A Dynamic Model of the Housing Market[*]

Patrick Bayer, Duke University

Robert McMillan, University of Toronto

Alvin Murphy, Duke University

Christopher Timmins, Duke University

July 13, 2007

Abstract

We use a unique dataset linking information about buyers and sellers to the complete census of housing transactions in the San Francisco metropolitan area for a period of 15 years to examine the microfoundations of housing market dynamics. We develop a tractable model of neighborhood choice in a dynamic setting along with a computationally straightforward estimation approach. This approach allows the observed and unobserved features of each neighborhood to evolve in a completely flexible way and uses information on neighborhood choice and the timing of moves to recover semi-parametrically: (i) preferences for housing and neighborhood attributes, (ii) preferences regarding the performance of the house as a financial asset (e.g., expected appreciation, volatility), and (iii) moving costs. This model and estimation approach is potentially applicable to the study a wide set of dynamic phenomena in housing markets and cities.

In this paper, we use the model to develop testable implications of housing market efficiency and in particular rational expectations on the part of home buyers. We begin by showing that when the model is restricted so that all households have identical preferences, rational expectations implies the absence of predictable returns, i.e., the absence of the positive persistence in housing prices shown in the literature following Case and Shiller (1989). Thus, as the houses considered in an analysis are closer substitutes for one another, the predictability of returns should fall to zero. We examine this hypothesis empirically by studying the dynamics of housing prices at various levels of aggregation across both geographic and

---

socioeconomic dimensions. The results of our analysis are generally inconsistent with rational expectations: there is significant positive persistence in appreciation across counties and Census PUMAs and significant negative persistence across Census tracts within counties or PUMAs within the San Francisco Bay Area. Collectively, these results suggest that a positive shock to prices in a given tract predicts significant positive returns in nearby tracts the following year. Future analysis will use the estimated structural model to formally test rational expectations.

Note: We intend to divide the analysis presented in this paper into two papers roughly corresponding to the paragraphs of the abstract above. In preparing papers for conferences at SITE, NBER, and the Econometric Society meetings in Summer 2007, we have found it useful to temporarily fold these ideas into a single paper that summarizes where we stand in this research project.

1 Introduction

The purchase of a primary residence is simultaneously the largest single consumption decision and largest single investment of the vast majority of US households; the typical household spends about 23 percent of its income on its house and its house constitutes two-thirds of its portfolio.[1] As a result, the housing market not only constitutes an important sector of the economy but also blends the features of consumption and financial markets in unique and interesting ways.

Relative to simpler consumption decisions, the home-buying decision is complicated by the sheer amount of money involved in the transaction and the associated transaction costs. The latter ensure that this decision is very costly to adjust and, as a result, that dynamic considerations including the expected performance of the house as an asset and expected evolution of the property and neighborhood have an important role in the decision. These dynamic considerations add to the complexity of the static decision, which already folds a number of important dimensions of consumption (e.g., housing characteristics, commuting time, local schools, crime, and other neighborhood amenities) into a single decision.

As opposed to many standard financial instruments, the existence of large transaction costs, the predominance of owner-occupancy in large segments of the market, and the inherent difficulty of holding short positions limit the ability of professionals to eliminate pricing inefficiencies in the housing market. As a result, housing prices exhibit time-series properties at both high and low frequencies that are inconsistent with the standard implications of the efficient market hypothesis. In particular, previous research has consistently documented that prices exhibit positive persistence (inertia) in the short-run (annually) and mean reversion in the longer run (five years).[2]

Because professionals cannot eliminate the predictability of future prices, it is well understood that this predictability alone does not imply that the economic agents operating in the housing market are irrational. In fact, whether individual agents act with rational expectations remains very much an open question. This question is at the heart of the contentious debate over

---

[1]According to the American Household Survey in 2005, the national median percentage of income spent on housing was 23 percent. Tracy, Schneider, and Chan (1999) report the portfolio share figure.

[2]A great deal of empirical research documents and explores these aspects of housing market dynamics. See, for example, Abraham and Hendershott (1996), Capozza et. al. (2002), Case and Mayer (1995), Case and Shiller (1989), Cho (1996), Clayton (1997), Englund, Gordon and Quigley (1999), Gelfand et. al. (2004), Himmelberg, Mayer, and Sinai (2005), and Lamont and Stein (2004).

whether the recent upsurge in housing prices in many US metropolitan areas is a bubble fueled by unrealistic expectations or perfectly understandable in terms of the fundamentals.[3]

In this paper, we develop an estimable model of the dynamic decision-making of individual home-owners with the aim of using the model to provide new insight into the microfoundations of housing market dynamics. In so doing, we seek to make explicit the link between the microeconomic primitives of the housing market (i.e., the factors governing individual buying and selling decisions) and the aggregate market dynamics characterized in the existing literature.

The starting point for our analysis is a unique dataset linking information about buyers and sellers to the complete census of housing transactions in the San Francisco metropolitan area for a period of 15 years (1.5 million transactions in all). In addition to demographic and economic information about buyers and sellers, this dataset contains information about the structure and lot (e.g., square footage, year built, lot size), transaction price, attributes of the mortgage, exact location, exact sales date, and a unique house ID that identifies repeat sales of the same property. In most cases, it is also possible to link sellers of one property to their newly purchased properties, provided they move within the same metropolitan area. By linking information about buyers and sellers to houses at a fine level of granularity in terms of both space and time, this dataset has significant advantages over the large-scale datasets that have been used in previous research to characterize housing market and neighborhood dynamics.

With this dataset in hand, we develop a tractable model of neighborhood choice in a dynamic setting, along with a corresponding estimation approach that is computationally straightforward. This approach, which combines and extends the insights of Rust (1987), Berry (1994), and Hotz and Miller (1993), allows the observed and unobserved features of each neighborhood to evolve in a completely flexible way and uses information on neighborhood choice and the timing of moves to recover semi-parametrically: (i) preferences for housing and neighborhood attributes, (ii) preferences regarding the performance of the house as a financial asset (e.g., expected appreciation, volatility), and (iii) moving costs. In order to accommodate a number of important features of housing market, this approach extends methods developed in the recent

---

[3]For examples of research that argue that recent price increases are not driven by bubbles, see McCarthy and Peach (2004), and Himmelberg, Mayer and Sinai (2005). For a contrasting view, see Shiller (2005, 2006) and Baker (2006). Case, Shiller and Quigley have done some direct surveys about expectations. Researchers have been able to test some implications of market efficiency. See, for example, Rosenthal (1999). At some level, it may also be worth noting that the predictability of housing prices is not a very well known thing. Also note that Glaeser and Gyourko (2006) have a hard time fitting high frequency price volatility with their calibrated model.

literature on the dynamic demand for durable goods in a number of key ways.[4]

The model and estimation method that we propose are potentially applicable to the study of a wide set of dynamic phenomena in housing markets and cities. These include, for example, the analysis of the microdynamics of residential segregation and gentrification within metropolitan areas.[5] More generally, the model and estimation approach can be extended straightforwardly to study the dynamics of housing and labor markets in a system of cities. A number of important lines of research within labor and urban economics draw intuitively on what would be a dynamic Roback (1982) framework and, yet, to our knowledge, there has been no attempt to estimate such a model directly.[6] In this way, an important goal of this paper is to provide a coherent and computationally feasible basis for the analysis of the dynamics of housing and labor markets from a microeconomics perspective.

In this version of our paper, we focus on developing testable implications of rational expectations in our dynamic model. We begin by showing that if all households were restricted to have identical preferences, rational expectations would imply the absence of the predictability of future prices. The reason is simple: because all houses for sale at a given time need to provide the same indirect utility to the set of current buyers, current buyers would arbitrage away any price inefficiencies. Thus, in the presence of homogeneous preferences, rational expectations on the part of individual home-buyers eliminate any predictable component of price without the need for professional investors.

This simple insight suggests that when the set of houses included in an analysis of the persistence of appreciation are closer substitutes for one another, the predictability of returns should fall to zero. In this way, while significant short-run predictable difference in expected returns might arise between two distant cities because the households in each of those cities are strongly attached to their respective labor markets in the short run, any predictable price differences between two neighborhoods that are reasonably close substitutes within a single city should be immediately

---

[4]We discuss this literature in more detail in Section 2.

[5]Recent theoretical research on aspects of the dynamic microfoundations of housing markets by Ortalo-Magne and Rady (2002, 2005, 2006) and Bajari, Benkard, and Krainer (2005) raise a number of additional interesting empirical questions that could be addressed using this framework.

[6]There a number of interesting spects of labor and housing market dynamics across cities at both high and low frequencies. Gyourko, Mayer, and Sinai (2004), for example, focus on low frequency dynamics of migration and housing prices across US cities. Glaeser and Gyourko (2006) calibrate a dynamic Rosen model and use it to explore both high and low frequency dynamics of the housing market. A long literature in labor economics following Blanchard and Katz (1992) explores both high and low frequency labor market dynamics and migration across cities and regions. Ultimately, all of these important dynamic features of housing and labor markets should be able to be viewed through the lens of a single dynamic Rosen framework.

arbitraged away by the set of current home buyers considering those neighborhoods.

We examine this hypothesis empirically by studying the dynamics of housing prices at various levels of aggregation across both geographic and socioeconomic dimensions. The results of our analysis are generally inconsistent with a pure version of rational expectations: there is significant positive persistence in appreciation across counties and Census PUMAs and significant negative persistence across Census tracts within counties or PUMAs within the San Francisco Bay Area. Collectively, these results suggest that a positive shock to prices in a given tract predicts significant positive returns in nearby tracts the following year. This pattern may be driven by the challenging informational problem that home buyers and sellers face in the market due to the significant heterogeneity in houses and neighborhoods and the price formation processes (e.g., the use of comparable sales) that individuals rely on to overcome this problem. We close the current version of the paper with a discussion of how the structural model can be used to formally test for rational expectations.

The remainder of the paper proceeds as follows. Section 2 briefly summarizes how our estimator relates to recent literature on dynamic demand for durable goods. Section 3 describes the dataset we develop. Our model, estimation strategy, and parameter estimates are presented in Sections 4-6, respectively. Section 7 uses the model to develop a new testable implication of market efficiency and presents a series of related analysis that tests of positive persistence in housing price appreciation at various levels of aggregation. Section 8 concludes.

2   Related Literature on Dynamic Demand

The model and estimation approach developed in this paper are related to a recent literature on the dynamic demand for durable goods. Much of this literature has focused on extending BLP style models to allow for forward looking behavior, while retaining the controls for unobserved product characteristics. Melnikov (2001) develops a tractable model without individual heterogeneity to estimate the demand for printers. Agents make two decisions: they decide what period (if any) to buy a printer and then which brand to buy conditional on buying a printer. All the dynamic behavior lies in the timing of purchase and the brand choice is a static discrete choice. Carranza (2007) looks at the digital camera market and extends the Melnikov (2001) model to allow for random coefficients and captures the dynamic decision using a reduced form

specification. By allowing consumers to make repeat purchases, Gowrisankaran and Rysman (2007) allow both the timing and product choices to be determined dynamically. They estimate the model by nesting a Rust (1987) style optimal stopping problem inside of the BLP style product choice model. Schiraldi (2007) extends the Gowrisankaran and Rysman (2007) model to include secondary markets and transaction costs.

Erdem, Imai, and Keane (2003) estimate a structural model of the demand for goods that are frequently purchased, branded, storable, and subject to frequent price fluctuations or promotions. They control for the effects of inventory build up and expectations about future price changes. The model, while computationally demanding, allows for individual heterogeneity. Using the market for laundry detergent, Hendel and Nevo (2006) estimate a similar model. They structure the model such that they can separate the brand choice and quantity choice. The quantity choice incorporates forward looking behavior and the brand choice is static. This separation of choices leads to computational simplifications, however, the model can not allow for individual heterogeneity.

A common issue in dynamic discrete choice models is the direct link between the size of the choice set and the size of the state space. Standard estimation approaches such as Rust (1987) quickly become infeasible with a large choice set. Melnikov (2001) proposed a potential solution to this problem where the logit inclusive value is treated as a sufficient statistic for predicting future continuation values. Tractability is maintained as the state space is reduced to one dimension by this assumption at a cost of a loss of information. Similar assumptions are made in Carranza (2007), Hendel and Nevo (2006), Gowrisankaran and Rysman (2007), and Schiraldi (2007).

Our model, which is based on individual level data, incorporates unobserved choice characteristics, endogenous wealth accumulation, and heterogeneous households. The static demand models of Berry (1994), and Berry, Levinsohn, and Pakes (1995) (BLP) introduced a framework for controlling for unobserved product characteristics while highlighting the importance of trying to capture individual heterogeneity. Given individual data, we capture heterogeneity by allowing individuals to value neighborhood attributes differently based on their observable characteristics. In addition to specifying a dynamic model, we also differ from BLP by allowing heterogeneity in the valuation of unobserved neighborhood characteristics.

Our approach differs from these models as it does not require the reduction of the state space to a

univariate statistic. We can avoid the inclusive value sufficiency assumption as the computational burden our estimator is not affected by the size of the state space. We build upon the literature by estimating a semiparametric model with a computationally very straightforward approach. Given the low computational burden of our estimator we place no restrictions on the size of state space or the size of choice set. We also allow heterogeneity in valuation of both observed and unobserved neighborhood characteristics. Finally, we treat the object of choice (housing) as an asset and, as such, the wealth of households changes endogenously.

3 DATA

In this section, we briefly describe the new dataset that we have assembled by merging information about buyers and sellers with the universe of housing transactions in the San Francisco metropolitan areas. We provide more details on the source data and demonstrate that the merge results in a high quality and representative dataset based on multiple diagnostic tests.

The dataset that we develop is drawn from two main sources. The first comes from a national real estate data company and provides information on every housing unit sold in the core counties of the Bay Area (San Francisco, Marin, San Mateo, Alameda, Contra Costa,and Santa Clara) between 1990 and 2004. The buyers' and sellers'names are provided along with transaction price, exact street address, square footage, year built, lot size, number of rooms,number of bathrooms, number of units in building, and many other housing characteristics. Overall, the housing characteristics are considerably better than the those that are provided in Census microdata. A key feature of this transaction dataset is that it also includes information about the buyer's mortgage including the loanamount and lender's name for all loans. It is this mortgage information which allows us to link information about buyers (and many sellers) to this transaction dataset.

The source of the economic and demographic information about buyers(and sellers) is the dataset on mortgage applications published in accordance with the Home Mortgage Disclosure Act (HMDA), which was enacted by Congress in 1975 and is implemented by the Federal Reserve Board's Regulation C.[7] The HMDA data provides information on the race, income, and gender

---

[7]The act requires lending institutions to report public loan data. The purpose of the act is to provide public loan data that can be used to determine whether financial institutions are serving the housing needs of their communities and whether public officials are distributing public-sector investments so as to attract private investment to areas where it is needed. Another purpose is to identify any possible discriminatory lending patterns. (see http://www.ffiec.gov/hmda for more details).

of thebuyer/applicant as well as mortgage loan amount, mortgage lender's name, and the census tract where the property is located. Thus, we are able to merge the two datasets on the basis of the following variables: census tract, loan amount, date, and lender name. Using this procedure, we obtain a unique match for approximately 70% of sales. Because the original transactions dataset includes the full names of buyers and sellers, we are also able to merge demographic and economic information about sellers into the dataset provided (i) a seller bought another house within the metro area and (ii) a unique match with HMDA was obtained for that house. This procedure allows us to merge information about sellers in for approximately35-40 percent of our sample.

To ensure that our matching procedure is valid we conduct two diagnostic tests. Using public access Census micro data from IPUMS,we calculate the distributions of income and race of those who purchased a house in 1999 in each of the six Bay Area counties. We compare these distributions to the distributions in our merged dataset in Table 1. As can be seen, the numbers match almost perfectly in each of the six counties suggesting that the matched buyers are representative of all new buyers.

A comparison of Tables 2 and 3 provides a second diagnostic check on the representativeness of the merged dataset in terms of housing characteristics. Table 2 provides sample statistics for a subset of the house level variables taken from the original dataset that includes the complete universe of transaction, while Table 3 presents sample statistics for the merged dataset. Both tables report variables in 2000 dollars. A comparison of the two tables suggests that the set of houses for which we have a unique loan record from HMDA are very representative of the complete sample of houses. The mean price for the houses in the matched sample is a little higher and the other means are very similar. Overall, our two diagnostic checks provide strong evidence in support the validity of our matching algorithm.

Finally, we close this brief data section by providing the reader with a sense of the variation in the evolution of prices across regions of the Bay Area. The precision of the estimation of the dynamic aspects of the model of neighborhood choice developed below likely depends critically on the fact that rates of house price appreciation are not uniform across census tracts. Figure 1 reports price levels by county from 1990 to 2004. Estimated price levels are derived from a repeat sales analysis in which the log of the sales price (in 2000 dollars) is regressed on a set of county-year fixed effects as well as house fixed effects. The values on the vertical axis indicate the real

Table 1: Comparison of Sample Statistics for Transactions Data/HMDA and IPUMS

	ALAM	C.C.	MARIN	S.F.	S.M.	S.C.
HMDA / Transactions Data						
Median Income	83000	78000	121000	103000	108000	101000
Mean Income	98977	99141	166220	147019	137777	123138
Std Dev Income	96319	97928	176660	225646	123762	125138
IPUMS						
Median Income	83400	76785	120000	100000	102400	100000
Mean Income	104167	99047	162322	137555	140447	124483
Std Dev Income	84823	83932	138329	121552	123451	99373
HMDA / Transactions Data						
% White	49.85	68.27	90.65	59.12	60.08	49.07
% Asian	28.68	10.55	4.68	31.47	26.57	34.21
% Black	6.45	6.01	0.67	2.08	1.22	1.45
% Hispanic	11.76	12.38	2.51	5.86	9.90	12.27
IPUMS						
% White	47.64	64.57	87.5	61.92	58.1	50
% Asian	27.34	11.37	3.3	23.37	25.41	33.51
% Black	7.77	6.05	2.3	2.8	1.24	1.16
% Hispanic	14.62	14.2	3.62	8.18	12.5	12.09

Table 2: Summary Statistics - Transactions Data

Variable	Obs	Mean	Std. Dev.	Median	Min	Max
Sale Price	12325756	352489	221364	306906	16094	1505635
Lot Size	1105557	6884	11385	5400	0	199940
Square Footage	1106305	1647	720	1481	400	10000
Number Bedrooms	1106360	2.94	1.13	3	0	8
Number Rooms	1045028	6.73	2.00	6	1	18

price level of house prices (in percentage terms) relative to 1990 - 1990 price levels are normalized to one for all counties. The figure reveals that by 1995, house prices reached their lowest point in Santa Clara county at 20 percent lower than 1990 levels. In contrast, other counties, such as Contra Costa, experienced larger price depreciation up to 1997 but faster appreciation from 1997 to 2004. Overall, house prices were nearlytwice as high (in real terms) in 2004 as they were in the mid 1990s.

Table 3: Summary Statistics - Transactions Data/HMDA

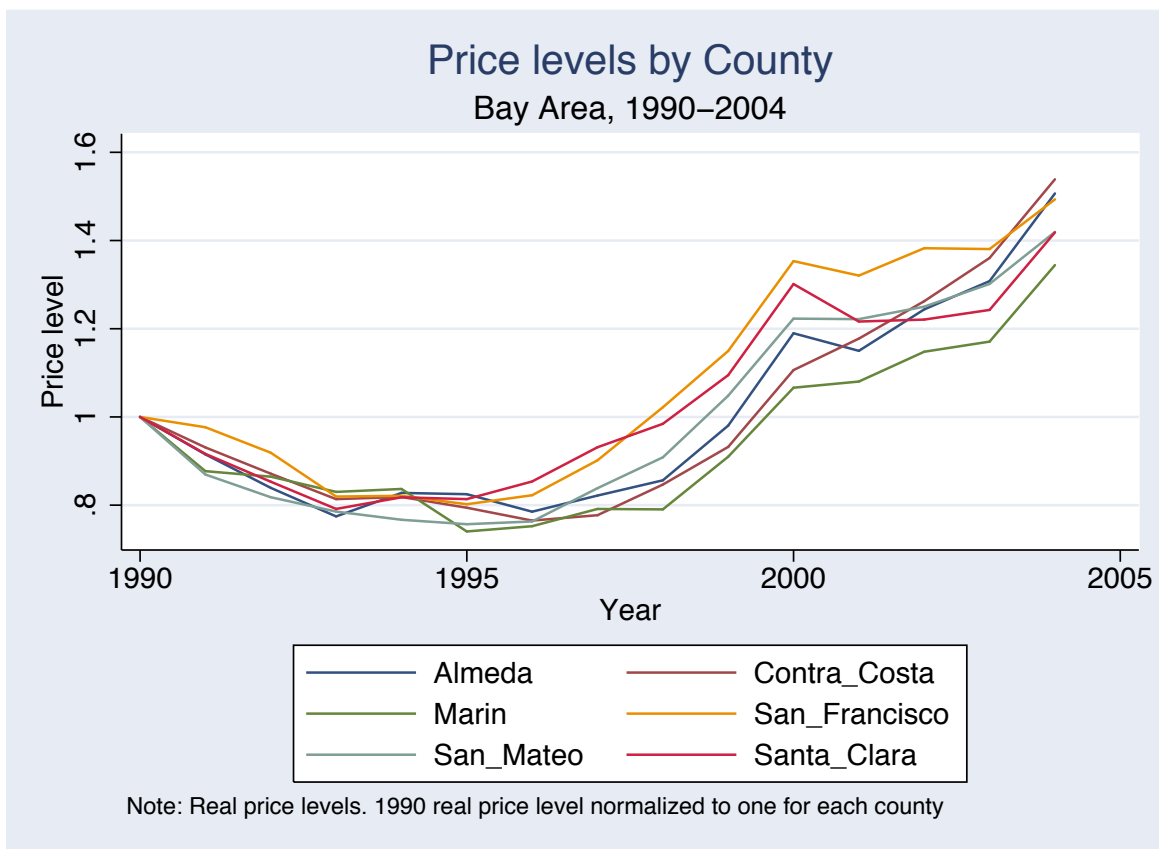| Variable	Obs	Mean	Std. Dev.	Median	Min	Max
Applicant Income	598472	119171	128649	97255	0	11200000
First Loan Amount	611859	298969	151014	266099	2363	2463707
Second Loan Amount	611859	15704	37879	0	0	1820630
White	510544	0.60	0.49	1	0	1
Asian	510544	0.24	0.43	0	0	1
Black	510544	0.03	0.18	0	0	1
Hispanic	510544	0.11	0.31	0	0	1
Male	591646	0.76	0.43	1	0	1
Co-Applicant	611856	0.62	0.48	1	0	1
Sale Price	611859	396475	218474	349697	16282	1505635
Lot Size	611265	6770	10822	5400	0	199940
Square Footage	611751	1646	694	1490	400	9943
Number Bedrooms	611779	2.93	1.13	3	0	8
Number Rooms	578448	6.72	1.99	6	1	18

4 A Dynamic Model of Neighborhood Choice

The previous literature that has explored the sorting of households across neighborhoods and communities has universally adopted a static approach.[8] We introduce the dynamics of the neighborhood choice problem through three channels: wealth accumulation, neighborhood dynamics, and moving costs. Households have expectations about appreciation of housing prices and may choose a neighborhood that offers lower per-period utility in the current period in return for the increase in wealth that would accompany price increases in that neighborhood. Similarly, households likely make trade-offs between current and future neighborhood attributes, choosing neighborhoods based in part on demographic or economic trends. The final component of the neighborhood choice problem that induces forward looking behavior on the part of households are moving costs. Because households typically pay 5-6 percent of the value of their house in real estate agent fees in addition to the non-financial costs of moving, it is clearly prohibitively costly to re-optimize every period. As a result, households will naturally account for their expectations of the future utility streams when deciding where to live.

We model households as making a sequence of location decisions that maximize the discounted

---

[8]See Epple and Sieg (1999), Bayer McMillan and Rueben (2004), Ferreyra (2006), Ekeland, Heckman, and Nesheim (2004) for some important recent examples. One exception is Kennan and Walker (2005), which analyzes interregional migration in the US in a dynamic context.

Figure 1:

Price levels by County
Bay Area, 1990–2004

Note: Real price levels. 1990 real price level normalized to one for each county

sum of expected per-period utilities. Our general model can be formulated in a familiar dynamic programming setup, where a Bellman equation illustrates the determinants of the optimal choice. We model households as choosing between neighborhoods, where a neighborhood is defined as a U.S. Census tract. Census tracts are small areas with approximately 1,500 housing units that are designed to be homogenous in terms of demographic characteristics.[9] Our data for the San Francisco Bay Area includes information on over one million house sales in approximately 800 census tracts between 1990 and 2004. Each period each household chooses whether to move or not. If they move, they incur a moving cost and then choose the neighborhood which yields the highest expected lifetime utility.

A key feature of our approach is that it controls for unobserved neighborhood heterogeneity in a dynamic model using a semi-parametric estimator that is computationally tractable. In addition, we have a novel way to capture the marginal utility of wealth that circumvents the

---

[9]See the Geographic Areas Reference Manual of the U.S. Census Bureau for more information.

traditional problem of the endogeneity of housing prices - thus avoiding the need to instrument for price. The model, as outlined below, temporarily abstracts from some important issues such as the decision whether to rent or to own as well as migration decisions. These are important features that will certainly be introduced into the model.

The observed state variables at time t are X_{jt}, Z_{it}, and H_{it}. X_{jt} is a vector of characteristics of the different choice options that affect the utility a household may receive from choosing neighborhood $j \in \{1, \ldots, J\}$. Z_{it} is a vector of characteristics of each household that potentially determine the per period utility from living in a particular neighborhood, as well as the costs associated with moving. For example, X may include variables such as price of housing, quality of local schools, or the average education level in the tract, and Z may include such variables as income, wealth, or race. Let H_{it} be another observable variable denoting the choice made in the previous period, i.e , $H_{it} = d_{it-1}$, where the decision variable, d_{it}, denotes the choice of household i in period t. Therefore, in the context of our model, H_{it} is the neighborhood in which household i resides before making a decision in period t.

In addition to the decision variable, d, and the observable variables, X_{jt}, Z_{it}, and H_{it}, there are three unobservable variables, ξ, ϵ_{ijt}, and ζ_{it}. We include and control for unobserved neighborhood characteristics, ξ.[10] ϵ_{ijt} is an idiosyncratic stochastic variable that determines the utility a household receives from living in neighborhood j and ζ_{it} affects moving costs. Note that we assume for simplicity that ζ_{it} is the same for all j. The decision variable, d_{it}, is given by the function $d_{it} = d(\cdot)$ where the arguments of $d(\cdot)$ are discussed below. For notational convenience, let $W_{ijt} = [X_{jt}, \xi_{jt}, Z_{it}]$, and let Ω_{it} denote an information set which includes all current characteristics, $\{W_{ijt}\}_{j=1}^{J}$ and anything that helps predict future characteristics.

The primitives of the model are $(\tilde{u}, p, \beta)$. $\tilde{u} = \tilde{u}(W_{ijt}, H_{it}, \zeta_{it}, \epsilon_{ijt})$ is the per period utility function, where the tilde denotes that this flow utility includes moving costs if applicable. $p = p(\Omega_{it+1}, H_{it+1}, \zeta_{it+1}, \epsilon_{it+1}, |\Omega_{it}, H_{it}, \zeta_{it}, \epsilon_{it}, j_{it})$ denotes the transition probabilities of the observables and unobservables. The transition probabilities are assumed to be Markovian. β is the discount factor.

Each household is assumed to behave optimally in the sense that its actions are taken to maximize

---

[10]We differ from previous work, such as Berry, Levinsohn, and Pakes (1995), that forces all individuals to have the same preferences for the unobserved neighborhood characteristic by allowing individuals to value the unobserved neighborhood characteristic differently depending on their demographic characteristics.

lifetime expected utility. d^* is the optimal decision rule and under the Markov structure of the problem is only a function of the state variables. That is, $d_{it} = d_{it}^*(\Omega_{it}, H_{it}, \zeta_{it}, \epsilon_{it})$. When the sequence of decisions, $\{d_i\}$, is determined according to the optimal decision rule, d^*, lifetime expected utility becomes the value function.

$$V_t = max_j\{E\sum_{s=t}^{T}\beta^s\big(\tilde{u}(W_{ijs}, H_{is}, \zeta_{is}, \epsilon_{is})\big)|\Omega_{it}, H_{it}, \zeta_{it}, \epsilon_{it}, d_{it} = j\} \tag{1}$$

We can break out the lifetime sum into the flow utility at time t and the expected sum of flow utilities from time $t+1$ onwards. This allows us to use the Bellman equation to express the value function at time t as the maximum of the sum of flow utility at time t and the discounted value function at time $t+1$.

$$V_t(\Omega_{it}, H_{it}, \zeta_{it}, \epsilon_{it}) = max_j\{\tilde{u}(W_{ijt}, H_{it}, \zeta_{it}, \epsilon_{it}) + E\beta V_{t+1}(\cdot)|\Omega_{it}, H_{it}, \zeta_{it}, \epsilon_{it}, d_{it} = j\} \tag{2}$$

We assume that the problem has an infinite horizon, $T = \infty$, which induces stationarity. By stationary, we mean $V_t(\Omega_{it}, H_{it}, \zeta_{it}, \epsilon_{it}) = V(\Omega_{it}, H_{it}, \zeta_{it}, \epsilon_{it})$ and $d_t(\Omega_{it}, H_{it}, \zeta_{it}, \epsilon_{it}) = d(\Omega_{it}, H_{it}, \zeta_{it}, \epsilon_{it})$. Under the assumptions of an infinite horizon and Markovian transition probabilities, we can rewrite the Bellman equation as:

$$V(\Omega_i, H_i, \zeta_i, \epsilon_i) = max_j\{\tilde{u}(W_{ij}, H_i, \zeta_i, \epsilon_i) + \beta\int V'(\cdot)p(d\Omega_i', dH_i', d\zeta_i', d\epsilon_i'|\Omega_i, H_i, \zeta_i, \epsilon_i, d_i = j)\} \tag{3}$$

Under certain technical assumptions, equation (3) is a contraction mapping in V. However, the difficulty is that V is a function of both the observed and unobserved state variables. Therefore, we follow Rust (1987) and make a series of assumptions which simplify the model. We make the assumptions that the flow utility is separable in the idiosyncratic error term and that this error term is distributed i.i.d. over time and options.

This allows us to recursively define the value function, $V(\Omega_i, H_i, \zeta_i, \epsilon_i)$, and the choice specific value function, $\tilde{v}_j(\Omega_i, H_i, \zeta)$.

$$V(\Omega_i, H_i, \zeta_i, \epsilon_i) = max_j[\tilde{v}_j(\Omega_i, H_i, \zeta_i) + \epsilon_{ij}] \tag{4}$$

$$\tilde{v}_j(\Omega_i, H_i, \zeta_i) = \tilde{u}(W_{ij}, H_i, \zeta_i) + \beta\int G(\cdot)\pi(d\Omega_i', dH_i', d\zeta_i'|\Omega_i, H_i, \zeta_i, d_i = j) \tag{5}$$

14

where $G(\cdot) = \int V(\Omega'_i, H_i, \zeta'_i, \epsilon'_i) q(d\epsilon'_i) = \int max_k[\tilde{v}_k(\Omega'_i, H'_i, \zeta'_i) + \epsilon'_{ik}] q(d\epsilon'_i)$

We break out the choice specific value function into two terms. The first term capturing the lifetime expected utility of choosing neighborhood j ignoring moving costs and the second term involves moving costs. The second term capturing the difference between the lifetime expected utility of choosing neighborhood j when the previous choice was j and when the previous choice was not j. In order to do this, we modify Rust's two assumptions.

Assumption (AS'): Additive Separability. We assume that the per period utility function can be broken out into two components: the flow utility from living in neighborhood j and a term that the household pays only if they move in that period. Therefore we can express $\tilde{u}(W_{ij}, H_i, \zeta_i) + \epsilon_{ij}$ as:

$$\tilde{u}(W_{ij}, H_i, \zeta_i) + \epsilon_{ij} = u(W_{ij}) - TC(Z_i, H_i, \zeta_i) \cdot I_{[j \neq H_i]} + \epsilon_{ij} \tag{6}$$

Assumption (CI'): Conditional Independence. We assume that the transition density for the Markov process $\{W, \epsilon, H, \zeta\}$ is given by:

$$p(d\Omega_{t+1}, d\epsilon_{t+1}, dH, d\zeta_{t+1}|W_t, \epsilon_t, \zeta_t, j_t) = q_\epsilon(d\epsilon_{t+1}) q_\zeta(d\zeta_{t+1}) q_H(dH_{t+1}|j_t) \pi(d\Omega_{t+1}|W_t, j_t) \tag{7}$$

Then it can be shown that with the exception of Z, the choice specific value function is separable in the variables that affect moving costs and those that affect the non-moving cost portion of per-period utility. Similarly to the flow utility, the tilde indicates that the choice specific value function incorporates possible moving costs.

$$\tilde{v}_j(\Omega_i, H_i, \zeta_i) = v_j(\Omega_i) - TC(Z_i, H_i, \zeta_i) \cdot I_{[j \neq H_i]} \tag{8}$$

5 Estimation

The estimation of the primitives of the model proceeds in three stages. In the first stage, we recover the non-moving cost component of lifetime expected utility. In the second stage, we recover moving costs and the marginal utility of wealth. While a number of standard options for estimating the marginal utility of wealth are available, we propose recovering the marginal utility of wealth by utilizing outside information on the financial costs of moves. Having recovered

moving costs and the marginal utility of wealth in the second stage, we estimate fully flexible estimates of the per-period utility in a final stage. With estimates of the per-period utility function it is straightforward to implement any of the applications discussed below. A key feature of our estimation strategy is its low computational burden.

5.1 Estimation - Stage One - Choice Specific Value Function

Consider the problem faced by a household that has chosen to move. It will choose the neighborhood $j \neq H$ which offers the highest utility by maximizing over the choice specific value functions $\tilde{v}$. Conditional on moving, the moving cost term, $TC(Z_i, H_i, \zeta_i) \cdot I_{[j \neq H_i]}$, is identical for all neighborhoods. As an additive constant, it simply drops out and, conditional on moving, each household chooses j to maximize:

$$v_j(\Omega_i) + \epsilon_{ij} = u(W_{ij}) + \beta \int \int \int G(\cdot) q_\zeta(d\zeta') q_H(dH', |j) \pi(dW', |W, j) + \epsilon_{ij} \qquad (9)$$

Under certain technical assumptions discussed in Rust 1994, we can show (9) is a contraction mapping with a unique fixed point v. Assuming that the idiosyncratic error term, ϵ_{ij}, is distributed i.i.d., Type 1 Extreme Value allows us to recover $v_j(\Omega_i)$ in a number of ways.

Previous methods for estimating dynamic discrete choice models in the presence of a large choice set will be plagued by a curse of dimensionality. We employ a variant of Hotz and Miller (1993) based on the contraction mapping in Berry (1994) which avoids this problem. Specifically, based on household characteristics such as income, wealth, and race, we divide households into distinct types indexed by τ. Let $\theta_{jt}^\tau = v_j(\Omega_i)$ when the characteristics of the household, Z_i, imply that they are of type τ. θ_{jt}^τ is then the choice specific value a household of type τ receives from choosing neighborhood j. Letting δ_{jt}^τ denote the deterministic component of flow utility for a household of type τ, we can rewrite (9) using lifetime utilities, θ_{jt}^τ.

$$\theta_{jt}^\tau = \delta_{jt}^\tau + \beta \int log\Big(exp(\theta_{jt+1}^{\tau'}) + \sum_{k \neq j} exp(\theta_{kt+1}^{\tau'} - TC^{\tau'} - \zeta_i')\Big) q(d\zeta') p(d\theta_{t+1}^{\tau'}|\theta_t) p(d\tau'|\tau, j) \quad (10)$$

Household i of type τ chooses neighborhood j if $\theta_j^\tau + \epsilon_{ij} > \theta_k^\tau + \epsilon_{ik} \forall k \neq j$. Therefore, the probability of any household of type τ choosing neighborhood j when ϵ_{ij} is distributed i.i.d.,

Type 1 Extreme Value can be expressed as:

$$P_j^\tau = \frac{e^{\theta_j^\tau}}{\sum_{k=1}^{J} e^{\theta_k^\tau}} \tag{11}$$

The vector of mean utilities, θ^τ, is unique up to an additive constant thus requiring some normalization for each τ. We temporarily normalize the mean (over neighborhoods) of the fixed effects to zero for each type in each time period. Denoting the number of types as M implies that we make M normalizations. Therefore, instead of recovering θ_j^τ for every neighborhood and type, we recover $\tilde{\theta}_j^\tau$ where $\tilde{\theta}_j^\tau = \theta_j^\tau - m^\tau$ and $m^\tau = 1/J \sum_j \theta_j^\tau$. Let S_j^τ and $S_j^\tau(\theta^\tau)$ denote the observed and predicted portion of households of type τ who reside in neighborhood j. $S_j^\tau(\theta^\tau)$ is given by P_j^τ. We can then easily calculate $\tilde{\theta}_j^\tau$ as:

$$\tilde{\theta}_j^\tau = log(S_j^\tau) - 1/J \sum_k log(S_k^\tau) \tag{12}$$

As the number of types, M, grows large relative to the sample size, we may face some small sample issues with observed shares. Therefore, instead of simply calculating observed shares as the portion of households of a given type who live in an area, we use a weighted measure to avoid zero shares. We do this to incorporate the information from those of a similar types when calculating shares for any given type. For example, if we want to calculate the share of households with an income of \$50,000 choosing neighborhood j, we would use some information about the residential decisions of those earning \$45,000 or \$55,000. Naturally, the weights will depend on how far away the other types are in type space. We denote the weights by $W^\tau(Z_i)$. The formula for calculating observed shares is given by:[11]

$$S_j^\tau = \frac{\sum_{i=1}^{N} I_{[d_i=j]} \cdot W^\tau(Z_i)}{\sum_{i=1}^{N} W^\tau(Z_i)} \tag{13}$$

where the weights are constructed as the product of K kernel weights, where K is the dimension of Z. Each individual kernel weight is formed using a standard normal kernel, N, and bandwidth, h_k.

$$W^\tau(Z_i) = \prod_{k=1}^{K} \frac{1}{h} N\left(\frac{Z_i - Z^\tau}{h_k}\right) \tag{14}$$

---

[11] If $W^\tau(Z_i) = I_{[Z_i=Z^\tau]}$, this results in the standard way for calculating shares.

5.2 Estimation - Stage Two - Moving Costs and the Marginal Utility of Wealth

Households behave dynamically by taking into account the effect their current decision has on future utility flows. In our model, the current decision affects future utility flows through two channels. Households are aware they will incur a transaction cost by re-optimizing in the future. In addition, the decision about where to live today affects wealth in the future. Equation (10) shows how the current action impacts both today's flow utility and the future utility. It also suggests that if θ_{jt}^{τ} (or $\tilde{\theta}_{jt}^{\tau}$) is known for all τ and j, we can estimate moving costs based on households decisions to move or stay in a given period.

Given estimates of $\tilde{\theta}_{jt}^{\tau}$ from the first stage, we can estimate moving costs in stage two by considering the move/stay decisions of households. From the model outlined above, we know that in any given period a household will move if the lifetime expected utility of staying in their current neighborhood is less than the lifetime expected utility of the best other alternative when moving costs are factored in.

We assume that moving costs, TC, are composed of financial costs, $F(H)$ and psychological costs, $\psi(Z_i) + \zeta_i$. The financial moving costs are a function of H as households pay financial costs based primarily on the property the sell. The psychological costs are a function of the observable characteristics that define type, Z, as well as the unobserved stochastic component, ζ_i. As the financial moving costs reduce wealth, choosing to move changes a households type. For example, if moving costs are \$10,000, then a given household with \$100,000 in wealth chooses where to live based on the utility of staying in their current neighborhood with wealth of \$100,000 and the highest alternative utility with a wealth of \$90,000. In practice, we treat financial moving costs as observable and set them equal to 6% of the value of housing in the neighborhood a household is leaving, *ie* $F(H_i) = 0.06 \cdot Price_{H_i}$

If a household of type τ living in neighborhood j moves, we denote their new type as $\bar{\tau}_j$. The new type following a move reflects the reduction in wealth by the amount of $F(H)$.

A household who chose j in the previous period, i.e. $H_i = j$, will choose to stay if:

$$Max_{k \neq j}[\theta_k^{\bar{\tau}_j} + \epsilon_{ik}] - (\psi(Z_i) + \zeta_i) < \theta_j^{\tau} + \epsilon_{ij} \tag{15}$$

However, from the first stage we only recover the demeaned choice specific value functions, $\tilde{\theta}_j^\tau$, where $\tilde{\theta}_j^\tau = \theta_j^\tau - m^\tau$. We can then rewrite (15) as:

$$Max_{k \neq j}[\tilde{\theta}_k^{\bar{\tau}_j} + \epsilon_{ik}] - (m^\tau - m^{\bar{\tau}_j}) - (\psi(Z_i) + \zeta_i) < \tilde{\theta}_j^\tau + \epsilon_{ij} \tag{16}$$

The term $m^\tau - m^{\bar{\tau}_j}$ is unobserved but can be estimated. In principle, we could estimate a separate term for each combination of τ and $F(H)$, however, we choose to flexibly parameterize it as a function of Z and $F(H_i)$. Recall that $m^\tau = 1/J \sum_j \theta_j^\tau$ and, as such, $m^\tau - m^{\bar{\tau}_j}$ is the difference (averaged across neighborhoods) between having the utility associated with being type τ and the having the utility from the reduced wealth after paying the financial moving costs.

Note that the three stochastic terms are $Max_{k \neq j}[\tilde{\theta}_k^{\bar{\tau}_j} + \epsilon_{ik}]$, ϵ_{ij}, and ζ_i. We estimate $m^\tau - m^{\bar{\tau}_j}$ and $\psi(Z)$ from a likelihood function based on the probability of a household staying in its current house

$$P_i^\tau(Stay|H_i = j) = \int_{-\infty}^{\infty} \frac{e^{\tilde{\theta}_j^\tau}}{e^{\tilde{\theta}_j^\tau} + \sum_{k \neq j} e^{\tilde{\theta}_k^{\bar{\tau}_j} - (m^\tau - m^{\bar{\tau}_j}) - \psi(Z) - \zeta_i}} \cdot \phi(\zeta_i) d(\zeta_i) \tag{17}$$

The first stage of our estimation approach involved making a normalization for each type of household (i.e., $\tilde{\theta}_j^\tau$ is mean zero across all locations j), where type could be defined by personal characteristics such as race, income, wealth. Once we set the mean choice specific utility from no wealth to zero, we only need to know these baseline differences, $m^\tau - m^{\bar{\tau}_j}$, to recover the unnormalized choice specific value functions. As we can estimate the baseline differences, we can simply recover the true choice specific value functions as $\theta_j^\tau = \tilde{\theta}_j^\tau + m^\tau$.

It is important to recover these baseline differences because they represent the extra utility a household would receive from extra wealth. A key aspect of the dynamic model is that the choice of neighborhood affects future type. Therefore, the baseline differences in utility across types represent potential future utility gains from wealth accumulation.

5.3 Estimation - Stage Three - Per-Period Utility

From stages one and two, we know the distribution of moving costs for each type, the marginal value of changing type and the true mean utility terms, θ_j^τ. We can then estimate the transition probabilities $p(d\theta_{t+1}^{\tau'}|\theta_t)$ and $p(d\tau'|\tau, j)$. In theory, we could estimate the transition probabilities

fully non-parametrically, as we have a time series for each type and neighborhood. However, to increase the efficiency of our estimates of the transition probabilities, we can impose some symmetry restrictions on the transition probabilities. For example, within each type we could assume that the neighborhood mean utilities, θ_{jt}^τ, evolve according to an auto-regressive process where some of the coefficients are common across neighborhoods.

In practice, we estimate transition probabilities separately for each type but pool information over neighborhoods. To account for different means and trends we include a separate constant and time trend for each neighborhood's choice specific value function for each type. We assume the transition of the choice specific value functions, θ_{jt}^τ, is given by:[12]

$$\theta_{jt}^\tau = \sum_{l=1}^{L} \alpha_{1,l}^\tau \theta_{jt-l}^\tau + \sum_{l=1}^{L} \alpha_{2,l}^\tau \prime X_{jt-l} + \kappa_{0,j} + \kappa_{1,j}\, t + \varepsilon_{jt}^\tau \tag{18}$$

We also need to know how housing wealth transitions to specify transition probabilities for types, $p(d\tau'|\tau, j)$. We use sales data to construct prices indexes for each type, tract, year combination. With these price indexes we use a similar method to the choice specific value functions, θ_{jt}^τ, to estimate transition probabilities on price levels. Given transition probabilities on price levels it is straightforward to estimate transition probabilities for wealth and type, τ.

Knowing θ^τ, ψ^τ, $p(d\theta_{t+1}^{\tau\prime}|\theta_t)$, and $p(d\tau'|\tau, j)$, allows us to calculate mean flow utilities for each type and neighborhood, δ_{jt}^τ, according to:

$$\delta_{jt}^\tau = \theta_{jt}^\tau - \beta \int log\Big(exp(\theta_{jt+1}^{\tau\prime}) + \sum_{k \neq j} exp(\theta_{kt+1}^{\bar{\tau}_j'} - \psi^{\tau\prime} - \zeta_i')\Big) q(d\zeta') p(d\theta_{t+1}^{\tau\prime}|\theta_t) p(d\tau'|\tau, j) \tag{19}$$

For each type, τ, neighborhood, j, and time, t, we have the necessary information to calculate the integral on the right hand side of (19). It is then straightforward to recover the $M \cdot J \cdot T$ values for the mean flow utilities, δ_{jt}^τ.

Once we recover the mean per-period utilities, we can decompose them into functions of the observable neighborhood characteristics, X_{jt}. We assume that ξ is uncorrelated with the other

---

[12]Depending on the number of regressors, we could make this specification more flexible by allowing the coeffients on the lags to be functions of the right-hand side variables. A straightforward way to do this would be to first detrend the θs and then use the local linear estimator of Fan (1992). Given the potentially large number of regressors we could follow Bajari and Khan (2005) and interpret the regression as flexible rather than truly non-parametric.

neighborhood characteristics and treat it as an error term in the following regression.

$$\delta_{jt}^{\tau} = g(X_{jt}; \chi) + \xi_{jt}^{\tau} \tag{20}$$

where $g(X_{jt}; \chi)$ is a flexible function of X_{jt} known up to parameter χ. This decomposition of the mean flow utilities is similar to Berry, Levinsohn, and Pakes (1995) or Bayer, McMillan, and Rueben (2004) with one important difference. In these models it was necessary to instrument for price in the regression equation (20). In our approach, we already know the coefficient on price as we have previously calculated the marginal utility of wealth.

6   Results

7   Testable Implications of Rational Expectations

In this section, we use the dynamic model proposed above to develop testable implications of housing market efficiency and in particular rational expectations. We begin by showing that when the model is restricted so that all households have identical preferences, rational expectations implies the absence of predictable returns, i.e., the absence of the positive persistence in housing prices shown in the literature following Case and Shiller (1989).

The easiest way to understand the implications of identical preferences for the model is to work through the equation that links lifetime and per-period utility, equation (19). First notice that, identical preferences imply that the lifetime utility provided by all neighborhoods must be identical in any given period. That is, the set of current home-buyers must get the same indirect utility from each available choice. Thus, the lifetime utility term (θ) effectively drops out of equation (19) - leaving a direct relationship between current per-period utility and the expectation of how lifetime utility and wealth are expected to evolve over the next period. Then notice that the lifetime utility provided by each neighborhood in the next period will again be identical. Thus, equation (19) essentially reduces to a relationship between current per-period utility and expected changes in wealth over the next period. In this way, any neighborhood with higher than average expected appreciation over the next period must simultaneously provide lower current per-period utility. Because this offset in current per-period utility comes by way

of higher prices today, given identical preferences and rational expectations, current prices must directly reflect any expected appreciation - and lagged appreciation should provide no predictive power regarding future appreciation.

The simple intuition derived from the extreme assumption of identical preferences suggests more generally that as the houses considered in an analysis are closer substitutes for one another, the predictability of returns should fall to zero. In this way, while significant short-run predictable differences in expected returns might arise between two distant cities because the households in each of those cities are strongly attached to their respective labor markets in the short run, any predictable price differences between two neighborhoods that are reasonably close substitutes within a single city should be immediately arbitraged away by the set of current home buyers considering those neighborhoods.

We examine this hypothesis empirically by studying the dynamics of housing prices at various levels of aggregation across both geographic and socioeconomic dimensions. Tables 4 and 5 report parameter estimates from a series of instrumental variables regressions of annual appreciation in house prices on lagged appreciation at various levels of geographic aggregation. The estimating equation can be written:

$$A_t^j = \beta_i A_{t-i}^j + \varepsilon_t^j \tag{21}$$

where j indicates the geographic area, t indicates the year, A_{t-i}^j is a vector of lagged appreciation measures and β_i is the coefficient on the ith lag. For each geographic level, results are reported for specifications that include one and two lags, respectively. For geographic levels below the metropolitan area, results are reported for specifications that include year dummies interacted with higher levels of geographic aggregation.

To address a series of measurement error problems related to the construction of price indices and appreciation measures, we follow the procedure outlined in Case and Shiller (1989). In particular, we first split the full sample of houses for which we have repeat sales information randomly into two subsamples (A and B). For each subsample, we estimate annual price indices for each geographic area using a repeat sales specification; the price index for each geographic area is derived from the coefficients on year dummies interacted with that geographic area in a log price regression that also includes fixed effects for each house. For each IV regression reported in the tables, the estimated lagged appreciation in the sample listed in the row heading

is instrumented with the lagged appreciation in the other sample. [13]

7.1 Discussion of Rational Expectations Results

The first column of Table 4 reports results when the entire San Francisco Bay Area is used as the geographic level. Aggregating in this way limits the number of observations in the 1-lag specifications to twelve - one less than the number of years that in the sample. The main point of this column is to show that the San Francisco Bay Area in this time period displays the positive persistence (although slightly smaller in magnitude) generally estimated at the metropolitan area level in previous studies. The following two columns in Table 4 show analogous results when the geographic unit is the county. The second of these columns includes year fixed effects, which control for the overall metropolitan level positive persistence.

These parameter estimates reveal that the level of positive persistence exhibited for the metropolitan area as a whole remains equally strong or even stronger across counties within the metropolitan area. The final three columns of Table 4 show analogous results when the geographic unit is the Census PUMA; there are 42 PUMAs in the Bay Area sample. The second of these columns again includes year dummies and the final column includes year*county dummies. The parameter estimates in the second column again reveal that the level of positive persistence exhibited for the metropolitan area as a whole remains equally strong or even stronger across PUMAs within the metropolitan area. Even within counties, there is substantial positive persistence in appreciation across PUMAs.

Overall, the pattern of results presented in Table 4 provides little evidence for decreasing positive persistence in a within- versus across-metropolitan analysis given rational expectations. Table 5 presents a series of results when the geographic unit is the Census tract, restricting attention for the time being to the 178 largest tracts - these average at least 40 transactions per year. At this level of geographic aggregation, a new pattern emerges. In particular, while positive persistence continues to hold across tracts within the metropolitan area as a whole, the pattern of dynamic dependence looks much different across tracts within counties or within PUMAs. In these cases, the coefficients are significantly negative. Thus, tracts with a relatively high appreciation rate

---

[13]It is clearly possible to significantly improve the efficiency of the underlying repeat sales regressions and the IV regressions reported in Tables 4 and 5. Splitting the sample in two, for example, is a rather inefficient way of addressing the measurement error issues associated with the fact that the appreciation measures used on both sides of our main estimating equations are estimated and not known exactly.

Table 4: IV Regressions of Current Appreciation on Lagged Appreciation at Various Geographic Levels

| Dependent Variable: | | Annual Appreciation in House Prices at t | | | | | |
Geographic Level:		Metro Area	County		Census PUMA		
1-Lag Specifications							
Sample A	Appreciation at t-1	0.58	0.56	1.01	0.54	0.84	0.41
		0.24	*0.06*	*0.11*	*0.03*	*0.05*	*0.10*
Sample B	Appreciation at t-1	0.58	0.57	1.05	0.54	0.85	0.33
		0.24	*0.06*	*0.11*	*0.03*	*0.05*	*0.12*
Observations		12	72	72	564	564	564
2-Lag Specifications							
Sample A	Appreciation at t-1	0.38	0.39	0.43	0.40	0.56	0.47
		0.34	*0.07*	*0.14*	*0.04*	*0.09*	*0.14*
	Appreciation at t-2	0.13	0.11	0.10	0.09	-0.07	-0.07
		0.32	*0.02*	*0.15*	*0.02*	*0.09*	*0.14*
Sample B	Appreciation at t-1	0.37	0.40	0.48	0.37	0.45	0.37
		0.34	*0.08*	*0.07*	*0.04*	*0.08*	*0.17*
	Appreciation at t-2	0.14	0.09	-0.02	0.12	0.04	0.01
		0.32	*0.03*	*0.12*	*0.03*	*0.08*	*0.14*
Observations		11	66	66	517	517	517
Includes:							
Year Dummies				X		X	
County*Year Dummies							X

Note: The full sample of houses for which we have repeat sales information was initially split randomly into two subsamples (A and B). For each subsample, price indices at the geographic level listed in the column heading were estimated using a repeat sales specification. The table reports results from a series of IV regressions of annual appreciation in house prices on lagged appreciation, where the estimated lagged appreciation in the sample listed in the row heading is instrumented with the lagged appreciation in other sample. For geographic levels below the metropolitan area, results are reported for specifications that include year dummies for higher levels of geographic aggregation. Standard errors were corrected for clustering at the level of geography reported in the column heading and are reported in italics.

in a given year generally experience slower rates of appreciation the following year than the other tracts in the same county or PUMA. When combined with the positive persistence at

Table 5: IV Regressions of Current Appreciation on Lagged Appreciation at Various Geographic Levels

| Dependent Variable: | Annual Appreciation in House Prices at t | | | |
| Geographic Level: | Census Tract | | | |

1-Lag Specifications

Sample A	Appreciation at t-1	0.38	0.20	-0.21	-0.30
		0.03	*0.05*	*0.05*	*0.06*
Sample B	Appreciation at t-1	0.41	0.26	-0.16	-0.25
		0.02	*0.04*	*0.05*	*0.06*
Observations		2136	2136	2136	2136

2-Lag Specifications

Sample A	Appreciation at t-1	0.23	0.01	-0.19	-0.31
		0.03	*0.04*	*0.05*	*0.06*
	Appreciation at t-2	0.19	0.12	0.02	-0.07
		0.03	*0.04*	*0.05*	*0.06*
Sample B	Appreciation at t-1	0.26	0.04	-0.17	-0.31
		0.03	*0.04*	*0.05*	*0.07*
	Appreciation at t-2	0.14	0.06	-0.05	-0.17
		0.02	*0.03*	*0.04*	*0.04*
Observations		1958	1958	1958	1958
Includes:					
Year Dummies			X		
County*Year Dummies				X	
PUMA*Year Dummies					X

Note: The full sample of houses for which we have repeat sales information was initially split randomly into two subsamples (A and B). For each subsample, price indices at the geographic level listed in the column heading were estimated using a repeat sales specification. The table reports results from a series of IV regressions of annual appreciation in house prices on lagged appreciation, where the estimated lagged appreciation in the sample listed in the row heading is instrumented with the lagged appreciation in other sample. For geographic levels below the metropolitan area, results are reported for specifications that include year dummies for higher levels of geographic aggregation. Standard errors were corrected for clustering at the level of geography reported in the column heading and are reported in italics

the PUMA and county level, the results as a whole are consistent with a spatial propagation of shocks, whereby a positive shock in a given tract is sent along to other nearby tracts in the following period. Appreciation in these nearby tracts can simultaneously explain the negative

persistence across tracts within PUMAs and the positive persistence for the PUMA as a whole.

Overall the magnitude and statistical significance of the parameter estimates presented in Tables 4 and 5 appears to be inconsistent with rational expectations in its purest form. The particular way in which rational expectations appears to fail here may be related to the way in which home buyers and sellers actually gather information from the market. In particular, it is well known that the heterogeneous nature of housing makes it difficult for both buyers and sellers to ascertain what an appropriate market price for a home should be. In most cases, recent sales of comparable properties are used by individuals on both sides of the market to help with this process. The use of comparable recent sales would give rise to a pattern for the propagation of a local shock very much like the one characterized in tables 4 and 5.

While the results presented in Tables 4 and 5 are very suggestive, they should be treated as preliminary, as more can be done to make them more precise. We are currently working on estimating specifications that (i) characterize the spatial pattern of appreciation across tracts more carefully, (ii) more efficiently deal with the series of measurement error issues mentioned above than the two subsample method used here, (iii) examine the pattern of dynamic dependence across tracts that are not only close in geographic space but also close in socioeconomic space, (iv) aggregate smaller tracts into larger ones rather than dropping them - so as to more completely utilize the data in the sample.

7.2 A Structural Test of Rational Expectations

We can also use the structural model directly to test for rational expectations. The test is quite straightforward in the context of our model. Included in the information set Ω_t are current prices and neighborhood characteristics, as well as lagged values of housing prices and other neighborhood characteristics. The lagged values are included as they help predict future values. Lagged characteristics, however, should enter the choice specific value function, $v_j(\Omega_i)$, only through the expected continuation value. If rational expectations hold, the flow utilities, δs, that we recover should not be a function of any lagged characteristics. Testing for rational expectations can then be simplified to testing whether or not the δs are a function of anything other than the variables that affect per-period utility. While this test requires the assumption that the model has been specified correctly, the fact that our estimation approach identifies

per-period utility in a flexible semiparametric way provides a good deal of assurance regarding about the validity of our proposed test.

8   Conclusion

References

ABRAHAM, J., AND P. HENDERSHOTT (1996): "Bubbles in Metropolitan Housing Markets," *Journal of Housing Research*, 7(2), 191–207.

BAJARI, P., C. L. BENKARD, AND J. KRAINER (2005): "House Prices and Consumer Welfare," *Journal of Urban Economics*, 58, 474–487.

BAJARI, P., AND M. KAHN (2005): "Estimating Housing Demand With An Application to Explaining Racial Segregation in Cities," *Journal of Business and Economic Statistics*, 23(1), 20–33.

BAYER, P., F. FERREIRA, AND R. MCMILLAN (2007): "A Unified Framework for Measuring Preferences for Schools and Neighborhoods," mimeo, Duke University.

BAYER, P., AND R. MCMILLAN (2006): "Racial Sorting and Neighborhood Quality," NBER Working Paper No. 11813.

BAYER, P., R. MCMILLAN, AND K. RUEBEN (2004): "An Equilibrium Model of Sorting in an Urban Housing Market," NBER Working Paper No. 10865.

BAYER, P., AND C. TIMMINS (2005): "On the Equilibrium Properties of Locational Sorting Models," *Journal of Urban Economics*, 57, 462–477.

——— (2007): "Estimating Equilibrium Models of Sorting Across Locations," *Economic Journal*, Forthcoming.

BERRY, S. (1994): "Estimating Discrete Choice Models of Product Differation," *Rand Journal of Economics*, 25, 242–262.

BERRY, S., J. LEVINSOHN, AND A. PAKES (1995): "Automobile Prices in Market Equilibrium," *Econometrica*, 63, 841–890.

CAPOZZA, D., P. HENDERSHOTT, C. MACK, AND C. MAYER (2002): "Determinants of Real House Price Dynamics," NBER Working Paper No. 9262.

CARD, D., A. MAS, AND J. ROTHSTEIN (2006): "Tipping and the Dynamics of Segregation in Neighborhoods and Schools," mimeo.

CARRANZA, J. E. (2007): "Estimation of demand for dierentiated durable goods," mimeo.

CASE, K., AND C. MAYER (1995): "Housing Price Dynamics within a Metropolitan Area," NBER Working Paper No. 5182.

CASE, K., AND R. SHILLER (1989): "The Efficiency of the Market for Single-Family Homes," *American Economic Review*, 79(1), 125–137.

CHO, M. (1996): "House Price Dynamics: A Survey of Theoretical and Empirical Issues," *Journal of Housing Research*, 7(2), 145–172.

CLAYTON, J. (1997): "Are Housing Prices Driven by Irrational Expectations?," *Journal of Real Estate Finance and Economics*, 14(3), 341–63.

CUTLER, D., E. GLAESER, AND J. VIGDOR (1999): "The Rise and Decline of the American Ghetto," *Journal of Political Economy*, 107, 455–506.

EASTERLY, W. (2004): "Empirics of Strategic Interdependence: The Case of the Racial Tipping Point," NYU Development Research Working Paper No. 5.

EKELAND, I., J. J. HECKMAN, AND L. NESHEIM (2004): "Identification and Estimation of Hedonic Models," *Journal of Political Economy*, 112(1), S60–S109.

ELLISON, G., AND E. GLAESER (1997): "Geographic Concentration in U.S. Manufacturing Industries: A Dartboard Approach," *Journal of Political Economy*, 105(5), 889–927.

ENGLUND, P., T. GORDON, AND J. QUIGLEY (1999): "The Valuation of Real Capital: A Random Walk down Kungsgatan," *Journal of Housing Economics*, 8, 205–216.

EPPLE, D., AND H. SIEG (1999): "Estimating Equilibrium Models of Local Jurisdictions," *Journal of Political Economy*, 107, 645–681.

ERDEM, T., S. IMAI, AND M. P. KEANE (2003): "Brand and Quantity Choice Dynamics Under Price Uncertainty," *Quantitative Marketing and Economics*, 1(1), 5–64.

FAN, J. (1992): "Design-Adaptive Non-Parametric Regression," *Journal of the American Statistical Assosciation*, 87(420), 998–1004.

FERREYRA, M. (2007): "Estimating the Effects of Private School Vouchers in Multi-District Economies," *American Economic Review*, Forthcoming.

GELFAND, A., M. ECKER, J. KNIGHT, AND C. SIRMANS (2004): "The Dynamics of Location in Home Price," *Journal of Real Estate Finance and Economics*, 29(2), 149–166.

GLAESER, E., AND J. GYOURKO (2006): "Housing Dynamics," mimeo.

GOWRISANKARAN, G., AND M. RYSMAN (2006): "Dynamics of Consumer Demand for New Durable Goods," mimeo.

HENDEL, I., AND A. NEVO (2006): "Measuring the Implications of Sales and Consumer Inventory Behavior," *Econometrica*, 74(6), 1637–1673.

HIMMELBERG, C., C. MAYER, AND T. SINAI (2005): "Assessing High House Prices: Bubbles, Fundamentals, and Misperceptions," *Journal of Economic Perspectives*, 19(4), 67–92.

HOTZ, V. J., AND R. A. MILLER (1993): "Conditional Choice Probabilities and the Estimation of Dynamic Models," *Review of Economic Studies*, 60, 497–529.

HOTZ, V. J., R. A. MILLER, S. SANDERS, AND J. SMITH (1994): "A Simulation Estimator for Dynamic Models of Discrete Choice," *Review of Economic Studies*, 61, 265–289.

KENNAN, J., AND J. R. WALKER (2005): "The Effect of Expected Income on Individual Migration Decisions," NBER Working Paper No. 9585.

LAMONT, O., AND J. STEIN (2004): "Leverage and House-Price Dynamics in U.S. Cities," mimeo, University of Chicago GSB.

McCARTHY, J., AND R. PEACH (2004): "Are Home Prices the Next Bubble?," *Economic Policy Review*, 10(3), 1–17.

MELNIKOV, O. (2001): "Demand for Differentiated Durable Products: The Case of the U.S. Computer Printer Market," mimeo.

MERLO, A., F. ORTALO-MAGNÉ, AND J. RUST (2006): "Models of Bargaining and Price Determination of Residential Real Estate, with and without Real Estate Agents," mimeo.

ORTALO-MAGNÉ, F., AND S. RADY (1999): "Boom In, Bust Out: Young Households and the Housing Price Cycle," *European Economic Review*, 43, 755–766.

——— (2005): "Heterogeneity Within Communities: A Stochastic Model with Tenure Choice," CESifo Working Paper No. 1465.

——— (2006): "Housing Market Dynamics: On the Contribution of Income Shocks and Credit Constraints," *Review of Economic Studies*, 73(2), 459–485.

ROBACK, J. (1982): "Wages, Rents, and the Quality of Life," *Journal of Political Economy*, 90(6), 1257–1278.

ROSENTHAL, S. (1999): "Residential Buildings and the Cost of Construction: New Evidence on the Efficiency of the Housing Market," *Review of Economics and Statistics*, 81(2), 288–302.

RUST, J. (1987): "Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher," *Econometrica*, 55(5), 999–1033.

——— (1994): "Structural Estimation of Markov Decision Processes," in *Handbook of Econometrics*, ed. by R. Engle, and D. McFadden, vol. 4. New York: North Holland.

SCHELLING, T. (1971): "Dynamic Models of Segregation," *Journal of Mathematical Sociology*, pp. 143–186.

SCHIRALDI, P. (2006): "Automobile Replacement: A Dynamic Structural Approach," mimeo, Boston University.

SHILLER, R. (2005): *Irrational Exuberance*. Princeton University Press, 2nd edn.

——— (2006): "Long-Term Perspectives on the Current Boom in Home Prices," *The Economists' Voice*, 3(4), Article 4.

TIMMINS, C. (2006a): "Estimable Equilbrium Models of Locational Sorting and Their Role in Development Economics," *Journal of Economic Geography*, 5, 83–100.

——— (2006b): "Estimating Spatial Differences in the Brazilian Cost-of-Living With House-hold Location Chocies," *Journal of Development Economics*, 80(1), 59–83.

——— (2007): "If You Can't Take the Heat, Get Out of the Cerrado... Recovering the Equilibrium Amenity Cost of Non-Marginal Climate Change in Brazil," *Journal of Regional Science*, Forthcoming.

TIMMINS, C., AND J. MURDOCK (2007): "A Revealed Preference Approach to the Measurement of Congestion in Travel Cost Models," *Journal of Environmental Economics and Management*, Forthcoming.

TRACY, J., H. SCHNEIDER, AND S. CHAN (1999): "Are Stocks Overtaking Real Estate in Household Portfolios?," *Current Issues in Economics and Finance*, 5(5), 1–6.