NBER WORKING PAPER SERIES

A DYNAMIC MODEL OF DEMAND FOR HOUSES AND NEIGHBORHOODS

Patrick Bayer
Robert McMillan
Alvin Murphy
Christopher Timmins

Working Paper 17250
http://www.nber.org/papers/w17250

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
July 2011

This paper is a revised version of NBER Working Paper No. 17025. We would like to thank Kelly Bishop, Morris Davis, Ed Glaeser, Phil Haile, Aviv Nevo, participants at the Econometric Society Summer Meetings, NBER Summer Institute, Regional Science Annual Meetings, Stanford Institute for Theoretical Economics, and seminar participants at the University of Arizona, Duke, UBC, Georgetown, Minnesota, NYU, Northwestern, Ohio State, Queen's, Rochester, St. Louis Federal Reserve, and Yale for many valuable suggestions. Co-editor Jean-Marc Robin and three anonymous referees provided numerous comments that have helped us improve the paper significantly. Thanks to Elliot Anenberg for excellent research assistance. Financial support from the National Science Foundation and SSHRC is gratefully acknowledged. All remaining errors are our own. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

A Dynamic Model of Demand for Houses and Neighborhoods
Patrick Bayer, Robert McMillan, Alvin Murphy, and Christopher Timmins
NBER Working Paper No. 17250
July 2011, Revised February 2015
JEL No. H0,H23,H41,H7,L85,R0,R14,R21,R31,R51

ABSTRACT

This paper develops a dynamic model of neighborhood choice along with a computationally light multi-step estimator. The proposed empirical framework captures observed and unobserved preference heterogeneity across households and locations in a flexible way. The model is estimated using a newly assembled data set that matches demographic information from mortgage applications to the universe of housing transactions in the San Francisco Bay Area from 1994- 2004. The results provide the first estimates of the marginal willingness to pay for several non-marketed amenities – neighborhood air pollution, violent crime and racial composition – in a dynamic framework. Comparing these estimates with those from a static version of the model highlights several important biases that arise when dynamic considerations are ignored.

Patrick Bayer
Department of Economics
Duke University
213 Social Sciences
Durham, NC 27708
and NBER
patrick.bayer@duke.edu

Robert McMillan
University of Toronto
Department of Economics
150 St. George Street
Toronto, ON M5S 3G7
CANADA
and NBER
mcmillan@chass.utoronto.ca

Alvin Murphy
Department of Economics
Arizona State University
PO Box 879801
Tempe, AZ 85287-9801
Alvin.Murphy@asu.edu

Christopher Timmins
Department of Economics
Duke University
209 Social Sciences Building
P.O. Box 90097
Durham, NC 27708-0097
and NBER
christopher.timmins@duke.edu

1 Introduction

Models of residential sorting and hedonic equilibrium provide the basis for a number of long-standing literatures in economics. A large theoretical literature in public and urban economics, for example, has used these models to characterize the equilibrium structure of cities and the provision of public goods in a system of political jurisdictions.[1] And empirical researchers have developed related models in order to provide theoretically-consistent estimates of household willingness to pay for a wide variety of non-marketed local goods (e.g., education, crime, and environmental amenities)[2] and as a tool for simulating how counterfactual policies would affect the housing market equilibrium, residential sorting, and, ultimately, household welfare.[3]

Despite making progress along many important dimensions, nearly all of the models developed and estimated in these literatures have adopted a static approach that assumes agents are not forward-looking. This has raised a concern, shared by critics and practitioners alike,[4] that empirical findings from static models might be subject to biases related to the inherently dynamic nature of household location decisions.[5]

That location decisions are dynamic follows directly from a number of important features of the housing market: (i) large transactions costs that make moves relatively rare, (ii) changing household tastes and needs over the life-cycle, and (iii) evolving local amenities and housing prices that give neighborhoods a dynamic character. Despite housing markets having these obvious characteristics, estimating dynamic models of location choice has proven difficult for two primary reasons. The first concerns data:

[1]Theoretical contributions to the residential sorting literature include papers by Ellickson (1971), Epple, Filimon, and Romer (1984), Epple and Romer (1991), Epple and Romano (1998), and Nechyba (1999, 2000). Related contributions to the literature analyzing hedonic equilibrium include Rosen (1974), Epple (1987), and Ekeland, Heckman, and Nesheim (2004).

[2]Empirical sorting papers include Epple and Sieg (1999), Bayer, Ferreira, and McMillan (2007), Ferreyra (2007), and Kuminoff (2008); for empirical analyses of hedonic equilibrium, see Bajari and Kahn (2005), Kuminoff and Jarrah (2010), and Bishop and Timmins (2012).

[3]See, for example, Sieg, Smith, Banzhaf, and Walsh (2004), Epple, Romano, and Sieg (2006), Walsh (2007), and Bayer, McMillan, and Rueben (2011).

[4]The dynamic nature of location decisions is often acknowledged by researchers and has, for instance, prompted a debate as to whether all households or just recent movers should be used when estimating preferences for amenities. See the discussion in Kahn (1995), Cragg and Kahn (1997), or Bayer, Keohane, and Timmins (2009).

[5]Static models are also limited in that they cannot distinguish how households value permanent versus temporary changes in amenities.

the estimation of sorting models typically requires the matching of a large sample of households along with their characteristics to the location and features of their housing choices. Given this core data need, most papers in the prior literature have used data from the decennial Census, which provides great detail about large cross-sections of households but very little information about the dynamics of decision-making or the continued evolution of households and neighborhoods.[6]

The second factor that makes estimating dynamic models difficult is the high dimensionality of the state space required to characterize the evolution of a system of neighborhoods (or cities). The resulting curse of dimensionality has made it exceedingly difficult to compute a reasonable dynamic model of residential location decisions that takes account of the heterogeneous and evolving nature of both households and neighborhoods.

The main goal of this paper is to provide a new approach for estimating a dynamic model of demand for houses and neighborhoods that is computationally tractable. The starting point for our analysis is a newly assembled data set that links information about buyers and sellers to the universe of housing transactions in the San Francisco metropolitan area over a period of 11 years. In addition to providing precise information about housing structure (e.g., square footage, year built, lot size, transaction price) and house location, a key feature of these data is that they provide important demographic and economic information about the buyers and sellers themselves, permitting us to follow households over time as they move within the metropolitan area.

With these data in hand, we develop a model of neighborhood choice in a dynamic setting, along with a multi-step estimation approach that is computationally light. This approach, which combines and extends the insights of Rust (1987), Berry (1994), and Hotz and Miller (1993), allows both the observed and unobserved features of each neighborhood to evolve over time in a completely flexible manner. It makes use of information about neighborhood choices and the timing of moves to recover: (i) preferences for hous-

[6]Recent papers by Epple, Romano, and Sieg (2010) and Caetano (2010) have made useful advances in the estimation of dynamic models, exploiting the limited dynamic information in the Census to shed new light on household dynamics over the lifecycle, assuming a stationary environment with respect to neighborhood and population evolutions.

ing and neighborhood attributes, incorporating unobserved preference heterogeneity, (ii) aspects of demand related to the performance of housing as a financial asset (e.g., expected appreciation, volatility), and (iii) moving costs.

Our paper is related to another important strand of research, namely recent advances in the industrial organization literature on dynamic demand for durable and storable goods.[7] Important early contributions to this literature have considered household demand for storable goods (Hendel and Nevo (2006)) and durable goods in markets where products are improving rapidly over time (see Gowrisankaran and Rysman (2012) and Melnikov (2013)). This literature has emphasized a number of important biases from ignoring dynamic considerations in demand estimation that arise because households substitute intertemporally or because early adopters of new technology have systematically stronger preferences for the improved features.[8]

Our model and estimator builds upon this durable demand literature in several ways. First, motivated by the fact that housing constitutes two-thirds of the typical homeowner's financial portfolio, we explicitly model housing as an asset and allow each household's wealth to evolve endogenously. Households in our model anticipate selling their homes at some point in the future and thus explicitly consider the expected evolution of neighborhood amenities and housing prices when deciding where and when to purchase (or sell) their house.

Second, we develop a novel approach for identifying the marginal utility of consumption, which has long been a thorny issue in the literature on demand estimation in applications in both industrial organization and urban economics. The main challenge faced by researchers stems from the strong correlation between a product's price and its unobserved quality, generally requiring instruments that are correlated with price but uncorrelated with unobservables. Such instruments are hard to come by. In our application, we exploit the fact that households face a monetary trade-off both in the standard sense of deciding which product (neighborhood) to purchase but also in terms of the

[7]See Aguirregabiria and Nevo (2010) for an excellent review of this literature.

[8]The paper by Schiraldi (2011) is the most similar research to ours in this literature, developing a model of automobile demand that extends Gowrisankaran and Rysman (2012) to account for the possibility of re-sale in a flexible way.

decision of when to move. Here, we take advantage of the fact that realtor fees during our sample period were quite uniform (6 percent of house value) in order to identify the marginal utility of consumption when estimating each resident's move-stay decision.

Third, we relax the index sufficiency assumption (see Hendel and Nevo (2006) and Melnikov (2013)) that has become a common feature of the dynamic demand literature. This assumption helps to deal with the computational challenges posed by the large state space typically arising in models of dynamic demand. Instead of treating the logit inclusive value as a sufficient statistic for predicting future continuation values, we construct the continuation value from underlying values and observed characteristics associated with each neighborhood in the subsequent period, letting those neighborhood values depend on the current state space in a flexible manner.[9]

We estimate a version of the model that allows both for observed household preference heterogeneity on the basis of race, income, and wealth, and also unobserved preference heterogeneity. We then use the estimated utility parameters to value marginal changes in non-marketed amenities, a central area of inquiry in public and urban economics. In particular, we estimate the way that neighborhood racial composition, violent crime, and air pollution affect the flow utility derived from a particular residential choice.

The findings from this exercise indicate that the preference estimates derived from our dynamic approach differ substantially from estimates derived from a comparable static demand model and from a dynamic model that excludes unobserved heterogeneity. In particular, we find that the static model understates marginal willingness to pay ('MWTP') to avoid pollution and crime; in terms of neighborhood racial composition, the static model overstates MWTP for low-income households and understates MWTP for high-income households. When estimating a dynamic model without unobserved heterogeneity, we find that this significantly understates willingness to pay for crime and neighborhood race.

While multiple factors are relevant for explaining these patterns, a primary expla-

[9] In an important recent contribution to the sorting literature, Kennan and Walker (2011) model inter-state migration decisions by forward-looking agents, focusing on the role of expected income in a job-search framework. Kennan and Walker's model does not consider housing markets; we focus on housing market dynamics within a single market.

nation involves the time-series properties of the respective amenities. In the case of a mean-reverting disamenity such as crime, a static demand model will incorrectly interpret the justifiable downweighting by households of a high value today as a low static valuation, thereby *understating* willingness-to-pay for the amenity. The reverse is true for an amenity such as neighborhood race that exhibits positive persistence: a high value of the amenity today predicts an even higher value in future, so households will appear to overweight current values of the amenity, seen through the lens of a static model. We discuss several other potential sources of bias below.

Beyond the current application, the model and estimation method that we propose provide a foundation for addressing a wide set of dynamic issues in housing markets and cities. These include, for instance, the analysis of the microdynamics of residential segregation and gentrification within metropolitan areas.[10] In addition, the kinds of transactions data required to estimate the model have become increasingly available for cities in the United States and elsewhere in recent years. Together, these make further exploration of dynamic issues ever more viable.

The remainder of the paper proceeds as follows: Section 2 describes the main components of our data set and the matching procedure used to construct it. Our model and estimation strategy are presented in Sections 3 and 4, and the estimates are presented in Section 5. Relative to our benchmark dynamic model that incorporates unobserved heterogeneity, Section 6 discusses the implications of both estimating static demand models when agents are actually behaving dynamically and ignoring individual unobserved heterogeneity. Section 7 concludes.

2 Data

In this section, we describe a new data set that we have assembled, merging information about buyers and sellers with the universe of housing transactions in the San Francisco metropolitan area. We discuss the source data and also demonstrate that the merge

[10]Recent theoretical research on aspects of the dynamic microfoundations of housing markets by Ortalo-Magné and Rady (2002, 2006, 2008) and Bajari, Benkard, and Krainer (2005) raise a number of additional empirical questions that could be addressed using this framework.

results in a high-quality and representative data set based on multiple diagnostic tests.

The data set that we construct is drawn from two main sources. The first comes from Dataquick Information Services, a national real estate data company, and provides information about each housing unit sold in the core counties of the Bay Area (San Francisco, Marin, San Mateo, Alameda, Contra Costa, and Santa Clara) between 1994 and 2004. The buyers' and sellers' names are provided, along with the 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 characteristics.[11] A key feature of this transactions data set is that it also includes information about the buyer's mortgage (including the loan amount and lender's name for all loans). It is this mortgage information that allows us to link the transactions data to information about buyers (and many sellers).

The source of the economic and demographic information about buyers and sellers is the data set 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.[12] The HMDA data provide information on the race, income, and gender of the buyer/applicant, as well as the mortgage loan amount, mortgage lender's name, and the census tract where the property is located.

We merge the two data sets 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 percent of sales. Because the original transactions data set includes the full names of buyers and sellers, we are also able to merge demographic and economic information about sellers into the data set, provided a seller bought another house within the metropolitan area and a unique match with HMDA was obtained for that house. This procedure allows us to merge in information about sellers for approximately 35-40 percent of our sample.

[11] By comparison, the list of housing characteristics is considerably more detailed than that available in Census microdata.

[12] 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.)

To ensure that our HMDA/Dataquick matching procedure is valid, we conduct several diagnostic tests. Using public-access Census micro data from IPUMS, we first 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 data set in Table A.1 in the Appendix. As can be seen, the numbers match almost perfectly in each of the six counties, indicating that the matched buyers are representative of all new buyers. Table A.2 provides a second diagnostic check, concerning the representativeness of the merged data set in terms of housing characteristics. We report sample statistics for a subset of the house-level variables taken from the original data set that includes the complete universe of transactions, as well as sample statistics for the merged data set.[13] A comparison of the two samples suggests that the set of houses for which we have a unique loan record from HMDA is representative of the universe of houses. Overall, our diagnostic checks provide strong evidence supporting the validity of our matching procedure.

In addition to merged data on households and the houses they choose, the estimation routine discussed below also requires that we follow households through time so that we can determine both when they buy and sell a property (if a sale occurs). Since the data set provides a complete census of all house sales with a unique code for every property, it is straightforward to determine if a household moves. And if an individual buys a house in a given period, we know that he/she will stay there until we see that house sell again.[14]

The unit of geography in the model discussed below is a neighborhood, where we define neighborhoods by merging nearby census tracts until there are approximately

[13] We drop outlying observations where reported sales price – in year-2000 dollars – is above the 99th or below the 1st percentile of sales prices. We also drop houses with reported values of lot size, square footage, number of bedrooms, and number of rooms higher (or lower) than their respective maximum (or minimum) reported in Table A.1.

[14] More difficult is determining *where* a household moves to, conditional on moving. The raw data do not provide a unique household identifier; however, they do provide the name of both the purchaser and the seller. We use the name information to create a household identifier by looking for a house purchase in a window of time around a sale for which the purchaser's name (in the purchase) matches the seller's name (in the sale). If we cannot find a new purchase within a year on either side of the sale, we assume that the household has either left the Bay Area or moved to a rental unit.

10,000 housing units in each neighborhood.[15] We drop a number of neighborhoods that have less than 6 sales in any year between 1994 and 2004 or where the ratio of maximum to minimum annual sales exceeds five,[16] leaving us with 218 neighborhoods in total. The corresponding neighborhood boundaries are shown in Figure 1, along with the county names.

Table 1 presents summary statistics for the merged data that we use for estimation. We report summary statistics for both household and neighborhood characteristics.

Table 1: Summary Statistics

Household Characteristics

Variable	Obs.	Mean	Std. Dev.	Min.	Max.
Income	220403	106.87	45.44	0.89	240.00
Down-payment	220403	82.46	51.92	0.00	240.00
Sales Price	220403	382.86	163.70	98.53	1536.71
White	220403	1	0	1	1
Year	220403	1999.04	3.17	1994	2004

Neighborhood Characteristics

Variable	Obs.	Mean	Std. Dev.	Min.	Max.
Percent White	2398	69.63	16.21	26.69	96.79
Violent Crime	2398	453.67	247.02	46.03	2011.05
Ozone	2398	2.17	2.57	0.002	18.25
Sales Price	2398	429.13	206.27	122.75	1792.01

Note: Income, Down-payment, and Sale Price are measured in $1000's.

Our household sample consists of over 220,000 observations.[17] The household-level characteristics we focus on are income, race, and wealth. The sample mean household income is approximately $107,000, with a standard deviation of $45,000. As income is only observed when a household makes a purchase, we assume that income does not

[15]The merging algorithm starts with the least populated census tract, and merges it together with the nearest tract such that the combined population does not exceed 25,000. The algorithm iterates until no possible combination of tracts would result in combined populations of less than 25,000. A population of 25,000 roughly corresponds to 10,000 housing units. The population and geographic data for each census tract come from the 2000 Census.

[16]Specifically, we drop 35 neighborhoods, equivalent to 14 percent of neighborhoods but only 7 percent of sales.

[17]As we estimate the model for white households, with income and down payments less than $240,000, we restrict our sample on a similar basis.

change over time. In terms of race, we only use white households when estimating the model, as discussed below; white households account for just under 70 percent of households in the Bay Area. Household wealth is measured as the difference between the household's current house value and the initial mortgage amount, with current house value being defined as sales price in the year the house is sold and an imputed price in subsequent years. The imputation uses the original house price and adjusts this according to an appreciation index generated from a repeat sales analysis, with the appreciation index calculated separately for each neighborhood.

The neighborhood characteristics we use are mean house price, air quality (ground-level ozone concentrations), violent crime rates, and the racial composition (percentage white) of home owners. We control for changing attributes of the houses that sold when calculating time variation in each neighborhood's mean price.[18] In terms of air quality, we use annual data from the California Air Resources Board (www.arb.ca.gov/adam/) that reports readings from thirty seven monitors in the Bay Area between 1994-2004. While several different measures of ground-level ozone pollution are reported in these data, we use information about the number of days each year that pollution exceeded the one-hour state standard (i.e., 90 parts per billion) to construct specific measures for each neighborhood centroid. In particular, we use the latitudes and longitudes of all monitors to construct a distance-weighted average of the number of 'exceedances' for each neighborhood.

Ozone is a convenient environmental disamenity to study in this context. Unlike many other pollutants, geography and weather are largely responsible for cross-sectional variation in ground-level ozone pollution. San Francisco, Oakland, and San Jose all face heavy traffic congestion. However, wind patterns mitigate much of the ozone pollution in San Francisco and Oakland, while worsening it in San Jose; and mountains ringing the southern end of the Bay Area block air flows and contribute to this effect.[19] At the

[18]To generate appreciation trends, we use the same repeat sales analysis used to impute individual house values. We regress log price on year and house dummies and create appreciation measures from the coefficients on the year dummies: the regressions and associated appreciation measures are estimated separately for every neighborhood. This procedure is effectively a simplification of the method described in Case and Shiller (1989). The cross-sectional variation in house prices is driven by differences in prices across neighborhoods averaged over all years.

[19]The mountains on the eastern side of the Bay are similarly responsible for high levels of pollution

same time, fog (which is especially prevalent in San Francisco) can lower temperatures and block sunlight, preventing the formation of ozone.

In addition to the cross-sectional variation just described, there is also significant variation in ozone pollution levels over time, much of which is due to a variety of programs initiated after California passed its Clean Air Act of 1988. Following several years of relatively low ozone pollution, the Bay Area experienced its worst year of air quality for a decade in 1995. In 1996, the vehicle Buyback Program for cars manufactured in 1975 or before was implemented, and this program contributed to the summer of 1997 being the cleanest season since the early 1960's.[20] While 1998 saw considerably more ozone pollution, the remaining years of our sample returned to relatively low levels. There is no reason to expect that any of these programs would have had special economic consequences for housing prices in any specific part of the Bay Area, aside from those operating through changing amenity values.

Data on violent crimes are taken from the RAND California data base.[21] These figures represent the number of "crimes against people, including homicide, forcible rape, robbery, and aggravated assault" per 100,000 residents and are organized by city. The data describe crime rates for 80 cities in the Bay Area between 1986 and 2008; and we impute crime rates at the centroid of each neighborhood using a distance-weighted average of the crime rate in each city. We focus our attention on violent (as opposed to property) crimes as they are likely to be less subject to reporting error (see Gibbons (2004)). With that in mind, it is possible that our measure of crime will, to some extent, proxy for other sorts of crimes as well.

Crime rates in the Bay Area (and in many other parts of the US) fell dramatically over the course of the 1990's. In the Bay Area, this is particularly evident in communities starting out with very high rates of violent crime (e.g., East Palo Alto), whereas low crime areas (e.g., Palo Alto) saw virtually no change in crime rates over the decade. In general, however, local crime rates have tended to fluctuate in the short run (annually), and even over longer periods.

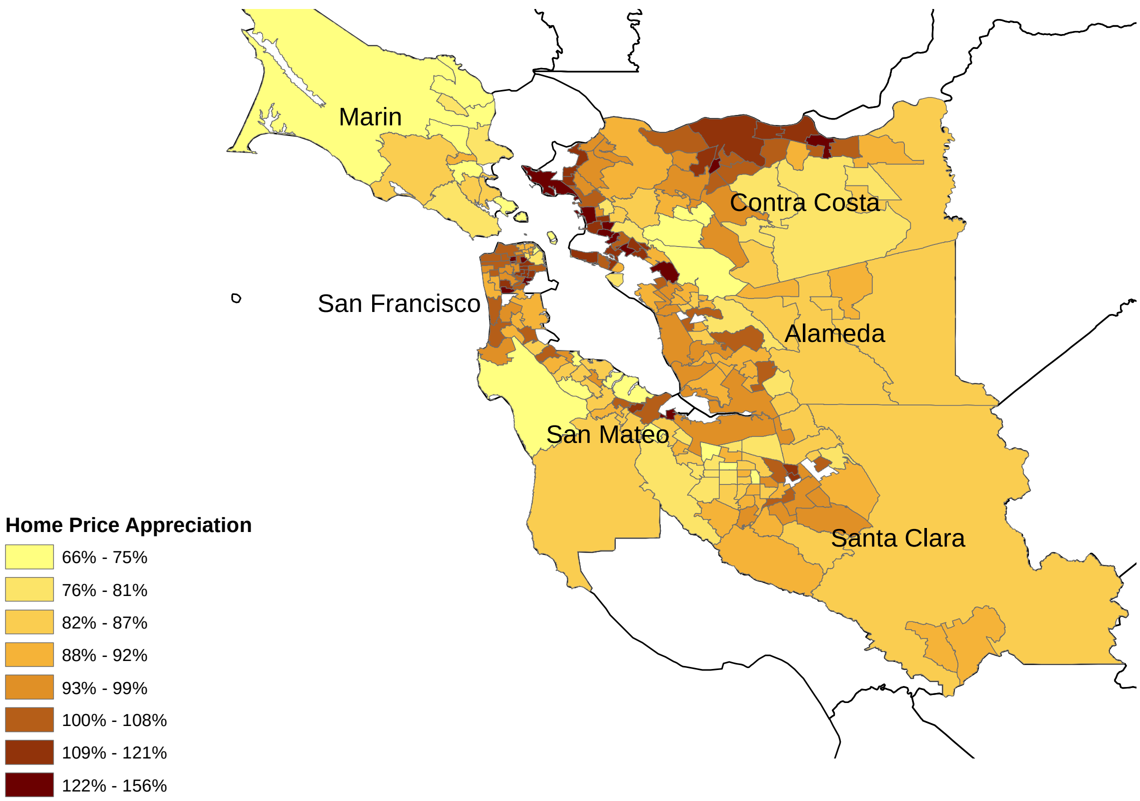
along the I-680 corridor in eastern Contra Costa and Alameda counties.

[20]Also relevant were the Lawn Mower Buyback and Clean Air Plan of 1997.

[21]http://ca.rand.org/stats/community/crimerate.html

The time-series variation in amenities just described may give rise to biases in static demand estimation, anticipating the application in Section 6. Both ground-level ozone and crime vary a great deal from year-to-year and mean-revert over very short time horizons. Neighborhood racial composition, in contrast, is positively persistent, with any change in composition today likely to persist into the future. If households anticipate either the mean reversion or the persistence, their responses will reflect not only the current change but also those expectations; and as a result, we would expect a static model to return biased estimates when valuing these amenities. Regressions exploring the time-series patterns of each (dis)amenity are shown in Table A.6 in the Appendix.

Figure 1: Appreciation rates by neighborhood



The precision of our model depends critically on the fact that rates of change in amenities and house prices are not uniform across neighborhoods. To illustrate the variation in the evolution of prices across regions of the Bay Area, Figure 1 shows real house price appreciation by neighborhood from 1994 to 2004. The estimated price levels are derived separately for each neighborhood using a repeat sales analysis in which the log of the sales price (in 2000 dollars) is regressed on a set of year fixed effects as well as house fixed effects. The figure makes clear the significant differences across neighborhoods in real house price growth over this time period.[22]

2.1 Preliminary Evidence on Dynamic Considerations

Before turning to our dynamic model of neighborhood choice, we present the results of a simple exercise designed to highlight the potential role of dynamic considerations in household location decisions. Specifically, Table 2 examines how neighborhood market shares for the households at time t vary with both contemporaneous and lagged measures of amenities in that neighborhood.

Table 2: Evidence of Dynamic Behavior

	Share	Share
Percent White	0.02479	0.02709
	(0.00026)	(0.00329)
Violent Crime	-0.00092	-0.00047
	(0.00002)	(0.00003)
Ozone	0.07284	0.04831
	(0.00183)	(0.00184)
Price	-0.01331	-0.00734
	(0.00017)	(0.00073)
Lagged Percent White		-0.00316
		(0.00328)
Lagged Violent Crime		-0.00034
		(0.00003)
Lagged Ozone		0.07092
		(0.00160)
Lagged Price		-0.00577
		(0.00078)

Note: The dependent variable is measured as Share*1000.

[22]Omitted neighborhoods in the study area are shaded white, as are the bordering counties.

The results reveal that lagged measures are statistically and economically significant, implying that agents consider both the current state of the neighborhood and how that state has been changing recently when making location decisions. These results also foreshadow our subsequent comparisons of MWTP estimates from the dynamic and static models. For crime and air quality measures, the results show that households put comparable weight on current and lagged measures, implying that their decisions are based more on a longer-run average than the observed measures for a given year. In contrast, the results for neighborhood race imply that households value both the current neighborhood composition and recent changes in composition (which predict future neighborhood composition) in a similar manner - i.e., the coefficient for the lagged value has the opposite sign of the coefficient on the current value.

3 A Dynamic Model of Neighborhood Choice

Previous research modeling the process of household sorting across neighborhoods has generally assumed a static environment.[23] In developing a dynamic sorting model, we introduce the dynamics of the neighborhood choice problem through two channels: wealth accumulation and moving costs. Households have expectations about the appreciation of housing prices and may rationally choose a neighborhood that offers lower current-period utility in return for the increase in wealth associated with house price appreciation in that neighborhood. Moving costs are the other component of the neighborhood choice problem that induce forward-looking behavior: Because households typically pay six percent of the value of their house in real estate agent fees, in addition to the non-financial costs of moving, it is prohibitively costly to re-optimize every period. As a result, households will naturally consider expectations about future utility streams when deciding where to live, making trade-offs between current and future neighborhood attributes and therefore choosing neighborhoods based in part on demographic or economic trends.

[23]See Epple and Sieg (1999), Ekeland, Heckman, and Nesheim (2004), Bajari and Kahn (2005), Ferreyra (2007), and Bayer, McMillan, and Rueben (2011) for several recent examples. Three exceptions are Kennan and Walker (2011) and Bishop (2007), who analyze interregional migration in the US in a dynamic context, and Murphy (2007), who examines the role of dynamic behavior in the supply of new housing.

The model we present is one of homeowner behavior.[24] Households are treated as making a sequence of location decisions that maximize the discounted sum of expected per-period utilities: formulated in a familiar dynamic programming setup, a Bellman equation captures the determinants of the optimal choice.

In each period, every household chooses whether or not to move. If a household moves, it incurs a moving cost and then chooses the neighborhood that yields the highest expected lifetime utility. The decision variable, $d_{i,t}$, denotes both of the choices made by household i in period t, namely (i) whether to move, and (ii) where to move, conditional on deciding to move. If a household decides to move, we denote that decision by $d_{i,t} = j \in \{0, 1, \ldots, J\}$, where j indexes neighborhoods, J denotes the total number of neighborhoods in the Bay Area, and 0 denotes the outside option. If a household decides not to move, we denote that decision by $d_{i,t} = J + 1$.[25]

The observed state variables at time t are $X_{j,t}$, $Z_{i,t}$, and $h_{i,t}$. $X_{j,t}$ is a vector of characteristics that affect the per-period utility a household may receive from living in neighborhood $j \in \{0, 1, \ldots, J\}$. For example, $X_{j,t}$ may include the price of housing and the quality of local attributes, such as air quality, crime, or the neighborhood's racial composition. $Z_{i,t}$ 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. This vector may include such variables as income, wealth, or race. And $h_{i,t} \in \{0, 1, \ldots, J\}$ denotes the neighborhood chosen in $t-1$, including the outside option.

In addition to the decision variable, $d_{i,t}$, and the observable variables, $X_{j,t}$, $Z_{i,t}$, and $h_{i,t}$, the model incorporates three unobservable variables, g_i, $\xi_{j,t}$, and $\epsilon_{i,j,t}$. Of

[24] We do not explicitly model the decision whether to rent or to own. We do, however, include an outside option that includes moving from home ownership in the Bay Area to either a rental property or a home outside the Bay Area.

[25] The number of choice options is therefore $J+2$. For a household who lived in neighborhood j in $t-1$, we distinguish between the choices of not moving ($d_{i,t} = J+1$) and of moving to a different house within neighborhood j ($d_{i,t} = j$), as there are a small number of observations for which households make such within-neighborhood moves. To simplify notation, we do not use a separate index for neighborhoods and choices. For choices $j \in \{0, 1, \ldots, J\}$, the household is choosing to move to neighborhood j; for choice $j = J + 1$, the household is choosing to not move and so to remain in the current neighborhood, which is in $\{0, 1, \ldots, J\}$.

these, g_i represents the unobserved type of the household;[26] $\xi_{j,t}$ represents unobserved neighborhood quality;[27] and $\epsilon_{i,j,t}$ is an idiosyncratic stochastic shock that determines the utility a household i receives from choosing option $j \in \{0, 1, \ldots, J+1\}$ in period t.[28] Let $s_{i,t}$ denote the states X_t, ξ_t, $Z_{i,t}$, and g_i, as well as any other information-set variables (such as lagged characteristics) that help predict future neighborhood or household characteristics.

The primitives of the model can be written (u, MC, q, β). Taking these in turn, $u_{i,j,t} = u(X_{j,t}, \xi_{j,t}, Z_{i,t}, g_i, \epsilon_{i,j,t})$ is the per-period utility function, excluding any moving costs, capturing the utility that household i receives from living in neighborhood j. $MC_{i,t} = MC(Z_{i,t}, X_{h_{i,t}})$[29] is the per-period moving cost function, which is only paid when a household moves. By assumption, moving costs are not a function of where within the metropolitan area the household moves to; however, they are assumed to be a function of the characteristics of the neighborhood the household is leaving, $X_{h_{i,t}}$, in order to capture the fact that realtor fees are proportional to the value of the house one sells. The full flow utility function, adjusting for moving costs in the event that they are incurred, is given by $u^{MC}_{i,j,t} = u_{i,j,t} - MC_{i,t}1_{[j \neq J+1]}$. The transition probabilities of the observables and unobservables are assumed to be Markovian and are given by $q = q(s_{i,t+1}, h_{i,t+1}, \epsilon_{i,t+1} | s_{i,t}, h_{i,t}, \epsilon_{i,t}, d_{i,t})$, where $\epsilon_{i,t}$ is the vector containing all choices. Finally, β is the discount factor.

Each household is assumed to behave optimally in the sense that its actions are taken to maximize lifetime expected utility. That is, the household's problem is to make

[26] The unobserved type captures a households' preferences for sub-regions of the Bay Area and is discussed in greater detail in Section 4.3 below.

[27] We allow households to derive different levels of utility from unobserved neighborhood quality based on their observable demographic characteristics. In so doing, we differ from previous work, such as Berry, Levinsohn, and Pakes (1995), in which all individuals have the same preferences for the unobserved choice characteristic.

[28] As the vector of idiosyncratic shocks contains $J+2$ elements, a household which moves but chooses to reside in the same neighborhood would receive a different draw for ϵ than if that household had chosen not to move.

[29] We write $X_{h_{i,t}}$ using a slight abuse of notation, omitting the distinct 'time' subscript.

a sequence of residential location decisions, $\{d_{i,t}\}$, to maximize

$$E\big[\sum_{r=t}^{T}\beta^{r-t}\big(u^{MC}(X_{j,r},\xi_{j,r},Z_{i,r},g_i,\epsilon_{i,j,r},X_{h_{i,r}})\big)\big|s_{i,t},h_{i,t},\epsilon_{i,t},d_{i,t}\big] \tag{1}$$

The optimal decision rule is given by d^*. Under the Markov structure of the problem, this is only a function of the state variables – that is, $d_{i,t}=d^*_{i,t}(s_{i,t},h_{i,t},\epsilon_{i,t})$. When the sequence of decisions, $\{d_{i,t}\}$, is determined according to the optimal decision rule, d^*, lifetime expected utility can be represented by the value function, which can be broken 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 $t+1$. We assume that the problem has an infinite horizon, allowing us to drop time subscripts on the value function, V.[30] Thus:

$$V(s_{i,t},h_{i,t},\epsilon_{i,t})=max_j\{u^{MC}_{i,j,t}+\beta E\big[V(s_{i,t+1},h_{i,t+1},\epsilon_{i,t+1})\big|s_{i,t},h_{i,t},\epsilon_{i,t},d_{i,t}=j\big]\} \tag{2}$$

While equation (2) is a contraction mapping in V under certain technical assumptions, V is a function of both the observed and unobserved state variables. Therefore, we make a series of assumptions similar to those in Rust (1987) in order to simplify the model.

Additive Separability Assumption: We assume that the per-period utility function, u, is additively separable in the idiosyncratic error term, $\epsilon_{i,j,t}$.

Thus we can express the full flow utility function, $u^{MC}_{i,j,t}$, as

$$u^{MC}_{i,j,t}=u(X_{j,t},\xi_{j,t},Z_{i,t},g_i)-MC(Z_{i,t},X_{h_{i,t}})1_{[j\neq J+1]}+\epsilon_{i,j,t} \tag{3}$$

Conditional Independence Assumption: This is in two parts, relating to the transition probabilities of the unobserved and observed states. The idiosyncratic choice-specific error term, $\epsilon_{i,j,t}$, is assumed to be distributed i.i.d. Type I Extreme Value (with

[30] Assuming an infinite horizon implies that $V_t(s_{i,t},h_{i,t},\epsilon_{i,t})=V(s_{i,t},h_{i,t},\epsilon_{i,t})$ and $d_t(s_{i,t},h_{i,t},\epsilon_{i,t})=d(s_{i,t},h_{i,t},\epsilon_{i,t})$.

density q_ϵ). Additionally, we assume that conditional on $s_{i,t}$ and $d_{i,t}$, the errors $\epsilon_{i,j,t}$ have no predictive power regarding future states $s_{i,t+1}$ or $h_{i,t+1}$. Based on the structure of the model and the assumption about $\epsilon_{i,j,t}$, it follows that, conditional on $s_{i,t}$ and $d_{i,t}$, the neighborhood chosen in the previous period, $h_{i,t}$, has no predictive power regarding any future states and that $d_{i,t}$ is sufficient to predict $h_{i,t+1}$ perfectly. We can therefore express the transition density for the Markov process, q, as[31]

$$q(s_{i,t+1}, h_{i,t+1}, \epsilon_{i,t+1}|s_{i,t}, h_{i,t}, \epsilon_{i,t}, d_{i,t}) = q_s(s_{i,t+1}|s_{i,t}, d_{i,t})q_h(h_{i,t+1}|d_{i,t})q_\epsilon(\epsilon_{i,t+1}), \forall t \tag{4}$$

This allows us to define the choice-specific value function, $v_j^{MC}(s_{i,t}, h_{i,t})$, as

$$v_j^{MC}(s_{i,t}, h_{i,t}) = u_{i,j,t} - MC_{i,t}1_{[j\neq J+1]} + \beta E\Big[log\big(\sum_{k=0}^{J+1} exp(v_k^{MC}(s_{i,t+1}, h_{i,t+1}))\big)\Big|s_{i,t}, d_{i,t} = j\Big] \tag{5}$$

where

$$log\big(\sum_{k=1}^{J+1} exp(v_k^{MC}(s_{i,t}, h_{i,t}))\big) = E_\epsilon\Big[V(s_{i,t}, h_{i,t}, \epsilon_{i,t})\Big] = E_\epsilon\Big[max_k[v_k^{MC}(s_{i,t}, h_{i,t}) + \epsilon_{i,k,t}]\Big]$$

Similarly to per-period utility, we break out the full choice-specific value function into a component capturing the lifetime expected utility of choosing neighborhood j ignoring moving costs and another component that involves moving costs – these two components are the focus of the first stage of our estimation approach, described in the next section. It is worth noting that the component of lifetime utility that ignores moving costs is not a function of $h_{i,t}$, the neighborhood in which the household lives before making the decision in period t. Thus we have

$$v_j^{MC}(s_{i,t}, h_{i,t}) = v_j(s_{i,t}) - MC(Z_{i,t}, X_{h_{i,t}})1_{[j\neq J+1]} \tag{6}$$

where

$$v_j(s_{i,t}) = u_{i,j,t} + \beta E\Big[log\big(\sum_{k=0}^{J+1} exp(v_k^{MC}(s_{i,t+1}, h_{i,t+1}))\big)\Big|s_{i,t}, d_{i,t} = j\Big] \tag{7}$$

[31]In Section 4, we will outline in detail our assumptions about the transitions of the observable states.

4 Estimation

The estimation of the model primitives proceeds in two main stages. In the first stage, we use the household location and mobility decisions to estimate the value of lifetime expected utility for each neighborhood, time period, and household type, where household type is characterized by race, income, and wealth, as well as an unobservable characteristic that captures a household's preference for sub-regions within the Bay Area. In the second stage, we recover fully-flexible estimates of the per-period utility and regress them on a set of observable attributes. We use a novel approach to control for the endogeneity of price in this second stage, utilizing outside information relating to the financial cost of moving to pin down the coefficient on house prices. As will become clear, an important feature of our estimation strategy is its low computational burden.

For simplicity of exposition, we first present the estimation routine without controlling for unobserved heterogeneity – this corresponds to not including g_i in the model above. We then outline how controlling for unobserved heterogeneity affects the estimation in Section 4.3.

4.1 Estimation Stage One – Household Location and Mobility Decisions

Stage one of our estimation approach focuses on household location and mobility decisions. To provide context, we describe the nature of the data we observe regarding household choices, i.e., the decisions we wish to capture when forming a likelihood. A household appears in our data set for the first time when we observe the household buying a house. This decision is only observed conditional on that household choosing to move to a neighborhood in the Bay Area in that period. We then observe whether or not that household chooses to move in each subsequent period. We observe T_i of these decisions, where T_i is the number of periods up to and including the period household i moves or is censored because the household has not moved before the end of our sample. If the household does move, we observe whether or not it subsequently chose the outside option, which is defined as not buying in the Bay Area (and thus includes renting). We first outline the location problem faced by a household that has chosen to move, and then the problem of moving or staying in any given period.

4.1.1 Household Location Decisions and Lifetime Expected Utility

A household who has decided to move will choose a neighborhood which offers the highest lifetime utility by maximizing over the set of choice-specific value functions v^{MC}. Conditional on moving, the moving cost term, $MC(Z_{i,t}, X_{h_{i,t}}) = MC(Z_{i,t}, d_{i,t-1})$, is assumed to be identical for all neighborhoods. As an additive constant, it simply drops out and each household that moves chooses neighborhood j to maximize $v_j(s_{i,t}) + \epsilon_{i,j,t}$, where $v_j(s_{i,t})$ is given in (7).

Based on household characteristics – income, wealth, and race in our actual implementation – we divide households into distinct types indexed by τ. Let $v^{\tau}_{j,t} = v_j(s_{i,t})$ when the characteristics $(Z_{i,t})$ of household i imply that the household is of type τ. $v^{\tau}_{j,t}$ is then the choice-specific value a household of type τ receives from choosing neighborhood j, ignoring any potential moving costs. Letting $u^{\tau}_{j,t}$ denote the deterministic component of flow utility for a household of type τ, we can rewrite (7) using lifetime utilities, $v^{\tau}_{j,t}$:

$$v^{\tau}_{j,t} = u^{\tau}_{j,t} + \beta E\Big[log\Big(\sum_{k=0}^{J+1} exp(v^{\tau_{t+1}}_{k,t+1} - MC^{\tau_{t+1}}_{j,t+1} 1_{[k \neq J+1]})\Big)\Big| s_{i,t}, d_{i,t} = j\Big] \tag{8}$$

We assume that agents use the state variables in s and the decision $d_{i,t}$ to directly predict future lifetime utilities, $v_{j,t+1}$ and future types, τ_{t+1}. We discuss exactly how they forecast in Section 4.2.

For any given time period, the vector of mean lifetime utilities, v^{τ}_t, is unique up to an additive constant, thus requiring some normalization for each τ. Therefore, instead of estimating $v^{\tau}_{j,t}$ for every neighborhood and type, we estimate $\tilde{v}^{\tau}_{j,t}$, where $\tilde{v}^{\tau}_{j,t} = v^{\tau}_{j,t} - m^{\tau}_t$ and m^{τ}_t is a normalizing constant, which we can estimate as it is identified by the mobility decisions discussed below. In practice, we set m^{τ}_t such that the normalized lifetime expected utilities have a zero mean for each type-year combination. Household i of type τ chooses option j if $\tilde{v}^{\tau}_{j,t} + \epsilon_{i,j,t} > \tilde{v}^{\tau}_{k,t} + \epsilon_{i,k,t}$, $\forall\, k \neq j$. Conditional upon moving to an inside option (i.e., for $d_{i,t} \neq \{0, J+1\}$), the probability of any household of type

τ choosing neighborhood j in period t when $\epsilon_{i,j,t}$ is distributed i.i.d., Type I Extreme Value can therefore be expressed as:

$$P^{\tau}_{j,t} = \frac{e^{\tilde{v}^{\tau}_{j,t}}}{\sum_{k=1}^{J} e^{\tilde{v}^{\tau}_{k,t}}} \tag{9}$$

Let $t_{1,i}$ denote the time period in which household i decides where to move (conditional on moving to an inside option). A household's likelihood contribution for this decision is denoted by $L_i^{neigh}(\tilde{v})$, where $\tilde{v}$ is the vector of all values of $\tilde{v}^{\tau}_{j,t}$ and is given by:

$$L_i^{neigh}(\tilde{v}) = \prod_{j=1}^{J} (P^{\tau}_{j,t_{1,i}})^{1_{[d_{i,t_{1,i}}=j]}} \tag{10}$$

We also want to estimate a lifetime utility term for the outside option. The probability that a household chooses the outside option in time period t, conditional on moving, is given by:

$$P^{\tau}_{0,t} = \frac{e^{\tilde{v}^{\tau}_{0,t}}}{\sum_{k=0}^{J} e^{\tilde{v}^{\tau}_{k,t}}} \tag{11}$$

Given data that allow us (at least partially) to follow individuals through time, we can form the likelihood that a household chooses the outside option conditional on moving. Let $t_{2,i}$ denote the time period in which household i is considering the outside option (conditional on moving). The likelihood, which is denoted by $L_i^{out}(\tilde{v})$, is given by:

$$L_i^{out}(\tilde{v}) = (P^{\tau}_{0,t_{2,i}})^{1_{[d_{i,t_{2,i}}=0]}} (1 - P^{\tau}_{0,t_{2,i}})^{1_{[d_{i,t_{2,i}} \in \{1,\ldots,J\}]}} \tag{12}$$

For the many households who never make a subsequent move during our sample, we define L_i^{out} as equal to one.

4.1.2 Household Mobility Decisions, Moving Costs, and the Marginal Utility of Wealth

In a housing market context, households behave dynamically by taking into account the effect that their current decision has on future utility flows. In our model, the current decision affects future flows through the two channels mentioned previously: households are aware that they will incur a transaction cost by re-optimizing in the future, and in addition, the decision of where to live today affects wealth in future. Equation (8) shows how the current decision affects both today's flow utility and future utility; it also makes clear that $v^{\tau}_{j,t}$ (or $\tilde{v}^{\tau}_{j,t}$) as well as moving costs determine households' decisions to move or stay in a particular period.

From the model outlined above, we know that in any given period, a household will move if the lifetime expected utility of staying in its current neighborhood is less than the lifetime expected utility of the best alternative when moving costs are factored in. We assume that moving costs, $MC^{\tau_t}_{j,t}$, are made up of two components: financial costs, $FMC(h_{i,t})$, and psychological costs, $PMC(Z_{i,t})$. The financial moving costs are a function of previous location decisions, $h_{i,t}$, as households pay financial costs based primarily on the property they sell. The psychological costs are assumed to be a function of the observable characteristics, $Z_{i,t}$, that define household type τ.

As the financial moving costs reduce wealth, choosing to move changes a household's 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 its 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 percent of the value of housing in the neighborhood that a household is leaving (i.e., $FMC(h_{i,t}) = 0.06 \cdot Price_{h_{i,t}}$).

If a household of type τ moves, we denote their new type as $\bar{\tau}$, the new type following a move reflecting the reduction in wealth by the amount of FMC. A household will choose to stay if:

$$v^{\tau}_{J+1,t} + \epsilon_{i,J+1,t} > max_k[v^{\bar{\tau}}_{k,t} + \epsilon_{i,k,t}] - PMC(Z_{i,t}) \tag{13}$$

Employing the definition of the normalized choice-specific value functions, $\tilde{v}^{\tau}_j$, where $\tilde{v}^{\tau}_j = v^{\tau}_j - m^{\tau}$, we can then rewrite (13) as:

$$\tilde{v}^{\tau}_{J+1,t} + \epsilon_{i,J+1,t} > max_k[\tilde{v}^{\bar{\tau}}_{k,t} + \epsilon_{i,k,t}] - (m^{\tau}_t - m^{\bar{\tau}}_t) - PMC(Z_{i,t}) \tag{14}$$

The term $(m_t^{\tau} - m_t^{\bar{\tau}})$ is unobserved but can be estimated as the difference between the value associated with being type τ and the value associated with the reduced wealth after paying financial moving costs. In principle, we could estimate a separate term for each combination of τ and FMC; in practice, we choose to parameterize it as a function of $Z_{i,t}$ and $FMC_{i,t}$, so:

$$m_t^{\tau} - m_t^{\bar{\tau}} = FMC_{i,t}\gamma_{fmc}^{\tau}$$

where $FMC_{i,t} = 0.06 \cdot Price_{h_{i,t}}$ and $\gamma_{fmc}^{\tau} = Z_{i,t}'\gamma_{fmc}$. We parameterize the psychological costs as:

$$PMC_{i,t} = Z_{i,t}'\gamma_{pmc}$$

Note that the stochastic terms are $max_{k \neq j}[\tilde{v}_{k,t}^{\bar{\tau}} + \epsilon_{i,k,t}]$, and $\epsilon_{i,J+1,t}$.[32] The probability of a household staying in its current house in a given period t is:

$$P_{stay,i,t}^{\tau} = \frac{e^{\tilde{v}_{J+1,t}^{\tau}}}{e^{\tilde{v}_{J+1,t}^{\tau}} + \sum_{k=0}^{J} e^{\tilde{v}_{k,t}^{\bar{\tau}} - FMC_{i,t}\gamma_{fmc}^{\tau} - Z_{i,t}'\gamma_{pmc}}} \tag{15}$$

The likelihood contribution of each household's sequence of move/stay decisions is denoted by $L_i^{stay}(\tilde{v}, \gamma_{fmc}, \gamma_{pmc})$ and is given by:

$$L_i^{stay}(\tilde{v}, \gamma_{fmc}, \gamma_{pmc}) = \prod_{t=t_{1,i}+1}^{t_{1,i}+T_i} (P_{stay,i,t}^{\tau})^{1[d_{i,t}=J+1]}(1 - P_{stay,i,t}^{\tau})^{1[d_{i,t} \neq J+1]} \tag{16}$$

4.1.3 Combined Likelihood

Combining the contributions from the location choice and move/stay decision, the full log-likelihood can then be expressed as:

$$\mathcal{L}(\tilde{v}, \gamma_{fmc}, \gamma_{pmc}) = \sum_{i=1}^{N} \Big(log(L_i^{neigh}(\tilde{v})) + log(L_i^{out}(\tilde{v})) + log(L_i^{stay}(\tilde{v}, \gamma_{fmc}, \gamma_{pmc})) \Big) \tag{17}$$

One estimation option would be choose $(\tilde{v}, \gamma_{fmc}, \gamma_{pmc})$ to maximize $\mathcal{L}$. This would be computationally prohibitive. A simple and commonly-used approach to circumvent

[32] It would be straightforward to allow for a shock to moving costs also, which would effectively allow all the idiosyncratic errors except ϵ_{J+1} to be correlated.

this difficulty is to take advantage of the separability of the log-likelihood function. In that vein, one could choose $\tilde{v}$ to maximize $\sum_{i=1}^{N}(log(L_i^{neigh}(\tilde{v}))+log(L_i^{out}(\tilde{v}))$ and then choose $(\gamma_{fmc},\gamma_{pmc})$ to maximize $\sum_{i=1}^{N}(log(L_i^{stay}(\widehat{\tilde{v}},\gamma_{fmc},\gamma_{pmc})))$. The second step would have a low computational burden because the dimensionality of $(\gamma_{fmc},\gamma_{pmc})$ is low and the gradient and Hessian have closed-form solutions.[33] The first step would also have a low computational burden, because the first-order condition associated with maximizing $\sum_{i=1}^{N}(log(L_i^{neigh}(\tilde{v}))+log(L_i^{out}(\tilde{v})))$ yields a closed-form solution for $\tilde{v}$. This closed-form solution for $\tilde{v}_{j,t}^{\tau}$ is given by:

$$\widehat{\tilde{v}}_{j,t}^{\tau} = log(\hat{P}_{j,t}^{\tau}) - \frac{1}{J+1}\sum_{k=0}^{J} log(\hat{P}_{k,t}^{\tau}) \tag{18}$$

where $\hat{P}_{j,t}^{\tau}$ denotes the empirical probability that households of type τ choose neighborhood j in period t.

In our application, instead of simply calculating observed shares as the portion of households of a given type who buy in an area, we use a weighted measure to avoid some small sample issues when the number of types, M, grows large relative to the sample size. We do this to incorporate the information from similar types when calculating shares for any particular type.[34] Naturally, the weights will depend on how far away the other types are in type space. Denoting the weights by $W^{\tau}(Z_{i,t})$, the formula for calculating observed shares (of inside choices) is given by:[35]

$$\hat{\hat{P}}_{j,t}^{\tau} = \frac{\sum_{i=1}^{N} 1_{[d_{i,t}=j]} \cdot W^{\tau}(Z_{i,t})}{\sum_{i=1}^{N} W^{\tau}(Z_{i,t})} \tag{19}$$

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,

[33] The second step would effectively be a standard binary-choice logit model with one of the parameters $(\widehat{\tilde{v}})$ known.

[34] For example, if we want to calculate the share of households with an income of \$50,000 choosing neighborhood j in period t, we would use some information about the residential decisions of those earning \$45,000 or \$55,000 in that period.

[35] If $W^{\tau}(Z_i) = 1_{[Z_i=Z^{\tau}]}$, $\hat{\hat{P}} = \hat{P}$.

N, and bandwidth, $b_k(\tau)$, determined by visual inspection:

$$W^{\tau}(Z_{i,t}) = \prod_{k=1}^{K} \frac{1}{b_k(\tau)} N\Big(\frac{Z_{i,t} - Z^{\tau}}{b_k(\tau)}\Big) \tag{20}$$

The outside option shares are estimated using the share of households who were owning in the Bay Area, sell, and then choose to not buy another house in the Bay Area.[36]

Formally, our estimation approach is given by:

$$(\widehat{\tilde{v}}, \hat{\gamma}_{fmc}, \hat{\gamma}_{pmc}) = argmax_{(\tilde{v},\gamma_{fmc},\gamma_{pmc})} \mathcal{L}(\tilde{v}, \gamma_{fmc}, \gamma_{pmc}) \tag{21}$$

subject to the constraint:

$$\hat{\hat{P}}^{\tau}_{j,t} = P^{\tau}_{j,t}(\tilde{v}), \ \forall \ j \in \{0, ...J\}. \tag{22}$$

In practice, the constraint (22) determines $\widehat{\tilde{v}}$, which can be found using a first-order condition analogous to (18).[37] Treating $\widehat{\tilde{v}}$ as known, $(\gamma_{fmc}, \gamma_{pmc})$ is then chosen to maximize $\sum_{i=1}^{N}(log(L_i^{stay}(\widehat{\tilde{v}}, \gamma_{fmc}, \gamma_{pmc})))$, which is effectively a standard binary-choice logit model.

Showing consistency is straightforward: as long as the weights in (20) are defined such that $\hat{\hat{P}} \rightarrow \hat{P}$ as the sample size increases, the constraint converges to the first-order condition (18) and the estimation approach becomes standard Maximum Likelihood.

4.2 Estimation Stage Two – Per-Period Utility

The second stage of the estimation procedure involves estimating per-period utility in several sub-steps, then decomposing it.

From the first stage, we know the distribution of moving costs for each type, the marginal value of changing type, and the mean utility terms, $\tilde{v}$. The first stage of our

[36] As there are fewer observations for households that we can follow over time, we do not estimate $\hat{\hat{P}}^{\tau}_{0t}$, separately for each year and type. Instead, for each type, we estimate a separate logit model including a linear time trend.

[37] The closed-form is given by: $\widehat{\tilde{v}}^{\tau}_{j,t} = log(\hat{\hat{P}}^{\tau}_{j,t}) - \frac{1}{J+1}\sum_{k=0}^{J} log(\hat{\hat{P}}^{\tau}_{k,t})$.

estimation approach involved making a normalization for each household type, where 'type' could be defined by personal characteristics (race, income, and wealth in our application). Once we set the mean choice-specific utility from having no wealth to zero, we only need to know these baseline differences, $m_t^{\tau} - m_t^{\bar{\tau}}$, in order to recover the unnormalized choice-specific value functions.[38] As we can estimate the baseline differences, we simply recover the true choice-specific value functions as $v_{j,t}^{\tau} = \tilde{v}_{j,t}^{\tau} + m_t^{\tau}$. It is important to recover these baseline differences because they represent the additional utility a household would receive from extra wealth, the marginal utility of wealth being a key output of the estimation. Given that the choice of neighborhood affects future household type, the baseline differences in utility across types represent potential future utility gains from wealth accumulation. In addition, we will also use the estimate of the marginal value of wealth as a novel way to deal with the endogeneity of house prices in the final sub-step.

4.2.1 Recovering Per-Period Utility

Given, $(v, \gamma_{fmc}, \gamma_{pmc})$ from the first stage, the next step is to specify and estimate the relevant transition probabilities. We assume that households use today's states to directly predict future values of the lifetime utilities, v, rather than predicting the values of the variables upon which v depends. As potential future moving costs are a function of the price of housing in the neighborhood chosen in this period, households also need to predict how the price of the house they currently occupy will 'transition.' Further, as both moving costs and lifetime utilities are determined by household type, households also need to predict how their types will change. The only determinant of type that changes endogenously is wealth, so we assume that knowing how house prices change is sufficient for knowing how wealth (and thus type) will change also.[39] We therefore only need to model transition probabilities for v and price.

In theory, we could estimate the transition probabilities for lifetime utility sepa-

[38]We set the mean choice-specific utility from no wealth to zero for each year/income/race combination. We are not imposing that these values are identically zero, however, and effectively undo this temporary normalization through the use of year and type dummies in the last sub-step.

[39]With access to richer data including other forms of household wealth, the definition of wealth that we use to define type could be expanded.

rately by type-neighborhood combination, as we have a time series for each type and neighborhood. To increase the efficiency of our estimates, we impose several restrictions.

Within each type, we could assume that the neighborhood mean utilities, $v^{\tau}_{j,t}$, 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 model the transition of the choice-specific value functions, $v^{\tau}_{j,t}$, as:

$$v^{\tau}_{j,t} = \rho^{\tau}_{0,j} + \sum_{l=1}^{L} \rho^{\tau}_{1,l} v^{\tau}_{j,t-l} + \sum_{l=1}^{L} X'_{j,t-l} \rho^{\tau}_{2,l} + \rho^{\tau}_{3,j}\, t + \omega^{\tau}_{j,t} \tag{23}$$

where the time-varying neighborhood attributes included in $X_{j,t}$ are price, racial composition (percentage white), pollution (the number of days that the ozone concentration exceeds the California state maximum threshold), and the violent crime rate.[40] Lagged value functions are also included as explanatory variables, implicitly allowing the transition probabilities be a function of the unobserved neighborhood attributes.

We also need to know how housing wealth changes in order to specify transition probabilities for types. We use sales data to construct price indices for each type-tract-year combination. Recalling that price is one of the variables in the set of neighborhood characteristics, X, we estimate transition probabilities for price levels according to:

$$price_{j,t} = \varrho_{0,j} + \sum_{l=1}^{L} X'_{j,t-l} \varrho_{2,l} + \varrho_{3,j}\, t + \varpi^{\tau}_{j,t} \tag{24}$$

Given these transition probabilities, it is straightforward to estimate transition probabilities for wealth and thus type, τ. In both cases, we use two lags of the dependent variable ($v^{\tau}_{j,t}$ or $price^{\tau}_{j,t}$) as well as two lags of the other exogenous variables in X.

Knowing v, PMC, and the transition probabilities allows us to calculate mean flow

[40]For the outside option, we do not observe any attributes and estimate only including lags of the choice-specific value function. That is, we estimate $v^{\tau}_{0t} = \rho^{\tau}_{0,0} + \sum_{l=1}^{L} \rho^{\tau}_{1,l} v^{\tau}_{0t-l} + \rho^{\tau}_{3,0}\, t + v^{\tau}_{0t}$.

utilities for each type and neighborhood, $u^{\tau}_{j,t}$, according to:

$$u^{\tau}_{j,t} = v^{\tau}_{j,t} - \beta E\Big[log\Big(\sum_{k=0}^{J+1} exp(v^{\tau_{t+1}}_{k,t+1} - MC^{\tau_{t+1}}_{j,t+1} 1_{[k \neq J+1]})\Big)\Big| s_{i,t}, d_{i,t} = j\Big] \tag{25}$$

where, in practice, s includes all the variables on the right-hand side of equations (23) and (24), and β is set equal to 0.95.

For each type, τ, neighborhood, j, and time, t, we now have the necessary information to simulate the expectation on the right-hand side of (25). To do this, we draw a large number of $v_{j,t+1}$'s and $price_{j,t+1}$'s according to their estimated distributions. Specifically, using r to index random draws, each $v_{j,t+1}(r)$ and $price_{j,t+1}(r)$ are generated by drawing from the empirical distribution of errors obtained when estimating (23) and (24) and using the observed values of the current states. The draws on $price_{j,t+1}$ are used to form τ_{t+1} and $MC^{\tau_{t+1}}_{j,t+1}$.[41] For each draw, r, we can then calculate a per-period flow utility $u^{\tau}_{j,t}(r)$ using (25). The simulated $u^{\tau}_{j,t}$ is then calculated as $\frac{1}{R}\sum_{r=1}^{R} u^{\tau}_{j,t}(r)$.[42]

4.2.2 Decomposing Per-Period Utility

Once we recover the mean per-period flow utilities, we can decompose them into functions of the observable neighborhood characteristics, $X_{j,t}$. We treat $\xi^{\tau}_{j,t}$ as an error term in the following regression:

$$u^{\tau}_{j,t} = \alpha^{\tau}_0 + \alpha^{\tau}_c + \alpha^{\tau}_t + X'_{j,t}\alpha^{\tau}_x + \xi^{\tau}_{j,t} \tag{26}$$

This decomposition of the mean flow utilities is similar to Berry, Levinsohn, and Pakes (1995) or Bayer, Ferreira, and McMillan (2007), though in these models, the choice-specific unobservable, $\xi_{j,t}$, was treated as a vertical characteristic that affected all households' utilities in the same way. In our application, we allow households who are different (based on observable demographic characteristics) to view the unobservable component

[41]Once we draw a value for $price_{j,t+1}$, we can calculate $wealth_{t+1}$ for someone living neighborhood j as $wealth_t + (price_{j,t+1} - price_{j,t})$ and $MC^{\tau_{t+1}}_{j,t+1}$ as 6 percent of $price_{j,t+1}$.

[42]The total number of draws, R, is chosen to be large enough such that the simulated $u^{\tau}_{j,t}$ does not change as R increases. In practice, setting R equal to 10,000 is sufficient.

differently, as in Timmins (2007) – hence the 'τ' superscript in $\xi^{\tau}_{j,t}$. In addition to the neighborhood characteristics already referred to, we include controls for type (τ), county (c), and year (t).

The user cost of owning a house is typically calculated as a percentage of house value. Here, we calculate the user cost for a given neighborhood as 5 percent of mean prices in the neighborhood. User costs are clearly endogenous, however. The traditional approach to this problem makes use of instrumental variables. Our approach is different: we use the marginal value of wealth estimated in Section 4.1.2 to recover the marginal disutility of user costs. We assume that the effect of a marginal change in wealth on lifetime utility is the same as the effect of a marginal change in income on one period's utility. In particular, the marginal utility of income (the negative of which can be interpreted as the coefficient on user cost) is given by γ^{τ}_{fmc}. To decompose mean flow utilities, we therefore estimate the following regression, where $\widehat{\gamma^{\tau}_{fmc}}$ is known from Stage 1 and $\tilde{X}$ denotes the non-user cost components of X:

$$u^{\tau}_{j,t} + \widehat{\gamma^{\tau}_{fmc}} usercost_{j,t} = \alpha^{\tau}_0 + \alpha^{\tau}_c + \alpha^{\tau}_t + \tilde{X}'_{j,t}\alpha^{\tau}_x + \xi^{\tau}_{j,t} \tag{27}$$

In principle, we could decompose the flow utilities separately for each type, τ, as written above. In practice, we estimate two versions of the model. In the first, we constrain $\alpha^{\tau}_c, \alpha^{\tau}_t, \alpha^{\tau}_x$ to be the same for all τ. In the second, we allow $\alpha^{\tau}_c, \alpha^{\tau}_t, \alpha^{\tau}_x$ to vary with income, but not with wealth. In both cases, no restrictions are placed on α^{τ}_0 and $\xi^{\tau}_{j,t}$.

4.3 Estimation – Unobserved Geographical Preferences

We model a form of individual unobserved heterogeneity that describes geographic preferences. These capture persistent sub-regional attachments associated with place of work or extended family ties, neither of which are observed in our data set[43] – information about place of work is one respect in which confidential Census data are unambiguously

[43] Other avenues through which unobserved attributes might affect the lifetime utility households received from different neighborhoods come to mind, though persistent unobserved geographic preferences are particularly natural.

better.

Specifically, based on a household's unobservable type, we allow households to have preferences for sub-regions within the Bay Area. A household's unobservable type is denoted by $g_i \in \{1, 2, 3\}$, where $g_i = 1$ denotes a household with a preference for San Francisco and Marin (specifically San Francisco County, Marin County, and San Mateo County), $g_i = 2$ denotes a household with a preference for the South Bay (Santa Clara County), and $g_i = 3$ denotes a household with a preference for the East Bay (Alameda County and Contra Costa County). The corresponding ex-ante probabilities that households are of a particular type are given by: $\{\pi_1, \pi_2, \pi_3\}$; analogously, the sub-region in which neighborhood j is located is denoted by $G_j \in \{1, 2, 3\}$, where $G_j = 1$ if neighborhood j is in {San Francisco County, Marin County, San Mateo County}, $G_j = 2$ if neighborhood j is in {Santa Clara County}, and $G_j = 3$ if neighborhood j is in {Alameda County, Contra Costa County}.

This richer preference specification leads to a change in the first stage of the estimation approach described above. We parameterize the value functions such that $v_{j,t}^{\tau,g} = v_{j,t}^{\tau} + \phi 1_{[G_j = g]}$. That is, the lifetime expected utility that a household of type (τ, g) gets from living in a given neighborhood is composed two terms: $v_{j,t}^{\tau}$, the common component that anyone of type τ would receive from living in that neighborhood, plus an additional term, ϕ, that they only receive if the neighborhood is located in their preferred sub-region. This specification captures sub-regional attachments, perhaps related to place of work (the concentration of the semiconductor industry in sub-region 2, for example).

Using this specification for $v_{j,t}^{\tau,g}$, we can define unobserved-type-specific choice probabilities $(P_{j,t}^{\tau,g}, P_{0,t}^{\tau,g}, P_{stay,i,t}^{\tau,g})$ and unobserved-type-specific likelihood contributions $(L_{i,g}^{neigh}(\tilde{v}, \phi), L_{i,g}^{out}(\tilde{v}, \phi), L_{i,g}^{stay}(\tilde{v}, \phi, \gamma_{fmc}, \gamma_{pmc}))$ that are analogous to the no-unobserved-heterogeneity case. These equations can be found in Appendix B.

The overall log-likelihood is then the sum of the log of each household's likelihood contribution, where a household's likelihood contribution is the weighted sum of the household's unobserved-type-specific likelihood contribution over all observed decisions:

$$\mathcal{L}(\tilde{v}, \phi, \gamma_{fmc}, \gamma_{pmc}, \pi) = \sum_{i=1}^{N} log\Big(\sum_{g=1}^{3} \pi_g L_{i,g}^{neigh}(\tilde{v}, \phi) L_{i,g}^{out}(\tilde{v}, \phi) L_{i,g}^{stay}(\tilde{v}, \phi, \gamma_{fmc}, \gamma_{pmc})\Big) \quad (28)$$

The unobserved heterogeneity estimator is given by:

$$(\widehat{\tilde{v}}, \hat{\phi}, \hat{\gamma}_{fmc}, \hat{\gamma}_{pmc}, \hat{\pi}) = argmax_{(\tilde{v},\phi,\gamma_{fmc},\gamma_{pmc},\pi)} \mathcal{L}(\tilde{v}, \phi, \gamma_{fmc}, \gamma_{pmc}, \pi) \quad (29)$$

subject to the constraint:

$$\hat{\hat{P}}_{j,t}^{\tau} = \sum_{g=1}^{3} \pi_g P_{j,t}^{\tau,g}(\tilde{v}, \phi), \ \forall \ j \in \{0, ...J\} \quad (30)$$

While estimation is less simple than in the 'no unobserved heterogeneity' case, it is still straightforward. With the addition of unobserved heterogeneity, observations are not independent over time, and the log-likelihood is no longer additively separable in L^{neigh}, L^{out} and L^{stay}. Thus estimation is not done sequentially. The computational burden is still low, however. While searching for the parameters that maximize $\mathcal{L}$, for each guess of $(\phi, \gamma_{fmc}, \gamma_{pmc}, \pi)$, a contraction mapping finds the corresponding optimal vector of $(\tilde{v})$. This Berry (1994)-style contraction mapping is given by:

$$\tilde{v}_t^{\tau,r+1} = \tilde{v}_t^{\tau,r} + \log(\hat{\hat{P}}_t^{\tau}) - \log(\sum_{g=1}^{3} \pi_g P_t^{\tau,g}(\tilde{v}_t^{\tau,r}, \phi)) \quad (31)$$

Two key features of the data help identify ϕ and π. Taking these in turn, suppose that a household would receive a low value from living in neighborhoods in a particular sub-region, based on observable characteristics (and other parameters). If we observed this household choosing a neighborhood in the given sub-region, we would infer that it had a high unobserved taste for that sub-region. Furthermore, if this household never subsequently moved, we would further infer a strong taste for that region. If many such households are observed, this will identify a high value for ϕ. In terms of the identification of π, given $(\phi, \gamma_{fmc}, \gamma_{pmc}, \tilde{v})$, if households are more likely to initially choose a given sub-region or are less likely to leave a given subregion, this will identify a higher π for that sub-region.

4.4 Discussion: Measuring Wealth and Broader Life-Cycle Modeling Issues

Before turning to the results of our analysis, it is important to point out the strengths and limitations of our measure of 'wealth' and, consequently, the limits of our model compared to a life-cycle model of consumption and savings.

To date, the literature on residential sorting has been primarily static precisely because it has been so difficult to build a data set that follows a sizable sample of individuals over time as they move within a metropolitan area. By linking housing transactions data with information from loan applications, our analysis is the first to be able to track such movements for a large sample of households. Our resulting data set includes information on race, income, and the down-payments made by households at the time of purchase. While this represents a significant innovation relative to data sets used in prior work, several limitations remain: (i) we do not observe other household attributes such as family size or age, and (ii) we do not observe annual measures of income, wealth, consumption, etc. The latter would be necessary for the development of a richer life-cycle model of behavior.

In light of this limitation, the evolution of wealth plays a more limited role in our model of household decisions than it would in a fully-specified life-cycle model. The main dynamic captured in our model is that a household's wealth increases as its home appreciates and decreases in the event of a move, given that moving costs are substantial. In this way, our model captures the natural trade-off arising when it comes to the decision of whether to move: relocating allows households to re-optimize their location given accumulated changes in their wealth and the set of available neighborhoods, but comes at the expense of a sizeable moving cost (usually 6% of house value), which effectively lowers household wealth.

Given our lack of information about savings and consumption, we formulate the decision problem in the form of an indirect utility framework – i.e., we allow per-period utility to vary completely flexibly by household type, which depends on race, income, and wealth. This formulation allows us to use the dynamic trade-off embedded in the decision of whether to move (re-optimize across locations) to recover the marginal utility of wealth, along with marginal willingness to pay measures for neighborhood amenities. The latter measures are important for a range of research questions associated with hedonic valuation or cost-benefit analysis linked to policies that affect neighborhood amenities. That said, our current modeling approach might not be appropriate for answering certain kinds of questions related to important life-cycle issues – i.e., how changes in interest rates or other aspects of saving/borrowing technology would affect the housing market equilibrium.

5 Results

In this section, we present our baseline estimates of the model with unobserved heterogeneity. The model can be estimated separately for any given time-invariant observable household characteristic. The results we report here are based on estimating the second stage of the model (in which we regress per-period utility on a set of observable attributes) for three income types – \$40,000, \$120,000, and \$200,000 – and for whites only.[44] The estimation requires that we include many categories of wealth as a household's wealth transitions endogenously. In total, we have 225 types in the second stage: 25 wealth types measured in \$10,000 increments – \$0 to \$240,000 – interacted with the three income types and the three unobserved types.[45] When, for comparison, we estimate a version of the model without the three unobserved types, this reduces the total number of types to 75.

While there are several steps in the estimation, the primary results of interest are

[44] The process could easily be replicated for other racial groups, although small-numbers problems may be an issue in the first stage for other races. This could be addressed by pooling racial groups, at least where researchers were not seeking to estimate the value of racial composition by racial group.

[45] In the first stage, we want to use as much data as possible. As this requires assigning households to the nearest type, we include a much larger number of types. In that case, we use 25 income types and 25 wealth types (both measured in \$10,000 increments from \$0 to \$240,000) as well as the three unobserved types, when applicable.

the marginal willingness to pay estimates recovered in the final stage.[46] Given that our approach to controlling for the endogeneity of user cost makes use of estimates of the marginal utility of wealth, we also include a brief discussion of the moving cost results, as follows.

5.1 Moving Costs and the Marginal Utility of Wealth

We use the binary move/stay decision faced by each household in every period to identify and estimate the psychological and financial components of moving costs; exploiting the fact that financial moving costs are 6 percent of the selling price allows us to recover the marginal value of wealth as well. The results of this estimation are given in Table 3. From the table, it is clear that the psychological costs of moving are large, they decrease slightly in household income, and are falling over time.[47]

Table 3: Moving Cost Estimates

Psychological Costs	
Constant	9.50612
	(0.04344)
Income	-0.00209
	(0.00038)
t	-0.15111
	(0.00392)
Financial Costs	
Constant*6% House Value	0.03515
	(0.00148)
Income*6% House Value	-0.00008
	(0.00001)

Note: Income and House Value are measured in $1000's.

The financial cost estimates are of particular interest, given that they relate to the

[46] Transition probability estimates of (23) and (24) are provided in Tables A.3 and A.4 in the Appendix. As (23) is estimated separately for every type and includes both neighborhood dummies and neighborhood-specific time trends, there are too many parameters to report. Therefore, with $L = 2$, we form $\sum_{l=1}^{L} \rho_l^{\tau}$ for each type and report the percentiles across types.

[47] The mean psychological costs are high as they represent the amount a household would pay to avoid moving to a randomly chosen neighborhood in a randomly chosen time period. See Kennan and Walker (2011) for an excellent discussion of the interpretation of moving costs in this class of models.

marginal value of wealth. The estimates suggest that the marginal value of wealth is positive but considerably lower for high-income types. The marginal per-period utility of income coefficient that we take to the estimation of the final stage is given by 0.0315 - 0.00008*income. Our estimate of the marginal per-period utility of income is decreasing in income, as expected, and is roughly half as small for households with an income of $200,000 compared to those with an income of $40,000.

5.2 Marginal Willingness to Pay for Neighborhood Attributes

In the process of decomposing the estimates of flow utilities in the second stage, we control for the endogeneity of user cost by estimating equation (27). We decompose the flow utilities in two ways. First, we pool the mean utilities for all types and restrict the preference parameters for percentage white, violent crime, and ozone (as well as the county and year effects) to be same for all households.[48] Later, we relax these restrictions, estimating willingness to pay separately by income.[49]

The raw coefficients resulting from this process are reported in Table A.5 in Appendix A, but are difficult to interpret by themselves. Therefore, to better understand the magnitude of the coefficients, we calculate per-period willingness to pay for changes in each neighborhood characteristic. Per-period marginal willingness to pay (in $1000's) is given by $\alpha_x^\tau / \gamma_{fmc}^\tau$, which measures how much a household would be willing to pay annually to receive a given change in each of the amenities, holding expectations about future amenities constant.

Table 4 reports willingness-to-pay (WTP) measures for a 10-percent change in each amenity across four different specifications, the WTP figures being given at the means of percent white (69.6), the violent crime rate (453.7 per 100,000 residents), and ozone (2.2 days exceeding the state pollution standard). As the marginal utility of income, γ_{fmc}^τ, varies by income, we report the willingness to pay measures for a household with income of $120,000, which is close to the mean.

Our preferred specification is reported in column I. In this specification, we exclude

[48]This corresponds to estimating $u_{j,t}^\tau + \widehat{\gamma_{fmc}^\tau} usercost_{j,t} = \alpha_0^\tau + \alpha_c + \alpha_t + \tilde{X}'_{j,t}\alpha_x + \xi_{j,t}^\tau$.

[49]This corresponds to estimating $u_{j,t}^\tau + \widehat{\gamma_{fmc}^\tau} usercost_{j,t} = \alpha_0^\tau + \alpha_c^{inc} + \alpha_t^{inc} + \tilde{X}'_{j,t}\alpha_x^{inc} + \xi_{j,t}^\tau$.

the two highest and two lowest wealth categories (out of 25) and use a Least Absolute Deviations (LAD) regression to limit the effect of outliers. The results show that households with income of $120,000 are willing to spend $2,256.09 per year to increase percentage white in the neighborhood by ten percent at the mean. The estimates are very precise.[50] Analogously, households with $120,000 in income are willing to pay $760.33 for a ten-percent reduction in violent crime.[51] For ozone, households are willing to pay $359.89 for a ten-percent reduction in the number of days that ozone exceeds the one-hour state standard of 90 parts per billion.[52]

To shed light on the robustness of these WTP estimates, we also estimate the model using OLS instead of LAD in column II: the results there are quite similar. In columns III and IV, we estimate the model without excluding the two highest and two lowest wealth categories using LAD and OLS regressions, respectively. As can be seen from the table, the results are reasonably similar to those in column I.

Table 4: Willingness to Pay for a 10-Percent Increase in Amenities

	I	II	III	IV
Percent White	2256.09	2470.99	2188.18	2349.79
	(88.16)	(116.17)	(85.45)	(112.86)
Violent Crime	-760.33	-620.10	-725.19	-573.48
	(43.16)	(43.96)	(41.02)	(42.50)
Ozone	-359.89	-315.50	-347.14	-299.36
	(22.16)	(23.80)	(21.36)	(23.42)
County Dummies	Yes	Yes	Yes	Yes
Year Dummies	Yes	Yes	Yes	Yes
Type Dummies	Yes	Yes	Yes	Yes
Estimator	LAD	OLS	LAD	OLS
Wealth Outliers	NO	NO	YES	YES

[50] All the standard errors reported here and elsewhere in the paper were obtained using a bootstrap procedure with 240 draws.

[51] This willingness to pay implies a value of a statistical case of violent crime (similar in construction to the familiar value of a statistical life) of $1.6 million. This amount is consistent with other research on the costs of crime (Linden and Rockoff (2008)) and is reasonable in magnitude (i.e., approximately one fifth the size of a typical VSL estimate).

[52] The corresponding willingness to pay figures for one-unit changes in the three amenities are: $323.96 per year to increase percentage white by one percentage point, -$16.76 for one additional violent crime per 100,000 residents, and -$1657.65, for one extra day of ozone exceeding the threshold.

As a supplement to the pooled estimates in Table 3, we also estimate the willingness-to-pay measures separately for each of the three income types: \$40,000, \$120,000, and \$200,000. The results are presented in Table 5 and point to significant heterogeneity in willingness to pay for neighborhood amenities by income. For ease of exposition, we only show results for our preferred specification – i.e. using median regression and excluding the extreme wealth types.

The implied income elasticities of demand for neighborhood race are substantial, with a five-fold increase in income raising WTP for percent white almost eight times. Similarly for crime, a five-fold increase in income is estimated to increase WTP slightly less than a factor of four. In contrast, the implied income elasticities of demand for ozone are much smaller, with a five-fold increase in income only increasing WTP to avoid ozone by 31 percent.

Table 5: Willingness to Pay for a 10-Percent Increase in Amenities by Income

	\$40,000	\$120,000	\$200,000
Percent White	612.14	2428.91	4888.42
	(84.45)	(116.72)	(277.96)
Violent Crime	-350.15	-962.19	-1298.80
	(48.66)	(71.46)	(94.06)
Ozone	-302.06	-380.03	-395.58
	(28.30)	(30.12)	(39.32)

Recalling that $g_i = 1, 2, 3$ denotes households with preferences for subregion 1 (San Francisco and Marin), subregion 2 (the South Bay), and subregion 3 (the East Bay) respectively, the estimated probabilities of a household being of type g are $\hat{\pi} = \{.3625, .2286, .4088\}$. The estimate of the own-sub-region taste parameter is $\hat{\phi} = 1.8664$. For a household with annual income of \$120,000, this means that if a given neighborhood is in one's preferred sub-region, it increases the lifetime utility from living in that neighborhood by an amount equivalent to a one-time payment of \$73,372.

5.3 Model Fit

In the context of the location choice literature, the estimated model does reasonably well at predicting neighborhood choices for households of a given type. In particular, 31 percent of households choose a neighborhood that would have been ranked in the top 5 percent of their choices and 48 percent choose one from the top 10 percent of ranked choices when the static idiosyncratic errors are excluded from the prediction.[53] Adding unobserved individual heterogeneity to the model improves the fit significantly. In particular, for the model without unobserved heterogeneity, 23 and 37 percent of households choose a neighborhood in the top 5 and top 10 percent of their ranked choices, respectively. Adding unobserved geographic heterogeneity to the model improves the fit because it naturally accounts for factors such as employment location that clearly affect individual location decisions.

Given that household types are distinguished only by race, income, and wealth in our empirical model, it would be surprising if the non-idiosyncratic components of the model were able to explain location decisions much better, given that many other factors (e.g., employment locations, education, family structure, age, etc.) certainly play a large role in individual location decisions. In this respect, our results are very close in spirit to those presented in Kennan and Walker (2011) for their analysis of inter-state migration decisions. In particular, Kennan and Walker show that idiosyncratic errors are also an important driver of individual location decisions in that context. In the location choice literature more generally, idiosyncratic errors naturally play an important role in any model with a large number of choices. While these models cannot perfectly predict the location decision of any individual, they can generally predict quite well the types of locations that households with certain characteristics tend to select, as our model does here.

[53] 75 percent of households choose a neighborhood from the top quartile of their ranked choices.

6 Interpreting the Estimates

In this section, we contrast our preferred estimates from a dynamic model including unobserved heterogeneity with estimates first from a static model, and then with a dynamic model without unobserved heterogeneity.

6.1 Dynamic Versus Static Approaches

In this sub-section, we compare the MWTP estimates from a dynamic model with those from a corresponding static model. The goal of this exercise is to highlight the nature of the potential biases that might results from ignoring dynamic considerations by using models that are otherwise identical. It is important to note, however, that this does not imply that specific parameter estimates that we report in this paper are necessarily better than some of the existing estimates in the literature based on static approaches. In particular, the static estimates that we report in this nested comparison fall short of the static 'gold standard' in at least two respects. First, several well-known static choice models are estimated using confidential census data; and the rich cross-sectional information there certainly affords a very detailed view of individual attributes – race, education, family structure, even work location – not available in the dynamic panel we have assembled for this paper. The second respect in which our comparable static estimates fall short is that the amenities are treated as exogenous in or analysis – i.e., no explicit research design is utilized to exploit exogenous amenity variation.

Equation (8) illustrates the difficulty associated with estimating a static model when the true model is dynamic: in essence, specifying a static model creates an omitted variables problem. In a dynamic setting, current neighborhood characteristics determine the choice-specific value functions in two ways: (i) they affect flow utility directly, and (ii) they help predict future neighborhood utility. Estimating a static model omits the latter effect.

The static model can be conveniently nested within our dynamic framework. The former effectively assumes that the household will always stay in the same location and that attributes will never change; therefore, the location-specific value function remains constant over time. As such, one interpretation of the static model is that $v^{\tau}_{j,t} = u^{\tau}_{j,t}/(1-\beta)$. The omitted variable can then be expressed as:

$$\beta E\Big[log\Big(\sum_{k=0}^{J+1} exp(v^{\tau_{t+1}}_{k,t+1} - MC^{\tau_{t+1}}_{j,t+1} 1_{[k \neq J+1]})\Big)\Big| s_{i,t}, d_{i,t} = j\Big] - \beta v^{\tau}_{j,t} \tag{32}$$

The specific way a given current characteristic predicts future utility will determine whether the static estimator over- or under-predicts the effect of that characteristic on per-period utility. If higher values of a given characteristic predict improvements in a neighborhood, then the marginal willingness to pay for that attribute will be biased towards positive infinity, which is the case for all three amenities we consider (public safety, air quality, and neighborhood percent white).[54]

To set out the relevant intuition, it is useful to think about the current value of an attribute predicting future exposure to that attribute, rather than future utilities. There are two primary components to this term: how the amenity will change if the household does not move, and what the amenity exposure will look like if the household does move. We first consider the non-move case. Take a disamenity, such as violent crime, that is mean-reverting; as in our data, a high level of crime today predicts falling crime in future. In this case, we would expect the static model to understate the disutility of crime. The argument is as follows: households may be willing to pay quite large amounts to avoid high levels of crime. However, when they see a neighborhood with a high value of this disamenity, they know that the value is likely to fall in the future; and they are therefore willing to pay much more for a house in that 'bad' neighborhood than they would be willing to pay if the high value of the disamenity were permanent. The upshot is that the estimated willingness to pay to avoid crime taken from a naive static model will tend to be biased downward. The same type of argument applies to air pollution – in our data, ozone levels are also mean-reverting.

There are other neighborhood attributes that are persistent over time, such as racial

[54]Given that our actual measures are for crime and pollution – disamenities rather than amenities – a positive bias in the relevant coefficient means that the absolute effect of crime and pollution on utility will be biased downwards – i.e. the static results will suggest households have a weaker distaste for those disamenities.

composition. In contrast to ozone and crime, seeing a high percentage of whites in a neighborhood today signals that the neighborhood is more likely to have an even higher percentage white in the future. If these are attributes that households value (and recall that we are only modeling the decisions of white households), they will be willing to pay more for a house in such a neighborhood than they would be if the high value of the attribute were only temporary – in other words, persistent amenities are likely to be worth more than fleeting ones. A naive static model ignores this fact and attributes all of the value to current preferences, thereby overstating the contribution to flow utility of high percentage white neighborhoods for white households.

The second determinant of future exposure is expected exposure to the amenity if one moves. The choice today can determine the probability of moving in future periods and can also influence the future neighborhood choice through wealth effects. Having said that, given a choice between two neighborhoods this period, positive predicted mobility will always push expected future exposure to an amenity to be more similar across the neighborhoods in the dynamic model compared with the static model where future mobility is assume to be zero. The effect of this is to *underestimate* the taste for the amenity.

To highlight the problems associated with ignoring forward-looking behavior, we estimate a static version of our model for comparison purposes. Under the assumption that agents are not forward-looking, a fraction of Stage 1 estimates (i.e. $v^{\tau}_{j,t}(1-\beta)$) can be interpreted as flow utilities. We can then decompose those flow utilities by running the same Stage 2 procedure used to decompose $u^{\tau}_{j,t}$ above. In particular, we estimate equation (27), replacing $u^{\tau}_{j,t}$ with $v^{\tau}_{j,t}(1-\beta)$. The marginal utility of income is recovered in Stage 1 and is still equal to γ^{τ}_{fmc}. By using the same marginal utility of income coefficient as in our dynamic specification, we keep the models as comparable as possible and limit any bias to the coefficients relating to the amenities. Here, even if the researcher were to incorrectly assume the model to be static, she would control for the correlation between price and unobserved neighborhood attributes using an Instrumental Variables approach. Yet if the true model is dynamic, the chosen instrument will typically not be valid.[55]

[55] The problem with the IV strategy is that if the true model is actually dynamic, any static instrument

Table 6: Willingness to Pay for a 10-Percent Increase in Amenities – Static versus Dynamic Estimates by Income

	Static			Dynamic		
	$40,000	$120,000	$200,000	$40,000	$120,000	$200,000
Percent White	1627.02	1901.43	2221.66	612.14	2428.91	4888.42
	(11.28)	(18.76)	(48.55)	(84.45)	(116.72)	(277.96)
Violent Crime	-291.14	-380.67	-448.88	-350.15	-962.19	-1298.80
	(7.68)	(11.08)	(19.02)	(48.66)	(71.46)	(94.06)
Ozone	-66.24	-80.71	-97.04	-302.06	-380.03	-395.58
	(2.13)	(2.43)	(3.15)	(28.30)	(30.12)	(39.32)

Table 6 reports the marginal willingness to pay for a 10-percent change in each amenity derived from the static version of the model, where willingness to pay varies with income. The earlier dynamic results from Table 5 are also included for ease of reference. As before, the marginal willingness to pay figures are reported at the means of the amenity levels.

The comparison of static and dynamic results in the table suggests that incorrectly estimating a static model in a dynamic context can lead to very biased estimates. The static model substantially overestimates willingness to pay for living in close proximity to neighbors of the same race for low-income households: the static estimate is $1,627.02 whereas the corresponding dynamic estimate is $612.14. For high-income households, the bias runs in the opposite direction and the static model underestimates the willingness to pay by a factor of more than two. The biases for both crime and air pollution are such that the static model always underestimates the willingness to pay. For low-income households, the static estimates are -$291.14 in the static case and -$350.15 in the dynamic case, respectively, for a 10-percent increase in violent crime. The magnitude of the bias grows significantly with income and the corresponding figures for high-income households are -$448.88 and -$1,298.80. In the case of pollution, for low-income house-

will be correlated with expected future utility, which is subsumed in the error term. In particular, any potential instrument must satisfy the condition that it should be correlated with the endogenous variable – in this case, price. Now, expected future utility is a function of all current attributes. Therefore, unless current price has no predictive power with respect to future utility, it will be impossible to find an instrument that is both correlated with price but also uncorrelated with expected future utility.

holds, the static estimates are -$66.24 versus -$302.06 in the dynamic case, again for a 10-percent increase in ozone, with little change in the relative magnitude of the bias as income changes. In each case, the differences are substantial and are precisely estimated. As can be inferred from the above discussion, the income elasticities implied by the static model are substantially smaller for both race and crime relative to the elasticities from the dynamic model, apparent from the much steeper profiles in the dynamic case; for ozone, the respective elasticities are fairly similar.

The signs of these biases are consistent with the discussion above. From a different angle, it is also interesting to see whether these biases can be partially explained by patterns in the actual variables themselves – instructive as the omitted variable given by the equation is only known once the full structural model has been estimated. Regressions exploring the time-series patterns of each (dis)amenity are shown in Table A.6 in the Appendix; also relevant to the bias discussion are the predicted move probabilities, which are 0.083, 0.12, 0.16 for the three income types, $40,000, $120,000, and $200,000, when averaged across neighborhoods, time, and wealth/geographic preferences types.

For all results, the time-series patterns are consistent with the biases found: for pollution and crime, higher ozone and crime rates this period predict larger falls (or smaller increases) in ozone and crime one period ahead. As the move probabilities are positive and get larger with income, we should predict that the static model would always be biased downward and that the bias would grow with income. This is almost exactly what we see.[56]

For race, higher percent white this period predicts larger increases (or smaller reductions) in percent white one period ahead. The race-persistence effect and the move-probability effect go in opposite directions, so one cannot predict the bias. However, the static model bias should vary with income and the static model will be most likely to understate the willingness to pay when income is high, which again is what we find.

It is worth emphasizing that the transition probability regressions shown in Table A.6 were not used in our estimation routine; there, we used lagged attributes to predict future value functions directly, rather than using them to predict future amenities. As

[56]The high-income bias for ozone is no bigger than the medium-income bias.

we do not use the time-series patterns of the amenities directly in our estimation, we draw additional confidence from the fact that our empirical results are consistent with these patterns.

6.2 Geographic Heterogeneity

Table 7 reports the marginal willingness to pay for a 10-percent change in each amenity derived from the dynamic version of the model where we do not control for unobserved geographic heterogeneity. Similar to before, the willingness to pay varies with income, and the earlier dynamic results from Table 5 are also included for ease of reference.

Table 7: Willingness to Pay for a 10-Percent Increase in Amenities – No Geographic Heterogeneity versus Geographic Heterogeneity Estimates by Income

	No Geographic Heterogeneity			Geographic Heterogeneity		
	\$40,000	\$120,000	\$200,000	\$40,000	\$120,000	\$200,000
Percent White	149.72	1361.77	3016.52	612.14	2428.91	4888.42
	(80.53)	(105.98)	(189.85)	(84.45)	(116.72)	(277.96)
Violent Crime	-304.50	-700.51	-847.92	-350.15	-962.19	-1298.80
	(45.63)	(69.49)	(86.71)	(48.66)	(71.46)	(94.06)
Ozone	-331.56	-406.07	-390.40	-302.06	-380.03	-395.58
	(27.66)	(31.06)	(42.43)	(28.30)	(30.12)	(39.32)

As can be seen from Table 7, with the exception of willingness to pay for ozone, failing to control for unobserved geographic heterogeneity results in non-trivial underestimates of willingness to pay in absolute terms. This bias-towards-zero feature is most pronounced for high-income households.

A revealed preference interpretation analogous to the static bias story above provides an intuitive way to think about these patterns. For a given set of neighborhood location decisions, a larger long-run difference in exposure to crime maps into a smaller estimated absolute distaste for crime. The overall transition probabilities for the neighborhood values are very similar in both models. However, the likely destination next period is different when we control for unobserved heterogeneity. In particular, if a household is placed in a random neighborhood, the heterogeneity model predicts that they are more likely to move, which would result in smaller differences in exposure to crime when geographic preferences are included. This suggests that removing the heterogeneity controls should reduce the willingness to pay in absolute value.

Most of the willingness to pay coefficients are substantially larger in the heterogeneity model, as noted. And the model does predict higher mobility when controlling for heterogeneity. The predicted income-specific move probabilities are 0.083, 0.12, 0.16 in the heterogeneity model, but only 0.038, 0.061, and 0.089 in the no-heterogeneity model, which is consistent with higher willingness to pay estimates in the heterogeneity model and the notion that this difference would increase in income.[57]

The results of the static model without controls for unobserved geographic heterogeneity are almost identical to the static results reported in Table 6 and can be found in Table A.7 in the Appendix. The similarity arises because, for any level of geographic sub-region preferences, the model predictions for within-sub-region shares are identical. Therefore, once sub-region dummies are included as controls, a regression of estimated utilities (simply log estimated shares in the static model) on neighborhood characteristics will yield the same coefficients on the neighborhood amenities and result in the same willingness to pay estimates.[58] In other words, if the true model is dynamic, then the regional preferences predict different choice probabilities through a mechanism that the static model omits and, as such, the static model cannot make the correct inference.

7 Conclusion

While models of residential sorting and hedonic equilibrium have a been the focus of a substantial body of research, almost all existing empirical studies based on these models have adopted a static estimation approach. This is with good with reason: compu-

[57]The actual mobility rates for the three groups are remarkably similar: 0.0767, 0.0752, and 0.0763. As these just represent data, they do not vary across dynamic models. However, as the identification comes from comparing counterfactual neighborhoods, the neighborhood-specific model-predicted mobility rates are what is important.

[58]The fact that we have any change in the static results occurs because we impose a different price coefficient (which we estimate from the move/stay decision) and because when we add unobserved heterogeneity, thus the move/stay decisions play an additional minor role in identifying $\tilde{v}$.

tational and data issues have made the estimation of dynamic models extraordinarily difficult. Yet location decisions are inherently dynamic, and this has led to concerns that the estimates from static models may be biased.

In this paper, we developed a tractable dynamic model of neighborhood choice that controls for unobserved household and neighborhood heterogeneity, along with a computationally straightforward semi-parametric estimation approach. Our neighborhood choice model and estimator adapt dynamic demand models for durable and storable goods for use in a housing market context, and build on this class of models in several ways: (i) treating the house as an asset and allowing household wealth to evolve endogenously, (ii) using stable, uniform realtor fees to estimate the marginal utility of consumption without the need for a price instrument, and (iii) relaxing the strong assumption about the evolution of the continuation value that is standard in the existing literature.

With our model and estimation approach in hand, we merged very rich transactions data with publicly-available mortgage application data to create a data set that matches the attributes of many buyers and sellers to homes. We then used this dynamic data set to estimate household preferences in a manner consistent with forward-looking behavior. The estimates we obtained indicate that the biases associated with static demand estimation are significant for three important non-marketed amenities: air quality, crime, and neighborhood race. Further, the signs of the biases we find are consistent with what one would expect, based on the time-series properties of each amenity and likely exposure to amenities upon moving in future. We also show that the willingness to pay estimates from a dynamic model without heterogeneity are understated, compared to our preferred specification, in a manner consistent with predicted mobility rates.

Given the importance of accounting for such dynamic considerations when estimating preferences for non-marketed goods, the model, data set, and estimation procedure presented in this paper have potentially broad applicability to the study of a range of dynamic phenomena in housing markets and cities. These include examining the microdynamics of residential segregation and gentrification within metropolitan areas – applications we intend to pursue in related research using the approach in this paper.

References

Aguirregabiria, V., and A. Nevo (2010): "Recent Developments in Empirical Dynamic Models of Demand and Competition in Oligopoly Markets," mimeo, University of Toronto.

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," *Journal of Political Economy*, 115(4), 588–638.

Bayer, P., N. Keohane, and C. Timmins (2009): "Migration and Hedonic Valuation: The Case of Air Quality," *Journal of Environmental Economics and Management*, 58(1), 1–14.

Bayer, P., R. McMillan, and K. Rueben (2011): "An Equilibrium Model of Sorting in an Urban Housing Market," mimeo, Duke University.

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.

Bishop, K. (2007): "A Dynamic Model of Location Choice and Hedonic Valuation," mimeo, Washington University in St. Louis.

Bishop, K. C., and C. Timmins (2012): "Estimating the Marginal Willingness to Pay Function Without Instrumental Variables," mimeo, Washington University in St. Louis.

Caetano, G. (2010): "Identification and Estimation of Parental Valuation of School Quality in the US," mimeo, University of Rochester.

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

Cragg, M., and M. Kahn (1997): "New Estimates of Climate Demand: Evidence from Location Choice," *Journal of Urban Economics*, 42(2), 261–284.

Ekeland, I., J. J. Heckman, and L. Nesheim (2004): "Identification and Estimation of Hedonic Models," *Journal of Political Economy*, 112(1), S60–S109.

Ellickson, B. (1971): "Jurisdictional Fragmentation and Residential Choice," *American Economic Review*, 61(2), 334–339.

Epple, D. (1987): "Hedonic Prices and Implicit Markets: Estimating Demand and Supply Functions for Differentiated Products," *Journal of Political Economy*, 95(1), 59–80.

Epple, D., R. Filimon, and T. Romer (1984): "Equilibrium among local jurisdictions: Toward an integrated treatment of voting and residential choice," *Journal of Public Economics*, 24(3), 281–308.

Epple, D., and R. Romano (1998): "Competition between private and public schools, vouchers, and peer-group effects," *American Economic Review*, 88(1), 33–62.

Epple, D., R. Romano, and H. Sieg (2006): "Admission, Tuition, and Financial Aid Policies in the Market for Higher Education," *Econometrica*, 74(4), 885–928.

——— (2010): "The Intergenerational Conflict over the Provision of Public Education," mimeo, Carnegie-Mellon University.

Epple, D., and T. Romer (1991): "Mobility and Redistribution," *Journal of Political Economy*, 99(4), 828–858.

Epple, D., and H. Sieg (1999): "Estimating Equilibrium Models of Local Jurisdictions," *Journal of Political Economy*, 107, 645–681.

Ferreyra, M. (2007): "Estimating the Effects of Private School Vouchers in Multi-District Economies," *American Economic Review*, 97(3), 789–817.

Gibbons, S. (2004): "The Costs of Urban Property Crime," *Economic Journal*, 114, F441–F463.

Gowrisankaran, G., and M. Rysman (2012): "Dynamics of Consumer Demand for New Durable Goods," *Journal of Political Economy*, 120, 1173–1219.

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

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

Kahn, M. (1995): "A Revealed Preference Approach to Ranking City Quality of Life," *Journal of Urban Economics*, 38(2), 221–235.

KENNAN, J., AND J. R. WALKER (2011): "The Effect of Expected Income on Individual Migration Decisions," *Econometrica*, 79(1), 211–251.

KUMINOFF, N. V. (2008): "Recovering Preferences from a Dual-Market Locational Equilibrium," mimeo, Arizona State University.

KUMINOFF, N. V., AND A. S. JARRAH (2010): "A New Approach to Computing Hedonic Equilibria and Investigating the Properties of Locational Sorting Models," *Journal of Urban Economics*, 67(3), 322–335.

LINDEN, L., AND J. ROCKOFF (2008): "Estimates of the Impact of Crime Risk on Property Values from Megan's Laws," *American Economic Review*, 98(3), 1103–1127.

MELNIKOV, O. (2013): "Demand for Differentiated Durable Products: The Case of the U.S. Computer Printer Market," *Economic Inquiry*, 52(2), 1277–1298.

MURPHY, A. (2007): "A Dynamic Model of Housing Supply," mimeo, Washington University in St. Louis.

NECHYBA, T. J. (1999): "School Finance Induced Migration and Stratification Patterns: The Impact of Private School Vouchers," *Journal of Public Economic Theory*, 1(1), 5–50.

——— (2000): "Mobility, Targeting, and Private-School Vouchers," *American Economic Review*, 90(1), 130–146.

ORTALO-MAGNÉ, F., AND S. RADY (2002): "Tenure Choice and the Riskiness of Non-Housing Consumption," *Journal of Housing Economics*, 11(3), 266–279.

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

——— (2008): "Heterogeneity Within Communities: A Stochastic Model with Tenure Choice," *Journal of Urban Economics*, 64(1), 1–17.

ROSEN, S. (1974): "Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition," *Journal of Political Economy*, 82(1), 34–55.

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

SCHIRALDI, P. (2011): "Automobile Replacement: A Dynamic Structural Approach," *RAND Journal of Economics*, 42, 266–291, mimeo, Boston University.

Sieg, H., V. K. Smith, H. S. Banzhaf, and R. Walsh (2004): "Estimating the General Equilibrium Benefits of Large Changes in Spatially Delineated Public Goods," *International Economic Review*, 45, 1047–1077.

Timmins, C. (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*, 47(1), 1–25.

Walsh, R. (2007): "Endogenous Open Space Amenities in a Locational Equilibrium," *Journal of Urban Economics*, 61(2), 319–344.

Appendix A: Tables

Table A.1: Comparison of Sample Statistics for Merged Data and IPUMS Data by Bay Area County

	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	125646	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						
Percent White	49.85	68.27	90.65	59.12	60.08	49.07
Percent Asian	28.68	10.55	4.68	31.47	26.57	34.21
Percent Black	6.45	6.01	0.67	2.08	1.22	1.45
Percent Hispanic	11.76	12.38	2.51	5.86	9.90	12.27
IPUMS						
Percent White	47.64	64.57	87.5	61.92	58.1	50
Percent Asian	27.34	11.37	3.3	23.37	25.41	33.51
Percent Black	7.77	6.05	2.3	2.8	1.24	1.16
Percent Hispanic	14.62	14.2	3.62	8.18	12.5	12.09

Table A.2: Comparison of Sample Statistics for Transactions Data and Merged Data

Transactions Data

Variable	Obs.	Mean	Std. Dev.	Min.	Max.
Sale Price	1045920	354915	220886	16500	1521333
Lot Size	1045920	6857	11197	0	199940
Square Footage	1045920	1647	714	400	10000
Number Bedrooms	1045920	2.98	1.10	0	8
Number Rooms	1045920	6.73	2.00	1	18

Merged HMDA/Transactions Data

Variable	Obs.	Mean	Std. Dev.	Min.	Max.
Sale Price	804699	372240	212938	16513	1521204
Lot Size	804699	6730	10605	0	199940
Square Footage	804699	1649	687	400	10000
Number Bedrooms	804699	3.01	1.08	0	8
Number Rooms	804699	6.77	1.98	1	18

Table A.3: Value Function Transition Probability Parameters

	25^{th} %ile	50^{th} %ile	75^{th} %ile
v^{τ}	-0.39205	-0.21532	-0.16618
Sales Price	-0.00300	-0.00178	-0.00084
Percent White	-0.03833	-0.03137	-0.02455
Violent Crime	-0.00015	-0.00008	-0.00004
Ozone	-0.01203	-0.01075	-0.00778

Note: %iles of $\sum_{l=1}^{2} \rho_l^{\tau}$ across τ as defined in (23).

Table A.4: Price Transition Probability Parameters

	Price_{t+1}
Sales Price_t	0.385
	(0.221)
Percent White_t	1.643
	(1.283)
Violent Crime_t	-0.045
	(0.012)
Ozone_t	0.362
	(0.447)
Sales Price_{t-1}	-0.245
	(0.021)
$\text{Percent White}_{t-1}$	1.438
	(1.484)
$\text{Violent Crime}_{t-1}$	0.053
	(0.013)
Ozone_{t-1}	1.916
	(0.393)
Neighborhood Dummies	Yes
Neighborhood Time Trends	Yes

Note: Estimates of ϱ as defined in (24).

Table A.5: Decomposition of Flow Utilities – Raw Parameter Estimates

	I	II	III	IV
Percent White	0.00824	0.00903	0.00799	0.00858
	(0.00026)	(0.00034)	(0.00025)	(0.00033)
Violent Crime	-0.00043	-0.00035	-0.00041	-0.00032
	(0.00002)	(0.00002)	(0.00002)	(0.00002)
Ozone	-0.04217	-0.03697	-0.04068	-0.03508
	(0.00297)	(0.00292)	(0.00286)	(0.00287)
County Dummies	Yes	Yes	Yes	Yes
Year Dummies	Yes	Yes	Yes	Yes
Type Dummies	Yes	Yes	Yes	Yes
Estimator	LAD	OLS	LAD	OLS
Wealth Outliers	NO	NO	YES	YES

Table A.6: Time-Series Properties of Amenities

	Δ Percent White$_{t+1}$	Δ Violent Crime$_{t+1}$	Δ Ozone$_{t+1}$
Percent White$_t$	0.0156	-0.4466	0.0089
	(0.0014)	(0.1123)	(0.0027)
Violent Crime$_t$	0.0010	-0.1520	-0.0016
	(0.0001)	(0.0083)	(0.0002)
Ozone$_t$	-0.0162	-0.5958	-0.6566
	(0.0105)	(0.8491)	(0.0201)
County Dummies	Yes	Yes	Yes
Year Dummies	Yes	Yes	Yes

Note: For the dependent variable, $\Delta X_{t+1} = X_{t+1} - X_t$.

Table A.7: Willingness to Pay for a 10-Percent Increase in Amenities by Income – Static Model with No Geographic Heterogeneity

	\$40,000	\$120,000	\$200,000
Percent White	1605.19	1840.00	2171.76
	(10.29)	(14.73)	(37.62)
Violent Crime	-277.59	-357.94	-434.80
	(7.02)	(9.70)	(15.77)
Ozone	-65.57	-78.23	-95.85
	(2.00)	(2.16)	(2.91)

Appendix B: Unobserved Heterogeneity – Choice Probabilities and Likelihood Contributions

The analogous unobserved-type-specific choice probabilities to the 'no unobserved heterogeneity' case are defined as follows:

$$P_{j,t}^{\tau,g} = \frac{e^{\tilde{v}_{j,t}^{\tau} + \phi 1_{[G_j = g]}}}{\sum_{k=1}^{J} e^{\tilde{v}_{k,t}^{\tau} \phi 1_{[G_k = g]}}} \tag{A.1}$$

$$P_{0,t}^{\tau,g} = \frac{e^{\tilde{v}_{0,t}^{\tau}}}{\sum_{k=0}^{J} e^{\tilde{v}_{k,t}^{\tau} + \phi 1_{[G_k = g]}}} \tag{A.2}$$

$$P_{stay,i,t}^{\tau,g} = \frac{e^{\tilde{v}_{J+1,t}^{\tau} + \phi 1_{[G_{J+1} = g]}}}{e^{\tilde{v}_{J+1,t}^{\tau} + \phi 1_{[G_{J+1} = g]}} + \sum_{k=0}^{J} e^{\tilde{v}_{k,t}^{\bar{\tau}} + \phi 1_{[G_k = g]} - FMC_{i,t} \gamma_{fmc}^{\tau} - Z_{i,t}' \gamma_{pmc}}} \tag{A.3}$$

Similarly, we can define unobserved-type-specific likelihood contributions as:

$$L_{i,g}^{neigh}(\tilde{v}, \phi) = \prod_{j=1}^{J} (P_{j,t_{1,i}}^{\tau,g})^{1_{[d_{i,t_{1,i}} = j]}} \tag{A.4}$$

$$L_{i,g}^{out}(\tilde{v}, \phi) = (P_{0,t_{2,i}}^{\tau,g})^{1_{[d_{i,t_{2,i}} = 0]}} (1 - P_{0,t_{2,i}}^{\tau,g})^{1_{[d_{i,t_{2,i}} \in \{1,\ldots,J\}]}} \tag{A.5}$$

$$L_{i,g}^{stay}(\tilde{v}, \phi, \gamma_{fmc}, \gamma_{pmc}) = \prod_{t=t_{1,i}+1}^{t_{1,i}+T_i} (P_{stay,i,t}^{\tau,g})^{1_{[d_{i,t} = J+1]}} (1 - P_{stay,i,t}^{\tau,g})^{1_{[d_{i,t} \neq J+1]}} \tag{A.6}$$