Segmented Housing Search*

Monika Piazzesi
Stanford & NBER

Martin Schneider
Stanford & NBER

Johannes Stroebel
University of Chicago
Booth School of Business

September 2012

Abstract

This paper asks how heterogeneous but interconnected housing markets adjust to shocks. We use a novel data set on email alerts from the popular real estate website Trulia.com to construct measures of buyer search patterns. We then build a quantitative model of housing market search that can account for the joint distribution of turnover, inventory and search patterns in San Francisco Bay Area. We use the model to infer the distribution of searcher preferences as well as the matching technology. We show that overlapping search patterns play an important role in the transmission of shocks across segments of the housing market.

*HIGHLY PRELIMINARY & INCOMPLETE. Email addresses: piazzesi@stanford.edu, schneidr@stanford.edu, johannes.stroebel@chicagobooth.edu. We thank Giuseppe Moscarini for helpful comments. We are greatful to Trulia. We also thank seminar participants at the Northwestern-Tsinghua Conference, Society of Economic Dynamics, and Yale University. This research is supported by a grant from the National Science Foundation.

1 Introduction

In recent years there has been substantial heterogeneity in US households' housing market experiences, even within the same metro area. Consider the San Francisco Bay Area: cheaper East Bay communities still have slow markets at depressed prices, whereas around Silicon Valley activity is high and prices are rising. Nevertheless, the two regions are close enough geographically that their housing markets are likely to be connected. Suppose, for example, that Silicon Valley locations become more desirable because of new jobs there. Some searchers might settle for a longer commute and buy in the East Bay. Conversely, suppose an East Bay neighborhood becomes less stable because the population there experiences more income shocks. Some houses might be bought by searchers who have previously looked largely around Silicon Valley.

This paper asks how heterogeneous but interconnected housing markets adjust to shocks. We start from micro data on market activity. In particular, we use a novel data set on email alerts from the popular real estate website Trulia.com to construct measures of buyer search patterns. We then build a quantitative model of housing market search that can account for the joint distribution of turnover, inventory and search patterns in San Francisco Bay Area. We use the model to infer the distribution of searcher preferences as well as the matching technology. We find that more expensive neighborhoods are more stable and are less desirable in the sense that there are fewer searcher per house. We also ask what happens if preferences change to affect the desirability or stability of a neighborhood. Here the main result is that overlapping search patterns play an important role in the transmission of shocks.

The paper proceeds in two steps. First, it uses a comprehensive micro data set to document stylized facts on housing market activity in the San Francisco Bay Area since 2008. We rely on county deeds and assessment records to develop measures of volume and housing stock, respectively. We use a feed of house listings to construct measures of inventory and time on market. In addition, the website Trulia.com allows searchers to set an alert that triggers an email whenever a house with the searcher's desired characteristics comes on the market. We consider a large sample of alerts and aggregate alerts by their searcher's identification number to infer the distribution of search profiles.

The most important criterion for search patterns is geography: many searchers look for houses in contiguous areas. To the extent that searchers look at houses that are physically distant from each other, those houses tend to lie in similar price ranges. Our data also shows that the range of houses considered varies substantially with searchers' desired area. In particular, searchers who scan expensive urban areas tend to look at substantially more inventory than searchers who look at less expensive or more suburban areas.

To analyze the effect of search behavior on activity, we pool the data at the zipcode level. We construct monthly series of turnover (volume divided by housing stock), inventory share (inventory divided by housing stock) and time on market (mean time between listing and sale for houses sold that month) for each of 183 Bay area zipcodes, We show that the cross sectional means of turnover, inventory and time on market are all mutually positively correlated, both over our entire sample 2008-2011 and for individual years. In particular, more expensive neighborhoods tend to have lower inventory and lower turnover than cheaper neighborhoods.

Expensive neighborhoods in dense urban areas – San Francisco, Berkeley and Silicon Valley – also have lower time on market than cheaper neighborhoods; this is what drives the overall correlations. At the same time, in some expensive suburbs – for example in Marin and Contra Costa counties – low inventory and turnover go along with higher time on market. Interestingly, expensive urban neighborhoods also see the largest number of search alerts.

Our model describes heterogeneous agent types who buy and sell houses in a set of market segments. An agent type is identified with a desired area – a subset of market segments – that the agent searches over in order to find his "favorite" house. An agent's search is faster the bigger the fraction of the housing stock that is listed in his desired area. Once a favorite house has been bought and provides utility, it may randomly fall out of favor. The agent then puts the house on the market and searches for a new favorite. The steady state equilibrium of the model delivers a mapping from the distribution of preferences into the joint distribution of turnover, inventory, time on market and price by segment, as well as the distribution of buyers by type.

When we quantify the model, the set of segments corresponds to Bay Area zipcodes, and the set of possible agent types is derived from our email alert data. Indeed, we code every search profile in the alert data as a subset of the set of all zipcodes. The distribution of preferences consists of the number of agents of each type, as well as the rates at which favorite houses fall out of favor. Aggregated to the segment level, these two sets of parameters capture the concepts of desirability and stability. A more desirable segment has a larger clientele of agents searching over it. In a more stable segment, houses fall out of favor more slowly.

The stability and desirability of segments can be identified from data on market activity. Indeed, in more desirable segments we should see more searchers (and hence more email alerts), higher turnover, lower inventory and lower time on market. Intuitively, a larger potential buyer pool more quickly absorbs any houses that come on the market, thus increasing turnover and leaving less inventory. Also, more stable segments have lower turnover and lower inventory. Since houses come on the market at a slower rate, there are fewer houses for sale and fewer transactions relative to the total housing stock.

The effect of stability on time on market depends on how strongly the rate of finding a favorite house responds to inventory. With a "strong demand effect", that is, an increase in inventory substantially speeds up search, then stable segments – where houses come on the market more slowly – also see slower trading. In contrast, with a weaker demand effect, the lower inventory in stable segments leads to faster trading.

The above intuition on identification guides our calibration strategy. We select the distribution of preferences to match the model implied distribution of search patterns, together with moments of time on market and volume. The model is tractable enough to infer a high dimensional parameter vector and check a large number of overidentifying restrictions. We find that the model does a reasonable job of matching the cross section of market activity. Moreover, the results indicate a strong demand effect. According to our parameter values, expensive areas tend to be more stable – agents there are less likely to move out. Moreover, expensive urban areas tend to be less desirable in the sense that there are at any point in time fewer buyers scanning inventory.

To illustrate the role of search patterns for the transmission of shocks, we then proceed to perform comparative statics exercises. In particular, we ask how time on market and inventory change if the supply of houses in a segment increases. The answer crucially depends on the number of searchers and what other markets those searchers look at. For example, shocks to a downtown San Francisco zipcode with many searchers who search broadly is transmitted widely across the city. In contrast, shocks to a suburban zipcode close to the San Francisco city boundary has virtually no effect on the market in the city itself.

Related Literature. Our paper provides the first model in which potential buyers search for a house in different segments of the market. Their search patterns may integrate different housing segments and thereby create commonality among these markets. Alternatively, the search patterns may lead to fully segmented housing markets that do not have common features. Our paper contributes to a literature that has investigated the implications of search models for a single market (e.g., Wheaton 1990, Krainer 2001, Caplin and Leahy 2008, Novy-Marx 2009, Piazzesi, and Schneider 2009, Burnside, Eichenbaum, and Rebelo 2011.)[1] Landvoigt, Piazzesi and Schneider (2012) develop an assignment model to study different housing segments. Their model has implications for the relative volume of various segments, but not for overall volume or the behavior of time on the market. Van Nieuwerburgh and Weill (2010) study the predictions of a dynamic spacial model for the dispersion of wages and house prices across U.S. metropolitan areas. Empirical studies (e.g., Poterba 1991, Bayer, Ferreira, and McMil-

[1] Recent models of a single housing market with frictions *other than* search include Piazzesi and Schneider (2012), Favilukis, Ludvigson and Van Nieuwerburgh (2012), and Glover, Heathcote, Krueger, and Rios-Rull (2012).

lan 2007, Mian and Sufi 2010) document the importance of determinants such as credit constraints, demographics, or school quality in different housing markets. More related to our paper, Genovese and Han (2012) document the number of homes that actual buyers have visited on their house hunt, but without knowing the location or other characteristics of these homes, which are key elements in our work.

2 Data Description and Summary Statistics

To conduct the empirical analysis, we combine a number of key datasets. The first dataset contains the universe of ownership-changing deeds in the Bay Area between 1994 and 2011. The property to which the deeds relate is uniquely identified via the Assessor Parcel Number (APN). The variables in this dataset that we use in this project include property address, transaction date, transaction price, type of deed (e.g. Intra-Family Transfer Deed, Warranty Deed, Foreclosure Deed), and the type of property (e.g. Apartment, Single-Family Residence). The second dataset contains the universe of tax-assessment records in the Bay Area for the year 2009. Properties are again identified via their APN. This dataset includes information on property characteristics such as construction year, owner-occupancy status, lot size, building size, and the number of bedrooms and bathrooms. The tax assessment records allow us to construct the housing stock within each segment at every point in time.

The third dataset includes the universe of all property listings on the online real estate search engine Trulia.com between October 2005 and December 2011. The key variables in this dataset that we use are the listing price and the listing address. First, by using the address of properties to match listings data to deeds data, we can construct a measure of the time on market for each property that eventually sells. In addition, by combining listings data and sales data, we can also construct a measure of the total inventory of houses listed for sale in a particular segment at a particular point in time.

Finally, we observe data on the user search behavior on Trulia.com, one of the largest U.S. online real estate search engines (with 24 million monthly unique users.) Visitors to Trulia.com can search for houses that are currently listed for sale by specifying a number of search criteria such as geography, price range and property type. After having refined their search criteria to fit their true preferences, users then have the option to set an email alert, which would send them an email every time a new house with their preferred characteristics would come on the market. Figure 1 shows part of a screen shot of Trulia's email alert sign-up page.

We obtained a dataset that includes a random sample of approximately $60,000$ such email alerts set by $30,800$ unique users searching for houses in the San Francisco Bay

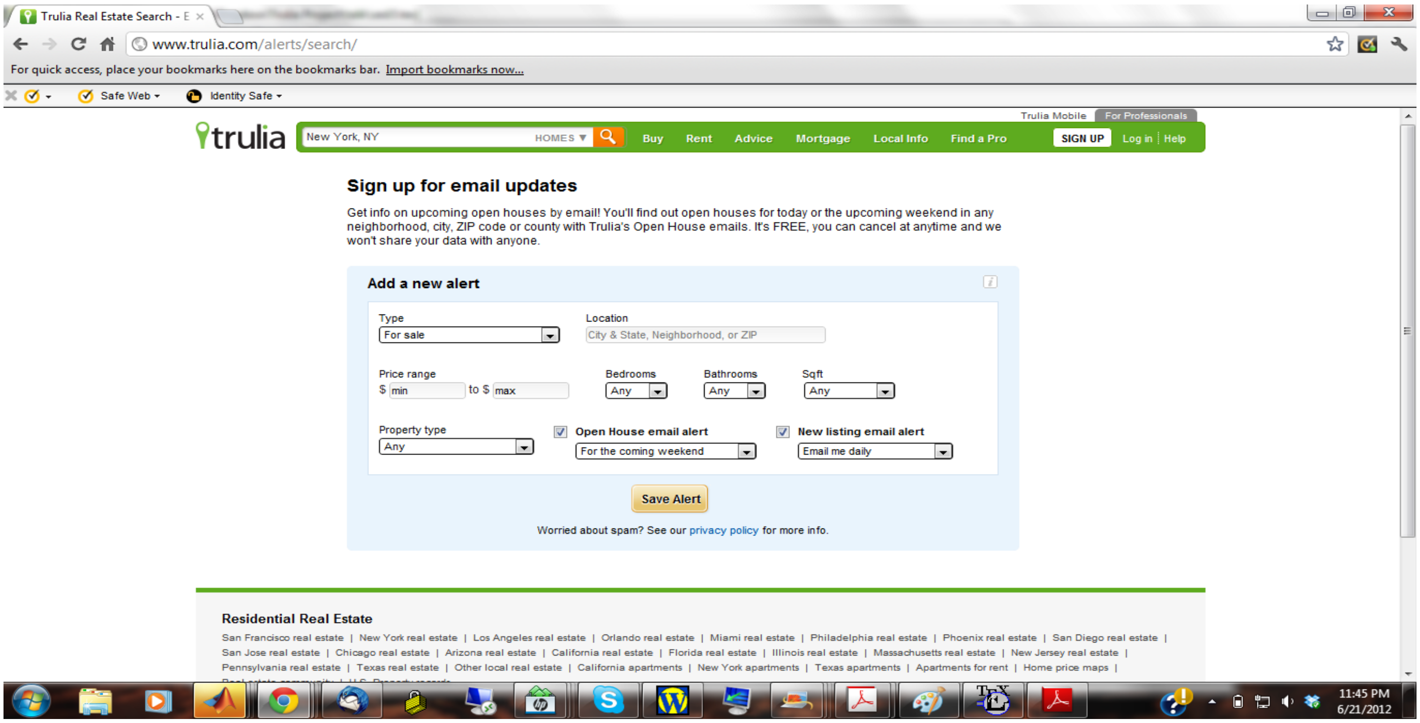

Figure 1: Screenshot – Email Alert Sign-Up. This is a screenshot of the website where users can set email alterts.

Area between March 2006 and April 2012.[2]

2.1 Construction of Key Variables

The purpose of this project is to determine how home buyers search across different segments of the housing market. In order to do so, a first step involves the definition of market segments. As a first pass, we define a segment of the housing market to be a zip code (alternative definitions could be zip code / price range combinations, or zip code / property type combinations). We can then construct our key variables of interest by zip code for each time period. Appendix A describes this construction in more detail.

[2] We believe that analyzing these email alerts is useful to learn about users' search behavior, in particular relative to considering the entire search behavior on the website. This is because search on the website involves some learning about supply in each market. People often start with a much broader or much narrower search than their final email alert - we hence interpret the final email alert to contain information about their true search radius. In addition, many people browse the website without an intention to purchase. Given the (small) cost of receiving an email alert, we believe that these alerts are more likely to include well-considered preferences from buyers with a serious possible interest.

2.1.1 Segment Specific Moments

For each segment and time period, we determine (1) the median sales price, (2) the total volume of transactions, (3) the average time on market for houses sold in that time period and (4) the total inventory of listings per segment and time period. For each time period, price by zip code is determined by finding the median observed transaction price in each zip code. Volume for each time period is constructed by calculating the total number of transactions in each zip code. We normalize volume in a zip code by dividing it by the total stock of residential housing in the zip code in 2009, as determined from the tax assessment records. Time on market for each property is calculated by measuring the time period between the initial listing date and the final sale. We then calculate the average time on market for all properties sold in that particular time period.

Inventory levels are first constructed at the monthly level. Annual levels are constructed by finding the average across the twelve months in the year. We start by adding all newly listed properties in that zip code to the inventory observed in the previous month, as long as the listing occurred in the first half of the month or in the second half of the previous month. In addition, all listings that result in a sale in the first half of the month as observed in the deeds data get removed from the inventory.

Table 1 shows summary statistics for our key time series variables. In 2011, the average zip code had a transaction volume of about 0.4% of the housing stock transacting each month. The average across our zip codes in terms of the median 2011 sales price was about $515,000. The average zip code had about 1.2% of the total housing stock listed for sale, and houses that eventually sell would be on the market for about 3.48 months.

Figure 2 illustrates the behavior of time on market, turnover, and inventory across the 183 Bay Area zip codes in our sample. Each dot in the figure is a zip code. The top panel shows time on market and turnover for each zip code, averaged over time. Time on market is measured in months, while turnover is measured in percent. The two variables are clearly positively correlated, with a correlation coefficient of 30% – zip codes in which houses once they are listed sell fast are zip codes in which many houses get sold overall. The colormap indicates the (log) house price in the zip code, ranging from expensive (pink) to cheap (blue). The colors reveal that more expensive zip codes have low turnover and time on the market, while the cheaper zip codes have high turnover and time on the market. This negative correlation between prices and turnover/time on the market is a new stylized fact about the cross section of housing markets. Interestingly, this stylized fact goes against conventional intuition about 'hot and cold housing markets', in which houses in hot markets sell fast, at high prices, and many are sold, while cold markets are dead – prices are low, nothing sells, and houses sit on the market for a long time.

The bottom panel in Figure 2 shows that inventory (in percent) is also positively related to time on the market, and negatively related to house prices. In expensive neighborhoods, few houses get listed and once a house is listed, it sells fast. In cheaper neighborhoods, there is more inventory and houses sit on the market for a longer time.

Table 1: Summary Statistics Across Zip Codes

	Inventory (% of total stock)	Volume (% of total stock)	Median Price (in $1,000 USD)	Average Time on Market (in months)
Year 2011				
Mean	1.32%	0.40%	515	3.48
Median	1.26%	0.37%	460	3.52
Std.	0.54%	0.12%	262	0.73
Min	0.05%	0.15%	100	1.59
Max	3.01%	0.81%	1,450	6.92
Year 2010				
Mean	1.50%	0.41%	546	3.28
Median	1.35%	0.36%	509	3.28
Std.	0.64%	0.14%	260	0.71
Min	0.13%	0.16%	110	1.60
Max	3.59%	0.85%	1,585	6.01
Year 2009				
Mean	1.57%	0.42%	530	4.06
Median	1.40%	0.37%	494	4.03
Std.	0.74%	0.18%	250	0.92
Min	0.09%	0.17%	85	1.17
Max	4.03%	1.00%	1,550	6.72
Year 2008				
Mean	1.93%	0.38%	616	3.74
Median	1.60%	0.33%	573	3.84
Std.	1.04%	0.15%	281	1.07
Min	0.19%	0.15%	130	1.59
Max	5.58%	0.98%	1,700	6.45

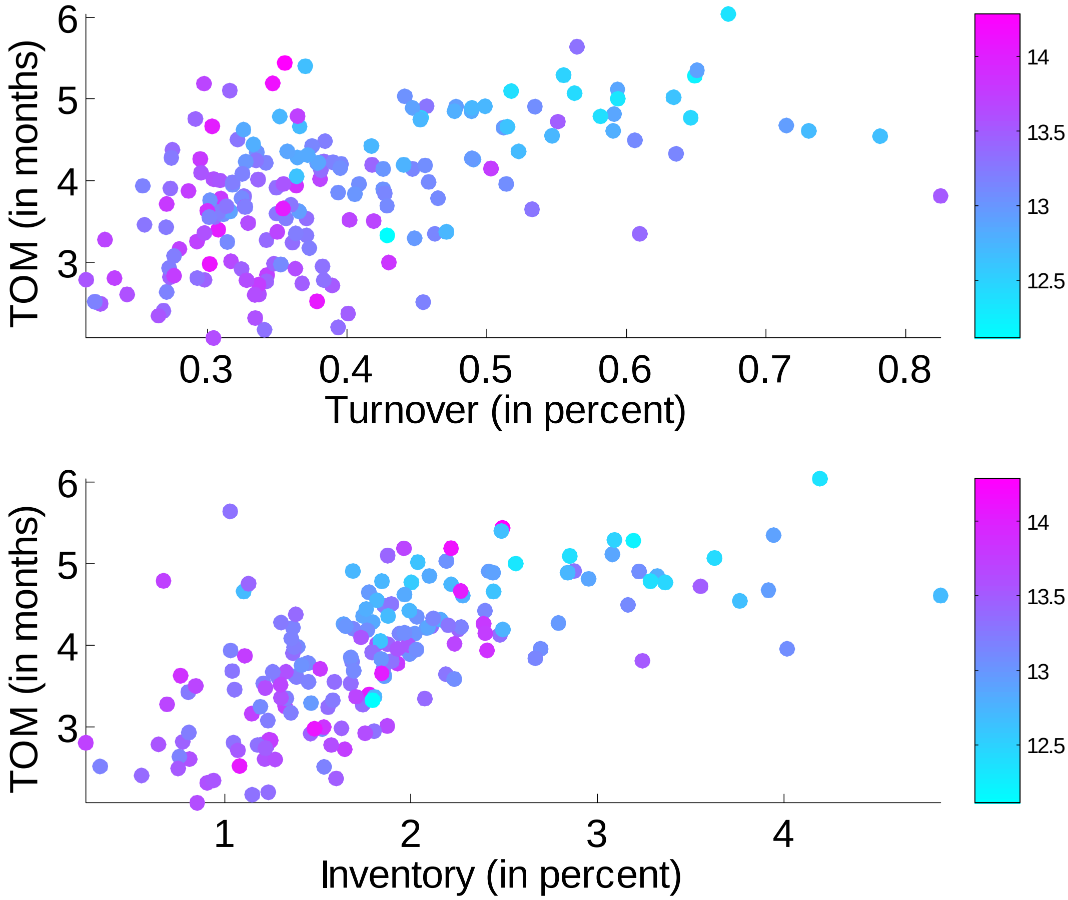

Figure 2: The top panel plots average time on market (in months) against turnover (as a fraction of total housing stock per month.) The bottom panel plots average time on market against inventory (as a fraction of the total housing stock.)

Figure 3 illustrates these cross sectional facts on a map of the Bay Area. The top left panel shows that turnover is low (blue) in dense, urban areas — the downtown areas of San Francisco, Berkeley, and San Jose. (In the map, downtown San Francisco is the blue large ball in the northwest, Berkeley is the blue area across the Bay and east of San Francisco, and San Jose is the large blue area in the southeast.) In contrast, turnover is high (pink) in other areas of the Silicon Valley (which on the map is between San Francisco and San Jose.) The top right panel shows time on the market. In the dense urban areas, houses sell fast, so time on the market is low (blue). In contrast, in expensive but rural areas – Marin and Contra Costa counties – houses sit on the market longer (pink dots with time on the market). On the map, Marin is north of San Francisco, while Contra Costa county is in the far northeast. The bottom left panel

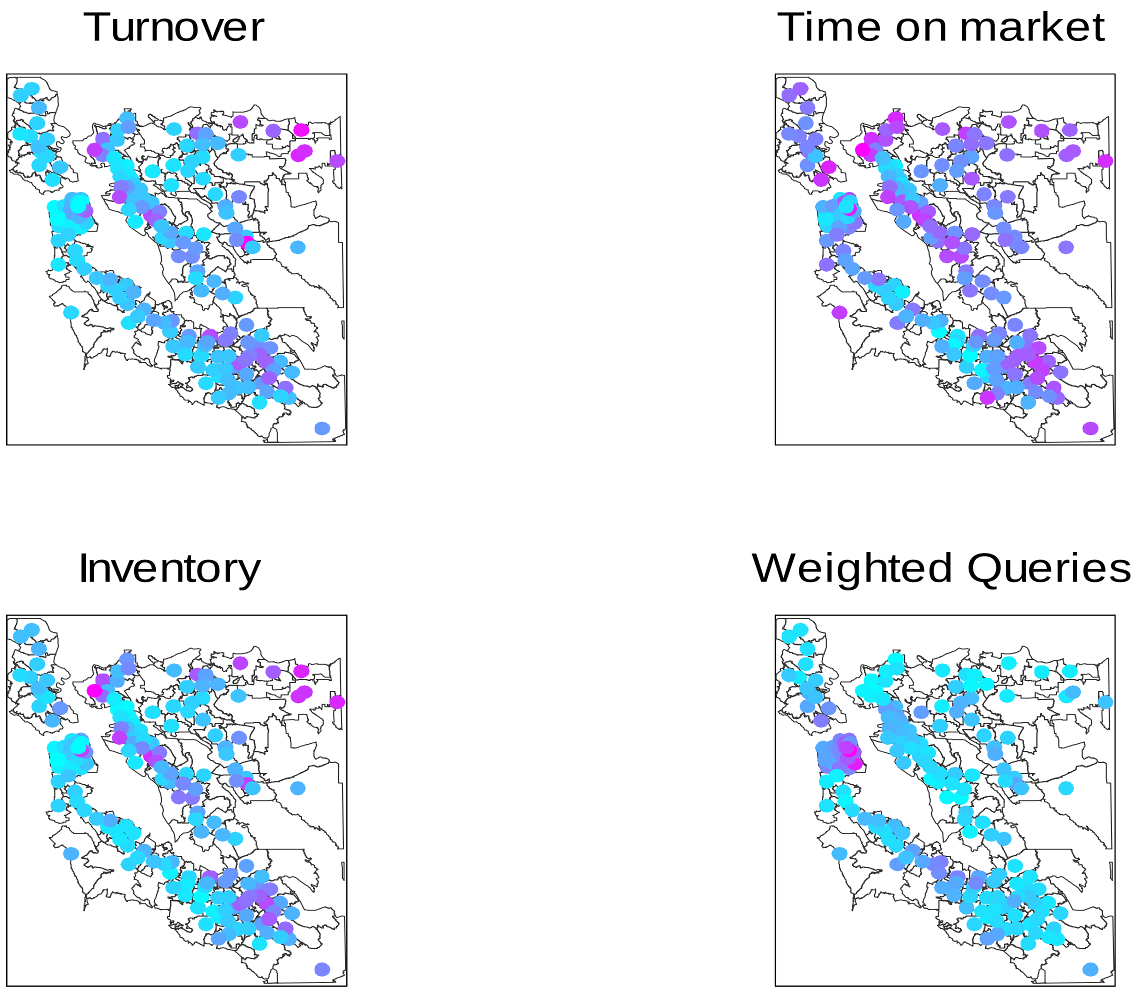

Figure 3: Top left panel: mean turnover in each zip code. Top right panel: mean time on the market in each zip code. Bottom left panel: mean inventory. Bottom right panel: number of search queries weighted by the inverse amount of housing stock that they scan.

shows inventory, which is low (blue) in the urban areas and thus behaves similar to turnover (in the top left panel.) The bottom right panel illustrates the number of clients weighted by the inverse of the amount of housing inventory they scan. This plot shows that the dense, urban areas – especially San Francisco – are searched most heavily.

2.1.2 Search Patterns

To construct the search pattern distribution, we determine for each email alert which zip codes are be covered by the alert. Alerts either specify that they are looking for properties that have recently been sold (16%), that have recently been listed for sale (68%) and that have recently been listed for rent (16%). Each alert defines the desired search geography by including one or more cities, zip codes or neighborhoods (e.g. the Mission district in San Francisco). About 71% of alerts include at least one geography in terms of cities (e.g. Palo Alto, Menlo Park, San Jose), 20% of searchers defined their search range by providing preferred zip codes and the remaining 17% of alerts specify at least one the neighborhood. Many queries include a number of cities, zip codes or neighborhoods in the query. Table 2 shows a number of cross tabulations of the alerts. 64.4% of alerts do not specify either a neighborhood or a zip code, but only a city. About 2.8% of queries specify one zip code and one city. About 0.2% of listings specify more than 5 neighborhoods and more than 5 cities.

Since we are taking the zip code as our measure of a segment, we need to deal with alerts that specify geography at a unit that might not perfectly overlap with zip codes. For those searchers who scan listings at the city level, we assign all zip codes that are at least partially within the range of the city to be covered by the search query (i.e. for a searcher who is looking in Mountain View, we assign the query to cover the zip codes 94040, 94041 and 94043). Neighborhoods and zip codes also do not line up perfectly, and so for each neighborhood we again all zip codes that are at least partially within the neighborhood to be covered by the search query (i.e. for a searcher who is looking in San Francisco's Mission District, we assign the query to cover zip codes 94103 and 94110). We then merge all queries by the same individual to get a sense of the entire geographic area considered by the searcher.

In addition to geography, 57% of listings also provide a price range of interest. A zip code is determined to be covered by an email alert if the price range overlaps with a range of 0.9 to 1.1 times the median transaction price in that zip code and quarter. To determine whether the price range quoted in "for rent" listings covers a particular zip code, we scale the price range assuming a price to annual rent ratio of 15.

Aggregating over all the email alerts set by a particular searcher, we can now analyze how many zip codes are considered by each searcher. When only considering geography, the average email alert considers about 8.9 zip codes, the median alert considers 4 zip codes. When also taking into account the price range specified by the agents, the average query covers 5.7 zip codes, the median 2. The 30,755 unique searchers generate 9,162 unique search patterns when including information in price, and 5,701 unique search patterns when discarding that information. This means that some searcher set alerts that cover identical areas – many of these are alerts that cover just one city, such as

San Jose. The difference between the unique patterns with and without price is driven by the fact that some queries that specify the same city might cover different zip codes within that city depending on the price range selected.

Table 2: Summary Statistics of Search Queries

	Number of Zip Codes					
# Neighborhoods	0	1	2	3	4	5+
0	64.44	16.61	0.75	0.34	0.22	0.62
1	7.96	0.47	0.05	0.02	0.01	0.03
2	0.69	0.06	0.03	0.01	0	0
3	0.68	0.06	0.02	0	0	0.01
4	0.64	0.06	0.01	0.01	0	0
5+	5.63	0.43	0.03	0.03	0.02	0.06

	Number of Cities					
# Neighborhoods	0	1	2	3	4	5+
0	14.81	61.37	1.54	0.07	0.95	3.24
1	6.81	1.33	0.05	0.06	0.07	0.21
2	0.60	0.06	0.04	0.02	0.01	0.07
3	0.63	0.05	0.02	0.01	0.01	0.05
4	0.62	0.04	0.02	0.01	0.01	0.04
5+	5.59	0.23	0.09	0.05	0.04	0.20

	Number of Cities					
# Zip Codes	0	1	2	3	4	5+
0	13.19	59.98	1.55	1.05	0.97	3.3
1	14.36	2.78	0.11	0.09	0.07	0.28
2	0.57	0.16	0.03	0.02	0.01	0.09
3	0.27	0.06	0.02	0.01	0.01	0.03
4	0.18	0.02	0.01	0.02	0.01	0.02
5+	0.48	0.08	0.03	0.03	0.02	0.10

Note: This table shows the percentage of alerts that select a certain number of geographic characteristics.

Figure 4 gives an impression of the breadth in housing search by the median client in a given zip code. Each dot in the figure represents a zip code. On the horizontal axis, the plot measures the inventory in that zip code. On the vertical axis, the plot measures the amount of overall inventory scanned by the median client active in that zip code. The black line is the 45-degree line. To understand the pattern that we see in the figure, it is useful to think about the following two extreme cases: full segmentation versus full integration. Suppose housing markets were fully segmented. In that case, all dots would be exactly on the 45-degree line. The reason is that the median client in any given zip code would be limiting his search to that zip code alone and thus only be scanning the inventory in that zip code. Suppose instead that housing markets were fully integrated. In that case, all dots would be on a horizontal line equal to the total housing inventory in the Bay Area. The reason is that the median client in any given zip code would be searching the entire Bay Area inventory. From Figure 4, we can see that the search data are somewhat in-between these two extremes. We can find zip codes in which the median client scans a large amount of housing inventory. There are other zip codes which are on the 45-degree line, where indeed the median client just scans the inventory in that zip code.

A key feature of the search alerts is that they allow us to uncover patterns of search that go beyond what would have been possible by only considering geographic distance between zip codes. In particular, a number of search alerts cover zip codes that are geographically far apart, but might be similar along dimensions that a particular searcher cares about (such as price.) One way to illustrate this is to define "clientele distance" between each pair of zip-codes: for each zip code pair we look at all the queries that cover at least one of the two zip codes, and then find the fraction of those queries that cover both. The maximum clientele distance is 1, when every person who searches in one of the two zip codes also searches in the other. The minimum clientele distance is 0, when nobody who searches in either zip code also searches in the other. A higher clientele distance means more people searching jointly across two zip codes.

Our data sample contains 183 zip codes, so that there are clientele distances between each of the (183x182)/2 zip code pairs. About 12% percent of the pairs in the Bay Area have a clientele distance of 0, suggesting that not a single query covers them both. An example of such a pair is 94024 (Los Altos Hill, south Bay, average home sale in 2010 above $1.7 million) and 94564 (Pinole, very north of Bay, average home sale in 2010 around $250,000). However, the other 88% of zip code pairs have at least one query covering them both. The density of clientele distance is relatively left-skewed, with many zip codes pairs having only very few queries covering them both. The mean clientele distance between a zip code pair is 0.044, which means that about 4.4% of queries covering either zip code actually cover both. About 5% of zip code pairs (at total of 832) have a clientele distance of at least 0.35, suggesting that they are regularly searched jointly.

To give a sense of the geographic range of queries, we calculated for each query the

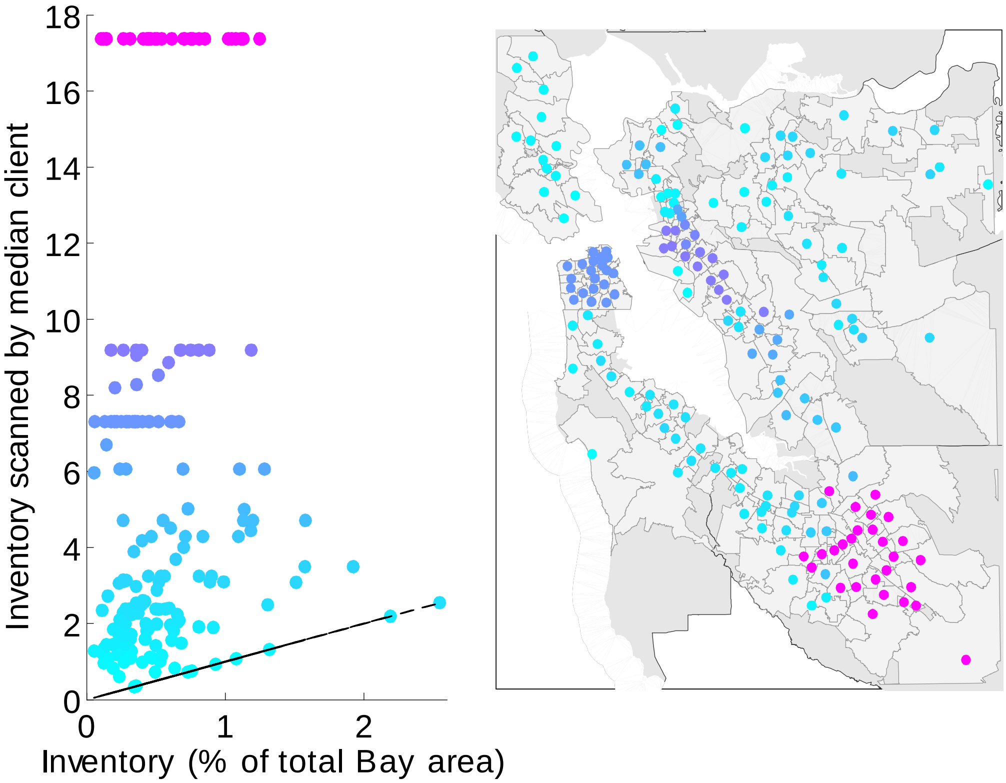

Figure 4: Breadth of Search. Each dot indicates the inventory of the zip code on the horizontal axis together with the total inventory scanned by the median client who is active in that zip code on the vertical axis. Both inventories are measured in percent of the total housing stock. The colors range from pink (expensive) to blue (cheap.)

distance (in kilometers) between each of the zip code centroids covered by the query (excluding those queries that only cover one zip code). We can then calculate the average and maximum distance between zip codes covered by each query. For example, if a query covers zip codes A, B and C, we can calculate the average distance of A and B, B and C and A and C. The median query covers zip codes that are, on average 5.73 kilometers apart. The average distance between zip codes covered by a query rises to about 15 kilometers for the 90th percentile of queries. We also look at the maximum distance between any two zip codes covered by a query. On average, the maximum distance of two zip codes in a query is about 12 kilometers. About 40% of queries cover two zip codes at least 10 kilometers apart, 17.3% of queries cover zip codes that are at least 20 kilometers apart, 6% of queries cover zip codes that are at least 30 kilometers apart and about 2.6% of queries cover zip codes that are over 40 kilometers apart. Apart from geographic distance, the range of price matters for queries. Indeed, when queries involve two zip codes that are far apart geographically, they are similar in price. Keeping clientele-distance fixed, a one-standard deviation (25km) increase in the distance between two zip code centroids reduces the price distance by about a third of a standard deviation.

3 Setup

Time is continuous and the horizon is infinite. The model describes a small open economy, such as the San Francisco Bay Area. Agents discount the future using the discount factor $\beta = e^{-r}$, where r is the riskless interest rate. Let H denote a finite set of market segments. It is helpful to think of a segment as a combination of location and quality (for example, all houses in Oakland in a certain price range) The measure μ^H on H counts the number of houses in each segment. We normalize the total number of houses to one.

Let Θ denote a finite set of agent types. Agents have quasilinear utility over two goods: numeraire ("cash") and housing services. Agents own at most one house. They obtain housing services if and only if they live in their favorite house, which they find through search. The house an agent lives in may randomly cease to be his favorite house; he can then put it on the market in order to sell it and search for a new favorite house. We assume that search is costless, whereas putting a house on the market costs c per period.

Agent θ is identified by a *search profile*, that is, a subset $\tilde{H}(\theta) \subset H$ of market segments that he is interested in. An agent of type $\theta \in \Theta$ knows that his favorite house lies in $\tilde{H}(\theta)$. For example, he might care about commuting distance to their employer, Facebook, and thus search only near the company's Menlo Park headquarters, subject to a certain price range. When a type θ agent searches for a favorite house, he scans all inventory available for sale in $\tilde{H}(\theta)$. Let $\mu^S(h)$ denote the number of houses that have been put on the market in segment $h \in H$. The *inventory scanned* by type θ is then

$$\tilde{\mu}^S(\theta) := \sum_{h \in \tilde{H}(\theta)} \mu^S(h) \tag{1}$$

Type θ agents find a favorite house at the rate $f_\theta\left(\tilde{\mu}^S(\theta)\right)$, where f_θ is a strictly increasing function with $f_\theta(0) = 0$. In other words, agents who scan more inventory find a favorite house more quickly.

The favorite house is equally likely to be any house in the scanned inventory. Once it has been identified, the agent learns that he will receive per period utility $\tilde{v}$ from living there, where $\tilde{v}$ is an iid draw from a cdf $F_{\theta,h}$ that can depend on both segment and type.

The agent contacts the seller, who observes $\tilde{v}$ and makes a take-it-or-leave-it offer. If the buyer rejects the offer, the seller keeps the house and the buyer continues searching. If the buyer accepts the offer, the seller starts to search, whereas the buyer moves into the house and begins to receive utility $\tilde{v}$. The house ceases to be his favorite house at the rate $\eta(h, \theta)$, and thereafter yields no utility.

The measure μ^Θ on Θ counts the number of agents of each type. The total number of agents is

$$\bar{\mu}^\Theta = \sum_{\theta \in \Theta} \mu^\Theta(\theta) > 1.$$

Since there are more agents than houses and agents own at most one house, some agents are always searching.

In equilibrium, agents make optimal decisions taking as given the distribution of others' decisions. In particular, owners decide whether or not to put their houses on the market, sellers choose price offers and buyers choose whether or not to accept those offers. In what follows, we focus on steady state equilibria in which (i) owners who live in their favorite house do not put it on the market, (ii) owners who do not live in their favorite house put their house on the market, (iii) all offers are accepted. Such an equilibrium exists as long as the utility flow $\tilde{v}$ from living in a favorite house is sufficiently high.

Characterizing equilibrium quantities

Since the model has a fixed number of agents and houses, the steady state distributions of agent states can be studied independently of the prices and value functions. We need notation for the number of agents who are in different states. Let $\mu^F(h;\theta)$ denote the number of type θ agents who live in their favorite house in segment h. Let $\mu^S(h;\theta)$ denote the number of type θ agents who put their house on the market in segment h. Finally, let $\tilde{\mu}^B(\theta)$ denote the number of type θ agents who are "buyers", that is, they are currently searching. In steady state, all these numbers are constant. We now derive a set of equations to determine them.

The number of houses in segment h that type θ agents put on the market for the first time is $\eta(h;\theta)\,\mu^F(h;\theta)\,dt$. Indeed, $\mu^F(h;\theta)$ is the number of type θ agents who live in a favorite house in segment h and those house fall out of favor at the rate $\eta(h;\theta)$. For $\mu^S(h,\theta)$ to remain constant over time, the number of houses brought on the market by type θ agents in segment h must equal the number of houses that are sold by type θ agents in segment h. Writing $m(h)\,dt$ for total volume in segment h, we have

$$\eta(h;\theta)\,\mu^F(h;\theta) = \frac{\mu^S(h;\theta)}{\mu^S(h)} m(h)\,, \tag{2}$$

where the right hand side is the number of houses $m(h)$ that are sold in segment h

times the fraction of the inventory in segment h that is held by agents of type θ. Since all sellers are equally likely to sell at any time, the fraction $\mu^S(h;\theta)/\mu^S(h)$ is also the probability that a house in segment h is sold by a type θ agent.

Moreover, for $\mu^F(h;\theta)$ to remain constant over time, the number of houses sold by type θ agents in segment h must equal the number of houses bought by type θ agents in market h. We write

$$\frac{\mu^S(h;\theta)}{\mu^S(h)} m(h) = \frac{\mu^S(h)}{\tilde{\mu}^S(\theta)} f_\theta\left(\tilde{\mu}^S(\theta)\right)\tilde{\mu}^B(\theta). \quad (3)$$

Here the right hand side uses the number $\tilde{\mu}^B(\theta)$ of type θ buyers, their likelihood $f_\theta\left(\tilde{\mu}^S(\theta)\right)$ of finding a favorite house, so that a number $f_\theta\left(\tilde{\mu}^S(\theta)\right)\tilde{\mu}^B(\theta)$ of type θ agents find a favorite house. The fraction $\mu^S(h)/\tilde{\mu}^S(\theta)$ is the inventory of houses in segment h relative to the entire inventory scanned by agent θ. Since the buyer is equally likely to find his favorite house in any market that he scans, this fraction is also the probability that agent θ finds his favorite house in segment h among all segments $\tilde{H}(\theta)$ that he scans.

It is helpful to define the *clientele* of segment h as all agents who scan segment h:

$$\tilde{\Theta}(h) := \left\{\theta \in \Theta : h \in \tilde{H}(\theta)\right\}$$

The number of agents and the number of houses must add up to their respective totals.

$$\begin{aligned} \mu^H(h) &= \sum_{\theta\in\tilde{\Theta}(h)} \left(\mu^F(h;\theta) + \mu^S(h;\theta)\right) \\ \mu^\Theta(\theta) &= \tilde{\mu}^B(\theta) + \sum_{h\in\tilde{H}(\theta)} \left(\mu^F(h;\theta) + \mu^S(h;\theta)\right) \end{aligned} \quad (4)$$

The ($2\ \#H\ \#\Theta + \#H + \#\Theta$) equations (2), (3) and (4) jointly determine the ($2\ \#H\ \#\Theta + \#H + \#\Theta$) numbers $\mu^F(h;\theta), \mu^S(h;\theta), m(h)$ and $\tilde{\mu}^B(\theta)$.

Equilibrium prices

Since sellers make a take-it-or-leave offers, they charge a price equal to the buyers' continuation utility. Denote by $V^F(h;\theta,\tilde{v})$ the utility of a type θ agent who lives in a favorite house in segment h and receives flow utility $\tilde{v}$. The Bellman equations of that agent as well as that of a seller are

$$\begin{aligned} rV^F(h;\theta,\tilde{v}) &= \tilde{v} + \eta(h;\theta)\left(V^S(h;\theta) - V^F(h;\theta,\tilde{v})\right), \\ rV^S(h;\theta) &= -c + \frac{m(h)}{\mu^S(h)}\left(E\tilde{p}(h) - V^S(h;\theta)\right), \end{aligned}$$

where $E\tilde{p}(h)$ is the mean of cross sectional distribution of prices in segment h.

Substituting, the mean price is

$$E\tilde{p}(h) = \frac{\left(r + m(h)/\mu^S(h)\right) E\left[\frac{\tilde{v}}{r+\eta}\right] - cE\left[\frac{\eta}{r+\eta}\right]}{r + m(h)/\mu^S(h)\, E\left[\frac{r}{r+\eta}\right]}$$

In general, the price depends on the moments of the distribution of preferences in the clientele of segment h. A tractable special case obtains when the separation rates η depend only on the segment. The price formula is then

$$E\tilde{p}(h) = \frac{E[\tilde{v}]}{r} - \frac{\eta(h)}{r + m(h)/\mu^S(h) + \eta(h)} \frac{E[\tilde{v}] + c}{r}$$

The first term is the present value of utility flow if the house remains always the favorite. This price obtains if houses never fall out of favor ($\eta = 0$) or if the market is frictionless in the sense that matching is infinitely fast ($m/\mu^S \to \infty$). More generally, the price incorporates a liquidity premium – the second term – that reflects foregone utility flow during search as well as the cost of search itself. The liquidity premium is larger if houses fall out of favor more quickly (η higher) and if it is more difficult to sell a house in the sense that time on market $\mu^S(h)/m(h)$ is longer.

Observables

We now define model counterparts for our observables. Segments are identified with zipcodes. We obtain the set of agent search profiles $\tilde{H}(\theta)$ from our email alert data. As described in Section 2.1.2, we construct for each agent the set of zipcodes he covers with his alerts. Since the alerts are a sample of buyers, we map the relative frequency of a search profile $\tilde{H}(\theta)$ to the share of type θ buyers in the total buyer pool:

$$\tilde{\beta}(\theta) := \frac{\tilde{\mu}^B(\theta)}{\bar{\mu}^\Theta - 1}.$$

Time on market in segment h is obtained by summing up (3) over the clientele[3] of segment h:

$$\frac{1}{T(h)} = \frac{m(h)}{\mu^S(h)} = \sum_{\theta \in \tilde{\Theta}(h)} f\left(\tilde{\mu}^S(\theta)\right) \frac{\tilde{\mu}^B(\theta)}{\tilde{\mu}^S(\theta)} \tag{5}$$

[3]Summing over the clientele gives on the left hand side of equation (3): $\sum_{\theta \in \tilde{\Theta}(h)} \left(\mu^S(h;\theta)/\mu^S(h)\right) m(h) = m(h)$. On the right hand side of equation (3), we get $\mu^S(h) \sum_{\theta \in \tilde{\Theta}(h)} f_\theta\left(\tilde{\mu}^S(\theta)\right) \tilde{\mu}^B(\theta)/\tilde{\mu}^S(\theta)$, which divided by $\mu^S(h)$ gives the RHS of equation (5).

Time on market is longer if buyers find their favorite house more quickly, and if there are more buyers relative to inventory scanned by those buyers. How time on market compares across segments depends on the respective clienteles. With heterogenous clienteles time on market will typically be different across segments. However, it will tend to be more similar if there is a strong common clientele.

Turnover in market h is given by

$$V(h) = \frac{m(h)}{\mu^H(h)} \tag{6}$$

In a steady state, the share of inventory in the total housing stock $\mu^S(h)/\mu^H(h)$ must equal the product of time on market and turnover $T(h)V(h)$. Below, we show that this condition is satisfied in our data to a good approximation even if we average over short periods such as one year.

Two special cases: segmentation vs integration

To illustrate the role of search profiles, we compare two special cases: full segmentation and full integration. First suppose there are exactly as many types as segments and each type scans exactly one segment. We use the label $\theta = h$ for the type scanning segment h and otherwise drop θ arguments. Equilibrium inventories are determined segment by segment by

$$\eta(h)\left(\mu^H(h) - \mu^S(h)\right) = f_h\left(\mu^S(h)\right)\left(\mu^\Theta(h) - \mu^H(h)\right) \tag{7}$$

The left hand side[4] is the rate at which houses come on the market. It is strictly decreasing in inventory: higher inventory means that fewer agents are living in their favorite house and thus fewer houses can come on the market each instant. The right hand side describes the rate at which houses are sold. It is strictly increasing in inventory: higher inventory means that buyers locate their favorite house more quickly. It follows that there is a unique equilibrium level of inventory μ^S – if inventory is too low, then too many houses come on the market whereas if inventory is too high, then too many houses are sold.

Consider how segments differ in the cross section under full segmentation. If we divide by $\mu^H(h)$, equation (7) shows the steady state turnover rate. Turnover is higher in markets where houses fall out of favor more quickly (higher η) or in which there are more buyers relative to houses (higher μ^Θ). Time on market is given by

$$T(h) = \frac{\mu^S(h)}{f(\mu^S(h))} \frac{1}{\mu^\Theta(h)/\mu^H(h) - 1}.$$

[4]Starting from equation (2), $\eta(h)\mu^F(h) = \eta(h)\left(\mu^H(h) - \mu^S(h)\right)$ using the adding up constraint for the number of houses. The right-hand side starts from equation (3), which boils down to $f_h\left(\mu^S(h)\right)\mu^B(h) = f_h\left(\mu^S(h)\right)\left(\mu^\Theta(h) - \mu^H(h)\right)$ because the number of agents also adds up.

An increase in the number of buyers relative to houses unambiguously reduces time on market. In contrast, the effect of an increase in the separation rate η on time on market depends on the shape of the function f. On the one hand, more separations mean that more houses come on the market, which tends to increase time on market. However, an increase in inventory implies that buyers find a favorite house more quickly, which tends to reduce time on market. The latter effect prevails if and only if $\mu/f(\mu)$ is decreasing as a function of μ, a case we refer to as a "strong demand effect".

Now consider integration, that is, all segments are exclusively scanned by a single type. The equilibrium inventories again equate the flow of house coming on the market to the volume of sales:

$$\eta(h)\left(\mu^H(h) - \mu^S(h)\right) = \frac{\mu^S(h)}{\sum_{h\in H}\mu^S(h)} f_\theta\left(\sum_{h\in H}\mu^S(h)\right)\left(\bar{\mu}^\Theta - 1\right)$$

Volume (the right hand side) depends on the segment only via inventory $\mu^S(h)$. It follows that with full integration time on market $T(h) = \mu^S(h)/m(h)$ is equated across all segments. Indeed, any house in inventory is equally likely to be a favorite house, so buyers flow into markets in proportion to inventory. As a result, if more houses come on the market in one segment (perhaps because η is larger there) then that segment will also attract more buyers in steady state.

Integration need not lead to more volume. For example, if the function f_θ is linear, then we obtain an aggregation result: the integrated market works as if every segment has a single type with house finding rate f_θ. More generally, the effect of integration on volume depends on the shape of f_θ. With a "strong demand effect" as discussed above, then integration implies faster matching. Intuitively, broader search implies a larger scanned inventory which allows agents to find favorite houses at a faster rate.

4 Quantitative Analysis

First, in steady state, the model predicts that the share of inventory in the total housing stock $\mu^S(h)/\mu^H(h)$ must equal the product of time on market and turnover $T(h)V(h)$. To check this implication of the model, Figure 5 plots the time series averages of both series. Each dot is a zip code and measures its average inventory share $\mu^S(h)/\mu^H(h)$ on the vertical axis (in percent), together with the product of its average time on market and its turnover $T(h)V(h)$ on the horizontal axis. The color of the dot indicates the average house price in the zip code. The colors range from blue (cheap) to pink (expensive.) The plot indicates that inventory and the product of time on market and turnover are positively related. The green line is the fitted value from a linear regression.

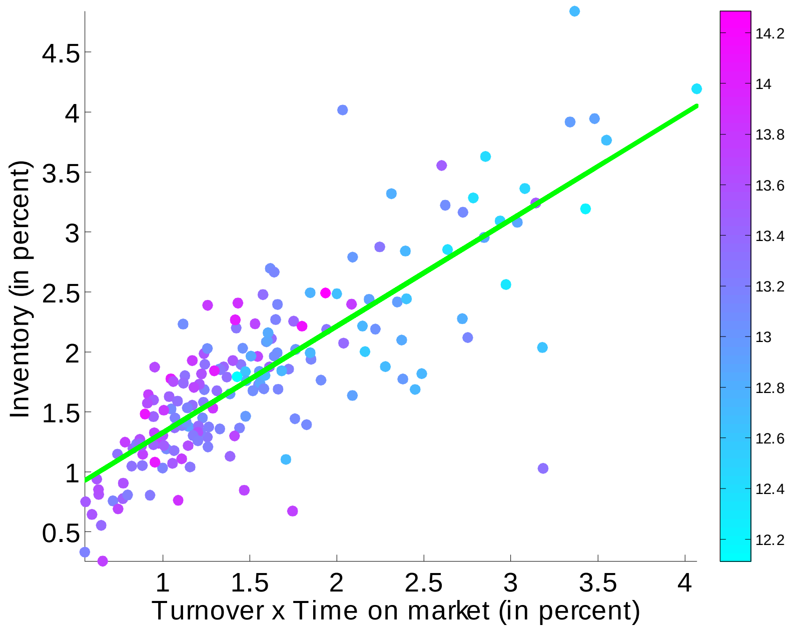

Figure 5: Time series averages of inventory against the product of turnover and time on the market. Inventory is measured as a fraction of the total housing stock. Each dot is a zipcode; the colors indicate average (log) house prices in the zip code ranging from cheap (blue) to expensive (pink). The green line is the fitted value of a linear regression.

The estimated intercept in this regression is 0.0037 and the estimated slope coefficient 0.966 with $R^2 = 59\%$.

Next, we select parameters by matching the steady state distributions of the model to moments of the cross section of San Francisco Bay Area zipcodes. For each of our 183 zipcodes, we form mean turnover and mean time on market over the period 2008-2011. We also use the empirical distribution of search alerts. There are 6,308 distinct search profiles, which delivers another 6,307 moment conditions.

We need three sets of parameters. The first is the cross section of separation rates η. We assume that η depends only on the segment and not on type θ. We thus need to find 183 parameters $\eta(h)$. We can use (2), (5) and (6) to derive

$$\eta(h) = \left(\frac{1}{V(h)} - T(h)\right)^{-1}.$$

We can thus identify $\eta(h)$ independently of the search profiles $\tilde{H}(\theta)$. The first step of our matching exercise fits $\eta(h)$ this way. Since in our data monthly turnover $V(h)$ is typically less than 1%, and time on the market is less than 6 months, the resulting

$\eta(h)$s closely track turnover. The backed out $\eta(h)$s are mapped in the left hand panel of Figure 6, with η increasing along the colormap from blue to pink. The main finding here is that more expensive neighborhoods that have low turnover are more stable, that is, they have lower $\eta(h)$.

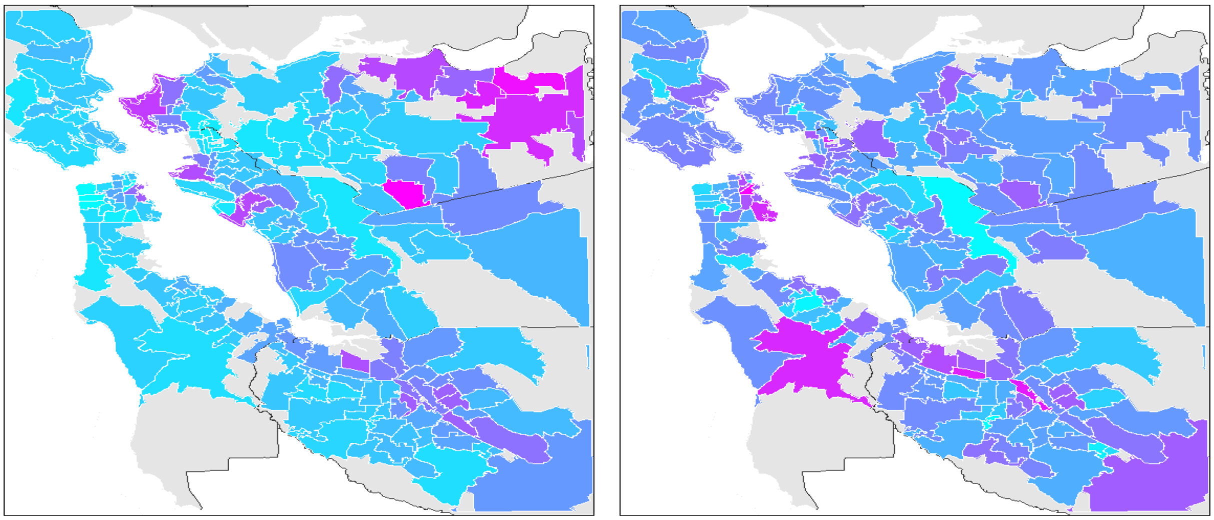

Figure 6: Left panel: Calibrated $\eta(h)$ increasing from blue to pink. Right panel: Number of search profiles covering a segment, weighted by the inverse amount of housing stock scanned by the profile, increasing from blue to pink.

The second set of parameters comes from the house finding rate. We use the functional form

$$f_\theta\left(\tilde{\mu}^S(\theta)\right) = \bar{f}\left(\frac{\tilde{\mu}^S(\theta)}{\sum_{h\in\tilde{H}(\theta)}\mu^H(h)}\right)^\delta,$$

where $\bar{f}$ and δ are strictly positive scalars. Inventory $\tilde{\mu}^S(\theta)$ is thus scaled with the housing stock of the area scanned by type θ. In other words, an agent finds a house more quickly if a larger share of the housing stock in his desired area is available for sale.

The third set of parameters is the distribution of agent types μ^Θ. We find μ^Θ, $\bar{f}$ and δ as follows. Given the parameters η and δ, we can select μ^Θ and $\bar{f}$ in order to (i) match exactly the buyer shares $\tilde{\beta}(\theta)$ from the distribution of search alerts, (ii) match the cross sectional average of time on market and (iii) set the cross sectional average of times it takes a buyer to find a house, that is, $\tilde{\mu}^B(\theta)/f_\theta\left(\tilde{\mu}^S(\theta)\right)$ equal to the average time it takes a seller to sell. This step determines 183 + 6307 parameters. We then select δ to minimize the sum of squared prediction errors for the inventory share. We find a value of $\delta = 2.93$, which implies a strong demand effect.

The right hand panel of Figure 6 maps a summary statistic of how desirable a segment is. In particular, the colors represent the number of agents scanning the zipcode, weighted by the inverse of the housing stock they scan. If the markets were fully segmented or fully integrated, then it would reduce to the ratio of the number of agents to the number of houses. The finding is that the downtown areas of San Francisco and San Jose are the most sought after areas. It is the counterpart of the summary statistics of search alerts in the bottom right panel of Figure 3. The difference is that the former concept is computed based on the unobservable number of agents $\mu^{\Theta}(\theta)$, whereas the latter is computed based on the observed number of search queries $\tilde{\beta}(\theta)$.

The above procedure leaves us with a large number of overidentifying restrictions. To illustrate the fit of the model, Figure 7 reports model predictions and data value for the cross sectional means of time on market and the inventory share. Since the calibration procedure assigns essentially no weight to time on market, the model does not fit particularly well in this dimension. However, the results are sensitive to the functional form of f: reducing δ increases the correlation between predicted and actual mean time on market, although it overstates the slope of the relationship. This suggests that one may be able to find a parsimonious functional form for f that improves the fit of the model further. The fit for the inventory share is much better: the model clearly captures the fact that more expensive (pinker) segments have lower turnover.

Figure 8 shows how the steady state equilibrium changes if the supply of houses in a zipcode is increased by one percent of the existing housing stock in one zipcode only. Formally, we recompute the steady state using the same parameters as above, but we increase $\mu^{H}(h)$ by one percent in one particular segment h. All panels are maps of only the tip of the San Francisco peninsula. The left two panels assume that the hypothetical change in the housing stock occurs in zipcode 94102 in downtown San Francisco, marked by a pin. The leftmost panel shows the change in time on market, and the second panel shows the change in inventory. Changes increase from blue to pink.

The result is that a change in 94102 has spillover effects: it increases time on market and inventory all over San Francisco as well as in the suburbs. This is because a large share of searchers scan all these segments jointly. In contrast, the two panels on the right show panels assume that the housing stock increases in the suburb of Daly City (zipcode 94015), again marked with a pin. Here the spillover effects are confined to neighboring zipcodes to the south and west. Remarkably, essentially nothing happens to the north, in the city of San Francisco itself. These results show that search patterns introduce asymmetries in the transmission of shocks.

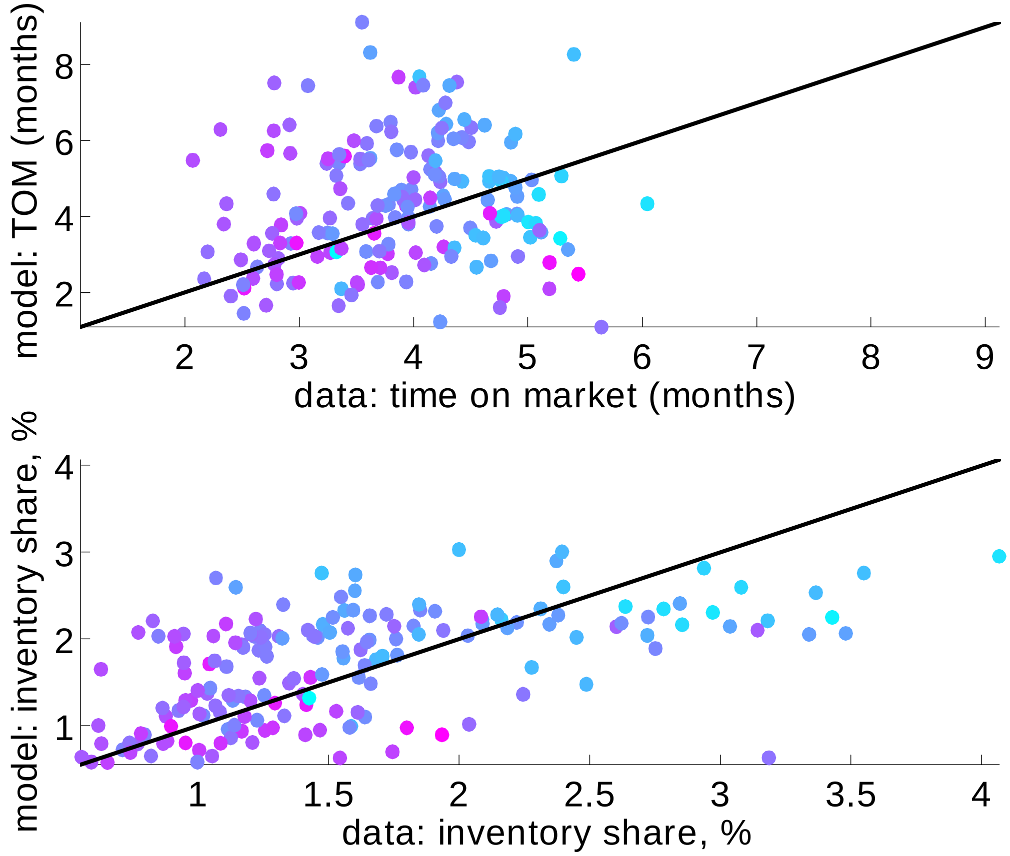

Figure 7: Top panel: mean time on the market (in months) by zip code plotted against model prediction. Bottom panel: mean inventory share by zip code plotted against model prediction.

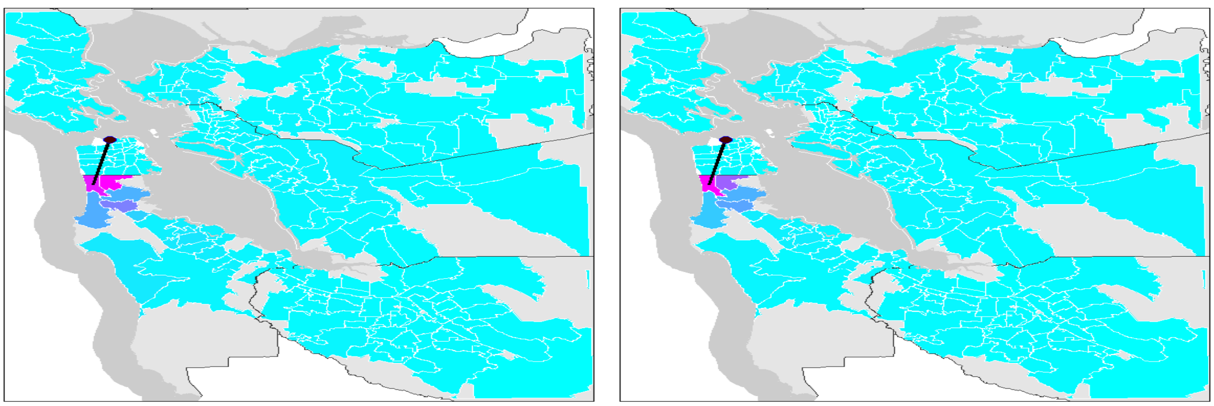

Figure 8: Reponses to one-percent increases in the housing supply. Panels 1 and 2: responses in time on market and inventory, respectively. The responses are to a one-percent increase in the housing supply in downtown San Francisco (zipcode 94102, marked by a pin). Panels 3 and 4: responses in time on market and inventory, respectively. The responses are to a one-percent increase in the housing supply in the suburb of Daly City (zipcode 94015).

References

Burnside, Craig, Martin Eichenbaum, and Sergio Rebelo 2011. "Understanding Booms and Busts in Housing Markets." *NBER Working Paper* 16734.

Caplin, Andrew and John Leahy 2008. "Trading Frictions and House Price Dynamics." NBER Working Paper 214605.

Favilukis, Jack, Sydney Ludvigson and Stijn van Nieuwerburgh 2012. "Macroeconomic Implications of Housing Wealth, Housing Finance, and Limited Risk-Sharing in General Equilibrium." Working Paper, NYU.

Glover, Andrew, Jonathan Heathcote, Dirk Krueger and Jose-Victor Rios Rull 2012. "Intergenerational Redistribution and the Great Recession." Working Paper, UPenn.

Krainer, John 2001. "A Theory of Liquidity in Residential Real Estate Markets." *Journal of Urban Economics* 49, pp. 32-53.

Landvoigt, Tim, Monika Piazzesi and Martin Schneider 2012. "The Housing Market(s) of San Diego." *NBER Working Paper* 17723.

Genovese, David and Lu Han 2012. "Search and Matching in the Housing Market." *Journal of Urban Economics* 72, pp. 31-45.

Mian, Atif and Amir Sufi 2010. "House Prices, Home Equity-Based Borrowing, and the U.S. Household Leverage Crisis." Forthcoming, *American Economic Review.*

Novy-Marx, Robert 2009. "Hot and Cold Markets." *Real Estate Economics* 37, pp. 1-22.

Piazzesi, Monika and Martin Schneider 2009. "Momentum traders in the housing market: survey evidence and a search model." *American Economic Review*, 99, pp. 406-411.

Piazzesi, Monika and Martin Schneider 2012. "Inflation and the Price of Real Assets." Working Paper, Stanford.

Poterba, Jim 1991. "House Price Dynamics: The Role of Tax Policy and Demography." *Brookings Papers on Economic Activity* 2, 143-202.

Stevens, Margaret 2007. "New Microfoundations for the Aggregate Matching Function." *International Economic Review* 48, pp. 847-868.

Van Nieuwerburgh, Stijn and Pierre-Olivier Weill 2010. "Why Has House Price Dispersion Gone up?" *Review of Economic Studies* 77, pp. 1567-1606.

Wheaton, William C. 1990. "Vacancy, Search and Prices in a Housing Market Matching Model." *Journal of Political Economy* 98, pp. 1270-1292.

A Data construction

The Bay Area has roughly 240 zip codes. Among these zip codes, some have low or no housing stock. For example, zip codes 95042 and 94802 have zero population.[5] We focus on ZIP codes that have a minimum of at least 30 transactions in each year, and at least 50 transactions per year on average over our sample period. Table A.1 lists all zip codes in our data sample.

To define price by zip code, volume, and time on market using arms-length transaction price in each zip code. Armslength transactions are defined as transactions in which both buyer and seller act in their best economic interest. This ensures that transaction prices reflect the market value of the property. We include all deeds that are one of the following: "Grant Deed," "Condominium Deed," "Individual Deed," "Warranty Deed," "Joint Tenancy Deed," "Special Warranty Deed," "Limited Warranty Deed" and "Corporation Deed." This excludes intra-family transfers and foreclosures. We drop all observations that are not a Main Deed or only transfer partial interest in a property.

In defining inventory, one empirical challenge is that we do not observe when listings that do not result in a sale get removed from the market. We remove all listings from the inventory if they haven't sold within 150 days after the initial listings (as a reference point, note that the average time on market for houses that do eventually get sold is about 100 days). Of course, if a house sells that was listed for more than 150 days, we record that as a sale.

Another empirical challenges is that many property listings of condos do not include information on the actual unit number, just the street address, which makes it hard to correctly match listings to sales when there are more than one condo listed at an address (e.g. in the same condo building) at the same time. We hence exclude condos from our measure of inventory and only calculate the inventory for single family residences. We then scale up this measure by the share of single family residences in the zip code (as observed in the tax assessment records) to get a measure of the total inventory.

Similarly, time on market is only calculated for single-family residences, and not for condos. This makes the implicit assumption that the patterns for single-family residences are somewhat representative of condos. An additional complication in the calculation of inventory levels is that (1) not all properties will be listed on online search enginges such as Trulia.com and (2) that Trulia expanded its coverage of listings over that period. This means that our inventory measure will only be a subset of the total inventory in

[5]For more information on these zip codes, see http://www.zip-codes.com/zip-code/95042/zip-code-95042.asp, http://www.zip-codes.com/zip-code/94802/zip-code-94802.asp

the market. To deal with this issue, we count for each month the number of sales we observed in the deeds data for which we did not observe a prior listing. Under the assumption that the listings we observe are somewhat representative of the universe of listings, this gives us a measure of the fraction of inventory we observe. We then scale our observed inventory by the fraction of sales without observed prior listing in that month.

Table A.1: Bay Area Zip Codes In Our Sample

94002	94089	94303	94538	94577	94618	94941	95118
94010	94102	94306	94539	94578	94619	94945	95119
94014	94103	94401	94541	94579	94621	94947	95120
94015	94107	94402	94542	94580	94702	94949	95121
94019	94108	94403	94544	94582	94703	94960	95122
94022	94109	94404	94545	94583	94704	94965	95123
94024	94110	94501	94546	94587	94705	95008	95124
94025	94112	94502	94547	94588	94706	95014	95125
94030	94114	94505	94549	94595	94707	95020	95126
94040	94115	94506	94550	94596	94708	95023	95127
94041	94116	94507	94551	94597	94709	95030	95128
94043	94117	94509	94552	94598	94801	95032	95129
94044	94118	94513	94553	94601	94803	95035	95130
94061	94121	94517	94555	94602	94804	95037	95131
94062	94122	94518	94556	94603	94805	95050	95132
94063	94123	94519	94560	94605	94806	95051	95133
94065	94124	94520	94561	94606	94901	95054	95134
94066	94127	94521	94563	94607	94903	95070	95135
94070	94131	94523	94564	94608	94904	95110	95136
94080	94132	94526	94565	94609	94920	95111	95138
94085	94133	94530	94566	94610	94925	95112	95139
94086	94134	94531	94568	94611	94930	95116	95148
94087	94301	94536	94572	94612	94939	95117	