Measuring the External Benefits of Homeownership

Edward Coulson
Herman Li
Department of Economics
Penn State University
University Park PA, 16802

This draft; September 2010

Abstract: The subsidization of homeownership through e.g. the mortgage interest deduction is justified on efficiency grounds only to the extent that it provides benefits to people other than the homeowner. We use the clustered neighborhoods in the American Housing Survey to measure that benefit in the form of higher housing prices in neighborhoods with higher ownership rates (and lower vacancies). We attempt to account for unobservable neighborhood and house attributes that may be correlated with occupancy and ownership through instrumental variables, switching regressions and panel methods. Estimates indicate that a housing transition from renting to owning creates approximately $1000 in measured benefits, which is more than the deadweight loss arising from the mortgage interest deduction.

Acknowledgments: We thank Sung-Jae Ju, Joris Pinkse, Mark Roberts, and Brent Ambrose as well as listeners at the Federal Reserve Bank of St. Louis, University of Southern California, Penn State, Conference of Urban and Regional Economics at FEEM (Milan), Regional Science Association Annual Meeting, AREUEA Midyear Meeting, and Israel Symposium on Real Estate and Urban Economics. Errors are ours.

1. Introduction

Homeownership is heavily subsidized by the federal government through various tax expenditures and other programs that directly or indirectly provide substantial encouragement for households to become homeowners. Prominent among these is the exemption from tax of the implicit rental income generated by owner-occupation, and the deductibility of mortgage interest payments from income. Capital gains from the sale of owner-occupied housing are also subject to exclusions and there are other subsidies in the tax code that accrue to owner-occupiers. Using data from 1990, Gyourko and Sinai (2003) estimate that the treatment of implicit rent and mortgage deductibility alone reached almost 200 billion dollars in tax expenditures. The level of tax subsidization has surely reached much higher levels since then given the increase in housing prices. Moreover there are numerous state and local programs designed to foster ownership, including homestead exemptions from local property taxes and other programs particularly aimed at neighborhoods with below average incomes and environments. All recent presidential administrations have been seen as fostering homeownership; the examples cited in Gabriel and Rosenthal (2004) bear witness to the political popularity of promoting homeownership

Considerable doubt has been cast on the desirability of these policies, especially in light of the recent economic downturn. While this paper is not the appropriate venue for a discussion of that downturn, it is appropriate to note that the usual recitation of its chronology lays some blame on the credit risk posed by households entering ownership without the financial means to do so, and that the public subsidization of homeownership did nothing to discourage such risk-taking. Quite the opposite. However, even before recent events focused US attention on the tax treatment of ownership, there were doubts about the interest deduction and similar subsidies. A prominent example of this is the Presidents Commission Advisory Panel on Tax Reform, which in its final report of November 2005 suggested that it be replaced with a 15% tax credit[1].

Perhaps worse, at least from an efficiency perspective, is the deadweight loss associated with the special tax treatment of owner-occupied housing. Poterba (1992) calculates the deadweight loss

---

[1] It is worth noting that in light of the exclustion of implicit rent, the mortgage interest deduction merely puts debt and cash purchases of houses on the same footing. Even in the absence of the MID, there would still be subtantial financial advantages to ownership.

associated with the income tax code's treatment of housing, and finds it to be substantial, especially so for those in higher tax brackets.

The justification for the tax treatment of housing, or any subsidization of ownership should not rest on its status as a merit good-- that ownership is part of the "American Dream" and thus "should" be accessible to any household-- but with the more compelling justification that ownership creates external benefits; that ownership not only creates private benefits, but benefits for the neighborhood and broader community. Over the past two decades or so a research program has grown around the identification of external effects that are created by ownership. Three sets of effects have been so identified and on that account debated in the literature:

a. Maintenance and appearance: Rossi-Hansberg, Sarte, and Owens (2010), among others, document the strong spillover effects of housing revitalization expenditures. However, in the absence of directed policy expenditures as described by these authors, the literature suggests that neighborhood maintenance is more likely to be undertaken by owner-occupiers than by renters.. Renters have little incentive to do perform maintenace directly, since the return on such investment accrues not to them but to the landlord, and the landlord cannot credibly commit to properly compensating tenants for proper maintenance, or what amounts to the same thing, punishing tenants for excessive wear and tear (Henderson and Ioannides (1982)). Landlords do have an incentive to maintain the property, but this comes at a higher cost when the landlord is an absentee one. Galster (1983) and Harding et al (2000) both come to the conclusion that owner-occupied properties are better maintained than rental properties. DiPasquale and Glaeser (1999) report some similar findings.

b. Family life: Green and White (1997) and Haurin, Parcels and Haurin (2002) both contain evidence to suggest that children growing up in owner-occupied dwellings have higher high school graduation rates and cognitive test scores. Aaronson (1998) notes that this seems to be due to the longer spells that owners have in their place of residence. It should be noted that a recent paper by Barker and Miller (2009) casts doubt on the link between childhood outcomes and ownership, stating that the regressions run by these authors is subject to omitted variable bias, and that car ownership is at least as important as homeownership in this regard.

c. Citizenship: In a widely-cited paper, DiPasquale and Glaeser (1999) provide substantial evidence that homeowners are more involved with local organizations and community, have greater knowledge of their local elected officials, and vote with greater frequency. Not all of this is necessarily productive. Fischel (2001) notes that homeowners will be more active adherents of NIMBYism (i.e. not in my back yard) and that this may devolve to mere rent-seeking. Contrary evidence is obtained by Englehart et al (2009) who use a randomized treatment to obtain exogenous shifting of households into ownership. They find little evidence of increased civic or neighborhood involvement by these new owners.

The lacuna in the above literature is that the benefits that accrue to one's neighbors are only indirectly measured; there is little or no sense that the behaviors identified by the above authors is at all valuable to those that live nearby. Conceptually, it would be a straightforward thing to calculate the externality value of homeownership, even if the behaviors are not directly observable. If ownership is valuable to the owner's neighbors then those neighbors should be willing to pay more to live near owner-occupiers. A hedonic regression, one that correlates housing prices (where this can be either the flow rent, or the asset price) to the numerous structural and locational characteristics embodied in them, could include some measure of the ownership propensity in the neighborhood, and that characteristic should, if the aforementioned externalities exist, have a positive coefficient in the regression. Indeed, Nelson (1979), Kohlhase (1991), and likely dozens of other authors have found that the ownership rate within a census tract has a positive influence on housing prices in that tract. However the obvious problem is that there is unobserved heterogeneity across neighborhoods that can cause the correlation to be spurious. Coulson, Hwang and Imai (2002, 2003) tried come to grips with the problem of consistently estimating the hedonic price of neighborhood homeownership in the presence of this heterogeneity. These authors found that even controlling for unobservable person and neighborhood effects (and tenure choice); neighborhood ownership had a positive impact on housing prices.

This paper moves beyond Coulson, Hwang and Imai in a number of dimensions. First, like Coulson, Hwang and Imai (2002) we use the "cluster samples" of the AHS (about which more below) to identify neighborhoods, but we also note a prominent number of vacancies in the sample. This has prompted us to address the (increasingly relevant) issue of the hedonic price of neighborhood vacancy. While not as prevalent as rental properties, unoccupied properties are potentially a drag on neighborhood property values (Ioannides and Zabel, 2008). But neighborhood vacancy is also subject to the same endogeneity concerns as neighborhood ownership, so it will be necessary to account for this in our estimation procedures.

Second, we estimate separate hedonic parameter vectors for rental and owner-occupied property. Hedonic modeling nearly always chooses one or the other to comprise the sample (or in the case of Coulson, Hwang and Imai (2002) or Bajari and Kahn (1998) constrains the marginal effects of housing attributes to be the same across the two tenure types). In this paper we model rent and value determination as the result of a switching regression process, with endogenous selection into each tenure mode[2]. But in light of the previous paragraph, we also model occupancy on the same principal. Importantly, it is only for occupied properties that we observe a price, thus, we follow Hotchkiss and Pitts (2005) among others[3], in creating a double selection model. First the house is selected into occupancy; conditional on occupancy we observe a price, and that price—rent or value-- is the result of endogenous switching into either the rental or ownership market.

Our third innovation is to exploit the panel nature of the neighborhood cluster data. Our first identification assumption is this: that the prices we observe in the data are a function of the current conditions of the house and neighborhood, but the ownership and occupancy status of those are due to conditions that existed prior to the time of the survey. Migration and tenure decisions are very costly and are the result of long and slow adjustment processes. Thus we have natural exclusion restrictions that the price of a unit at time t does not depend on conditions at t-1; but the ownership and vacancy rate do. This allows for the creation of natural instruments that we use in the estimation of the model.

In the next section, we describe the data set we use in our estimation. We do this first because the nature of the data informs the construction of the empirical model. We then describe that empirical model and our estimation strategy. This is followed by the model estimation, and with these estimates we provide some back-of-the-envelope cost-benefit calculations.

Data

---

[2] Charlier et al (2001) develop a semiparametric estimator for this model.

[3] Hotchkiss and Pitts estimate a model of labor force participation and wages. The wage is observed only those who participate, and then there is selection into the full-time or part-time market, whose data generating processes are distinct.

Our data comes from the American Housing Survey (AHS). While the AHS takes a couple of different forms, our data is from the National Sample, which is a biennial survey of over 50,000 housing units from across the USA. It is important to note that this sample repeatedly surveys the same units, so it is a panel of the units, but not necessarily the same occupants. The AHS records data on the price and physical structure of the units (size, assortment of rooms and other characteristics) and the occupants (including income and some limited financial data, as well as numerous demographic characteristics) as well as the quality of the location and housing unit, as evaluated by the occupants.

There are both rental and owner-occupied units in the sample. The price given in each record corresponds to the tenure type: rental units report the monthly rent, while the owners provide an estimate of the current market value of the unit. This estimate is of course subject to error (as is perhaps the rent reportage), however Kiel and Zabel (1997) note that while the error has a positive bias, given owners' optimism about the value of their assets, this bias is uncorrelated with housing attributes, so that only the intercept term is affected.

Importantly for our purposes, in the 1985, 1989 and 1993 waves of the survey, for a limited number of respondents (called "kernel" respondents) the sample also included "neighborhood clusters". These clusters are normally the 10 housing units that are nearest to the respondent in question. These contiguous units are only sampled in the given years and in particular not surveyed in 1987 or 1991 waves. We assemble a panel data set consisting of the kernels and the surrounding cluster for each of the three surveys. The units are classified as being vacant, rented, or owner-occupied.

In the United States, owner-occupied housing is strongly associated with single family structures. The reasons for this are open to debate (Coulson and Fisher (2009)) but Glaeser and Sacerdote (2000) note that the types of social interactions that occur in the former type of housing can be different than those in the latter. For this combination of reasons, we limit our analysis to clusters which are entirely composed of single family households. To include multiple structure types would complicate our analysis immensely, and to use anything other than single family housing as our sample would unduly limit the size of our database.

Table 1 presents the count of clusters. As can be seen, there are as few as six and as many as sixteen houses in a cluster. As hinted above, 11 is the most prominent number, but as noted by Myers (2004) there are exceptions. For example, when construction contiguous to the cluster occurs, new units may be added to that cluster. Also, demolition or conversion to non-residential use may occur, removing the unit from the sample. However, in order to keep the definition of a cluster constant over time, we only include units if they are present in all three waves[4]. There are 5688 observations in all.

Given our focus on single family structures, and the aforementioned correlation of structure and tenure types, it is of interest to note that there is in fact substantial variation in ownership rates across clusters. Tables 2a and 2b provides evidence on that point. Note that there are a large number of neighborhoods with 100% occuancy and ownership, and an aggregate ownership rate (.81) that is higher than the overall US ownership rate, but both of these facts befit our sample of single family structures. Nonetheless Table 2b demonstrates a wide variety of ownership rates that appear in our sample.

Model

The preceding considerations suggest an empirical model for housing prices- values (P) and rents (R) based on the following log linear approximation:

$$\ln P_{i\,jt} = X_i\varphi + \bar{Z}_{jt}\gamma + \bar{H}_{jt}\delta + \bar{O}_{jt}\theta + \alpha_{it} + \tau_{jt} + \varepsilon_{ijt} \quad (1)$$

$$\ln R_{i\,jt} = X_i\varphi' + \bar{Z}_{jt}\gamma' + \bar{H}_{jt}\delta' + \bar{O}_{jt}\theta' + \alpha'_{it} + \tau'_{jt} + \varepsilon'_{ijt} \quad (2)$$

The log-linear form is a convenience in that it will allow us to compare the sizes of the coefficients in the rental and owner equations. Note that i=1…n indexes housing units, j=1… J indexes neighborhood clusters, and t indexes time periods. Also

[4] There was no subtraction from the sample between 1985 and 1993, and 18 removals between 1989 and 1993. Seven of these are from a single cluster, which was removed from the estimates, entirely. There were many more additions. There were 82 units added to 1989 clusters, of which only 64 were included in 1993. There were 115 additions in 1993, out of the several thousand units that are in our final sample.

P_{ijt}= price of housing unit i in neighborhood j at time period t

R_{ijt}=rent of housing unit i in neighborhood j at time period t

X_{it}= vector physical characteristics of housing unit i in year t.

$\bar{z}_{jt}$=vector of demographic characteristics of neighborhood j. This consists of averages of demographic characteristics of the cluster residents. That is, with n_j as the number of units in cluster j,

$$\bar{Z}_{jt} = \sum_{i=1}^{n_j} Z_{ijt} \ /n_j$$

O_{ijt}=1 if house i in neighborhood j at time t is occupied by an owner or a renter (and =0 if vacant).

H_{ijt} =1 if house i in neighborhood j at time t is owner-occupied (and =0 if occupied by a renter) and is only observed if the house is occupied.

$\bar{O}_{jt}$= occupancy rate in neighborhood j at time t. This is calculated in the same way as $\bar{Z}_{jt}$.

$\bar{H}_{jt}$= homeownership rate in neighborhood j at time t. This is calculated in the same way as $\bar{Z}_{jt}$. Note that for purposes of this calculation $\bar{H}$ is calculated as the percentage of units (and not just occupied units) that are inhabited by their owners.

These regressors have associated coefficients φ,γ,δ, and θ respectively in the hedonic equation describing the owner-occupied market. The prime modifiers indicate analogous parameters in the rental market.

α_{it}=unobserved characteristics of the housing unit that are possibly changing over time.

τ_{jt}= unobserved characteristics of the neighborhood that are possibly changing over time.

ε_{ijt}=random error term

Again, primes indicate parameters from the renter equation. As noted, we observe only the monthly rent for properties that are occupied by renters, and the owner's estimate of property value for those units that are owner-occupied, and neither for vacant units. For reasons noted below, we will limit the estimation to the two latter years in our sample, 1989 and 1993. The appropriate estimation procedure depends critically on the assumptions made about the behavior of α and τ. This is important because the occupancy and ownership status of individual dwellings likely depend on unobserved attributes of the house and neighborhood, and omitting them from the model would bias the coefficients of the ownership and occupancy rates.

Under the usual fixed effects assumption, the α and τ components are constant over time, and therefore observable with panel data. The estimation of the above equations would provide consistent estimates of the parameters. We could identify the price impact of neighborhood ownership and occupation by the usual method of observing changes in $\bar{O}$ and $\bar{H}$ in individual neighborhoods over time, controlling for the unchanging α and τ via fixed effects. (Note that controlling for an unchanging α *ipso facto* controls for unchanging τ.)

However the very fact that units become occupied or vacant in a neighborhood, or switch their tenure status between owning and renting can be related to the fact that these unobservable features are changing as well. A change in $\bar{H}$ is presumably correlated to a changing τ. In this case fixed effect estimation will not yield consistent parameter estimates. But in fact the problem is somewhat more complex than that.

First consider the *switching problem.* We note that P and R are only observed when the unit is occupied, and the respective prices are only observed when the corresponding tenure choice is made. Thus both the occupancy and tenure decision for a given property impacts whether a price is observed for that property and nature of the price that we do observe. Thus we potentially face a two part selection issue. The first selection is into occupied status, and the second selects into the rental or owner market, with the resulting hedonic equation ((1) or (2))being relevant.

In order for a building to be occupied, the owner of the property and the (new) resident must meet and agree to make the appropriate transaction. In the case of rental property, the landlord must meet and match with a renter, while in the ownership market the old owner must do the same with a new owner. We consider the propensity of a building to be occupied can be modeled by a "matchability function":

$$O^*_{ijt} = F_{ijt}\varphi + w_{ijt} \qquad (4)$$

where F is a vector of variables that affect the time on the market that a house remains vacant. We assume that this vector of parameters is the same, regardless of its eventual tenure. This is an explicit assumption about the decision tree that market participants make about housing transactions[5]. In order to not belabor the notation too much, we let w_{ijt} contain all the unobservable elements that are specific to the housing unit and the neighborhood. We will come to the specification of F shortly but it is worthwhile to recall that this data is from an era when the default rate was very low, "foreclosure contagion" (Harding, Rosenblatt and Yao, 2009) was not an issue, and housing vacancies were only rarely due to foreclosure, but rather due to the simple frictions that arise in a market with significant search costs.

As is usual in this kind of context, O* is unobserved, but we do observe the indicator function for the unit being occupied (in which case O=1) :

$$O_{ijt} = 1 \; if \; O^*_{ijt} > 0 => F_{ijt}\varphi > -w_{ijt} \qquad (5)$$

$$= 0 \; if \; O^*_{ijt} < 0 => F_{ijt}\varphi < -w_{ijt}$$

Similarly, we consider the propensity of an occupied building to be owner-occupied. For any given unit, the process through which it comes to be owner-occupied or rented is a function of the size of the building (for tax purposes) and due to its "risk". In the late 1980s and early 1990s, homeowners did not practice asset diversification (Caplin et al,) but rather had upwards of 90% of their portfolio tied up in a single asset, their home. Thus owner-occupiers were presumably gravitating towards purchasing those housing units that were deemed less risky assets. We define a "risk" function, as a latent variable H*:

$$H^*_{ijt} = K_{ijt}\omega + u_{ijt} \qquad (6)$$

and if H=1 if the unit is owner-occupied:

---

[5] Alternative decisions trees are of course plausible. Properties may be owner-occupied, and if not owner-occupied, then the owner must decide whether to leave them vacant or rent them out. This latter tree pre-supposes that all vacant houses are part of the owner-occupied market. The one we use does not make that assumption.

$$H_{ijt} = 1 \; if \; H^*_{ijt} > 0 => K_{ijt}\omega > -u_{ijt} \qquad (7)$$

$$= 0 \; if \; H^*_{ijt} < 0 => K_{ijt}\omega < -u_{ijt}$$

Again, there are unit and neighborhood specific elements to u. Now consider the mean functions that arise from (1) and (2):

$$\mathrm{E}(\ln P_{i\,jt} \,|P_{ijt} observed) = \mathrm{E}(\ln P_{i\,jt} \,|F_{ijt}\rho < -w_{ijt} \; and \; K_{ijt}\omega > -u_{ijt})$$

$$= X_i\varphi + \bar{Z}_{jt}\gamma + \bar{H}_{jt}\delta + \bar{O}_{jt}\theta + E(\tau_{jt} + \alpha_{it}|F_{ijt}\rho < -w_{ijt} \; and \; K_{ijt}\omega > -u_{ijt}) \qquad (8)$$

and for rents:

$$\mathrm{E}(\ln R_{i\,jt} \,|R_{ijt} observed) = \mathrm{E}(\ln R_{i\,jt} \,|F_{ijt}\rho < -w_{ijt} \; and \; K_{ijt}\omega < -u_{ijt})$$

$$= X_i\varphi' + \bar{Z}_{jt}\gamma' + \bar{H}_{jt}\delta' + \bar{O}_{jt}\theta' + E(\tau'_{jt} + \alpha'_{it}|F_{ijt}\rho < -w_{ijt} \; and \; K_{ijt}\omega < -u_{ijt}) \qquad (9)$$

The two-step estimator would be constructed as follows:

1. A bivariate probit model of O and V is estimated. The log likelihood for this is:

$$\log L = \textstyle\sum_{O=0} \log\,[1 - \Phi(-F_{ijt}\rho)] + \sum_{O=1,H=0} \log\,[\Phi_2(-F_{ijt}\rho, K_{ijt}\omega, \sigma_{uw}) + \sum_{O=1,H=1} \log\,[\Phi_2(-F_{ijt}\rho, -K_{ijt}\omega, \sigma_{uw})] \qquad (10)$$

where the first term is the contribution of the unoccupied units to the likelihood, the second term is that of renter homes, and the third is the contribution of owner units. From Poirier (1980) and followers (e.g. Hotchkiss and Pitts (2005), the hedonic regressions become:

$$\ln P_{i\,jt} = X_i\varphi + \bar{Z}_{jt}\gamma + \bar{H}_{jt}\delta + \bar{O}_{jt}\theta + \sigma_{ug}\lambda_{P1} + \sigma_{wg}\lambda_{P2} + \epsilon_{ijt} \qquad (11)$$

$$\ln R_{i\,jt} = X_i\varphi' + \bar{Z}_{jt}\gamma' + \bar{H}_{jt}\delta' + \bar{O}_{jt}\theta' + \sigma_{ug'}\lambda_{R1} + \sigma_{wg'}\lambda_{R2} + \epsilon'_{ijt} \qquad (12)$$

where g=τ+α and the final terms are zero-mean noise encompassing both ε and the deviations of u and w from their respective expectations. The σ terms are the covariances between g and the residuals in the two selection equations. Since F and K (as will be seen) come from the lagged wave of the sample, this is in effect the ability of potential occupants to forecast neighborhood quality at time t. The conditional expectation terms are given as:

$$\lambda_{P1} = \frac{\varphi(-F_{ijt}\rho)[1-\Phi[\frac{(K_{ijt}\omega-\sigma_{uw}F_{ijt}\rho)}{(1-\sigma_{uw}^2)^{\frac{1}{2}}}]}{\Phi_2(-F_{ijt}\rho,-K_{ijt}\omega,\sigma_{uw})} \qquad (13)$$

$$\lambda_{P2} = \frac{\varphi(K_{ijt}\omega)\Phi[\frac{(F_{ijt}\rho-\sigma_{uw}K_{ijt}\omega)}{(1-\sigma_{uw}^2)^{\frac{1}{2}}}]}{\Phi_2(-F_{ijt}\rho,K_{ijt}\omega,-\sigma_{uw})} \qquad (14)$$

$$\lambda_{R1} = \frac{\varphi(-F_{ijt}\rho)\Phi[\frac{(K_{ijt}\omega-\sigma_{uw}F_{ijt}\rho)}{(1-\sigma_{uw}^2)^{\frac{1}{2}}}]}{\Phi_2(-F'_{ijt}\rho,K'_{ijt}\omega,-\sigma_{uw})} \qquad (15)$$

$$\lambda_{R2} = \frac{\varphi(-K'_{ijt}\omega)\Phi[\frac{(F_{ijt}\rho-\sigma_{uw}K_{ijt}\omega)}{(1-\sigma_{uw}^2)^{\frac{1}{2}}}]}{\Phi_2(-F_{ijt}\rho,-K_{ijt}\omega,\sigma_{uw})} \qquad (16)$$

where each λ term is proportional to the conditional expectation of τ or α conditional on the observational status of O and H.

Now consider *the endogeneity problem.* The decisions that come together in neighborhood j to create the cluster occupancy and ownership rate depend in part on the house and neighborhood unobservables. But the neighborhood unobservable is part of the error term for the ith house, at least if τ changes over time. In that case the ownership and occupancy rates in neighborhood j will be correlated with the error term even under fixed effects. Therefore it is necessary to instrument for the ownership and occupancy rates. Think of $\overline{H}$ as n_j endogenous binary regressors under the constraint that they have the same coefficient (δ/n_j). The instrument for each of these binaries would be the predicted probability of ownership. Under the constraint, we clearly have a natural instrument for the ownership rate, which is the predicted ownership rate for neighborhood j. Treating the occupancy rate in analogous fashion these instruments are:

$$\overline{\overline{H}}_{jt} = \frac{\sum_{i \in j} \Phi(-F_{ijt}\rho)}{n_j} \qquad (18)$$

$$\overline{\overline{O}}_{jt} = \frac{\sum_{i \in j} \Phi(-K_{ijt}\omega)}{n_j} \qquad (19)$$

We will discuss in detail the specifications of F and K later, but for the moment we simply note that we rely on data observed at wave t-1 to provide instruments. This will be valid, however, only if the neighborhood effect τ_{jt} is uncorrelated over time. This too is implausible; there is likely to be some--though not universal-- persistence in neighborhood quality. We resolve this by assuming that

$$\tau_{jt} = \tau_j + \tilde{\tau}_{jt} \qquad (20)$$

where τ_j is a standard time-invariant fixed effect and

$$E(\tilde{\tau}_{jt}) = 0 \; for \; all \; t; \;\; E(\tilde{\tau}_{jt-1}\tilde{\tau}_{jt}) = 0 \quad (21)$$

The idea is that τ_i continues to be modeled as a neighborhood fixed effect, which provides for some persistence in unobservable neighborhood quality, but the residual is a random effect, which if correlated with the ownership and occupancy rates, will necessitate the use of instrumental variables. Under the condition implied by (21) lagged variables will be available as instruments in the vectors K and F, as long as the fixed effect (τ_j) is included in the model.

Two issues now arise that will impose limits on our ability to estimate all of the parameters in equations (11) and (12). One is that with the use of lags, we observe three waves, but only two are thereby used for estimation. Thus the unobservable housing-specific characateristic α, will be difficult to estimate for each household. The second is that observing the neighborhood effect as it applies to renters, τ', is also going to be problematic, because for a large number of neighborhoods, the number of renters is zero, or one. Moreover, in such a case the separate observation of α is not possible. These problems are exacerbated when we attempt to estimate the external effects of the occupancy rate, which as shown in Table 2a, much less sample variability than the ownership rate. For that reason, while evidence will be provided on all these dimensions, we are going to concentrate below on the impact of owners on owners. This is appropriate both from the viewpoint of obtaining reliable estimates, with policy importance, and because owners are far more numerous in both the sample and among US households.

Specification and Estimation

Table 3 presents variables means and standard deviations for the variables used in the hedonic regressions (1) and (2), stratified by year and tenure status; for convenience we present this for occupied units only. The first pair of variables is rent and value means for the respective tenures. These are logarithms: the corresponding average value is around $90,000 and the average rent is approximately $400/month. The next set of variables includes the individual demographic variables which are aggregated to the neighborhood level to create measures of neighborhood attributes. (These are also reported in the table at the individual level.) Household income (ZINC) and number of children (CHILD) variables are self-explanatory. RACE=1 if the head of household is white, and zero otherwise. The measures of school quality (SCH), public transportation (PUBTRANS), and shopping (SHPADEQ) are =1 if the locational attribute in question is "adequate". All other responses, including apathetic or ignorant ones are set to zero[6]. By this measure shopping is largely seen as adequate, and public services not so.

---

[6] "Don't know" and "Unable to state". We use these fairly gross measures of locational quality in the belief that more finely measured ratings (also available in the survey) might be contaminated by self-selection of people into neighborhoods based on unobservable person characteristics. These concerns are hopefully minimized if there is general agreement on what 'adequate' means.

Structural characteristics include BATHS, the number of bathrooms. In matching observations across time, we inspected the coded number of full baths and the number of half-baths in the unit. Somewhat surprisingly, we found that these two variables were often not the same from observation to observation of the same house, but that the *sum* of the two was (almost always) identical. We conclude that there is some confusion about the distinction between full and half baths in the minds of the survey respondents and we correspondingly just add up the two numbers for every observation and use the total as the measure of BATHS. There is a very slight increase over time in this variable that can be accounted for by renovation. AGEHOUSE is the age of the unit. This is the year of the survey minus the year that the house was built. The latter is coded into the AHS as a categorical variable, so we take the middle year of the category as our measure of construction year. GARAGE, AIRSYS and the three heating indicators are all equal to 1 if the given attribute exists. Regional and center city dummies are included in our variable list, but are only used in those specifications where fixed effects are not used, and therefore not included in the tabled results. Finally, two measures of size, LOT (square feet of the lot) and UNITSF (interior square feet) , are included in the model. A time dummy (for 1993) is also included.

Table 4 presents initial estimation results . The first two columns merely provide a benchmark by presenting the results of a regression of value and rent on the two key variables, the ownership rate and the occupancy rate. As can be seen, in both equations, both of the indicators are positive and statistically significant at the usual levels. The dependent variable is the log of the price, and the two rates are listed in decimal terms so the coefficients of these two variables can be read as approximately the percentage change in housing price when moving the rate from zero to 100%. Thus the stated increase in homeownership raises the value of said owner-occupied property by about 95%. Increasing the occupancy rate by 10 percentage points yields a 13% increase in housing values in the cluster. The effect on rental property are smaller, an increase in the ownership rate by 10 percentage points leads to a 2.3 % increase in rents.

The next two columns present the estimates from the OLS estimation of equations (1) and (2). Again, this is mostly for benchmarking purposes, and to demonstrate that our sample corresponds for the most part to received wisdom on hedonic price indices. To that end, note that the important structural characteristics have positive coefficients and have standard errors that are estimated pretty precisely. An additional bathroom adds about 13% to property values and 10% to rent. Both measures of size have statistically significant coefficients in the value equation, but lot has unexpected negative sign in the rental market, although this coefficient does not meet the usual criterion for precision. A garage adds about 13% to value and 9% to rent. Age is negative as expected, and the coefficient suggests for both owned and rented units a depreciation rate between .1 and .2 percent per year. This is in line with other estimates although there is always the possibility of survivor bias contributing to what looks like very slow depreciation (see also Coulson and McMillen, 2008). These specifications also contain binaries for location. These are measures of proximity to the central city of metropolitan areas based on the setting of the cluster, and move from the omitted category of central city, through adjacent suburban locations (category 2) through rural locations not adjacent to metropolitan areas (category 6). Compared to central cities, suburban locales have higher prices, but rural locations have lower values and rents.

Turning to the measures of cluster quality, the adequacy of schools has a negative weight (surprisingly) but the coefficients of shopping and transportation adequacy both are positive and significant. Neighborhood demographics, average income and percentage white, are significant in the value equation—somewhat at odds with other studies, the latter has a negative sign—but insignificant in the rental equation. The most surprising finding is that the ownership rate has a negative weight in the determination of both value and rent. This is at odds with the first two columns and with other studies. The coefficient of the occupancy rate is positive in both.

The next four columns present the estimates for fixed effects regressions. In the columns labeled "Fixed House Effects" the fixed effect is at the level of the house. That is, it is a more or less standard panel data model with two observations (1989 and 1993) per panel member. Thus the coefficients should be read as the effect of changes in the independent variables on the change in house price between those two dates. Note first of all that (the change in) age disappears from the regression because it is perfectly collinear with (the change in year of observation). The structural variables (unitsf, lot, baths, garage and the heating and cooling binaries) are only identified to the extent that such improvements were carried out in the sample units. Obviously this is fairly rare, so the sample variation in this model is not particularly large, and the standard errors are correspondingly high. Even the neighborhood variables, which do have somewhat more sample variability, are not estimated with any great precision in this model. However, and importantly, note that the coefficient on the homeownership rate in the value equation is now positive and statistically significant (t=3.7). Combining this observation with the previous model which lacked fixed effects. the inference we draw is that homeowners are more or less selecting into neighborhood with low quality. Despite the inclusion of regional indicators in the previous model, our suspicion is that this is due to ownership mostly occurring outside of high-priced – meaning high-access—locations. The coefficient indicates that an increase in the cluster ownership rate of 10 percentage points increases property values in the cluster by about 4.8%. It is worth noting however that none of the other key coefficients has a precisely estimated parameter. This experience will be often repeated in other specifications; once fixed effects have been controlled for, there seems to be relatively little explanatory power for the occupancy rate, as befits the lower sample variability of this indicator. Moreover there are fewer rental properties than owner properties, and this seems to have an impact on the ability to explain variation of rental prices. Because most properties are owner-occupied, and because our interest is primarily in the effect of the ownership rate, this is not particularly damaging to our ability to draw interesting conclusions, as will be seen.

In the next two columns we revert to using neighborhood, rather than house fixed effects. There can be as many as 22 (and occasionally more) observations (using both periods) for a given neighborhood. This has a number of advantages over house fixed effects. For one thing, although this is not our main concern, it allows for far more precise estimates of the parameters of the structural housing characteristics. All of these parameters are positive and statistically significant in the owner equation, and many have these properties in the rental equation. However, as might be expected, the neighborhood demographic variables are still imprecisely estimated. Interestingly, the key parameter estimate, the coefficient of the ownership rate in the owners' equation, hardly changes at all. Its standard error remains the same as well.

The use of neighborhood rather than house fixed effects does not appear to be critical for the evaluation of the effect of neighborhood ownership, and will allow for separate identification of the selection parameters, as we discuss below, so it appears to be an appropriate modeling strategy. Nevertheless it is helpful to see whether this might justified on statistical grounds. We do this by estimating a model (not presented) which includes both neighborhood fixed effects and house fixed effects. Thus the house fixed effects can be thought of as deviations from the neighborhood fixed effects that are nevertheless constant across time. We test for the joint significance of these terms. The appropriate F statistic has a prob-value of .03 in the value equation and of .32 in the rental equation. Thus our treatment of the fixed effects is certainly appropriate in the latter, and marginally so under the former if a small value for Type I error is used. Given the desirability of using neighborhood fixed effects this seems acceptable.

We turn, then, to the specification of the selection terms and instruments. We are guided by two principles. The first is that Lahiri and Song (2002) stress that functional form restrictions are not enough, in a practical sense to identify parameters in the hedonic function, especially using the two-step approach we are using here. Second, recall that we define F and K as measures of "matchability" and "risk". Haurin (1988) noted that the "atypicality" of a housing unit contributed to its time on the market, and by extension, the probability of it being vacant at any particular point in time. The atypicality is defined as the weighted distance of a housing unit's attributes from the "average". Haurin (1988) discussed an "atypicality" metric that used hedonic prices as weights. In our procedure, the determination of atypicality must precede the hedonic calculation (and this is congruent with our identification assumption above) so we use the somewhat simpler Mahalanobis distance to measure this. Letting A be the entire vector of variables in the hedonic regression, the Mahalanobis distance, M_i is the quadratic distance from the average:

$$M_i = (A_i - \bar{A})' cov(A)^{-1} (A_i - \bar{A}) \qquad (22)$$

Thus the measure not only increases when a housing attribute is far from the mean, but also when unusual combinations of attributes occur. For each house we define the distance both with respect to the mean of the entire sample (*Mahal)* and with respect to the mean of the cluster (*Localmahal*). . These variables are included in the F vector that comprises the determinants of occupancy probability, but can reasonably be excluded from the price equations; thin markets do not necessarily lower (or raise) the price of a commodity. We further speculate that these measures are determinants of risk, in the sense that a more atypical property also has more variable future price path, and therefore include these measures in K, the vector of variables associated with the probability of ownership. We also include lagged values of the neighborhood occupancy rate in the probability of occupancy equation and the lagged neighborhood ownership rate in the probability of ownership equation. The exclusion of the lagged occupancy rate from K presumably aids in the identification of the parameters of the tenure selection equation.

Finally, we include in the vector K a variable called *tax*. This variable is the average marginal tax rate as it applies to the mortgage interest deduction and is specific to the state of residence and the time period of the housing unit. Thus it is the sum of the average federal marginal tax rate for the state in question, plus the average state marginal tax rate if that state also allows for the deductibility of mortgage interest. This data is available online from the National Bureau of

Economic Research. The difficulty we encounter is that the American Housing Survey does not identify the state of residence of the housing unit. It does, however, identify the CMSA or SMSA of the unit, if it is in a metropolitan area. Thus, if this datum is in the record we assume that the residence is in the same state as the central city of the metropolitan area, and include the corresponding tax rate in the model for ownership probability. Thus, going forward the sample excludes rural clusters[7].

Table 5 presents the results of the estimation of the selection models. As can be seen, the lagged rate (whether occupancy or ownership) plays a significant role in determining the status of units, as is natural. The "global" Mahalanobis measure is negative in both the ownership and occupancy equations, and is statistically significant in the former. This is congruent with expectations: the more unusual the house, the thinner the market, presumably, and the longer it would take to find a buyer/renter, as surmised by Haurin(1988). And of those that are occupied, unusual houses are more likely to be rental properties. Caplin et al (1997) note that households who are homeowners in the early 1990s had more than 90% of their portfolio in this one (illiquid) asset. It makes sense for such households to avoid highly risky properties. Landlords, who are more likely to have a portfolio of real estate assets, are better equipped to diversify away this idiosyncratic risk. This line of reasoning does not however extend to the local Mahalanobis measure. The coefficient of *localmahal* is positive in both equations and is marginally significant in the ownership model. Conditional on having chosen a neighborhood, people evidently seek out unusual houses in that neighborhood. Incentives certainly exist in both directions. They may do so for fiscal advantage: the smallest house in the neighborhood shares the common resources while paying the lowest property taxes, or they may seek to be conspicuous consumers on their block (Turnbull, Dombrow and Sirmans (2006), Leguizamon (2010)). Finally note that the tax variable is positive as expected—higher tax subsidies for ownership encourages ownership—but is not estimated particularly precisely. In any case the effect is modest. The parameter suggests that with a marginal tax rate of about 30% the interest deductibility or rental exclusion increases the ownership probability by about 2.7%

The first two columns of Table 5 present the estimates using both fixed (neighborhood) effects and instrumental variables, treating both the ownership and occupancy rate and endogenous.

[7] The results from the OLS and fixed effect regressions are not substantially different with this data exclusion. Results available on request. Note that there is a case for including *tax* in the price equation. Doing so yielded a negative coefficient with a very small t, so we retain its use as an excluded instrument.

The instruments include the predicted ownership and occupancy rates as suggested by equations (21) and (22). We also include *mahal* and *localmahal* as separate instruments, given that these two variables enter (21) and (22) nonlinearly. The Hausman test for exogeneity is easily rejected (in this and all subsequent specifications) suggesting that using instruments for these two variables is appropriate. The value of the key parameter estimate rises very slightly, to 0.49. The ownership rate coefficient is 0.68 and its t-ratio is nearly two in the rental equation. The occupancy rate is negative but insignificant for both value and rent.

The selection terms described above (13) through (16) were inserted into the hedonic as suggested by equations (11) and (12). As noted previously, the absence of house-specific fixed effects (net of neighborhood effects) means that by accounting for the selection terms we are merely assuming that these house-specific effects are zero mean, random effects that are uncorrelated across time. The selection term coefficients appear in the third and fourth columns in the rows labeled *lambdai* and *lambda2i,* where i=v,r as appropriate. The first of these is the selection term for occupancy, and the second is the selection term for ownership. As can be seen, the former is statistically insignificant but the second one is fairly precisely estimated (though smaller in absolute value). The coefficient on the ownership rate in the owners equation increases rather dramatically to 1.47. An increase in the ownership rate of 10% increases property values by about 14%. None of the other four coefficients of interest is estimated particularly precisely.

In the final specification, we take the natural next step of treating the two neighborhood demographic averages, *avgincome* and *white* as endogenous. Since we have used four instruments in the previous specifications the model remains identified. The results are surprisingly similar to the previous specification. In particular the ownership rate coefficient in the owners equation is 1.56, barely changed from the previous specification. The other three coefficients are all imprecisely estimated. The coefficients on the two new endogenous variables increase in absolute value, but are also imprecisely estimated. Clearly it is becoming difficult to separately identify the various effects under consideration here but the important conclusion remains, that neighborhood ownership affects owners.

We consider here three variants whose results are summarized. First (letting *avgincome* and *white* remain exogenous) we include the square of ownership in the list of endogenous variables and re-estimate the owners equation. When this change in the specification is made, the linear part of ownership has a coefficient of 0.72 and the quadratic term has a coefficient of 0.77. Neither of these two have a t-ratio greater than 1.3 but this does suggest that the benefits to neighborhood ownership increase at an increasing rate. Second, we replace the quadratic term with a binary that equals one if the ownership rate equals one. There appears to be a jump of about 9% at this discontinuity. This binary has a t-ratio of about unity. The conclusion is somewhat similar, that high ownership neighborhoods benefit more than proportionately from the homeownership externality. Finally we interact the ownership rate with lotsize. The hypothesis is that neighborhood externalities mean more when houses are closer together; this does not appear to be the case, since the coefficient on the interaction term is positive (again with a t-stat about one). Since "better" neighborhoods will presumably have bigger lots (and more ownership), this again suggests an increasing return to neighborhood ownership.

Given the robustness of our main result it is worth performing a back of the envelope calculation. Note that the typical neighborhood in our sample has 11 houses. We ignore the effect of neighborhood ownership on rental units as being too imprecisely estimated. So assuming nine of these eleven are owner-occupied, the transition of the tenth unit from rental to owner would raise the ownership rate by 9.1%. Let us assume that the ownership effect is as estimated in the penultimate specification of Table 5. Then each house of the other nine houses experiences a price increase of approximately .091*1.47= 13.3%. Assuming a typical sample property value of $90,000, this amounts to about $12040 per housing unit. Thus the externality benefit of ownership in a neighborhood is 9*3680=$108,353. If a 3% annual capitalization rate is applied (assuming an infinite lived asset), this yields an annuity of approximately $3250 per year in externality value. If we take the more modest and more precise estimate of 0.49, the externality value is $1083.

Poterba (1992) calculates the deadweight loss that accompanies the use of the mortgage interest tax deduction for 1990 taxpayers (i.e. under the 1986 Tax Reform Act). This date is quite congruent with our use of data from 1989 and 1993. This loss varies considerably across income groups and Poterba gives an estimate for those with income of 30, 50 and 250 thousand dollars. The annualized deadweight loss is $53, $326, and $1631 respectively. Thus, the calculations above suggests (subject to considerable variation, obviously) that if the mortgage interest deduction happens to induce additional ownership, the benefits from that marginal owner outweigh the deadweight loss of the deduction for all but the highest income households.

Conclusions

We have attempted to estimate the externality value of homeownership in the context of the small neighborhood clusters constructed by the American Housing Survey in 1989 and 1993. Our estimation strategy considers a wide variety of assumptions on the nature of the unobservable housing and neighborhood effects. The estimates range rather substantially, but a benchmark model with modest assumptions about those effects suggest that transiting a home from rental to ownership in a typical neighborhood would create $1000 -$3000 per year in externality value, suggesting that the mortgage interest deduction may very well pass the cost-benefit test.

References

D. Aaronson (2000) "A Note on the Benefits of Homeownership" *Journal of Urban Economics*, 47, 356-369

David Barker and Eric Miller (2009) "Homeownership and Child Welfare" R*eal Estate Economics*, 37, 279-303

Andrew Caplin*,* Charles Freeman*,* Joseph Tracy*,* Sewin Chan (1997) *Housing Partnerships : A New Approach to a Market at a Crossroads,* MIT Press

Erwin Charlier, Bertrand Melenberg, and Arthur van Soest, (2001) "An analysis of housing expenditure using semiparametric models and panel data," *Journal of Econometrics*, . 101, 71-107

Edward Coulson and Lynn Fisher (2009) "Structure and Tenure" unpublished

Edward Coulson, Seok-joon Hwang and Susumu Imai (2002) "The Value of Owner-Occupation in Neighborhoods" *Journal of Housing Research,* 13, 153-174

Edward Coulson, Seok-joon Hwang and Susumu Imai (2003) "The Benefits of Owner-Occupation in Neighborhoods" (with Susumu Imai and Seok Joon Hwang), *Journal of Housing Research*, 14, 21-48

Edward Coulson and Daniel McMillen (2009) "Estimating Time, Age and Vintage Effects in Housing Prices" *Journal of Housing Economics*, 17, 138-151

D. DiPasquale and E. Glaeser (1999) "Incentives and Social Capital: Are Homeowners Better Citizens?" *Journal of Urban Economics,* 45, 354-384

Gary V. Engelhardt, Michael D. Eriksen, William G. Gale, Gregory B. Mills (2009) "What Are the Social Benefits of Homeownership? Experimental Evidence for Low-Income Households" *Journal of Urban Economics,* forthcoming

William Fischel (2001) *The Homevoter Hypothesis: How Home Values Influence Local Government Taxation, School Finance, and Land-Use Policies*. Cambridge: Harvard University Press, 2001

Stuart Gabriel and Stuart Rosenthal (2004) "Homeownership in the 1980s and 1990s: aggregate trends and racial gaps" *Journal of Urban Economics*, 57, 101-127

George Galster (1983) "Empirical Evidence on Cross-Tenure Differences in House Maintenance and Conditions" *Land Economics ,* 59, 107-113

Edward Glaeser and Bruce Sacerdote (2000) "The Social Consequences of Housing" *Journal of Housing Economics,* 9, 1-23

Richard Green and Michelle White (1997) "Measuring the Benefits of Homeowning: Effects on Children," *Journal of Urban Economic*s, 41.

Joseph Gyourko and Todd Sinai (2003) **"**The Spatial Distribution of Housing-Related Ordinary Income Tax Benefits, *Real Estate Economics,* 31, 527-575

John Harding Thomas J. Miceli, and C. F. Sirmans.(2000) "Do Owners Take Better Care of Their Housing Than Renters?" *Real Estate Economics*, 28,: 663 – 681

John Harding, Eric Rosenblatt and Vincent Yao (2009) "The Contagion Effect of Foreclosed Properties", *Journal of Urban Economics*, 66, 164-178

Donald Haurin (1988) "The Duration of Marketing Time of Residential Housing" *American Real Estate and Urban Economics Association Journal,* 16, 396-410

D. Haurin, T. Parcells, and R. Haurin (2002) "The Impact Of Home Ownership On Child Outcomes" *Real Estate Economics*, 30, 635-666.

J. V. Henderson and Y. M Ioannides (1982) "A Model of Housing Tenure Choice" *American Economic Review ,* 72, 98-113

Katherine A. Kiel and Jeffrey E. Zabel (1997) "Evaluating the Usefulness of the American Housing Survey for Creating House Price Indices" *Journal of Real Estate Finance and Economics* 189-202

Janet Kohhase (1991) "The Impact of Toxic Waste Sites on Housing Values" *Journal of Urban Economics*, 30, 1-26

Kajal Lahiri and Jae Song (2000) "The Effect of Smoking on Health Using a Sequential Self-Selection Model" *Health Economics*, 9, 491-511

Caitlin Knowles Myers (2004) "Discrimination and neighborhood effects: understanding racial differentials in US housing prices" *Journal of Urban Economics,* 56*, 279-302*

Suzanne Leguizamon (2010) "The Effect of Reference Group House Size on House Price" *Real Estate Economics,* forthcoming

Jon Nelson (1979) "Airport Noise, Location Rent, and the Market for Residential Amenities" *Journal of Environmental Economics and Management,* 6, 320-331

Dale J. Poirier (1980) "Partial observability in bivariate probit models" *Journal of Econometrics*, 12, 209-217

James Poterba (1992) "Taxation and Housing: Old Questions, New Answers" *American Economic Review,* 82, 237-242

Geoffrey Turnbull, Jonathan Dombrow, and C.F. Sirmans (2006) "Big House, Little House: Relative Size and Value," *Real Estate Economics*, 34, 439-456.

Table 1.

Number of Units in Neighborhood	Number of Neighborhoods
6	1
7	1
8	3
9	11
10	31
11	309
12	8
13	2
16	1

Table 2a: Distribution of Occupancy Rates

Occupancy Rate in Cluster	Number of Neighborhoods in 1989	Number of Neighborhoods in 1993
0	1	1
0.09091	0	0
0.18182	2	0
0.36364	0	1
0.45455	1	0
0.54545	2	4
0.6	0	0
0.63636	3	6
0.7	1	1
0.72727	6	14
0.75	2	2
0.76923	0	1
0.77778	3	6
0.8	5	7
0.8125	0	0
0.81818	19	23
0.83333	0	2
0.85714	0	1
0.875	1	1
0.88889	2	3
0.9	11	6
0.90909	84	75
0.91667	5	3
0.92308	0	1
0.9375	1	1
1	219	208

Table 2b: Distribution of Ownership Rates

Ownership rate in cluster	Number of Neighborhoods in 1989	Number of Neighborhoods in 1993	Ownership rate in cluster	Number of Neighborhoods in 1989	Number of Neighborhoods in 1993
0	9	9	0.533333	1	0
0.090909	1	1	0.538462	0	0
0.1	1	2	0.545455	8	8
0.111111	1	0	0.555556	3	2
0.125	1	0	0.571429	2	2
0.142857	0	1	0.583333	0	0
0.153846	1	0	0.6	13	4
0.166667	1	0	0.625	2	2
0.181818	0	1	0.636364	8	7
0.2	2	2	0.666667	6	7
0.222222	0	1	0.7	11	11
0.25	0	1	0.714286	1	2
0.272727	1	2	0.727273	14	14
0.285714	0	1	0.75	3	8
0.3	0	1	0.777778	9	6
0.307692	0	0	0.8	22	16
0.333333	2	2	0.818182	27	23
0.363636	0	2	0.833333	3	4
0.375	1	0	0.857143	3	3
0.4	3	2	0.875	5	5
0.416667	0	1	0.888889	6	9
0.428571	3	2	0.9	22	30
0.444444	2	4	0.909091	48	37
0.454546	4	3	0.916667	0	0
0.466667	0	1	1	109	120
0.5	8	11			

Table 3: Means and Standard Deviations, by year and tenure status (occupied units only)

	rent 1989		rent 1993		owner 1989		owner 1993	
Variable	mean	sd	mean	sd	mean	sd	mean	sd
value= *log of owners estimate of value*	--	--	--	--	11.34	0.80	11.41	0.73
rent= *log monthly rent*	5.95	0.65	6.06	0.64	--	--	--	--
zinc =*household income*	30871.72	22734.27	33985.39	27462.27	43090.19	31954.48	46968.33	34224.04
race= 1 *if white*	0.74	0.44	0.71	0.45	0.83	0.38	0.82	0.39
sch= 1 *if schools are "adequate"*	0.33	0.47	0.41	0.49	0.27	0.44	0.25	0.43
child= *number of children*	1.07	1.38	1.16	1.38	0.69	1.12	0.64	1.06
pubrans = *1 if public trans. is "adequate"*	0.32	0.47	0.33	0.47	0.31	0.46	0.31	0.46
shpadeq = 1 *if shopping "adequate"*	0.85	0.36	0.90	0.30	0.89	0.31	0.91	0.28
baths= *number of bathrooms*	1.44	0.67	1.51	0.66	1.93	0.87	1.98	0.91
agehouse= *age of the housing unit*	35.19	23.50	37.90	23.52	33.26	21.07	35.25	21.21
garage = 1 if *attached garage*	0.59	0.49	0.60	0.49	0.80	0.40	0.81	0.39
airsys= 1 *if central A/C*	0.25	0.43	0.31	0.46	0.47	0.50	0.51	0.50
heatgas = 1 *if gas heat*	0.67	0.47	0.65	0.48	0.67	0.47	0.68	0.47

heatoil= 1 *if oil heat*	0.07	0.25	0.08	0.26	0.12	0.32	0.11	0.32
heatelec = 1 *if electric heat*	0.19	0.39	0.19	0.40	0.17	0.37	0.18	0.38
Regionne	0.09	0.22	0.08	0.28	0.19	0.40	0.20	0.40
regionmwnc	0.20	0.40	0.19	0.39	0.23	0.42	0.23	0.42
regionwest *=1 for indicated region*	0.32	0.47	0.34	0.47	0.22	0.41	0.22	0.41
centercity= *1 if in center city of msa*	0.41	0.49	0.47	0.50	0.37	0.48	0.37	0.48
lot = *square ft of lot*	14739.74	40330.27	13910.00	30059.29	16823.56	34800.86	18252.40	47824.71
unitsf = *interior square ft.*	1437.56	698.40	1429.27	6.04	1948.54	847.07	1953.84	844.61
mahal = *mahalanobis distance*	4.29	1.26	4.30	1.02	4.19	1.11	4.18	1.21

	OLS		OLS		Fixed house effects		Fixed Neigh. Effects	
	Value	Rent	Value	Rent	Value	Rent	Value	Rent
homown	0.95	0.23	-0.14	-0.31	0.48	0.30	0.45	0.24
	0.06	0.09	0.05	0.08	0.13	0.31	0.13	0.28
occupied	1.32	1.70	0.24	0.38	-0.02	0.13	-0.10	0.01
	0.12	0.22	0.11	0.23	0.15	0.32	0.15	0.30
avgincome			1.85E-05	1.66E-05	1.03E-06	1.04E-05	1.01E-06	1.26E-05
			4.99E-07	1.47E-06	1.04E-06	4.41E-06	1.08E-06	4.15E-06
white			-0.10	0.10	0.07	-0.23	0.13	0.01
			0.03	0.08	0.12	0.33	0.13	0.31
schadeq			-0.40	0.08	0.02	-0.11	0.00	-0.23
			0.05	0.10	0.07	0.22	0.07	0.20
shopadeq			0.08	0.25	0.13	0.13	0.14	0.11
			0.05	0.12	0.06	0.20	0.07	0.19
trans			0.28	0.11	-0.09	-0.15	-0.09	-0.12
			0.03	0.08	0.05	0.17	0.05	0.16
baths			0.13	0.10	0.03	-0.05	0.10	0.12
			0.01	0.04	0.02	0.13	0.01	0.05
unitsf			5.78E-05	5.56E-05	-6.26E-05	-1.22E-03	6.58E-05	1.50E-04
			1.06E-05	3.11E-05	1.05E-04	7.01E-03	9.52E-06	4.67E-05
lot			3.05E-07	-1.27E-06	9.89E-08	-9.49E-07	4.68E-07	-5.58E-07
			1.82E-07	5.20E-07	3.01E-07	3.79E-06	1.53E-07	7.75E-07
garage			0.13	0.09	-0.02	-0.05	0.04	0.04
			0.02	0.04	0.04	0.11	0.02	0.05
airsys			-0.06	0.02	0.03	-0.13	0.07	0.00
			0.02	0.05	0.04	0.15	0.02	0.07
heatgas			-0.09	0.20	0.10	0.03	-0.03	0.04
			0.04	0.08	0.05	0.11	0.04	0.08
heatoil			0.20	0.29	0.01	0.32	-0.03	0.19
			0.05	0.10	0.08	0.20	0.04	0.12
heatelec			-0.05	0.22	0.06	0.25	0.00	0.21
			0.05	0.09	0.06	0.13	0.04	0.09
agehouse			-1.57e-03	-1.77e-03	(omitted)	(omitted)	-6.47e-04	-1.21e-03
			4.21e-04	8.98e-04			4.72e-04	1.30e-03
year			-0.03	-0.01	-0.01	-0.01	-0.01	-0.01
			0.00	0.01	0.00	0.01	0.00	0.01
_Imetro93_2			0.21	0.09				
			0.02	0.05				
_Imetro93_3			0.19	-0.02				
			0.03	0.07				
_Imetro93_4			0.34	0.34				
			0.32	0.50				
_Imetro93_5			0.05	0.02				
			0.06	0.12				
_Imetro93_6			-0.11	-0.26				
			0.03	0.07				
_cons	9.12	4.13	12.30	5.18	11.80	7.90	11.70	5.25
	0.11	0.19	0.36	0.87	0.37	9.85	0.33	0.84
N	5681.00	774.00	5681.00	774.00	5681.00	774.00	5681.00	774.00
r2	0.08	0.10	0.49	0.45	0.02	0.06	0.05	0.08

Table 4: Coefficients from specified regressions. Standard errors in gray. The R^2 in the fixed effect regression is the within group goodness of fit.

ownership	
homown	2.60E+00
	1.30E-01
mahalanob	-9.36E-11
	4.85E-11
localmahal	2.01E-11
	1.20E-11
tax	5.97E-03
	8.57E-03
_cons	-9.46E-01
	2.64E-01
status	
occupied	2.38E+00
	2.45E-01
mahalanob	-3.70E-12
	5.58E-11
localmahal	1.03E-11
	1.42E-11
_cons	-7.32E-01
	2.33E-01

Table 5 The coefficients from the heckman selection into occupancy and ownership. Standard errors in gray.

	Fixed Effects, IV		Fixed Effects, IV, and Selection		Fixed Effects, IV for all neighborhood demographics, and Selection	
	Value	Rent	Value	Rent	Value	Rent
homown	0.49	0.68	1.47	-1.93	1.56	-1.92
	0.20	0.35	0.65	1.68	0.94	3.39
occupied	-0.31	-0.42	1.10	-4.81	3.08	-5.29
	0.24	0.39	1.46	3.05	5.57	5.59
avgincome	1.08E-06	6.98E-06	9.53E-07	5.48E-06	9.87E-06	1.05E-07
	1.29E-06	4.82E-06	1.29E-06	4.80E-06	7.91E-05	8.79E-05
white	-0.09	-0.03	-0.17	0.08	-2.05	0.58
	0.16	0.42	0.17	0.41	2.97	5.38
schadeq	-0.02	-0.49	-0.01	-0.64	-0.11	-0.64
	0.09	0.22	0.09	0.24	0.60	0.24
shopadeq	0.28	0.06	0.25	0.32	0.40	0.24
	0.10	0.25	0.10	0.27	0.77	0.83
trans	-0.13	-0.08	-0.14	-0.02	-0.24	0.06
	0.06	0.20	0.06	0.19	0.12	0.95
baths	0.08	0.15	0.08	0.16	0.08	0.16
	0.01	0.06	0.01	0.06	0.02	0.06
unitsf	5.41E-05	1.48E-04	5.38E-05	1.44E-04	5.38E-05	1.45E-04
	1.24E-05	6.35E-05	1.24E-05	6.24E-05	1.27E-05	9.20E-05
lot	2.22E-07	2.26E-07	1.70E-07	1.53E-08	2.39E-07	1.20E-07
	2.24E-07	1.14E-06	4.08E-07	1.38E-06	9.81E-07	1.76E-06
garage	0.02	0.07	0.02	0.07	0.02	0.07
	0.02	0.06	0.02	0.06	0.04	0.06
airsys	0.06	0.04	0.06	0.04	0.06	0.03
	0.02	0.08	0.02	0.08	0.04	0.11
heatgas	-0.05	-0.14	-0.05	-0.13	-0.05	-0.14
	0.05	0.11	0.05	0.10	0.10	0.18
heatoil	-0.05	0.12	-0.05	0.15	-0.05	0.16
	0.06	0.14	0.06	0.14	0.06	0.23
heatelec	-0.07	0.13	-0.08	0.15	-0.08	0.14
	0.05	0.12	0.05	0.12	0.07	0.23
agehouse	4.94E-04	-2.36E-03	4.15E-04	-2.24E-03	4.32E-04	-2.21E-03
	6.48E-04	1.99E-03	6.50E-04	1.96E-03	6.72E-04	2.53E-03
year	-0.02	0.00	-0.02	0.01	-0.04	0.02
	0.00	0.01	0.00	0.01	0.06	0.07
lambdap			1.77		2.91	
			1.91		5.34	
lambda2p			0.42		0.56	
			0.28		0.37	
lambdar				-6.64		-7.22
				4.80		7.60
lambda2r				-1.88		-1.89
				1.28		2.43
_cons	12.60	5.52	11.10	11.90	11.80	11.70
	0.46	1.12	1.72	3.60	7.13	4.19
N	3608	447	3608	447	3608	447
R2	.038	.136	.039	.174	.000	.167

Table 6: notes see Table 4.