

The research program of the Center for Economic Studies (CES) produces a wide range of theoretical and empirical economic analyses that serve to improve the statistical programs of the U.S. Bureau of the Census. Many of these analyses take the form of CES research papers. The papers are intended to make the results of CES research available to economists and other interested parties in order to encourage discussion and obtain suggestions for revision before publication. The papers are unofficial and have not undergone the review accorded official Census Bureau publications. The opinions and conclusions expressed in the papers are those of the authors and do not necessarily represent those of the U.S. Bureau of the Census. Republication in whole or part must be cleared with the authors.

**LOCATION, LOCATION, LOCATION:
THE 3L APPROACH TO HOUSE PRICE DETERMINATION**

by

Katherine A. Kiel*
College of the Holy Cross

and

Jeffrey E. Zabel *
Tufts University

CES 04-06 May, 2004

All papers are screened to ensure that they do not disclose confidential information. Persons who wish to obtain a copy of the paper, submit comments about the paper, or obtain general information about the series should contact Sang V. Nguyen, Editor, Discussion Papers, Center for Economic Studies, Washington Plaza II, Room 206, Bureau of the Census, Washington, DC 20233-6300, (301-763-1882) or INTERNET address snguyen@ces.census.gov.

Location, Location, Location: The 3L Approach to House Price Determination

Latest Version: March 2004

Katherine A. Kiel
Department of Economics
College of the Holy Cross
kkiel@holycross.edu

Jeffrey E. Zabel
Department of Economics
Tufts University
jeff.zabel@tufts.edu

The authors appreciate the help of Housing and Household Economics Statistics (HHES) Division staff (U.S. Bureau of the Census) in compiling the data. The research in this paper was conducted while Zabel was a research associate at the Boston Data Research Center. Research results and conclusions expressed are those of the authors and do not necessarily indicate concurrence by the Bureau of the Census. This paper has been screened to insure that no confidential data are revealed.

Location, Location, Location: The 3L Approach to House Price Determination

Abstract

The immobility of houses means that their location affects their values. This explains the common belief that three things determine the price of a house: location, location, and location. We use this notion to develop the 3L Approach to house price determination. That is, prices are determined by the Metropolitan Statistical Area (MSA), town, and street where the house is located. This study creates a unique data set based on data from the American Housing Survey (AHS) consisting of small ‘clusters’ of housing units with information on their housing characteristics and resident characteristics that is merged with census tract-level attributes. We use this data to verify the 3L Approach: we find that all three levels of location are significant when estimating the house price hedonic equation. This indicates that individuals care about their local neighborhood, i.e. the general upkeep of their street and possibly their neighbors’ characteristics (cluster variables), a broader area such as the school district and/or the town (tract variables) that account for school quality and crime rates, and the particular amenities found in their MSA.

1. Introduction

Housing is an unusual good in three dimensions: heterogeneity, durability, and immobility. It is this last factor that causes the location of the house to be an important determinant of its value, since the purchaser buys both the dwelling and the site where the house is located. In fact, there is an oft-used phrase by real estate agents that three things determine the price of a house: location, location, and location.

On the broadest level, house prices are influenced by the MSA in which the house is located. MSA specific amenities such as temperature, proximity to bodies of water, and cultural attractions will cause similar houses to have different prices across MSAs. These amenities are explicitly considered in the quality of life literature (e.g. Roback (1982) and Gyourko and Tracy (1991)). On a smaller scale, Tiebout (1956) first suggested that individual residential location decisions are motivated by the quality of local public goods. Hence the quality of local public goods should also influence house prices. Finally, on a very local level, Schelling (1971) developed a model where decisions about where to live are affected by the characteristics of the other residents in the neighborhood. In particular, he posed a model of tipping behavior where a critical mass of blacks residing in a neighborhood would cause whites to exit.

The phrase “location, location, location,” (3L) could be applied to these three levels of geography that affect house prices: MSA, town, and neighborhood. In this paper, we develop the 3L Approach to house price determination and empirically verify that these three levels of geography simultaneously affect house prices.

An analysis of the different levels of geography that affect house prices ties into the literature that attempts to clarify the concept of “neighborhood.” In particular, when included in

house price hedonics, town-wide and street-level variables are often referred to as neighborhood characteristics. Different definitions are used in the literature and our results shed light on what a ‘neighborhood’ is in terms of what area matters to households.

The main source of our data is the American Housing Survey (AHS) which has provided information on house and owner characteristics for thousands of housing units in the United States since the early 1970s. However, the public-use version of the AHS lacks accurate information on the units’ locations, indicating only what metropolitan area they are located in, and whether they are in the central city of that area. Also, the only measure of neighborhood quality is an index that is reported by the resident. In an effort to rectify this problem, some units were selected at random in 1985, 1989 and 1993 and the ten nearest housing units were ‘interviewed’.¹ These ‘clusters’ are the source of the street-level data in this analysis.

The data set used for this analysis has information on these ‘clusters’ of houses including the structure attributes of up to eleven houses in the cluster, the characteristics of the residents, the AHS index, and averages over the houses and residents in the cluster. Census tract-level measures are included by accessing proprietary AHS data.

In this paper, we test the 3L Approach to house price determination by jointly testing the significance of variables at these three levels of geography in the house price hedonic. We are not interested in the significance of particular variables because we do not have information on specific town, or street-level amenities. Instead, we have proxies for these variables at the different geographic levels. Thus, while there are studies that confirm the importance of two of

¹This data set has been used by, among others, Hoyt and Rosenthal (1997), Gabriel and Rosenthal (1996a, 1996b), Hardman and Ioannides (1997), Ioannides (2000), and Ioannides and Zabel (2003a, 2003b).

these levels by finding that specific MSA, town, and/or street-level amenities are significant (e.g. Li and Brown 1980), it has not been shown that all three are jointly significant on a broad geographic scale. One reason for this is that it is very difficult to obtain information on specific town and street-level amenities for a large number of MSAs.

Our results show that the cluster, tract, and MSA variables are jointly significant and hence provide support for the 3L Approach. Individuals care about their local neighborhood, i.e. the general upkeep of their street and possibly their neighbors' characteristics (cluster variables). They also value amenities at the town-level (tract variables). Finally, individuals are willing to pay different amounts for houses with the same structure, block-level, and town characteristics that are located in different MSAs. This is consistent with the theory that they value the different amenities provided across MSAs.

Given that the cluster variables are not generally available, we estimate a model that excludes them. We compare the coefficient estimates for the tract, MSA, and structure characteristics that result from this model with those from the full model. When using the preferred estimator (fixed effects), we find that the coefficient estimates for the tract and structure characteristics and for the price index are not significantly affected by the exclusion of the cluster variables though there is some evidence that the coefficients for the regional and possibly the MSA specific dummy variables are biased.

The paper is structured as follows. Section 2 contains a literature survey and a discussion of the concept of 'neighborhood.' Section 3 provides the theory underlying the 3L Approach. A description of the data is given in Section 4 and some basic statistics are provided. Section 5 presents the model and empirical results and concluding remarks are given in Section 6.

2. Literature Survey

This paper ties into the research on neighborhoods that has occurred in many disciplines such as sociology, urban planning, political science, and economics. Despite the intense analysis, there is little agreement as to what constitutes a neighborhood. The notion of neighborhood that is closest to this analysis is the existence of a ‘collective externality’ that is felt by a distinct group of individuals in a given area (Segal 1979). Galster (1986) uses the term ‘externality space’ which he defines as “the area over which environmental changes initiated by others are perceived as altering the well-being (psychological or financial) a given individual derives from the given location.” (Pg. 246) Galster notes that one way to operationalize this concept of neighborhood is through hedonic regressions using house values. Variables would need to be “carefully specified as proxies for the existence of a variety of externalities at various distances from the given observation.” (Pg. 255)

Galster’s reliance on externalities as a means for determining neighborhoods relates to the literature that attempts to measure “neighborhood effects” as the coefficients associated with distances to various residential and nonresidential land uses in hedonic house price regressions. Strange (1992) surveys this literature and notes that these studies have had minimal success in finding significant effects. However, most of these studies do not include other variables to capture the impact of location on house values (such as the provision of public goods and neighborhood appearance) so the estimates of the neighborhood effects are likely biased.

These results imply that not only will neighborhood characteristics be significant in explaining house prices but that the particular variables that are significant will convey information about what homeowners consider to be their “neighborhood.” These variables can

include the measures of local public services like schools and security (as measured by crime rates), accessibility to work, highways, and commercial centers, distance to facilities such as schools, parks and other recreation facilities, and industrial centers, the characteristics of neighbors such as their race and education level, the quality of other houses in the area, noise levels, and general measures of aesthetics. The extent to which these variables are included in a given analysis depends crucially on the data that are being analyzed. Generally, it is quite difficult to get accurate measures of many of these variables. Much of these data are not available at a disaggregated enough level to be useful.

One study that includes most of the above variables is by Li and Brown (1980). They estimate a house price hedonic that includes structure and site attributes, census tract summary variables (such as median income, the percent of those 16-21 years old who are high school dropouts, residential density, and air pollution), and local public services and costs (such as school test scores and the local property tax, and an accessibility measure - distance to the Central Business District (CBD)). They also include “micro-neighborhood” characteristics such as noise pollution, a local aesthetic quality index, and distances to the ocean, river, expressway interchange, school, industry, and commercial areas. The data are 781 sales of single-family houses in 1971-1972 in 15 suburban towns in the southeast sector of the Boston MSA.

Their results show that adding the micro-neighborhood characteristics significantly reduces the impact of median income while the impact of distance to the CBD becomes negative and significant. This is evidence that aggregated socioeconomic variables like median income are proxies for local characteristics such as aesthetics and noise level rather than indicators of the desirability of neighborhoods with high income residents.

Two papers that look at the impact of using census data at different levels of aggregation are Goodman (1977) and Schultz and King (2002). Neither paper includes variables at different levels of aggregation in the same regression. Goodman (1977) estimates a hedonic model using sales data from 1967-1969 in New Haven, CT. He includes neighborhood variables aggregated to the tract and block group levels in separate regressions. The R^2 for the latter model is slightly higher (0.8388 vs 0.8353). Generally, the coefficient estimates are comparable across regressions. One important difference is the impact of the racial variables which are significant with the expected signs when aggregated at the block group level but not at the tract level.

Schultz and King (2002) use a model of hedonic house prices to estimate the marginal prices of open space amenities and non-residential land use. They use house prices for Tucson AZ aggregated at the census block level from the 1990 Decennial Census. They look at whether the coefficient estimates for the open space amenities and non-residential land use variables are affected when these variables are aggregated at the block, block group, and census tract level. They find that aggregation does not impact the coefficient estimates for the open space amenity variables but they recommend aggregation of land use variables to the block group level rather than the block or tract level.

Clapp and Wang (2003) use the technique of Classification and Regression Trees (CART) to determine the boundaries of neighborhoods. As the authors note, there are numerous reasons to question the use of census tracts or school districts as the basis for determining neighborhoods. The authors state that the goal of their paper is to “replace census tract boundaries with a statistical method.” (Pg. 7) These neighborhoods can then be aggregated to form sub-markets. In their application using sales transactions from West Hartford CT, Clapp

and Wang find fewer neighborhoods using the CART method than existing census tracts. They also find that the boundaries are more logical since they run behind houses and not down the middle of streets.

3. Theory

Let individual i 's utility function depend on non-housing composite consumption (C_i), structure characteristics of the house (sc_i), individual characteristics (z_i), income (y_i), local geographic characteristics (i.e. street-level - S) that depend on one's neighbors' characteristics ($z_{n(i)}$) and other amenities associated with street s , A_s , town-wide characteristics (T) that include locally provided public goods (G_w) and other town-wide amenities (A_w) for town w , and MSA-wide characteristics (M) that are made up of amenities A_m for MSA m . Then an individual maximizes utility subject to a budget constraint that includes expenditures on non-housing consumption C_i (price is numeraire) and housing ($r \cdot P(sc, S, T, M)$) where r is the interest rate and $P(\bullet)$ is the price of a house h with characteristics sc , S , T , and M

$$\max U\left(C_i, sc, S(z_{n(i)}, A_s), T(G_w, A_w), M(A_m), z_i\right) \quad (1)$$

$$\text{s.t. } C_i + r \cdot P\left(sc, S(z_{n(i)}, A_s), T(G_w, A_w), M(A_m)\right) = y_i.$$

To motivate the link between individual preferences for S , T , and M and the house price hedonic, define the expenditure function $E(z_i, sc, S(z_{n(i)}, A_s), T(G_w, A_w), M(A_m), u)$. $E(\cdot)$ is the solution to the equation

$$V(y_i - E(z_i, sc, S, T, M, u), sc, S(z_{n(i)}, A_s), T(G_w, A_w), M(A_m), z_i) = u \quad (2)$$

where V is the indirect utility function. $E(\cdot)$ is the maximum amount an individual is willing to spend for a house with characteristics sc , $S(z_{n(i)}, A_s)$, $T(G_w, A_w)$, and $M(A_m)$ given u while $P(sc, S(z_{n(i)}, A_s), T(G_w, A_w), M(A_m))$ is the minimum price that the individual must pay to obtain this house. In equilibrium we have

$$\begin{aligned} & E(z_i, sc^*, S(z_{n(i)}^*, A_s^*), T(G_w^*, A_w^*), M(A_m^*), u) \\ & = r \cdot P(sc^*, S(z_{n(i)}^*, A_s^*), T(G_w^*, A_w^*), M(A_m^*)) \end{aligned} \quad (3)$$

where the “*”s indicate the optimal level of the good that comes from individual utility maximization. This condition equates the optimal housing expenditure with the hedonic house price function. The optimal allocation occurs where the bid function is tangent to the hedonic price function, e.g.

$$\frac{\partial E}{\partial A_s} = r \cdot \frac{\partial P}{\partial S} \cdot \frac{\partial S}{\partial A_s} \quad (4)$$

There is, therefore, a relationship between the individual’s willingness to pay for a particular (street-level) amenity and the hedonic house price function. The right hand side is a function of the coefficients in the hedonic house price equation. Finding significant coefficients related to A_s (or A_w or A_m for that matter) indicates that individuals value A_s in terms of the utility it provides them by locating on that street (town or MSA). Thus, which level of location is important to individuals can be determined by empirically estimating this hedonic house price equation.

Further, it is possible to express the expenditure function as the product of the price of housing services, $p(A_m)$, and the amount of housing services, $H(sc, z_{n(i)}, A_s, G_w, A_w)$ if it is assumed that H is homogeneous of degree one in its attributes (Sieg et al 2002). $p(A_m)$ is the unit price of housing services in a given market, here assumed to be the MSA. Substituting into (3) and taking logs gives

$$\ln P = -\ln r + \ln p(A_m) + \ln H(sc, z_{n(i)}, A_s, G_w, A_w). \quad (5)$$

The price of the house, P , depends on street-level characteristics, $z_{n(i)}$ and A_s , town-wide characteristics, G_w and A_w , and the unit price of housing services that varies across MSAs. Thus the price of the house is affected by these three levels of geography, street, town, and MSA. We call this the “3L Approach” to house price determination.

In this paper, we are interested in determining which levels of location, S , T , and/or M are significant rather than the specific elements of these levels of location. This is important because we do not have information on relevant town amenities. Instead, we have proxies for these variables. At the street-level, we have the neighbors’ characteristics but we do not have measures of noise, upkeep, or distances to local land uses as do Li and Brown (1980). Thus, while there are studies that confirm the importance of one or two of these levels by finding that specific MSA, town, and/or block-level amenities are significant, it has not been shown that all three are jointly significant on a broad geographic scale. One possible reason for this is that it is very difficult to obtain information on specific town and street-level amenities for a large number of MSAs.

4. Data

The main data source for this study is the National version of the American Housing Survey (NAHS). Beginning in 1974, the AHS contains detailed information on particular houses through time that includes the current owner's evaluation of the house price, house characteristics, and self-reported information on the house's current occupants. The NAHS is conducted every two years. In 1985, 680 housing units (kernels) were selected at random, and the ten nearest housing units (cluster) were interviewed, yielding a data set of 7,350 housing units. This was repeated in 1989 and 1993 resulting in a three period panel of houses.²

We use this data set to create multiple measures of neighborhood quality. One measure is based on the averages over the houses in the cluster.³ We create six variables: mean household income, median age, the proportion of residents who have graduated from high school, the proportion of residents who are nonwhite, the proportion of units vacant, and the proportion of houses that changed hands in the past five years. These variables will serve as proxies for the quality of the area closely surrounding the house.

A census tract is defined as a homogeneous area in regard to the characteristics of the population, their economic status, and living conditions. A tract generally has between 2,500 and 8,000 residents. This defines a much larger neighborhood than the cluster data. Detailed information from the decennial Censuses is available in the Summary Tape Files (STF). The STF provides information for different geographical units (including census tracts) that are weighted

²In 1989 there were 769 kernel houses (with regular interview status) and in 1993 there were 1018 kernel houses (with regular interview status).

³The kernel house is not included in these averages.

to represent the total population. The tract information includes the same six variables that we created from the cluster data. This provides another set of proxies for various neighborhood characteristics that affect house prices.

The internal AHS files identify the census tract in which each house is located and through special arrangements with the U.S. Census Bureau we have access to these proprietary data. We merge the information from the 1980 and 1990 STF's with the AHS data by census tract. We use weighted averages of the 1980 and 1990 STF data to create the tract variables for the 1985, 1989, and 1993 surveys.⁴

For most years, the AHS lacks objective information on neighborhood quality. The only publicly available variable is the owner's self-assessment of the neighborhood. Starting in 1984 owners were asked to rate their neighborhood on a scale of 1 to 10. The subjective nature of this measure of neighborhood quality is problematic because people are less likely to down-rate their own neighborhood due to a stigma effect; they do not want to admit they live in an undesirable neighborhood. Also, satisfaction with a neighborhood may be more reflective of the individual's outlook on life rather than the physical characteristics of the local area (Keller 1982). Another problem is that this measure of neighborhood quality is affected by Tiebout sorting. People make their choice of residence based on their particular tastes for the services the neighborhood provides; they do not randomly choose the location.

The advantages of the AHS measure are: it is an index of neighborhood quality that is

⁴One problem with the STF data for 1990 is that it is based on 1990 census tracts while the AHS is based on 1980 census tracts. There have been a number of changes in census tract designations for each decennial census, particularly the splitting of census tracts in the peripheral areas. We use a data file from the Urban Institute that provides information from the 1990 STF for 1980 defined census tracts.

useful for making comparisons across neighborhoods and time, one can estimate a price for neighborhood quality by including the index in a hedonic house price regression, and it is available in all survey years. Since the cluster and tract variables are exogenous, they can be used as instruments and the house price hedonic that includes the AHS index can be consistently estimated using instrumental variables (IV).

The data set used for this analysis has information on kernel houses including the owner's valuation and characteristics of the owner and house. In addition, we have three measures of neighborhood quality: 1) cluster variables, 2) tract variables, and 3) the owner's rating. The first two are objective measures of the quality of different geographical areas that include the house, while the third is a subjective measure that covers an area that is similar to the cluster data. These three measures allow us to compare different concepts of neighborhood. In this section, we provide some descriptive statistics of these measures. In the next section, we estimate the impact that the neighborhood variables have on house prices through hedonic analysis.

For each survey date, owners are asked to estimate how much their property (house plus lot) would sell for if it were for sale. Goodman and Ittner (1992) and Kiel and Zabel (1999) have analyzed the accuracy of these owner provided values and they find that the average owner tends to over-estimate the value of their house by 5%. Kiel and Zabel find that this over-valuation is greatest for new owners and declines with length of tenure. They therefore recommend that tenure be included in the house price hedonic. Other than tenure, though, the over-valuation is not related to the observed characteristics of the owner, house, or neighborhood.⁵

⁵Since we use the owner's valuation rather than the sales price, we actually estimate house value rather than house price hedonics. But we refer to house prices rather than house values to be consistent throughout the paper and with the literature.

A house from the AHS is included in our sample only if it is a kernel house and there are at least 4 other houses in the cluster with non-missing data⁶, there is a regular occupied interview, and the unit is owner occupied. Since we are using census tract information for MSAs only, we lose houses that are in rural areas. The final numbers of observations with non-missing values for the relevant variables for the three years are 208, 238, and 318 for a total of 764 observations.

The names and definitions of the variables used in this study are given in Table 1. These include the house and owner characteristics, regional indicators, and neighborhood quality proxies. The means and standard deviations for these variables for all three years combined are also given in Table 1.

We create seven variables at the cluster level: the natural log of mean permanent income (LNPERMINC_c) the natural log of median age (LNMEDAG_c), the proportion of residents who have graduated from high school (PRHSc) and the proportion that are nonwhite (PRNONW_c), the proportion of units vacant (PRVAC_c), the proportion of units that are owner-occupied (PROWN_c), and the proportion of houses that changed hands in the past five years (PRCH5_c) and. These are characteristics of neighbors and the neighborhood that individuals are likely to care about.

Note that we use the cluster mean of permanent income instead of actual income. This is because household income will be endogenously determined with house values if households sort by income. One can view permanent income as an instrument for actual income since it is the predicted value from an income regression. First, we regress the natural log of household income

⁶We do not use other houses in the cluster (other than to create the cluster-level variables) because they clearly are not a random sample from the population.

on a third degree polynomial in age and education, and central city, male, married, black, Hispanic, and MSA dummies. We then use the estimated coefficients to form the (natural log of) permanent income for the households in the cluster data.

We also create the same seven cluster variables at the tract level (names end in “t” rather than “c”). We follow the same procedure as above when estimating permanent income at the tract level. We also include as regressors the median number of beds, the poverty and unemployment rates, and information on commuting times at the tract level.

The canonical correlation between the two groups of neighborhood characteristics is 0.89. As summarized by these variables, the neighborhoods as defined by the ten closest neighbors and the census tract are quite similar. The individual correlations between the seven cluster variables and the corresponding census tract variables are quite high (these are highlighted in bold in Table 2). They are higher than the correlations with all of the variables from the same group. For example, the correlation between $PRNONWc$ and $PRNONWt$ is 0.89. This correlation is greater than the correlation between $PRNONWc$ and any of the other six cluster variables and the same is true for $PRNONWt$. One would expect that the same characteristics would be positively correlated since the kernel house is a random draw from the census tract. The very high level of correlation for race suggests a high level of racial homogeneity at the tract level. The correlation between $PRHSc$ and $PRHSt$ is 0.59. This indicates that education is less homogeneous (relative to race) at the tract level.

Another way to look at the level of homogeneity of neighborhoods is to consider the cluster and tract means conditional on owner characteristics. We create seven owner characteristics: the natural log of permanent income ($LNPERMINC$), the natural log of the

owner's age (LNAGE), and whether the owner is nonwhite (NONWHITE), has graduated from high school (HS), was married (MARRIED), is male (MALE), and moved in the last five years (MOVEDL5). Given that the owner is nonwhite, the means of PRNONWc and PRNONWt are 0.69 and 0.65, respectively. These means when the owner is white are 0.07 and 0.15. Thus, nonwhites tend to cluster together, both at the cluster and tract levels. The canonical correlation between the cluster variables and the owner characteristics is 0.79. This indicates that the owner characteristics are quite similar to those of the other members of the cluster though, again, there is some heterogeneity within the cluster. The canonical correlation between the tract and owner variables is 0.72. This also implies a similarity of residents within the tract.

As previously mentioned, the AHS survey asks respondents to rank their neighborhood on a scale from one to ten. The frequency distribution for this index is given in Table 2. It indicates that most owners find their neighborhoods to be of high quality; 35.4% rate their neighborhood as a 10 (out of 10) while 73.0% rate it as an 8 or higher.⁷ The canonical correlation between the neighborhood quality index and the cluster and tract variables is only 0.35 and 0.29, respectively. These low correlations (relative to that between the cluster and tract variables) reinforce the subjective nature of the AHS neighborhood quality index (given that the cluster and tract characteristics are objective neighborhood measures).

⁷The full distribution of the AHS index variable cannot be presented for disclosure reasons.

5. Hedonic Regression Analysis

In this section, we present a hedonic model of house prices so that we can empirically apply the 3L Approach and compare the impact of the different measures of location on house prices. This will provide some evidence about the relative importance of the different geographical areas in terms of their impact on house prices. First, the nature of the dataset raises some empirical issues that must be addressed. Second, we estimate the benchmark model that includes proxies for all three locations: street, town, and MSA. Third, we investigate the impact on the coefficient estimates when some of these levels of location are not controlled for in the hedonic model.

Let P_{kswmt} be the price of house k , on street s in town w in MSA m at time t . Assume that the natural log of P_{kswmt} is a linear function of house characteristics (sc_{kt}), street-level neighborhood characteristics (S_{st}), and town-level public goods (T_{wt})

$$\ln P_{kswmt} = \beta_{0mt} + \beta_{1mt} sc_{kt} + \beta_{2mt} S_{st} + \beta_{3mt} T_{wt} + v_{kswmt} \quad (6)$$

$$k=1,\dots,K_t, \quad t=85,89,93, \quad m=1,\dots,NM.$$

where K_t is the number of clusters in period t and NM is the number of MSAs. The structure variables in the vector sc are the number of bedrooms, bathrooms, and total rooms, the natural logs of lot size and interior size, whether there is a garage or air conditioning and whether there are leaks, holes, cracks, or broken plaster or peeling paint over one square foot in the unit. We also include the natural log of the property tax rate and whether or not the house is in the central city of the MSA (these variables are regularly included in the AHS). As derived in Section 3, the MSA specific intercepts, β_{0mt} 's, capture the price of housing across MSAs.

5.1 Data Issues

Lot size and interior size are missing for 260 and 94 observations, respectively. Our options are to exclude these variables, only include observations with non-missing values, or to substitute the mean of the observed values for these variables and include a variable that indicates which observations have missing values. We are particularly interested in including as many significant structure characteristics as possible so as to minimize the bias in the contribution of the cluster variables (which may be correlated with the excluded structure variables). Using only observations with non-missing values for lot size and interior size will substantially reduce the sample size which is not very large to begin with. Thus, we choose the third option. If the missing observations are random, the missing indicator variables will not be significant.⁸

S and T are vectors of neighborhood characteristics that include measures of accessibility, land use, local public services, and other neighborhood and environmental quality variables. Obtaining data on these variables at a disaggregated enough level to be useful, particularly for the whole United States, is very difficult. We use area averages of owner and house characteristics which are considered to be proxies for these variables. In this study, data for S are based on the 10 nearest neighbors and for T are based on census tract data. We will also look at the AHS neighborhood index as a proxy for neighborhood quality.

⁸We also do the same thing for property taxes and leaks, though, there are significantly fewer missing values for these two variables (a little more than a dozen and a couple, respectively - actual values cannot be reported for disclosure reasons).

The size and scope of the data set put restrictions on our ability to accurately estimate the house price hedonic equation (6). In particular, it is generally believed that there is at least one separate housing market for each MSA. This requires that the parameters be allowed to vary across MSAs. The problem is that there are not enough observations per MSA for this to be feasible. One solution is to pool observations across MSAs. While this a common practice when using the NAHS (e.g. Goodman and Ittner (1992) and DiPasquale and Somerville (1995)), a thorough justification is lacking.⁹

We provide evidence for pooling by carrying out a test of whether the structure and neighborhood characteristics vary across MSAs. To do this, we access the non-cluster data of the NAHS 85, 89, and 93. This constitutes 90% of these data sets. We pool the data across years and then estimate separate regressions for each MSA with at least 100 observations (we do allow the intercept to vary across time). This is possible for 64 MSAs with a total number of 22,768 observations. We have two justifications for pooling across years. First, we need enough observations to be able to estimate regressions for each MSA. Second, DiPasquale and Somerville (1995), Kiel and Zabel (1999) and Ioannides and Zabel (2003a) provide evidence that the coefficients for the structure and neighborhood variables do not vary across time (since the dependent variable is the log of price the coefficients are measured in percent terms).

We need to estimate two models in order to test whether the coefficients vary across MSAs. The restricted model assumes constant structure and neighborhood coefficients but allows the intercept to vary across MSAs. The unrestricted model involves estimating a separate regression for each MSA. Note that we cannot include the cluster variables in these regressions

⁹This is also true in quality of life literature, e.g. Gyourko and Tracey (1991).

since we are using the non-cluster data but we do include the census tract variables. Given the large number of observations, it should not be too surprising that we reject the null hypothesis of constant parameters for the structure and neighborhood variables either individually or jointly. We also use the Ohta-Griliches procedure which is a less restrictive version of the F-test that is not affected by sample size that has been used in the literature (e.g., Palmquist 1982, Kiel and Zabel 1999). This test favors the pooled model (constant coefficients) if the percent increase in the standard error between the unrestricted and restricted models is less than 10%. The percent increase in the standard error between the varying coefficient and pooled models is 8.03%. Hence, this evidence supports the constant structure and neighborhood coefficients. Further evidence is in Mills and Simenauer (1996), who find that the coefficients on the structure variables do not vary significantly across regions.

Based on these results, we pool the data across MSAs, leaving only the intercept to vary across time and MSAs. Still, there are not enough observations in the cluster data to estimate the β_{0mt} 's. But the ability to pool across MSAs means that we can use the non-cluster data to get these estimates. Since we do not have to estimate a separate equation for each MSA, we relax the inclusion restriction of at least one hundred observations to at least ten per year. This results in a larger dataset that includes all 102 MSAs in the cluster dataset. Based on the analysis in Section 3, the intercepts provide the price index that allows us to estimate the (relative) price of living in a given MSA. We leave out Denver in 1985 and set this price to 100. All other prices are relative to this value. The mean of the price variable is 88 and the standard deviation is 31.

We use this index to estimate the β_{0mt} 's in equation (6) using the cluster data.

$$\ln P_{kswmt} = \beta_{01} + \beta_{02} \ln \hat{p}_{mt} + \beta_1 sc_{kt} + \beta_2 S_{st} + \beta_3 T_{wt} + v_k + \varepsilon_{kswmt} \quad (7a)$$

where $\ln \hat{p}_{mt}$ is the natural log of the price index. We denote (7a) as the Index Model. Note that because the same house can appear in more than one year, we have included the individual house effect, v_k . We will estimate this model using random effects to account for the correlation across observations for the same house. We could impose the restriction that $\beta_{02} = 1$ but we allow this coefficient to vary so that we can test that the 3 levels of location are significant. When we estimate this model we will test the hypothesis that $\beta_{02} = 1$.

One problem with this approach is that this does not allow us to see how the estimates of the individual β_{0mt} 's are affected by the exclusion of the cluster variables. To do so, we restrict the β_{0mt} 's to vary only across the four census regions. This results in enough observations for each census region to estimate the intercepts directly using the cluster dataset. This model is specified as

$$\ln P_{kwsrt} = \sum_{t=85,89,93} \sum_{r=1}^4 \beta_{rt} + \beta_1 sc_{kt} + \beta_2 S_{st} + \beta_3 T_{wt} + v_k + \varepsilon_{kwsrt} \quad (7b)$$

where r now indexes the census region in which the house is located. We denote (7b) as the Region Model. The estimates for equation 7b should be viewed only in that they serve to give some idea of how the exclusion of one or more sets of location variables might impact the

estimation of house price indices. We can also compare the estimates of the structure and neighborhood coefficients from the Index and Region Models to see if the use of the latter model biases these estimates.

5.2 Benchmark Model: Cluster Variables Included

We consider the model that includes both the cluster and census tract variables as the benchmark since, ideally, one would like to be able to include the full set of neighborhood variables in the model. The results for the Index and Region Models estimated using random effects (RE) with robust standard errors are included in columns 2 and 4 of Table 3.

One issue of concern in hedonic models is missing variables bias. In particular, if missing structure characteristics are correlated with the cluster variables, the significance of these variables will be misstated. We do include a large number of structure characteristics in our model but we miss design factors that are unique to a given house that significantly impact the price. Zabel (1999) recommends the use of the fixed effects estimator since it excludes time-invariant unobservable factors, such as structure characteristics, that are correlated with the regressors.¹⁰ We test for the significance of the correlated effects using the Hausman test. For both models, we reject the null hypothesis that the individual unobserved house effect, ν_k , is uncorrelated with the regressors in the house price hedonic (p-values < 0.001). This justifies the use of the fixed effects estimator.

¹⁰ Another way of alleviating this type of bias is to use the repeat sales model. This will difference out missing structure characteristics that are present at both sales.

Houses must appear at least twice in the dataset to be included in the fixed effects model. There are 54 houses that appear twice and 159 houses that appear three times in our data set. 179 houses must be excluded since they only appear once. The results for the Index Model using fixed effects (FE) with robust standard errors are given in column 5 of Table 3. There are a number of significant differences between the FE and RE estimates. First, while a number of the structure characteristics are significant when RE is used, none are individually significant when FE is used. This is expected since for FE the identification of the structure parameters is based on across time and within unit variation in the structure variables. Thus, only houses with changes in the structure variables can contribute to this variation. Second, the FE estimate of LNPERMINC_t is much larger than that for RE while the opposite is true for LNPERMINC_c . This is because the FE estimator controls for unobserved time-invariant factors that affect house prices. These are likely to include structure and cluster-level characteristics. These variables will be correlated with cluster income and hence will bias the estimate of LNPERMINC_c when RE is used. Third, the estimated coefficient for LNPRICE is 0.904 and it is not significantly different from 1 at the 10% level. The corresponding estimate is 0.742 when RE is used and it is significantly less than 1. Recall, that we expect this coefficient to be 1, so this is further evidence in support of the FE model.

Since the cluster and tract variables are proxies for street-level and town-wide characteristics, we care most about the impact of each group as a whole. For the Index Model, the p-values for the F-statistics used to test the joint significance of the cluster and tract variables are 0.108 and 0.003, respectively. Thus, the tract variables appear to have a more significant role in determining prices than do the cluster variables. It is important to note, though, that when the

Index Model is estimated using RE, both the cluster and tract variables are jointly significant with p-values of less than 0.001. Thus, when not controlling for time-invariant unobservable factors, the cluster variables have a much more significant impact on prices. In essence, the fixed effects are controlling for the time-invariant factors at the cluster level. Hence the RE estimator provides a better framework for testing for the joint significance of the cluster variables.

Given that LNPRICE is highly significant, it appears that each geographical measure includes significant independent information about house prices (the three sets of location measures are also jointly significant, $p\text{-value} < 0.001$); there is significant heterogeneity within the census tract (despite the fact that the canonical correlation is 0.89). That is, holding the characteristics of the census tract (and the house) constant, there is still significant street to street variation in neighborhood quality. By holding the census tract characteristics constant, we are essentially holding school quality and crime constant (assuming that the tract variables are good proxies for these two public goods). Thus, we are picking up other unmeasured factors, like accessibility to green spaces and other amenities, and also the general upkeep of the street and the neighbors' characteristics.

These results support the 3L approach since the cluster, tract, and MSA variables are all highly significant in the house price hedonic. Indeed, "location, location, and location" does matter in determining house prices. To get an idea of the relative importance of the different components of location in determining house prices, we calculate the percent increase in the standard error when each set of cluster, tract, and MSA variables are excluded from the regressions. We make this calculation for the structure variables as well to measure the relative importance of location versus structure in determining house prices. For the tract variables, we

consider the decrease when only the variables that come from the decennial censuses are excluded and when the town-level variables that are taken from the AHS are also excluded. The latter variables include the central city indicator and the natural log of the property tax rate. For the MSA variables, we calculate the decrease based on two models. Both exclude the MSA variables but one excludes and one includes time dummies. The latter model accounts for the average increase in house prices across periods that is not MSA specific and can be considered to be the appropriate alternative when the MSA does not affect house prices.

This exercise provides for a similar decomposition of house prices as Zabel (1999) does for house price returns: national, market, sub-market (neighborhood), and structure characteristics. Here, the time dummies capture the national (inflation) effect, the MSA variables capture the market effect and we divide the sub-market effect into tract and cluster impacts.

The results are given in Table 4. We present results for both RE and FE since RE gives some idea of how the fixed effects are apportioned to the different structure and location components of house price. When the Index Model is estimated using RE, the 3L's all have about the same importance in determining house prices. That is, the percent increase in the standard error is similar when the cluster variables, the same set of tract variables, and the price index (not including the time dummies) are excluded from the model: 2.2%, 2.3%, and 2.7%, respectively. When the Index Model is estimated using FE, as expected, the importance of the cluster variables is diminished as compared to the other location factors. Also note that the importance of the structure characteristics is greater than any one single location factor when either FE or RE is used to estimate the Index Model.

Finally, the use of the Region Model produces coefficient estimates for the structure characteristics that are similar to those from the Index model estimated using FE (results given in Table 3). In contrast, the coefficient estimates for the cluster and tract characteristics are quite different from those for the Index model. Thus, not allowing for a separate price per MSA does bias appear to bias the estimates of the other location variables. For the Region Model, the estimated appreciation rates vary considerably across the four regions (Northeast 85 is the excluded variable).¹¹ The coefficient for NE8589 is the appreciation rate between 1985 and 1989 in the Northeast (NE) which is estimated to be 30.0%. For all four regions, prices showed substantial increases between 1985 and 1989 and were essentially flat between 1989 and 1993.

5.3 Practical Model: Cluster Variables Excluded

The model that is most likely to be estimated by most researchers will not include the cluster variables since there are unique to the 1985, 1989, and 1993 waves of the NAHS. This it is worthwhile seeing how this model compares with the baseline model discussed in the last subsection. The results from estimating the Index and Region Models using random and fixed effects and excluding the cluster variables are given in Table 5. For the most part, the coefficient estimates for the structure variables (when using either FE or RE) are unaffected by the exclusion of the cluster variables. When using FE, the coefficient estimates for the census tract variables are also not significantly affected. At first glance, this is somewhat surprising given the high correlation between the cluster and tract variables. But given that these results are for the FE

¹¹Actually, these are not appreciation rates but the estimated coefficients on the region/time dummies. To obtain appreciation rates, one must make the transformation $\exp(\beta)-1$ where β is the region/time coefficient.

estimator, it is likely that the cluster variables are picked up by the fixed effects. We do get the expected results when using the RE estimator. In this case, the coefficient estimates for the tract variables (LNPERMINCt in particular) and the price index are significantly different than those when the cluster variables are included. We do see a significant impact on the appreciation rates in the Region Model when the cluster variables are excluded. There is particularly evident when the RE estimator is used.

To summarize, when using fixed effects, the coefficients for the structure and tract variables and the price index are not significantly affected when the cluster variables are excluded, but there is a significant impact on the regional appreciation rates. On the other hand, only the coefficient estimates for the structure variables are unaffected when the RE estimator is used. The conclusion to be drawn from this exercise is that the FE estimator is preferred when the cluster variables are unavailable and, for the most part, the estimates are still reliable.

5.4 AHS Neighborhood Quality Index

In this section, we consider models that include neighborhood proxies that are based only on variables that are available in the public version of the usual AHS survey that does not have the cluster sample. This means that we exclude the cluster and tract variables but include the neighborhood quality index. As previously mentioned, the nice features of this latter variable are: (1) it provides a single measure of neighborhood quality, (2) its coefficient indicates how much house prices will change when quality is increased, and (3) it is available in all AHS surveys. The disadvantage is that this index is a subjective measure of neighborhood quality. We estimate the Index Model using both random and fixed effects. While the latter estimator is

preferable from a consistency standpoint, there is not likely to be much variation in the neighborhood index for a given house and hence the use of the fixed effects estimator might result in an insignificant impact on house prices.

The results are given in Table 6. Using FE, the estimated coefficient for the AHS index is 0.003 with a standard error of 0.012. We test for the exogeneity of this variable using the Hausman test and we do not reject the null hypothesis that this variable is exogenous. The identifying instruments are the tract variables. For FE, the IV estimates are listed in column five of Table 6. The estimated coefficient for the AHS index is -0.103 with a standard error of 0.060. Thus, in neither case is the AHS index variable significant. The estimates of the structure variables are significantly affected by the exclusion of the tract and cluster variables. In particular, the coefficient estimates for CENCITY, HAGE, and HAGE² are now significant. Using RE, the estimated coefficient for the AHS index is 0.130 and it is significant at the 5% level. In this case, the Hausman test statistic is significant at the 1% level and hence we reject the null hypothesis that the AHS index is exogenous. For RE, the IV estimates are listed in column four of Table 6. The estimated coefficient for the AHS index is 0.016 with a standard error of 0.009. Thus once we control for the endogeneity of the index, its impact on house values is greatly diminished. As is the case for the FE estimates, the coefficients for a number of the structure variables are significantly affected when the AHS index variable replaces the tract and cluster variables. The conclusion is that the AHS index is endogenous and even when one corrects for this endogeneity, the coefficient estimates of the structure variables are suspect.

6. Conclusion

The immobility of houses means that their location affects their values. This explains the common belief that three things determine the price of a house: location, location, and location. We use this notion to develop the 3L Approach to house price determination where prices are affected by the MSA, town, and street where the house is located. This study creates a unique data set consisting of small ‘clusters’ of housing units with information on their characteristics, resident characteristics, and census tract-level attributes. We use these data to verify the 3L Approach. That is, we find that all three levels of location are jointly significant when estimating the house price hedonic equation. This indicates that individuals care about their local neighborhood, i.e. the general upkeep of their street and possibly their neighbors’ characteristics (cluster variables) a broader area such as the school district and/or the town (tract variables) that account for school quality and crime rates, and the particular amenities found in their MSA.

Our preferred model includes an MSA-wide price index that is estimated using fixed effects. We find that not controlling for time-invariant unobserved characteristics (random effects) results in biased estimates of variables at all three location levels. Including only regional indicators (instead of the MSA-wide price index) also results in biased estimates of the location variables. Finally, leaving out the street-level indicators (as proxied by the cluster variables) does not appear to significantly affect the coefficient estimates for the structure and town-wide characteristics though there is some evidence that this might bias the MSA-wide price index. Finally, we provide some evidence that supports pooling across MSAs when estimating house price hedonics (but allowing for separate intercepts across MSAs). This is useful when data considerations do not allow one to estimate a separate model for each MSA.

References

- Clapp, J.M. and Y. Wang. 2003. "Defining Neighborhood Boundaries: The Use of Transactions Data," unpublished mimeo, University of Connecticut.
- DiPasquale, Denise and C. Tsurriel Somerville. 1995. "Do House Price Indices Based on Transacting Units Represent the Entire Stock? Evidence from the American Housing Survey," Journal of Housing Economics 4: 195-229.
- Gabriel, Stuart A. and Stuart S. Rosenthal. 1996a. "Commute Times, Neighborhood Effects and Earnings: An Analysis of Compensating Differentials and Racial Discrimination." Journal of Urban Economics 40:61-82.
- , 1996b. "Location and the Effect of Demographic Traits on Earnings: The Tradeoff Between Omitted Variable and Simultaneity Bias." Working Paper, University of British Columbia.
- Galster, George C. 1986. "What is Neighborhood: An Externality-Space Approach." International Journal of Urban and Regional Research 10: 243-263.
- Goodman, Alan C. 1977. "A Comparison of Block Group and Census Tract Data in a Hedonic Housing Price Mode," Land Economics 53: 483-487.
- Goodman J.L. and J.B. Ittner. 1992. "The Accuracy of Home Owners' Estimates of House Value." Journal of Housing Economics 2:339-357.
- Gyourko, Joseph and Joseph Tracey. 1991. "The Structure of Local Public Finance and the Quality of Life," Journal of Political Economy 99: 774-806.
- Hardman, Anna and Yannis M. Ioannides. 1997. "Income Mixing and Housing in U.S. Cities: Evidence from Neighborhood Clusters of the American Housing Survey." Working Paper, Department of Economics, Tufts University.
- Hoyt, William H. and Stuart S. Rosenthal. 1997. "Household Location and Tiebout: Do Families Sort According to Preferences for Locational Amenities," Journal of Urban Economics 42: 159-178.
- Ioannides, Yannis M. 2000. "Residential Neighborhood Effects," Tufts University Working Paper 2000-12.
- Ioannides, Yannis M. and Jeffrey E. Zabel. 2003a. "Neighborhood Effects and Housing Demand," Journal of Applied Econometrics 18: 563-584.

- Ioannides, Yannis M. and Jeffrey E. Zabel. 2003b. "Interactions, Neighborhood selection, and Housing Demand," unpublished mimeo, Tufts University.
- Keller, Suzanne. 1982. "The Neighborhood." in Ronald H. Baylor Ed. Neighborhoods in Urban America. Port Washington N.Y.: Kennikat Press.
- Kiel, Katherine A. and Jeffrey E. Zabel. 1996. "House Price Differentials in U.S. Cities: Household and Neighborhood Racial Effects". Journal of Housing Economics 5: 143-165.
- , 1999. "The Accuracy of Owner Provided House Values: The 1978-1991 American Housing Survey," Real Estate Economics 27: 263-298.
- Li, Mingche M. and James Brown. 1980. "Micro-Neighborhood Externalities and Hedonic Housing Prices." Land Economics 56: 125-141.
- Roback, Jennifer. 1982. "Wages, Rents, and the Quality of Life," Journal of Political Economy 90: 1257-1278.
- Rosen, Sherwin. 1974. "Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition." Journal of Political Economy 82:34-55.
- Schelling, T.C. Models of Segregation. Santa Monica: The Rand Corporation, RM-6014-RC.
- Segal, David. 1979. The Economics of Neighborhood. New York: Academic Press.
- Shultz, Steven D., and David A. King. 2002. "The Use of Census Data for Hedonic Price Estimates of Open Space Amenities and Land Use." Mimeo, North Dakota State University.
- Strange, William. 1992. "Overlapping Neighborhoods and Housing Externalities," Journal of Urban Economics 32:17-39.
- Tiebout, Charles M. 1956. "A Pure Theory of Local Public Expenditures." Journal of Political Economy 64: 416-424.
- Zabel, Jeffrey E., 1999. "Controlling for Quality in House Price Indices," Journal of Real Estate Finance and Economics 19:223-241.

Table 1
Variable Names , Definitions and Summary Statistics

Name	Definition	Mean (std)
VALUE	owner-estimated value of the house	112,679 (80,002)
PRICE	MSA-wide house price index	88.033 (31.000)
Structure Characteristics		
HAGE	the age of the house in years	38.382 (19.998)
GARAGE	=1 if the house has a garage, =0 otherwise	0.803 (0.397)
BEDROOMS	number of bedrooms in the house	2.982 (0.903)
FULLBATHS	number of full bathrooms in the house	1.637 (0.774)
ROOMS	total number of rooms	6.353 (1.665)
AIRCOND	=1 if the house has either central or room air conditioning, =0 otherwise	0.754 (0.431)
TENURE	years that owner has lived in house	15.822 (12.649)
LOTSIZE	lot size in square feet	17,742.9 (35,751.5)
UNITSF	interior size in square feet	1,848.84 (721.22)
HOLES	= 1 if holes in floor, = 0 otherwise	***
LEAK	= 1 if water leaked into home from outside in last twelve months	***
CRACKS	= 1 if open cracks or holes in walls or ceiling, 0 otherwise	***
BIGP	= 1 if broken plaster or peeling paint over one square foot	***
Owner Characteristics		
PERMINC	permanent household income	33,233.5 (22,969.9)
AGE	owner's age	53.622 (16.731)
HS	=1 if owner graduated from high school, =0 otherwise	0.835 (0.371)
NONWHITE	=1 if owner is nonwhite, =0 otherwise	0.160 (0.367)
MARRIED	=1 if owner is married, =0 otherwise	0.613 (0.487)
MALE	=1 if owner is male, =0 otherwise	0.674 (0.469)

Table 1 - Continued		
MOVEDL5	=1 if owner moved in last five years, =0 otherwise	0.228 (0.420)
Cluster and Tract Characteristics		
CENCITY	=1 if house in central city of SMSA, =0 otherwise	0.484 (0.500)
TAX	property tax per \$1,000	1.523 (2.365)
INDEX	Owner's rating of neighborhood quality on a scale of one (worst) to ten (best)	8.245 (1.872)
LNPERMINC(c,t)	(cluster, tract) mean of the natural log of permanent household income	10.190 (0.528) c 10.535 (0.290) t
LNMEDAG(c,t)	natural log of median age of individuals in the house's (cluster, census tract)	3.926 (0.142) c 3.433 (0.163) t
PRNONW(c,t)	proportion of nonwhite individuals in the house's (cluster, census tract)	0.167(0.303) c 0.225 (0.271) t
PRHS(c,t)	proportion of individuals over 25 who have completed high school in the house's (cluster, census tract)	0.829 (0.191) c 0.720 (0.119) t
PRCH5(c,t)	proportion of houses in the (cluster, census tract) that have changed hands in the last five years	0.359 (0.201) c 0.454 (0.131) t
PRVAC(c,t)	percent of housing units vacant in the house's census tract	0.055 (0.089) c 0.060 (0.050) t
MEDBEDt	median number of bedrooms in the unit in the census tract	2.666 (0.586)
POVRATEt	poverty rate in the census tract	0.100 (0.091)
UNEMPLOYt	unemployment rate in the census tract	0.059 (0.039)
COMMUT12t	percent in census tract whose commute was less than 20 minutes	0.613 (0.153)
COMMUT24t	percent in census tract whose commute was between 20 and 40 minutes	0.258 (0.099)

**Table 2 - Correlations Between Cluster, Tract, Owner,
and AHS Neighborhood Quality Variables**

Cluster Variables							
	LNPERMINC _c	LNMEDAG _c	PRNONW _c	PRHSc	PRVAC _c	PRCH5 _c	PROWNC
Cluster Variables							
LNMEDAG _c	-0.138						
PRNONW _c	-0.220	0.009					
RRTHSc	0.496	-0.126	-0.239				
PRVAC _c	-0.175	0.039	0.089	-0.108			
PRCH5 _c	-0.066	-0.467	-0.068	0.032	0.142		
PROWN _c	0.223	0.309	-0.123	0.240	-0.133	-0.438	
Tract Variables							
LNPERMINC _t	0.607	0.002	-0.346	0.467	-0.230	-0.070	0.315
LNMEDAG _t	0.112	0.338	-0.233	0.256	0.116	-0.051	0.180
PRNW _t	-0.210	-0.016	0.885	-0.302	0.121	-0.030	-0.163
PRHSt	0.413	0.007	-0.392	0.591	-0.203	-0.016	0.276
PRVAC _t	-0.332	0.196	0.117	-0.176	0.419	0.088	-0.139
PRCH5 _t	-0.031	-0.185	-0.157	0.060	0.030	0.265	-0.206
PROWN _t	0.112	0.088	-0.182	0.127	-0.097	-0.159	0.392
Owner Characteristics							
LNPERMINC	0.648	-0.072	-0.162	0.338	-0.143	0.002	0.134
LNAGE	-0.116	0.228	0.012	-0.078	0.072	-0.124	-0.038
NONWHITE	-0.160	0.012	0.749	-0.163	0.089	-0.103	-0.042
HS	0.221	-0.002	-0.102	0.331	-0.084	0.025	0.198
MOVEDL5	0.032	-0.079	-0.046	0.069	0.013	0.155	-0.036
AHS Index	0.139	0.158	-0.144	0.209	-0.135	-0.206	0.244
Tract Variables							
	Ininhtmt	Inmedagt	pctnwt	pcthst	pctvact	pctch5t	pctownt
Tract Variables							
LNMEDAGT	0.353						
PCTNWT	-0.429	-0.351					
PCTHST	0.653	0.363	-0.502				
PCTVACT	-0.389	0.072	0.162	-0.209			
PCTCH5T	-0.144	-0.328	-0.075	0.088	0.278		
PCTOWNT	0.474	0.319	-0.282	0.287	-0.242	-0.556	
Owner Characteristics							
LNINCHAT	0.451	0.120	-0.171	0.315	-0.258	-0.050	0.064
LNAGE	-0.049	0.078	-0.001	-0.037	0.121	-0.031	-0.003
NONWHITE	-0.227	-0.181	0.677	-0.280	0.107	-0.123	-0.127
HS	0.264	0.166	-0.148	0.276	-0.112	-0.031	0.072
MOVEDL5	0.002	-0.008	-0.013	0.056	0.062	0.154	-0.057
AHS Index	0.214	0.187	-0.190	0.247	-0.062	-0.088	0.140

Table 3 - Random and Fixed Effects Results				
Dependent Variable: Ln of Owner's Valuation				
	Region Model		Index Model	
	RE	FE	RE	FE
LNTAX	-0.242**	-0.291**	-0.247**	-0.285**
	(0.036)	(0.059)	(0.031)	(0.055)
CENCITY	0.096*	0.081	0.029	-0.103
	(0.047)	(0.060)	(0.043)	(0.053)
HAGE	0.001	-0.006	0.005	0.004
	(0.004)	(0.013)	(0.004)	(0.011)
HAGE ²	-0.003	0.001	-0.006	-0.006
	(0.004)	(0.011)	(0.004)	(0.010)
GARAGE	0.044	0.041	0.073	0.044
	(0.050)	(0.104)	(0.047)	(0.095)
BEDROOMS	0.054*	0.096*	0.051*	0.092
	(0.025)	(0.049)	(0.025)	(0.049)
FULLBATHS	0.085**	0.081	0.112**	0.084
	(0.024)	(0.044)	(0.023)	(0.044)
ROOMS	0.018	-0.013	0.024	-0.018
	(0.014)	(0.020)	(0.014)	(0.020)
AIRCOND	-0.119**	-0.082	-0.078	-0.093
	(0.043)	(0.092)	(0.042)	(0.084)
TENURE	-0.013**	-0.014*	-0.012**	-0.014*
	(0.004)	(0.007)	(0.004)	(0.006)
TENURE ²	0.030**	0.028*	0.026**	0.029*
	(0.009)	(0.013)	(0.009)	(0.014)
LNLOTSIZE	0.022	-0.043	0.038	-0.06
	(0.022)	(0.042)	(0.021)	(0.041)
LNUNITSF	0.178**	0.293	0.182**	0.287
	(0.055)	(0.155)	(0.052)	(0.150)
LNPERMINCt	0.06	0.265	0.237**	0.742**
	(0.099)	(0.357)	(0.090)	(0.264)
LNMEDAGt	0.236	-0.113	0.129	-0.18
	(0.157)	(0.476)	(0.150)	(0.484)
PRNWt	-0.249	-0.648	-0.300*	-0.54
	(0.155)	(0.587)	(0.142)	(0.618)
PRHSt	0.521*	-0.877	0.419	-1.113
	(0.228)	(0.786)	(0.221)	(0.742)
PRVACt	-0.384	0.926	0.212	2.161
	(0.544)	(1.582)	(0.512)	(1.317)
PRCH5t	-0.336	-0.246	-0.419	-0.371
	(0.286)	(0.614)	(0.264)	(0.610)

Table 3- Continued				
	Region Model		Index Model	
	RE	FE	RE	FE
LNMEDAGc	0.147	-0.145	-0.032	-0.097
	(0.147)	(0.293)	(0.144)	(0.288)
PRNWc	0.126	0.532	0.019	0.614
	(0.124)	(0.322)	(0.122)	(0.328)
PRHSc	-0.267*	-0.015	-0.003	0.09
	(0.121)	(0.225)	(0.121)	(0.230)
PRVACc	-0.003	0.157	0.103	0.175
	(0.192)	(0.293)	(0.195)	(0.288)
PRCH5c	0.035	-0.122	0.073	-0.05
	(0.094)	(0.123)	(0.093)	(0.114)
PROWNc	-0.032	0.301	0.125	0.245
	(0.128)	(0.266)	(0.125)	(0.256)
NE8589	0.217**	0.262**		
	(0.060)	(0.090)		
NE8993	-0.139*	-0.064		
	(0.063)	(0.104)		
SNE85	-0.075	-0.344**		
	(0.083)	(0.086)		
S8589	0.228**	0.180*		
	(0.051)	(0.088)		
S8993	-0.102	0.049		
	(0.055)	(0.082)		
MWNE85	-0.154	-0.382**		
	(0.079)	(0.080)		
MW8589	0.213**	0.205*		
	(0.055)	(0.102)		
MW8993	-0.081	0.049		
	(0.052)	(0.087)		
WNE85	-0.177	-0.17		
	(0.114)	(0.090)		
W8589	0.382**	0.365**		
	(0.067)	(0.087)		
W8993	-0.154*	0.025		
	(0.062)	(0.105)		
LNPRICE			0.742**	0.904**
			(0.082)	(0.163)
Observations	764	585	764	585
Number of Houses	392	213	392	213
R-Squared Within	0.469		0.474	
R-Squared Between	0.699		0.702	
R-Squared Overall	0.707		0.715	
Standard errors in parentheses				
* significant at 5%; ** significant at 1%				

Table 4 - Relative Impact of 3L's				
Percent Change in RMSE				
	Random Effects Models		Fixed Effects Models	
Variables Excluded	Regional Dummies	Price Index	Regional Dummies	Price Index
Structure	4.9	6.8	2.6	3.1
Cluster	13.1	2.2	1.7	0.7
Tract - Variables Same as Cluster	1.4	2.3	0.3	2.1
Tract - Includes AHS Variables	11.2	13.7	16.0	17.0
Price Index	3.3	3.3	3.1	4.9
Price Index - Time Effect Excluded	0	2.7	0.2	2.0

**Table 5 - Random and Fixed Effects Results:
Cluster Variables Excluded
Dependent Variable: Ln of Owner's Valuation**

Variable	Region Model		Index Model	
	RE	FE	RE	FE
LNTAX	-0.242** (0.038)	-0.295** (0.057)	-0.240** (0.031)	-0.280** (0.054)
CENCITY	-0.037 (0.055)	0.034 (0.069)	-0.017 (0.042)	-0.096 (0.053)
HAGE	0.002 (0.005)	-0.005 (0.012)	0.005 (0.004)	0.005 (0.011)
HAGE ²	-0.002 (0.005)	0.001 (0.010)	-0.006 (0.004)	-0.005 (0.010)
GARAGE	0.07 (0.062)	0.034 (0.111)	0.094 (0.049)	0.034 (0.099)
BEDROOMS	0.056* (0.029)	0.083 (0.049)	0.053* (0.025)	0.082 (0.048)
FULLBATH	0.105** (0.026)	0.081 (0.044)	0.128** (0.022)	0.091* (0.045)
ROOMS	0.006 (0.015)	-0.015 (0.020)	0.024 (0.014)	-0.021 (0.020)
AIRCOND	-0.104* (0.048)	-0.088 (0.091)	-0.071 (0.043)	-0.098 (0.084)
TENURE	-0.014** (0.005)	-0.014* (0.006)	-0.012** (0.004)	-0.014* (0.006)
TENURE ²	0.029** (0.011)	0.026 (0.013)	0.027** (0.009)	0.030* (0.014)
LNLOTSIZE	0.014 (0.026)	-0.04 (0.044)	0.038 (0.022)	-0.059 (0.042)
LNUNITSF	0.265** (0.065)	0.232 (0.150)	0.213** (0.052)	0.241 (0.146)
LNPERMINCt	0.814** (0.080)	0.25 (0.364)	0.399** (0.074)	0.800** (0.252)
LNMEDAGt	0.299 (0.171)	-0.046 (0.454)	0.099 (0.146)	-0.152 (0.465)
PRNWt	0.009 (0.096)	-0.32 (0.619)	-0.310** (0.078)	-0.195 (0.640)
PRHSt	0.386 (0.245)	-0.817 (0.808)	0.538* (0.215)	-1.141 (0.772)
PRVACt	-0.636 (0.611)	0.672 (1.558)	0.152 (0.523)	2.025 (1.199)
PRCH5t	-0.359 (0.348)	-0.19 (0.624)	-0.381 (0.268)	-0.378 (0.603)
PROWNt	-0.735** (0.247)	1.906* (0.839)	-0.476* (0.202)	0.406 (0.813)

Table 5 - Continued				
	Region Model		Index Model	
	RE	FE	RE	FE
NE8589	0.149*	0.264**		
	(0.065)	(0.094)		
NE8993	-0.095	-0.012		
	(0.067)	(0.099)		
SNE85	-0.332**	-0.528**		
	(0.092)	(0.099)		
S8589	0.061	0.149		
	(0.051)	(0.077)		
S8993	0.042	0.146*		
	(0.055)	(0.072)		
NWNE85	-0.435**	-0.595**		
	(0.088)	(0.093)		
MW8589	0.042	0.169		
	(0.052)	(0.102)		
MW8993	0.057	0.175*		
	(0.054)	(0.082)		
WNE85	-0.172	-0.233*		
	(0.117)	(0.103)		
W8589	0.258**	0.362**		
	(0.066)	(0.091)		
W8993	0.02	0.134		
	(0.066)	(0.093)		
LNPRICE			1.034**	0.960**
			(0.063)	(0.156)
Observations	764	585	764	585
Number of Houses	392	213	392	213
R-Squared Within	0.457		0.467	
R-Squared Between	0.564		0.681	
R-Squared Overall	0.539		0.687	
Robust standard errors in parentheses				
* significant at 5%; ** significant at 1%				

Table 6 - Random and Fixed Effects Results: AHS Index Variable Included Dependent Variable: Ln of Owner's Valuation				
	Actual Index		Predicted Index	
Variable	RE	FE	RE	FE
LNTAX	-0.233**	-0.274**	-0.236**	-0.273**
	(0.032)	(0.055)	(0.032)	(0.055)
CENCITY	-0.044	-0.398**	-0.073	-0.450**
	(0.052)	(0.058)	(0.044)	(0.064)
HAGE	0.008	0.023**	0.011**	0.024**
	(0.005)	(0.009)	(0.004)	(0.008)
HAGE ²	-0.008	-0.014	-0.012**	-0.018*
	(0.005)	(0.010)	(0.004)	(0.009)
GARAGE	0.159**	0.046	0.156**	0.039
	(0.056)	(0.098)	(0.053)	(0.098)
BEDROOMS	0.051	0.09	0.05	0.074
	(0.031)	(0.051)	(0.026)	(0.048)
FULLBATH	0.113**	0.086*	0.125**	0.056
	(0.029)	(0.043)	(0.026)	(0.047)
ROOMS	0.015	-0.018	0.016	-0.009
	(0.015)	(0.019)	(0.014)	(0.021)
AIRCOND	-0.006	-0.091	-0.016	-0.097
	(0.050)	(0.087)	(0.048)	(0.086)
TENURE	-0.016**	-0.016**	-0.015**	-0.015**
	(0.005)	(0.006)	(0.004)	(0.005)
TENURE ²	0.033**	0.034**	0.032**	0.031**
	(0.011)	(0.012)	(0.009)	(0.012)
LNLOTSIZE	0.072*	-0.064	0.100**	-0.051
	(0.028)	(0.045)	(0.021)	(0.046)
LNUNITSF	0.408**	0.168	0.506**	0.185
	(0.068)	(0.140)	(0.043)	(0.148)
INDEX	0.130*	0.003	0.016	-0.103
	(0.056)	(0.012)	(0.009)	(0.060)
LNPRICE	1.371**	1.165**	1.348**	1.063**
	(0.063)	(0.145)	(0.058)	(0.143)
Observations	764	585	764	585
Number of Houses	392	213	392	213
Adjusted R-Squared	0.997	0.997	0.925	0.926
Robust standard errors in parentheses				
* significant at 5%; ** significant at 1%				