INTERACTIVE PROPERTY VALUATIONS

Yannis M. Ioannides[1]

Department of Economics

Tufts University

Medford, MA 02155

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

This version: June 19, 2002

Abstract

This paper estimates models of social interactions within residential neighborhoods using data on neighborhood clusters for standard metropolitan areas in the United States from the American Housing Survey for 1985 and 1989. It examines effects of social interactions in the form of reaction functions for homeowners' valuation of their properties at the level of the *immediate residential* neighborhood, with neighborhoods consisting of a randomly chosen dwelling unit and about ten nearest neighbors.

The paper identifies the effect of endogenous social interactions to be significant and large, ranging from 0.587 to 0.770, and much more important then the dynamic (autoregressive) structure of the model when both variables are present and both are significant. The interactive regressions that it reports improve upon commonly used hedonic regressions as well. The paper provides empirical support for the notion, common in the real estate world, of the importance of neighboring properties in property valuations.

JEL classification codes: R21, C39
Keywords: Neighborhood effects, housing, social interactions, property valuations.
Suggested Running Head: Interactive Property Valuations.

[1]I thank two referees for very insightful comments. I am especially indebted to the editor, Jan Brueckner, for numerous corrections and suggestions that helped me improve the paper. 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, Stuart Rosenthal, Holger Sieg, Jeff Zabel, participants at the 1997 ESEM (Toulouse), and seminar participants at Brown and Harvard for comments to earlier versions of the paper, and Pamela Jia, Annie K. Thompson and Hui Zhang for excellent research assistance. I am grateful to the National Science Foundation and the John D. and Catherine T. MacArthur Foundation for research support.

1 Introduction

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.[2] 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 capital gains due to proximity to other valuable property or 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 likely benefit from careful attention to the dynamics of interaction within residential neighborhoods. Residential capital is important: 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]

This paper estimates models of social interactions within residential neighborhoods using data on neighborhood clusters for standard metropolitan areas in the United States from the American Housing Survey for 1985 and 1989. It examines effects of social interactions in the form of reaction functions for homeowners' valuation of their properties at the *immediate residential* neighborhood level, with neighborhoods consisting of a dwelling unit and about ten nearest neighbors. The paper identifies the effect of endogenous social interactions to be significant and large, ranging from 0.587 to 0.770, and much more important then the dynamic (autoregressive) structure of the model when both variables are present and both are significant. The interactive regressions that it reports improve upon commonly used hedonic regressions as well. It thus provides empirical support for the notion, common in the real estate world, of the importance of neighboring properties in property valuations using a novel but natural empirical setting.

The paper relies on data from the American Housing Survey (AHS) (U.S. Bureau of the Census [51]), which are collected from a *panel* of dwelling units and their current occupants. It makes use of a little known feature of the survey: for roughly one out of a hundred of dwelling units sampled,

[2]Economists have shown interest in the phenomenon of the formation of jurisdictions, especially as reflected in the Tiebout model [Tiebout [50]]. Benabou [3] and Durlauf [14] have reconsidered the fundamental underpinnings of this model with special emphasis to inequality. Lack of empirical attention to key ideas underpinning Tiebout's theory would have been astonishing were it not for a recent revival of interest recently; see Hoyt and Rosenthal [26].

[3]Source: Statistical Abstract of the United States [52]), T. 1213 and 866. See also, T. 745.

up to ten of their nearest neighboring units are also sampled. The literal notion of a *residential* neighborhood may be central to a variety of social interactions. A plethora of phenomena, such as individuals' attitudes towards race, income inequality, crime and ethnicity factors are both causes and effects of the composition of individuals' immediate physical and human environment. [4]

I examine empirically the extent in which individuals' valuations of their own properties depend upon those of their neighbors. I distinguish between the impact of characteristics of neighboring dwelling units and of characteristics of their occupants from those of one's neighbors' valuations. When neighborhood-average property values are included as regressors as well, the latter emerge as a more important determinant than own lagged values. This result suggests, as we shall see, that endogenous social 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. The findings pertaining to the specific ways in which neighborhood-average magnitudes matter is robust to various changes in the specification. These results are novel in the context of social interactions literature as well.

Most of the research to date that employs contextual effects has used geographic detail which is no smaller than census tracts. A fair amount of research uses data for Metropolitan Statistical Areas (MSAs, for short), which is a much larger geographical unit. Both census tracts and MSAs are arguably too large for studying phenomena of the sort that are emphasized in this paper, like neighborhood interactions. Data on individual dwelling units, their occupants and their immediate neighbors allow us a glimpse at the workings of many processes which are likely to be averaged out at higher levels of aggregation.

The remainder of the paper is organized as follows. Section 2 discusses the literature on urban neighborhood interactions. Section 3 discusses the data and Section 4 develops econometric models for estimating the behavioral model in the presence of neighborhood interactions. Section 5 presents our empirical results and section 6 concludes.

[4] Recent research has provided theoretical foundations for an understanding of the emergence of a variety of economic institutions from local interactions; see Durlauf [15]. Glaeser, Sacerdote and Scheinkman [22] examine social-interactions based explanations of the incidence of crime. Gladwell [21] highlights an epidemic theory of crime explanation of the recent decrease in crime.

2 The Literature on Urban Neighborhood Interactions

Neighborhood interactions have attracted relatively little attention. 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* [49], p. 1901]. A pioneering piece by Pollak [41] emphasizes the empirical implications of the assumption that preferences of individual members of a group are interdependent, though not necessarily in a neighborhood context, the more recent literature has invoked Nash equilibrium of strategic interactions. Strange [48], for example, 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 [4] adapts a linear-quadratic version of Pollak's model for the presence of social interactions and show that under certain conditions the model's predictions imply equivalence with the case of self-centered individuals.

Within 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. Of course, a neighborhood may evolve around a local public good. Ellickson [16] 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 assumes initial economies of scale, which are exhausted after some minimum neighborhood size. 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. Werczberger and Berechman [53] 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 [38], who addresses household location choice and stability properties of mixed-city equilibrium in the context of city-wide interactions.

There has been some empirical research on the economic consequences of social interactions on individuals as they pass through various neighborhoods. Kremer [35] finds significant linear effects on individuals' education from average education in the census tracts where individuals grew up, and Ioannides [29] significant nonlinear ones, as well. With the notable exception of Coulson and Bond [10], empirical research on the impact of neighborhood effects on residential succession is very limited. These authors test a model, due to Bond and Coulson [5], 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 [2] models the behavior of suppliers in the presence of exogenous neighborhood effects.[5]

Manski [36] shows how difficult it is to distinguish, by relying entirely on data, among alternative models of interaction, i.e. situations where individuals' actions appear to be in response to their neighbors' *actions* rather than their neighbors' *characteristics.* Therefore, even if suitable data were available, careful statistical analysis may have to go beyond just appending statistical descriptions of a person's neighborhood to her own personal economic and social characteristics and then just running regressions.

A number of studies of urban neighborhood interactions originate in the context of evaluating the urban renewal projects of the 1960's. Davis and Whinston [13], Rothenberg [42] and Schall [43] study housing maintenance behavior. Stahl [47] is an exhaustive study of the consequences of neighborhood effects for replacement/rehabilitation of housing and housing maintenance. Spatial proximity is important in understanding neighborhood dynamics [Strange [48]], including such phenomena as neighborhood tipping, where except for the path-breaking work by Schelling [44] there is little analytical work that can be used to structure empirical investigation. There appears to be no research in the hedonic analysis if housing markets literature on interactive property valuations, perhaps because of the lack of data. A recent exhaustive survey by Sheppard [45] mentions

[5] I am aware of two other works that examine neighborhood interactions empirically, both of which use local data. Galster [20] 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. 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: "There is no evidence that lower income or black homeowners are less likely to undermaintain their homes' exteriors than are higher-income or white homeowners. ... Yet, the dynamics of succession and transition can generate expectations ... that the physical and/or the socioeconomic quality of the neighborhood will fall." [Galster *op. cit.,* , p. 240.], and that evidence would exonerate the behavior of the "in-migrating" households as the key cause of neighborhood deterioration. Spivack [46] uses data on code violations from Providence, Rhode Island, and finds some impact of neighborhood variables: ownership patterns and vacancies are the most influential determinants of maintenance and upkeep.

spatial interactions only in the context of the possible importance of spatial autocorrelation [*ibid.,* p. 1618], but cites no research on social interactions as such. The only related paper is Kiel and Zabel [34], except that their emphasis is on comparing the relative performance of cluster-level vs. census-tract level variables by means of reduced-form models only. The present paper is closely related to Ioannides [30], which estimates models of interdependence of maintenance decisions among neighbors, and to Ioannides and Zabel [31] and [32], which estimate models of housing demand, when demand decisions by neighbors are interdependent.[6]

A lively literature has addressed issues of strategic interactions at higher levels of aggregation, like among local governments. For an example, see Brueckner [7] who examines urban growth controls as a case in point, and Brueckner [8] for a comprehensive methodological review of the literature on empirical studies of strategic interactions among governments.

3 Data

The American Housing Survey (AHS) is a panel of dwelling units, which was redesigned in 1985 to involve more than 50,000 dwelling units that are interviewed each two years. This paper explores an additional but neglected feature of the data, that is, data on neighborhood clusters, which are available for years 1985, 1989, and 1993 [51]. In those years only, a random sample of originally 680 — and subsequently more, as we will see shortly below — 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 [24], p. 1-51]. The cluster may contain fewer than 10 units, because of the pattern of urban development or non-response. Appendix A provides details on sample structure and data availability. The empirical investigation reported here is based only on data from the 1985 and 1989 waves of the AHS data.[7]

[6]Only a small number of papers, including Coulson *et al.* [11], Gabriel and Rosenthal [18], [19], Hoyt and Rosenthal [26], Hardman and Ioannides [25], Ioannides [27], Ioannides [30], Ioannides and Zabel [31] [32] and Kiel and Zabel [34] have utilized the AHS clusters data to date. The latter four papers involve the only previous uses of the 1993 clusters data. Hardman and Ioannides 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. Ioannides and Zabel [31], [32] aim at estimating housing *demand* in the presence of social interactions, which requires use of additional data, beyond what the present paper is employing.

[7]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

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.[8]

Data are missing for a variety of reasons. Units 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 Table 1. Details on the construction of variables are given in Appendix A.[9]

Referring to Table 1, the mix of socioeconomic characteristics of the members of neighborhoods is of particular interest. 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. 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 suggests considerable heterogeneity within neighborhoods and does not contradict self-selection in neighborhoods (Hardman and Ioannides [25] and Ioannides [27]).

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

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.

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

[9] See also Ioannides [30], Table 1, who 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. Appendix A, in *ibid.,* gives additional information on the data.

principle, because individuals may 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 pair of successive 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.

4 Estimation of Neighborhood Interaction

As this paper is purely empirical, I postulate that the valuations of property values within small residential neighborhoods are interdependent. There are many reasons why this might be so. First, different neighbors' maintenance decisions may reflect, in part, decisions by their neighbors, that is, individuals might react directly to maintenance decisions of their neighbors or to the outcome of such decisions. That is, in a Nash equilibrium, the value of each property is a function of the vector of shocks affecting all neighborhood properties and of the vector of individual wealths. One would expect that, typically, residential neighborhoods would include neighbors of various "vintages," that is, households that have moved into the neighborhood at different times. Also, neighborhoods are mixed in terms renters and owners. While renters make decisions in terms of current prices and conditions, owners' valuations are likely to be serially correlated. Therefore, by 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. Nash equilibrium in each neighborhood reflects the fact that new residents have chosen a neighborhood because it offered higher utility than all of their alternative courses of action. Similarly, old residents have 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. Therefore, 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 (Ioannides [27]). In this paper, I take the composition of different neighborhoods as given.[10] I also take housing prices as given.[11]

Following a (by now standard) typology of social interaction models proposed by Manski [36] and [37], 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 so-called "social multiplier," through which, as Manski [36] 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.[12] 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. Distinguishing between endogenous and exogenous social effects is important. [13] 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.[14]

Let $y_{i\kappa ht}$ denote the valuation by individual h who occupies specific housing unit i in cluster κ, $\kappa = 1, \ldots, K_t$, at time t; Y_t denotes the vector made up of all the $y_{i\kappa ht}$'s, the vector of endogenous

[10] A straightforward implication of a residential sorting model, like that of Epple and Sieg [17], 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.

[11] Miyao [38] and Durlauf [14] offer models where community-specific housing prices reflect the socioeconomic characteristics of their residents. Ioannides and Zabel [32] explore empirically the extent in which prices reflect such characteristics.

[12] 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.

[13] See for example, Manski's critique of Crane [12] regarding confusion over those two types of effects. 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.

[14] 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.

variables here.[15] The above intuitive discussion suggests that $y_{i\kappa ht}$ 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 vector 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 a neighborhood's socioeconomic and geographic characteristics, and on dwelling unit characteristics, $(z_{ht}|x_{\kappa t}, q_{it})$. I take these characteristics as given and do not attempt to correct for sample selection bias associated with individual characteristics and neighborhood characteristics. [16] The previous discussion allows me to specify the empirical model so as $y_{i\kappa ht}$ is 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, $E[z_{ht}|x_{\kappa t}, q_{it}]$:

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

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. With the data at my disposal, I cannot measure the term $E[z_{ht}|x_{\kappa t}, q_{it}]$, cannot identify γ, and therefore set $\gamma = 0$. The error term $u_{i\kappa ht}$ in the RHS of (1) captures the impact of factors, over and above observable ones, which are observed by individuals but unobserved by the econometrician, which I assume to be conditional on neighborhood and individual dwelling unit characteristics:

$$E[u_{i\kappa ht}|x_{\kappa t}, q_{it}] = \delta_x x_{\kappa t} + \delta_q q_{it}, \tag{2}$$

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. (1) 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 $\gamma E[z_{ht}|x_{\kappa t}, q_{it}]$ would have expressed a *contextual effect*, that is, 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 would give the effect of

[15] See Kiel and Carson [33] for previous work on owner valuations.
[16] Ioannides and Zabel [32] pursues that line of inquiry.

the distributions of variables of potential interest, like racial and ethnic composition, within the neighborhood. The conditional mean of $u_{i\kappa ht}$, from (2), $\delta_x x_{\kappa t} + \delta_q q_{it}$, expresses *correlated effects* pertaining to valuation: 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. We shall see shortly below that by setting $\gamma = 0$, such a term is excluded, and in effect my specification of the error introduces something akin to a contextual effect. 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. However, to the extent that selection is present, z_{ht} proxies for the socioeconomic characteristics of neighbors. As Manski [36] emphasizes, if $\gamma = 0$, then the remaining (endogenous) social effect may be readily identified.

I note that it is difficult to distinguish the above empirical model from a hedonic model of property values with social interactions. Yet, it is important to do so. A pure hedonic model of property values would be based on a regression like Equ. (1) along with an error specification according to (2), subject to the following conditions. First, $\mu = 0$, as there is no reason why current dwelling unit characteristics should not be sufficient to determine market value. Second, $\eta = 0$, as the owner-occupant's own characteristics are not a market attribute, at least in theory; however, the characteristics of one's neighbors, represented here by $\gamma E[z_{ht}|x_{\kappa t}, q_{it}]$, are. If they are not identifiable, then presence of a term like ηz_{ht} suffices for bringing in the effect of the characteristics of one's neighbors either because the characteristics or neighbors are correlated (through sorting), or through the endogenous effect $\beta \Pi_i Y_t$, as we shall see shortly. With this in mind, we proceed with examining the impact of the spatial structure of the data.[17]

4.1 Spatial Interaction

If the kernel and all neighbors are treated symmetrically and interaction is global *within* each cluster, the spatial weighting matrix Π, employed in Equ. (1) above, is block-diagonal of size $I \times I$,

[17] This model combines certain features of Manski [36], especially its spatial model, *ibid.*, p. 537, Equ. (7), who examines estimation problems for social interaction models. 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 " [*ibid.*, p. 537]. In contrast to the principal model in the latter, in (1) social interactions are expressed in terms of *actual* behavior, Y_t, instead of *expected* behavior of one's neighbors, conditional on observables $[x_{\kappa t}, q_{it}]$, $E[Y_t|x_{\kappa t}, q_{it}]$.

with elements in each row summing up to 1. Its entries are defined as

$$\pi_{ij} = \frac{1}{n_i - 1}, \; i = 1, \ldots, I, j \in n(i), i \neq j, \quad \text{and } \pi_{ii} = 0, \text{otherwise}, \tag{3}$$

where n_i denotes the set of i's neighbors and n_i its size, $n_i = |n(i)|$. 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. [18] When no confusion arises, I use n and refer to cluster size as a constant, even though it does vary within the data.

The simplest such model obtains when in (1), I set $\mu = \eta = \tau = 0$, and similarly $\delta_x = \delta_q = (0, \cdots, 0)$, and ΠY_t is simply the vector that assigns to each unit the mean valuation of all other units in the neighborhood. That is, Equ. (1) simplifies to:

$$y_{i\kappa h} = \alpha + \beta \frac{1}{n-1} \sum_{j \in n(i), j \neq i} y_{j\kappa h'} + \lambda_i + \varepsilon_i. \tag{4}$$

I assume, like Glaeser and Scheinkman [23], that λ_i above is a cluster-specific random effect, with mean zero and variance σ_λ^2, which is uncorrelated with ε_i. This implies that across the data, individual valuations have mean and variance given by:

$$\bar{y} = \frac{\alpha}{1-\beta}, \;\; \sigma_y^2 = \frac{\sigma_\lambda^2}{(1-\beta)^2} + \sigma_\epsilon^2 + \sigma_\epsilon^2 \left(\frac{\beta}{1-\beta} \right)^2 \frac{3(n-1) - 2\beta(n-2) - \beta^2}{(n-1+\beta)^2}. \tag{5}$$

This model can be estimated with maximum likelihood. It will be referred to below as the *mean field* model of neighborhood interactions.

In contrast to the mean field case of interactions, where each individual is affected by the average behavior of all of her neighbors, the literature has investigated the consequences of alternative

[18] 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:

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

patterns (topologies) of local interactions; see [28]. Of particular interest is the *circular* interaction pattern (*ibid.,* and [22]), where each individual observes the behavior of only one other one, so that the connections form a circle. For such a case, drawing from [28], we have for mean and variance of the individual valuations, respectively:

$$\bar{y} = \frac{\alpha}{1-\beta}, \quad \sigma_y^2 = \frac{1+\beta^n}{1-\beta^n}\frac{\sigma_\epsilon^2}{1-\beta^2}. \tag{6}$$

This model can be estimated with maximum likelihood. It will be referred to below as the *circular* model of neighborhood interactions. To my knowledge, neither the mean field nor the circular models have been estimated before.

4.1.1 Spatial Stochastic Structure

I refer to (1), set $\gamma = 0$, 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 (1). 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.

The most general model would 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 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. (1) may be rewritten in vector form as:

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

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. 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:

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

$$+[\mathcal{I} - \beta\Pi]^{-1}\left[\delta_x X_t + \delta_q Q_t + \eta Z_t\right] + [\mathcal{I} - \beta\Pi]^{-1}[\mathcal{I} - \tau\Pi]^{-1}\varepsilon_t. \tag{8}$$

After elementary but tedious transformations, [19] Equ. (8) may be transformed further to yield a reduced- form, which under the assumption that cluster size is constant, $n_i = n$, is written as follows:

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

$$+\frac{n-1}{n-1+\beta}\mu Y_{t-1} + \frac{n}{n-1+\beta}\frac{\beta}{1-\beta}\left[\mu\bar{Y}_{t-1} + \bar{X}_t'\delta_x + \bar{Q}_t'\delta_q + \bar{Z}_t'\eta\right] + \bar{\varepsilon}_t, \tag{9}$$

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

[19] 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)]:

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

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

of each of the corresponding matrices. This may be simplified, again by elementary transformations, to yield:[20]

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

where $\bar{\varepsilon}^{\nu}_t$ 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^2_ε, β, τ, and n, which is exogenous and fixed at 10.[21] So, in sum, the most general model implies correlated random effects that derive from spatial interaction in unobserved components. Ignoring them leads to inefficiency. Unfortunately, I have not been able to estimate this model. I offer this analysis here in the hope that it would receive further attention in the future.

4.2 Estimation Strategy

Pausing first to summarize, I have derived the implications of Nash equilibrium within each neighborhood by means of a system of equations in structural form, Equ. (7), and in reduced form, Equ.'s (9) and (10). 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 models allow one to identify the effect of social interactions. This formulation excludes 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. Because of this and considering the complexity of the above model, I will test below the spatial interaction in the error structure by means solely of cluster-specific random effects without spatial autocorrelation.

I proceed first with a sequence of simplified models, which I discussed in subsection 4.1 above,

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

[21] Manski's result still applies, of course, in that separate identification of γ from β, when γ is not equal to 0, 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 [6] emphasize. Ioannides and Zabel [32] pursue this line of research.

and which derive from (1). Next I turn to the reduced-form model, Equ. (9), and impose the condition $\mu = 0$. This model may be estimated by OLS, if the constraints associated with the presence of β in the coefficients for X, Q, Z, and $\bar{X}, \bar{Q}, \bar{Z}$, are ignored, or by GLS, in order to account for random effects associated with different units belonging the same cluster. This model may also be estimated by maximum likelihood, where we allow for n_i, the number of units in each cluster to vary by cluster, as is indeed the case in the data.

The full model according to Equ. (9) may also be estimated. The presence of the lagged term comes with an additional parameter, μ. However, inspection of that equation reveals that terms $\frac{n-1}{n-1+\beta}\mu Y_{t-1}$ and $\frac{n}{n-1+\beta}\frac{\beta}{1-\beta}\mu \bar{Y}_{t-1}$, provide an additional route to the identification of β from the 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 (1). However, the variation in n_i in the data is substantial and, therefore, this method may be relied upon only as an approximation.

Finally, I can estimate the model by working with Equ. (7) as a structural form. 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. However, as I mentioned above, I will set $\tau = 0$, and thus ignore the spatial autocorrelation in the errors and instead estimate just a random effects specification.[22]

[22]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, estimation of β involves both the error structure and the coefficients of several RHS variables. Unfortunately, it is not possible to identify the spatial autocorrelation coefficient

5 Empirical Results

I discuss first estimates of the mean-field and circular interactions models, according to Equ. (5) and (6) by means of maximum likelihood, which are given on Table 2. The estimates for the mean-field model without and with a random effect, respectively in columns 2 and 3 of Table 2, imply a highly significant estimate for β, ranging from .683 to .618, but an insignificant cluster random effect. The estimates for 1993 are very similar and therefore not reported here. Column 4 reports the estimates for the circular interactions model. This yields an even higher estimate for β, 0.770, which is, again, highly significant. These models are significant in terms of the maximum likelihood ratio test. As I indicated above, these are the first estimates of these models ever reported.

I discuss next estimates obtained along the lines of Equ. (7) and (9), where I set $\tau = 0$. It is appropriate to summarize how those equations differ. Equation (7) is a structural form, where the dependent variable $y_{i\kappa ht}$ is a function of its own lagged value, $y_{i\kappa ht-1}$, 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, $\delta_x x_{it}$, and of unit i's own characteristics, $\delta_q q_{it}$. Identification of the latter effect requires a richer data set that allows one to study matching of individuals with neighborhoods and dwelling units (see Ioannides and Zabel [32]). Therefore, I set $\gamma = 0$.

The social interactions effect β may be identified as the coefficient of the predicted mean of the dependent variable among a unit's neighbors. 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 $y_{i\kappa ht}$ 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.

The estimates along the lines of Equ. (7), reported in Table 3, are typical of the entire set of regressions I have 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

without carrying out maximum likelihood estimations, which are extremely tedious in this setting and at the end did not work out for me. That the covariance structure of (9) may also aid identification has been noted by Case [9] and Moffitt [39].

1985, when appropriate. 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 [30], Appendix A. Specifically, including all three waves while retaining the panel structure would introduce considerable heterogeneity and reduce the sample size.

The dependent variable is the log of the self-reported value of owner-occupied dwellings. Table 3 reports estimation results for a conventional and an *augmented* hedonic regression, and interactive regressions along the lines of Equations (9) and (7). Column 1 is a hedonic regression, with the cluster and unit characteristics as regressors, that is variables X_t, Q_t, only. Columns 2 and 3 report a hedonic model, where the traditional explanatory variables have been augmented to include in addition to cluster characteristics, X_t, and dwelling unit own characteristics, Q_t, the characteristics of neighboring units, $\bar{Q}_t$, and the own socioeconomic characteristics of the owner and of her neighbors, $Z_t, \bar{Z}_t$, respectively. Note that to save space, the entries of Column 3 for the Q's and Z's correspond to $\bar{Q}_t, \bar{Z}_t$, respectively, that is, to one's neighbors' average characteristics. E.g., the entry for age in column 3 is for $\overline{\text{age}}$, the average age of neighboring dwelling units. My ability to run such a regression depends entirely on the availability of the neighborhood clusters data. Columns 4, 5 and 6 report interactive valuation regressions. That is, Columns 4 and 5 report results for the reduced-form model (9) with cluster-specific random effects using GLS. Finally, Column 6 reports results for the structural model (7) using 2SLS. The presentation of the results in Table 3 is organized according to groups of explanatory variables, that is of cluster-specific variables, the X's, of dwelling-unit variables, the Q's, and of individual-specific variables, the Z's.

Both groups of regressors, cluster-specific variables, the X's, and dwelling-unit variables, the Q's, are important explanatory variables in the hedonic regressions and the interactive valuation regressions. Hedonic regressions, like the one reported on Column 1, reflect market valuations and therefore condition only on cluster characteristics and dwelling unit characteristics. They exclude individual occupant characteristics. Some of the neighborhood characteristics may be interpreted as exogenous social, or contextual, effects, like per cent of owner-occupants, of household heads who are White, and of vacancies in the neighborhood. Neighborhood (cluster) specific variables performed quite well in several regressions. Some of them imply nonlinear effects, such as, in particular, cluster-averages for race and for vacancies, for which I have estimated cubic polynomial structures. All of these groups are significant.

In the hedonic regressions, reported respectively in Columns 1 and 2-3, I treat observations belonging to the same cluster as independent. When I add, in the regression reported in Columns 2-3, contextual information associated with the characteristics of neighboring units and of neighbors, I also include the occupant's own characteristics. This is not standard for hedonic regressions. I justify the presence of individual own characteristics in hedonic regressions because of high correlation with the respective variables for neighbors.[23] This makes it harder to interpret the regression coefficients but does improve the overall fit, raising the R^2 from 0.428 to 0.530. It is therefore more appropriate to examine the performance of the explanatory variables in Column 1. All estimated coefficients generally accord with intuition.

As I discussed in the previous section, spatial interactions may 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. I have, however, chosen to report only the model with cluster-specific random effects. Columns 4-5 present results with cluster-specific random effects along the lines of reduced-form model (9). All t statistics reported are robust with respect to heteroscedasticity associated with the neighborhood clusters. Using fixed effects appears to over-parameterize the model[24], yielding an implausibly high R^2 of 0.9979. While a substantial improvement in the quality of fit is, of course, to be expected, the fact that the social interactions variable remains significant is highly supportive of the notion of neighborhood effects. The significance of cluster-specific effects of either type is, in and of itself, an indication of the significance of social effects.[25]

Fixed versus random effect specifications may be tested by means of the Hausman specification test. I test the null hypothesis that the random effects are uncorrelated with the regressors, which

[23] Unlike Kiel and Zabel [34], I do not have access,for the purpose of this paper, 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, 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 [31] use occupant data from a larger sample of the AHS, the entire metropolitan sample, as proxies for neighborhood effects.

[24] I owe this remark to a referee.

[25] See Munshi and Myaux [40] for a related application of fixed effects as social effects.

is required by GLS theory. Under that null hypothesis, both the fixed effects and the random effects estimators are consistent, but the fixed effects estimator is inefficient. The Hausman test does reject, in my case, the null hypothesis very strongly. 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 otherwise fits reasonably well. I should note that for the property valuation model, the within and between R^2's are .116 and .787, respectively, with the fraction of the overall variance that is due to the random effect is 0.394. I interpret the significance of the random effects model on its own as lending support to the spatial stochastic structure that was introduced in the preceding section.

The reduced form model contains both the lagged dependent variable, Y_{85}, and the lagged average value among each individual's neighbors in the cluster, $\bar{Y}_{85}$. Both these variables are highly significant and improve the overall fit. In fact, the estimate of the effect of the latter, at 0.317, is larger than that of the former, 0.190, while both are highly statistically significant, implying a more important role for the social interaction effect. I interpret these results as evidence of significant social interaction in neighborhoods, where individuals are affected by the valuation behavior of their neighbors.

Recall the discussion in the previous section that a point estimate of the social interactions coefficient may be recovered from the ratio of the above two coefficients, $\frac{n}{n-1}\frac{\beta}{1-\beta}$. This would, of course, be only approximate because cluster sizes do vary across the sample. Using a value for n equal to the average cluster size, $\bar{n} = 6.7$,[26] the implied value of $\hat{\beta}$ is 0.587, and is reported on Table 4. Also reported there are, for ease of comparison, the estimates for β that are obtained directly from the mean field and circular interactions models, as well as the structural form model, which we discuss further below. The reduced form model also includes as regressors the averages of the neighboring units' characteristics and of their occupants, variables $\bar{Q}_t, \bar{Z}_t$. Note that to save space, the entries of Column 5 for the Q's and Z's correspond to the neighbors' average characteristics, $\bar{Q}$'s and $\bar{Z}$'s, respectively.

It is interesting to note that the estimated coefficients for those of the X's, the $(Q, \bar{Q})$'s and the $(Z, \bar{Z})$'s that are present in both Columns 2-3 and 4-5 are quite similar, except that those

[26] This number differs from 10 considerably, because of missing values, of different cluster sample sizes and of the fact that observations for renters may not be used in these regressions.

in Columns 2-3 have higher statistical significance than their counterparts in Columns 4-5. This is obviously caused by the presence in the reduced-form regression, according to Equ. (9) of the previous section, of the lagged value of the dependent variable, of the average valuation among one's neighbors, and of cluster-specific random effects. This regression is reported in Columns 4-5 above. In fact, from my perspective, it is remarkable that the coefficients of the X's, the $(Q, \bar{Q})$'s and the $(Z, \bar{Z})$'s retain any significance at all.

I note that while the t statistics I report in Columns 4-5 are obtained from GLS, I have also been concerned about correcting for the fact that the social interactions term is a predicted value. Unfortunately, the model is very difficult to estimate by means of 2SLS with standard econometric packages while still allowing for random effects. 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.[27] If housing markets do price properties correctly, we would expect that social interactions would make their presence felt even if we were to exclude the lagged value of the dependent variable. Pure curiosity suggests that this is worth a try. Leaving out the lagged value makes the estimate of the social interactions coefficient much larger but does not affect the estimates of the other coefficients.

A noteworthy result is that the estimated coefficient for income for the property valuation model are .031 and .020, from Columns 4 and 6, 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.[28]

Next I turn to the estimation of the structural form model according to Equ. (7). I note that the predicted value among one's neighbors is included as an explanatory variable. This is clearly endogenous and is instrumented by means of all exogenous variables. The second-stage regressions exclude the mean values among one's neighbors of dwelling unit structural characteristics and

[27] See Ioannides and Zabel [31] for an explanation of the necessary correction.

[28] The estimates of the income elasticity of housing valuations obtained here are very similar to those of Ioannides and Zabel [31], who aim at estimating a housing demand model with neighborhood effects, and to results by others who have also used the AHS data. Accounting for neighborhood selection, however, does raise the estimates of income elasticity of housing demand. See [32].

neighbor household characteristics, that is variables $\bar{Q}_t, \bar{Z}_t$. The results, which are obtained using 2SLS, are reported on Table 3, Column 6, and the estimated standard errors for the structural form model are suitably corrected. The estimated social interactions coefficient is 0.671 and thus quite high, even though the own lagged value is also included in the regression and has an estimated coefficient of .161, which is small but highly significant. It is particularly interesting to compare the structural form results with those of the augmented hedonic. While the same data are used when one considers both stages of the structural form estimation, the structural form gives a better fit in terms of R^2 and delivers the attractive interpretation of the social interactions coefficient. That is, one's neighbors' valuations is an important explanatory variable of one's own property valuations.

Again, it is interesting to compare the estimated coefficients for the X's, the $(Q, \bar{Q})$'s and the $(Z, \bar{Z})$'s across the models. These variables continue to be statistically significant as a group, but very few of them are individually. This must be due to the overwhelming role of the endogenous social effect, predicted mean among one's neighbors.

Table 4 offers a juxtaposition of all direct estimates of β, Columns 2, 3, 4, and 6, and of the estimated coefficients on which indirect inference on β rests, Column 5. It thus confirms that all of our estimates of the social interactions coefficient β are fairly near one another and generally very significant. It is noteworthy that such completely different models as the mean field and the circular interactions models yield similar results. Since both the property valuation and its predicted mean among one's neighbors are in logs, coefficient β may be interpreted as an elasticity.

6 Conclusions

I explore a relatively neglected feature of data from the American Housing Survey for 1985 and 1989, 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 in neighboring property valuations 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.

Using a variety of models, I find the impact of social interactions to be quite substantial: the estimated coefficient ranges from .587 to .770. The social interaction effect is found to be both stronger and more significant then that of the own lagged value, when both are included.

The results provide empirical support for the notion of interactions in residential property valuations. That is, individuals' valuations of their properties are influenced by those of their neighbors. As a positive finding, this may be interpreted as supportive of the notion, common in the real estate world, that neighborhood is very important as a determinant of property valuations. It is also supportive of the notion that public policy interventions may bring about urban neighborhood change through a social multiplier.

References

[1] L. Anselin, Spatial Econometrics, School of Social Sciences, University of Texas at Dallas, April, mimeo (1999).

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

[3] R. Benabou, Heterogeneity, Stratification, and Growth, *American Economic Review,* 86, 3, 584-609 (1996).

[4] M. Binder and M. H. Pesaran, Optimal Consumption Decisions under Social Interactions, working paper, University of Maryland, September (1997).

[5] E. W. Bond and N. E. Coulson, Externalities, Filtering and Neighborhood Change, *Journal of Urban Economics,* 26, 231–249 (1989).

[6] W. A. Brock and S. N. Durlauf, Interactions –Based Models, in "Handbook of Econometrics," Volume 5 (J. Heckman and E. Leamer, Eds.), North-Holland, Amsterdam, forthcoming, March (2000).

[7] J. K. Brueckner, Testing for Strategic Interaction among Governments: The Case of Growth Controls, *Journal of Urban Economics,* 44, 438–467 (1998).

[8] J. K. Brueckner, Strategic Interaction among Governments: An Overview of Empirical Studies, Department of Economics, University of Illinois, Champaign-Urbana, April, mimeo (2001).

[9] A. C. Case, Neighborhood Influence and Technological Change, *Regional Science and Urban Economics,* 22, 491-508 (1992).

[10] N. E. Coulson and E. W. Bond, A Hedonic Approach to Residential Succession, *Review of Economics and Statistics,* LXXII, 3, August, 433–444 (1990).

[11] N. E. Coulson, S.-J. Hwang and S. Imai, The Value of Home-Ownership in Neighborhoods, Pennsylvania State University, Department of Economics, working paper, September (2001).

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

[13] O. A. Davis and A. Whinston, The Economics of Urban Renewal, *Journal of Law and Contemporary Problems,* 26, 105-117 (1961).

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

[15] S. N. Durlauf, Statistical Mechanics Approaches to Socioeconomic Behavior, 81–104, in "The Economy as a Complex Evolving System II," (W. A. Arthur, S. N. Durlauf and D. Lane, eds.), Addison-Wesley, Reading, MA. (1997).

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

[17] D. Epple and H. Sieg, Estimating Equilibrium Models of Local Jurisdictions, *Journal of Political Economy,* 107, 4, 645–681 (1999).

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

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

[20] G. C. Galster, "Homeowners and Neighborhood Reinvestment," Duke University Press, Durham, North Carolina (1987).

[21] M. Gladwell, The Tipping Point, *The New Yorker,* June 3, (1996).

[22] E. L. Glaeser, B. Sacerdote and J. A. Scheinkman, Crime and Social Interactions, *Quarterly Journal of Economics,* 507–548 (1996).

[23] Glaeser, Edward L., and Jose A. Scheinkman, Non-Market Interactions, paper presented at the World Congress of the Econometric Society, Seattle (2000).

[24] L. Hadden and M. Leger, "Codebook for the American Housing Survey," Abt Associates, Cambridge, MA (1990).

[25] A. M. Hardman and Y. M. Ioannides, 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).

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

[27] Y. M. Ioannides, Neighborhood Income Distributions, working paper, Department of Economics, Tufts University, June (2001a).

[28] Y. M. Ioannides, Topologies of Social Interactions, working paper, Tufts University, June (2001b).

[29] Y. M. Ioannides, Nonlinear Neighborhood Interactions and Intergenerational Transmission of Human Capital, forthcoming in "Essays in the Honour of Emmanuel Drandakis," (G. Bitros and Y. Katsoulacos, Eds.) Edward Elgar, Cheltenham, UK (2001c).

[30] Y. M. Ioannides, Residential Neighborhood Effects, *Regional Science and Urban Economics,* 32, 2, 145–165, (2002).

[31] Y. M. Ioannides and J. E. Zabel, Neighborhood Effects and Housing Demand, presented at the Conference on Empirics of Social Interactions, Brookings Institution, Washington DC, January; forthcoming in *Journal of Applied Econometrics* (2000).

[32] Y. M. Ioannides and J. E. Zabel, Interactions, Neighborhood Selection and Housing Demand, working paper, Department of Economics, Tufts University; presented at the AEA Meetings, Atlanta, GA, January (2002).

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

[34] K. A. Kiel and J. E. Zabel, 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 (1997).

[35] M. Kremer, How Much Does Sorting Increase Inequality? *Quarterly Journal of Economics,* February, 115–139 (1997).

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

[37] C. F. Manski, Economic Analysis of Social Interactions, *Journal of Economic Perspectives,* 14, 115–136 (2000).

[38] T. Miyao, Dynamic Instability of a Mixed City in the Presence of Neighborhood Externalities, *American Economic Review,* 68, 3, 454–463 (1978).

[39] R. A. Moffitt, Policy Interventions, Low-Level Equilibria and Social Interactions, 45–82, in "Social Dynamics," (S. N. Durlauf and H. Peyton Young, Eds.), MIT Press, Cambridge, MA (2001).

[40] K. Munshi and J. Myaux, Social Effects in the Demographic Transition: Evidence from Matlab, Bangladesh, presented at the NEUDC Conference, Economic Growth Center, Yale University (1998).

[41] R. A. Pollak, Interdependent Preferences, *American Economic Review,* 78, 745–763 (1976).

[42] J. Rothenberg, "Economic Evaluation of Urban Renewal," Brookings, Washington DC (1967).

[43] L. D. Schall, Urban Renewal Policy and Economic Efficiency, *American Economic Review,* 66, 4, 612–628 (1976).

[44] T. C. Schelling, Dynamic Models of Segregation,*Journal of Mathematical Sociology,* 1, 143–186 (1971).

[45] S. Sheppard, Hedonic Analysis of Housing Markets, in " Handbook of Regional and Urban Economics," Vol. 3, "Applied Urban Economics," (P. Cheshire and E. S. Mills, Eds.,) 1595–1635, North-Holland, Amsterdam, (1999).

[46] R. Spivack, The Determinants of Housing Maintenance and Upkeep: A Case Study of Providence, Rhode Island, *Applied Economics,* 23, 639–646 (1991).

[47] K. O. Stahl, Market Failures and the Allocation of Resources in Cities, Ph. D. Dissertation, University of California, Berkeley (1974).

[48] W. Strange, Overlapping Neighborhoods and Housing Externalities, *Journal of Urban Economics,* 32, 17–39 (1992).

[49] "The New Shorter Oxford English Dictionary," Oxford University Press, Oxford (1993).

[50] C. Tiebout, A Pure Theory of Local Public Expenditure, *Journal of Political Economy,* 64, 416–424 (1956).

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

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

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

Table 1:

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 2:

Alternative Social Interaction Models, 1989

Model	Mean Field	Mean Field with Cluster RE	Circular Interaction
1	2	3	4
Intercept	3.604	4.337	2.637
	(5.80)	(2.16)	(4.99)
Social interaction	.683	.618	.770
β	(12.47)	(3.49)	(16.64)
Random Shock	.277	.234	.154
σ_ϵ^2	(4.43)	(1.26)	(3.28)
Cluster Random Effect		.027	
σ_λ^2		(1.27)	

Table 3

Interactive Regressions for Owners, 1989

Variable	LV_{89}	LV_{89}	LV_{89}	LV_{89}	LV_{89}	LV_{89}
Column	1	2	3	4	5	6
Variables	Hedonic	Augmented Hedonic		Reduced Form		Structural Form
Mean	11.364	11.286		11.286		11.286
Observations	2968	2180		2179		2179
Number Clusters	324	324		324		324
Obs per cluster	6.7	6.7		6.7		6.7
R^2, within				.116		
R^2, between				.787		
R^2 overall	.428	.530		.626		.617
F or Wald χ^2	101.98	48.22		1278		95.65
MSE	.620	.534		.387		.482
Cluster Effects	No	No		Random Effects		No
S.D. of RE	.			.312		
Fraction of variance due RE	.			.394		
Model	OLS	OLS		GLS		2SLS
Intercept	11.80	13.75		1.164		-.087
	(3.58)	(11.05)		(0.08)		(.99)
LV_{85}				.190		.161
				(11.78)		(8.55)

Table 3

Interactive Regressions for Owners, 1989 (Continued)

Variable	LV_{89}	LV_{89}	LV_{89}	LV_{89}	LV_{89}	LV_{89}
Column	1	2	3	4	5	6
Variables	Hedonic	Augmented Hedonic		Reduced Form		Structural Form
Cluster data — X Variables						
Pred. Mean neighbors$_{89}$						.671 (18.17)
Mean neighbors$_{85}$				.317 (8.37)		
Central City SMSA	.293 (7.14)	.171 (4.18)		.146 (2.15)		.047 (1.26)
Suburb SMSA	.391 (9.71)	.272 (6.76)		.206 (3.04)		.065 (1.76)
Region North East	.635 (17.74)	.487 (12.73)		.385 (5.76)		.129 (3.40)
Region South	.034 (0.77)	.116 (2.56)		.019 (.25)		-.031 (.79)
Region West	.520 (13.20)	.484 (11.23)		.310 (4.06)		.073 (1.80)
Degrees	.018 (1.42)	-.004 (0.32)		-.013 (.56)		.002 (.20)

Table 3 Continued

Variable	LV_{89}	LV_{89}	LV_{89}	LV_{89}	LV_{89}	LV_{89}
Variables	Hedonic	Augmented Hedonic		Reduced Form		Structural Form
Column	1	2	3	4	5	6
Own	-1.574	.271		-.235		-.560
	(.56)	(2.90)		(1.75)		(.08)
Own^2	.295	.904		2.431		.244
	(.38)	(.45)		(.23)		(.14)
Own^3	-.011	-.079		.026		-.281
	(.15)	(.46)		(.11)		(.19)
Race	-.392	.028		-.039		-.006
Head White=1	(2.41)	(.66)		(.58)		(.17)
$Race^2$	.386	.0001		.0002		.0000
	(4.56)	(2.11)		(1.44)		(.79)
$Race^3$	-.063	-.0000		-.0000		-.0000
Head White	(5.58)	(2.83)		(1.66)		(.92)
Change in		-.108		-.093		-.015
$Race_{89,85}$		(2.29)		(1.28)		(.37)
Vacant	.310	-.045		.760		.052
	(.92)	(.12)		(1.30)		(.15)
$Vacant^2$	-.838	-.014		-.175		-.012
	(1.45)	(.16)		(1.43)		(.17)
$Vacant^3$	.331	.002		.023		.001
	(1.43)	(.15)		(1.37)		(.08)
Change in		.010		.009		.008
$Vacant_{89,85}$		(.72)		(.43)		(.62)

Table 3 Continued

Columns 3 and 5 report coefficients of neighbors' average values for regressions in Columns 2 and 4, respectively.

Variable	LV_{89}		LV_{89}	LV_{89}	LV_{89}	LV_{89}
Variables	Hedonic	Augmented Hedonic		Reduced Form		Structural Form
Column	1	2	3	4	5	6
Dwelling Unit Data — Q Variables			$\bar{Q}$ Variables	Dwelling Unit Data — Q Variables		
Age	-.001	-.092	.153	-.088	.058	-.030
	(.40)	(2.63)	(3.64)	(3.34)	(1.24)	(1.54)
Not detached	-.252	.141	-.449	.156	-.340	-.018
	(4.67)	(1.30)	(3.63)	(1.90)	(2.87)	(.32)
Unit area	.0002	.0001	.0001	.0001	.00001	.0000
	(11.75)	(4.52)	(4.33)	(4.23)	(1.53)	(2.76)
Rooms	.062	.369	-.092	.043	-.047	.036
	(6.94)	(3.74)	(5.19)	(5.12)	(1.92)	(4.29)
Baths	.176	.088	.222	.056	.077	.035
	(8.25)	(3.58)	(5.05)	(2.78)	(1.22)	(1.68)
Additions	.043	.049	.170	.033	.103	.025
	(3.05)	(3.28)	(5.04)	(2.45)	(2.19)	(1.84)

Table 3 Continued

Columns 3 and 5 report coefficients of neighbors' average values for regressions in Columns 2 and 4, respectively.

Variable	LV_{89}	LV_{89}	LV_{89}	LV_{89}	LV_{89}	LV_{89}
Variables	Hedonic	Augmented Hedonic		Reduced Form		Structural Form
Column	1	2	3	4	5	6
Household Data — Z Variables			$\bar{Z}$ Variables	Household Data — Z Variables		
Moved in since 1985		.0026 (.08)	-.208 (2.82)	-.006 (.20)	-.246 (2.37)	.019 (.62)
Age		.465 (.46)	-2.28 (1.07)	.713 (.83)	-1.46 (.50)	.772 (.87)
Age ^{2}		-.047 (.37)	.313 (1.15)	.85 (.77)	.189 (.51)	-.094 (.82)
Head White		-.062 (.95)	-.049 (.29)	-.075 (1.33)	-.067 (.32)	-.062 (1.07)
Education		.016 (3.42)	.029 (3.45)	.001 (2.50)	.001 (.11)	.008 (1.91)
HH Size		-.002 (.20)	.032 (1.38)	-.008 (.89)	-.010 (.32)	-.004 (.42)
Head married		.055 (1.58)	-.113 (1.51)	.058 (1.91)	-.054 (.52)	.055 (1.75)
Head male		-.077 (2.20)	-.038 (.51)	-.084 (2.70)	-.137 (1.33)	-.066 (2.08)
Cars		.021 (1.40)	.153 (4.43)	.004 (.27)	.096 (1.91)	-.006 (.42)
Income		.045 (3.72)	.146 (5.61)	.031 (2.89)	.074 (2.02)	.020 (1.80)

Table 4:

Alternative Estimates of the Social Interaction Parameter β

1	2	3	4	5	6
	Property Valuations				
Model	Mean Field		Circular Interaction	Reduced Form	Structural Form
μ					.161 (8.55)
β	.683 (12.47)	.618 (10.01)	.769 (17.72)		.671 (18.17)
$Value_{85}$				.190 (11.78)	
$\overline{Value}_{85}$				.317 (8.37)	
Implied $\hat{\beta}$				.587	

APPENDIX A: 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.

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 individual variables in levels and key variables in logarithms, as well.

Variable	Observations	Mean	Standard Deviation	Min	Max
Additions,Log 1989	2183	.500	.802	0	5
Age of Head 1989	2183	54.1	15.6	18	91
Age of dwelling 1989	2183	33.8	18.7	0	75
Baths 1989	2183	1.6	.69	0	5
Cars 1989	2183	1.6	.88	0	7
Central city of MSA 1989	2183	.36	.48	0	1
Degrees, 1989 (categorical,1–6)	2183	3.2	1.4	1	6
Dwelling not detached, 1989	2183	.96	.20	0	1
Education of Head, years	2183	13.3	3.4	0	18
Head Male, =1, if yes	2183	.76	.43	0	1
Head Married, =1, if yes	2183	.69	.46	0	1

Variable	Observations	Mean	Standard Deviation	Min	Max
Head White =1, if yes	2942	.892	.310	0	1
% White Heads (Race) in cluster	2183	88.51	25.71	0	100
ΔL Race$_{89,85}$	2183	-.03	.28	-2.47	3.15
%Owners in cluster	2183	84.92	17.56	14.29	100
Change in log % Owners 89-85	2183	.008	.140	-.81	1.25
Income, 1989	2183	43333	31487	0	205000
Income, log, 1989	2183	10.33	1.14	0	12.23
Moved since last year, 1989	2183	.069	.254	0	1
Number of rooms, 1989	2183	6.54	1.66	3	14
Number of vacant units in cluster	2183	.29	.90	0	5.88
Other urban, in MSA, 1989	2183	.12	.33	0	1
Persons in household	2183	2.77	1.45	1	11
Region Midwest	2183	.25	.43	0	1
Region Northeast	2183	.20	.40	0	1
Region South	2183	.32	.47	0	1
Region West	2183	.23	.42	0	1
Suburb of MSA, 1989	2183	.53	.50	0	1
Unit area, sf, 1989	2183	1959.2	839.3	152	4000
Value, 1985 $	2183	84702.7	53395.0	0	250001
Units Vacant	2183	.30	.90	0	5
ΔLog Vacant$_{t,t-1}$	2183	.33	.93	-1.94	4.80
Value, log, 1985	2183	11.13	.79	0	12.43
Value, 1989 $	2183	111271.2	82533.2	1000	350000
Value, log, 1989	2183	11.35	.78	6.91	12.77
Value, log, predicted of neighbors, 1989	2179	11.35	.69	8.29	12.77