NEIGHBORHOOD EFFECTS AND HOUSING DEMAND

Yannis M. Ioannides and Jeffrey E. Zabel

Department of Economics

Tufts University

Medford, MA 02155

yannis.ioannides@tufts.edu 617-627-3294 617-627-3917 (F)
jeff.zabel@tufts.edu 617-627-2318 617-627-3917 (F)

September 18, 2001

Abstract

In this paper, we estimate a model of housing demand with neighborhood effects. We exploit special features of the National sample of the American Housing Survey and properties of housing markets that allow us to create "natural" instruments and therefore identify the impact of social interactions. We find evidence of both endogenous and contextual neighborhood effects. We report two alternative sets of estimates for neighborhood effects that differ in terms of the instruments we use for estimating the model. When the endogenous neighborhood effect is large the respective contextual effects are weak, and vice versa. The elasticity of housing demand with respect to the mean of the neighbors' housing demands (the endogenous effect) ranges from 0.19 to 0.66 and is generally very significant. The contextual effects are also very significant. A key such effect, the elasticity with respect to the mean of neighbors' permanent incomes ranges from 0.17 to 0.54.

JEL Classification Codes: R21, C31.
Keywords: Housing demand, endogenous neighborhood effects, social interactions.

We are grateful to two referees and to the editor, Steven Durlauf, for very perceptive comments. This is a revised version of the paper that was presented at the Conference on Empirics of Social Interactions, Brookings Institution, January 14–15, 2000. We acknowledge insightful comments by the other conference participants and by our colleague Gib Metcalf. We thank Timothy Harris and Pamela Jia for expert research assistance. Ioannides acknowledges generous financial support from the National Science Foundation and the John D. and Catherine T. MacArthur Foundation.

1 Introduction

The increase in income equality in the 1990s has led to a new focus on measuring the impact of social effects on economic behavior. This rise is due, in part, to the belief that the rich benefit from a better social environment more than do poorer individuals. It is natural to expect that such social effects originate in one's place of residence and are often referred to as neighborhood effects. The impact of neighborhood on economic behavior can manifest itself through numerous channels. First, one's behavior can be influenced by the behavior or the characteristics of one's neighbors. Schelling (1978) has changed our thinking in this context. Desirable social interactions and beneficial local community "social capital" are known to be attractive features of different communities. Second, one can be affected by the public services that the town provides such as school quality and security. Third, one can be impacted by locally undesirable land uses such as toxic waste sites, incinerators, or electricity generating plants. The decision about where to locate and how much housing to consume will be influenced by these and other related factors.

In this paper, we derive and estimate a model of housing demand that incorporates social effects that originate in one's residential neighborhood. Such social effects will be referred to as neighborhood effects from now on. They include individuals' valuing the socioeconomic characteristics of their neighbors and their consumption of housing, which in turn show up as determinants of individual housing demand. Following a standard typology due to Manski (1993), they may be classified as contextual and endogenous neighborhood effects, respectively. The endogenous effect expresses a notion of "keeping up with the Joneses," whereby individuals who view their neighbors' decisions to maintain, renovate, repair, or make additions to their houses will strive to keep up by making similar decisions and hence increase their own housing consumption. The contextual effect will arise when owners view their neighbors' characteristics, e.g. income, as a signal of their future housing consumption and thus alter their own consumption accordingly. To our knowledge, no other researchers have estimated a model of housing demand with neighborhood effects.

Such social effects can result in reinforcing behavior that may lead to a divergence in quality across neighborhoods. That is, decisions to maintain, renovate, repair, or make additions to houses will lead other neighbors to do the same and to increasingly higher quality neighborhoods compared to those where these decisions are not undertaken. The higher quality neighborhoods will foster increased access to social capital that provides its residents with more opportunities for socioeconomic success. This will increase the (economic) gap between those individuals in the good and bad neighborhoods.

This paper treats the model of housing demand for a group of neighbors as a system of simultaneous equations. This approach follows Moffitt (2001) and helps clarify the identification conditions for neighborhood effects first articulated by Manski, *op. cit.*. These conditions have been developed further and generalized by Brock and Durlauf (2000). The endogenous variables in the system of equations are the set of individual housing demands in each neighborhood. We provide natural instruments for these variables by exploiting the intuition of the hedonic approach to the determination of house prices. The hedonic house price equation includes variables that our theory suggests that should be excluded from the demand equations. This identifies the system.

We estimate our model using micro data from suitably chosen subsamples of the national sample of the American Housing Survey (AHS). These subsamples include information on dwelling units and their occupants and on those of some of their nearest (in a geographical sense) neighbors, and together constitute about one-tenth of the data. This feature allows us to use this part of the data as a source of contextual information, and it is a key advantage of the AHS data. The remainder of the data are used for estimating the house price index that is included in the demand equation. This constitutes an important aspect of our approach. For specific applications, house price indices can be hard to define and calculate and their availability often determines the quality of housing research.

Our empirical analysis follows a four-step process that involves estimating the price of housing services, permanent income, the predicted value of the mean of neighbors' housing demands, and finally the structural housing demand equation. We first estimate a reduced-form version of the housing demand equation to show the existence of neighborhood effects without distinguishing between contextual and endogenous effects. Estimation of the structural model allows us to decompose these effects into endogenous and contextual effects. We use two alternative sets of instruments to identify the model. These are, first, the means of the structural characteristics of the neighbors' houses, and second, those of a single (randomly chosen) neighbor. We justify the use of these instruments through the link to the hedonic theory of house prices and also test for their exogeneity using a standard test for over-identification. The results support the presence of both endogenous and contextual effects. The elasticity of housing demand with respect to the mean of the neighbors' housing demands, the endogenous effect, ranges from 0.19 to 0.66. The elasticity with respect to the mean of neighbors' permanent income, a contextual effect, ranges from 0.17 to 0.54.

The paper is organized as follows. In Section 2 we describe the data and provide summary statistics of the relevant variables. In Section 3 we derive the model of housing demand with neighborhood effects. The empirical results are presented in Section 4 and Section 5 concludes.

2 The AHS Data

We start with a discussion the data, because its special features motivate our approach and some of our estimation procedure. The main data source used for this study is the national version of the American Housing Survey (NAHS). The NAHS is an unbalanced panel of more than 50,000 housing units that are interviewed every two years. It serves as the basis for US housing statistics. The NAHS contains detailed information on dwelling units and their occupants houses through time. The information includes the current owner's evaluation of the unit's market value, various structural characteristics of dwelling units, and self-reported information on the house's current occupants. Since the mid 1980s, the NAHS is conducted every two years.

In 1985, 630 dwelling units, which will be referred to as *kernels* in the remainder of the paper, were selected at random, and up to ten nearest housing units, to be referred to together with the kernel as neighborhood *clusters*, were interviewed. Additional observations come from larger clusters, making the total number of clusters equal to 680, yielding a data set of 7,350 housing units. This was repeated in 1989, when 769 kernel units were selected, and in 1993, when 1018 kernel units were selected. Additional units in existing clusters were included in 1989 to reflect additional units that had been added within the perimeter of the 1985 "neighborhood." By 1993, a maximum of 20 neighboring units were allowed per cluster. [1] The result is an unbalanced three-wave panel of dwelling units. Tables 1–5 provide extensive details on the structure of the data, including observation counts on new clusters, new households and new units, etc., and their

[1] We are grateful to Barbara T. Williams, US Bureau of the Census, for this clarification.

geographic distribution.

By working with this special subsample of the NAHS, we generate a data set that includes information on the value and characteristics for all dwelling units in the cluster. The owner-occupant's characteristics that we work with include, in particular, the owner's years of schooling, whether the owner is white, whether the owner is married, the number of persons in the household, household income, and whether the house has changed hands in the last five years.

For each survey, owners are asked to estimate how much their property (and, in addition, its lot, if appropriate) would sell for if it were for sale. Goodman and Ittner (1992) and Kiel and Zabel (1999) find that while, on average, owners over-estimate their value by 5%, this bias is not systematically related to the observed characteristics of the owner, house, or neighborhood. Kiel and Zabel find that the overvaluation is greatest for new owners and declines with length of tenure. They recommend that length of tenure be included in the house price hedonic when using the AHS data.

An observation on a dwelling unit from the AHS is included in our sample only if it: is associated with a regular occupied interview, is owner occupied, lies in a metropolitan statistical area (MSA), is valued by the owner to be worth at least $10,000, and is not missing any information on the unit's or occupant's characteristics that are included in our analysis. Since we are using the information about the neighbors to measure neighborhood effects, we require that there are at least four other dwelling units in the cluster after the above selection criteria have been employed. This reduces the number of observations for 1985, 1989, and 1993 to, respectively, 1947, 2318, and 2909.

Of the 365 neighborhood clusters included in our analysis, 196 are present in all three years. These clusters are located in 100 MSAs. Full information on the distribution of clusters across the three years is given in Tables 1–5. The names and definitions of the variables used in this study are given in Table 6. The means and standard deviations for these variables by year and for all three years combined are given in Table 7.

3 The Neighborhood Model

We study housing demand by treating each neighborhood cluster as a group of interacting agents. A cluster's kernel is randomly selected within the MSA and all neighborhood-specific interactions are construed as conditional on the specific cluster to which a dwelling unit belongs. Under the Nash equilibrium assumption, an individual takes her neighbors' decisions as given. An individual's presence in a cluster implies dependence on her neighbors' housing consumption decisions and their observable individual characteristics. For example, Ioannides (2001b) documents a range of correlation coefficients among neighbors' income from 0.3 to 0.5, depending upon year and geographic location within the United States. Neighbors may act on information about one another which is not directly observable by the econometrician. Some of these variables may be correlated with observable characteristics because of cluster-specific housing market clearing may sort similar people to similar neighborhoods. For the purpose of this paper, characteristics of neighbors that affect housing demand and are measured at the cluster level are treated as sources of *contextual effects*, while neighbors' housing consumption decisions are treated as sources of *endogenous* social effects, where we have invoked the terminology of Manski (1993).[2] These labels are clarified further by the behavioral model to which we turn next.

3.1 The Behavioral Model

An individual h cares about the quantity of housing services, y_h, produced by her property i with structural characteristics q_i in cluster κ, whose attributes are denoted by x_κ. Let an individual's utility function depend on composite nonhousing consumption, c_h, consumption of housing services, y_h, own demographic characteristics that might affect preferences, z_h, the vectors of housing consumptions in the neighborhood, $y_{n(i)}$,[3] and on the observable socioeconomic characteristics of neighbors, $z_{n(i)}$:

$$U_h = U(c_h, y_h; z_h; y_{n(i)}, z_{n(i)}). \tag{1}$$

We assume that the utility function $U_h(\cdot)$ is increasing and quasi-concave with respect to (c_h, y_h). We also assume that it depends on one's neighbors' housing consumption and on their socioeconomic

[2] In a follow-up study [Ioannides and Zabel (2000)], we also attempt to separately identify *correlated effects,* that is, an individual's tendency to behave like others with similar unobservable characteristics. This requires that we look at neighborhood choice. This, in turn, necessitates that we distinguish between residential neighborhood interactions, which are analyzed at the cluster level and require conditioning on the characteristics of a small *group* of individuals with possibly correlated characteristics, and community interactions which require conditioning on the characteristics of a *population*, such as that of a census tract, which individuals take as given. For the purpose of the present paper, we have not linked neighborhood clusters with census tracts. This could be accomplished only by access to privileged information. Indeed, Ioannides and Zabel (2000) is based on such information. In the present paper, we may condition only on cluster-specific variables.

[3] Preferences may depend directly upon the consumption of others, as in the formalization of Pollak (1976).

characteristics, which are expressed in vector form as $(y_{n(i)}, z_{n(i)})$. Nothing in the theory suggests a priori a specific functional dependence of U on $(y_{n(i)}, z_{n(i)})$. Assuming a static-equivalent setting, individual h chooses how to allocate her permanent income I_h to nonhousing consumption, c_h, and housing consumption, y_h, subject to

$$c_h + p y_h = I_h, \tag{2}$$

where nonhousing consumption is the numeraire and p is the price of housing services. Neighbors' consumption and socioeconomic characteristics are taken by individual h as given, as neighborhood choice is given.

Maximization of (1) with respect to nonhousing and housing consumption, subject to (2), yields the standard first-order condition:

$$p = \frac{\frac{\partial u}{\partial y_h}}{\frac{\partial u}{c_h}}. \tag{3}$$

Solving equations (2) and (3) yields the optimal consumption of housing services, housing demand. This is all standard except for the presence of social interactions, represented here by an individual's utility being a function of $(y_{n(i)}, z_{n(i)})$. Unless nonhousing and housing consumption, (c_h, y_h), and the remaining arguments of the utility function $(z_h, y_{n(i)}, z_{n(i)})$ enter separably, the marginal rate of substitution between housing and nonhousing consumption includes, in general, the full set of variables $(c_h, y_h; z_h, y_{n(i)}, z_{n(i)})$. Therefore, housing demand in general would reflect dependence on own demographic characteristics, housing consumption among neighbors and socioeconomic characteristics of neighbors, $(z_h; I_h; y_{n(i)}, z_{n(i)})$, in addition to price and income. We suppress the dependence of housing consumption on the actual unit and the cluster in which it lies, (i, κ), unless it is necessary for clarity.

Of particular interest in the context of social interactions is whether the utility function (1) depends on *actual* or on *expected* housing consumption among neighbors. We assume the former and therefore the demand functions that emanate from our model will contain a *spatial autoregressive* component. We justify this as follows. As Manski (1993) notes,[4] the spatial autoregressive model "implies that the sample members know who each other are and choose their outcomes only after

[4] "Thus, the spatial correlation model assumes that an endogenous effect is present within the researcher's sample rather than within the population from which the sample is drawn. This makes sense in studies of small-group interactions, where the sample is composed of clusters of friends, co-workers, or household members;" [Manski, *op. cit.*, , p. 537.]

having been selected into the sample" [Manski, *op. cit.*, , p. 537]. We assume that individuals react when neighbors increase their housing consumption through maintenance, addition, alteration, and/or repair to their housing structures by changing their own consumption in a similar fashion. Thus, we would like to know what happens when individuals observe their neighbors increasing their housing consumption. We would like to test what the simple notion of "keeping up with the Joneses" suggests, namely that they feel pressure to keep up with their neighbors and increase their own housing consumption.

We define the quantity of housing services as a scalar measure of the flow of services that arise from the structure and the neighborhood in which the house is located. This approach to housing demand has been used extensively in the literature and is useful for calculating price and income elasticities of housing demand; see Zabel (2001) and references therein. While there is also a large literature devoted to the demand for housing as derived from the demand for particular housing characteristics, we find that the use of housing services as a scalar quantity allows us to measure better the social interactions that we are trying to capture by our model of housing demand. That is, the social interactions reflect the impact on individuals' overall housing consumption of an increase in their neighbors' housing consumption via maintenance, addition, alteration, and/or repair of the housing structure.

The quantity of housing services is equal to the value of these services divided by the price per unit of service. Since the value of a house is the present discounted value of the stream of services provided by that house, the annualized value of these services is $r \cdot W_{i\kappa}$ where r is the user cost of housing and $W_{i\kappa}$ the value of unit i in cluster κ. Thus, we define the quantity of housing services consumed by individual h who occupies unit i in cluster κ as the ratio of housing expenditure, defined as the annualized value of her property, $r \cdot W_{i\kappa}$, divided by price, p_κ. It is convenient to express this definition in logs:

$$\ell n y_{i\kappa h} = \ell n W_{i\kappa} - \ell n p_\kappa. \tag{4}$$

Note that we have suppressed the term $\ln r$ from this equation since it will be subsumed in the constant term in the housing demand equation. Thus, because we work with logs, we need not make any assumptions about the rate at which properties produce the flow of housing services.[5]

[5]Because of the tax treatment of owner-occupied housing in the US, the user cost may be defined as net of tax. We refrain from doing so in this paper. However, one of the variables that we control for in estimating the price of

Given a particular specification of the utility function in (1), it is possible to derive the housing demand equation by solving the first-order condition (3) along with the budget constraint (2). Instead, we follow much of the literature by approximating the housing demand function using a log-linear specification:

$$\ell ny_{i\kappa h} = \alpha + \mu \ell np_\kappa + \xi \ell nz_h + \delta \ell nI_h + \beta\Pi_y(\ell ny_{n(i)}) + \gamma\Pi_z(z_{n(i)}) + v_\kappa + \epsilon_{\kappa h}, \qquad (5)$$

where α, μ, ξ, δ are scalar and β and γ vectors of parameters to be estimated, v_κ is an unobservable cluster random effect and, $\epsilon_{\kappa h}$ is an unobservable random variable that is assumed to be independently and identically distributed over all observations. This model combines features of Case (1991; 1992) and of the spatial model in Manski (1993), p. 537, Equ. 7, which we discuss above. It differs from the principal model in the latter in that in (5), social interactions work through *actual* behavior, $\ell ny_{n(i)}$, instead of the *expected* behavior of her neighbors. Unlike the main model in Manski (1993), there is no need to solve for the expectation of endogenous variables, because they are not present as explanatory variables. Furthermore, if we had specified the model in terms of the expected consumption of one's neighbors, then it would be necessary to use the full set of equations for all neighbors as a system of simultaneous equations to solve for the expectations, in the style of Moffitt, *op. cit.*. A straightforward consequence of our modelling choice would be that all neighbors' socioeconomic characteristics would show up in the reduced forms for the expectation. Their presence there should not, of course, be taken to imply evidence that they constitute contextual effects. Such a presence would simply follow from the logic of the simultaneous estimation system.

We now link our terminology with that used by the literature. The term $\beta\Pi_y(\ell ny_{n(i)})$ on the RHS of (5) denotes a function of neighbors' incomes. It reflects an *endogenous social effect*. Such a social effect is, of course, central to the notion of neighborhood effects: a person's behavior depends on the *actual* behavior of her neighbors, rather than the expected behavior of the population from which the sample is drawn. For example, when one's neighbors maintain their property, one keeps up with them. The term $\gamma\Pi_z(z_{n(i)})$ denotes a function of neighbors' characteristics. It expresses a *contextual effect*, an exogenous social effect which gives the effect of the neighbors' characteristics of potential interest, like racial and ethnic composition, income levels, and educational attainment. An individual's demand for housing reflects her neighbors' characteristics as a matter of taste.

housing is the property tax rate. See section 3.2 below.

However, this could be a proxy for an individual's expectations of the future development of the neighborhood. The term $\xi \ell n z_h$ reflects the *direct* effect of occupants' own characteristics upon housing expenditure, again as a matter of taste. However, income is handled separately by $\delta \ell n I_h$. In contrast, the term $\gamma \Pi_z \ell n z_{n(i)}$ reflects the impact of the neighbors' characteristics. 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. [6]

Were it not for the terms $\beta \Pi_y(\ell n y_{n(i)})$ and $\gamma \Pi_z(z_{n(i)})$ in the demand equation, our approach would be entirely conventional. Also, any theory that encompasses this model would not necessarily imply that the dependence of an individual's demand upon those of her neighbors is through a loglinear function of $y_{n(i)}$ and of $z_{n(i)}$. Obviously, it could be more general. Of particular interest are such extreme cases as social effects being a function of the maximum or minimum income, or of housing consumption, or more generally of their distribution within an individual's neighborhood. We plan to deal with such extensions in our follow-up study, Ioannides and Zabel, *op. cit.,* that incorporates neighborhood choice.

The analytical consequence of imposing Nash equilibrium within the neighborhood requires considering a simultaneous equations system along the lines of (5) for all owner-occupants of dwelling units in each neighborhood cluster κ, to which unit i belongs. For numerous reasons, including missing data and varying numbers of owners in clusters, the number of units in each cluster in the data can vary across clusters, and the number of equations in each simultaneous system will differ across clusters. This is a complication which is avoided by our use of average values among one's neighbors, which are much simpler to compute irrespective of varying neighborhood cluster sizes across the data.

It is important to clarify that ours is the first study that allows for endogenous neighborhood effects in a housing demand model with a properly specified price. Ioannides (2000) and Kiel and Zabel (1998) have estimated models of housing decisions using American Housing Survey data along with clusters data for 1985, 1989, and 1993. Ioannides emphasizes the simultaneity of property valuations in a model where an individual's valuation of a dwelling unit is related to its valuation

[6] We ignore this aspect of the model in the present paper, in that it would require a model of neighborhood choice. We address this issue in Ioannides and Zabel (2000).

in the previous wave and to those of the neighboring units within clusters, in addition to a unit's own characteristics and to those of its neighbors in the cluster. Because of conditioning on the lagged value of the endogenous variable, his model may be referred to as a pseudo demand model. Kiel and Zabel carry out hedonic house price regressions where in addition to characteristics of individual units they use characteristics of the clusters and of the census tracts within which the units in the sample lie, but do not attempt to estimate a demand model as such. Ioannides (2001a) estimates a model of neighborhood effects in maintenance decisions.

3.2 The Price of Housing Services

Estimation of the housing demand equation requires knowledge of the price of housing, p_κ. Note that p_κ appears (implicitly) on the LHS of equation (5) as well as on the RHS as a regressor. Once a measure of p_κ is available, we can calculate housing services using equation (4). A number of methods have been used to obtain the price of housing services. We follow the common approach of estimating the hedonic house price function. This results in an index of the relative value of a "constant quality" house for different housing markets.

The operation of the housing market may be visualized as a process of bidding[7] for different housing units and packages of neighborhood amenities by different individuals, so that the maximum valuations of each set of characteristics prevail as prices. Housing units, on the other hand, are produced by profit seeking entrepreneurs, and therefore cost considerations enter through the supply side and make values of houses be a function of the structural and other characteristics, q_i, including characteristics of the neighborhood, x_κ, where i indexes dwelling units and κ indexes neighborhood clusters. Invoking standard hedonic theory of housing markets [Rosen (1974)], we postulate that the relationship between housing values and housing and neighborhood characteristics, $W(q_i, x_\kappa, y_{n(i)}, z_{n(i)})$, may be estimated if suitable data are available.[8]

[7] This is a standard concept with a vast literature. See also Ellickson (1981) and Lerman and Kern (1983), who link explicitly the definition of house price hedonics with bidding models of housing markets.

[8] Formally, the hedonic function may be visualized as the outer envelope of the individual expenditure functions where different expenditure functions are parameterized by income. To see this, let housing services be produced from structural characteristics and neighborhood characteristics, $(q_i, x|\kappa, y_{n(i)}, z_{n(i)})$, by means of a household production function. Therefore, the expenditure function for housing, $V_{i\kappa h} = V(q_i, x_\kappa; y_{n(i)}, z_{n(i)}, u, I_h)$ may be defined as

$$U(I_h - V_{i\kappa h}, q_i, x_\kappa; y_{n(i)}, z_{n(i)}) = u;$$

$V_{i\kappa h}$ is the amount an individual is willing to pay for different values of housing and neighborhood characteristics, given a fixed level of utility, u, and conditional on a given choice of neighborhood with socioeconomic characteristics

In accordance with the above discussion, we specify a log-linear hedonic house price function

$$\ell n W_{i\kappa} = a_0 + a_1 \ell n q_i + a_2 \ell n x_\kappa + a_y \Pi_y \ell n y_{n(i)} + a_z \Pi_z \ell n z_{n(i)} + u_{i\kappa}, \tag{6}$$

where: the intercept a_0 absorbs the user cost, the parameters a_1, a_2, a_y, a_z are the marginal valuations of the respective characteristics, which may be estimated; Π_y and Π_z denote proximity matrices that define the relevant neighborhood for each unit and attribute an effect from the neighborhood $n(i)$ to the individual unit i; and $u_{i\kappa}$ denotes an unobserved random variable. As discussed above, the terms $\Pi_y \ell n y_{n(i)}$ and $\Pi_z \ell n z_{n(i)}$ can be simplified to express dependence of housing values on the average housing consumption and the average socioeconomic characteristics among the members of the neighborhood cluster [see Ioannides (2000)]. We note that the presence of both unit and neighborhood characteristics, (q_i, x_κ), and of the characteristics of neighbors $(y_{n(i)}, z_{n(i)})$, simply reflect the role of the latter as neighborhood amenities and therefore contextual effects. We emphasize that the characteristics of a unit's own occupant do not enter as determinants of a unit's value. It is those of the neighbors that do.

In order to obtain estimates of the price of housing services, we extend Equ. (6) to multiple housing markets which we define as MSAs.[9] Thus, Equ. (6) becomes

$$\ell n W_{i\kappa j} = a_0$$

$$+ \sum_{j=1}^{J-1} a_{0j} \mathrm{MSA}_{ij} + a_1 \ell n q_{ij} + a_2 \ell n x_{\kappa j} + a_{yj} \Pi_y \ell n y_{n(i)} + a_{zj} \Pi_z \ell n z_{n(i)} + u_{i\kappa j}, \; i = 1, \ldots, N_j, j = 1, \ldots, J, \tag{7}$$

where i indexes dwelling units; j indexes MSAs; N_j is the number of observations in MSA j; MSA_{ij} is dummy variable, which is equal to 1, for all units in MSA j, and to zero, otherwise; and the α_{0j}'s are the MSA-specific intercepts which we interpret as housing price index values. By leaving MSA J out of the equation, all other prices are interpreted as relative to MSA J. It is these MSA-specific numbers, the $\{\alpha_{01}, ..., \alpha_{0J}\}' \, s$, that we use as an instrument for prices, the p_κ's, that appears on the RHS of the demand Equ. (5) and on the RHS of Equ. (4) that defines housing services.

$(y_{n(i)}, z_{n(i)})$). Note that, in equilibrium

$$V(q_i^*, x_\kappa^*; y_{n(i)}, z_{n(i)}, u^*, I_h) = W(q_i^*, x_\kappa^*, y_{n(i)}, z_{n(i)}) \equiv W_{i\kappa}.$$

That is, the optimal expenditure V is identified with the house price hedonic function, defined in (6) below.

[9] See Mills and Simenauer (1996) for a related application in housing markets.

Note that it is important to include all observable determinants of housing in the house price hedonic, Equ. (7), for two reasons. First, regardless of the factors one wishes to account for in the house price index, it is necessary to include all observable determinants of house prices in (7) in order to obtain the most accurate estimates of the parameters in this equation. Second, since we wish to construct a housing price index for metropolitan areas of the U.S., it is important to compare a standard dwelling unit across different housing markets, which in our case are identified as MSAs. This entails that we account for both structural and neighborhood characteristics in our housing price index [Zabel (1999)].

3.3 Identifying the Model: The Endogeneity of Neighbors' Housing Demand

As is clear from the housing demand equation, the endogenous social effect, $\Pi_y \ell n y_{n(i)}$, is correlated with the error term since it includes the unobserved cluster effect (v_κ). In this paper, we use the average of the log of the neighbors' housing demands, $\overline{\ln y_{n(i)}}$, as this endogenous social effect. Thus, it is necessary to instrument for $\overline{\ln y_{n(i)}}$. Using the housing demand equations for all members of a neighborhood cluster (5) as a system to solve for $\overline{\ln y_{n(i)}}$, one can see that valid instruments will include $\ln p_\kappa$ and the means of neighbors' characteristics, including permanent income. To identify the model, we need instruments that are correlated with $\overline{\ln y_{n(i)}}$ but uncorrelated with the disturbance term, $v_\kappa + \epsilon_{\kappa h}$. We use the link to hedonic theory (as discussed in the previous subsection) to select appropriate instruments. In particular, we include the mean of the neighbors' exogenous structural characteristics as instruments. Thus we estimate the following equation for the mean of neighbors' housing services

$$\overline{\ln y_{n(i)}} = \pi_0 + \pi_1 \ln p_\kappa + \pi_2 \overline{\ell n z_{n(i)}} + \pi_3 \overline{\ell n q_{n(i)}} + \eta_{n(i)}, \tag{8}$$

where $\overline{\ell n z_{n(i)}}$ and $\overline{\ell n q_{n(i)}}$ denote, respectively, the means of neighbors' characteristics and the means of neighboring units' characteristics, and $\eta_{n(i)}$ is the unobserved error term. While one might question the validity of own structural characteristics as instruments, it should be noted that it is the means of one's neighbors' structural characteristics that appear on the RHS of (8), which does not include one's own structural characteristics. Also, we conduct an over-identification test for the validity of these instruments when we carry out the empirical analysis.

4 Empirical Results

This section presents the estimation results for the housing demand function (5). We estimate our model in four steps: one, we obtain estimates of the price of housing services, p_κ; two, we estimate permanent income; three, we estimate the reduced form for the mean of the neighbors' housing demand; and four, we use the predicted values from step three to estimate the structural housing demand equation (5). We discuss the implementation of this procedure and ensuing results from these four steps in turn.

4.1 A House Price Index

As is clear from the discussion in Section 3.2, we need to estimate market prices for housing. Recall that the clusters data in our sample come from 100 MSAs for the years 1985, 1989, and 1993. Ideally, we would like to have a price index that corresponds to the price of a "standard" dwelling unit and comes from a different data source that the actual data we use to estimate the neighborhood model. We were unable to find such a price index that covered all 100 MSAs for these three years. To underscore this difficulty, consider for example the Case-Shiller index [Case and Shiller (1987)]. It is available for 90 MSAs, but the use of this index results in a loss of about one third of the data since not all MSAs covered by the Case-Shiller index are among the 100 MSAs in our data set. Also, the Case-Shiller index would have to be made comparable across MSAs, which would require a technique like the one we implement below.

We follow the procedure outlined in Section 3.2 and estimate the hedonic house price model (6). We use the non-neighborhood clusters subsample of the NAHS data to estimate this model. We do so because the non-cluster data make up approximately 90% of the NAHS data and this results in a much larger data set. Also, this means that the prices are obtained from a different data set than the one that is used to estimate the housing demand equation.

The regressors in the house price hedonic, Equ. (6), include the structural characteristics of units, q_i, and neighborhood characteristics that are constant within each cluster, x_κ. In addition, the regressors must account for neighbors' property values and characteristics, $\Pi_y y_{n(i)}$ and $\Pi_z z_{n(i)}$, respectively. The non-neighborhood clusters subsample of the NAHS data does *not* include any such neighborhood characteristics. Therefore, we use individual characteristics of occupants to proxy for $\Pi_y y_{n(i)}$, and $\Pi_z z_{n(i)}$. We appeal to Kiel and Zabel (1998) for support of this step: they show that when actual neighborhood characteristics are unavailable, the use of owner characteristics is a reasonable alternative. Also Hardman and Ioannides (1998) show that the socioeconomic characteristics of neighbors in the cluster are highly correlated.

In order to estimate the housing demand equation for each of the three years (1985, 1989, and 1993), we need to estimate a separate house price hedonic for each of these years. We also estimate a housing demand model using the pooled data. In order for the price index to be comparable across years, we estimate the house price hedonic using the pooled data. Given that we define each MSA in each year to be a separate housing market, it might make sense to allow the coefficients to vary across time and space. This would lead to a large number of parameters in the model. Instead, we restrict the coefficients on all variables but the intercept to be constant. This is consistent with Mills and Simenauer (1996) who find very little variation in coefficients over time and space when they estimate a house price hedonic using national data from the House Financing Transaction Database, collected by the National Association of Realtors over the period 1986-1992. We do allow the coefficients to vary over time (though not across MSAs) but this only increases the estimated standard error of the regression by 0.05% so we use the restricted model for our analysis. Thus, for the pooled data, the model that we estimate is

$$\ell n W_{ijt} = a_0 + \sum_{j=1}^{J-1} a_{0jt}\mathrm{MSA}_{ijt} + a_1 q_{it} + a_2 x_{\kappa t} + a_z z_{ht} + u_{ijt}, \tag{9}$$
$$i = 1, \ldots, N_j, j = 1, \ldots, J,\ t = 1985,\ 1989,\ 1993,$$

where q_{it} includes the age of the unit and its square, the number of full baths, the number of bedrooms and whether or not there is a garage; $x_{\kappa t}$ includes a dummy variable that indicates whether or not the unit lies in the central city of the MSA; and z_{ht} includes household income, age of the owner, highest grade attained, and dummy variables that indicate if the owner is married, male, Black, or Hispanic. As mentioned in Section 2, we also include length of tenure and its square to capture the overvaluation of house prices by owners. Generally, one would like to include such variables in $x_{\kappa t}$ as school quality and crime rates, but these unfortunately are not available in the AHS. We include the owner's characteristics as proxies for neighborhood quality since they are likely to be correlated with these omitted variables and hence will reduce the bias in the coefficient estimates for the MSA dummy variables.

There are 140 potential MSAs but we require that there are at least ten observations in an MSA in a given year for observations in that MSA to be included in the regression. This leaves 10135, 10712, and 10815 observations for 1985, 1989, and 1993, respectively. The adjusted R^2s for these regressions are 0.505, 0.586, and 0.583, respectively. As a group, the neighborhood proxies are very significant. These regression results are not reported here but are available from the authors upon request. We use the estimated MSA dummies to calculate a price index for the MSA where unit i lies, $\hat{p}_\kappa$. For the yearly regressions, we set the price for Denver, the excluded MSA, to be 100 and those for the other MSAs to be 100 times the antilog of the corresponding coefficient estimate. For the pooled regression, Denver in 1985 is the reference point.

From the descriptive statistics reported in Table 7, we note that the dispersion in our estimate of housing prices, as measured by the coefficient of variation, increased considerably from .30 in 1985 to .38 in 1993. This increase is not inconsistent with the observed increased regional disparities in the dynamics of housing prices in the US over roughly that period [Poterba (1991)].

4.2 Estimating Permanent Income

It is standard practice to use permanent rather than current income in housing demand equations [Olsen (1987)]. Also, it is reasonable to assume that individuals are better able to predict their neighbors' permanent rather than current income given the larger fluctuations in the latter measure. We define permanent income as the predicted value from the following model of (the natural log of) current income:

$$\ell n I_{it} = \theta_0 + \sum_{j=1}^{J-1} \theta_{0jt} \mathrm{MSA}_{ijt} + \theta_1 c_{it} + \nu_{ijt},$$

where c_{it} includes a cubic polynomial in age and years of education, dummy variables that indicate if the owner is married, male, Black, or Hispanic and whether or not the unit lies in the central city of the MSA. We include the MSA dummy variables that are included in the house price hedonic regressions (9) to capture differences in the cost-of-living across MSAs. We use the non-neighborhood clusters subsample of the NAHS to estimate the model of income for the same reasons that led us to use this data to estimate the price index. We estimate a separate regression for each of the three years. The adjusted R^2's for 1985, 1989, and 1993 are 0.401, 0.374, and

0.293, respectively. Results are available from the authors on request. We then use the parameter estimates from these regressions, along with the data from the cluster subsample to estimate the permanent income variable that is included in the housing demand equations.

4.3 Estimating the Reduced Form for the Mean of the Neighbors' Housing Demand

As discussed in Section 3.3, the mean of the neighbors' housing demands, $\overline{\ln y_{n(i)}}$, is correlated with the disturbance term in the housing demand equation (5). Thus, we first estimate the reduced form equation for this variable according to Equ. (8). Table 8, column 1, reports the results for the pooled data. The crucial variables in this regression are the mean of the neighbors' structural characteristics since they identify the housing demand equation. Clearly they are very significant. The predicted values and residuals from this regression are referred to as $\widehat{\overline{\ell ny}}_{n(i)}$ and $\widehat{\eta_{n(i)}}$, respectively.

We test for the endogeneity of the mean of the neighbors' housing demand using a Hausman test. We estimate the housing demand equation using the observed mean of the neighbors' housing demand and include the residual from the reduced form equation, $\widehat{\eta_{n(i)}}$, as an additional regressor. The null hypothesis that the mean of the neighbors' housing demand is exogenous is rejected if the $t-$statistic for $\widehat{\eta_{n(i)}}$ is significantly different from zero [Hausman (1978)]. The $p-$value for the $t-$statistic is less than 0.0001 so the null hypothesis of exogeneity is rejected and the evidence indicates that the mean of the neighbors' housing demand is endogenous in the housing demand equation.

4.4 Estimating the Housing Demand Equation

As a point of reference when using the pooled data, we first estimate the standard housing demand equation that includes the price and own characteristics as the only regressors. We include only one observation per cluster to provide a data set that is comparable to those used in previous studies of housing demand. The results are presented in Table 8, column 2. The price elasticity is -0.199 and it is significantly different from zero. The income elasticity is 0.309 and also significant. While the price elasticity is somewhat low in magnitude, these results are generally comparable to those in the housing demand literature.[10]

As a next step, we estimate the reduced form equation for own housing demand using the pooled data. This equation results from substituting the mean neighbors housing demand equation (8) into the structural housing demand equation (5). This results in the means of the neighbors' socioeconomic and house characteristics becoming additional regressors. While the neighbors' consumption is obviously absent from this reduced form regression, the results are interesting in their own right, as they may be given a more limited interpretation of housing demand with neighborhood effects. That is, significant presence of contextual effects is evidence of neighborhood effects, even if endogenous effects may not be identified separately. The result is reported in Table 8, column 3. As a group, both the neighbors' socioeconomic characteristics and house characteristics are very significant. The former result is evidence of neighborhood effects, although it does not distinguish endogenous and/or contextual social effects, per se. As a group, the neighborhood effects are very important in explaining housing demand. The percentage of the explained sum of squares that is attributable to the neighborhood effects is 65.5. Note that the price elasticity is now large (relative to the estimate from the standard housing demand equation in Table 8, column 2) in magnitude, -0.427, and highly significant. The income elasticity is small, 0.119, but significant.

Finally, we estimate the structural housing demand equation (5). We include the predicted value from the reduced form regression for the mean neighbors' housing demand ($\widehat{\ell ny}_{n(i)}$). This allows us to identify the endogenous neighborhood effect. The means of the following neighbors' characteristics are treated as contextual neighborhood effects: household income and whether the owner is white, is married, graduated from high school, and whether moved into the house in the last five years. Excluded from the housing demand regressions are the means of the structural characteristics of the neighbors' units.

We estimate a log-linear housing demand model using cluster-specific random effects. The results are reported in Table 9. We provide results for each year of the three waves, 1985, 1989, and 1993. The coefficient estimates are quite stable over this period so we also provide the results for the pooled sample. While the test for parameter constancy is rejected (F-statistic= 2.4, p-value

[10]The study that is closest to this empirical example is Goodman (1988). He estimates a model of tenure choice and housing demand using the 1978 NAHS. The price and income elasticities are approximately -0.5 and 0.25, respectively.

= 0.00004), the percent increase in the residual sum of squares by imposing the restrictions is only 0.94%. Thus we focus our discussion, in the remainder of the section, on the results for the pooled model.[11] The price elasticity, −0.244, is significantly different from zero at 1%. The elasticity of own permanent income is 0.119 and is highly significant. The estimate of the price elasticity is comparable to that of the standard housing demand model (-0.199) but the income elasticity has decreased considerably (was 0.309). The remaining own characteristics are not significant in this regression, neither individually nor as a group.

The elasticity of housing demand with respect to the mean of the neighbors' housing demands is 0.660 and it is highly significant. This is evidence of a strong endogenous social effect. This result is in broad agreement with the finding by Ioannides (2001a) of an effect of maintenance by one's neighbors on own maintenance of 0.487. The means of neighbors' socioeconomic characteristics are jointly significant at the 3% level. Thus, this is weaker evidence of contextual effects. The only variable that is individually significant is the mean of neighbors' permanent incomes. The elasticity with respect to neighbors' permanent income is 0.175. One interpretation of this effect is that individuals see their neighbors' permanent income as an indication of their "permanent" housing consumption. Thus, a rise in the neighbors' permanent income would signal a rise in their future housing consumption which individuals anticipate by increasing their own current housing demand. It is somewhat surprising that this effect is larger than that of the owner's permanent income.

We test for the validity of the identifying instruments, that is the means of the neighbors' structural characteristics, using the, by now, standard test for overidentifying restrictions [see, for example, Wooldridge (2000)]. This test is carried out by first estimating the housing demand equation using only one of the mean neighbors' structural characteristics, so that the model is just identified. Let this variable be denoted q_1. The residuals from this regression are then regressed on all the exogenous variables, including the remaining mean neighbors' structural characteristics. The test of the null hypothesis that all instruments are uncorrelated with the disturbance term from

[11] Since the predicted, rather than the actual, value of the mean of neighbors' demand is included in the housing demand equation, it is necessary to adjust the standard errors for the random effects estimator. In this case, this results in a small change in the standard errors. The ratio of the adjusted standard error to the unadjusted standard error is 1.060, 1.061, 1.059, and 1.062 for the random effects regressions for 1985, 1989, 1993, and the pooled sample, respectively. Appendix A clarifies the steps involved in the necessary correction.

the structural equation is rejected if the test statistic (n$\cdot R^2$) is greater than the critical value χ^2_c where c is the number of overidentifying instruments (in this case c=4). The R^2 from this regression is 0.0003 and the test statistic has a p-value of 0.684. Thus the null hypothesis is not rejected and the mean neighbors' structural characteristics appear to be valid identifying instruments.

Note that it is necessary to assume that q_1, the one instrument that is used to just identify the housing demand equation, is exogenous in order to carry out the overidentification test. Thus, the fact that the mean of the neighbors' structural characteristics pass the overidentification test, does not totally confirm their validity as instruments. Recall that we have not accounted for residential choice in this model, so the error term might include neighbors' common (unobservable) characteristics that led them to choose the same location that might be correlated with the mean of the structural characteristics in the cluster. Because of such model uncertainty, we tried using a different, but related, set of instruments to compare results. Rather than using the means of the neighbors' structural characteristics, we use instead the structural characteristics for only one of each individual's (randomly chosen) neighbors in the cluster. An argument in favor of this choice is that the structural characteristics of one neighbor are less likely to be representative of the neighborhood than the mean of all the neighbors in the cluster, and hence will be more likely to be uncorrelated (or, at least, less correlated) with the unobserved cluster component of the error term in the housing demand equation. A drawback of using these instruments is that to the extent that the mean is a valid instrument, the use of only one neighbor's structural characteristics will lead to a less efficient estimator. Another drawback is the arbitrary nature of choosing as instruments a particular neighbor's structural characteristics (versus those of some other resident in the neighborhood). To deal with this latter problem, we carried out this exercise four times (using four different sets of neighbor's structural characteristics). We expected and did get approximately the same result each time.

In order to estimate the housing demand equation using the alternative sets of instruments, we must first estimate the reduced form equation for the mean of the neighbors' housing demand. The R^2s for the four regressions range from 0.226 to 0.243. These are substantially less than the R^2 for the reduced form equation when the means of the neighbors' structural characteristics are used (0.517). Next, we use the predicted values from these regressions to estimate the housing demand equation. As expected, the results for the four cases are quite similar. We present the results for a representative case in Column 5 of Table 9. This set of instruments also passes the test of overidentifying restrictions. We refrain from reporting details for reasons of brevity.

There are some important differences between these results and the ones presented earlier which use the means of the neighbors' structural characteristics as instruments. First, the endogenous effect has fallen from 0.660 to 0.188 (the range is 0.150 to 0.210) though it is still significant. Second, the coefficient on the mean of (the log of) neighbors' income has increased from 0.175 to 0.539 (the range is 0.520 to 0.570). Third, three of the five coefficients on the other means of the neighbors' characteristics are significant at the 5% level. As compared to the results based on the means of the neighboring units' structural characteristics, using only one of neighboring unit's structural characteristics as instruments results in a much larger portion of the neighborhood effects being accounted for by the contextual effect than by the endogenous effect. Still, the endogenous effect remains a significant factor in this latter case.

The conclusion to be drawn from this exercise is that *both* endogenous and contextual effects are present in the housing demand model though which one dominates is not clear. Also, the fact that the two sets of instruments produce such different results indicates that they may well be correlated with unobservables in the structural equation. As mentioned above, one source of such unobservables is the fact that we have not accounted for residential choice. In view of the considerable model uncertainty affecting our model, we feel more comfortable presenting both set of results, which provide a range of possibilities.

5 Conclusion

In this paper, we estimate a model of housing demand with neighborhood effects. Neighborhood effects produce a high degree of interdependence among neighbors's demands. We exploit special features of the National sample of the American Housing Survey and properties of housing markets that allow us to create natural instruments that allow us to estimate the model. We believe that we are the first researchers to estimate such a model. We find evidence of both endogenous and contextual neighborhood effects. The endogenous effect implies that individual housing demand is affected by the mean housing demand of one's neighbors. This is consistent with the concept of "keeping up with Joneses" which manifests itself in individual maintenance, repair, addition, and renovation decisions. That is, neighbors' decisions to maintain, repair, renovate, or make additions to their homes induce individuals to keep up by increasing their own housing consumption. The contextual effect implies that individual housing demand is influenced by the neighbors' characteristics. This could be a pure preference effect. Alternatively, it could be rationalized as owners' viewing their neighbors' characteristics, e.g. income, as signals of their future housing consumption which leads them to alter their own consumption accordingly.

A difficult aspect of models with social effects is identifying the endogenous and contextual effects. We attempt to identify these effects using as instruments the means of the structural characteristics of the dwelling units occupied by neighbors. We claim that these variables are correlated with the endogenous variable (the mean of the neighbors' housing demand) but uncorrelated with the error term in the own structural housing demand equation. One justification for this latter claim is that these means include only the neighbors' structural characteristics. In fact, when we use the own structural characteristics as instruments, the over-identification test is rejected. We also try another set of instruments that include the structural characteristics for the dwelling unit occupied by only one neighbor. Given that they are less representative of the neighborhood than the means of all neighbors' structural characteristics, these instruments may well be less correlated with the unobserved cluster effect in the structural own housing demand equation. The two sets of instruments provide a range of significant estimates for the endogenous and contextual effects. When the endogenous neighborhood effect is large the respective contextual effects are weak, and vice versa.

One factor that we have not accounted for in our model is residential choice. Ignoring residential choice can lead to a form of sample selection bias. For example, Rapaport (1997) shows that accounting for community choice in a model of housing demand can significantly affect the estimated price elasticity. The data that we have used for this analysis do not allow us to account for residential choice since they contain very limited information about the location of houses within the MSA. By gaining access to the US Bureau of the Census' own (but confidential) version of the NAHS, we plan to extend our model to include neighborhood choice. Not only will this allow us to alleviate the potential sample selection bias, but it also provides additional instruments for identifying the endogenous and contextual effects [Brock and Durlauf (2000)].

6 References

Angrist, Joshua D., and Alan B. Krueger (1995), "Split- Sample Instrumental Variables Estimates of the Return to Schooling," *Journal of Business and Economic Statistics,* 13, 2, 225–235.

Brock, William A., and Steven N. Durlauf (2000) "Interactions-Based Models," forthcoming in Heckman, James J, and Edward Leamer, eds., *Handbook of Econometrics,* 5.

Case, Anne C. (1991), "Spatial Patterns in Household Demands," *Econometrica,* 59, 4, July, 953-965.

Case, A. C. ,1992. Neighborhood Influence and Technological Change. Regional Science and Urban Economics 22, 491–508.

Case Karl E., and Robert J. Shiller (1987) "Prices of Single-Family Homes Since 1970: New Indices for Four Cities," *New England Economic Review,* 45–56.

Dusansky, Richard and John D. Wilson (1993), "The Demand for Housing: Theoretical Considerations," *Journal of Economic Theory,* 61, 1, 120–138.

Ellickson, Brian (1981), "An Alternative Test of the Hedonic Theory of Housing Markets," *Journal of Urban Economics,* 9, 56–79.

Goodman, A.C., (1988), "An Econometric Model of Housing Price, Permanent Income, Tenure Choice, and Housing Demand," *Journal of Urban Economics,* 23, 237-353.

Goodman, J. L., and J. B. Ittner (1992), "The Accuracy of Home Owners' Estimates of House Value," *Journal of Housing Economics,* 2, 339–357.

Hardman, Anna M., and Yannis M. Ioannides (1998), "Income Mixing and Housing in U.S. Neighborhoods," Department of Economics, presented at the AREUEA meetings, Chicago, January.

Hausman, Jerry A. (1978), "Specifications Tests in Econometrics," *Econometrica,* 46, 1251–1271.

Ioannides, Yannis M. (2000) "Interactive Property Valuations," Department of Economics, Tufts University, working paper.

< `http://ase.tufts.edu/econ/papers/index.html` >

Ioannides, Yannis M. (2001a) "Residential Neighborhood Effects," Department of Economics, Tufts University, working paper. Forthcoming *Regional Science and Urban Economics.*

< `http://ase.tufts.edu/econ/papers/index.html` >

Ioannides, Yannis M. (2001b) "Neighborhood Income Distributions," Department of Economics, Tufts University, June.

< `http://ase.tufts.edu/econ/papers/index.html` >

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

Kiel, Katherine, and Jeffrey E. Zabel (1998) "The Impact of Neighborhood Characteristics on House Prices: What Geographic Area Constitutes a Neighborhood?" Department of Economics, Wellesley College, Working paper 98-04, presented at the AREUEA meetings, Chicago, January.

Kiel, Katherine, and Jeffrey E. Zabel (1999), "The Accuracy of Owner Provided House Values: The 1978-1991 American Housing Survey," *Real Estate Economics, the Journal of the American Real Estate and Urban Economics Association,* 27, 263–298.

Lerman, Steven R., and Clifford R. Kern (1983), "Hedonic Theory, Bid Rents, and Willingness-to-Pay: Some Extensions of Ellickson's Results," *Journal of Urban Economics,* 358–363.

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

Mills, Edwin S., and Ronald Simenauer (1996), "How Hedonic Estimates of Regional Constant Quality House Prices," *Journal of Urban Economics,* 39, 209–215.

Moffitt, Robert A. (2001), "Policy Interventions, Low-Level Equilibria and Social Interactions," 45–82, in Durlauf, Steven N., and H. Peyton Young, eds., *Social Dynamics,* MIT Press, Cambridge, MA.

Olsen, Edgar O., (1987), "The Demand and Supply of Housing Service: A Critical Survey of the Empirical Literature," in Edwin S. Mills ed., *Handbook of Regional and Urban Economics*, Volume 2, Amsterdam: North-Holland.

Pollak, Robert A. (1976), "Interdependent Preferences," *American Economic Review,* 66, 309–320.

Poterba, James M. (1991), "Housing Price Dynamics: The Role of Tax Policy and Demography," *Brookings Papers on Economic Activity,* No. 2, 143–183.

Rapaport, Carol (1997), "Housing Demand and Community Choice: An Empirical Analysis," *Journal of Urban Economics,* 42, 243–260.

Rosen, Sherwin (1974), "Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition," *Journal of Political Economy,* 82, 34–55.

Wooldridge, Jeffrey M. (2000), *Introductory Econometrics: A Modern Approach,* Southwestern.

Zabel, Jeffrey E. (1999), "Controlling for Quality in House Price Indices," *Journal of Real Estate Finance and Economics,* 19, 223 241.

Zabel, Jeffrey E. (2001), "Continuous Models of Housing Demand," working paper, Tufts University, July.

APPENDIX A:

Calculation of the standard errors for the random effects estimator of the housing demand model with neighborhood effects

In this appendix, we show how the standard errors for the random effects estimator of the housing demand model (5) with neighborhood effects must be adjusted. This adjustment is similar to the one associated with the instrumental variables estimator and is dictated by the fact that we instrument for the endogenous neighborhood effect on the RHS of (5).

For this analysis, the endogenous neighborhood effect is the mean of the log of neighbors' housing demand, $\overline{\ell ny}_{n(i)}$. As defined in the text, $\widehat{\overline{\ell ny}}_{n(i)}$ and $\widehat{\eta_{n(i)}}$, are the predicted values and the residuals from the reduced form regression for the mean of the log of neighbors' housing demand (equation 8). Thus

$$\overline{\ell ny}_{n(i)} = \widehat{\overline{\ell ny}}_{n(i)} + \widehat{\eta_{n(i)}}.$$

By substituting into the demand Equ. (5), we have

$$\begin{aligned}\ell ny_{i\kappa h} &= \alpha + \mu \ell np_{\kappa} + \xi \ell nz_{h} + \delta \ell nI_{h} + \beta\left(\widehat{\overline{\ell ny}}_{n(i)} + \widehat{\eta_{n(i)}}\right) + \gamma \overline{\ell nz}_{n(i)} + \upsilon_{\kappa} + \epsilon_{\kappa h}, \\ &= \alpha + \mu \ell np_{\kappa} + \xi \ell nz_{h} + \delta \ell nI_{h} + \beta \widehat{\overline{\ell ny}}_{n(i)} + \gamma \overline{\ell nz}_{n(i)} + \upsilon_{\kappa} + \epsilon_{\kappa h} + \beta \widehat{\eta_{n(i)}},\end{aligned}$$

where $\overline{\ell nz}_{n(i)}$ stands for $\Pi_z \ell nz_{n(i)}$. Estimating this model by random effects results in an estimator of the variance of the regression equal to :

$$\hat{\sigma}^2 = \frac{1}{\mathrm{DF}} \sum_{\kappa=1}^{N} \sum_{h=1}^{n_\kappa} r_{\kappa h}^2,$$

where N is the number of clusters, n_κ is the number of units in cluster κ, DF is the degrees of freedom in the housing demand equation, and $r_{\kappa h}$ is the residual from the random effects regression. This will lead to the wrong standard errors because of the fact that the predicted rather than the actual mean of the log of housing demand is included on the RHS of the housing demand equation. It is possible to fix the standard errors for this estimator by following the approach taken in the standard instrumental variables estimator case. That is, the corrected standard error is calculated by using the following estimator of the variance of the regression,

$$\tilde{\sigma}^2 = \frac{1}{\mathrm{DF}} \sum_{\kappa=1}^{N} \sum_{i=1}^{n_\kappa} r_{i\kappa}^{*2},$$

where

$$r^*_{i\kappa} = r_{i\kappa} - \hat{\beta}(\widehat{\eta_{n(i)}} - \hat{\theta}_\kappa \overline{\widehat{\eta_{n(i)}}}),$$

and

$$\hat{\theta}_\kappa \equiv 1 - \sqrt{\frac{\hat{\sigma}^2_\epsilon}{\hat{\sigma}^2_\epsilon + n_\kappa \hat{\sigma}^2_\upsilon}},$$

and $\overline{\widehat{\eta_{n(i)}}}$ is the cluster mean of $\widehat{\eta_{n(i)}}$.

One other issue is that $\hat{\theta}_\kappa$ will also be incorrect because it depends on $\hat{\sigma}^2_\epsilon$, the estimated regression variance from the fixed effects model, which in turn is affected by the fact that the predicted rather than the actual value of the mean housing demand is included in the housing demand equation. But this impact should be minor and can probably be ignored. This follows from the fact that the mean of the log of housing demand among one's neighbors is nearly constant across the units that make up the cluster. That is, the fixed effects residuals must be corrected by subtracting out $\beta \frac{\widehat{\eta_{n(i)}} - \overline{\widehat{\eta_{n(i)}}}}{n_\kappa}$. This adjustment should make a small difference, particularly given the division by n_κ and hence it should make little difference in calculating $\hat{\sigma}^2_\epsilon$.

Table 1: **Neighborhood Cluster Counts: A**

Clusters Present Only in 1985	6
Clusters Present Only in 1985 and 1989	14
Clusters Present Only in 1985 and 1993	9
Clusters Present in 1985, 1989, and 1993	196
Total Clusters in 1985	225
Clusters Present Only in 1989	10
Clusters Present Only in 1989 and 1985	14
Clusters Present Only in 1989 and 1993	45
Clusters Present in 1985, 1989, and 1993	196
Total Clusters in 1989	265
Clusters Present Only in 1993	85
Clusters Present Only in 1993 and 1985	9
Clusters Present Only in 1993 and 1989	45
Clusters Present in 1985, 1989, and 1993	196
Total Clusters in 1993	335

Table 2: **Neighborhood Cluster Counts: B**

	1985	1989	1993	Total
Clusters Present in Only One Year	6	10	85	101
	1985/1989	1985/1993	1989/1993	Total
Clusters Present in Two Years	14	9	45	68
Clusters Present in All Three Years				196
Total Number of Clusters				365

Table 6: DEFINITIONS OF REGRESSION VARIABLES

Table 3: **Neighborhood Frequency Counts**

	Total Units in Cluster		Total Owner-Occupied Houses in Cluster	
Number in Cluster	Clusters	Frequency of Number of Clusters	Clusters	Frequency of Number of Clusters
6	2	0.55	54	14.79
7	10	2.74	55	15.07
8	29	7.95	69	18.90
9	53	14.52	67	18.36
10	107	29.32	59	16.16
11	141	38.63	58	15.89
12	19	5.21	2	0.55
13	3	0.82	1	0.27
14	1	0.27		

Table 4: **Neighborhood Unit Counts: A**

Houses Present Only in 1985	133
Houses Present Only in 1985 and 1989	224
Houses Present Only in 1985 and 1993	142
Houses Present in 1985, 1989, and 1993	1448
Total Houses in 1985	1947
Houses Present Only in 1989	153
Houses Present Only in 1989 and 1985	224
Houses Present Only in 1989 and 1993	493
Houses Present in 1985, 1989, and 1993	1448
Total Houses in 1989	2318
Houses Present Only in 1993	826
Houses Present Only in 1993 and 1985	142
Houses Present Only in 1993 and 1989	493
Houses Present in 1985, 1989, and 1993	1448
Total Houses in 1993	2909

Table 5: **Neighborhood Unit Counts: B**

	1985	1989	1993	Total
Units Present in Only One Year	133	153	826	1112
	1985/1989	1985/1993	1989/1993	Total
Units Present in Two Years	224	142	493	859
Units Present in All Three Years				1448
Total Number of Units				3379

	HOUSING DEMAND VARIABLES
LNHDEM	natural log of housing demand
PRICE	House price index obtained from the hedonic regression
LNINCOME	Natural log of household income
LNPINCOME	Natural log of permanent household income
HIGH SCHOOL	=1 if owner graduated from high school, =0 otherwise
WHITE	= 1 if owner is white, =0 otherwise
NPERSONS	Number of persons in the household
MARRIED	= 1 if owner is married, = 0 otherwise
CHANGED HANDS	= 1 if house changed hands in last five years, = 0 otherwise
LNHDEMM	Cluster mean of LNHDEM
LNINCOMEM	Cluster mean of LNINCOME
LNPINCOMEM	Cluster mean of LNPINCOME
HIGH SCHOOLM	Cluster mean of HIGH SCHOOL
WHITEM	Cluster mean of WHITE
NPERSONSM	Cluster mean of NPERSONS
MARRIEDM	Cluster mean of MARRIED
CHANGEHANDM	Cluster mean of CHANGEHAND
Dummy89	Dummy89 = 1 if 1989 obs, = 0, otherwise
Dummy93	Dummy93 = 1 if 1993 obs, = 0, otherwise
	HOUSE PRICE HEDONIC EQUATION VARIABLES
LNVALUE	Natural log of owner-estimated value of the house
CENCITY	=1 if house in central city of SMSA, =0 otherwise
HAGE	the age of the house in years
HAGESQ	square of HAGE
GARAGE	=1 if the house has a garage, =0 otherwise

BEDROOMS	number of bedrooms in the house
FULLBATHS	number of full bathrooms in the house
TENURE	years that owner has lived in house
GRADE	Highest grade attained by owner
MALE	=1 if owner is male, = 0 otherwise
HISP	=1 if owner is Hispanic, = 0 otherwise
HAGEM	Cluster mean of HAGE
HAGESQM	Cluster mean of HAGESQ
GARAGEM	Cluster mean of GARAGE
BEDROOMSM	Cluster mean of BEDROOMS
FULLBATHSM	Cluster mean of FULLBATHS

Table 7: MEANS AND STANDARD DEVIATIONS FOR REGRESSION VARIABLES

Variable	1985	1989	1993	POOLED
VALUE/1,000	93.56	138.10	133.27	124.06
	55.89	94.23	85.23	83.73
PRICE	82.11	122.57	119.03	105.11
	24.92	51.67	45.46	43.80
INCOME/1,000	40.52	47.79	51.40	47.28
	28.10	34.83	37.23	34.46
HIGH SCHOOL	0.836	0.860	0.870	0.858
	0.37	0.35	0.34	0.35
CHANGED HANDS	0.289	0.317	0.278	0.294
	0.45	0.47	0.45	0.46
WHITE	0.879	0.866	0.848	0.862
	0.325	0.341	0.359	0.344
NPERSONSM	2.851	2.768	2.738	2.779
	1.439	1.470	1.436	1.448
MARRIED	0.713	0.654	0.645	0.666
	0.452	0.476	0.479	0.472
Observations	1947	2318	2909	7174

Table 8: REGRESSION RESULTS FOR LOG-DEMAND EQUATION :

Reduced Forms for Pooled Data

(Standard Errors in Parentheses)

Variable	Mean of Neighbors' Housing Demand	Standard Housing Demand - No Neighborhood Effects	Housing Demand Reduced Form
LNPRICE	-.382** (.028)	-.199** (.071)	-.427** (.082)
LNINCOME		.309** (.055)	.119** (.015)
HIGH SCHOOL		.057 (.060)	.003 (.016)
CHANGED HANDS		.022 (.040)	.012 (.011)
WHITE		.230** (.053)	.013 (.016)
NPERSONS		.009 (.014)	.006 (.004)
MARRIED		-.051 (.050)	.011 (.013)
Dummy89	-.127** (.010)	-.081 (.049)	-.129** (.031)
Dummy93	-.191** (.011)	-.086 (.048)	-.192** (.034)
LNPINCOMEM	.494** (.026)		.422** (.069)
HIGH SCHOOLM	.108** (.026)		.134 (.071)

Table 8: CONTINUED

Variable	Mean of Neighbors' Housing Demand	Standard Housing Demand - No Neighborhood Effects	Housing Demand Reduced Form
CHANGED HANDSM	.168** (.027)		-.038 (.05)
WHITEM	.206** (.014)		.165** (.042)
NUMBER PERSONM	-.052** (.007)		-.043** (.019)
MARRIEDM	.033 (.024)		.116 (.065)
HAGEM	.004** (.001)		.008** (.003)
HAGESQM	-.006 (.001)		-.009** (.003)
FULLBATHSM	.259** (.008)		.230** (.023)
BEDROOMSM	.155** (.009)		.032 (.025)
GARAGEM	.210** (.013)		.177** (.037)
CONSTANT	2.107** (.131)	4.401** (.334)	2.285** (.362)
Observations	7174	7174	7174
R-SQUARE: WITHIN R-SQUARE: BETWEEN R-SQUARE: OVERALL	 ..617	 .133	.001 .588 .411
S.D. of RE S.D. Regression error Per cent of variance due to RE	 .279 		.307 .307 .427

Note: *, ** indicate significant at the 5, 1 percent significance levels.

Table 9

REGRESSION RESULTS FOR LOG-DEMAND EQUATION:
Cluster Variables Included

(Standard Errors in Parentheses)

Variable	1985	1989	1993	POOLED	POOLED–alt.instr
LNHDEMM	.710**	.737**	.621**	.660**	.189**
	(.078)	(.074)	(.072)	(.047)	(.037)
LNPRICE	-.230	-.160	-.281*	-.244*	-.584**
	(.145)	(.133)	(.129)	(.079)	(.081)
LNPINCOME	.136**	.084**	.140**	119**	.148**
	(.026)	(.024)	(.022)	(.014)	(.015)
HIGH SCHOOL	-.028	.037	-.002	.004	.009
	(.026)	(.027)	(.024)	(.015)	(.016)
CHANGED HANDS	.032	.038	-.025	.014	.009
	(.017)	(.018)	(.015)	(.010)	(.010)
WHITE	.077*	.008	-.009	.015	.006
	(.030)	(.029)	(.022)	(.015)	(.017)
NUMBER PERSONS	-.004	.003	.010	.004	.003
	(.006)	(.006)	(.006)	(.004)	(.004)
MARRIED	-.005	.028	-.006	.010	.018
	(.023)	(.022)	(.019)	(.012)	(.013)
LNPINCOMEM	.146	.128	.191	.175*	.539**
	(.124)	(.117)	(.113)	(.071)	(.071)
HIGH SCHOOLM	.039	.064	.053	.062	.157*
	(.110)	(.118)	(.113)	(.067)	(.071)
CHANGED HANDSM	-.005	-.047	-.069	-.032	-.061
	(.064)	(.077)	(.079)	(.043)	(.046)

Note: *, ** indicate significant at the 5, 1 percent significance levels.

Table 9: CONTINUED

WHITEM	-.075 (.069)	.008 (.068)	.071 (.061)	.019 (.038)	.078 (.041)
NUMBER PERSONM	-.038 (.029)	-.018 (.028)	-.008 (.026)	-.023 (.016)	-.044* (.017)
MARRIEDM	.041 (.111)	.037 (.107)	.058 (.096)	.054 (.061)	.144* (.065)
Dummy89				-.065* (.029)	-.137** (.031)
Dummy93				-.089** (.032)	-.193** (.033)
CONSTANT	.228 (.576)	.176 (.564)	.278 (.529)	.381 (.316)	1.018** (.340)
Observations	1947	2318	2909	7174	7174
R-SQUARE: WITHIN	.0006	.0005	.0011	.0003	.0072
R-SQUARE: BETWEEN	.6150	.6165	.5914	.5921	.4584
R-SQUARE: OVERALL	.4273	.4293	.4166	.4136	.3280
S.D of RE	.2393	.2740	.2675	.2644	.2653
S.D. Regression Error	.2888	.3291	.3033	.3075	????
Per cent Variance due to RE	.4081	.4094	.4376	.4250	.4194
Per cent explained Variation due to Cluster vars					
P-Value for cluster vars	.6678	.8180	.2326	.0279	.001