban Community.” Reprinted in James F. Short (ed). 1971. *The Social Fabric of the Metropolis: Contributions of the Chicago School of Urban Sociology*. Chicago: University of Chicago Press

Menjívar, Cecilia. 2000. *Fragmented Ties: Salvadoran Immigrant Networks in America.* Berkeley: University of California Press

Petrescu-Prahova, Miruna. 2007. “Discrete Exponential Family Models for the Study of Ethnic Residential Segregation”. In Adrian Bejan (ed.), *Constructal Theory of Social Dynamics*, Springer

Portes, Alejandro and Min Zhou. 1993. “The New Second Generation: Segmented Assimilation and Its Variants Among Post-1965 Immigrant Youth.” *Annals of the American Academy of Political and Social Science*. 530:74-98

Reardon, Sean F. and Glenn Firebaugh. 2002. "Measures of Multigroup Segregation". *Sociological Methodology*. 32:33-67

Reardon,Sean F. and David O'Sullivan. 2004. "Measures of Spatial Segregation". *Sociological Methodology*, 34:121-162

Sakoda, James M. 1971. "The checkerboard model of social interaction." *Journal of Mathematical Sociology* 1:119-132

Schelling, Thomas C. 1969. "Models of segregation". American Economic Review, 59(2):483-493.

Schelling, Thomas C. 1971. "Dynamic models of segregation." *Journal of Mathematical Sociology* 1:143-186

South, Scott J., Kyle Crowder, and Erick Chavez. 2005. "Geographic Mobility and Spatial Assimilation among U.S. Latino Immigrants". *International Migration Review*. 39(3):577-607

Tauber, Karl E. and Alma F. Tauber. 1965. *Negroes in Cities: Residential Segregation and Neighborhood Change.* Chicago: Aldine.

Thomas, W.I .1921. "The Immigrant Community". Reprinted in James F. Short (ed). 1971. *The Social Fabric of the Metropolis: Contributions of the Chicago School of Urban Sociology.* Chicago: University of Chicago Press

Waldinger, Roger (ed.). 2001. *Strangers at the Gates: New Immigrants in Urban America.* Berkeley: University of California Press

Waters, Mary C. 1999. '*Black Identities: West Indian Immigrant Dreams and American Realities."* Cambridge, MA: Harvard University Press

White, Michael J. and Jennifer Glick. 1999. "The impact of immigration on residential segregation", in F. Bean and S. Bell-Rose (Eds), *Immigration and Opportunity: Race, Ethnicity, and Employment in the United States*, New York: Russell Sage

Wilson, Franklin D. and Roger B. Hammer. 2001. "The causes and consequences of racial residential segregation." Pp. 272-303 in Alice OConnor, Chris Tilly, and Lawrence Bobo (Eds.) *Urban inequality in the United States: Evidence from four cities.* New York: Russell Sage Foundation.

Zhang, Junfu. 2004. "A dynamic model of residential segregation". *Journal of Mathematical Sociology* 28:147-170

Zhou, M., 1992, *Chinatown: The Socioeconomic Potential of an Urban Enclave.*Philadelphia, PA: Temple University Press

Zubrinsky Charles, Camille. 2001. "Processes of Racial Residential Segregation."Pp. 217-71 in Alice O'Connor, Chris Tilly, and Lawrence Bobo (Eds.),*Urban Inequality: Evidence From Four Cities.* New York: Russell Sage Foundation