The Pennsylvania State University
The Graduate School

ESSAYS IN HOUSING AND HOMEOWNERSHIP

A Dissertation in
Economics
by
Herman Li

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

August 2011

The dissertation of Herman Li was reviewed and approved* by the following:

N. Edward Coulson
Professor of Economics
Dissertation Advisor, Chair of Committee

Mark Roberts
Professor of Economics

Sung Jae Jun
Assistant Professor of Economics

Brent Ambrose
Smeal Professor of Real Estate

Vijay Krishna
Distinguished Professor of Economics
Director of Graduate Studies in Economics

*Signatures are on file in the Graduate School.

Abstract

CHAPTER 1:

Measuring the External Beneifts of Homeownership (with N. Edward Coulson)

This chapter analyzes homeownership externailities in a neighborhood setting. 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. The subsidization of homeownership, 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. 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.

CHAPTER 2:

Analyzing Home Improvement Behavior in a Dynamic Setting

This chapter concentrates on the determinants of home improvement behavior in a dynamic setting both with and without neighborhood effects. In the evolution of the body of home improvement literature, there have been two basic ways of approaching home improvement behavior. The first method looks at moving behavior in that a homeowner can "change" housing capital simply by moving. The other method, which becomes my focus, is the physical improvement of a homeowner's own unit. Early studies first analyzed improvement behavior using a static framework, and until recently, only very basic dynamic analyses of home improvement behavior. This is an extension of the current body of literature concerning home improvement behavior with an emphasis on dynamic and neighborhood aspects. While using data publicly available from the American Housing Survey (AHS), I conduct analyses that examine the home improvmement behavior in various dynamic settings including one with neighborhood effects. According to the US Census Bureau, \$135 billion was spent on home improvement activities in 2007 alone, so the analysis is not only looks at a vital part of the economy, but also a vital part of homeownership.

Table of Contents

List of Tables

Acknowledgments

There are many who helped me complete this dissertation. First and foremost, I would like to thank my advisor, Ed Coulson. Without his guidance and support, none of this would have been possible, and I am truly honored to have studied under his direction. I would like to thank my committee members, Mark Roberts, Sung Jae Jun, and Brent Ambrose for providing me with extremely helpful comments and suggestions for my work. I would especially like to thank Ayşe Özgür Pehlivan for incredible support and help throughout everything. I would also like to thank my family for their love and support. I am also grateful to David and Carolyn Grow for their endowment to the College of Liberal Arts as well as funding from the Lincoln Institute of Land Policy. I would also like to thank Vijay Krishna who has helped me many times during my entire graduate career. Last, but certainly not least, I would like to thank all of the friends I made during my time here.

Dedication

For my mother and father.

Chapter 1

Measuring the External Benefits of Homeownership

1.1 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 .

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 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 . 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 , 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.

1.2 Data

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 1.1 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 . 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 1.2-1.4 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 Tables 1.3 and 1.4 demonstrates a wide variety of ownership rates that appear in our sample.

1.3 Model

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

$$\ln(P_{ijt}) = \varphi X_i + \gamma Z_{jt} + \delta H_{jt} + \theta O_{jt} + \alpha_{it} + \tau_{jt} + \varepsilon_{ijt}(1)$$

$$\ln(R_{ijt}) = \varphi' X_i + \gamma' Z_{jt} + \delta' H_{jt} + \theta' O_{jt} + \alpha'_{it} + \tau'_{jt} + \varepsilon'_{ijt}(2)$$

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 $i = 1 \ldots N$ indexes housing units, $j = 1 \ldots J$ indexes neighborhood clusters, and t indexes time periods. Also

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.

$\overline{Z}_{jt}$=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,

$\overline{Z}_{jt} = \frac{\sum_{i=1}^{n_j} Z_{itj}}{n_j}$

$O_{ijt} = 1$ 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.

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

$\overline{H}_{jt}$ = homeownership rate in neighborhood j at time t. This is calculated in the same way as $\overline{Z}_{jt}$. Note that for purposes of this calculation $\overline{H}_{jt}$ 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 $\overline{O}$ and $\overline{H}$ 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 $\overline{H}$ 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_{ijt} and R_{ijt} 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":

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

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 . 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 $O = 1$) :

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

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

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^*:

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

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

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

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

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

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

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

and for rents:

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

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

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:

$$\begin{aligned}\ln L &= \sum_{O=0} \ln[1-\Phi(-F_{ijt}\rho)] + \sum_{(O=1,H=0)} \ln[\Phi_2(-F_{ijt}\rho, K_{ijt}\omega, \sigma_{uw}) + \\ &\sum_{(O=1,H=1)} \ln[\Phi_2(-F_{ijt}\rho, -K_{ijt}\omega, \sigma_{uw})] \ (10)\end{aligned}$$

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:

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

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

where g=$\tau + \alpha$ 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:

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

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

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

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

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 nj endogenous binary regressors under the constraint that they have the same coefficient ($\frac{\delta}{n_j}$). 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:

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

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

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

$$\tau_{jt} = \tau_j + \widetilde{\tau}_{jt} \quad (20)$$

where τ_j is a standard time-invariant fixed effect and

$$E(\widetilde{\tau}_{jt}) = 0 \;\forall\; t;\;\; E(\widetilde{\tau}_{jt}\widetilde{\tau}_{jt-1}) = 0 \;(21)$$

The idea is that τ_j 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 1.2, 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.

1.4 Specification and Estimation

Tables 1.5 and 1.6 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 . By this measure shopping is largely seen as adequate, and public services not so.

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.

Tables 1.7 and 1.8 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 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, Mi is the quadratic distance from the average:

$$M_i = (A_i - \overline{A})' Cov(A)^{-1}(A_i - \overline{A}) \ (22)$$

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 .

Table 1.9 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 1.10 present the estimates using both fixed (neighborhood) effects and instrumental variables, treating both the ownership and occupancy rate and endogenous. 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.

1.5 Back of the Envelope Calculations

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 1.10. 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.

1.6 Conclusion

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.

1.7 References

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

Victor Aguirregabiria & Pedro Mira, 2002. "Swapping the Nested Fixed Point Algorithm: A Class of Estimators for Discrete Markov Decision Models," Econometrica, Econometric Society, vol. 70(4), pages 1519-1543, July.

Victor Aguirregabiria & Pedro Mira, 2007. "Sequential Estimation of Dynamic Discrete Games," Econometrica, Econometric Society, vol. 75(1), pages 1-53, 01.

Victor Aguirregabiria, 2009. "A Method for Implementing Counterfactual Experiments in Models with Multiple Equilibria," Working Papers tecipa-381, University of Toronto, Department of Economics.

Aguirregabiria, Victor & Mira, Pedro, 2010. "Dynamic discrete choice structural models: A survey," Journal of Econometrics, Elsevier, vol. 156(1), pages 38-67, May.

Robert Aumann, “Acceptable points in general cooperative n-person games”, in R. D. Luce and A. W. Tucker (eds.), Contributions to the Theory 23 of Games IV, Annals of Mathematics Study 40, 287–324, Princeton University Press, Princeton NJ.

Patrick Bajari & C. Lanier Benkard & Jonathan Levin, 2007. "Estimating Dynamic Models of Imperfect Competition," Econometrica, Econometric Society, vol. 75(5), pages 1331-1370, 09.

Patrick Bajari & Phoebe Chan & Dirk Krueger & Daniel Miller, 2010. "A Dynamic Model of Housing Demand: Estimation and Policy Implications," NBER Working Papers 15955, National Bureau of Economic Research, Inc

Baker, K. and Kaul, B. (2002), Using Multiperiod Variables in the Analysis of Home Improvement Decisions by Homeowners. Real Estate Economics, 30: 551–566.

David Barker and Eric Miller (2009) “Homeownership and Child Welfare” Real Estate Economics, 37, 279-303

Patrick Bayer, Robert McMillan, Alvin Murphy, and Christopher Timmins (2009) “A Dynamic Model of Demand for Houses and Neighborhoods,” mimeo, Duke University.

Kelly Bishop (2008), "A Dynamic Model of Location Choice and Hedonic Valuation", Working Paper Washington University of St. Louis

Thomas P. Boehm & Keith R. Ihlanfeldt, 1986. "The Improvement Expenditures of Urban Homeowners: An Empirical Analysis," Real Estate Economics, American Real Estate and Urban Economics Association, vol. 14(1), pages 48-60.

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

Davis, O.A. and Whinston, A.B. (1961) ‘The economics of renewal’, in Wilson, J.Q. (ed.) Urban renewal: The Record and the Controversy, Cambridge, MA: Harvard University Press 2

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 Economics, 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.

James Heckman(1981b). “The Incidental Parameter Problem and the Problem of Initial Conditions,” in C. Manski and D. McFadden, eds, Structural Analysis of Discrete Data, MIT Press.

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

V. Joseph Hotz and Robert A. Miller, "Conditional Choice Probabilities and the Estimation of Dynamic Models" The Review of Economic Studies Vol. 60, No. 3 (Jul., 1993), pp. 497-529

Ioannides, Yannis M., 2002. "Residential neighborhood effects," Regional Science and Urban Economics, Elsevier, vol. 32(2), pages 145-165, March.

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

Linton, Oliver B. and Srisuma, Sorawoot, Semiparametric Estimation of Markov Decision Processes with Continuous State Space (August 2010). LSE STICERD Research Paper No. EM550

Mendelsohn, Robert, 1977. "Empirical evidence on home improvements," Journal of Urban Economics, Elsevier, vol. 4(4), pages 459-468, October

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

Ariel Pakes, 1991. "Dynamic Structural Models: Problems and Prospects. Mixed Continuous Discrete Controls and Market Interactions," Cowles Foundation Discussion Papers 984, Cowles Foundation for Research in Economics, Yale University.

Martin Pesendorfer & Philipp Schmidt-Dengler, 2008. "Asymptotic Least Squares Estimators for Dynamic Games," Review of Economic Studies, Wiley Blackwell, vol. 75(3), pages 901-928, 07

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

Esteban Rossi-Hansberg & Pierre-Daniel Sarte & Raymond Owens, 2010. "Housing Externalities," Journal of Political Economy, University of Chicago Press, vol. 118(3), pages 485-535, 06.

John Rust (1994) Structural Estimation of Markov Decision Processes in R. Engle and D. McFadden (eds.) Handbook of Econometrics, North Holland: Amsterdam.

John Rust & Christopher Phelan, 1997. "How Social Security and Medicare Affect Retirement Behavior in a World of Incomplete Markets," Econometrica, Econometric Society, vol. 65(4), pages 781-832, July

Srisuma, Sorawoot, Minimum Distance Estimation for a Class of Markov Decision Processes, Working Paper

Charles M. Tiebout, "A Pure Theory of Local Expenditures", The Journal of Political Economy, Vol. 64, No. 5. (Oct., 1956), pp. 416-424.

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

Wolpin, Kenneth I, 1984. "An Estimable Dynamic Stochastic Model of Fertility and Child Mortality," Journal of Political Economy, University of Chicago Press, vol. 92(5), pages 852-74, October.

Jeffrey M. Wooldridge, 2005. "Simple solutions to the initial conditions problem in dynamic, nonlinear panel data models with unobserved heterogeneity," Journal of Applied Econometrics, John Wiley & Sons, Ltd., vol. 20(1), pages 39-54

Ziegert, A. L. (1988), The Demand for Housing Additions: An Empirical Analysis. Real Estate Economics, 16: 479–496

1.8 Tables for Chapter 1

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

Table 1.1. Frequency of Neighborhood Sizes

Occupancy Rate in Cluster	No. of Neighborhoods in 1989	No. 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 1.2. Distribution of Occupancy Rates

Ownership Rate in Cluster	No. of Neighborhoods in 1989	No. of Neighborhoods in 1993
0	9	9
0.09091	1	1
0.1	1	2
0.1111	1	0
0.125	1	0
0.142857	0	1
0.153846	1	0
0.16667	1	0
0.181818	0	1
0.2	2	2
0.2222	0	1
0.25	0	1
0.272727	1	2
0.285714	0	1
0.3	0	1
0.307692	0	0
0.33333	2	2
0.363636	0	2
0.375	1	0
0.4	3	2
0.41667	0	1
0.428571	3	2
0.44444	2	4
0.454545	4	3
0.46667	0	1
0.5	0	1

Table 1.3. Distribution of Ownership Rates

Ownership Rate in Cluster	No. of Neigbhorhoods in 1989	No. of Neighborhoods in 1993
0.53333	1	0
0.538462	0	0
0.545455	8	8
0.55556	3	2
0.571429	2	2
0.58333	0	0
0.6	13	4
0.625	2	2
0.636364	8	7
0.66667	6	7
0.7	11	11
0.714286	1	2
0.727273	14	14
0.75	3	8
0.77778	9	6
0.8	22	16
0.818182	27	23
0.83333	3	4
0.857143	3	3
0.875	5	5
0.88889	6	9
0.9	22	30
0.909091	48	37
0.916667	0	0
1	109	120

Table 1.4. Distribution of Ownership Rates Continued

	rent 1989		rent 1993	
Variable	mean	sd	mean	sd
value = log of owners estimate of value	—	—	—	—
rent = log monthly rent	5.95	0.65	6.06	0.64
zinc = household income	30871.72	22734.27	33985.39	27462.27
race = 1 if white	0.74	0.44	0.71	0.45
sch = 1 if schools are "adequate"	0.33	0.47	0.41	0.49
child = number of children	1.07	1.38	1.16	1.38
pubtrans = 1 if public trans is "adequate"	0.32	0.47	0.41	0.49
shpadeq = 1 if shopping is "adequate"	0.85	0.36	0.90	0.30
baths = number of bathrooms	1.44	0.67	1.51	0.66
agehouse = age of housing unit	35.19	23.50	37.90	23.52
garage = 1 if attached garage	0.59	0.49	0.60	0.49
airsys = 1 if central A/C	0.25	0.43	0.31	0.46
heatgas = 1 if gas heat	0.67	0.47	0.65	0.48
heatoil = 1 if oil heat	0.07	0.25	0.08	0.26
heatelec = 1 if electric heat	0.19	0.39	0.19	0.40
regionne	0.09	0.22	0.08	0.28
regionmwnc	0.20	0.40	0.19	0.39
regionwest	0.32	0.47	0.34	0.47
=1 for indicated region				
centercity = 1 if in center city of msa	0.41	0.49	0.47	0.50
lot = sq ft of lot	14739.74	40330.27	13910.00	30059.29
unitsf = interior sq ft	1437.56	698.40	1429.27	6.04
mahal = mahalanobis distance	4.29	1.26	4.30	1.02

Table 1.5. Means and Standard Deviations

	owner1989		owner1993	
Variable	mean	sd	mean	sd
value = log of owners estimate of value	11.34	0.80	11.41	0.73
rent = log monthly rent	—	—	—	—
zinc = household income	43090.19	31954.48	46968.33	34224.04
race = 1 if white	0.83	0.38	0.82	0.39
sch = 1 if schools are "adequate"	0.27	0.44	0.25	0.43
child = number of children	0.69	1.12	0.64	1.06
pubtrans = 1 if public trans is "adequate"	0.27	0.44	0.25	0.43
shpadeq = 1 if shopping is "adequate"	0.89	0.31	0.91	0.28
baths = number of bathrooms	1.93	0.87	1.98	0.91
agehouse = age of housing unit	33.26	21.07	35.25	21.21
garage = 1 if attached garage	0.80	0.40	0.81	0.39
airsys = 1 if central A/C	0.47	0.50	0.51	0.50
heatgas = 1 if gas heat	0.67	0.47	0.68	0.47
heatoil = 1 if oil heat	0.12	0.32	0.11	0.32
heatelec = 1 if electric heat	0.17	0.37	0.18	0.38
regionne	0.19	0.40	0.20	0.40
regionmwnc	0.23	0.42	0.23	0.42
regionwest	0.22	0.41	0.22	0.41
=1 for indicated region				
centercity = 1 if in center city of msa	0.37	0.48	0.37	0.48
lot = sq ft of lot	16823.56	34800.86	18252.40	47824.71
unitsf = interior sq ft	1948.54	847.07	1953.84	844.61
mahal = mahalanobis distance	4.19	1.11	4.18	1.21

Table 1.6. Means and Standard Deviations Continued

	OLS		OLS	
	Value	Rent	Value	Rent
homown	0.95	0.23	-0.14	-0.31
	(0.06)	(0.09)	(0.05)	(0.08)
occupied	1.32	1.70	0.24	0.38
	(0.12)	(0.22)	(0.11)	(0.23)
avgincome			1.85E-05	1.66E-05
			(4.99E-07)	(1.47E-06)
white			-0.10	0.10
			(0.03)	(0.08)
schadeq			-0.40	0.08
			(0.05)	(0.10)
shopadeq			0.08	0.25
			(0.05)	(0.12)
trans			0.28	0.11
			(0.03)	(0.08)
baths			0.13	0.10
			(0.01)	(0.04)
unitsf			5.78E-05	5.56E-05
			(1.06E-05)	(3.11E-05)
lot			3.05E-07	-1.27E-06
			(1.82E-07)	(5.20E-07)
garage			0.13	0.09
			(0.02)	(0.04)
airsys			-0.06	0.02
			(0.02)	(0.05)
heatgas			-0.09	0.20
			(0.04)	(0.08)
heatoil			0.20	0.29
			(0.05)	(0.10)
heatelec			-0.05	0.22
			(0.05)	(0.09)
agehouse			-1.57E-03	-1.77E-03
			(4.21E-04)	(8.98E-04)
year			-0.03	-0.01
			(0.00)	(0.01)
_cons	9.12	4.13	12.30	5.18
	(0.11)	(0.19)	(0.36)	(0.87)
N	5681	774	5681	774
r^2	0.08	0.10	0.49	0.45

Table 1.7. Coefficients from specified regressions

	Fixed House Effects		Fixed Neigh. Effects	
	Value	Rent	Value	Rent
homown	0.48	0.30	0.45	0.24
	(0.13)	(0.31)	(0.13)	(0.28)
occupied	-0.02	0.13	-0.10	0.01
	(0.15)	(0.32)	(0.15)	(0.30)
avgincome	1.03E-06	1.04E-05	1.01E-06	1.26E-05
	(1.04E-06)	(4.41E-06)	(1.08E-06)	4.15E-06
white	0.07	-0.23	0.13	0.01
	(0.12)	(0.33)	(0.13)	(0.31)
schadeq	0.02	-0.11	0.00	-0.23
	(0.07)	(0.22)	(0.07)	(0.20)
shopadeq	0.13	0.13	0.14	0.11
	(0.06)	(0.20)	(0.07)	(0.19)
trans	-0.09	-0.15	-0.09	-0.12
	(0.05)	(0.17)	(0.05)	(0.16)
baths	0.03	-0.05	0.10	0.12
	(0.02)	(0.13)	(0.01)	(0.05)
unitsf	-6.26E-05	-1.22E-03	6.58E-05	1.50E-04
	(1.05E-04)	(7.01E-03)	(9.52E-06)	(4.67E-05)
lot	9..89E-08	-9.49E-07	4.68E-07	-5.58E-07
	(3.01E-07)	(3.79E-06)	(1.53E-07)	(7.75E-07)
garage	-0.02	-0.05	0.04	0.04
	(0.04)	(0.11)	(0.02)	(0.05)
airsys	0.03	-0.13	0.07	0.00
	(0.04)	(0.15)	(0.02)	(0.07)
heatgas	0.10	0.03	-0.03	0.04
	(0.05)	(0.11)	(0.04)	(0.08)
heatoil	0.01	0.32	-0.03	0.19
	(0.08)	(0.20)	(0.04)	(0.12)
heatelec	0.06	0.25	0.00	0.21
	(0.06)	(0.13)	(0.04)	(0.09)
agehouse	(omitted)	(omitted)	-6.47E0-04	-1.21E-03
			(4.72E-04)	(1.30E-03)
year	-0.01	-0.01	-0.01	-0.01
	(0.00)	(0.01)	(0.00)	(0.01)
_cons	11.80	7.90	11.70	5.25
	(0.37)	(9.85)	(0.33)	(0.84)
N	5681	774	5681	774
r^2	0.012	0.06	0.05	0.08

Table 1.8. Coefficients from specified regressions Continued

ownership	
homown	2.60
	(0.13)
mahalanob	-9.36E-11
	(4.85E-11)
localmahal	2.01E-11
	(1.20E-11)
tax	5.97E-03
	(8.57E-03)
_cons	-0.946
	(0.264)
status	
occupied	2.38
	(0.245)
mahalanob	-3.70E-12
	(5.58E-11)
localmahal	1.03E-11
	(1.42E-11)
_cons	-0.732
	(0.233)

Table 1.9. Heckman Selection into Occupancy and Ownership

	Fixed Effects	and IV	Fixed Effects, IV	and Selection
	Value	Rent	Value	Rent
homown	0.49	0.68	1.47	-1.93
	(0.20)	(0.35)	(0.65)	(1.68)
occupied	-0.31	-0.42	1.10	-4.81
	(0.24)	(0.39)	(1.46)	(3.05)
avgincome	1.08E-06	6.98E-06	9,53E-07	5.48E-06
	(1.29E-06)	(4.82E-06)	(1.29E-07)	(4.80E-06)
white	-0.09	-0.03	-0.17	0.08
	(0.16)	(0.42)	(0.17)	(0.41)
schadeq	-0.02	-0.49	-0.01	-0.64
	(0.09)	(0.22)	(0.09)	(0.24)
shopadeq	0.28	0.06	0.25	0.32
	(0.10)	(0.25)	(0.10)	(0.27)
trans	-0.13	-0.08	-0.14	-0.02
	(0.06)	(0.20)	(0.06)	(0.19)
baths	0.08	0.15	0.08	0.16
	(0.01)	(0.06)	(0.01)	(0.06)
unitsf	5.41E-05	1.48E-04	5.38E-05	1.44E-04
	(1.24E-05)	(6.36E-05)	(1.24E-05)	(6.24E-05)
lot	2.22E-07	2.26E-07	1.70E-07	1.53E-08
	(2.24E-07)	(1.14E-06)	(4.08E-07)	(1.38E-06)
garage	0.02	0.07	0.02	0.07
	(0.02)	(0.06)	(0.02)	(0.06)
airsys	0.06	0.04	0.06	0.04
	(0.02)	(0.08)	(0.02)	(0.08)
agehouse	4.94E-04	-2.36E-03	4.15E-04	-2.24E-03
	(6.48E-04)	(1.99E-03)	(6.50E-04)	(1.96E-03)
year	-0.02	0.00	-0.02	0.01
	(0.00)	(0.01)	(0.00)	(0.01)
lambdap			1.77	
			(1.91)	
lambda2p			0.42	
			(0.28)	
lambdar				-6.64
				(4.80)
lambda2r				-1.88
				(1.28)
_cons	12.60	5.52	11.10	11.90
	(0.46)	(1.12)	(1.72)	(3.60)
N	3608	447	3608	447
r^2	0.038	0.136	0.039	0.174

Table 1.10. Coefficients of Specified Regressions

	Fixed Effects, IV for all Nbhd Demographics and Selection	
	Value	Rent
homown	1.56	-1.92
	(0.94)	(3.39)
occupied	3.08	-5.29
	(5.57)	(5.59)
avgincome	9.87E-06	1.05E-07
	(7.91E-05)	(8.79E-05)
white	-2.05	0.58
	(2.97)	(5.38)
schadeq	-0.11	-0.64
	(0.60)	(0.24)
shopadeq	0.40	0.24
	(0.77)	(0.83)
trans	-0.24	0.06
	(0.12)	(0.95)
baths	0.08	0.16
	(0.02)	(0.06)
unitsf	5.38E-05	1.45E-04
	(1.27E-05)	(9.20E-05)
lot	2.39E-07	1.20E-07
	(9.81E-07)	(1.76E-06)
garage	0.02	0.07
	(0.04)	(0.06)
airsys	0.06	0.03
	(0.04)	(0.11)
agehouse	4.32E-04	-2.21E-03
	(6.72E-04)	2.53E-03
year	-0.04	0.02
	(0.06)	(0.07)
lambdap	2.91	
	(5.34)	
lambda2p	0.56	
	(0.37)	
lambdar		-7.22
		(7.60)
lambda2r		-1.89
		(2.43)
_cons	11.80	11.70
	(7.13)	(4.19)
N	3608	447
r^2	0.000	0.167

Table 1.11. Coefficients of Specified Regressions Continued

Chapter 2

Analyzing Home Improvement Behavior in a Dynamic Framework

2.1 Introduction

According to the U.S. Census Bureau, in 2007, homeowners spent approximately $135 Billion on repairing and improving the dwellings in which they live. An additional total of about $34 Billion alone is spent on routine maintenance. With such a large amount spent each year on housing, there has been a surprisingly small amount of literature devoted to the study of home improvement activity. Even so, within this body of literature, previous studies have shown that there are certain characteristics of units, homeowners, and neighborhoods that affect the decision to make improvements to one's dwelling. What this paper seeks to do is to take advantage of the panel nature of the publicly available American Housing Survey (AHS) dataset and analyze these decisions in a dynamic framework. For the most part, the current body of literature has only concerned itself with determining the significance of certain tangible variables that are observed in the data. In addition, most of this literature has been conducted in a static framework, with dynamic models appearing only very recently.

As was alluded to in the previous paragraph, what we seek to do in this paper is to answer the question, "How does a homeowner decide to improve upon her own living space?" In the process of doing so, we realize that these improvement decisions are not taken lightly. Home improvements are a costly, time-consuming, and have an (almost) irreversible effect on one's existing structure. As a result, we realize that home improvements are an inherently dynamic decision process. Conducting an improvement has not only an effect on the household's intertemporal utility by improving existing capital, but also on the unit as an asset for future sale. It is with this in mind that we believe home improvement behavior should be analyzed in a dynamic framework, with decisions dependent upon the current conditions as well as expectations on the future.

One also cannot ignore the fact that most homeowners live in communities, even more narrowly, neighborhoods. Tiebout (1956) presented a model in which residents sort themselves into communities that provide for a public good most favorable to those residents, with the moral of the story being that residents of the same community would usually have similar preferences and self-selected themselves to live together. In addition, the residents of the communities opt to provide themselves with optimal provision of public goods. While it is beneficial to examine the decision making process of a random homeowner, we cannot exclude the notion that most homeowners are affected by the decisions of neighboring households, especially if there are externality effects. For example, Rossi-Hansberg et al. (2011) looks at directed subsidies to units in impoverished neighborhoods in Richmond, VA and do indeed find an external effect. Even in the absence of subsidies, the notion of "keeping up with the Joneses" has been extremely prevalent among Americans especially with respect to real estate. Anecdotal evidence suggests just this. During one summer in 2008, the city of Kent, OH experienced a small boom in deck installations. Homeowners were more inclined to build decks when their adjacent neighbors had decks as well (Fredmonsky 2008). In Wilkes-Barre, PA, a resident gladly repairs and repaints his home simply because he shares a bond with his neighbors and he feels a sense of pride since his neighbors maintain their property as well (Biebel 2010). While motives may be altruistic, motivation for housing capital improvements can also be self-serving. McMansions have been thought to be overly pretentious with respect to housing capital and in some instances present a negative externality to the surrounding houses in the neighborhood.. It becomes even worse when neighbors compete with respect to expansions and improvement in lavish displays of wealth which can be seen in wealthy town such as Westchester, NY (Podoshen 2007). In light of these observations, we would like to model capital improvements not only with respect to dynamics, but with respect to neighborhood peer effects as well.

The ultimate goal of this analysis is to be able to conduct ex-ante counterfactuals on fundamental policy changes to the environment. We can look at the effects of subsidization to improvements or changes in property taxes on the actions of the individual homeowner. To accomplish this, we utilize the often-used Conditional Choice Probability (CCP) estimator in a dynamic discrete choice framework in both in a single agent and two-player environment. We find that in single agent model, subsidization increases the propensity to conduct improvements while increased property taxes decrease the propensity to conduct improvements as we would expect. In the two-player game, we find that neighborhood utility from an asymmetric subsidy is at least as good as directed cash subsidies of the same amount.

2.2 Related Literature

Home improvements have not been studied to the same extent as other areas in housing such as new construction, tenure, and mortgages. The earliest known study of housing improvements comes from Robert Mendelsohn in 1977. Mendelsohn (1977) wishes to answer two main questions. The first question he asks is exactly what are the determinants of homeowner improvement expenditure? The second question he asks is: given that improvements are made, does the homeowner hire workers to do the job, or does he do them himself? To answer the first question, Mendelsohn first breaks the improvement process into two distinct parts. In the first part, the homeowner decides whether or not to make a nonzero amount of improvement, and in the second part, given that it is nonzero, the homeowner decides on how much will he spend on those improvements? He uses a logit estimator to model the first decision, and OLS to calculate the second decision using the same home, homeowner, and geographical characteristics for both analyses. The only significant relationship found in this exercise is that homeowners with higher incomes not only have a higher likelihood of making a nonzero expenditure but also would spend more money on improvements. To answer the second question (do-it-yourself versus hiring contractors), Mendelsohn again uses a logit model to determine the factors that influence the decision of hiring outside help. He finds that again income plays an important role where those with higher income tend to hire workers. Also, he finds that age plays a role in that younger homeowners tended to do more of the work themselves. This initial study lacks one main feature in that outside of an indicator of urban or rural status, it does not take into account locational characteristics. A natural question to ask would be whether or not homeowners are willing to invest in improvements if the neighborhood quality is good and vice versa.

Boehm and Ihlanfeldt(1986) attempt to remedy Mendelsohn's (1977) lack of neighborhood effects by incorporating neighborhood characteristics such as public school quality, crime, and clean streets and considering those as additional utility bearing attributes of living in a particular unit. Since neighborhood conditions play an important role in the valuation of housing capital, the analysis of neighborhood quality variables attempts to control for other potential determinants of home improvement behavior. For example, a new family may be more inclined to increase their current housing stock if they happen to live in a neighborhood with superb public schooling. Also in this analysis, Boehm and Ihlanfeldt specify a few new variables that may have an impact on improvement behavior. The first is an instrument that tries to predict the likelihood of moving out within the next two years which is calculated using a logit model with dependent variables such as homeowner characteristics, condition of the unit, and previous number of moves. The addition of this instrument attempts to control for the fact that households that move may be more inclined to renovate (To increase sale value) or less inclined (we're moving anyway so why bother?) The other new variables in Boehm and Ihlanfeldt (1986) are crowding and excess space which are calculated to indicate whether or not there is more than 1 person per room (crowding) or less than 0.4 people per room (excess space). This was created to give a measure of over-crowdedness or undercrowdedness where an overcrowded unit is more likely to make modifications. Also, they include the Dodge Construction Cost Index and the Price of non-housing goods. This paper, however, does not model the sequential decision-making process as in Mendelsohn (1977). Using OLS, Boehm and Ihlanfeldt find that input prices and neighborhood variables are indeed statistically significant. The indication that neighborhood quality matters is one of the driving forces behind this paper.

Returning to the Mendelsohn (1977) framework, Ziegert (1988) limits herself to analyzing only improvements where a room is added to an existing structure. She analyzes random samples of homes from three year periods between 1968 and 1981 and changes in housing and household composition over those three years. Because she uses five three year panels, she is then able to determine whether or not a household member was added or subtracted (change in household composition). Also she creates a variable, HOUSENEED which indicates the difference between the number of rooms a particular unit has and the average of the number of rooms of similar units with similar demographic and geographic characteristics. By using those variables, the literature is pushed more towards a dynamic framework, where household needs are evolving. In her results, Ziegert finds that HOUSENEED and wealth are statistically significant in the initial decision of making a nonzero amount of alterations.

Trying to capitalize on the results of Ziegert (1988), Baker and Kaul (2002) determine that not only are the evolving household needs important, but past spending on improvement activities may be important as well since spending a lot on improvements in the recent past would affect current period spending. Much like Ziegert (1988), Baker and Kaul (2002) limit themselves to examining additions only when a room or structure is added. Using data from two consecutive periods (separated by two years), Baker and Kaul assign variables which indicate changes in household composition, more specifically whether a child or an adult was added or subtracted to the household. They also added previous spending on additions to the model in an effort to see what effect that has on current spending, making this more of a dynamic model than in Ziegert (1988). Baker and Kaul also only carry out the logit estimation of the choice of making any addition. In their results, they find that previous spending and increasing size have positive signs, which gives evidence that decisions are made with consideration given towards past improvements.

As in Boehm and Ihlanfeldt (1986), who show that neighborhood characteristics are important, Ioannides(2002) tries to determine the effect of actions taken by an immediate neighbor on a household's maintenance decisions. Ioannides(2002) is the closest paper in relation the work I present here. Ioannides uses the AHS cluster data (the same data we use in the first essay) to test the hypothesis that the maintenance decisions of one's neighbors affects the maintenance decisions of one's own unit. The Ioannides model presents a very clever way to capture neighborhood effects and dynamics together. By assuming that neighbors in the cluster play Nash Equilibrium strategies in a one-shot game, Ioannides assumes the equilibrium best response function depends on neighbors' equilibrium actions, homeowners socioeconomic characteristics, previous improvements as well as the previous value of the unit. He finds that indeed, social interaction is indeed present among homeowners with a significant and positive effect, indicating that an increase in neighboring maintenance activity increases a unit's maintenance activity implying that homeowners base their decisions on the expected decisions of others in the neighborhood. What Ioannides lacks in this analysis are the expectations of the future. Since housing decisions (construction, tenure, and improvement) are all costly and difficult to change, future expectations should be taken into account for an analysis of home improvement behavior.

The natural gap that appears in this body of literature is a treatment of home improvement behavior in a dynamic, forward looking environment with peer effects. Truly dynamic models are slowly making their way into housing literature. Bajari et al. (2010) attempt to estimate demand for housing in a dynamic setting where agents increase their housing capital by buying homes (not from each other) and house prices, interest rates, income evolve stochastically over time. Bishop (2008) forms a dynamic model of location choice remarking that a static framework would bias willingness-to-pay estimates since homeowners may appear to "overvalue" an amenity in the present since they expect improvements in the future. Bayer et al. (2010) present a dynamic model of housing and neighborhoods and finds that a static model understates the willingness to pay to avoid air pollution and crime. All of these papers use variants of the Hotz and Miller (1993) estimator.

The obvious next step to build upon this dynamic modeling of housing is to involve peer effects. The models listed above are single agent models which lack any interaction between neighbors or actions of agents within the neighborhoods. As mentioned previously, anecdotal evidence points to the notion that neighbors influence each other's decisions and Ioannides (2002) shows that neighborhood effects are significant. Even though Ioannides (2002) places lagged variables in the best response function (lagged house price, lagged improvement expenditures), the game theoretic model is indeed a static model. In this paper, we wish to extend the model by incorporating forward looking behavior. In other words, homeowners' current state of being, expectations on their future state of being, their neighbors' state of being as well as expectations on their neighbors' future state of being will affect the current decision of improving the home.

2.3 A Reduced Form Model of Dynamic Home Improvement With Unobserved Heterogeneity

We first present a model that is a reduced form estimation of current period improvements regressed on lagged improvements as well as other variables while also taking into account unobserved heterogeneity. As far as we know, this is the first attempt at incorporating unobserved heterogeneity into an analysis of home improvement behavior. The most common method of incorporating dynamic effects into a model of home improvement is introducing a lagged dependent variable into the econometric model. The most notable piece of literature in the intersection of home rehabilitation and dynamic panel data models is from Baker and Kaul (2002) who present a discrete choice model in which they estimate the probability of current period (non-zero) spending on improvements over a four-year period on a host of other variables including, most importantly the existence of previous period spending. The model they use follows:

$$y_i = I(y_i^* > 0) where\ y_i^* = \beta x_i + \gamma y_{i1993} + u_i$$

where y_i^* is the latent or unobserved variable in a threshold-crossing model and $I(\cdot)$ is the indicator function. While they cannot observe the latent variable (i.e. the difference in utility between improving or not improving), they do observe the decisions that those households have made. They denote y_i as the binary choice of having made a discretionary expansionary improvement within the time period of

1994-1997, x_i are exogenous variables pertaining to the house and the household, y_{i1993} is the observed binary choice of having conducted discretionary expansionary improvements (in this case, the addition of a room) within the time period of 1992-1993, and u_i are the unobservable characteristics. In fact, y_{i1993} is the key covariate of interest. Baker and Kaul (2002) adopt this lagged dependent variable model in order to test for the presence of dynamic effects: that past behavior can affect current behavior. The rationale is that discretionary expansions to a house usually occurs in phases, and prior improvement behavior may be an indicator of current period expenditures. Using maximum likelihood methods, Baker and Kaul (2002) obtain results that suggest the presence of dynamic effects. However, their estimator seems to be inconsistent, as the model suffers from endogeneity issues. The conclusion that homeowners conduct discretionary partial adjustments in phases, may be weakened by the presence of unobserved heterogeneity. For example, a homeowner may have an (unobserved) expensive taste that makes her more likely to undertake improvements. The homeowner would still have that same (unobserved) expensive taste when she made a decision during the 1992-1993 period. While this is not the only way in which inconsistency can arise, it illustrates the possibility of correlation between $y_{i1993}$and u_i, causing an endogeneity problem, and making the estimator inconsistent. We will try to control for this unobserved heterogeneity by assuming that each household i, has a permanent unobservable effect, c_i which will account for and correlation between the lagged dependent variable and the unobserved random component. We will make two amendments to the model: in this section, estimate the model to account for the endogeneity, and later we try to truly take advantage of the panel data to estimate the parameters of a dynamic discrete choice model.

One major hurdle in the implementation of this type of dynamic discrete choice model is the "initial conditions problem", first examined in Heckman (1981) where in dynamic panel data models, treating the lagged dependent variable as an exogenous independent variable gives rise to an inconsistent estimator. The main concern stems from the fact that most panels of data have a starting point not included in the data. Using a lagged dependent variable as a covariate poses a problem when dealing with the initial lag. In the model specified above, even that initial observation depends on the unobserved previous period lag, other exogenous variables and unobservables. Treating the initial lag as exogenous poses consistency problems. Even if the econometrician observes all lags, Wooldridge (2005) notes that we cannot assume that the initial lag dependent variable is independent of unobservables. For example, he gives the example of the initial wage after graduating college. Even though the initial wage is observed, it will still be correlated with unobservables since the initial post-graduation wage will undoubtedly be correlated with unobserved individual attributes such as innate ability. While the model in Baker and Kaul (2002) is not exactly the same model as ours, one can still see that this model suffers from endogeneity in using y_{i1993} as an explanatory variable. I will now describe a dynamic panel data model while also correcting for the initial conditions problem using a method described in Wooldridge (2005).

2.4 Data

The US Census Bureau conducts a biennial survey called the American Housing Survey(AHS) National Sample. Every two years, the AHS selects around 50,000 households to interview and has those households answer a series of questions. This survey contains a lot of data and contain much of what we need to conduct this experiment. Included in the survey are data about the household including household income, race and gender of the primary homeowner, education obtained, etc. There is also data on the unit itself and most importantly data that deal with home improvement activities. Luckily though, many of these houses (not households) are selected to be reinterviewed for many consecutive periods, in effect creating a panel (of units) for use. Of these 50,000 households, I must eliminate certain observations for this exercise. First, I keep only those units which are interviewed every two years from 1997-2003, so each unit will have observations from 4 periods. There does not seem to be any bias as to which units are included or excluded from consecutive surveys. Second, I must narrow down the dwellings to include only single detached units occupied by the homeowner. This distinction is made to filter out units in which occupants have a relatively harder time remodeling (apartments, duplexes, and mobile homes), and also observations in which owners do not live in the unit themselves (investment properties). One can therefore see this analysis as one focused on homeowners who live in single detached units.

Additionally, we will consider only observations where occupants have answered a certain set of key questions that contain variables essential to our analysis. Finally, it is important to note that there is considerable turnover in these units in that the ownership of these units change hands throughout time. I do not restrict the analysis to only units that have had the same owner throughout the panel.

Previous literature has shown that characteristics of the homeowner (especially income) are important determinants of conducting improvements (Mendelsohn,1977). For this analysis, we use a specific set of controls for the homeowner (some of which have been modified from the AHS dataset) that include income ($zinc$), years lived in unit ($yearslived$), age of the homeowner (age), gender of the homeowner (sex), race of the homeowner (white, black, asian, hispanic, indian), and approximate years of education ($educ$). The literature has also shown that changes in household composition are important, so we will record the number of people living in the household (per) every period. It also stands to reason that characteristics of the unit itself will have an effect on the decision as well. Therefore, we will use variables such as the number of bathrooms and half bathrooms ($baths\ and\ halfb$), bedrooms ($bedrms$), age of the unit ($agehouse$), total square footage of the unit ($unitsf$), lot size (lot), value of the unit ($value$), total number of rooms ($rooms$). Boehm and Ilhanfeldt (1986) show that neighborhood characteristics are important as well, so we include a variable that indicates the region within the US where the unit is located ($region$) and a variable that indicates whether or not the unit is in a rural, urban, or center city area (metro). The AHS does offer a variable that indicates the Consolidated Metropolitan Statistical Area (CMSA) or the Standard Metropolitan Statistical Area (SMSA) however, the public use file censors many of the responses citing confidentiality issues so this variable is not included. An important additional control that we include is the binary indicator of a turnover. Since we are analyzing units which may or may not have experienced turnover during the length of the panel, the binary indicator of a turnover controls for the notion that improvements may occur more frequently during a period of turnover. Either the exiting homeowner fixes up the unit before selling, or the incoming homeowner does so after buying. While it may be a strong assumption to assume this to be an exogenous variable, it is not at all unreasonable. Since we are dealing with owner-occupied properties and not investment properties, turnover is more likely to be due to a random shock such as relocation because of a new job as opposed to turnover for profitability. Another concern is that some of these houses may be subject to flipping, which are subject to more renovation than a normal unit. However, since we restrict ourselves to properties that a reported as owner-occupied in all periods, the "flipper" will have to buy and sell the unit in-between survey periods. In addition the work done by these "flippers" would not be reported since they would not have been interviewed. If anything, the unobserved presence of house flippers will only bias the coefficient on turnover downwards. We also will include time (year) dummy variables in the estimation which will account for variations in market conditions from period to period.

Summary statistics are provided for the variables to be used as well as a summary of household composition changes between each period (Tables 2.1, 2.2, 2.3, 2.4). The dependent variable is the binary decision of making an expansionary and substantial improvement during the current period. In other words, the improvement done in 2001 are the improvements carried out between 2001-2003. Exogenous variables include homeowner, household, and housing characteristics which have been standard in the literature. For the homeowner this includes age, race, marital status (married), and education level (educ). For the household, we have income (zinc), and number of residents (per). For the housing unit we have the square footage of land, unit square footage, the number of rooms, and the number of bathrooms.

2.5 Estimation

The main question of this section is: Does previous improvement affect the current decision of improvement? While controlling for changing household characteristics, we will follow the previous literature in analyzing a reduced form econometric model of the discrete choice of either making or not making an improvement with the addition of unobserved heterogeneity.

We begin with a dynamic discrete choice model in which we observe an action, but we cannot observe the underlying utility. While we do not specify the form of the homeowner's utility function, we can see that the homeowner's observed action reveals her preference. In other words, in the current time period, if the homeowner has made an improvement, the (future expected) utility from making that improvement is greater than her (future expected) utility from not making an improvement. If we let x_{it} be the exogenous variables(characteristics of the homeowner, household, and home in time period t), y_{it-1} be the homeowner's decision on improvement observed in the previous period, and c_i be some individual random effect that does not vary with time. The literature has long established that household demographic characteristics as well as characteristics of the unit itself play a role in home improvement decisions so we make sure to include those covariates.

The Initial Conditions Problem: Found in many nonlinear dynamic panel data models, the initial conditions problem poses difficulty in obtaining consistent parameter estimates. As described in Heckman (1981), the problem lies in that the "initial condition" or the first lag used on the RHS of the regression, called y_{i0}, is not exogenously determined. Assuming that the initial lag is exogenously determined is an issue since it is inconsistent, especially in our scenario to assume that for y_{i0} but that the future y_{it}, $t > 0$ are not. Wooldridge (2005) proposes a method for estimating a dynamic probit model.

Consider the modified model of what we have above:

$$y_{it} = 1(y^*_{it} > 0) \ where \ y^*_{it} = x_{it}\beta_1 + \gamma y_{it-1} + c_i + u_{it}$$

x_{it} are exogenous variables that vary over time and we assume strict exogeneity. These variables include household demographic characteristics as well as unit characteristics and year dummy variables.

y_{it-1} is the lagged indicator of an expansionary project in the previous period.

c_i is an unobserved effect.

Also let $x_i = [x'_{i1} \ x'_{i2} ... x'_{iT}]$ be all of the exogenous covariates in all time periods minus the (redundant) time-invariant covariates.

Since this is a nonlinear model, we cannot use first differencing to difference out the unobserved effect and we also cannot treat the initial condition y_{i0} as exogenous since it is not independent of c_i. Naturally, the initial lag y_{i0} which are expansionary projects in the first period of observation will be correlated with the unobserved effect. Wooldridge (2005) introduces a method that approximates the distribution of the unobserved effect, c_i conditional on the initial condition y_{i0} and x_i which is the set of all covariates in all time periods. Here, we are assuming that the mean of the homeowner's unobserved effect is based on the initial lag as well as other exogenous variables in all periods, which implies that the initial expenditure is one of the main indicators of the specific unobserved household characteristics. In other words, positive improvements in the first period may indicate that this household has a preference to carry out constant improvements. While previous literature has tried to approximate the conditional distribution of y_{i0} (Heckman 1981), Wooldridge(2005) proposes a somewhat simpler method that requires a small set of assumptions.

We first assume that we have correctly specified the model in that the conditional density of y_{it} contains the correct dynamics (at most one lag) and that the x_i are strictly exogenous conditional on c_i:

$$f(y_{it}|x_{it}, z_i, y_{it-1}, c_i) = f(y_{it}|x_i, y_{it-1}, y_{it-2}, ..., y_{i0}, c_i)$$

We next assume that we have correctly specified that the conditional distribution above is normal (probit model)

We finally assume that the density $h(c_i|y_{i0}, x_i)$ is normally distributed, i.e., $c_i|y_{i0}, x_i \tilde{} Normal(\alpha_0 + \alpha_1 y_{i0} + x_i\alpha_2, \sigma_a^2)$. Then the random variable, $c_i = \alpha_0 + \alpha_1 y_{i0} + x_i\alpha_2 + a_i$ where $a_i|y_{i0}, x_i \tilde{} Normal(0, \sigma_a^2)$

Thus, the latent equation becomes $y_{it}^* = x_{it}\beta + \gamma y_{it-1} + \alpha_0 + \alpha_1 y_{i0} + x_i\alpha_2 + a_i + u_{it}$ and a_i is a random effect. By specifying the density of the unobserved effect, c_i which we assume to be our source of unobserved heterogeneity, conditional on the initial lag and all other exogenous variables, what effectively remains is an uncorrelated random effect a_i and thus ameliorates the issue with the estimation strategy in Baker and Kaul (2002).

Here we will use the *xtprobit* command in STATA that yields estimates for a traditional random effects probit model.

Note that there may be an identification problem with the time-invariant variables that belong in x_{it} since $c_i = \alpha_0 + \alpha_1 y_{i0} + x_i\alpha_2 + a_i$ and there exist time-constant covariates in x_i as well. However we are more concerned here with analyzing the importance or even existence of dynamic behavior in home improvement activity, thus the parameter of interest is still identified.

2.6 Results

The results of this exercise are given in Table 2.5. When analyzing the results of the dynamic model, we see that lagged improvement expenditures yields a small, but positive and significant coefficient, implying that if a homeowner had carried out improvements in the previous period, there is a slightly higher probability that she will carry out improvements this period. We calculate the marginal of effect of previous period improvement evaluated when the other covariates are set to zero, as is done in Baker and Kaul (2002). We find that carrying out improvements in the previous period increases the probability of improvement in the current period by 2.5619×10^{-7}, which is a relatively small increase. Another covariate of some importance is again the homeowner's self reported measure of satisfaction with the neighborhood, *hown*. A homeowner is more likely to carry out improvements if she has high satisfaction with her neighborhood. A similar variable, *howh* gives the homeowner's satisfaction with their own housing capital. Naturally if the homeowner is very satisfied with her own capital, she will be less likely to undertake improvements. One drawback here is that since many covariates across time are used as instruments in this specification, it is hard to identify the effects of objects that are unchanging throughout the panel or are relatively constant throughout the panel. What is most important here, however, is that we have tried to alleviate the initial conditions problem to produce consistent estimates of the parameters of interest. The parameter of most interest, the lagged improvement expenditures is both positive and significant. We compare our result to that found in Baker and Kaul (2002) although their result from their logit model is much larger in magnitude where they find a 73% chance of improvement in the current period given there was important based on their results, which is significantly larger than what we have found. Therefore, in the presence of unobserved heterogeneity, the probability that a homeowner will improve on a unit, given that she made an improvement in the previous period is extremely small and is far different than previous results from Baker and Kaul (2002). We have also shown that there are some statistically significant household characteristics and housing characteristics that affect the decision to undertake improvements as well as to show that previous improvements are also a causal factor as well. While this analysis can answer questions such as how an increase in household size would affect the propensity to undertake improvements, we cannot ask questions for example on how a home improvement subsidy would affect homeowners' behavior. To be able to answer questions like that, we first try to model home improvement behavior as a Markov decision process.

2.7 Home Improvement as a Markov Decision Process

Since we have now some preliminary evidence to suggest that agents behave in a dynamic fashion with respect to home improvement behavior or at least show some linkage between past behavior and present behavior, the logical next step is to construct a structural model of home improvement behavior. Let us first break down the household's problem. The household has a certain set of characteristics such as income, family size, the value of their home, the mortgage rate, the property tax rate, etc. We would expect that the household has certain expectations on how these characteristics would evolve into the future. For example, the household may put some implicit probability distribution on next period's income based on the current period income. If the probability is high that income will increase, it could be possible that home improvement expenditures increase as well. We can make a similar argument for any other time-varying characteristics as well. If family size is expected to increase, or indeed does increase, that could also lead to home expansions and improvements. It is with this notion, that characteristics and needs of the household are every changing, and therefore, based on these changes, or even expectations of these changes, the improvement decision is greatly affected.

In order to carry out our analysis, we rely on the dynamic optimization of a Markov decision process (MDP), which has long been used as a work horse in industrial organization and labor economics, but has not received as much usage in housing models. While capital investment models have long been commonplace in economics, this is to our knowledge the first attempt at modeling home rehabilitation as an MDP (which in effect is an investment model itself). There are many concerns when dealing with a MDP in both a single agent setting as well as a multi-player environment (game). As outlined in Rust (1994) as well as Pakes (1991), with the exception of a certain class of problems for which one can derive the stochastic Euler equation (first order necessary condition), techniques for the most part rely on approximation methods to approach these problems. Rust(1987) provides the framework for the analysis of a single agent dynamic discrete choice model. Hotz and Miller (1993) provide a great reduction to the computational complexity of Rust (1987) with the introduction of the conditional choice probability (CCP) estimator (which will be our workhorse) that provides a link between choice specific value functions and choice probabilities. While these discrete choice techniques are currently being used in industrial organization and in labor economics, less work has been done concerning continuous choice models and continuous state spaces. Srisuma (2010) provides a framework for single agent optimization with a continuous control and a discretized state space while Srisuma and Linton (2010) tackles the problem of a continuous state space and with discrete choice. Bajari, Benkard, and Levin (2007) look at continuous choice in an oligopoly (game) setting and Srisuma (2010) looks at continuous choice in a single agent optimization with a discrete state space.

In a similar vein to the papers that applied these techniques we also wish to model home improvement behavior as an MDP. Doing so will provide us with a vehicle for counterfactual analysis. As an initial step, we first apply a single agent dynamic discrete choice model to home improvement where homeowners pick a level of improvement, and conditional on that choice and the current state (characteristics) of the household, the next period state variables (and payoff) evolve in a stochastic manner (which we recover from the data). For each single agent model we present, we also present an analogous two-player model, which has the same utility structure, with the exception that an agent in this model also makes improvements with respect to the state and anticipated actions of her opponent. In other words, the state space in the single agent model is the agent's own state, whereas in the two-player model, the state space is the agent's own state *and* her neighbor's state. After we recover primitive utility parameters, we can perform counterfactual exercises such as examining the effect of subsidies or increases in taxation.

2.8 A Single Agent Model

We first begin with a simple model of home improvement behavior. As opposed to the previous reduced form portion, one of the key restrictions here is that we must base our analysis on a smaller subset of characteristics. In effect, we must assume that the decision is based on a few key characteristics that are assumed to be completely representative of the homeowner, since increasing the state space becomes quickly infeasible. We assume that an agent derives utility in three ways: from her unit (divided by the number of household members), the increase in value to her unit between periods, as well as non-housing consumption which we define to be the biennial household income less the amount spent on improvement minus property taxes and mortgage interest payments. Note that we do not allow for savings or borrowing such as home equity loans.

Let there be a total of I individuals who maximize over an unknown time horizon of T time periods

Let $x_{it} \in X$ be the observable state variables (current housing stock (price) h_{it}, the previous period housing stock (price) h_{it-1}, household members m_{it}, current household income Y_{it}, property tax rate τ_{it}, and mortgage interest rate r_i). Let a_{it} be the action at time t taken by individual i $a_{it} \in A$, where A is the set of possible actions.

$$\begin{aligned} u(x_{it}, a_{it}) &= \theta_1 \log(0.15 h_{it}/m_{it}) + \theta_2 \log(h_{it} - h_{it-1}) + \\ & \quad \theta_3 \log(Y_{it} - p_{ait} a_{\,it} - (\tau_{it} + r_i) h_{it}) + \varepsilon_{it}(a_{it}) \end{aligned}$$

ε_{it} is an iid shock.

p_{ait} is the cost to a particular improvement level.

A is the set of actions, of which the homeowner can choose $A = \{\$0, \$2500, \$10,000\}$. These levels signify the improvement options of none, moderate, and major respectively.

θ_1, θ_2 and θ_3 are the utility parameters to be estimated.

$x_{it} \in X$ are the state variables (both endogenous and exogenous). Here h_{it} is the housing stock, m_{it} is the number of household members, h_{it-1} is the previous period house value, τ_{it} is the property tax rate and r is the mortgage interest rate. The first utility term gives an indication of the flow utility from housing consumption. We multiply the value of housing by 0.15 to denote housing flow. There is literature to support approximating the flow value of utility from housing at 7.5% of the unit value (Bajari et al. (2010)). We also divide the flow of capital stock by the number of household members. This is to represent the fact that most improvements are done when there is a change in household size, as well as the fact that household capital is a congestible commodity. The second term gives the utility to the payoff of an increase in house value from the previous period. This term represents the fact that improvements can also be done for investment purposes. Since we are assuming (for now) that the homeowner lives in that unit forever, this term is a way of capturing the utility from capital gains. If we had allowed for the homeowner to be able to sell the property in some future period, the decision would be to sell if the price of the unit exceeds the future expected utility. If the homeowner decides to sell, then she would derive utility from the sale price which would include any capital gains. The third utility term is non-housing consumption which is biennial income less the improvement, less the property tax and mortgage interest payments. We assume that the household exhausts its entire remaining income without the option to save for the future. We let the state variables $h_{it}, m_{it}\ and\ \tau_{it}$ vary, however we let r_i be constant. Since we are assuming that the homeowners live in their houses forever and that there is no mortgage refinancing, we assume that the initial reported mortgage interest rate, r_i does not change. The other state variables, however, are time varying and we will go into more detail about how they vary later.

From the point of view of the homeowner and the utility function we have specified, there are now two ways in which utility increases from an increase in house value. She gains utility not only from just having more valuable capital to use in that specific time period, but also that the house has increased in asset value. Note that this does not imply she receives cash for the increase in value (or else it would have been added to income), but simply receives utility from seeing it increase in value. The addition of the property tax can also generate different effects. Higher property taxes may prevent homeowners from undertaking improvements since if the house value increases, higher property taxes would have to be paid. However, higher taxes may also lead to public good increases.

2.9 Value Function

The standard assumptions for dynamic discrete choice models (Aguirregabiria and Mira (2010)):

I. Additive Separability: The unobserved state variable (shock) is additively separable from the rest of the utility function. i.e., $u(x_{it}, a_{it}) = \widetilde{u}(x_{it}, a_{it}) + \varepsilon_{it}(a_{it})$, where $\widetilde{u}(x_{it}, a_{it}) = \theta_1 \log(0.15 h_{it}/m_{it}) + \theta_2 \log(h_{it} - h_{it-1}) + \theta_3 \log(Y_{it} - p_{ait} a_{\,it} - (\tau_{it} + r) h_{it})$

II. Conditional Independence: ε_{it} is independent of the transition probabilities of the observed state variables and (as mentioned before ε_{it} is iid. Thus, the transition probability, $f(x_{it+1}|x_{it}, a_{it}, \varepsilon_{it}) = f(x_{it+1}|x_{it}, a_{it})$

III. Independent Private Values: $\varepsilon_{it}(a_{it})$ is independently distributed across players and alternatives. They are also distributed Type I extreme value, $\varepsilon_{it}\tilde{}G(\varepsilon_{it})$.

These assumptions are necessary to ensure that we have a closed form solution for the conditional choice probability estimator. These assumptions generate a basic multinomial logit framework which is computationally feasible.

For a given agent, i, we can write the value function (the sum of the total expected future stream of utility) as:

$$\max_{a_i^* = (a_{i1}^*, .., a_{iT}^*)} V(x_{it}) = u(x_{it}, a_{it}^*) + \varepsilon_{it}(a_{it}^*) + \sum_{\tau=t}^{T} \beta^{\tau} (u(x_{i\tau}, a_{i\tau}^*) + \varepsilon_{i\tau}(a_{i\tau}^*))$$

This is a standard dynamic programming problem. $V(x_{it})$ is the maximum utility that agent i can achieve, in that the optimal individual policy a_i^* gives rise to $V(x_{it})$. The first term on the RHS represents the current period utility given the optimal action that will maximize future expected utility. The second term is the i.i.d. error. The third term represents the discounted sum of all future utilities given the optimal sequence of actions into the future. One of the keys to the estimation will be to find a closed form solution for $\sum_{\tau=t}^{T}\beta^{\tau}(u(x_{i\tau},a^*_{i\tau})+\varepsilon_{i\tau}(a^*_{i\tau}))$. When only dealing with single-agent optimization, the single action a^*_i maximizes $V(x_{it})$ at each t, the expected future discounted utility under the assumption that expectations are rational. What is missing from the value function is that the state variables evolve in a stochastic manner every period via a Markov process. In other words, how does the action taken, a_{it} affect x_{it}?

After an action is chosen in a period, the state variables, x_t evolve via a Markov process given by the probability distribution $f(x_{t+1}|x_t,a_t)$. For example, consider a homeowner with an annual income of $\$100,000$ living in a $\$100,000$ house with one other person (two people total), paying 2% in property taxes per year. If she chooses no improvement, then next period's state variables: income, house value, household size, and property taxes, will evolve in a certain manner based on the distribution, $f(x_{t+1}|x_t,a_t)$. The state variables will evolve differently if a different action (moderate or major improvement) is taken. We assume that $f(x_{t+1}|x_t,a_t)$ is known by all of the agents and is the same for all agents. We are able to recover these probability distributions from the data and we assume that those are also the beliefs of the agents who carry out the optimization. This means that if we find in the data that a homeowner in the previously mentioned hypothesized state will have a 1/3 probability of having a house valued at $\$90,000$ if she takes no action, then this is the same belief that she has when making her optimal choice.

Given this known (recovered) probability distribution, we can rewrite the value function as

$$\begin{aligned} V(x_{it}) &= u(x_{it},a^*_{it})+\varepsilon_{it}(a^*_{it})+ \\ &\quad \beta\int\sum_{x_{it+1}}V_{t+1}(x_{it+1},a^*_{it+1})f(x_{it+1}|x_{it},a^*_{it})dG(\varepsilon_{it+1}) \end{aligned}$$

Here, we see that the third term on the RHS shows that the future period value function is weighted by the probability of being in a certain state x_{it+1} conditional on being in state x_{it} and having taken action a^*_{it}

2.10 State Space

We denote all individual level state variables, $(h_{it-1}, h_{it}, m_{it}, Y_{it}, \tau_{it}, r_i) = x_{it} \in X$. When a homeowner makes a decision, the individual state variables are sufficient to summarize the observable information available to the homeowner and is the vector of her own current housing stock, household size, and current period household income, current property tax rates, and the individual mortgage interest rate. We assume that house value h_{it} evolves endogenously via the action taken in the previous period. This is a key assumption in that this is the mechanism by which the homeowner's action affects her utility. We assume that the agent commits to a course of action and pays for it in the current period, but does not see the payoff to her investment until the next period (since it takes some adjustment time for the capital to be built). For example, if a homeowner decides to carry out a major $\$10,000$ investment, she will pay for that in the current period, and in the next period, there will be a realized value of the house, and the probability distribution of that realized house value is based on $f(x_{it+1}|x_{it}, a_{it})$. The other state variables will also follow a similar process, although the other state variables do not have to be dependent on the action taken. For example, household income should not depend on the action taken in the previous period, nor should household size. We will determine the transition probabilities empirically with a more detailed explanation below. We also do not model depreciation since we are taking house values at every period and that these updated house values should reflect the state of the housing stock net of any improvements or depreciation.

Another issue that arises is the fact that we must discretize the state space (as well as the action space).[1] As mentioned before, the action space contains three discreet actions, none, moderate and major, while we must also discretize the state space, which aids in the estimation of this model. The state variable that we are most concerned with is that of house value, which we will discretize into 10 different states in increments of \$30,000. This will be the finest discretization that we make. For household size, we have three states: a single person, two people, and a family (more than two people). In other words, the household state is the same whether there are 3 household members or 10 household members. For income, we have 3

---

[1] Rust and Phelan (1997) also discretize the variables in their analysis of social security, health-care, and labor decisions; choices and states which are naturally continuous.

different states in increments of $75,000. For property taxes, we have two states, one in which there are property taxes below 2% and one where there are property taxes above 2%. For the mortgage rate, there are two states: a positive mortgage rate (which we set to be 7%), and no mortgage at all. The result is that there are 3600 possible states, however we only see 1461 states being reached by at least one household in the data. One issue here is that given the data, over 2000 specific states are never reached by the data. In addition there may also be an issue where there are very few observations for certain states. For example if there is only one observation for a state, $x \in X$, the transition we see implies that if this state transitions to $x' \in X$, then in the future, any unit that ends up in state x will always transition to x' no matter what, which seems implausible. While this may affect the estimates of the transition and choice probabilities, in the future we will apply smoothing techniques to approximate continuous transition and choice probabilities which will hopefully ameliorate this issue.

2.11 Rewriting the Value Function

Until very recently, the estimation of dynamic models was computationally very challenging. With the CCP estimator pioneered by Hotz and Miller (1993), much of the computational burden has been alleviated. We make the assumption that the homeowner has three discrete choices of improvement, none, moderate, major. This leads us to take on a dynamic discrete choice framework, since we have modeled the homeowner to be limited to three choices. This model is similar to literature in industrial organization regarding firm behavior in which a retail firm decides how many outlets to open up in a certain geographical area. Other dynamic discrete choice models were concerned with fertility choice (Wolpin 1984), bus engine replacement (Rust (1987)), and retirement choice (Rust and Phelan (1997). More recently in the housing literature, Bishop (2008) has used a dynamic discrete choice framework to model location choice and as mentioned previous, Baker and Kaul (2002) analyze the addition of a new room.

Since our utility function is linear in parameters, we can rewrite the expression $\theta_1 \log(0.15 h_{it}/m_{it}) + \theta_2 \log(h_{it} - h_{it-1}) + \theta_3 \log(Y_{it} - p_{ait} a_{it} - (\tau_{it} + r) h_{it}) = z(x_{it}, a_{it})'\theta$ where $z(x_{it}, a_{it})$ is the column vector $[\log(0.15 h_{it}/m_{it}) \;\; \log(h_{it} - h_{it-1}) \;\; \log(Y_{it} -$

$p_{ait}a_{\,it} - (\tau_{it} + r)h_{it})]'$ and θ is the vector $[\theta_1\ \theta_2\ \theta_3]'$. Following the method summarized in Aguirregabiria and Mira (2010), we wish to look at choice specific value functions. In other words: $v(x_{it}, a_{it}, \theta)$ which summarizes the future expected utility given that the homeowner starts with state variables, x_{it} and chooses specific action a_{it} at time t. We would like to calculate the choice specific value functions for all choices.

Since we assume the shock is additively separable (Assumption I) and independent over time, we can let $v(x_{it}, a_{it}, \theta) = \widetilde{z}(x_{it}, a_{it}, \theta) + \widetilde{e}(x_{it}, a_{it}, \theta)$ where the function $\widetilde{z}(x_{it}, a_{it}, \theta)$ be the discounted sum of expected future $z(\cdot)$ vectors where the starting point is (x_{it}, a_{it}). Similarly, $\widetilde{e}(x_{it}, a_{it}, \theta)$ is the discounted expected sum of the per period shocks, ε_{it}.

If we let $P(a|x)$ be the conditional choice probability of an agent choosing action a when the state is x, by Hotz and Miller (1993),

$$\begin{aligned}\widetilde{z}(x_{it}, a_{it}, \theta) &= z(x_{it}, a_{it}) + \\ &\quad \beta \sum_{x_{it+1} \in X} f(x_{it+1}|x_{it}, a_{it}) \left[\sum_{a \in A} P(a|x_{it+1}) \widetilde{z}(x_{it+1}, a, \theta) \right]\end{aligned}$$

$$\begin{aligned}\widetilde{e}(x_{it}, a_{it}, \theta) &= \beta \sum_{x_{it+1} \in X} f(x_{it+1}|x_{it}, a_{it}) * \\ &\quad \left[\sum_{a \in A} P(a|x_{it+1})(e_{t+1}(a, x_{it+1}) + \widetilde{e}(x_{it+1}, a, \theta)) \right]\end{aligned}$$

$e_{t+1}(a, x_{it+1})$ is the expected value of the shock $\varepsilon_{it+1}(a)$ given that action a is optimal conditional on the state x_{it+1}.Thus, the choice specific value function $v(x_{it}, a_{it}, \theta) = \widetilde{z}(x_{it}, a_{it}, \theta)'\theta + \widetilde{e}(x_{it}, a_{it}, \theta)$

With additive separability (Assumption I), we are able to separate the stream of future expected utilities (without shocks) and the stream of future expected shocks. Assumption II also appears in that the conditional probability distribution of the future state variables, $f(x_{it+1}|x_{it}, a_{it})$ does not depend on $\varepsilon_{it}(a_{it})$.

The first RHS term of $\widetilde{z}(x_{it}, a_{it}, \theta)$ is read as the current utility of choosing action a_{it} given state x_{it}. Since this is a stationary problem (the next period problem is the same as the current period problem up to the state), the choice specific future stream of utility is the same form as that on the LHS. For each of these choice-specific future streams of utility, there is a probability that a certain action $a \in A$ will be chosen by the agent, $P(a|x_{it+1})$, therefore, $\sum_{a \in A} P(a|x_{it+1})\widetilde{z}(x_{it+1}, a, \theta)$ is the expected future stream of utility. Given that the household is in state x_{it} there is a probability distribution over which state the household will enter into in the future, $x_{it+1} \in X$ and that is given by $f(x_{it+1}|x_{it}, a_{it})$ which is a column vector of transition probabilities. Analogous arguments can be made for $\widetilde{e}(x_{it}, a_{it}, \theta)$ which is the future expected stream of i.i.d. shocks.

2.12 Estimation of the Single Agent Model

To estimate this model, we follow the dynamic discrete choice literature pioneered by Hotz and Miller (1993).

Step 1: Discretize the action as well as the state space. We have described how we discretize the state space in a previous section where we get a total of 3600 possible states. We assume that no improvement implies that the homeowner spent nothing on home improvement. If the owner spent between \$0 and \$5000, we call that a moderate or small improvement. If the homeowner spent any amount over \$5000 we call that a large scale improvement. We also carry out a discretization for the state space. We then estimate the transition probabilities of the stochastic state variables. We do this by using a simple frequency estimator, which is similar to that used in Pesendorfer and Schmidt-Dengler (2008). In other words,

$$\widehat{f}(x'|x, a) = \frac{\sum_{I}\sum_{T} 1(x')1(x)1(a)}{\sum_{I}\sum_{T} 1(x)1(a)}$$

where $1(\cdot)$ is the indicator function.

The assumption here is that the observed transition probabilities that we observe in the data are the actual transition probabilities in the world, and also that these transition probabilities are common knowledge for all households. Note that since we assume this is a stationary problem, we can pool time periods. The denominator of the frequency estimator counts the total number of households who chose a particular action $a \in A$ and reside in a particular state $x \in X$. The numerator counts the total number of those who chose a particular action $a \in A$ and reside in a particular state $x \in X$ *and then* transitioned to a particular state $x' \in X$.

Step 2: We estimate the conditional choice probabilities (CCP), $P^0 = P(a|x)$. These CCP estimates give the probability that an agent would choose action a given state x. . This again is estimated using a simple frequency estimator. In other words,

$$\widehat{P}(a|x) = \frac{\sum\limits_{I}\sum\limits_{T} 1(x)1(a)}{\sum\limits_{I}\sum\limits_{T} 1(x)}$$

Similar to the previous frequency estimator, the denominator here is total number of households that were in a particular state $x \in X$ at any given time period. The numerator is the total number of those households who chose a particular action $a \in A$ given that they were in state x.

Step 3: In the above section, the choice specific value function is separated into two parts, $\widetilde{z}(x_{it}, a_{it}, \theta)$ and $\widetilde{e}(x_{it}, a_{it}, \theta)$. The estimation relies on the calculation of these the two expressions.

Following Hotz and Miller (1993), Aguirregabiria and Mira (2002) and Aguirregabiria and Mira (2010), we get closed formed expressions for these expressions which follow a multinomial logit form. Given estimates $\widehat{f}$ and $\widehat{P}$ of f and P and given Assumption III (that errors are i.i.d. Type I extreme error) Aguirregabiria and Mira (2002) provide the pseudo-likelihood function:

$$L(\theta, \widehat{P}, \widehat{f}) = \sum_{I}\sum_{T} \log \frac{\exp(\widetilde{z}(x_{it}, a_{it})'\theta + \widetilde{e}(x_{it}, a_{it},))}{\sum\limits_{a \in A} \exp(\widetilde{z}(x_{it}, a)'\theta + \widetilde{e}(x_{it}, a))}$$

2.13 Data

The data we use again comes from the AHS. See Section 2.3.1 for an explanation of the data. However in this setting, we assume that homeowners have ex-ante identical preferences with respect to their utilities, and we assume that the new homeowners are identical to previous homeowners with respect to preferences, so turnover is left for future work.

With a discrete control, we break improvement levels off into three categories: No improvement, moderate improvement, and major improvement. For our exercises we set no improvements to be \$0, moderate improvements to be between \$0 and \$5000, and anything over \$5000 to be a major improvement. We experimented with different cutoffs and found that there was no significant change in the results. The other variables that we use are household size, household income and house values. We also must discretize these variables and a description of this process is described above. As with all data in the AHS, these are self-reported values, and household income and house value are usually presented rounded to the nearest \$1000 dollars. This also helps since we must discretize the state space as well in order to apply the CCP estimator. While explicit discretization methods exist such as Tauchen (1986), we simply discretize the state space in equal intervals and experimented with interval sizes and state space sizes that would be relatively easy to handle in a third-party optimization program such as MATLAB.

2.14 Summary Statistics

The summary statistics are given in Tables 2.1-2.4. There is also an addendum in Table 2.6, where we give summary statistics for additional variables used for this analysis such as property tax, mortgage interest, and improvement expenditures. Recall that this is the same data that is used in the reduced form analysis. Here we also give some raw statistics for the proportion of those who choose a given action in different time periods (Table 2.7). Throughout the panel, the aggregate improvement level vary little from year to year. No improvement and moderate improvement are more or less equal throughout the panel, with major improvements being less than half of any other type of improvement.

2.15 Results of the Single Agent Model

Using the estimation techniques developed from Hotz and Miller (1993) and Aguirregabiria and Mira (2002), we use the two step estimator with first stage frequency estimates of the choice probabilities as well as the transition probabilities. The second stage involves maximization of a pseudo likelihood function given by the multinomial logit framework given the assumption of Type I Extreme errors. This model was estimated using the optimization packages in MATLAB. These likelihood functions based on the linear-in-parameters utility model were maximized fairly easily given that the size of the state space and the transition matrices were within MATLAB computational boundaries. The results for the single agent model are below. The first parameter reflects the utility of housing stock, the second parameter reflects the utility gained from an increase in house value, and the third parameter reflects the utility of non-housing consumptions.

	Single Agent
θ_1	$-0.0141(0.0063)$
θ_2	$0.0264(0.0087)$
θ_3	$0.0808(0.0255)$
Log-likelihood	-5322.01

The signs for the last two parameters are as we would expect, that there is a positive effect from an increase in house value, and also a positive effect of non-housing expenditures. However, we find a negative value for the parameter corresponding to the flow of housing utility. Even so, since h_{it} (house value) appears in both utility terms connected to θ_1 and θ_2 the net effect of an increase in house value may still be positive. Also, the flow of housing utility is divided by the number of household members which would imply that an increase in household members would have a positive effect on utility, so perhaps household gain utility from having more members.

2.16 The Two-Player Model: A Motivational Example

Consider a simple (albeit abstract) example of a neighborhood with only two residents. Every year they both have an option to either carry out improvements to their homes or not. At some random point in the future, one, or both neighbors would like to sell their housing units off and move away, but that information is unknown to either resident. The value of these units encompass both the physical characteristics of the house as well as the surrounding area (neighborhood), and if improvements are made consistently to both units, the neighborhood as a whole will be considered "nice" and property values will be high. However, if both residents refuse to improve every period, both units will become run-down and will be viewed as "poor" and property values will be low. This is what is known as the Slumlord's Dilemma (Davis and Whinston (1962)), where two neighboring landlords end up in a prisoner's dilemma situation where the dominant strategy is for neither landlord to carryout improvements. The case remains where one resident improves and the other does not. In this case, the resident that does not improve gains benefits by being located next to a "nice" house, whereas the resident who actually did improve bears not only the cost of the improvement but of also living near an unimproved unit. In one shot game, this leads to the static prisoner's dilemma which whose payoffs we normalize below:

	Build	Don't Build
Build	2,2	0,3
Don't Build	3,0	1,1

In the one shot prisoner's dilemma, the dominant strategy for both players is {Don't Build} and therefore have a Nash Equilibrium at {Don't Build, Don't Build}. However, if we assume that this game is repeated over time, with an unknown time horizon, then repeated cooperation as well as repeated defection is a sustainable outcome.(Aumann (1959)). We are able to make the assumption of an unknown time horizon for two reasons. Rarely do homeowners know with perfect foresight exactly when they will move out, and if one homeowner leaves, another will come in and take her place.

Analogous to the previous single agent model, we will take the same utility function. However, decision are now made with respect to the other player's state and anticipated actions.

2.17 A Basic Two-Player Model

In the neighborhood model, a homeowner can change her capital for the same reasons as the single agent model except now, she can also base her decision on her neighbors' states and actions as well. It is possible that neighbors "compete" with one another with respect to building capital. A homeowner may look over to her neighbor and see new construction, or a homeowner may anticipate that the neighbors will undertake a certain action in the present, or in the future and will choose her sequence of actions accordingly. Analyzing neighbors' actions in a dynamic framework gives us the opportunity to capture the fact that it may take time for a homeowner to respond to the actions of their neighbors. In one of the examples alluded to earlier, a homeowner considered building a deck because all of the nearby homes had one, which indicates a response due to neighbors' previous actions. The homeowner may also anticipate future conditions of the neighborhood or their neighborhoods which we would consider forward-looking behavior. In order to capture responses to neighbor's past or anticipated behavior, we will model home improvement behavior as a dynamic game.

As mentioned previously, there is a small subset of past AHS data that allows us to look at "neighborhood" clusters. Ioannides (2002) uses these clusters to investigate a similar problem to ours here. He estimates a reduced form "best response function" in which the action of that period depends on prior improvements, those of the unit's neighbors, as well as various demographic and capital controls. While he is able to determine, ceteris paribus, the change of the best response in light of a marginal change to a control, the approach we take here is to estimate primitive utility parameters within the confines of a dynamic game so that we can conduct counterfactuals on any number of possible different ex ante modifications. His analysis also falls short in taking future expectations into account.

As with the single agent model, we will be able to apply the CCP estimator. As a starting point, as well as for computational simplicity, we break up each neighborhood into pairs, creating at first a two-player game. Thus, we assume that each homeowner is playing this capital adjustment game with only one other homeowner. As usual, every period the homeowner chooses an action that maximizes the expected discounted sum of utility. Also in this model we do not include members of the household as a state variable (although we plan to do so in future work).

Similar to the single agent model, a homeowner still receives intertemporal utility from flow utility from housing, as well as non-housing consumption. The main difference here is that the maximization is carried out with respect to expectations on neighbors' actions as well as neighbors' current period state variables. In this setting, we assume that within each neighborhood, all neighbors' state variables are common knowledge for all. Every period neighbors choose a level of improvement, pay that cost in the current period, and see stochastic returns via capital and land values which depend on the actions taken by the entire neighborhood in the previous period.

2.18 Basic Two-Player Model Utility Function

To make this portion of our analysis more clear, it would be wise to compare the differences between the two-player model and the single agent model. As before, we will let there be a total of I individuals who each live in a neighborhood $j \in J$ and maximize over (as before) an unknown time horizon of T time periods. The j subscript is the newly added subscript which we will use for all members of neighborhood j.

Let $x_{jt} \in X$ be the *all* observable state variables for the neighborhood(current housing value $h_{jt} = \{h_{ijt}, h_{-ijt}\}$ for each member of the neighborhood, household members of the neighborhood $m_{jt} = \{m_{ijt}, m_{-ijt}\}$, property tax rates τ_{jt} (which we assume to be the same for both neighbors) and mortgage interest rates $r_j = \{r_{ij}, r_{-ij}\}$ for each member of the neighborhood, and current household income for each member of the neighborhood $Y_{jt} = \{Y_{ijt}, Y_{-ijt}\}$). For example, a homeowner in a neighborhood knows not only her own house value, income, household size, property tax rates and mortgage rates but also the analogous characteristics of her neighbor. Let a_{it} be the action at time t taken by individual i $a_{it} \in A$. We also note $a_{jt} = \{a_{ijt}, a_{-ijt}\}$ to be the vector of actions taken by each member of neighborhood j at time t.

The intertemporal utility is specified as follows:

$$\begin{aligned} u(x_{jt}, a_{jt}) &= \theta_1 \log(0.15 h_{ijt}/m_{ijt}) + \theta_2 \log(h_{ijt} - h_{ijt-1}) + \\ &\quad \theta_3 \log(Y_{ijt} - p_{aijt} a_{\,ijt} - (\tau_{ijt} + r_{ij}) h_{ijt}) + \varepsilon_{ijt}(a_{ijt}) \end{aligned}$$

ε_{ijt} is an iid shock.

p_{ait} is the cost to a particular improvement level, $p_{ait} \in \{\$0, \$2500, \$10000\}$

x_{jt} is the vector of each homeowner's individual state variables (both endogenous and exogenous) in neighborhood j. Since this is a two-player model, it is natural to wonder how the an opponent's action affects utility. We assume that house value is an endogenous state variable, which evolves dependent on the simultaneous actions taken by both neighbors. In other words, $a_{jt} = \{a_{ijt}, a_{-ijt}\}$ will affect the individual house values h_{ijt} and h_{-ijt}. Note however, that only the homeowner's individual level variables factor into the utility function. The exogenous state variables here are household size, m_{ijt} and income Y_{ijt} and property taxes τ_{it}. Similar to the single agent model, these state variables are allowed to evolve in a pattern independent of the action taken in the previous period. p_{ait} does not evolve at all. While the utility depends only on individual level variables, it is worth reiterating that the transition probabilities of the payoff (next period housing value) depends on the actions taken by the entire neighborhood in the previous period.

2.19 Markov Perfect Equilibrium

In the single agent model, one single household conducted a maximization against nature (state transition probabilities). In the two-player game, actions are taken not only against transition probabilities of both the homeowner and the opponent's state, but also the strategy of the opponent. We therefore have to make an additional assumption to the three that we initially used in the single agent model.

The standard assumptions for dynamic discrete choice models (Aguirregabiria and Mira (2010)):

I. Additive Separability: The unobserved state variable (shock) is additively separable from the rest of the utility function. i.e., $u(x_{ijt}, a_{ijt}) = \widetilde{u}(x_{ijt}, a_{ijt}) +$

$\varepsilon_{ijt}(a_{ijt})$, where $\widetilde{u}(x_{ijt}, a_{ijt}) = \theta_1 \log(0.15h_{ijt}/m_{ijt}) + \theta_2 \log(h_{ijt} - h_{ijt-1}) + \theta_3 \log(Y_{ijt} - p_{aijt}a_{\,ijt} - (\tau_{ijt} + r_{ij})h_{ijt})$

II. Conditional Independence: ε_{it} is independent of the transition probabilities of the observed state variables and (as mentioned before ε_{ijt} is iid. Thus, the transition probability, $f(x_{jt+1}|x_{jt}, a_{jt}, \varepsilon_{ijt}) = f(x_{it+1}|x_{it}, a_{it})$

III. Independent Private Values: $\varepsilon_{ijt}(a_{ijt})$ is independently distributed across players and alternatives. They are also distributed Type I extreme value.

IV. One Markov Perfect Equilibrium: We make a somewhat strong assumption that there is no unobserved heterogeneity in that all neighbor pairs are playing the same equilibrium strategy (the equilibrium strategy we see in the data). In dynamic games, the presence of multiple equilibria is sometimes an issue, therefore, we let the data determine which equilibrium is being played in our model.

Define an individual's strategy function as $\alpha_i(x_j, \varepsilon_{ij})$, which depends on the observed neighborhood state variables as well as the individual level shock. Let $\alpha = \{\alpha_i(x_j, \varepsilon_i), \alpha_{-i}(x_j, \varepsilon_{ij})\}$ be strategy functions for the neighborhood.

For a given agent, i we can write the value function (the sum of the total expected future stream of utility) as:

$$\max_{a_i^* = (a_{i1}^*, .., a_{iT}^*)} V(x_{jt}) = u(x_{jt}, a_{jt}^*) + \varepsilon_{ijt}(a_{ijt}^*) + \sum_{\tau=t}^{\infty} \beta^{\tau}(u(x_{j\tau}, a_{j\tau}^*) + \varepsilon_{i\tau}(a_{i\tau}^*))$$

After actions are chosen by the neighborhood in a certain period, the state variables, x evolve via a Markov process given by the probability distribution $f(x_{jt+1}|x_{jt}, a_{jt})$. We assume that $f(x_{jt+1}|x_{jt}, a_{jt})$ is known by all households and the same for all neighborhoods. The Markov transition probabilities are similar to those of the single agent model. Conditional on the action taken by *both* neighbors, state x_{jt} transitions to state x_{jt+1} via $f(x_{jt+1}|x_{jt}, a_{jt})$, which we will recover empirically.

We can rewrite the value function thus as

$$V(x_{jt}) = u(x_{ijt}, a_{ijt}) + \varepsilon_{it}(a_{ijt}) + \beta \int \sum V(x_{jt+1}) f(x_{jt+1}|x_{jt}, a_{jt}) dG(\varepsilon_{it+1})$$

Here, we see that the second term on the RHS shows that the future period value function is weighted by the probability of being in a certain state x_{jt+1}.

Aguirregabiria and Mira (2010) define a Markov Perfect Equilibrium to be a set of actions, $\alpha^* = \{\alpha_i^*(x_{jt}, \varepsilon_{ijt}), \alpha_{-i}^*(x_{jt}, \varepsilon_{-ijt})\}$ such that

$$\begin{aligned} \alpha_i^*(x_{jt}, \varepsilon_{ijt}) &= \arg\max_{a_{ijt} \in A} E_{\varepsilon_{-ijt}}[u(x_{jt}, a_{ijt}, \alpha_{-i}^*(x_{jt}, \varepsilon_{-ijt})) + \varepsilon_{it}(a_{it})] + \\ & E_{\varepsilon_{-ijt}}[\beta \int \sum V(x_{it+1}) f(x_{it+1} | x_{it}, a_{it}, \alpha_{-i}^*(x_{jt}, \varepsilon_{-ijt})) dG(\varepsilon_{ijt+1})] \end{aligned}$$

Here, we see that the stationary equilibrium strategy of player i, α_i^* is a best response to player $-i$'s equilibrium strategy of α_{-i}^*. In other words, given the equilibrium strategy of her opponent, and the commonly known transition probabilities, the homeowner's actions $a_i^* = (a_{i1}^*, .., a_{iT}^*)$ maximize future expected utility. Notice that player i takes into account the expectation of her opponents shock, the opponents state variables as well as the action α_{-i}^*

As in the single agent model, we can compute choice specific value functions. We assume that all players follow equilibrium strategies α^*. Given that players follow α^*, we assume that observed conditional choice probabilities reflect the equilibrium strategy played. Therefore, we denote the equilibrium conditional choice probability P^{α^*} The choice probabilities are in effect a result of the equilibrium strategy. What we will observe from homeowners are the equilibrium choice probabilities, where homeowners randomize over the three choices. In order to avoid the problem of having multiple equilibria, we assume that the data is generated by one unique Markov perfect equilibrium (Assumption IV). What this means is that for any neighborhood in state $x \in X$ the neighbors will always follow the same equilibrium choice probability. We have also abstained from any local effects or unobserved heterogeneity (which we leave for future work), and assume that if two neighborhoods share the same state, then they are identical. We reiterate that we assume that each player is already playing their equilibrium strategies for this particular equilibrium and this is reflected in the conditional choice probabilities. Following the method in Aguirregabiria and Mira (2010) we can again compute choice specific value functions.

$$\begin{aligned}\widetilde{z}(x_{jt}, a_{ijt}, \theta) &= \prod_{a_{-ijt}\in A} P(a_{-ijt}|x_{jt}) * z(x_{jt}, a_{ijt}) + \\ &\quad \beta \sum_{x_{jt+1}\in X} f(x_{jt+1}|x_{jt}, a_{jt}) \left[\sum_{a_{ijt}\in A} P(a_{ijt}|x_{jt+1})\widetilde{z}(x_{jt+1}, a_{ijt}, \theta)\right]\end{aligned}$$

$$\begin{aligned}\widetilde{e}(x_{jt}, a_{ijt}, \theta) &= \prod_{a_{-ijt}\in A} P(a_{-ijt}|x_{jt}) * \beta \sum_{x_{jt+1}\in X} f(x_{jt+1}|x_{ij}, a_{ij}) * \\ &\quad \left[\sum_{a_{ijt}\in A} P(a_{ijt}|x_{it+1})(e_{t+1}(a_i, x_{jt+1}) + \widetilde{e}(x_{it+1}, a_{ijt}, \theta))\right]\end{aligned}$$

The choice specific value function is: $v(a_{ijt}, x_{jt}, \theta) = \widetilde{z}(x_{jt}, a_{ijt}, \theta)'\theta + \widetilde{e}(x_{jt}, a_{ijt}, \theta)$

The main difference here is that the discounted sums are weighted by the conditional choice probabilities of the opponent. When looking at $\widetilde{z}(x_{jt}, a_{ijt}, \theta)$, the homeowners has knowledge of $P(a_{-ijt}|x_{jt})$ meaning that she knows (conditional on the state of the neighborhood), the probability that her neighbor will choose a certain action. She also has knowledge of the evolution of the neighborhood state variables and how they evolve. She also knows that probability that she herself will choose a certain action given the her state and her neighbors. Therefore, as in the single agent model, we can calculate choice specific utility separately calculating the future expected stream of utility for a given choice and the sum of future expected stream of i.i.d. choice-specific shocks, $\widetilde{e}(x_{jt}, a_{ijt}, \theta)$.

2.20 Estimation of the Two Player Game

As a simplification in order to reduce the state space, we first analyze a game between one unit and one neighbor since the size of the state space exponentially increases with the number of players in the game. We use the same discretization as in the single agent model. To refresh, we will discretize house value into 10 different states in increments of $30,000. This will be the finest discretization that we make. For household size, we have three states: a single person, two people, and a family (more than two people). In other words, the household state is the same whether there are 3 household members or 10 household members. For income, we have 3 different states in increments of \$75,000. For property taxes, we have two states, one in which there are property taxes below 2% and one where there are property taxes above 2%. For the mortgage rate, there are two states: a positive mortgage rate (which we set to be 7%), and no mortgage at all. The result is that there are 3600^2 possible states. Empirically, however, we only see 2580 states being reached by at least one neighborhood in the data. We randomly select pairs out of the neighborhood and assume that these two players are in a neighborhood on their own. We realize that in the data, the neighborhoods are usually around 10 units. However, if we had a 10 player game, the state space would become 3600^{10}, which is extremely hard to deal with. Again, in future work, for the same reasons as in the single agent model, we will, in future work, apply smoothing techniques to estimate the transition and choice probabilities.

Step 1: Estimate the transition probabilities. This is similar to the single agent model, except, now state variables are housing, household size and income for all (both) homeowners within the neighborhood. Also, the endogenous state variables, capital and land, depend on the actions taken by the neighborhood in the previous period.

$$\widehat{f}(x'|x,a) = \frac{\sum\limits_{I}\sum\limits_{T} 1(x')1(x)1(a_i,a_{-i})}{\sum\limits_{I}\sum\limits_{T} 1(x)1(a_i,a_{-i})}$$

where $1(\cdot)$ is the indicator function, x and x' are state vectors consisting of the state variables for all the agents in a neighborhood. a is the vector of actions taken by all agents in the neighborhood. The denominator of the frequency estimator is the total number of neighborhoods who took actions $(a_i,a_{-i}) \in A \times A$ and were in state $x \in X$. The numerator are those who took actions $(a_i,a_{-i}) \in A \times A$ and were in state $x \in X$ and transitioned to state x'. The result is the state transition probability for that particular neighborhood state and actions taken.

Step 2: Estimate the conditional choice probabilities (CCP), $P^0 = P(a_i|x)$. These CCP estimates give the probability that an agent would choose action a_i given state x. Notice first that this is the CCP of the action of a single agent conditional on the neighborhood level state variables. This again is estimated using a simple frequency estimator. In other words,

$$\widehat{P}(a_i|x) = \frac{\sum\limits_{I}\sum\limits_{T} 1(x_i)1(a)}{\sum\limits_{I}\sum\limits_{T} 1(x_i)}$$

The denominator is the total number of neighborhoods in a given state and because of stationarity, we can pool time periods together. The numerator is the total number of households that chose an action $a \in A$, given that they were in state $x \in X$.

Step 3: In the above section, the choice specific value function is separated into two parts, $\widetilde{z}(x_{it}, a_{it}, \theta)$ and $\widetilde{e}(x_{it}, a_{it}, \theta)$. The estimation relies on the calculation of the two expressions.

Following Aguirregabiria and Mira (2002) and Aguirregabiria and Mira (2010), we get closed formed expressions for these expressions. Given estimates $\widehat{f}$ and $\widehat{P}$ of f and P and given Assumption III (that errors are iid Type I extreme error) Since the errors are iid across time and are assumed to be Type I extreme value The best response function is in the multinomial logit form and yields the maximum likelihood function:

$$L(\theta, \widehat{P}, \widehat{f}) = \sum_{J}\sum_{T}\sum_{I_j} \log \frac{\exp(\widetilde{z}(x_{jt}, a_{ijt})'\theta + \widetilde{e}(x_{jt}, a_{ijt}))}{\sum\limits_{a \in A} \exp(\widetilde{z}(x_{jt}, a)'\theta + \widetilde{e}(x_{jt}, a))}$$

2.21 Data and Summary Statistics

The data that we use is the same set of data from Essay 1. We restrict ourselves to the usage of units that are owner-occupied, detached single family units that reside in a known neighborhood "cluster". The summary statistics are in Table 2.8. We also look at the break down of neighborhood pairs and what actions they have taken over time. Pooling all periods together, we see that the highest frequencies are when residents play {none, none} or {none, medium}. By far the lowest frequency is {high, high} leading us to believe that most residents opt not to do improvements. (Table 2.9). This is very similar to the single agent model where homeowners carried out none or medium improvements with far more frequency than a high level of improvement.

2.22 Results of the Two-Player Game

Using the estimation techniques developed from Hotz and Miller (1993) and Aguirregabiria and Mira (2002), we use the two step estimator with first stage frequency estimates of the choice probabilities as well as the transition probabilities. The second stage involves maximization of a pseudo likelihood function given by the multinomial logit framework given the assumption of Type I Extreme errors. Like the single agent model, this was estimated using the optimization packages in MATLAB. These likelihood functions based on the linear-in-parameters utility model were maximized fairly easily given that the size of the state space and the transition matrices were within MATLAB computational boundaries. The results for the two-player game are below. The first parameter reflects the utility of housing stock, and the second parameter reflects the utility of non-housing consumptions. While both are positive and significant, the weight placed on housing consumption is much larger than that of non-housing consumption. This is somewhat strange considering that both are measured in a dollar amounts and the two quantities do not differ greatly in magnitude. Standard errors are in parentheses and were calculated via 100 bootstrap samples.

	Two Player
θ_1	$-0.1308(0.03255)$
θ_2	$0.0036(0.0021)$
θ_3	$0.0888(0.0118)$
Log-likelihood	-4106.85

Here, as in the single agent model, we get a negative and significant result for the parameter on housing consumption. The other signs are what we would expect. Particulary notable is the parameter related to the utility gained from an increase in house price, indicating that homeowners do receive utility from increases in house values. Again, this may be mitigated by the fact that since house value, h_{it} appears in the second utility term, the net impact of housing is positive.

2.23 Policy Experiments

One of the main reasons for carrying out a structural estimation is the opportunity to do counterfactual policy estimations. For these experiments, we rely on the third model and carryout policy changes on subsidies and property taxes. For the first counterfactual policy experiment, we look at the effect of subsidization to home improvement activity. Because of the existence of multiple equilibria in dynamic games, counterfactual experiments prove to be very difficult in the two-player setting. We base our method of looking at policy experiments on the methodology laid out in Aguirregabiria and Mira (2002) who provide a Nested Pseudo Likelihood (NPL) algorithm for dynamic discrete choice problem. The NPL algorithm relies on convergence to a fixed point in the space of choice probabilities, where candidate choice probabilities are updated with the best response function with updated parameter estimates from previously updated choice probabilities. Eventually the parameter estimates and choice probabilities converge to a fixed point with this algorithm. With the presence of multiple equilibria, one cannot be sure which equilibrium arose from the policy change (Aguirregabiria 2009). However, what we can do is apply the "outer loop" portion of the NPL algorithm that updates the conditional choice probabilities. Using the parameters that are recovered, we alter the environment, and then we iterate the best response operator to obtain until a fixed point is achieved. In other words, we iterate:

$$P^K(a|x_m) = \frac{\exp(\widetilde{z}^{P^{K-1}}(x_m, a)'\widehat{\theta} + \widetilde{e}^{P^{K-1}}(x_m, a))}{\sum\limits_{a \in A} \exp(\widetilde{z}^{P^{K-1}}(x_m, a)'\widehat{\theta} + \widetilde{e}^{P^{K-1}}(x_m, a))}$$

where P^{K-1} are the choice probabilities from the previous iteration and the $\widetilde{z}^{P^{K-1}}(x_m, a)'\widehat{\theta} + \widetilde{e}^{P^{K-1}}(x_m, a)$ are calculated with P^{K-1}. Note that we calculate updated CCPs for every state, x_m to find the equilibrium counterfactual choice probabilities. The iteration process is very intuitive. In the two-player model, each iteration updates the best response choice probabilities, each player reacting to successive best responses of her opponent. Once we arrive at a fixed point, neither player would like to change her best response of her opponents strategy, and thus we arrive at a new equilibrium. We can thus view the single agent model as a game against nature or what the homeowner would view as the transition probabilities. She will maximize and choose a strategy that follows the best response form as given above in the multinomial logit framework, until she arrives at a fixed point where she does not want to deviate from what nature has determined to be her state transition probabilities.

2.24 Subsidization and Taxation

We first begin with the single agent model. We first assume that there is a flat \$2000 subsidy for those who decide to make an improvement, regardless of whether it is a minor improvement or a major improvement. In other words, a minor improvement now only costs \$500, and a major improvement now costs only \$8000. To make the analysis more clear, we look at the effect of a \$2000 on all households, but we report in this section specific behavioral changes with certain types of households.

Household 1: Consider a household with 2 members living in a \$45,000 house with an annual income of \$35,000, a 7% mortgage rate and paying 0% in property taxes. Empirically, before the subsidization, we observe that a household in this state carries out no improvements at all ($p = 1$). However, in light of the subsidy, the new equilibrium choice probabilities indicate that this household would randomize between carrying out a moderate level of improvement ($p = 0.5$) and a high level of improvement ($p = 0.5$).

Household 2: We now consider a household with 2 members living in a \$165,000 house with an annual income of \$150,000, no mortgage paying 2% in property taxes. Empirically, before the subsidization, we observe that a household in this state randomize between no improvements ($p = 0.5$) and a moderate level of improvement ($p = 0.5$). After the subsidization, the new equilibrium choice probabilities are calculated and the household will choose to undertake a moderate level of improvement with ($p = 1$).

Household 3: Finally, we consider a household with 2 members, living in a \$255,000 house with an annual income of \$35,000, no mortgage, and paying no property taxes. Empirically, before the subsidization, we observe that this household carries out a high level of improvement with ($p = 1$). After the subsidization, the new equilibrium choice probabilities do not change the household's behavior at all. They will still carry out a high level of improvement with ($p = 1$).

We also conduct the experiment of tripling property taxes across the board. Again, to make the analysis more clear and to look at the differences between the tax and the subsidy from the previous experiment, we will look at the same three households used in the previous experiment.

Household 1: Under the same state as before, but now under a (permanent) tripling of property taxes, the household still does not carry out any improvements ($p = 1$)

Household 2: Under the same state as in the previous exercise, a (permanent) tripling of property taxes, the household randomizes between no improvements with ($p = 0.87$) and moderate improvements with ($p = 0.13$).

Household 3: Under the same state as in the previous exercise, a (permanent) tripling of property taxes, the household will now randomize between no improvements ($p = 0.5$) and a moderate level of improvements ($p = 0.5$)

These three households illustrate the effect of a subsidy and a tax increase, and the results are as expected. In light of a subsidy, households will have a higher propensity to carry out improvements. The tax has an opposite effect. Given that more improvements lead to higher property values, increases in the property tax serve to dissuade homeowners from undertaking improvements. The results are summarized in Table 2.10.

For the two-player model, we carry out the same experiments. To make the analysis more clear, we look at particular neighborhood pairs and see how the same policy experiments (subsidization and taxation) play out in the two-player setting. The resultant of handing out a flat \$2000 subsidy for three neighborhood pairs:

Neighborhood 1: The actual characteristics of this neighborhood are summarized in Table 2.11. The initial choice of these two players in this particular state was to choose "no improvement" with ($p = 1$). After the subsidy, however, the new equilibrium choice probabilities for one player are no improvement ($p = 0.396$),moderate improvement ($p = 0.302$) and high level improvement ($p = 0.302$). For her opponent, the choice probabilities are no improvement ($p = 0.414$), moderate improvement ($p = 0.293$), and a high level of improvement ($p = 0.293$). We can see that with the subsidy, both agents wish to increase the probability that they carry out a moderate or a high level of improvement.

Neighborhood 2: The actual characteristics of this neighborhood are summarized in Table 2.11. The initial choice of these two players in this particular state was to choose "high level of improvement" with ($p = 1$). After the subsidy, the new equilibrium choice probabilities for one player are no improvement ($p = 0.280$),moderate improvement ($p = 0.280$) and high level improvement ($p = 0.440$). For her opponent, the choice probabilities are no improvement ($p = 0.285$), moderate improvement ($p = 0.285$), and a high level of improvement ($p = 0.430$).

Neighborhood 3: The actual characteristics of this neighborhood are summarized in Table 2.11. The empirical choice probabilities of the first player was no improvement ($p = 0.166$),moderate improvement ($p = 0.667$) and high level improvement ($p = 0.166$), and the second player was no improvement ($p = 0.571$),moderate improvement ($p = 0.429$) and high level improvement ($p = 0$) After the subsidy, the new equilibrium choice probabilities for the first player are no improvement ($p = 0.218$),moderate improvement ($p = 0.435$) and high level improvement ($p = 0.347$). For her opponent, the choice probabilities are no improvement ($p = 0.655$), moderate improvement ($p = 0.198$), and a high level of improvement ($p = 0.147$).

We also conduct the experiment where property taxes are (permanently) tripled.

Neighborhood 1: With the same characteristics as listed in Table 2.11. After (permanently) tripling property taxes, the new choice probabilities are for the first player: no improvement ($p = 0.395$),moderate improvement ($p = 0.303$) and high level improvement ($p = 0.302$) and for the second player: no improvement ($p = 0.410$),moderate improvement ($p = 0.295$) and high level improvement ($p = 0.295$)

Neighborhood 2: With the same characteristics as listed in Table 2.11. After (permanently) tripling property taxes, the new choice probabilities are for the first player: no improvement ($p = 0.285$),moderate improvement ($p = 0.285$) and high level improvement ($p = 0.430$) and for the second player: no improvement ($p = 0.292$),moderate improvement ($p = 0.291$) and high level improvement ($p = 0.417$)

Neighborhood 3: With the same characteristics as listed in Table 2.11. After (permanently) tripling property taxes, the new choice probabilities are for the first player: no improvement ($p = 0.230$),moderate improvement ($p = 0.427$) and high level improvement ($p = 0.343$) and for the second player: no improvement ($p = 0.620$),moderate improvement ($p = 0.220$) and high level improvement ($p = 0.160$)

Recall that the most frequent observations were that moderate improvements were being carried out, and not improving at all was a common outcome for both neighbors. We assume that when the best response operator is iterated until a fixed point it reached, that these constitute the new equilibrium choice probabilities. If we compare the two regimes, subsidies, versus taxation, we see that for only Neighborhood 2 does behavior fall where we would predict it, in that we should see an increased propensity to build as a result of the subsidy. In Neighborhood 3, we see a higher propensity to undertake improvements under higher taxes as opposed to subsidization, which is a somewhat counter intuitive result. In Neighborhood 1, the outcomes in both experiments are similar. The odd result may be due to the negative parameter that we recovered for housing consumption. In neighborhoods 1 and 2, occupants also play almost symmetric strategies, which makes sense given that the opponents in those neighborhoods have similar characteristics.

In addition, we run one more experiment, where one of the neighbors is offered a $2000 subsidy and the other is not. After iterating the best response operator to a fixed point, we find that there is no substantial change in the equilibrium conditional choice probabilities as compared to the equilibrium conditional choice probabilities when both neighbors are subsidized. This may be an indication that this equilibirium is very stable and cannot be changed by perturbations of the primitives, or that the true counterfactual equilibrium choice probabilities are very far away (in CCP space). This seems to imply that subsidies do not need to be given to both of the neighbors since the same result would occur if the government only subsidized one of them.

As mentioned before, one thing we would like to do is compare our neighborhood pairs to the slumlord's dilemma, that in the absence of intervention, homeowners would opt not to make improvements. If we look at Neighborhood 1, where we observed both neighbors choosing to make no improvements, under the subsidy, the new equilibrium indicates that both neighbors place a positive probability on moderate to major improvements. The indication is that the subsidy does provide an incentive to make improvements. One of the rationales for subsidies is that the external benefits of improvements will more than offset the subsidy itself. For the two-player neighborhoods, we compared two regimes. The first regime is the regime with asymmetric subsidization in which one of the two neighbors receives a $2000 subsidy for improvements. The second regime is one where a one time infusion of $2000 in cash is given to both households in the neighborhoods. A comparison of expected utilities indicates that the expected utility of households in the first regime is at least as much as under the second regime.

2.25 Conclusion

The main goal of this chapter was to provide an analysis of dynamic home improvement behavior in a structural setting. Obviously there are many obstacles to a model of this nature. In general, dynamic models, both single agent and multi-player are difficult to model and are often difficult to estimate. What we have done is to approximate the problem of a homeowner who carries out home improvements both in a single agent framework and a two-player game using the Hotz and Miller (1993) CCP estimator (and similar estimators to it). We have also tried to use the primitive utility parameters recovered and attempted to analyze best responses in light of different counterfactual policy changes. According to our estimates, in the single agent model, behavior is what we would expect: a higher propensity to improve in the presence of subsidies, and a lower propensity to improve in the presence of a property tax. In the two-player model, the results are slightly less intuitive. While the subsidization and taxation affect the equilibrium choice probabilities, it is not overly obvious as to how the interaction affects the counterfactual equilibrium. However we do find that subsidizing one household is as valuable as giving both households the same amount of the subsidy in cash. Future work calls for possibly more advanced methods such as Aguirregabiria and Mira (2002) and Aguirregabiria and Mira (2007) Nested Pseudo Likelihood method or continuous control methods such as those found in Bajari Benkard and Levin (2007) or Srisuma (2010).

2.26 References

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

Victor Aguirregabiria & Pedro Mira, 2002. "Swapping the Nested Fixed Point Algorithm: A Class of Estimators for Discrete Markov Decision Models," Econometrica, Econometric Society, vol. 70(4), pages 1519-1543, July.

Victor Aguirregabiria & Pedro Mira, 2007. "Sequential Estimation of Dynamic Discrete Games," Econometrica, Econometric Society, vol. 75(1), pages 1-53, 01.

Victor Aguirregabiria, 2009. "A Method for Implementing Counterfactual Experiments in Models with Multiple Equilibria," Working Papers tecipa-381, University of Toronto, Department of Economics.

Aguirregabiria, Victor & Mira, Pedro, 2010. "Dynamic discrete choice structural models: A survey," Journal of Econometrics, Elsevier, vol. 156(1), pages 38-67, May.

Robert Aumann, “Acceptable points in general cooperative n-person games”, in R. D. Luce and A. W. Tucker (eds.), Contributions to the Theory 23 of Games IV, Annals of Mathematics Study 40, 287–324, Princeton University Press, Princeton NJ.

Patrick Bajari & C. Lanier Benkard & Jonathan Levin, 2007. "Estimating Dynamic Models of Imperfect Competition," Econometrica, Econometric Society, vol. 75(5), pages 1331-1370, 09.

Patrick Bajari & Phoebe Chan & Dirk Krueger & Daniel Miller, 2010. "A Dynamic Model of Housing Demand: Estimation and Policy Implications," NBER Working Papers 15955, National Bureau of Economic Research, Inc

Baker, K. and Kaul, B. (2002), Using Multiperiod Variables in the Analysis of Home Improvement Decisions by Homeowners. Real Estate Economics, 30: 551–566.

David Barker and Eric Miller (2009) “Homeownership and Child Welfare” Real Estate Economics, 37, 279-303

Patrick Bayer, Robert McMillan, Alvin Murphy, and Christopher Timmins (2009) “A Dynamic Model of Demand for Houses and Neighborhoods,” mimeo, Duke University.

Kelly Bishop (2008), "A Dynamic Model of Location Choice and Hedonic Valuation", Working Paper Washington University of St. Louis

Thomas P. Boehm & Keith R. Ihlanfeldt, 1986. "The Improvement Expenditures of Urban Homeowners: An Empirical Analysis," Real Estate Economics, American Real Estate and Urban Economics Association, vol. 14(1), pages 48-60.

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

Davis, O.A. and Whinston, A.B. (1961) ‘The economics of renewal’, in Wilson, J.Q. (ed.) Urban renewal: The Record and the Controversy, Cambridge, MA: Harvard University Press 2

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 Economics, 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.

James Heckman(1981b). “The Incidental Parameter Problem and the Problem of Initial Conditions,” in C. Manski and D. McFadden, eds, Structural Analysis of Discrete Data, MIT Press.

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

V. Joseph Hotz and Robert A. Miller, "Conditional Choice Probabilities and the Estimation of Dynamic Models" The Review of Economic Studies Vol. 60, No. 3 (Jul., 1993), pp. 497-529

Ioannides, Yannis M., 2002. "Residential neighborhood effects," Regional Science and Urban Economics, Elsevier, vol. 32(2), pages 145-165, March.

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

Linton, Oliver B. and Srisuma, Sorawoot, Semiparametric Estimation of Markov Decision Processes with Continuous State Space (August 2010). LSE STICERD Research Paper No. EM550

Mendelsohn, Robert, 1977. "Empirical evidence on home improvements," Journal of Urban Economics, Elsevier, vol. 4(4), pages 459-468, October

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

Ariel Pakes, 1991. "Dynamic Structural Models: Problems and Prospects. Mixed Continuous Discrete Controls and Market Interactions," Cowles Foundation Discussion Papers 984, Cowles Foundation for Research in Economics, Yale University.

Martin Pesendorfer & Philipp Schmidt-Dengler, 2008. "Asymptotic Least Squares Estimators for Dynamic Games," Review of Economic Studies, Wiley Blackwell, vol. 75(3), pages 901-928, 07

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

Esteban Rossi-Hansberg & Pierre-Daniel Sarte & Raymond Owens, 2010. "Housing Externalities," Journal of Political Economy, University of Chicago Press, vol. 118(3), pages 485-535, 06.

John Rust (1994) Structural Estimation of Markov Decision Processes in R. Engle and D. McFadden (eds.) Handbook of Econometrics, North Holland: Amsterdam.

John Rust & Christopher Phelan, 1997. "How Social Security and Medicare Affect Retirement Behavior in a World of Incomplete Markets," Econometrica, Econometric Society, vol. 65(4), pages 781-832, July

Srisuma, Sorawoot, Minimum Distance Estimation for a Class of Markov Decision Processes, Working Paper

Charles M. Tiebout, "A Pure Theory of Local Expenditures", The Journal of Political Economy, Vol. 64, No. 5. (Oct., 1956), pp. 416-424.

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

Wolpin, Kenneth I, 1984. "An Estimable Dynamic Stochastic Model of Fertility and Child Mortality," Journal of Political Economy, University of Chicago Press, vol. 92(5), pages 852-74, October.

Jeffrey M. Wooldridge, 2005. "Simple solutions to the initial conditions problem in dynamic, nonlinear panel data models with unobserved heterogeneity," Journal of Applied Econometrics, John Wiley & Sons, Ltd., vol. 20(1), pages 39-54

Ziegert, A. L. (1988), The Demand for Housing Additions: An Empirical Analysis. Real Estate Economics, 16: 479–496

2.27 Tables For Chapter 2

Variable	Obs.	Mean	Std.
agehouse1 = age of the house in years	10637	36.1567	22.684
turnover1 = did the unit turnover since the last survey?	10637	0.18736	0.39022
totb1 = total number of bathrooms	10637	2.07972	0.88159
educ1 = education in years of the homeowner	10637	13.7576	3.18652
sex1 = 1 if the homeowner is male	10637	0.78753	0.40907
married1 = 1 if the homeowner is married	10637	0.77465	0.41783
white1 = 1 if the homeowner is white	10637	0.8619	0.34502
hispanic1 = 1 if the homeowner is hispanic	10637	0.0486	0.21505
black1 = 1 if the homeowner is black	10637	0.06017	0.23781
indian1 = 1 if the homeowner is Indian	10637	0.00301	0.05477
asian1 = 1 if the homeowner is asian	10637	0.01965	0.1388
bedrms1 = number of bedrooms in unit	10637	3.19667	0.83674
per1 = number of people	10637	2.90646	1.40675
lot1 = lot size in square feet	10637	74945.6	190070
unitsf1 = unit size in square feet	10637	2101.46	926.519
cstmnt1 = cost of routine maintenance	10637	488.677	733.424
zinc1 = annual household income	10637	61473.3	42590.4
rooms1 = total number of rooms	10637	6.69869	1.68894
hown1 = 1-10 Likert scale of neighborhood satisfaction	10637	7.94106	2.86088
howh1 = 1-10 Likert cale of house satisfaction	10637	8.20692	2.69361
age1 = age of homeowner	10637	51.159	14.399
yearslived1 = number of years lived in house	10637	13.891	12.8168
value1 = value of house	10637	136998	83388.6

Table 2.1. Summary Statistics for 1997 AHS Data

Variable	Obs.	Mean	Std.
agehouse2	10637	38.1567	22.684
turnover2	10637	0.11874	0.32349
totb2	10637	2.09301	0.93155
educ2	10637	13.7847	3.19127
sex2	10637	0.77691	0.41634
married2	10637	0.77729	0.41609
white2	10637	0.90096	0.29873
hispanic2	10637	0.04983	0.2176
black2	10637	0.06158	0.2404
indian2	10637	0.00348	0.05888
asian2	10637	0.02059	0.14201
bedrms2	10637	3.19536	0.79594
per2	10637	2.87722	1.40178
lot2	10637	73250.5	185437
unitsf2	10637	2196.86	1501.4
cstmnt2	10637	498.031	884.863
zinc2	10637	71797.1	65101.2
rooms2	10637	6.59528	1.60407
hown2	10637	7.89508	3.00821
howh2	10637	8.11526	2.90233
age2	10637	52.4244	14.5378
yearslived2	10637	15.1338	13.0817
value2	10637	154353	117052

Table 2.2. Summary Statistics for 1999 AHS Data

Variable	Obs.	Mean	Std.
agehouse3	10637	40.1567	22.684
turnover3	10637	0.09561	0.29407
totb3	10637	2.10191	0.93858
educ3	10637	13.8014	3.19492
sex3	10637	0.76826	0.42196
married3	10637	0.77024	0.4207
white3	10637	0.84996	0.35713
hispanic3	10637	0.05105	0.22011
black3	10637	0.06336	0.22363
indian3	10637	0.0032	0.05645
asian3	10637	0.02181	0.14607
bedrms3	10637	3.20476	0.78882
per3	10637	2.8289	1.40356
lot3	10637	73369.1	190179
unitsf3	10637	2157.37	1500.92
cstmnt3	10637	519.586	863.807
zinc3	10637	81928.1	93610.5
rooms3	10637	6.66184	1.74338
hown3	10637	7.90298	2.92995
howh3	10637	8.16292	2.81494
age3	10637	53.5987	14.6764
yearslived3	10637	16.2693	13.3383
value3	10637	177611	159224

Table 2.3. Summary Statistics for 2001 AHS Data

Variable	Obs.	Mean	Std.
agehouse4	10637	42.1567	22.684
turnover4	10637	0.0832	0.2762
totb4	10637	2.12015	0.9475
educ4	10637	13.8261	3.23628
sex4	10637	0.75952	0.4274
married4	10637	0.76093	0.43114
white4	10637	0.843	0.36382
hispanic4	10637	0.06233	0.24176
black4	10637	0.0613	0.23988
indian4	10637	0.00301	0.05477
asian4	10637	0.02096	0.14327
bedrms4	10637	3.21783	0.80312
per4	10637	2.78105	1.40109
lot4	10637	73102.2	186613
unitsf4	10637	2180.81	1594.61
cstmnt4	10637	523.999	893.5
zinc4	10637	78715.1	169956
rooms4	10637	6.76431	1.93875
hown4	10637	7.82091	3.16156
howh4	10637	8.05349	3.07085
age4	10637	54.7234	14.9071
yearslived4	10637	17.4615	13.6305
value4	10637	190422	162604

Table 2.4. Summary Statistics for 2003 AHS Data

Variable	Coefficient	Standard Error
lag = lagged dependent variable	2.07E-06	6.77E-07
initial lag	0.1042482	0.0172907
agehouse = age of the house in years	0.0273304	0.0099282
turnover = 1 if unit experienced turnover	0.0382506	0.0307179
totb= total number of bathrooms	-0.0822229	0.0266454
educ = education in years of the homeowner	0.0022539	0.0078486
sex = 1 if the homeowner is male	-0.0034501	0.0517144
married = 1 if the homeowner is married	-0.0126556	0.0425699
white = 1 if the homeowner is white	-0.0284208	0.1147811
hispanic = 1 if the homeowner is hispanic	-0.0042707	0.1321595
black = 1 if the homeowner is black	0.0419513	0.1588716
indian = 1 if the homeowner is Indian	-0.1081255	0.2738173
asian = 1 if the homeowner is asian	-0.200378	0.1555025
bedrms = number of bedrooms in unit	-0.0311829	0.0225655
per = number of people	0.0000422	0.0128691
lot = lot size in square feet	-4.45E-08	7.38E-08
unitsf = unit size in square feet	-0.0000185	0.0000109
cstmnt = cost of routine maintenance	-0.0000299	0.0000103
zinc = annual household income	-4.93E-08	8.41E-08
rooms = total number of rooms	-0.01998	0.0080229
hown = 1-10 Likert scale of neighborhood satisfaction	0.0247452	0.0061901
howh = 1-10 Likert cale of house satisfaction	-0.0312967	0.0063392
age = age of homeowner	-0.0074012	0.0022022
yearslived = number of years lived in house	0.0010079	0.0030362
value = value of house	-7.12E-08	1.08E-07
year1 = 1997 year dummy	0.1080217	0.0520649
year2 = 1999 year dummy	0.1112401	0.0339211
year3 = 2001 year dummy	(omitted)	
year4 = 2003 year dummy	(omitted)	
constant	-1.531177	0.1593785

Table 2.5. Reduced Form Results

variable	mean	SD
amtx1 = property tax rate in 1997	0.0174	0.0195
amtx2 = property tax rate in 1999	0.0147	0.0227
amtx3 = property tax rate in 2001	0.0113	0.0089
amtx4 = property tax rate in 2003	0.0141	0.0202
rac1 = $ spent on improvements in 1997	3125	8955.6
rac2 = $ spent on improvements in 1997	3655.7	13692
rac3 = $ spent on improvements in 1997	3942.4	14918
rac4 = $ spent on improvements in 1997	4659.3	22305
int1 = initial mortgage interest rate	0.0448	0.0394

Table 2.6. New Variables For Single Agent Analysis

	No Improvement	Moderate Improvement	Major Improvement
Improvements between 1995-1997	4510	4474	1653
Improvements between 1999-2001	4460	4348	1829
Improvements between 2001-2003	4793	4068	1776
Improvements between 2003-2005	4837	3841	1959

Table 2.7. Proportion of Improvement Levels Across Time Periods

variable	mean	SD
zinc85 = household income in 1985	39341	28649
zinc89 = household income in 1989	43686	32064
zinc93 = household income in 1993	47889	35748
rac85 = $ spent on improvements in 1985	1849	3372.2
rac89 = $ spent on improvements in 1989	2043.1	3999.3
rac93 = $ spent on improvements in 1993	1562.8	3411.7
per85 = household members in 1985	2.8602	1.3603
per89 = household members in 1989	2.7274	1.3484
per93 = household members in 1993	2.6003	1.3339
value85 = value of unit in 1985	88534	55498
value89 = value of unit in 1989	118354	87642
value93 = value of unit in 1993	122090	84955
amtx85 = property tax rate in 1985	0.0198	0.0499
amtx85 = property tax rate in 1985	0.0195	0.0783
amtx85 = property tax rate in 1985	0.0201	0.0586
int85 = initial mortgage rate	0.0701	0.135
n = 1416		

Table 2.8. Summary Statistics for the Two-Player Game

	None	Middle	High
None	512	560	190
Middle		270	160
High			35

Table 2.9. Pooled Frequencies of Actions in the Two-Player Game

Characteristics		No Improvement	Moderate Improvement	Major Improvement	Treatment
House Value	$45,000	p=1	p=0	p=0	Empirical
Household Income	$35,000	p=0	p=0.5	p=0.5	After Subsidy
Household Size	2	p=1	p=0	p=0	After Tax
Property Tax Rate	0%				
Mortgage Rate	7%				
House Value	$165,000	p=0.5	p=0.5	p=0	Empirical
Household Income	$150,000	p=0	p=1	p=0	After Subsidy
Household Size	2	p=0.87	p=0.13	p=0	After Tax
Property Tax Rate	2%				
Mortgage Rate	0%				
House Value	$255,000	p=0	p=0	p=1	Empirical
Household Income	$35,000	p=0	p=0	p=1	After Subsidy
Household Size	2	p=0.5	p=0.5	p=0	After Tax
Property Tax Rate	0%				
Mortgage Rate	0%				

Table 2.10. Single Agent Policy Experiments

Characteristics Neighborhood 1	Agent 1	Agent 2
House Value	$20,000	$20,000
Household Income	$15,000	$15,000
Household Size	3	1
Property Tax Rate	0.50%	0.50%
Mortgage Rate	0%	7%
Neighborhood 2		
House Value	$220,000	$220,000
Household Income	$45,000	$105,000
Household Size	4	2
Property Tax Rate	2%	2%
Mortgage Rate	0%	0%
Neighborhood 3		
House Value	$20,000	$20,000
Household Income	$15,000	$15,000
Household Size	2	1
Property Tax Rate	0.50%	0.50%
Mortgage Rate	0%	0%

Table 2.11. Two-Player Game Policy Experiments

Treatment Neighborhood 1	No Improvement	Moderate Improvement	Major Improvement	Treatment
Empirical	p=1, p=1	p=0, p=0	p=0, p=0	Empirical
After Subsidy	p=0.396, p=0.293	p=0.302, p= 0.293	p=0.302, p=0.293	After Subsidy
After Tax	p=0.395, p=0.410	p=0.303, p=0.295	p=0.302, p=0.295	After Tax
Neighborhood 2				
Empirical	p=0, p=0	p=0, p=0	p=1, p=1	Empirical
After Subsidy	p=0.280, p=0.290	p=0.280, p=0.290	p=0.440, p=0.420	After Subsidy
After Tax	p=0.285, p=0.292	p=0.285, p=0.291	p=0.430, p=0.417	After Tax
Neighborhood 3				
Empirical	p=0.166, p=0.571	p=0.667, p=0.429	p=0.166, p=0	Empirical
After Subsidy	p=0.218, p=0.655	p=0.435, p=0.198	p=0.347, p=0.147	After Subsidy
After Tax	p=0.230, p=0.620	p=0.427, p=0.220	p=0.343, p=0.160	After Tax

Table 2.12. Two-Player Game Policy Experiments Continued

Herman Li Vita

Herman was born in Pittsburgh, Pennsylvania on September 19, 1983. He went on to graduate from the University of Pennsylvania in May 2005 with a BA in mathematics and economics. Immediately afterwards, he began his doctoral studies at The Pennsylvania State University where he completed his degree in August 2011. His research interests include Real Estate/Housing Economics and Urban Economics. He was awarded two departmental teaching awards, the first in the Spring of 2006 and the second in the Summer of 2009. In 2010, he was also awarded the Harold F. Martin Graduate Assistant Outstanding Teaching Award as well as the Lincoln Institute of Land Policy C. Lowell Harriss Dissertation Fellowship.