Empirically Testing a Computational Model:
The Example of Housing Segregation

R.H. Sander, School of Law, UCLA
D. Schreiber, Department of Political Science, UCLA
J. Doherty, Empirical Research Group, UCLA

ABSTRACT

Thomas Schelling's famous model of housing segregation started with a few coins on an eight by eight grid and some very simple assumptions about individual preferences. Using the SWARM programming environment, we have extended Schelling's concept to examine the contemporary debate about the nature and causes of housing segregation. We begin with basic preferences functions derived from empirical data on neighborhood racial composition and add a variety of putative factors in housing decisions. The result is a computational model of racial housing segregation that provides insight into empirical patterns of segregation and desegregation in the late twentieth century.

INTRODUCTION

Thirty years ago, the economist Thomas Schelling advanced a theory to explain the persistence of racial segregation in an environment of growing racial tolerance (Schelling, 1971, 1978). Schelling posited a simple model that made a straightforward point: if the racial makeup of one's neighbors is a decisive factor in choosing housing, then the collective interaction of individual preferences will tend to produce segregation, even if many individuals tolerate or even prefer integration.

Schelling's use of "micromotives" to explain "macro" phenomena has become a familiar concept, but for many years it did not advance very far as a practical tool for studying segregation. Mathematically, it is much easier to analyze the aggregate behavior of individuals in market models, in which everyone is engaged in the same transaction, than in Schelling's "tipping" model, in which individuals react to their local environment rather than an aggregated market. But the spread and increased accessibility of computational modeling makes it possible to experiment with much more complex models, incorporating both neighborhood-based racial preferences and market-based housing choices into a more complete representation of urban forces.

In this paper, we describe our development of a computational model of housing segregation over the past year. We particularly focus here on a problem that has often bedeviled model-builders: how to "test" models with empirical data. We believe our experiments show some progress towards the goal of rigorously testing alternate theories of segregation.

The Paradox of Modern Housing Segregation

Housing segregation -- particularly of African Americans -- continues to be a dominant feature of most American cities, and it has been linked to a wide range of urban ills (Massey and Denton, 1994). Scholars in the field generally agree that income differences between blacks and whites only explain a small fraction of current segregation levels (Muth, 1986; Clark, 1986); but beyond this, there is no consensus and little in the way of convincing evidence to demonstrate why housing segregation has remained so high.

As an historical matter, it is obvious and widely agreed that black housing segregation came about through organized, mostly private efforts to ghettoize blacks in the early twentieth century -- particularly the years between the world wars (Sander, 1988). One can chart the rise of segregation through the index of dissimilarity, which measures the proportion of one group (e.g., blacks) who would have to move to different neighborhoods to achieve the same metropolitan distribution as a second group (e.g., whites). Where an index of 1.00 equals complete apartheid, black/white segregation rose in American cities from an average of .6 in 1910 to .90 in 1940.

Intense levels of housing discrimination persisted throughout metropolitan America during the 1940s and 1950s, and it is thus unsurprising that segregation remained extremely high into the 1960s. But in the late 1960s, things changed rather dramatically. The civil rights movement brought a striking change in the answers whites gave to survey questions about integration; many more whites expressed marked levels of tolerance. The same movement brought the Civil Rights Act of 1968, which outlawed a wide range of discriminatory conduct in housing markets. Moreover, some real changes in American cities followed in the 1970s. Measured levels of housing discrimination fell sharply by 1977 (Wienk et al, 1979). Black migration to previously-white housing accelerated (Sander, 1998). Although inner-city poverty remained acute, the black middle-class grew and in many ways flourished.

Given these changes, many observers predicted rapid declines in housing segregation. Generally, however, the declines were almost imperceptible. The average black/white index of dissimilarity across a range of cities fell from .88 to .81. Most of the decline was concentrated in a fraction of cities; in many of the nation's largest cities, the black/white index had fallen only a couple of points by 1990 (Sander, 1998).

This, then, is the central paradox of segregation: why have levels of black/white separation changed so little when so many related factors changed a lot? Related to the paradox are a number of smaller but important puzzles: why has housing in predominantly black areas "flipped," from being more expensive than white housing prior to 1970 to being less expensive afterwards? Why are Latinos, who experience levels of discrimination similar to those experienced by blacks, and who have about the same incomes as blacks, much less segregated from whites? And why have a few areas (e.g., Santa Clara County in California or Seattle, Washington) experienced sharp declines in segregation?

Scholars of segregation have not been able to provide robust theories that account for these related puzzles. While an extended discussion of the literature is beyond the scope of this paper, we believe that there are two central problems: a paucity of attempts to rigorously specify testable hypotheses derived from alternative theories of segregation; and the difficulties noted earlier in operationalizing Schelling's insights. We turned to computational modeling as a way of overcoming both problems.

Building a Computational Model of Segregation

Although Schelling described his segregation model as a "thought experiment" and focused on simple numerical examples, it lends itself so readily to computational modeling that its has been a standard computational demonstration for many years. In these models, agents are arrayed on an open grid, with each interior agent surrounded by eight squares. Each agent is either black or white, and they are either "happy" or "unhappy", depending on whether no more than some critical threshold of their neighbors are persons of the other race. Some random squares are vacant, and in a series of rounds, each agent who is unhappy moves to a vacant square. The surprising report, as we noted earlier, is that for a wide range of preference levels, integrated neighborhoods will "tip" towards one group or another, leading the outnumbered group to flee and thus producing segregation.

We wanted to bring this simple model closer to the real world in two ways. First, rather than have agents be either "happy" or "unhappy", we wanted them to evaluate and compare a wide range of possible states. Second, we wanted racial preferences to be only one of many factors agents used to compare neighborhoods. We achieved both goals by creating a multivariate utility model through which agents are periodically asked to compare a range of neighborhoods, evaluate the overall utility they would achieve at each location (considering several variables), and decide if they would like to move.

Since we sought to explore the evolution of segregation after the civil rights revolution, we modeled our schematic city on a prototypical large American metropolis in 1970.[1] The city is biracial (with "reds" and "blues" standing in for "whites" and "blacks"), and the minority group is systematically clustered near the center of the city (that is, the initial index of dissimilarity is 1.00). In the models described in this paper, the population of the city is 2,375 (a number that seems to us, so far, large enough to capture the key

---

[1] The federal Fair Housing Act went fully into effect on January 1, 1970 and decennial census conducted in April 1970 thus captures nicely the state of segregation as society embarked on its experiment in housing desegregation.

Please note that though we use the term "city" to describe our model, we envision it as a representation of a metropolitan area, in which the movement choices of agents are limited to the region of the model.

dynamics of the simulation). In each period, a randomly selected tenth of the population is given the option of moving to one of five alternative locations (which are randomly selected from the entire city) or remaining at the agent's current location. The agent compares the sites using five different criteria: housing cost, distance of the new site from the present location, local discrimination, the racial makeup of the immediate neighborhood (the eight adjacent agents using Moore's definition), and the racial makeup of the surrounding community. Each factor is closely related to a competing theory of housing segregation. We operationalized each variable as follows:

Housing cost: Five percent of the city's cells are vacant at any time, and the 2500 cells in the city at large are divided into 25 communities (demographically analogous to census tracts) each containing 100 cells. Popular communities have fewer vacant cells than unpopular ones. The housing price then becomes a simple inverse function of the vacancy rate. In actual runs of the model, neighborhood vacancies range from zero to over 20%.

Distance from existing home: We calculate the distance of each of the five sites to which an agent might move from the agent's present location, using the Pythagorean theorem. We assume that, other things being equal, utility monotonically declines with the distance an agent moves, because we also (simplistically) assume that an agent's relatives, job, friends and church are close to its current location.

Discrimination: In our model, some of the majority-group (red) agents discriminate against minority-group (blue) agents. We capture this by allowing a specified proportion of red agents to impose a utility cost on adjacent squares if they are occupied by blues.

Racial composition of neighbors: Survey data show substantial variation in what blacks say is their "ideal" neighborhood racial mix; the same is true of whites. We therefore created six different utility functions (three for each race) in which each possible neighborhood racial makeup is characterized by a unique utility for the agents. In evaluating moves, the agent calculates the racial makeup of its Moore neighborhood and considers the associated utility.

Racial composition of community: In much the same fashion, agents evaluate the overall racial makeup of their current and alternative communities. Again, we define the community in tracts of 100 agents. Agents use the same utility functions they apply to their Moore neighbors, but to larger groups of neighboring agents.

With these combined factors, we generate an overall utility function like this:

$$Utility = w_1 * neigh\_pref + w_2 * tract\_pref + w_3 * \left(\frac{occupancy_i}{totalhousing_i}\right) + w_4 * distance + w_5 * discrimination$$

Figure 1

Note that each element of the equation has an associated weight. We can thus investigate a wide variety of model specifications by only changing a few values. For instance, we can set w_1 to 1.0 and all the others to 0.0 and run a model akin to Schelling's original (only the race of immediate neighbors matters). The model's flexibility turns out to be crucial in generating empirically testable predictions.

THE BASIC MODEL

We sought, as noted, to create a simulation patterned after a prototypical city in 1970. In our basic model, the minority-group blues, comprising 20% of the overall population, are clustered in an inner-city ghetto. As in most 1970 cities, the black district is more crowded, and has higher housing prices (controlling for quality) than do similar to white neighborhoods (thus, the model starts with no vacancies in the ghetto). We used real-world survey data from the 1960s and 1970s to model the racial preferences of reds and blues (Pettigrew, 1973; Farley, 1979, bearing in mind that survey data does not necessarily measure true preferences. We tried to capture the diversity of real-world preferences by assigning one-third of the agents of each race a set of preferences that roughly described the "most tolerant third", "least tolerant third", or "middle third" of real-world preferences within each race. Finally, we experimented with varying levels of discrimination in order to simulate that measured in the first large-scale studies of discrimination after passage of the Fair Housing Act (Wienk et al, 1979).

From this starting point, we "run" the model through a sequence of time periods. In each period, one-tenth of the agents compare their current location with four randomly-selected alternatives; agents move to whatever location from the available options maximizes their utility. A variety of programmed metrics monitor, over time, various indicia of the city's condition: the level of red/blue segregation, the proportion of agents choosing to move; the utility of the average red or blue agent, and so on.

As a general matter, this "basic model" behaves in a way that broadly mirrors the typical evolution of large American cities after 1970. A significant number of the most-tolerant blacks take advantage of lower discrimination levels to move into white neighborhoods; a smaller number of the most tolerant whites migrate towards the inner-city. As the black population expands into white areas (mostly areas adjacent to the existing ghetto), the price of housing in mostly-black areas falls relative to housing costs in mostly-white areas. Segregation falls modestly, but significantly.

PUSHING THE MODEL: PRIMITIVE EMPIRICAL TESTS

We were delighted to create a complex computational model that could mimic the real world in some simple ways. But this achievement, by itself, did little to solve the puzzles of housing segregation with which we began. How could we reliably determine which elements of our model drove particular results? How could we use the model to compare theories?[2]

A related problem is the choice of parameter weights. Although we could model individual variables (e.g., racial preferences and housing prices) with close attention to real-world data, the relative weights given to the variables in our model were intrinsically arbitrary. One might, through the empirical housing literature, get some idea of how "distance" interacts with "price"; but how much relative weight might migrants give to price, racial makeup, and discrimination?

It is of course possible to get an intuitive feeling about model dynamics by changing parameters: if one increases the frequency of discrimination, for example, the runs of the model tend to produce less desegregation. The challenge is to make this experimentation systematic enough to produce clear and measurable results.

As of this writing, we have made modest but interesting progress on this front. Essentially, we have adopted the method of systematically varying parameters, measuring specific output variables, comparing input variation with output variation, and then comparing model results with empirical, demographic data. Here we give, very briefly, two examples of this technique.

Income Inequality and Segregation

The most common lay explanation of housing segregation blames income inequality -- the inability of blacks to afford housing in white communities. Most segregation scholars discount this explanation, pointing to the substantial overlap in black and white income distributions and a range of data showing that different economic clusters of blacks experience nearly identical levels of segregation (Muth, 1986; Farley, 1986). Nonetheless, there has been surprisingly little theoretical or empirical work explaining exactly why this should be so.

Within our model, it is possible to vary "income" across agents by simply varying the weight different agents give to housing cost. Our assumption is that affluent agents can make housing choices with less regard to housing cost than poor agents. Using this idea, we created three different simulation environments: (1) a world where all agents have the same income; (2) a world where agent incomes vary along a uniform distribution, with the maximum income three times the minimum income; (3) a world like (2), except that minority-group (blue) incomes vary across a narrower range and are, on average, about 20% lower than red incomes. We ran simulations of each world one hundred times and measured the index of dissimilarity after each run. The results are shown in Figure 2.

---

[2] This is a common curse of computational modeling: because the models are designed to illustrate the complexity of particular kinds of interactive phenomena, it is difficult to rigorously derive and "prove" particular results from the model. In general, if proofs could be derived from the model specifications, the computational model itself would be superfluous.

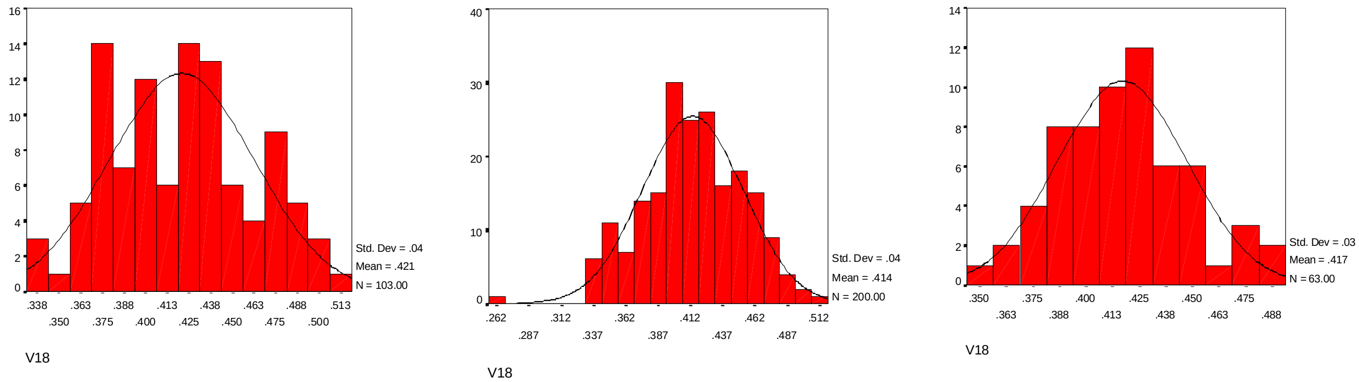


Figure 2. Economic differences do not drive segregation in the model

Housing cost = 50 mean dissimilarity = .42	25 < Housing Cost < 75 mean dissimilarity = .41	Blue/Red economic disparity mean dissimilarity = .42

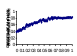
The basic result is clear. In all three models, the average dissimilarity level (over 100 model runs) that exists after 100 time periods is nearly identical (.41 or .42). The income configuration does not seem to matter. This is so even when the housing price (and implicitly, income) variable is itself playing a critical role in the model.

Neighborhood Identity and Segregation

Another theoretical view of segregation emphasizes the role of neighborhood boundaries (Sander 1998). It suggests that urban residents are sensitive not only to the racial makeup of their immediate neighborhood, but also of broader boundaries -- of ethnic enclaves, school districts, and suburbs -- that have crucial significance for some aspect of their social or political life. These boundaries are more likely than other locations to be "defended" by some majority-group members through discrimination, and to be avoided by minority-group members.

In our model, agents consider the racial composition of their immediate "Moore" neighborhood and their broader ("tract") community. We explored the role of neighborhood boundaries by systematically varying the relative weight that agents gave these two definitions of community. As Figure 3 illustrates, we found that the potential desegregation in the model city declines dramatically when the tract weight passes some threshold. At that critical point, tracts as a whole tend to become segregated and remain so. As in many real-world cities, an ordinary street can become a long-term placeholder between white and black communities if the boundary denotes some well-established distinction by which one of the separated communities defines itself.

Figure 3. Stronger geographic borders increase segregation in our model



CONCLUSION

The computational modeling of housing segregation is still in its infancy. However, we believe that our first steps have been fruitful. A utility-based model seems to work effectively to combine many real world factors shaping housing and neighborhood choice. A manageable agent space and iterated time sequences produces computational runs that strongly resemble actual urban dynamics. And systematic variation and outcome measurement allow us to develop data that can be directly analyzed and compared with empirical measurements of the real world. We are hopeful that, in time, these methods will make it possible to test robustly and modify competing theories of housing segregation.

REFERENCES

Works cited

Clark, W.A.V., 1991, "Residential Preferences and Neighborhood Racial Segregation: A Test of the Schelling Segregation Model," *Demography* 28: 1.

Farley, Reynolds, et al, 1979, "Barriers to the Racial Integration of Neighborhoods: The Detroit Case," *Annals of the American Academy* 441: 97.

Farley, Reynolds, 1986, "The Residential Segregation of Blacks from Whites: Trends, Causes, and Consequences," *Issues in Housing Discrimination*, U.S. Commission on Civil Rights: 14-28.

Massey, Douglas, and Denton, Nancy, 1994, *American Apartheid*.

Muth, Richard F., 1986, "The Causes of Housing Segregation," Issues in Housing Discrimination, U.S. Commission on Civil Rights: 3-13.

Pettigrew, Thomas, 1973, "Attitudes on Race and Housing: A Social-Psychological View," *Segregation in Residential Areas*, National Academy of Sciences: 25.

Sander, Richard H. 1988. "Individual Rights and Demographic Realities: The Problem of Fair Housing," *Northwestern University Law Review* 82: 874-928.

Sander, Richard H. 1998. "Housing Segregation and Housing Integration: The Diverging Paths of Urban America," *University of Miami Law Review* 52(4):977-1010.

Schelling, Thomas C., 1971, Dynamic Models of Segregation, *Journal of Mathematical Sociology* 1: 143.

Schelling, Thomas C. , 1978, *Micromotives and Macrobehavior*. New York: W.W. Norton.

Wienk, Reid, Simonson and Eggers, 1979, *Measuring Discrimination in American Housing Markets: The Housing Market Practices Survey*, U.S. Department of Housing and Urban Development.