RESIDENTIAL NEIGHBORHOOD EFFECTS

Yannis M. Ioannides[1]

Department of Economics

Tufts University

Medford, MA 02155

Email: yioannid@tufts.edu Voice: 617-627-3294 Fax: 617-627-3917

April 21, 2000

Abstract

This paper develops a model of housing decisions which allows for social interactions within small residential neighborhoods. It focuses on two decisions: homeowners' valuation of their own property and their maintenance decisions. It reports an empirical investigation with data from the American Housing Survey for 1985 and 1989. It explores a neglected feature of the data, namely the availability of data of neighborhood clusters for metropolitan areas in the United States, with neighborhoods consisting of a dwelling unit and its ten nearest neighbors. The paper identifies an important, and statistically very significant, effect of social interactions for both decisions examined, while individual and dwelling unit characteristics are accounted for.

JEL classification codes: R21, C39
Keywords: Neighborhood effects, housing, social interactions.

[1] I thank two referees and the editor for very insightful comments on an earlier version. I also thank the members of the MacArthur Social Interactions and Inequality Research Network, and Jeff Cohen, Anna Hardman, Vernon Henderson, Gib Metcalf, Andrew Oswald, Holger Sieg, Jeff Zabel, participants at the 97 ESEM (Toulouse), and seminar participants at Brown and Harvard for comments to earlier versions of the paper, and Pamela Jia and Annie K. Thompson for excellent research assistance. I am grateful to the National Science Foundation and the John D. and Catherine T. MacArthur Foundation for financial support.

1 Introduction

Residential neighborhoods are where most people spend a large fraction of their lives and where many of their social and economic interactions take place. In the United States, an extraordinary set of formal and informal organizations of civil society evolves around residential neighborhoods. Local community control of public schools and substantial power of local political organizations in the formation of community jurisdictions is one of key characteristics of the flexibility of U.S. economic and social institutions.[2]

The institutional flexibility goes hand-in-hand with the extraordinary residential mobility of the U.S. population: 17% of the total resident population (with the percentage being the same for 30 to 44 year-olds) moved homes from 1991–1992, with moves being more frequent in the West (21%) and less in the Northeast (12%). This process, together with the forces of urban growth and decay, facilitates the spatial arbitrage which is crucial for the valuation of U.S. residential capital. The latter, in 1991, at 7,889 billion dollars was nearly three times the 2,688 billion dollars of total assets of U.S. manufacturing corporations. [3]

Housing is a major component both of the consumption bundle and of personal wealth, and the single most important component of the tax base of primarily residential communities. A fair amount of research has addressed the way in which individuals accumulate wealth. However, past research has not considered in depth either the spatial aspects of the process nor its interaction with neighborhood change. These two are of course interdependent. The value of a particular house may go up because of its proximity to other valuable property and to other types of desirable developments in its vicinity. A full understanding of the microeconomic underpinnings of the determinants of the market value of housing (and thus of residential capital) will benefit from careful attention to the dynamics of interaction within residential neighborhoods.

This paper presents an empirical investigation of residential neighborhood effects. It relies on data from the American Housing Survey (AHS), which are collected on a *panel* of dwelling units and

[2]Economists have shown interest in the phenomenon of local interactions, as addressed by Schelling (1971), and in the formation of local communities, especially as reflected in the Tiebout model [Tiebout (1956)]. Benabou (1996) and Durlauf (1996; 1997) have reconsidered the fundamental underpinnings of this model. Lack of empirical attention to key ideas underlying Tiebout's theory would have been astonishing were it not for a recent revival of interest; see, for example, Hoyt and Rosenthal (1997).

[3]Source: U.S. Bureau of the Census (1994), T. 1213 and 866. See also, T. 745.

their current occupants. The paper makes use of a little known feature of the survey: for roughly one out of a hundred of dwelling units sampled, up to ten of their nearest neighboring units are also sampled. The decisions on which I focus are basically dictated by data availability. The concept of the neighborhood used here is a literal one based on physical proximity of dwellings.[4] The notion of the *residential* neighborhood is central to a variety of social interactions. A plethora of phenomena, such as individuals' attitudes towards race, income inequality, crime and ethnicity factors may be both causes and effects of the composition of their immediate physical and human environment. [5]

I examine empirically the extent in which individuals' valuations of their properties and their decisions about maintenance depend upon those of their neighbors, while individual, neighbor and neighborhood characteristics are controlled for. It should be no surprise that the evolution of property values over time exhibits dynamic dependence. However, when the average property value among one's neighbors is included as an explanatory variable as well, the latter emerge as a more important determinant than the own lagged value. This finding suggests, as we shall see, that social interaction effects are present. It could mean that exogenous changes which affect neighborhood-average magnitudes, as do some policy-based interventions, are likely to have numerically large effects.

Most research to date on income and race segregation has used geographic detail which is no smaller than census tracts [White (1987)], which are geographical units with 3,500 to 7,000 inhabitants. A fair amount of research uses data for Metropolitan Statistical Areas (MSAs, for short), which are large geographical units. Census tracts and MSAs are arguably too large for studying residential neighborhood interactions, although they are quite appropriate for other types of interactions. The AHS data on individual dwelling units, on their occupants and on their immediate neighbors allow us a glimpse at the workings of many processes which are likely to be obscured out at higher levels of aggregation.

The remainder of the paper is organized as follows. Section 2 discusses briefly the literature on urban neighborhood interactions. Section 3 outlines a model of individual behavior in the presence

[4]Recent research has provided theoretical foundations for an understanding of the emergence of a variety of economic institutions from local interactions [Durlauf (1997)].

[5]See Glaeser, Sacerdote and Scheinkman (1996) for social-interactions based explanations of the incidence of crime. See also Gladwell (1996) for an epidemic theory crime explanation of the decrease in crime.

of interdependence and of neighborhood equilibrium. Section 4 discusses the data and Section 5 develops econometric models for estimating the behavioral model in the presence of neighborhood interactions. Section 6 presents our empirical results and section 7 concludes.

2 The Literature on Urban Neighborhood Interactions

It is often forgotten that the notion of a neighborhood involves not only spatial proximity but also "a district [...] esp. considered in reference to the character or circumstances of its inhabitants; a small but relatively self-contained sector of a larger urban area" [*The New Shorter Oxford English Dictionary* (1993), p. 1901]. Surprisingly, neighborhood interactions have attracted relatively little attention. A number of studies of urban neighborhood interactions originated in the context of evaluating the urban renewal projects of the 1960's in the United States. Davis and Whinston (1961), Rothenberg (1967) and Schall (1976) study housing maintenance behavior. Stahl (1974) is an exhaustive study of the consequences of neighborhood effects for replacement/rehabilitation of housing and housing maintenance. It is, to the best of my knowledge, the first formal model of residential neighborhood interactions in the context of housing decisions that invokes a symmetric Nash equilibrium setting. Strange (1992) examines the role of distance and negative feedback in neighborhood effects using an interactive neighborhoods model where spillovers occur because individuals are affected by the densities of neighboring areas. Binder and Pesaran (1997) have adapted a linear-quadratic version of Pollak (1976) for the presence of social interactions and shown that under certain conditions the model maps into an equivalent one with "selfish" individuals.

In the urban economics literature, the contribution of neighborhood interactions to the evolution of residential patterns and neighborhood characteristics has received much less attention than the role of local public goods. The notion of a neighborhood evolving around a local public good was proposed by Ellickson (1979).[6] Werczberger and Berechman (1988) incorporate neighborhood effects into a multinomial model of spatial location decisions of individuals and firms and give some numerical simulation results. Certain aspects of urban interactions have been discussed by Miyao,

[6]Ellickson provides an explicit model of neighborhood formation, in which individuals care about nonhousing consumption and neighborhood quality, measured as the *average* housing consumption in each neighborhood. Ellickson contrasts cooperative behavior, where neighborhood quality is treated as a (local) public good and the outcome is the optimal configuration, with noncooperative behavior, which leads to a suboptimal configuration.

who has investigated the stability properties of mixed-city equilibrium in the context of city-wide interactions [Miyao (1978)].

Recent empirical research has examined the economic consequences of residence-based social interactions on individuals as they pass through neighborhoods. In particular, Kremer (1997) finds significant linear effects from neighbors' education on individuals' education, and Ioannides (1999) significant nonlinear ones, as well. With the notable exception of Coulson and Bond (1990), empirical research on the impact of neighborhood effects on residential succession is very limited. [7] I am aware of only two other works that examine neighborhood interactions empirically, both of which use local data. Galster (1987) reports empirical results using data from special surveys conducted in Wooster, Ohio, and Minneapolis, Minnesota. He shows that social interactions are very important in explaining home upkeep behavior.[8] Spivack (1991), who uses data on code violations from Providence, Rhode Island, finds some impact of neighborhood variables: ownership patterns and vacancies are the most influential determinants of maintenance and upkeep.

The lack of recent theoretical research on explicitly spatial models of urban interaction is surprising in view of its relevance. For example, decisions about house maintenance often reflect neighborhood considerations and they in turn contribute to the fundamental dynamics of neighborhood stability. The prisoners dilemma models which were invoked by the older urban renewal literature of the 1960's, as in Davis and Whinston (1961) and Schall (1976), have dramatized this issue.

3 A Model of Neighborhood Equilibrium

The notation which I use in presenting the model and in describing the data is as follows. Let $\kappa = 1, \ldots, K$, denote each of the neighborhoods, which are identified by their *kernels,* the members of the

[7]These authors test a model, due to Bond and Coulson (1989), of the inverse demand for dwelling unit and neighborhood characteristics, by using data on FHA loans and contextual data from census tracts. They show that high-income groups are willing to pay more to live in high-income neighborhoods, but find little evidence of an effect of income on the demand for racial composition. Anas (1980) models the behavior of suppliers in the presence of exogenous neighborhood effects.

[8]Homeowners in "most-cohesive neighborhoods" spend 28–45% more on upkeep. There is also evidence of social-threshold effects, in that social interactions are important only if "collective solidarity sentiments result." More importantly, Galster indicates that he has found evidence in favor of powerful self-fulfilling expectations[9], and that evidence would exonerate the behavior of the "in-migrating" households as the key cause of neighborhood deterioration.

original national sample, who were selected randomly and whose neighbors were also interviewed. A time subscript t for each of the waves is used when appropriate. Let $n(i)$ denote the set of the neighbors of unit i, a definition which also applies to the kernel unit κ; J_i denote unit i's entire neighborhood including itself: $J_i = \{i, n(i)\}$. I shall use $|n(\kappa)|$ to denote the size of kernel κ's neighborhood, which because of missing values may be different from 11. This, somewhat cumbersome, notation is necessary for expressing the full flavor of our setting.

An individual h cares about the quality of housing services produced by her property i, Y_{hit}, which is assumed to also be equal to the number of units of quality it contains, "its size;" its value is denoted by $V_{hit} = p_{it}Y_{hit}$, where the price may vary across the sample because units belong to different neighborhoods. As housing units are durable goods and individuals move in and out of neighborhoods, there coexist in each neighborhood new entrants and old timers. In the presence of transactions costs, the former may optimize with respect to a different set of exogenous variables then the latter. The quality of a property may undergo random deterioration, which is represented in terms of a vector of deterioration shocks, $\Psi_{\{J_i t\}}$. The effects of the deterioration may be dealt with by means of maintenance. The remainder of this section summarizes what one would expect to get from a fully specified model of consumer behavior.[10]

Let $u_{ht}(Y_{hit}, c_{ht}; Y_{\{n(i)t\}}; z_{ht}, Z_{\{n(i)t\}})$ denote individual h's period t utility function, as a function of consumption of housing services, , Y_{hit}, nonhousing consumption, c_{ht}, the individual's own characteristics, z_{ht}, and those of her neighbors, $Z_{\{n(i)t\}}$, and the vector of qualities of all neighboring properties, $Y_{\{n(i)t\}}$. We assume that utility $u_{ht}(\cdot)$ is increasing and concave with respect to housing services and nonhousing consumption. I will return to discuss further its properties with respect to $Y_{\{n(i)t\}}$.

I consider first the housing decision of a new entrant into a neighborhood, an individual who moves at time t into the neighborhood by buying property i of size Y^e_{hit} at time t. A standard utility maximization for housing as a durable good leads to an Euler equation, from which an expression for Y^e_{hit} may be obtained as a function of total wealth as of the beginning of period t, Ω_{ht}, and of

[10] A more detailed model is available in the working paper version of the paper, available from the author's web site.

prices p_{it}, p_{it+1}. That is:

$$Y^e_{hit} = \mathcal{Y}^e(\Omega_{ht}; p_{it}, p_{it+1}; Y_{\{n(i)t\}}, Z_{\{n(i)t\}}). \tag{1}$$

The dependence of $\mathcal{Y}^e$ on $Y_{\{n(i)t\}}$ expresses the fact that the individuals cares about the qualities of the neighborhood properties. Since period $t+1$ deterioration shock may not be known, dwelling units may be subject to random capital gains. Anticipation of capital gains or losses may reflect neighborhood conditions, which is particularly interesting in our context.[11] Even if the next period housing value is known with certainty, the deterioration shocks $\Psi_{\{J_i t\}}$ are sufficient to induce uncertainty on neighborhood conditions in period $t+1$, as perceived as of time t.

The housing decision of an individual h who already owns a dwelling unit i of value Y^o_{hit-1} prior to the deterioration shock that is realized in the beginning of period t. She decides whether or not stay in the neighborhood, and conditional on staying, she chooses maintenance so as to maximize lifetime utility. Depending upon the precise formulation of the model, it is possible that the value of her unit after maintenance, Y^o_{hit}, depends only upon Y^o_{hit-1} and the current deterioration shock. In any case, we would expect that

$$Y^o_{hit} = \mathcal{Y}^o(Y^o_{hit-1}, \Psi_{it}; Y_{\{n(i)t\}}, Z_{\{n(i)t\}}), \Psi_{it}). \tag{2}$$

3.1 Neighborhood Equilibrium

In a Nash equilibrium setting, individual h takes as given her neighbors' decisions and ignores the impact of her own decisions upon $Y_{\{n(i)t\}}$. I note that the dependence of Y^e_{hit} on $Y_{\{n(i)t\}}$, a dependence that I refer to as *neighborhood effects,* would follow, in general, from the specification of the utility function. However, even if the utility function were assumed to be decomposable with respect to the consumption bundle, on one hand, and neighborhood effects, on the other, dependence on neighborhood effects would still arise because of "spatial" (neighborhood) arbitrage, as new residents enter, and some of the old residents choose to leave.

Let J^e_i and J^o_i denote the subsets of J_i consisting of units, occupied by new entrants and by continuing residents of a particular neighborhood, respectively. The solutions for housing demand, (1) and (2), when considered as a system, imply a mapping, in general, from the characteristics of

[11] Dysanski and Wilson (1992) have shown that such random capital gains may cause housing demand to exhibit unconventional properties.

a resident h of unit $i \in J_i^e$, and the vector of housing stocks in the neighborhood. Therefore, the dynamic evolution of the vector of housing qualities in a neighborhood possesses a switching law structure, as differences in the decision problems of the two types of agents imply differences in the optimal solutions. By treating the optimal decisions for new entrants and for continuing residents, respectively, as a simultaneous system we have:

$$Y_{\{J_i t\}} = \mathcal{N}\left(Y_{\{J_i^o t-1\}}, Y_{\{J_i^e t\}}, \Psi_{\{J_i t\}}; \Omega_{\{n(i)t\}}, Z_{\{n(i)t\}}\right), \tag{3}$$

where $Y_{\{J_i t\}} = (Y_{\{J_i^e t\}}, Y_{\{J_i^o t\}})$. When construed for $\forall i \in J_i$ and under the assumption that all individuals observe the vector of neighborhood housing qualities, the mapping (3) defines an evolving Nash equilibrium for the vector of housing qualities in neighborhood J_i. That is, in a Nash equilibrium, the quality of each housing property is a function of the vector of shocks affecting all neighborhood properties and of the vector of individual wealths and socioeconomic characteristics in the neighborhood, and is conditional on dwelling unit qualities in the previous period. This equation may be rewritten in terms of property values, which are observable, rather than qualities, which are unobservable. Therefore, because of equilibrium in each neighborhood, the presence of continuing residents causes prices and thus property values in the neighborhood to reflect maintenance shocks and lagged property values. Furthermore, the inflow of new residents causes prices and thus property values to reflect neighborhood effects through newcomers' valuations.

Equ. (3) describes schematically the joint determination of neighborhood housing quality through neighborhood interactions. Several important conclusions follow from neighborhood Nash equilibrium. To name two, consider first that (3) implies a spatial dependence in the intertemporal evolution of individuals' wealth. Second, individuals' and their offspring's nonhousing wealth are interdependent through common schooling as well as other shared factors that may develop through physical proximity and are relevant to productivity.

I eschew a complete the description of the intertemporal evolution of neighborhood stochastic structure, and turn next to a qualitative description of the model. Equilibrium in each neighborhood is construed as conditional on new residents' having chosen a neighborhood because it offered higher utility than all of their alternative courses of action, and on old residents' having chosen to remain in a neighborhood because it offered higher utility than all of their alternative courses of action. Neighborhood composition based on choice is critical in understanding self-selection.

The occupants of each neighborhood cluster are not a random sample of the population. Or put differently, individuals' incomes in each neighborhood (and other characteristics) may be neither identical nor distributed according to the national income distribution.

In this paper, I take the composition of different neighborhoods as given. However, see Epple and Sieg (1999) for an approach to equilibrium neighborhood selection. A straightforward implication of that model is that even if individual utility is separable in the influence of the neighborhood housing stock for both new and continuing residents, in which case housing demand by new residents is independent of the neighborhood housing stock, selection introduces such dependence. This follows from the fact that associating neighborhood choice with utility comparisons implies bounds which themselves depend on neighborhood housing stocks. I also take housing prices as given.[12]

4 Data

The AHS is a panel of housing units, which was redesigned in 1985 and involves more than 50,000 dwelling units that are interviewed each two years. This paper explores an additional, and neglected (in spite of its rarity) dimension of the data, data on neighborhood clusters, which are available for years 1985, 1989, and 1993. In those years only, a random sample of originally 680 (and subsequently more) urban units were selected and for each one of them (up to) ten neighbor units were interviewed. Each such cluster includes the randomly chosen member of the national file (which is an urban AHS unit), the so-called *kernel,* and the ten homes closest to it [Hadden and Leger (1990), p. 1-51]. The cluster may contain fewer than 10 units if some could not be interviewed. Appendix A provides details on sample structure and data availability.

The empirical investigation reported here is based on data from the 1985 and 1989 waves of the American Housing Survey (AHS) data.[13] Only a small number of papers, that is, de Bartolome and Rosenthal (1996), Gabriel and Rosenthal (1996a, 1996b), Hoyt and Rosenthal (1997), Ioannides and Hardman (1998), Ioannides and Zabel (2000a,b) and Kiel and Zabel (1998) have utilized the AHS

[12] See Miyao (1978) and Durlauf (1996) for models where community-specific housing prices reflect the socioeconomic characteristics of their residents.

[13] I conducted extensive econometric analyses with data from the 1993 wave, as well, but at the end decided to report results with 1993. Basically, the greater increase in the number of observations from 1989 to 1993 over that from 1985 to 1989 may not be exploited, primarily because it cannot be translated into an increase in the number of data, as availability of retrospective information is restricted by 1989 data.

clusters data to date. The latter four papers involve the only previous uses of the 1993 clusters data. Ioannides and Hardman are exploring neighborhood income distributions. Kiel and Zabel compare the performance of clusters data against mean census tract-level attributes by utilizing (privileged) access to census-tract coding of the data. The present application by estimating a model of neighborhood *interactions* is a novel use of those data.[14]

I note, however, that because of its design, the randomly selected sample of kernels their immediate neighbors delivers a snapshot of clusters of urban dwellings in the United States. The data tell us nothing about proximity to urban/metro centers and to physical barriers; and the frequency of sampling over time is less than ideal. It does allow us to study the outcomes of several economic processes which evolve around spatial interactions. As such, it is a valuable setting for testing empirically fundamental aspects of local interaction, as it enables us to capture the essential feature of combined spatial and dynamic interdependence.

The total number of cross-sectional observations for 1985, 1989, and 1993, respectively, are: 7350, 8433, and 11293. I have a theoretical maximum of 27076 observations, of which at most $3 \times 7322 = 21966$ are available in a panel of dwelling units. However, selecting on the basis of regular interviews leaves us with 22446 observations over the three waves. More observations are deleted because they pertain to renters, who comprise about 45% of the data and are excluded from the study. Also, the need for retrospective information on units that are in the sample in both 1985 and 1989 further reduces the number of observations available for regressions.

The 1985 data contain observations mainly from 630 clusters (neighborhoods) of at most 11 units each. Additional observations come from larger clusters, making the total number of clusters equal to 680. Additional details on the structure of the data for 1989 and 1993, such as observation counts on new clusters, new households and new units, etc., and their geographic distribution are given in Appendix A. Additional units in existing clusters were included in 1989 to reflect additional units that had been added within the perimeter of the "neighborhood." By 1993, a maximum of 20 neighboring units were allowed per cluster.[15] Data are missing for a variety of reasons. Units

[14]Ioannides and Zabel (2000a,b) aim at estimating housing demand in the presence of social interactions, which requires use of additional data, beyond what the present paper is employing. Additional details on those papers are given further below.

[15]I am grateful to Barbara T. Williams, US Bureau of the Census, for this clarification.

may be vacant, about 10% in all waves. Even in occupied units, interviews could not be completed in some instances.

A basic set of descriptive statistics are given in Tables 1 and 2. Table 1 compares data from the Statistical Abstract of the United States and from my own processing of the AHS data for the purpose of establishing the representativeness of the AHS data. Table 2 reports descriptive statistics for the AHS data for all three waves of available data. Details on the construction of variables are given in Appendix B.

Referring to Table 2, the mix of socioeconomic characteristics of the members of neighborhoods is of particular interest to us. In 1985, 1989, and 1993, respectively, 84.1%, 83.2%, and 81.3% of the kernels have household heads who are White. When one looks at housing tenure, 55.5%, 55.2% and 51.5% of all kernels are owner-occupied, while the corresponding numbers for the entire sample are 54.0%, 54.0% and 53.1%.

Not surprisingly, the dispersion of the cluster-averaged data is smaller than that of the full sample. Still, this obscures a fair amount of heterogeneity across neighborhoods. While the mean value of household income for the kernels, which make up a random subsample of the main AHS sample of the U.S. population, and that of cluster means are very close to one another, as one would indeed expect, the dispersion is much larger than one would expect from statistical sampling theory. Roughly speaking, random samples of size ten should produce a standard deviation of roughly one-third of that of the kernels. The observed standard deviations are at least twice as much as that, which implies that the distribution of income within neighborhoods is much more dispersed than what random sampling would imply. This aspect of the data is in accordance with notions of self selection in neighborhoods and is explored further in Ioannides and Hardman (1998) and Ioannides (2000).

There is substantial turnover within the four-year span between two successive waves that I am working with. Moves, on one hand, are beneficial in making the sample more representative, in principle, because individuals reassess their information and units get revalued by the market. They do, on the other, cause sample selection problems. Because of the pattern of new entrants (clusters, units and individuals) there is actually little data left with a structure which may be amenable to estimation with panel techniques. After I had performed a number of econometric experiments with the two cross-sections that are available for estimating a dynamic model, I decided to present only one, with data from two successive periods, 1985 and 1989. Still, the period covered by the data offers some distinct advantages. Great real estate appreciation during the 1980's, gave way to depreciations during the late 1980's and the early 1990's, and both episodes exhibited pronounced regional variations.

5 Estimation of Models of Neighborhood Interaction

Following a (by now standard) typology of social interaction models proposed by Manski (1993), one may identify two types of *social* effects. An *endogenous social* effect is present, if an individual's behavior is affected by the *actual*, (or *expected*), behavior of her neighbors. This "keeping up with the Joneses" effect gives rise to a a so-called "social multiplier," through which, as Manski (1993) notes, policy intervention works to impact the behavior of an entire social group. Another type of social effect refers to agents' responding to the average (or some other measure of aggregation for the distribution) of various individual attributes of interest within the neighborhood, such as racial and ethnic composition of the neighborhood, neighborhood income distribution and the like.[16] This is the so-called *exogenous social,* or *contextual*, effect, whereby one cares about, or reacts to, one's neighbors' attributes, rather than one's neighbors' actions. Endogenous and exogenous social effects are often confused with one another.[17] There may also be a *correlated* effect among residential neighbors, if all dwelling units in a neighborhood tend to be occupied by individuals of similar socioeconomic characteristics.[18]

Let $y_{i\kappa ht}$ denote an endogenous variable, associated with a specific housing unit i in cluster κ; Y_t denotes the vector made up of all the $y_{i\kappa ht}$'s. I use this framework below with two alternative

[16]Such an effect could reflect a variety of motivations. E.g., the fact that a neighborhood may becoming occupied by higher income people is perceived as a good omen for a neighborhood's future, by higher income people, but a bad one, by lower income people.

[17]See for example, Manski's critique of Crane (1991). Crane poses an epidemic model of endogenous neighborhood effects, where dropout and childbearing behavior by teenagers is influenced by the frequency of such behavior within the neighborhood. However, Crane estimates a contextual-effects model, where a teenager's behavior is influenced by the occupational composition of her neighborhood.

[18]This may come about through selection, which as I argued above, may involve more than one characteristic. Consequently, selection may produce imperfect segregation in terms of, say, wealth or income. Similar could be an effect caused by response to an unobserved shock, such a change in the vicinity of the urban area, or by an unobserved individual characteristic.

endogenous variables, maintenance behavior and a household's own valuation[19] of dwelling unit i, located in neighborhood κ and occupied by household h at time t. According to (1), (2) and (3), $y_{i\kappa ht}$ may be specified as a function, in general, of the subvector of Y_t that is made up of the *endogenous* variables associated with h's neighbors, of a household's own socioeconomic characteristics, z_{ht}, and of a number of additional factors, such as variables reflecting socioeconomic characteristics of one's neighbors, conditional on neighborhood characteristics and on dwelling unit characteristics, $(x_{\kappa t}, q_{it})$. given and do not attempt to correct for sample selection bias associated with individual characteristics and neighborhood characteristics. [20] That is, I specify the empirical model corresponding to (1) and (2), given (3), as a function of the endogenous variable's own lagged value, $y_{i\kappa ht-1}$, of neighbors' housing consumption, $\Pi_i Y_t$, of own socioeconomic characteristics, z_{ht}, and of socioeconomic characteristics of neighbors conditional on neighborhood and dwelling unit characteristics, $E[z_{ht}|x_{\kappa t}, q_{it}]$:

$$y_{i\kappa ht} = \alpha + \mu y_{i\kappa ht-1} + \beta \Pi_i Y_t + \eta z_{ht} + \gamma E[z_{ht}|x_{\kappa t}, q_{it}] + u_{i\kappa ht}, \tag{4}$$

where Π denotes a known weighting matrix of dimensions $I \times I$ that defines spatial interaction (and is discussed further below), and Π_i is its ith row; α, β and μ denote scalar unknown parameters, and η and γ vectors of unknown parameters. The error term $u_{i\kappa ht}$ in the RHS of (4) captures the impact of unobserved factors, conditional on neighborhood and individual dwelling unit characteristics, for which I assume that:

$$E[u_{i\kappa ht}|x_{\kappa t}, q_{it}] = \delta_x x_{\kappa t} + \delta_q q_{it}, \tag{5}$$

where δ_x, δ_q, denote vectors of unknown parameters.

Referring again to the Manski typology, the term $\beta \Pi_i Y_t$ in the RHS of Equ. (4) reflects an *endogenous social effect.* Such a social effect is central to the notion of neighborhood effects: a person's behavior depends on the *actual* behavior of her neighbors. The term $\gamma E[z_{ht}|x_{\kappa t}, q_{it}]$ expresses a *contextual effect*, an exogenous social effect: given the characteristics $x_{\kappa t}$ of the neighborhood κ where unit i is located and unit i's own characteristics q_{it} , this term gives the effect of the distributions of variables of potential interest, like racial and ethnic composition, within the neigh-

[19]See Kiel and Carson (1990) for previous work on owner valuations.
[20]Ioannides and Zabel (2000b) pursues that line of inquiry.

borhood. The conditional mean of $u_{i\kappa ht}$, from (5), $\delta_x x_{\kappa t} + \delta_q q_{it}$, expresses *correlated effects*: units in the same neighborhood with characteristics $x_{\kappa t}$ and individual dwelling units with characteristics q_{it} tend to have similar unobserved individual characteristics. The term ηz_{ht} reflects the *direct* effect of the owner's characteristics upon the valuation of the dwelling they occupy, in part because of taste, income etc., or the decisions about maintenance that they make. In contrast, the term $\gamma E[z_{ht}|x_{\kappa t}, q_{it}]$ reflects the impact of the *expected* characteristics of occupants, conditional on cluster characteristics $x_{\kappa t}$, and unit characteristics q_{it}. Such dependence follows as an outcome of sorting features of the matching process of households with dwelling units, whereby individuals' interest in the socioeconomic profile of their neighborhood is mediated in the residential matching process. Unless matching is perfectly random, this expectation is likely to depend upon $(x_{\kappa t}, q_{it})$. As Manski (1993) emphasizes, if this is not present, $\gamma = 0$, then the remaining (endogenous) social effect may be readily identified. [See also footnote 26 below.]

This model combines certain features of Case (1992), who studies the propagation of innovation adoption, and Manski (1993), especially its spatial model, *ibid.*, p. 537, Equ. (7), who examines estimation problems for social interaction models. Unlike Case (1992), I work with continuous dependent variables. The spatial interaction model "implies that the sample members know who each other are and choose their outcomes only after having been selected into the sample " [Manski, p. 537]. In contrast to the principal model in the latter, in (4) social interactions are expressed in terms of *actual* behavior,[21] Y_t, instead of *expected* behavior of one's neighbors, conditional on observables $[x_{\kappa t}, q_{it}]$, $E[Y_t|x_{\kappa t}, q_{it}]$.

5.1 Spatial Interaction

The spatial weighting matrix Π, employed in Equ. (4) above, is block-diagonal of size $I \times I$, with elements in each row summing up to 1. Its entries are defined as

$$\pi_{ij} = \frac{1}{n(i) - 1}, \; \forall i, j \in n(i), i \neq j, \quad \text{and } \pi_{ii} = 0, \text{otherwise.} \tag{6}$$

The endogenous effect is generated within the neighborhood sample consisting of the kernel and its neighborhood cluster, rather than within the entire population from which the sample was drawn.

[21] However, one should not exclude the possibility that individual members of a cluster in our sample also interact with other individuals outside the cluster. Such influences must be treated as omitted variables.

[22]

5.2 Spatial Stochastic Structure

I refer to (4) and define the unobserved component of the correlated effect, $\epsilon_{i\kappa ht}$, as the deviation of $u_{i\kappa ht}$ from its mean, conditional on neighborhood and dwelling unit characteristics,

$$\epsilon_{i\kappa ht} = u_{i\kappa ht} - E[u_{i\kappa ht}|x_{\kappa t}, q_{it}] = u_{i\kappa ht} - (\delta_x x_{\kappa t} + \delta_q q_{it}).$$

Let ϵ_t and u_t denote the vectors of size I obtained by stacking up in the obvious way errors $\epsilon_{i\kappa ht}$, defined above, and $u_{i\kappa ht}$, defined in (4). Unfortunately, I may not define a richer stochastic structure, where we would distinguish a time-invariant unit-specific effect associated with unit i in neighborhood κ, and a time-invariant individual-specific effect associated with individual h, as individuals are not separately identified from units. Specific units are inseparably associated with their neighborhoods, and thus likely to share a individual effect that is common to all units belonging to the same neighborhood.

I assume that ϵ_t, the unobserved component of the correlated effect defined above, consists of a neighborhood interactions term and a random error,

$$\epsilon_t = \tau \Pi \epsilon_t + \varepsilon_t,$$

where ε_t is a $I \times 1$ vector of purely random errors, with $E(\varepsilon_t) = \iota_I 0$, ι_I is the unit column vector of size I, and $\text{Var}(\varepsilon_t) = \sigma_\varepsilon^2 \mathcal{I}$, where $\mathcal{I}$ is the unit diagonal matrix of dimension I. The

[22] As an example consider three kernels, $\kappa = 1, 2, 3$, with associated cluster sizes $n_1 = 3, n_2 = 4, n_3 = 3$. Variable neighborhood cluster sizes are due to missing values. The weighting matrix has size: $I = n_1 + n_2 + n_3 = 10$. In writing the respective matrix I assume that the vector Y is formed by stacking neighborhood by neighborhood, and within each neighborhood first the variables associated with kernels and then those of each kernel's neighbors. Specifically:

$$\Pi = \begin{bmatrix} 0 & 1/2 & 1/2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1/2 & 1/2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1/3 & 1/3 & 1/3 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1/3 & 0 & 1/3 & 1/3 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1/3 & 1/3 & 0 & 1/3 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1/3 & 1/3 & 1/3 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1/2 & 1/2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1/2 & 0 & 1/2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1/2 & 1/2 & 0 \end{bmatrix}.$$

term $\tau\Pi\epsilon_t$ has the interpretation that the error terms for all observations contain τ times the average error realized by all of each unit's neighbors. Spatial correlation in errors, represented here by the spatial autocorrelation coefficient τ, may be present when unobserved spatially correlated variables, possibly due to self-selection, affect the endogenous variable of interest. *Consistency of social interaction* requires $\epsilon_t = [\mathcal{I} - \tau\Pi]^{-1}\varepsilon_t$, provided that the matrix $[\mathcal{I} - \tau\Pi]$ is invertible.

Equ. (4) may be rewritten in vector form as:

$$Y_t = \alpha\mathcal{I}\iota_I + \mu Y_{t-1} + \beta\Pi Y_t + \eta Z_t + \gamma E[Z_t|X_t, Q_t] + \delta_x X_t + \delta_q Q_t + [\mathcal{I} - \tau\Pi]^{-1}\varepsilon_t, \qquad (7)$$

where the vector Y_t stacks the individual observations, and the matrices X_t, Q_t, Z_t are defined in terms of the respective vectors of characteristics $x_{\kappa t}, q_{it}, z_{ht}$ in the obvious way. Equ. (7) represents the endogenous variables as a system of simultaneous equations.[23] It expresses the condition for Nash equilibrium in neighborhood interactions as a structural form.

Under the Nash assumption that individuals take their neighbors' actions as given and that $[\mathcal{I} - \beta\Pi]$ is invertible, I solve (7) as a simultaneous system for Y_t to obtain:

$$Y_t = \alpha[\mathcal{I} - \beta\Pi]^{-1}\iota_I + [\mathcal{I} - \beta\Pi]^{-1}\mu Y_{t-1}$$

$$+ [\mathcal{I} - \beta\Pi]^{-1}[\delta_x X_t + \delta_q Q_t + \eta Z_t] + [\mathcal{I} - \beta\Pi]^{-1}\gamma E[Z_t|X_t, Q_t] + [\mathcal{I} - \beta\Pi]^{-1}[\mathcal{I} - \tau\Pi]^{-1}\varepsilon_t. \qquad (8)$$

After tedious but elementary transformations, [24] Equ. (8) may be transformed further to yield a reduced- form as follows:

$$Y_t = \frac{\alpha}{1-\beta}\iota_I + \frac{n-1}{n-1+\beta}\left[X_t'\delta_x + Q_t'\delta_q + Z_t'\eta + E[Z_t|X_t, Q_t]'\gamma\right]$$

[23] The first version of this paper preceded Moffitt (1998), who clarified the identification of social interactions models by reinterpreting Manski's model as a system of simultaneous equations.

[24] I note that $[\mathcal{I} - \beta\Pi]$ is block-diagonal, with blocks corresponding to neighborhoods ; for any neighborhood of size n, the respective block may be written as: $[\mathcal{I}_n - \beta\Pi_n] = \left(1 + \frac{\beta}{n-1}\right)\mathcal{I}_n - \frac{\beta}{n-1}\iota_n\iota_n' = \theta_1\mathcal{I}_n - \theta_2\iota_n\iota_n'$, where ι_n denotes the unit column vector of size n, $\mathcal{I}_n$ denotes the unit diagonal matrix of size $n \times n$, and θ_1, θ_2 are defined as: $\theta_1 \equiv 1 + \frac{\beta}{n-1}$, $\theta_2 \equiv \frac{\beta}{n-1}$. Working similarly with matrix $[\mathcal{I} - \tau\Pi]$ whose diagonal blocks are of the form $[\mathcal{I}_n - \tau\Pi_n]$, I define $\xi_1 = 1 + \frac{\tau}{n-1}$, $\xi_2 = \frac{\tau}{n-1}$ and write $[\mathcal{I}_n - \tau\Pi_n] = \xi_1\mathcal{I}_n - \xi_2\iota_n\iota_n'$. The inverses of those matrices may be written as follows [Case (1992)]:

$$[\mathcal{I}_n - \beta\Pi_n]^{-1} = \frac{1}{\theta_1}\left[\mathcal{I}_n + \frac{\theta_2}{\theta_1 - n\theta_2}\iota_n\iota_n'\right].$$

$$[\mathcal{I}_n - \tau\Pi_n]^{-1} = \frac{1}{\xi_1}\left[\mathcal{I}_n + \frac{\xi_2}{\xi_1 - n\xi_2}\iota_n\iota_n'\right].$$

$$+\frac{n-1}{n-1+\beta}\mu Y_{t-1}+\frac{n}{n-1+\beta}\frac{\beta}{1-\beta}\left[\mu\bar{Y}_{t-1}+\bar{X}_t'\delta_x+\bar{Q}_t'\delta_q+\bar{Z}_t'\eta+E[\bar{Z}_t|X_t,Q_t]'\gamma]\right]+\bar{\varepsilon}_t, \quad (9)$$

where $\bar{\varepsilon}_t=[\mathcal{I}-\beta\Pi]^{-1}[\mathcal{I}-\tau\Pi]^{-1}\varepsilon_t$, and vectors and matrices in the RHS of (9) with bars indicate, for each observation i, the average, within i's neighborhood $n(i)$, values of the entries in the ith row of each of the corresponding matrices. This may be simplified, again by elementary transformations, to yield:[25]

$$\bar{\varepsilon}_t=\frac{1}{\theta_1\xi_1}\left[\left(1+\frac{\theta_2}{\theta_1-n\theta_2}+\frac{\xi_2}{\xi_1-n\xi_2}\right)\varepsilon_t+n\frac{\theta_2}{\theta_1-n\theta_2}\frac{\xi_2}{\xi_1-n\xi_2}\bar{\varepsilon}_t^{\nu}\right], \quad (10)$$

where $\bar{\varepsilon}_t^{\nu}$ denotes the vector of size I obtained by replacing each component i of ε_t by the average value of ε_t among unit i's neighbors including itself. Inspection of the RHS of Equ. (10) suggests that its variance-covariance matrix may be written in terms of σ_ε^2, β, τ, and n, which is exogenous and fixed at 10.[26]

5.3 Identification and Estimation

Pausing first to summarize, I have expressed Nash equilibrium within each neighborhood by means of a system of equations in structural form, (7), and in reduced form, (9). Both those systems may be estimated with the AHS neighborhood clusters data, where care must be taken to allow for spatial stochastic dependence. Both those equations allow one to identify the effect of social interactions.

Inspection of the RHS of Equ. (9) reveals two terms, $\frac{n-1}{n-1+\beta}\mu Y_{t-1}$ and $\frac{n}{n-1+\beta}\frac{\beta}{1-\beta}\mu\bar{Y}_{t-1}$, which are due to the presence in Equ. (7) of the own lagged value of the dependent variable. These terms provide an additional route to the identification of μ and β.[27] I note that the presence of

[25] $\bar{\varepsilon}_t=\frac{1}{\theta_1\xi_1}\left[\left(1+\frac{\theta_2}{\theta_1-n\theta_2}+\frac{\xi_2}{\xi_1-n\xi_2}\right)\mathcal{I}+\frac{\theta_2}{\theta_1-n\theta_2}\frac{\xi_2}{\xi_1-n\xi_2}[\iota_n\iota_n']\right]\varepsilon_t$.

[26] Manski's result still applies, of course, in that separate identification of γ from β requires knowledge of $E[Z_t|X_t,Q_t]$. This follows from (7) by solving for the expectation of the average Y_t in each neighborhood conditional on (X_t,Q_t,Z_t). The coefficient of $E[\bar{Z}_t|X_t,Q_t]$ in that expression becomes

$$(\eta+\gamma)\left(\frac{n-1}{n-1+\beta}+\frac{n}{n-1+\beta}\frac{\beta}{1-\beta}\right)=\frac{\eta+\gamma}{1-\beta},$$

which agrees with Manski. By substituting back in the expression for $E[Y_t|X_t,Q_t,Z_t]$, I have that the coefficient of $E[\bar{Z}_t|X_t,Q_t]$ is equal to $\frac{\gamma+\beta\eta}{1-\beta}$, which agrees with Manski, again. As we argued above, $E[\bar{Z}_t|X_t,Q_t]$ depends upon properties of the matching mechanism in the housing market and is likely to depend upon (X_t,Q_t), possibly in a complicated non-linear manner, which itself may buy identification, as Brock and Durlauf (1999) emphasize. Ioannides and Zabel (2000b) pursue this line of research.

[27] Although not specifically derived, this result was clearly anticipated by Manski: "If one observes the process out of equilibrium, the recursive structure of [the dynamic version of the linear model] opens new possibilities for identification" [Manski, *op. cit.*, p. 540].

the lagged dependent variable in (9) naturally follows the neighborhood interactions model. An interesting implication of the model is that ratio of the coefficient of the neighborhood average to that of the own lagged value, $\frac{n}{n-1}\frac{\beta}{1-\beta}$, is independent of μ. This ratio exactly identifies the social interaction coefficient β, as the neighborhood size n is exogenous. It is larger the larger is β, and is not bounded upwards by 1. This confirms the critical role, alluded to above, of the presence of the own lagged value in the RHS of (4).

One can estimate the model by working either with Equ. (7) as a structural form, or with Equ. (9) as a reduced form, where the spatial interaction structure also contributes the stochastic structure of the error. Starting with Equ. (7), I note that if $\beta = 0$, then the corresponding social interaction terms vanish and only each unit's own regressors are present. In that case, the only influence of the spatial interaction structure is through the error structure, where the spatial interaction is present in the definition of the error according to (10), provided that $\tau \neq 0$. Specifically, if $\beta = 0$, then $\bar{\varepsilon}_t = \frac{1+\frac{1}{1-\tau}\frac{\tau}{n-1}}{1+\frac{\tau}{n-1}}\varepsilon_t$. That is, even in the absence of social interaction, spatial autocorrelation has the effect of magnifying the effect of the individual i.i.d. stochastic shocks ε_t. If τ, the spatial autocorrelation coefficient, is small, $\bar{\varepsilon}_t$ is a multiple of ε_t, with a factor of proportionality close to, but greater than, 1. The factor of proportionality is increasing in τ but the variance-covariance matrix is diagonal. Unfortunately, there is no way to identify τ in the absence of social interaction.

If social interactions are present, that is if $\beta \neq 0$, (10) implies that the variance-covariance matrix of the error structure in (9) is non-diagonal and it contains both β and τ. If spatial autocorrelation is absent, then the variance-covariance matrix is diagonal. GLS may be adapted in order to estimate τ from the variance-covariance matrix of neighboring units. If, on the other hand, β is close to 1, then the social interaction terms become dominant. Whereas the estimation of τ rests entirely on the error structure[28], estimation of β involves both the error structure and the coefficients of several RHS variables.[29]

[28]Unfortunately, it is not possible to identify the spatial autocorrelation coefficient without carrying out maximum likelihood estimations, which are extremely tedious in this setting. Therefore, the inclusion of the autocorrelation coefficient in the above model serves to highlight the spatial stochastic structure.

[29]That the covariance structure of (9) may also aid identification has been noted by Case (1992) and Moffitt (1998).

6 Empirical Results

I performed a number of econometric experiments along the lines of Equations (7) and (9). It is appropriate to summarize how those equations differ. Equation (7) is a structural form, where the dependent variable y_{hit} is a function of its own lagged value, y_{hit-1}, of the mean of the dependent variable among i's neighboring units (that is, the ith row of $\beta\Pi Y_t$), of individual h's own characteristics, ηz_{ht}, of the characteristics of unit i's neighborhood cluster, $\delta_x x_{it}$, of unit i's own characteristics, $\delta_q q_{it}$, and of the socioeconomic characteristics of the neighborhood cluster, conditional on cluster and unit characteristics, $\gamma E[Z_t|X_t, Q_t]$. Identification of the latter effect requires considering neighborhood choice and may not be handled by the techniques and data of this paper. That may only be accomplished by means of a richer data set that allows us to study matching of individuals with neighborhoods and dwelling units [Ioannides and Zabel (2000b)]. Therefore, I set $\gamma = 0$, and the social interactions effect β may be identified as the coefficient of the predicted mean of the dependent variable among a unit's neighboring units. This requires 2SLS estimation in the presence of correlated disturbances, the latter being induced the spatial stochastic structure, and is subject to the usual identification restrictions. Equ. (9), on the other hand, is a reduced form, where the dependent variable y_{hit} is a function of a similar set of regressors as in the case of the reduced form, except that identification of the social interactions effect now rests in part on the own lagged value coefficient. In the remainder of this section I report first results along the lines of Equ. (7), and then I turn to results obtained with Equ. (9).

An important feature of the present approach as a way to identify social interactions is that this model may benefit from a straightforward application of hedonic theory of housing markets [Rosen (1974)]. That theory suggests that different dwelling units are priced by the market according the valuations of their characteristics. Therefore, individuals' knowledge of market valuations implies identifying restrictions as follows. One could use estimates of hedonic regressions of unit valuations as functions of characteristics of dwelling units and of their neighborhoods and use them as first-stage regressions for the endogenous variable Y_t in the RHS of Equ. (7) in the case of the valuation model. In the second stage regressions, inclusion of own socioeconomic characteristics (which have been excluded from the first stage regression) and, in addition, of the own lagged value ensure identification. Normally, in the first stage of a 2SLS procedure, one regresses the endogenous variables against all exogenous variables.

I have chosen to report a set of results which are consistent with the basic intuition of the model as well as consistent with the data. The estimates along the lines of Equ. (7), reported in Table 3, are typical of the entire set of regressions I performed with both pairs of consecutive waves of data, 1985 to 1989 and 1989 to 1993. I have chosen to concentrate on the 1989 cross-section with retrospective information for 1985. This choice was dictated by the fair amount of turnover, in both units and households, and the addition of new clusters and dwelling units in 1989 and in 1993, relative to 1985 and to 1989, respectively, which is documented in the table of Appendix A. Table 3 reports estimation results for interactive regressions along the lines of the model in (7). Columns 1, 2, 3 and 4 report results for property valuations as of 1989, and columns 5, 6, 7 and 8 report results for maintenance spending from 1985 to 1989, all in logs.

A notable omission from the RHS of the above equations is prices. While, of course, a key element of the neighborhood equilibrium model (1)–(3), I have omitted prices from the estimated equations and from the remainder of my exposition, because I have not been able to find a useful set of price indices that performed reasonably well in the context of my data. Yet, as the endogenous variables are conditioned on own lagged values, an argument could be made that the model estimates *pseudo*-demand functions in a life cycle setting. I will invoke this basic intuition and refrain from formalizing this argument in terms of lifecycle theory.

I have performed extensive experimentation with hedonic regressions for property valuations and hedonic-type regressions for maintenance. The dependent variable of the valuation regressions is the self-reported value of owner-occupied dwellings. The dependent variable of the maintenance regressions is the reported costs of additions and repairs over the previous four years. Regarding the valuation regressions, both groups of regressors, cluster-specific variables, the X's, and dwelling-unit variables, the Q's, are important as explanatory variables. Those particular hedonic regressions condition only on cluster characteristics and dwelling unit characteristics and exclude individual occupant characteristics.[30] Some of the neighborhood characteristics are interpreted

[30] One could justify the presence of individual occupant characteristics in hedonic regressions as proxies for census tract characteristics as contextual effects. Unlike Kiel and Zabel (1998), I do not have access to information on census tracts to which clusters belong. Therefore, I cannot link the X_t variables to census tract level data. However, it is noteworthy that Kiel and Zabel find means and standard deviations of the cluster and census tract level neighborhood variables to be fairly similar and the correlation between the two sets of variables is .82. As they put it, "overall,

as exogenous social, or contextual, effects, like per cent of owners, of household heads who are White, and of vacancies in the neighborhood. Neighborhood (cluster) specific variables performed quite well in several regressions and imply nonlinear effects associated, in particular, with such variables as cluster-averages for race, and for duration of vacancies, for which I have estimated cubic polynomial structures. The natural use of hedonic theory to justify the first-stage regressions mitigates the need to regress the endogenous RHS variable against all exogenous variables in the model, as it is the usual approach in 2SLS. However, there is no obvious similar interpretation of the hedonic type regressions for the maintenance model. Nonetheless, I performed hedonic-type regressions for maintenance, which include cluster characteristics, unit characteristics and individual characteristics, and this is what I report in Table 3, Column 5.

The presentation of the results in Table 3 is organized according to groups of cluster-specific variables, the X's, of dwelling-unit variables, the Q's, and of individual-specific variables, the Z's. All of these groups are significant. In the first group of regressions, reported respectively in Columns 1 and 2, and in 5 and 6, I treat observations belonging to the same cluster as independent. However, as the preceding section makes clear, spatial interactions induce a stochastic structure within each cluster which may be naturally modelled by means of cluster-specific random effects. I also estimate cluster-specific fixed effects, and test those two stochastic structures, as well. Columns 3 and 4, and Columns 7 and 8 present results with cluster-specific effects: Columns 3 and 7 allow for fixed cluster-specific effects, and Columns 4 and 8 allow for random cluster-specific effects. All t statistics reported are robust with respect to heteroscedasticity associated with the neighborhood clusters.

The regressions in columns 2, 3 and 4, and 6, 7 and 8 include as an explanatory variable the average predicted value of the dependent variable among a unit's neighbors. These predicted values are used to instrument $\beta \Pi Y_t$ in the RHS of (7). They have been computed using the estimated coefficients from hedonic regressions like the ones reported in Columns 1 and 5, where a household's own characteristics have been excluded. Once both the own lagged value and the average predicted value of the dependent variable among each individual's neighbors in the cluster has been

there does not seem to be a great deal of difference between the cluster and the tract measures. ... When comparisons are made in a regression context, we find evidence that the cluster variables have greater explanatory power than do tract level variables." [*ibid.,* p. 23]. Also, Ioannides and Zabel (2000a) use occupant data from a larger sample of the AHS, the entire metropolitan sample, as proxies for neighborhood effects.

included as regressors, most other regressors associated with a unit's cluster are no longer significant. Unfortunately, this implies that I cannot estimate the effects of socioeconomic characteristics of one's neighbors as contextual effects.[31] Several structural characteristics of the unit and some household characteristics, like education of the household and household income, always remain very significant. All of the estimation results accord with intuition. The regressors used are defined in Appendix B. Their descriptive statistics are also given there.

I note, in particular, that when both the lagged dependent variable and the predicted mean value among each individual's neighbors in the cluster are included they are highly significant and improve the overall fit. In fact, the estimate of the effect of the latter is larger than that of the former, implying a more important role for the social interaction effect. This is particularly much more pronounced for the case of the maintenance decision where the social interaction effect is greater than the own lagged effect by an entire order of magnitude. Specifically, for the property valuation model the coefficients of the social interaction term are .651 and .716, respectively, for the fixed effects and the random effects model, with both being highly statistically significant. For the maintenance model, the coefficients of the social interaction term are .271 and .203, respectively, for the fixed effects and the random effects model, with both being highly statistically significant. Similarly, for the property valuation model the coefficients of the own lagged term are .301 and .157, respectively, for the fixed effects and the random effects model, with both being highly statistically significant. For the maintenance model, the coefficients of the own lagged term are .029 and .028, respectively, for the fixed effects and the random effects model, with both being statistically insignificant. I interpret these results as evidence of significant social interaction in neighborhoods, where individuals are affected by the maintenance behavior of their neighbors.

I note that while the t statistics I report are obtained from OLS, I have also tried to correct for the fact that the social interactions term is a predicted value. In principle, this could be done with 2SLS or simultaneous equations methods. Unfortunately, the model is very difficult to estimate by means of 2SLS with standard econometric packages. I have also tried to estimate the model as a simultaneous system, with the hedonic equations for all neighbors estimated at a first-stage

[31] This is data rather than a substantive problem. When another source of data is used for arriving at the predicted values of the dependent variable among each individual's neighbors, then Ioannides and Zabel (2000a) show that both endogenous social effects and contextual effects may be estimated.

regression and the valuation as a second-stage regression. Unfortunately, a very large number of parameters is involved while the degrees of freedom are substantially reduced because each cluster must be treated as an observation unit. While the full correction in the presence of individual effects is quite complicated, it turns out not to matter in this case, and the t statistics I report are actually accurate.[32]

Inclusion of fixed effects is highly significant for both regressions, which are reported in Table 3, Columns 3 and 7. Their inclusion in both models increases enormously the adjusted R^2 of the regression, from .577 to .9979, for the property valuation model, and from .067 to .451, for the maintenance model. This is, of course, to be expected.[33] However, and as I mentioned above, the random effects model fits the the stochastic structure better than the fixed effects model, that simply accounts for fixed cluster-specific characteristics. As is well known, fixed- vs. random effect specifications may be tested by means of the Hausman specification test. The Hausman test allows one to test the null hypothesis that the random effects are uncorrelated with the regressors, which is required by GLS theory. Under that null hypothesis, both the fixed effects and the random effects estimators are consistent, but the fixed effects model is inefficient. The test rests on the difference between the two estimators. The Hausman test does reject, in our case, the null hypothesis, very strongly for the valuation model but marginally for the maintenance model. I think that an appropriate interpretation of this rejection is that omitted variables in the specification of the model with random effects is the culprit, as the random effects model fits reasonably well. I should note that for the property valuation model, the within and between R^2's are .103 and .645, respectively, with the fraction of the overall variance that is due to the random effect is 0.410. Correspondingly, for the maintenance model, the within and between R^2's are .011 and .304, respectively, with the fraction of the overall variance that is due to the random effect is 0.014. I interpret the significance of the random effects model on its own as supporting the spatial stochastic structure that was introduced in the preceding section. I should note that the Hausman specification test may not used if cluster-specific socioeconomic characteristics are included as a source of contextual effects, because they may not coexist with cluster-specific fixed effects.

[32] See Ioannides and Zabel (2000a) for an explanation of the necessary correction.

[33] Fixed effects do make sense as a device to model unobserved components of social effects and have been used by previous authors, too. See Munshi and Myaux (1998) for a related application of fixed effects as social effects.

A particularly noteworthy result of the maintenance regressions is the performance of log income, which is highly significant and numerically large, .336 and .335, respectively from Columns 7 and 8. In contrast, the estimated coefficient for income for the property valuation model are .048 and .028, from Columns 3 and 4, and both statistically significant. These coefficients are a bit puzzling, because they may be interpreted as income elasticities of housing consumption for owners. In an effort to examine whether the numerically weak performance of income is due to the inclusion of the own lagged value in the property valuation regressions, I also estimated those models by excluding the own lagged values, and the results were quite similar.[34]

Next I turn to estimations along the lines of Equ. (9), which is the reduced-form counterpart of Equ. (7). Recall the discussion following Equation (9), which shows that estimates of the social interactions coefficient may be obtained from the estimates of the coefficients of the own lagged value and of the predicted mean of that variable among one's neighbors. I have also carried out such regressions, but report in Table 4 only a subset of the estimates, for reasons of brevity. For the purpose of comparison, I reproduce at the top portion of Table 4 the key results from the structural form estimates reported in Table 3. The fits obtained with the reduced-form regressions are not as good as those with the structural-form ones. Again, the random-effects model gives a better fit. From the ratio of the coefficients of $\bar{Y}_{85}$ and Y_{85}, which is equal to $\frac{n}{n-1}\frac{\beta}{1-\beta}$, the point estimates of $\hat{\beta}$ for the property valuation model are 0.179 and 0.209. These estimates[35] are obtained by using the average number of observations per cluster, $n = 6.9$. (This number differs from 10 considerably, primarily because of missing values and of the fact that observations for renters may not be used in these regressions). They are quite different from those for the social interactions coefficient obtained with the structural form. The maintenance model does not perform well at all in these regressions. While these results are clearly disappointing, they should not overshadow the important findings obtained with the structural-form model.

[34]The estimates of the income elasticity of housing valuations obtained here are very similar to those of Ioannides and Zabel (2000a), who aim at estimating a housing demand model with neighborhood effects.

[35]I have also tried, but failed, to estimate β by means of nonlinear least squares.

7 Conclusions

I explore a relatively neglected feature of data from the American Housing Survey for 1985, 1989 and 1993, namely the availability of data on neighborhood clusters in urban areas of the United States. This feature of the data allows me to estimate a model of social interactions at the neighborhood level. The concept of a neighborhood invoked here is quite literally that of a residential neighborhood that consists of a dwelling unit and its ten nearest neighbors. Therefore, these are novel results in the neighborhood effects literature. Most previous work is based on using contextual information associated with the census tract where a unit of observation belongs.

The model allows me to identify the effect of social interactions. Using a structural form equation, the impact of social interactions is found to be quite substantial, with the respective coefficient ranging from .651 to .716, for the property valuation model, and from .272 to .203, for the maintenance model. The social interaction effects is found to be more significant then that of the own lagged value. The results provide empirical support for the notion of residential neighborhood effects. That is, individual valuations of their properties and their maintenance behavior is influenced by those of their neighbors. As a positive finding, this may be interpreted as supportive of the notion, which was implicit in much of the literature in support of urban renewal, that public policy interventions may bring about urban neighborhood change through a social multiplier.

BIBLIOGRAPHY

Anas, Alex (1980), "A Model of Residential Change and Neighborhood Tipping," *Journal of Urban Economics,* 7, 358–370.

Benabou, Roland (1996), "Heterogeneity, Stratification, and Growth," *American Economic Review,* 86, 3, 584-609.

Binder, Michael, and M. Hashem Pesaran (1997), "Optimal Consumption Decisions under Social Interactions," working paper, University of Maryland, September.

Bond, Eric W., and N. Edward Coulson (1989), "Externalities, Filtering and Neighborhood Change," *Journal of Urban Economics,* 26, 231–249.

Brock, William A., and Steven N. Durlauf (1999), "Interactions –Based Models," Department of Economics, University of Wisconsin–Madison, March.

Case, Anne C. (1992), "Neighborhood Influence and Technological Change, " *Regional Science and Urban Economics,* 22, 491-508.

Coulson, N. Edward, and Eric W. Bond (1990), "A Hedonic Approach to Residential Succession," *Review of Economics and Statistics,* LXXII, 3, August, 433–444.

Crane, Jonathan (1991), "The Epidemic Theory of Ghettos and Neighborhood Effects on Dropping Out and Teenage Childbearing," *American Journal of Sociology,* 96, 1226-1259.

Davis, Otto A., and Andrew Whinston (1961), "The Economics of Urban Renewal," *Journal of Law and Contemporary Problems,* 26, 105-117.

de Bartolome, Charles and Stuart S. Rosenthal (1996), "Property Tax Capitalization in a Model With Tax Deferred Assets and Standard Deductions," working paper, University of British Columbia.

Durlauf, Steven N. (1996), "Neighborhood Feedbacks, Endogenous Stratification, and Income Inequality," in W. Barnett, G. Gandolfo and C. Hillinger, *Dynamic Disequilibrium Modelling,* Cambridge University Press.

Durlauf, Steven N. (1997), "Statistical Mechanics Approaches to Socioeconomic Behavior," 81–104, in Arthur, W. Brian, Steven N. Durlauf and David Lane, eds., *The Economy as a Complex Evolving System II,* Addison-Wesley, Reading, MA..

Ellickson, Bryan (1979), "Local Public Goods and the Market for Neighborhoods," in *The Economics of Neighborhoods,* Academic Press, New York.

Epple, Dennis, and Holger Sieg (1999), "Estimating Equilibrium Models of Local Jurisdictions," *Journal of Political Economy,* 107, 4, 645–681.

Gabriel, Stuart A., and Stuart S. Rosenthal (1996a), " Commute Times, Neighborhood Effects, and Earnings: An Analysis of Compensating Differentials and Racial Discrimination," *Journal of Urban Economics,* 40, July, 61-82.

Gabriel, Stuart A., and Stuart S. Rosenthal (1996b), " Location and the Effect of Demographic Traits on Earnings: The Tradeoff Between Omitted Variable and Simultaneity Bias," working paper, University of British Columbia.

Galster, George C. (1987) *Homeowners and Neighborhood Reinvestment,* Duke University Press, Durham, North Carolina.

Gladwell, Malcolm (1996), "The Tipping Point," *The New Yorker,* June 3, 1996.

Glaeser, Edward L., Bruce Sacerdote and Jose A. Scheinkman (1996), " Crime and Social Interactions, " *Quarterly Journal of Economics,* 507-548.

Hadden, Louise, and Mireille Leger (1990), *Codebook for the American Housing Survey,* Abt Associates, Cambridge, MA.

Hoyt, William H., and Stuart S. Rosenthal (1997), " Household Location and Tiebout: Do Families Sort According to Preferences for Locational Amenities?" *Journal of Urban Economics,* 42, 2, 159–178.

Ioannides, Yannis M. (1999), "Nonlinear Neighborhood Interactions in Local Communities and Intergenerational Transmission of Human Capital," working paper, Department of Economics, Tufts University, June.

Ioannides, Yannis M. (2000), "Neighborhood Income Distributions," working paper, Department of Economics, Tufts University, March.

Ioannides, Yannis M., and Anna M. Hardman (1998), "Income Mixing in Neighborhoods, Part I: Theory and Evidence from Neighborhood Clusters of the American Housing Survey," Department of Economics, Tufts University, June, presented at the AREUEA meetings, Chicago, 1998.

Ioannides, Yannis M., and Jeffrey E. Zabel (2000a), "Neighborhood Effects and Housing Demand," presented at the Conference on Empirics of Social Interactions, Brookings Institution, Washington DC, January.

Ioannides, Yannis M., and Jeffrey E. Zabel (2000b), "Interactions, Neighborhood Selection and Housing Demand," Department of Economics, Tufts University, in progress.

Kiel, Katherine A., and Richard T. Carson (1990), "An Examination of Systematic Differences in the Appreciation of Individual Housing Units," *The Journal of Real Estate Research,* 5, 3, 301-318.

Kiel , Katherine A., and Jeffrey E. Zabel (1997), "The Impact of Neighborhood Characteristics on House Prices: What Geographic Area Constitutes a Neighborhood?" Department of Economics, Tufts University, presented at the AREUEA meetings, Chicago, 1998.

Manski, Charles F., (1993) "Identification of Endogenous Social Effects: The Reflection Problem," *Review of Economic Studies*, 60, 531-542.

Miyao, Takahiro (1978), "Dynamic Instability of a Mixed City in the Presence of Neighborhood Externalities," *American Economic Review,* 68, 3, 454-463.

Moffitt, Robert A. (1998), "Policy Interventions, Low-Level Equilibria and Social Interactions," presented at the Colloquium for Social Dynamics, Brookings Institution, January.

Munshi, Kaivan, and Jacques Myaux (1998), "Social Effects in the Demographic Transition: Evidence from Matlab, Bangladesh," paper presented at the NEUDC Conference, Economic Growth Center, Yale University.

Pollak, Robert A. (1976), "Interdependent Preferences, " *American Economic Review,* 78, 745-763.

Rothenberg, Jerome (1967), *Economic Evaluation of Urban Renewal,* Brookings, Washington DC.

Schall, Laurence D. (1976), "Urban Renewal Policy and Economic Efficiency," *American Economic Review,* 66, 4, 612-628.

Schelling, Thomas. C., (1971), "Dynamic Models of Segregation," *Journal of Mathematical Sociology,* 1, 143-186.

Spivack, Richard (1991), "The Determinants of Housing Maintenance and Upkeep: A Case Study of Providence, Rhode Island," *Applied Economics,* 23, 639-646.

Stahl, Konrad O. (1974), " Market Failures and the Allocation of Resources in Cities," Ph. D. Dissertation, University of California, Berkeley.

Strange, William (1992), "Overlapping Neighborhoods and Housing Externalities," *Journal of Urban Economics,* 32, 17-39.

Tiebout, C. (1956), "A Pure Theory of Local Public Expenditure," *Journal of Political Economy,* 64, 416-424.

U.S. Bureau of the Census (1985), *American Housing Survey,* National Core and Supplement File, CD-ROM (recordable) for 1985, 1987 and 1989, and 1989, 1991, and 1993.

U.S. Bureau of the Census (1994), *Statistical Abstract of the United States,* 114th Edition, Washington, DC.

Werczberger, E., and J. Berechman (1988), "Incorporating Neighborhood Effects into Spatial Allocation Models," *Environment and Planning,* A, 20, 595-607.

Table 1:

Comparison of Incomes between American Housing Survey and National Data by Regions, 1985 and 1993

Sample: Kernels and Neighbors

U.S.- designated statistics are obtained from the *Statistical Abstract of the United States* [U.S. Bureau of the Census (1987; 1995)] and apply to the entire U.S. and regions, as appropriate, and not just urban areas. U.S. median housing costs and property values also apply to the entire U.S. and regions and are obtained from the AHS [U.S. Bureau of the Census (1985; 1993)]. All other statistics are based on the author's own processing of the American Housing Survey data [U.S. Bureau of the Census (1996)].

Year	1985					1993				
Regions	All SMSAs	Mid West	North East	South	West	All SMSAs	Mid West	North East	South	West
Summary Statistics										
Mean Income ($)	29410	26658	31140	28934	30928	37490	34085	41001	35893	39470
CV Income	.846	.818	.858	.859	.827	.854	.849	.850	.874	.819
Median Income ($)	23000	21700	24000	22145	24565	28248	26000	30075	26312	30336
U.S. Mean Income ($)	29066	28149	31146	27044	31475	41428	39442	45319	38249	45284
U.S. Median Income ($)	23618	23551	25485	21397	25782	31241	31400	33747	28441	33739
U.S. Median Monthly Housing Costs ($)	357	330	388	322	427	487	424	551	445	579
U.S. Median Property Values ($)	63211	45108	76224	47310	81913	86529	71898	116102	70376	134430

Table 2:

American Housing Survey: Descriptive Statistics

	Mean 85	Mean 89	Mean 93	Cv85	Cv89	Cv93
Cluster-averaged data, regular interview						
Household income ($)	29140	34282	36503	.574	.569	.557
CPI-Urban (all)	107.6	124.0	144.5			
Monthly rent ($)	347	423	485	.470	.496	.465
Property value ($)	76033	100599	105231	.628	.750	.693
CPI-Urban (housing)	107.7	123.0	141.2			
Household data (same units)						
Date head moved in (19 - -)	74.9	78.3	81.5	.155	.153	.153
Age of head (years)	48.52	49.30	49.68	.362	.355	.354
Highest grade (years)	12.53	12.77	12.94	.279	.267	.253
Race (%-age White)	84.1	83.2	81.3			
Household size	2.62	2.60	2.56	.571	.594	.579
Household income ($)	29549	35161	37499	.840	.844	.844
Dwelling unit data						
Number of rooms	5.47	5.50	5.50	.345	.336	.334
Unit area (ft^2)	1612.5	1621.2	1614.8	.586	.542	.543
Appreciation rate$_{t,t-1}$ (owners)		.061	.025		2.62	5.87
Monthly rent (renters)	323	405	465	.520	.522	.484
Property value ($, owners)	79684	107476	111546	.670	.788	.721

Table 3

Interactive Regressions for Owners, 1985 to 1989

Variable	LV_{89}	LV_{89}	LV_{89}	LV_{89}	$LMaint_{89}$	$LMaint_{89}$	$LMaint_{89}$	$LMaint_{89}$
Column	1	2	3	4	5	6	7	8
Mean	11.364	11.286	11.286	11.286	3.203	3.203	3.203	3.203
Observations	2968	2396	2396	2396	2967	2396	2396	2396
Number Clusters	348	348	348	436	348	348	348	348
Obs per cluster	6.9	6.9	6.9	6.9	6.9	6.9	6.9	6.9
R^2, within				.103				.011
R^2, between				.645				.304
R^2	.428	.578	.9979	.548	.5112	.051	.451	.069
F	101.98	105.91	22857		104.40	6.84	98.47	
MSE	.620	.520	.511	.396	2.69	3.74	3.72	2.98
S.D. of RE	.			.331				.361
Cluster FE/RE	No	No	FEs	RE	No	No	FE	RE
Hausman test			206.49				30.99	
Significance			.0000				.0738	
Intercept	11.80	-.519		-1.840	-11.52	1.226		-24.52
	(3.58)	(.98)		(1.18)		(.65)		(1.94)
LV_{85}		.301	.301	.157		-.057	-.031	-.010
		(4.48)	(3.60)	(10.80)		(.54)	(.33)	(.09)
$LMaint_{85}$						.043	.029	.028
						(2.07)	(1.39)	(1.33)
Cluster data — X Variables								
Pred. Mean		.612	.651	.716		.815	.272	.203
neighbors		(11.05)	(9.64)	(17.43)		(5.65)	(3.59)	(3.05)
CC-SMSA	.293	-.005			.571	.009		
	(7.14)	(.13)			(3.19)	(.03)		
Suburb-SMSA	.391	.037			.471	-.053		
	(9.71)	(0.99)			(2.68)	(.19)		
Region-NE	.635	.060			.247	.094		
	(17.74)	(1.47)			(1.59)	.38		

Table 3 Continued

Variable	LV_{89}	LV_{89}	LV_{89}	LV_{89}	LMaint$_{89}$	LMaint$_{89}$	LMaint$_{89}$	LMaint$_{89}$
Column	1	2	3	4	5	6	7	8
Region-S	.034				.498	.242		
	(0.77)				(2.60)	.85		
Region-W	.520				.432	.173		
	(13.20)				(2.52)	(.66)		
Degrees	.018				-.025	-.067		
	(1.42)				(.46)	(.78)		
Δ Own $_{t-1,t}$						.111		
						(.21)		
Own$_t$	-1.574				6.842			
	(.56)				(.56)			
Own$_t^2$	.295				-2.199			
	(.38)				(.64)			
Own$_t^3$	-.011				.232			
	(.15)				(.75)			
Head White	-.392				1.313			
	(2.41)				(1.85)			
Head White^2	.386				-.823			
	(4.56)				(2.22)			
Head White^3	-.063				.114			
	(5.58)				(2.30)			
ΔL Race$_{t,t-1}$		-.109				-.108		
		(2.16)				(.39)		
Vacant	.310				.994			
	(.92)				(.68)			
Vacant^2	-.838				-1.275			
	(1.45)				(.51)			
Vacant^3	.331				.437			
	(1.43)				(.43)			
ΔL Vacant$_{t,t-1}$		.023				.052		
		(1.74)				(.62)		
L Quality index					-.060	-.009		
					(1.38)	(.13)		

Table 3 Continued

Variable	LV_{89}	LV_{89}	LV_{89}	LV_{89}	$LMaint_{89}$	$LMaint_{89}$	$LMaint_{89}$	$LMaint_{89}$
Column	1	2	3	4	5	6	7	8
Dwelling Unit Data — Q Variables								
Age	-.001	.012	.012	-.055	.255		.489	.545
	(.40)	(.59)	(.42)	(2.46)	(3.01)		(4.17)	(4.13)
Not detached	-.252	-.120	-.137	-.043	.145		.345	.413
	(4.67)	(2.17)	(1.89)	(.66)	(.61)		(1.01)	(1.07)
Unit area	.0002	.146	.149	.171	.0001		.011	.051
	(11.75)	(3.76)	(3.13)	(6.13)	(.91)		(.05)	(.24)
Rooms	.062	.019	.018	.049	.056		.163	.157
	(6.94)	(2.39)	(1.56)	(6.26)	(1.40)		(2.58)	(2.55)
Baths	.176	.050	.040	.046	-.084		.096	.097
	(8.25)	(2.35)	(1.49)	(2.43)	(.90)		(.68)	(.67)
Additions	.043	.050	.050	.030	3.237			
	(3.05)	(3.41)	(3.75)	(2.62)	(52.37)			

Table 3 Continued

Variable	LV_{89}	LV_{89}	LV_{89}	LV_{89}	$LMaint_{89}$	$LMaint_{89}$	$LMaint_{89}$	$LMaint_{89}$
Column	1	2	3	4	5	6	7	8
Occupant Household Data — Z Variables								
Moved in since 1985		-.022 (.78)	-.026 (.89)	.022 (.84)		-.393 (1.82)	-.390 (1.77)	-.285 (1.29)
Age		.033 (.68)	-.413 (1.48)	.812 (1.05)	.531 (2.83)	-.982 (2.93)	-2.088 (1.88)	10.22 (1.58)
Age ^{2}			.056 (1.46)	-.098 (.98)			.128 (.79)	-1.450 (1.74)
Head White		-.066 (1.53)	-.078 (1.23)	-.043 (1.05)	.084 (.32)	.110 (.41)	.210 (.81)	.258 (.98)
Education		.014 (3.08)	.014 (2.76)	.008 (2.31)	.008 (.47)	.001 (.04)	-.0004 (.02)	-.006 (.20)
HH Size		-.009 (1.02)	-.008 (.78)	-.009 (1.07)	.035 (.79)	.078 (1.17)	.043 (.60)	.031 (.45)
Head married		.027 (.75)	.028 (.75)	.045 (1.68)	.138 (.95)	.227 (1.00)	.267 (1.12)	.247 (1.09)
Head male		-.058 (1.67)	-.060 (1.60)	-.051 (1.93)	-.091 (.63)	-.074 (.33)	-.091 (.40)	-.085 (.38)
Cars		.008 (.55)	.009 (.62)	-.009 (.77)	.027 (.41)	-.005 (.05)	.004 (.40)	-.015 (.16)
Income		.048 (3.17)	.048 (2.80)	.028 (3.02)	.148 (3.14)	.337 (4.39)	.336 (5.24)	.335 (4.36)

Table 4:

Alternative Estimates of Parameter β

1	2	3	4	5
	FE	RE	FE	RE
	Property Valuation		Maintenance	
Structural Form Equ. (7)				
μ	.301	.157	.029	.028
	(3.60)	(10.80)	(1.39)	(1.33)
β	.651	.716	.272	.203
	(9.64)	(17.43)	(3.59)	(3.05)
Reduced Form Equ. (9)				
Y_{85}	.106	.172	.000	-.001
	(6.98)	(11.68)	(0.02)	(0.08)
$\bar{Y}_{85}$	.027	.053	.010	.011
	(1.39)	(2.94)	(.011)	(0.21)
Income_{89}	.021	.036	.189	.173
	(2.05)	(3.78)	(2.84)	(3.11)
$\overline{\text{Income}}_{89}$	.021	.070	.216	-.094
	(0.44)	(1.91)	(0.69)	(.57)
R^2 overall	.3111	.5679	.5007	.5340
R^2 within	.1248	.1076	.5088	.5045
R^2 between	.3614	.7228	.4114	.6472
S.D. of RE		.295		.228
Implied $\hat{\beta}$	.179	.209		

APPENDIX A: American Housing Survey Observation Counts

These are based on the author's own calculations with the AHS data from National Core and Supplement CD-ROM for 1985, 1987 and 1989, and for 1989, 1991, and 1993. A few discrepancies were found when checking the data for 1989 from the two products. Information only from the later source of data is used. Regular interviews are defined as those with ISTATUS=1.

N-M denotes non-missing. The terms head and reference person are used interchangeably in the documentation.

CC-SMSA: central city, SMSA. Suburb: urbanized, plus other urban plus rural. UA-NM: urban area, non-metro. RA-NM: rural area, non-metro.

Categories	Total	1985	1989	1993
Observations	27076	7350	8433	11293
Regular interview	22446	6215	7024	9207
New units in sample			886	2218
Of these, in new clusters			857	2186
Of these, new units in existing clusters			25	24
Lost units			73	28
New households in sample			2239	2443
Clusters		670	805	1014
Regular interview		581	651	812
New in sample			93	245
Region–Northeast	6376	1674	2232	2470
Regular interview		1411	1859	2090
Region–Midwest	5785	1566	1717	2502
Regular interview		1373	1476	2071
Region–South	8493	2365	2392	3736
Regular interview		1941	1889	2873
Region–West	6422	1745	2092	2585
Regular interview		1490	1800	2173

Urban-rural Geography, 1993	Total	CC-SMSA	Suburb	UA-NM
Region–Northeast	2485	1167	1194	78
Regular interview		976	1035	70
Region–Midwest	2526	1152	919	410
Regular interview		926	809	336
Region–South	3759	1689	1496	542
Regular interview		1290	1190	393
Region–West	2597	1129	1231	219
Regular interview		934	1071	168

APPENDIX B: Definition of Variables and Descriptive Statistics

The first group of regressors pertain to cluster-specific information. These are the X_t variables in the discussion of the model. They are defined as follows. CC-SMSA denotes whether observation belongs to a central city of a Standard Metropolitan Statistical Area. Suburb-SMSA denotes whether observations belongs to a suburb of a Standard Metropolitan Statistical Area. The variables Region-NE, Region-S, and Region-W denote whether observation belongs, respectively, to the Northeastern, Southern or the Western regions of the US, as defined by the US Bureau of the Census. Degrees measures heating degree day indicates an additional geographical detail. Own is the logarithm of the average rate of ownership in the neighborhood cluster. Similarly, Head White is the logarithm of average number of owners in the cluster, and Vacant is the logarithm of average vacancy rate in the cluster. Quality index is a categorical variable that indicates neighborhood conditions, as assessed by the respondent, with greater values indicating worse conditions.

The second group of regressors pertain to dwelling unit characteristics. These are the Q_t variables in the discussion of the model. Age is the age of the dwelling unit in years. Not detached is a dummy variable indicating whether a unit is not detached. Unit area is the square footage of the dwelling unit. Rooms the number of its rooms, and Baths that of its bathrooms. Additions is a dummy variable indicating that the owner has performed renovations that have added to the size of the unit.

The third group of observations are the characteristics of the household that owns the dwelling unit, and its head, if appropriate. These are the Z_t variables in the discussion of the model. Moved since 1985 is a dummy variable indicating whether the household observed has moved into the dwelling unit since 1985. Age is the head's age in years, Head White is dummy variable indicating whether the household head is White. Education is the head's schooling in years. HH Size is the size of the household. Head Married is a dummy variable indicating whether the household head is married and Head Male whether it is male. Cars is a dummy variable indicating whether the household owns cars and is intended to measure wealth. Income is the logarithm of the household's total income.

The table that follows reports all variables in levels and key variables in logarithms, as well.

Variable	Observations	Mean	Standard Deviation	Min	Max
Additions,Log, 1985	2942	.4133	.755	0	5
Additions,Log 1989	2942	.5228	.833	0	5
Age of Head 1989	2942	53.40	15.98	16	91
Age of dwelling	2942	34.06	19.56	0	75
Baths 1989	2942	1.57	.702	0	10
Cars 1989	2942	1.540	.876	0	7
Central city of MSA, 1989	2942	.360	.480	0	1
Cost of Additions, 1985–89 ($)	2942	1604.5	3576.7	0	45488
Cost of Additions, Log, 1985–89	2942	3.206	3.849	0	10.73
Degrees, 1989 (categorical,1–6)	2942	3.17	1.38	1	6
Dwelling has heat,1989	2942	.994	.078	0	1
Dwelling not detached	2942	.942	.234	0	1
Education of Head, years	2942	13.27	3.33	0	18

Variable	Observations	Mean	Standard Deviation	Min	Max
Head Male, =1, if yes	2942	.745	.436	0	1
Head Married, =1, if yes	2942	.674	.468	0	1
Head White =1, if yes	2942	.892	.310	0	1
Income, 1989	2942	43,616	32,707	0	400,000
Income, log, 1989	2941	10.32	1.219	0	12.90
Moved since last year, 1989	2942	.069	.254	0	1
Moved here since 1985	2942	.273	.446	0	1
Number of rooms, 1989	2942	6.410	1.677	1	17
Number of vacant units in cluster, 1989	2942	.310	.958	0	6.95
Other urban, in MSA, 1989	2942	.110	.313	0	1
Persons in household	2942	2.73	1.44	1	11
Quality index, 1989	2942	.930	1.218	0	9
Region Midwest	2942	.236	.424	0	1
Region Northeast	2942	.213	.410	0	1
Region South	2942	.290	.454	0	1
Region West	2942	.261	.439	0	1
Suburb of MSA, 1989	2942	.530	.500	0	1
Unit area, sf, 1989	2942	1905.6	847.5	100	4000
Value, 1985, $	2397	84,409	54,284	0	250,000
Value, log, 1985	2397	11.12	.835	0	12.43
Value, 1989 $	2942	115,686	88,023	1000	350,000
Value, log, 1989	2942	11.37	.816	6.91	12.77
Value, log, predicted of neighbors, 1989	2942	11.37	.502	9.777	12.774