Segmented Housing Search*

Monika Piazzesi
Stanford & NBER

Martin Schneider
Stanford & NBER

Johannes Stroebel
New York University

May 2014

Abstract

This paper considers housing search, trading and valuation in interconnected housing market segments with heterogeneous buyers. We use a novel data set on online housing search to measure buyer search ranges for the San Francisco Bay Area. We document the cross section of turnover, inventory and search activity in a large number of market segments. We find substantial variation within narrow geographic areas that is critical for understanding market activity: for example, search activity and inventory covary positively within cities and zipcodes, but negatively across those units. A quantitative search model of the housing market shows how market activity at different levels of aggregation depends on the interaction of heterogeneous clienteles. It also implies liquidity discounts in house prices are large and vary widely across market segments.

1 Introduction

Home buyers typically look for properties in a search range that depends on their geographic preferences, budget, or family size. For example, they might focus on houses in a certain price range that are also in reasonable commuting distance from their workplace. A family with children might in addition require that the house be located in a good school district. An individual property that comes on the market is then considered by a clientele of potential buyers whose search ranges contain that property. The interaction of clienteles determines how turnover, inventory and prices differ across segments of the housing market.

*Addresses: piazzesi@stanford.edu, schneidr@stanford.edu, johannes.stroebel@nyu.edu. We thank Ed Glaeser, Giuseppe Moscarini, Robert Shimer, and Silvana Tenreyro for helpful comments. We are grateful to Trulia for providing data. We also thank seminar participants at the AEA meetings, Arizona State, Berkeley, the NBER Summer Institute, Conference on Heterogeneous-Agent Models in Macroeconomics in Mannheim, Macroeconomic Dynamics Workshop in London, Northwestern-Tsinghua Conference, Philadelpha Macro Conference, Society of Economic Dynamics, SITE, Stanford, Yale, UCLA, USC, NYU Stern, Chicago Booth and the Minneapolis Fed. This research is supported by a grant from the National Science Foundation.

Existing studies of housing search typically assume that clienteles are homogeneous. In particular, two common assumptions are that the market under consideration is either fully integrated – that is, the clientele for each house consists of all potential buyers – or that it is perfectly segmented, that is, it can be partitioned into submarkets, each with its own buyer type who considers all houses in the submarket. However, when studying, say, a metro area, homogeneity of clienteles is not a priori obvious. For example, a neighborhood with good schools might see competition between families with children – who search narrowly in that neighborhood – and other potential buyers who search more broadly. More generally, the distribution of workplaces and commuting costs in the population is likely to generate clienteles whose search ranges only partially overlap.

This paper considers housing search, trading and valuation in interconnected housing market segments with heterogeneous clienteles. We introduce a novel dataset on housing search behavior in the San Francisco Bay Area to document stylized facts on search ranges. When we divide the Bay Area into market segments based on observed search ranges, we find substantial heterogeneity not only for market outcomes across segments but also for clienteles both within and across segments. We then use a search model with multiple segments to relate market outcomes to the distribution of preferences and the matching technology. We show that the interaction of heterogeneous clienteles is a quantitatively important force in the housing market.

We infer search ranges from online search via the real estate website trulia.com. Searchers on trulia.com can set an email alert that triggers an email whenever a house with their desired characteristics comes on the market. We find that housing search occurs predominantly along three dimensions: geography, price and, to a lesser extent, house size as captured by the number of bathrooms. Most searchers look for houses in contiguous areas, but differ in geographic breadth. In cheaper urban areas, there are fewer searchers per house, and those who do search broadly for low prices. In contrast, clienteles in more expensive and more suburban areas tend to be larger but also more selective.

To analyze market activity, we divide the San Francisco Bay Area into 576 distinct market segments along the dimensions suggested by the search alerts. We then measure the cross section of turnover and inventory at the segment level by matching search alert data to deeds and assessment records as well as feeds of listings for sale. We find that about half of the variation in market activity occurs within zip codes, our finest geographic unit. Inventory and turnover comove strongly, both at the segment level and when we aggregate to the zipcode or city level. In particular, in cheaper areas or segments, houses turn over faster, but there is also more inventory for sale.

To relate market and search activity, we express search ranges as subsets of the set of all segments, resulting in about 9000 distinct ranges. We measure search activity at the segment level in terms of searchers per house. We find that the relationship between inventory and search activity depends critically on the level of aggregation. Across cities, inventory and search activity are inversely related – in other words, the "Beveridge curve" slopes *down across cities*. For example, in expensive cities like San Francisco, many people search scarce inventory, while in cheaper cities like San Jose, plenty of inventory is considered by few searchers. In contrast, the Beveridge curve slopes *up within most cities*: for example, cheaper segments within San Francisco have higher inventory and are considered by more searchers.

Our model exercise builds on a version of the Diamond-Mortensen-Pissarides random matching model with fixed numbers of both houses and agents. Moving shocks induce agents to sell their current house (at a cost) and search for another house. What is new in the model is the presence of multiple market segments as well as heterogeneous agent types identified by search ranges – subsets of the set of all segments as in our data. While matching is random, agents are more likely to match in those segments within their search range where inventory is higher. Prices reflect the present value of housing services less a discount due to search and transaction costs.

The equilibrium of the model relates the cross sectional distribution of turnover, inventory, price and search activity to the distribution of preferences, moving shocks and the matching technology. The distribution of preferences – including search ranges – allows the model to capture the rich clientele patterns we measure in the data. The key theoretical effect added by heterogeneous clienteles is that broad searchers flow to high inventory segments and compete with narrow searchers there. It is stronger in more integrated areas, for example within cities that have a larger share of broad searchers. It implies that the nature of clientele patterns then matters both for how market activity responds to changes in the environment and for what we can infer about parameters from the cross section of market activity.

If there is perfect segmentation, then identification of the three forces that drive segment heterogeneity is relatively simple. In more stable segments – where moving shocks arrive less frequently – turnover and inventory are both lower. In more liquid segments – where matching is faster holding fixed the buyer and seller pools – turnover is also higher but inventory is lower. The same is true in more popular segments that have more potential buyers per house. However, more popular segments also see more search activity.

Our quantitative exercise suggests that patterns at the city level are driven by differences in popularity and stability. More expensive cities like San Francisco are both more stable and more popular than cheaper cities like San Jose. The former explains why turnover and inventory are both lower in San Francisco. The latter explains why search activity is higher there and helps generate a downward sloping Beveridge curve in the cross section of cities.

Within cities, the Beveridge curve is affected not only by the correlation of exogenous forces, but also by the endogenous interaction of heterogeneous clienteles. Indeed, consider two segments that are equally popular and differ only in stability. Broad searchers who scan both segments will tend to flow to the less stable segment where inventory is higher. As a result, narrow searchers in the unstable segment find it harder to find a house and must search more. Within partially integrated areas such as cities, differences in stability alone thus generate an upward sloping Beveridge curve. We show that the endogenous response of broad searchers is quantitatively important.

We also use our estimated parameters to infer liquidity discounts for houses in various segments. These discounts are quantitatively large, between 10 and 40 percent of the frictionless house value (defined as the present discounted value of future housing services by the house.) The liquidity discounts are large in segments that are less stable, where houses turn over more often. They are also large in illiquid segments, where houses take a long time to sell for whatever reason. High turnover and high time on market increase the value of the trading frictions that the current and future buyers face, which amount to the liquidity discount.

Finally, we illustrate the role of search patterns for the transmission of shocks with 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 segment with many searchers who search broadly is transmitted widely across the city. In contrast, shocks to a suburban segment 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 perfectly 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 (2011), Novy-Marx (2009), Ngai and Tenreyro (2009), Piazzesi and Schneider (2009), Burnside, Eichenbaum and Rebelo (2011), and Han and Strange (2013).[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, Weil and Shiller (1991), Bayer, Ferreira and McMillan (2007), Mian and Sufi (2009)) document the importance of determinants such as credit constraints, demographics, or school quality in different housing markets.

More related to our paper, Genesove 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. A number of papers have considered how to divide housing markets into segments; Islam and Asami (2009) survey the literature. Most of these paper discuss how to split housing markets into mutually exclusive segments based on similarity along a number of characteristics. Goodman and Thibodeau

[1]Recent models of a single housing market with frictions *other than* search include Piazzesi and Schneider (2012), Floetotto and Stroebel (2012), Favilukis, Ludvigson and Van Nieuwerburgh (2010) and Glover, Heathcote, Krueger and Ríos-Rull (2011). Empirical analyses of frictions in the real estate market include Glaeser and Gyourko (2003), Levitt and Syverson (2008), Garmaise and Moskowitz (2004) and Stroebel (2012).

(1998) define housing markets as geographical areas based on a consistent price per unit of housing services. Leishman (2001) argues that housing markets can be segmented both spatially and structurally.

Perhaps the closest paper to ours is by Manning and Petrongolo (2011) who estimate a search and matching model for local labor markets. While the study does not have data on where unemployed workers look for jobs (as we have for home buyers), it uses their home addresses, the addresses of job vacancies in their local area and puts more structure on how workers compare jobs with different commuting times (e.g., workers are indifferent about commuting within some radius.)

2 Dimensions of housing search

In this section we document search behavior in the San Francisco Bay Area using email alerts set on trulia.com. We first describe the data and then provide summary statistics on the major dimensions of housing search. The results here provide answers to three broad questions. First, can search ranges inferred from trulia alerts can be plausibly interpreted as reflecting the considerations of a typical home buyer? Second, how much heterogeneity do we observe in search behavior? Finally, is there a simple way to describe search ranges as subsets of a space of characteristics (including geography and quality)?

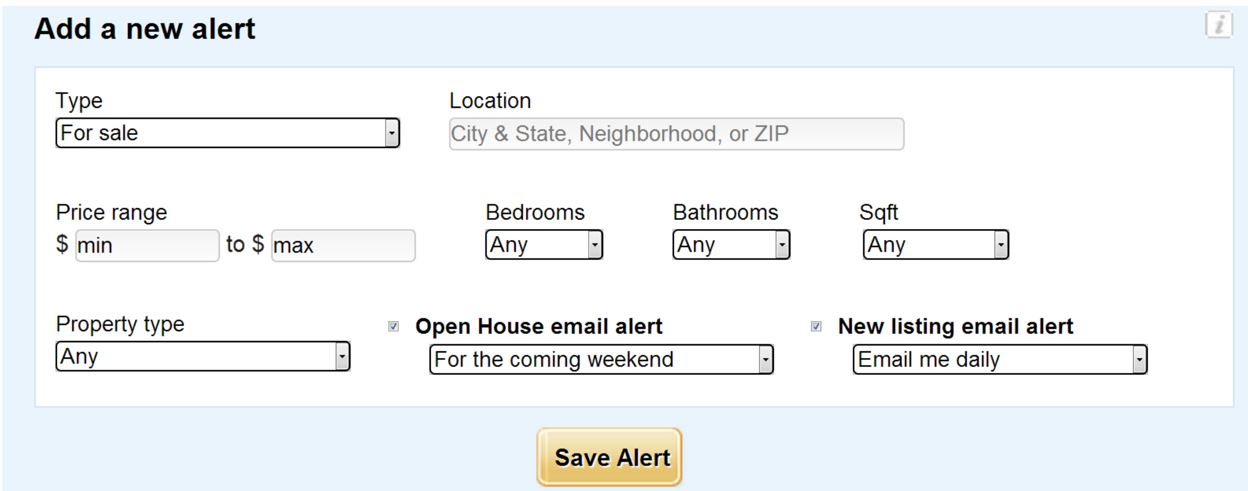
The San Francisco Bay Area is a major urban agglomeration in Northern California that includes the cities of San Francisco, San Jose and Oakland. Our analysis combines data on two Metropolitan Statistical Areas (MSAs) bordering San Francisco Bay. The San Francisco-Oakland-Hayward, CA Metropolitan Statistical Area comprises Alameda, Contra Costa, San Francisco, San Mateo, and Marin counties. The San Jose-Sunnyvale-Santa Clara, CA Metropolitan Statistical Area consists of Santa Clara and San Benito counties. As of the 2010 Census, these counties were home to about 6 million people who live in about 2.2 million housing units.

2.1 Email alerts

Visitors to trulia.com can set alerts that trigger regular emails when houses with certain characteristics come on the market. The web form for setting alerts is shown in Figure 1. Every alert must specify the fields in the first line: "Type" is either "For sale", "For rent", or "Recently sold". The field "Location" allows for a comma-delimited list of zipcodes, neighborhoods, or cities. Neighborhoods are geographic units commonly listed on realtor maps that are often, but not always, aligned with zipcodes. When users fill our the form, an autocomplete function suggests names of neighborhoods or cities.

The second row in the form provides the option of specifying house characteristics beyond geography. Price ranges may be set by providing a lower bound, an upper bound or both. For bedrooms and bathrooms, there is the option to set an integer lower bound from one to five ("1+", "2+" up to "5+"). Finally, for the house size in square feet, the lower bound is one of seventeen value between 250 square feet and 10,000 square feet. In the third row, "Property type" allows narrowing the search to "Single family home", "Condo" and several smaller categories. Finally, the remaining fields govern how emails are processed: for the "New listing email alerts" relevant for us, the options are "Email me daily" or "Email me weekly".

Figure 1: Setting email alerts on Trulia.com



Pooling alerts to obtain search ranges

We observe a random subset of 39,617 "For sale" search alerts between March 2006 and April 2012. Those alerts were set by 24,125 unique Trulia users, identified by the (scrambled) email address to which emails triggered by the alert are sent. Given the layout of the web form, it makes sense to set multiple alerts for example when searching according to different criteria in different cities. Almost 70 percent of searchers set only one alert, and more than 90 percent of individuals set 3 or fewer alerts.

We are interested in search ranges rather than individual alerts and thus pool alerts by searcher. In particular, we take the geographic area to be the union of all areas covered in individual alerts. For the purposes of this section, we also take the price range to be the maximal range considered across alerts.

Representativeness

The interpretation of our findings depends to some extent on whether searchers who use trulia are a special subset of the overall searcher pool. In particular, their use of the internet in home search might signal that they are younger and richer than the average home buyer. While we do not have direct demographic information on the searchers in our sample, recent surveys conducted by the National Association of Realtors provide some useful background information on modern home search.

The internet has now become the most important tool in the home buying process, with over 90 percent of homebuyers using the internet in their homesearch process (National Association of Realtors, 2013). In particular, for 35 percent of home buyers, looking online is the first step taken in the home purchase process. The fraction of people who deemed real estate websites "very important" as a source of information was 76 percent, larger than the fraction 68 percent who found real estate agents "very important". Moreover, use of the internet is not concentrated among younger buyers: 86 percent of home buyers between the ages of 45 and 65 go online to search for a home. The median age of homebuyers using the internet is 42, the median income is $83,700 (National Association of Realtors, 2011). This is only slightly younger than the median of all home buyers (which is 45) and slightly wealthier (the median income of all home buyers was $80,900).

In addition to showing that online real estate search is almost universal, this suggests that we can learn from online real estate search about overall search behavior. Moreover, trulia.com, with approximately 24 million unique monthly visitors (71 percent of whom report to plan to purchase in the next 6 months), has similar demographics to those of the overall online home search audience (Trulia, 2013).

Major dimensions of search

Table 1 shows that roughly a third of the queries does not specify any fields in addition to geography. The other fields that are specified regularly include listing price and the number of bathrooms. Just under a third of queries specifies both price and the number of bathrooms, while another third specifies just a price range. The remaining 5 percent of queries specifies just a bathroom criterion in addition to the geographic restriction. Other fields in Figure 1 are used much less. For example, only 1.3 percent of queries specify square footage while 2.7 percent of queries specify the number of bedrooms. While the latter two fields are alternative measures of size, the minimum number of bathrooms is a commonly used filter to place restrictions on the size of homes.

Table 1: Distribution of alert parameters

	Price not specified	Price specified	Total
Baths not specified	13,019	13,777	26,796
Baths specified	1,848	11,881	13,729
Total	14,867	25,658	40,525

Note: This table shows the distribution of alert parameters that Trulia users specify in addition to geography in our query sample.

2.2 Search by geography

Each alert defines the desired search geography by selecting one or more city, zip code or neighborhood. About 61 percent of alerts define the finest geographic dimension in terms of cities, 18 percent in terms of zip codes, and the remaining 21 percent of alerts specify the finest geographic dimension in terms of neighborhoods. Some queries include geographies in terms of cities, zip codes, and neighborhoods in the same query.

Distance

To summarize how search ranges reflect geographic considerations, we consider measures of size. 28 percent of searchers consider only a single zipcode. For the remaining searchers, we measure the maximal distance between zip codes contained in their search ranges. If a range is not directly defined in terms of zipcode, this requires converting city or neighborhood information to zipcode information, as described in the appendix. We then focus on distances between population-weighted zipcode centroids. Population weighting is useful since we are interested in distance between agglomerations within zipcodes that might reflect searchers' commutes.[2]

Table 2 reports maximum and mean distance between zipcode centroids for the ranges of searchers who consider more than one zipcode. We compare three measures of distance. Geographic distance is measured in miles and is direct "as the crow flies". Travel and transport time are calculated using Google Maps; they represent distance in minutes by car or public transport, respectively, as of 8am on Wednesday, March 20, 2013.[3] The distribution is over 17,488 searchers who select more than one zip code. We conclude from these numbers that the size of the typical search range is consistent with reasonable commuting times guiding geographic selections. Moreover, there is sizable heterogeneity in geographic breadth.

Table 2: Distribution of distances across search alert zip codes

	Population-Weighted Zip Code Centroids					
	Min	Bottom Decile	Median	Top Decile	Max	Mean
Max Geographic Distance	0.5	2.3	6.8	21.1	103.3	9.7
Mean Geographic Distance	0.5	1.8	3.2	8.9	74.0	4.7
Max Car Travel Time	4.0	9.5	20.5	38.5	143.5	22.8
Max Public Transport Time	3.8	8.9	13.1	19.7	132.5	14.0
Mean Public Transport Time	10.5	40.5	79.0	375.0	573.5	140.1

Note: This table shows the summary statistics across searchers who select more than one zip code (N = 17,488) of travel time and geographic distances between the centroids (population-weighted) of all zip codes selected by that query. Travel times are measured in minutes. Geographic distances are measured in miles.

[2]For each zipcode, we start from geographic centroids of all census blocks contained in the zip code, as provided by the Census Bureau, and then calculate their population-weighted arithmetic mean. For robustness, we also check results with geographic zipcode centroids and find similar results.

[3]A few zip code centroids are inaccessible by public transport as calculated by Google. Public transport distances to those zip code centroids were replaced by the 99th percentile of travel times between all zip code centroids for which this was computable. This captures that these zip codes are not well connected to the public transport network.

Contiguity & circularity

To guide our modeling of clientele heterogeneity, we ask whether there is a simple organizing principle for observed search ranges, namely that searchers consider contiguous areas, possibly centered around a focal point such as a place of work or a school. We say a search range is contiguous if it is possible to drive from between any two zipcode centroids in the range without ever leaving the range. Here we allow for travel across one of the six Bay Area bridges. Details are contained in Appendix A.1.

Table 3 shows summary statistics by the number of zipcodes defining the search range. The second column reports the share of searchers who select contiguous geographies. While overall only 18 percent of searchers have non-contiguous search ranges, they tend to be broad searchers who consider more than five distinct zipcodes and hence provide market integration across neighborhood and city boundaries. The third and fourth columns report the mean and max number of contiguous areas covered by a search range. Broad searchers often consider multiple distinct contiguous areas. Preference for certain cities plays a role here: the increase in the share of contiguous queries for the group with 21-30 zip codes selected can be explained by the prevalence of searches for "San Francisco" and "San Jose" in that category.

Table 3: Contiguity analysis – summary statistics

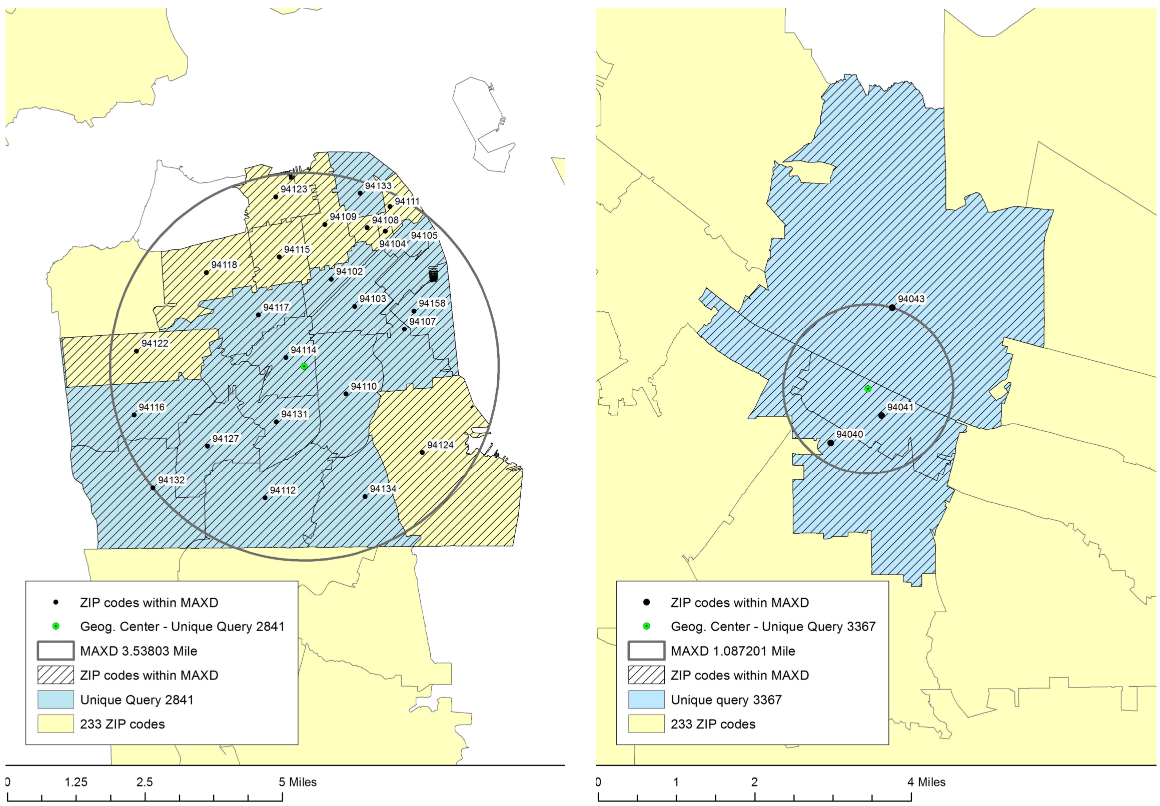
		Contiguous Segments		
Number of Zips Covered	Share contiguous	Mean	Max	Total Number
2	91%	1.09	2	2,927
3	83%	1.18	3	1,761
4	91%	1.10	3	2,248
5	67%	1.37	4	844
6-10	71%	1.38	5	2,612
11-20	74%	1.38	8	2,071
21-30	91%	1.13	10	4,213
30+	48%	1.94	9	798
Total	82%	1.24	10	17,474

Note: This table shows summary statistics for contiguity measures across queries that select different number of zip codes.

A stylized model of geographic search might view a range as a circle around a central point. We ask whether our search ranges can be suitably approximated by such a model. We thus compute, for each searcher, the geographic center of range: the average longitude and latitude of all (geographic) zipcode centroids selected by that searcher. We then determine the maximum distance to this center of any zip code centroid contained in the search range. On average, the maximum distance is 3.95 miles, while the 10th percentile is 1.31 miles and the 90th percentile is 12.78 miles. We then compute the number of zip code centroids (not necessarily contained in the search range) that are within maximum distance to the center.

We say a search range is circular if all zip codes within maximum distance to the center are also contained in the search range. Figure 2 illustrates this procedure.

Figure 2: Explanation of circularity test

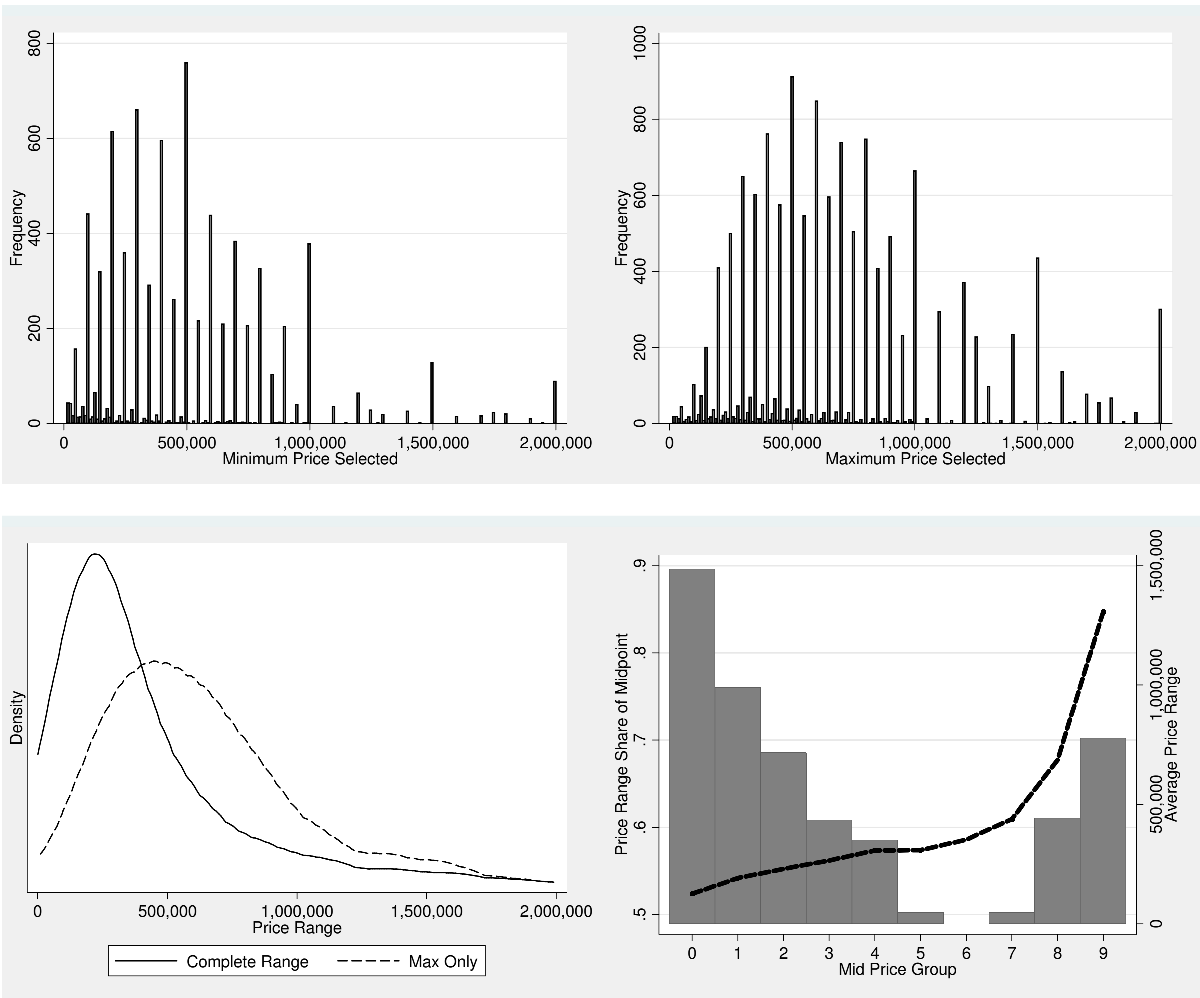


Note: The two panels are examples of the circularity analysis. All zip codes that are part of the search set are shown in blue. The geographic center of each search set is given in green. The circle is centered around this geographic center and has radius equal to the furthest distance of any zip code centroid in the search set to the search set center. All zip codes whose center lies within the circle (and who are thus at least as close as the furthest zip code center in the search set) are shaded.

Overall, 47 percent of all searchers have circular search ranges. This number is highest, at 83 percent, for ranges that only cover two zip codes, and declines for queries that cover more zip codes. In addition, for search sets with a larger maximum distance, the proportion of searches that cover all zip codes within this maximum distance from the center declines. On average, searchers cover 78 percent of all zip codes within maximum distance of their search range center. However, for non-contiguous ranges, the share of zip codes covered falls to 33 percent.

We conclude from these results that it is difficult to come up with a parsimonious description of the geographic selections defining search ranges. In particular, a modeling approach that describes ranges in terms of contiguous and/or circular subsets of the plane will fail to account for the behavior of broad searchers who integrate markets. This finding guides our approach in the next section, where we define a discrete grid of market segments, using zipcode as the geographic units. Geographic selection can then be represented as subsets of the set of all zipcodes, and it is straightforward to accommodate non-contiguous and non-circular search patterns.

Figure 3: Price cutoff analysis



Note: This figure shows a histogram in steps of $10,000 of the minimum and maximum listing price parameters selected by home searchers in their email alerts. The bottom left panel of this figure shows the distribution of price ranges across queries both for queries that only select a price upper bound as well as for those queries that select an upper bound and a lower bound. The bottom right panel shows statistics only for those alerts that select an upper and a lower bound. The line chart shows the average price range by for different groups of mid prices, the bar chart shows the average of the price range as a share of the mid price.

2.3 Search by price and size

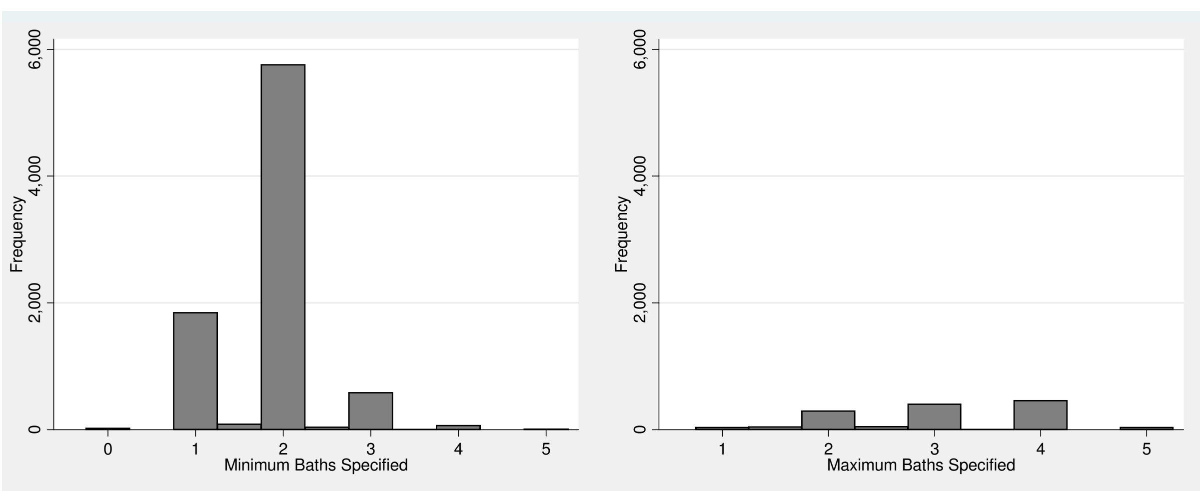
Out of the 61 percent of searchers who fill out the price field for their email alert, 50 percent specify both an upper and a lower bound, whereas 48 percent specify only an upper bound and only 2 percent select only a lower bound. The top panel of Figure 3 shows the distribution of minimum and maximum prices selected in the email alerts. Price range bounds are typically multiples of $\$50,000$, with particularly pronounced peaks at multiples of $\$100,000$.

There is significant heterogeneity in the breadth of the price ranges selected by different searchers. Among those who set both an upper and a lower bound, the 10th percentile selects a price range of $100,000, the median a price range of $300,000 and the 90th percentile a price range of $1.13 million. The bottom left panel of Figure 3 shows the distribution of price ranges both for those agents that select an upper and a lower bound, as well as for those agents that only select an upper bound.

The bottom right panel shows that searchers who consider more expensive houses specify wider price ranges. We bin the midprice of price ranges into 10 groups. The dashed line (with values measured along the right-hand vertical axis) shows that the price range considered increases monotonically with the midpoint of the price range. One simple hypothesis consistent with this is that searchers set price ranges choosing a fixed percentage range around a benchmark price. The bar chart (with percentages measured on the left hand vertical axis) shows that this is not the case: the percentage range is in fact U-shaped in price.

The third dimension that is regularly populated in the email alerts is a constraint on the number of bathrooms. Figure 4 shows the distribution of bathroom cutoffs selected for the Bay Area. 68% of all bathroom limits are set a value of 2, most of them as a lower bound. This setting primarily excludes 1 and 2 bedroom apartments and very small houses.

Figure 4: Bathrooms cutoffs selected



Note: This figure shows a histogram in steps of 0.5 of the minimum and maximum bathroom parameters selected by home searchers in their email alerts.

Tradeoffs between search dimensions

The three major search dimensions we have identified are not necessarily orthogonal. For example, one can search for houses in a particular price range by looking only at zip codes in that price range or only at homes of a certain size. Table 4 provides evidence on how different search dimensions interact. It shows in particular that searchers who are more specific on price or home size search more broadly geographically. For example, searchers who specify a price restriction cover an average of 10.3 zip codes with an average maximum distance between centroids of 7.9 miles, while other searchers cover only 7.3 zipcodes with an average maximum distance 1.06 miles. As we show below, discretizing the space of search characteristics can deal easily with searchers expressing their budget constraint or size preferences via geographic restrictions.

Table 4: Geography, price and bath parameter interaction

	No Price		Price		No Bath		Bath	
	Mean	N	Mean	N	Mean	N	Mean	N
# Zips Covered	7.3	8,725	10.3	15,400	8.8	15,716	10.0	8,409
Max Dist. (Mil)	7.9	5,375	10.6	12,113	8.9	10,899	11.1	6,589
Max Car (Min)	20.8	5,375	24.5	12,113	22.5	10,899	24.7	6,589
Max Public Trans. (Min)	75.8	5,375	92.4	12,113	82.1	10,899	95.9	6,589
Is Continguous	54%	8,725	62%	15,400	59%	15,716	60%	8,409

Note: This Table shows summary statistics across queries that cross-tabulate moments across different search parameters.

3 Market segments

In this section, we divide the San Francisco Bay Area housing market into a finite number of segments, motivated by search ranges inferred from email alerts. We then use summary measures of market and search activity at the segment level to establish stylized facts.

3.1 Data

To measure housing market activity, we combine three main datasets. We start from the universe of ownership-changing deeds in the Bay Area between 1994 and 2011. The property to which a deed relate is uniquely identified at the county level by the Assessor Parcel Number (APN). From the deeds data, we obtain the 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). We identify armslength transactions and foreclosure transactions using information on the type of deed and transaction price.

We also have the universe of tax assessment records in the Bay Area for the year 2009. Properties are again identified by 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.

Finally, we use a dataset of all property listings on trulia.com between October 2005 and December 2011. The variables we use here are listing date, listing price, and the listing address. The latter can be used to match listings data to deeds data. We can then construct a measure of time on market for each property that eventually sells.

Throughout we pool observations for the period 2008-2011. The goal of this paper is to understand the cross section of market activity. Pooling observations across years helps us achieve a finer description of cross sectional heterogeneity. In particular, there are sufficiently many observations to measure separately what happens in segments with low listing and housing turnover rates. To make prices comparable across years, we convert all prices in 2010 dollars using zipcode level repeat sales price indices.

3.2 Defining segments

We are looking for a partition of Bay Area houses into market segments. The finest partition that can be motivated by search data is obtained by joining all search ranges in our sample. Any division of houses into segments would then be motivated by the preferences of at least one searcher. Moreover, the preferences of any one searcher could be expressed exactly through a subset of the set of all segments. However, the problem with this approach is sample size: the number of houses per segment would be too small to accurately measure moments such as time on the market, inventory and buyer interest.

Our approach is to get as close as possible towards the finest partition, but subject to the constraint that segments must be sufficiently large in terms of volume and housing stock. This leads us to a set H of 576 segments as well as a set Θ of 9091 search ranges that can each be represented as a subset of H. These segments contain houses within a zipcode that are of similar quality (based on price) and size (based on bathrooms). We provide a detailed description of the algorithm in the appendix. In what follows we only sketch the main steps.

We start from our earlier result that people search mostly according to (i) quality, by specifying price ranges (ii) geography, where in particular zip code is the finest unit and (iii) size, by specifying the number of bathrooms, typically either "up to 2" or "more than 2" bathrooms. Facts (ii) and (iii) lead us to first divide the Bay Area by zipcode and then divide each zipcode into two size categories.

In order to use zipcode as the basic geographical unit, we need to deal with search ranges that specify geography at a unit that does not perfectly overlap with zipcodes. For search ranges that select listings at the city or neighborhood level, we assign all zip codes that are at least partially within the range of the city or neighborhood to be covered by the search range. This provides us, for each search range, with a list of zip codes that are covered by that search range. Using this method of expressing geographic selections in terms of zipcodes, the search ranges in our dataset cover a total of 191 unique Bay Area zipcodes.

To further accommodate search by quality, we further divide – zipcode by zipcode – each size group into four price groups. Here we start from a set of candidate price cutoffs: \$200K, \$300K, \$400K, \$500K, \$750K and \$1 million. We then select three cutoffs from these candidates that are most often close to price cutoffs appearing in our email alerts. The idea is that the resulting set of segments is close to the ideal partition implied by the ranges. In particular, high prices zipcodes will typically have higher cutoffs than lower priced zipcodes.

At this point, we have divided each zipcode into eight size-price groups. It is possible, however, that some of the groups are too small to provide accurate measures of segment level moments. Our criteria here are that a segment must have enough number of transactions as well as a sufficiently large housing stock. If this is not the case, we merge candidate segments to form a larger joint segment. As a result, some zip codes that have very thin housing markets might have very few segments.

Given a final set of segments we express each search range as a subset. Here we start from the raw search range, specified along the dimensions quality, size and geography, ignoring other dimensions. We then determine the set of segments that is approximately covered by the specified range. We also exclude ranges containing segments with very different median segment prices, relative to the distribution of ranges with an explicit price selection. Since some detail is lost at this step, the number of distinct patterns drops from about 30K to about 9K.

3.3 Market activity and search activity

The following notation is useful to organize facts reported at the segment level. Let H denote the set of all segments. The measure μ^H counts houses, so $\mu^H(h)$ is the housing stock in segment h. Let $V(h)$ denote the average monthly turnover rate in segment h defined as the number of transactions divided by total housing stock. Let $T(h)$ denote the mean time on the market in segment h defined as months between listing and sales date, less one month for the typical escrow period. Our measure of average inventory in segment h is $\mu^S(h) := T(h)V(h)\mu^H(h)$.[4] We also define the inventory share $I(h) = \mu^S(h)/\mu^H(h)$ which is the fraction of all houses that is currently for sale.

Every search range in our sample is a subset of the set of all segments H. We index the ranges by $\theta \in \Theta$ and refer to the set Θ as the set of searcher "types".[5] A searcher of type θ scans inventory in the set of segments $\tilde{H}(\theta) \subset H$. The total housing stock that is of interest to searcher θ is

$$\nu^H(\theta) = \sum_{h \in \tilde{H}(\theta)} \mu^H(h).$$

Similarly, we define the total inventory considered by searcher θ, denoted $\nu^S(\theta)$, as the sum over all inventory $\mu^S(h)$ for sale in segments in θ's search range $\tilde{H}(\theta)$.

[4]This measure of inventory for houses conditions on houses that are eventually sold, since T is based on actual sales. Alternatively, one can construct measures of inventory directly from listings data. The resulting series are noisy because they require assumptions on when listings are removed.

[5]For the presentation of facts in this section, "type" is no more than a label for search ranges. The notation is motivated by our model below where each search range will indeed correspond to a different type of agent (with the search range a feature of preferences).

The *clientele* of segment h consists of all searchers who consider segment h as part of their search range, that is,

$$\tilde{\Theta}(h) = \left\{\theta \in \Theta : h \in \tilde{H}(\theta)\right\}. \tag{1}$$

The pattern of clienteles reflects the interconnectedness of segments. As an extreme example, in a perfectly segmented market, there are $\#H$ types with search ranges each consisting of a single segment, and each segment has a homogenous clientele of one type who searches only that segment, that is $\tilde{H}(\theta) = \{h\}$. In contrast, in a perfectly integrated market there is a single type with $\tilde{H}(\theta) = H$ and all clienteles are identical and contain only that type. More generally, clienteles are heterogeneous and may consist of distinct types with only partially overlapping search ranges. Let $\beta(\theta)$ denote the relative frequency of search ranges θ in the data sample. The distribution of searchers interested in segment h follows by computing the marginal of β on $\tilde{\Theta}(h)$.

For example, as one summary statistic of overall search activity in segment h, we compute the weighted number of searchers per house

$$\sigma(h) = \sum_{\theta \in \tilde{\Theta}(h)} \frac{\beta(\theta)}{\nu^H(\theta)}. \tag{2}$$

Weighting here captures the idea that search effort is somewhat diluted if it is broader. Indeed, if every searcher were looking only at one segment, then $\sigma(h)$ simply reflects the number of searchers per house in h. More generally, some searchers θ in the clientele of h may consider segments other than h. Dividing the number $\beta(\theta)$ of type θ by the housing stock $\nu^H(\theta)$ that this type is interested in makes broader searchers (who are interested in more housing stock) count less towards search activity in h.

So far, all summary statistics have been defined at the segment level only. We are also interested in how market and search activity vary at different levels of aggregation. Since V, I and σ are all defined as ratios relative to housing stock, aggregation uses housing stock as weights. For example, the turnover rate over some subset $G \subset H$, such as a zipcode or city, is computed as

$$\sum_{h \in G} \frac{\mu^H(h) V(h)}{\sum_{h \in G} \mu^H(h)}.$$

We aggregate inventory share and search activity in the same way.

In addition to administrative geographic units, we are also interested in aggregating to sets of segments that are more closely integrated, in the sense that there is is a sufficiently large common clientele. We define the area connected to h as the set of segments $\tilde{h}$ such that the weighted share of searchers scanning both h and $\tilde{h}$ is at least a fraction ϕ of searchers scanning h,

$$A_\phi(h) = \left\{\tilde{h} \in H : \sum_{\theta : h, \tilde{h} \in \tilde{H}(\theta)} \frac{\beta(\theta)}{\nu^H(\theta)} \geq \phi \sum_{\theta : h \in \tilde{H}(\theta)} \frac{\beta(\theta)}{\nu^H(\theta)}\right\}. \tag{3}$$

The distribution of market and search activity

Table 5 presents summary statistics on market and search activity for the Bay Area as a whole as well as by segment. The housing market is illiquid: on average only 1.7 percent of Bay Area housing stock is for sale and the average turnover rate is .34 percent, so the typical house turns over once every 25 years. At the segment level, relative variation is substantial: at the 75th percentile for inventory share there is more than 2.5 times as much inventory than at the 25th percentile. At the 75th percentile for volume houses turn over twice as fast as at the 25th percentile.

Table 5: Summary statistics of market and search activity

	inventory share I (in percent)	turnover rate V (in percent)	search activity σ	mean price (in thous.)	housing stock
Bay Area	1.70	0.34	1.00	645	1,565,259
min	0.09	0.01	0.05	80	1,005
q25	0.81	0.21	0.52	313	1,564
q50	1.33	0.29	0.82	526	2,271
q75	2.17	0.41	1.31	805	3,341
max	12.46	1.75	4.31	2,490	11,178

The measure of search activity (2) has an average of one by construction. Its distribution is positively skewed: the majority of segments have less than one weighted searcher per house. The minimum of .05 is achieved in Martinez in the Sacramento Delta. In contrast, some segments have substantially more search activity, all the way to a maximum of 4.32 in a segment in central San Francisco.

Variation across submarkets

Table 6 reports cross sectional variation in observables at different levels of aggregation. The three left-hand panels show volatilities and correlation coefficients *across* segments, zipcodes and cities, respectively. Comparison of volatilities shows that there is substantial variation in segments which is below the zipcode and city level. Indeed, the zipcode-level movements account for only 53, 59, and 55 percent of the segment-level variance in inventory share I, turnover rate V, and search activity σ, respectively.

Our market activity indicators – inventory share and turnover rate – comove strongly across any type of "submarket": segment, zipcode, or city. Both variables also tend to be higher in cheaper submarkets. In contrast, the comovement of search activity and market activity depends crucially on the level of aggregation. While it is close to zero at the segment level, it is turns negative at the zipcode and even more at the city level. At the same time, the relationship with price also changes: while more expensive segments do not see higher search activity on average, more expensive zipcodes and cities are searched more.

Principal component analysis on the observables in Table 6 further clarifies the multivariate patterns at work in the data. At the segment level, the first principal components explains 61 percent of normalized variation and loads with equal sign on inventory and turnover and (with the opposite sign) on price. In contrast, the second principal component explains 22 percent and loads almost exclusively on search activity. The situation changes at broader units of aggregation: for example, at the city level, loadings make the first principal component (now explaining 68 percent) induce negative correlation between market and search activity.

Variation within submarkets

The three right hand panels of Table 6 consider segment-level variation within submarkets. The bottom two panels report volatilities and correlations within the average zipcode and city. There are 192 zipcodes and 96 cities in our data. All moments are weighted using housing stock.[6] The top panel shows segment level variation within areas connected by significant common clienteles $A_{\phi}(h)$, evaluating (3) with $\phi = .3$. There are as many such areas as there are segments.[7]

For market activity indicators and prices, the nature of covariation across and within submarkets is essentially the same. Indeed, inventory share and turnover rate move together and are both negatively correlated with price. For zipcode and city, the "within" correlation coefficients in the right hand panels are also quantitatively close to the "across" correlation coefficients in the left hand panels.

In contrast, the sign of comovement between search activity on the one hand and market activity and price on the other depends on whether we look within or across submarkets. Indeed, for both zipcodes and cities, search activity moves together with market activity and against price across units, but it moves against market activity and with price within units. The signs of within correlations are the same for connected areas that defined on the basis of common clienteles as opposed to geographic closeness.

The relationship between inventory and measures of search activity is reminiscent of the "Beveridge curve" that relates vacancies and unemployment in labor market statistics. The stylized fact here is that the housing-market Beveridge curve is *downward sloping* across broad units of aggregation, while it is on average *upward sloping* within broad units. In fact, the Beveridge curve is upward sloping within 81 out of 96 cities that represent 68 percent of the total housing stock.

The within-city Beveridge curve also slopes up for 15 of the largest 20 cities. One notable exception among the latter is San Jose, which has a correlation coefficient of $-.08$. The fact is not primarily driven by small cities, however. Indeed, the 15 cities in the top 20 that have an upward sloping Beveridge curve have an average slope of .53 and a 25th percentile slope of .38. The slope for San Francisco is .66. At the same time, negative slopes among the top

[6] The unweighted median number of segments for both zipcodes and cities is equal to 3. However, the distribution of cities is highly skewed. For example, San Francisco and San Jose contain 101 and 70 segments, respectively.

[7] There are as many connected areas as there are segments. The median and 75th percentile area by number of segments consist of 5 and 10 segments, respectively. The main qualitative message below is not sensitive to the choice ϕ.

Table 6: Cross sectional variation in market and search activity

	variation across units				avg variation within units			
	I	V	σ	$\log(p)$	I	V	σ	$\log(p)$
segment					connected area			
vol	1.26	0.19	0.68	0.66	0.80	0.12	0.56	0.34
corr	1	.95	.01	−.67	1	.89	.33	−.45
		1	−.01	−.57		1	.29	−.13
			1	.05			1	−.15
				1				1
zipcode					zipcode			
vol	0.92	0.14	0.52	0.52	0.90	0.13	0.43	0.42
corr	1	.95	−.21	−.73	1	.87	.51	−.76
		1	−.17	−.64		1	.36	−.61
			1	.42			1	−.47
				1				1
city					city			
vol	0.78	0.13	0.40	0.47	0.84	0.13	0.38	0.43
corr	1	.96	−.39	−.80	1	.84	.51	−.74
		1	−.30	−.71		1	.39	−.55
			1	.53			1	−.46
				1				1

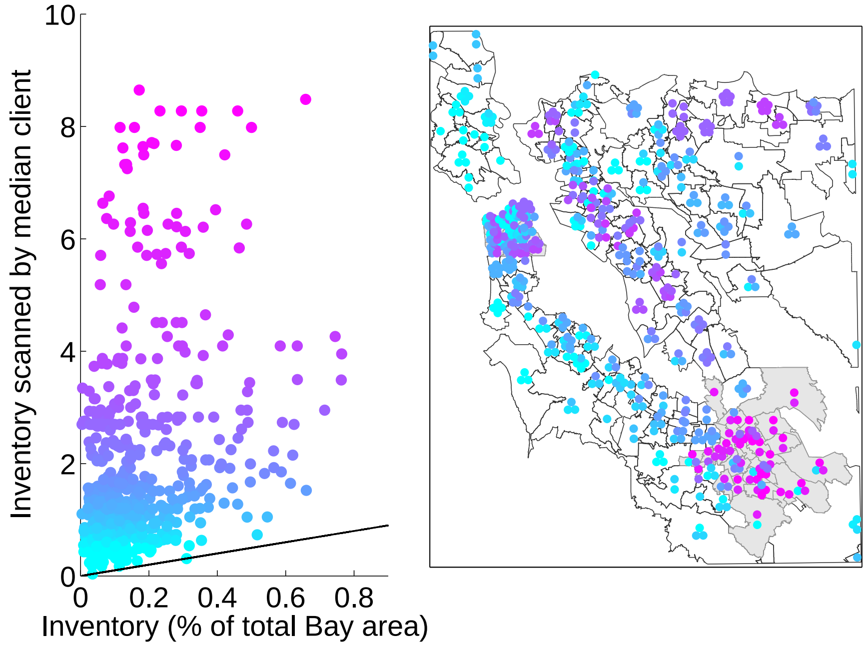
5 cities range between −.08 and −.26.

Breadth of search & integration

The summary measure $\sigma(h)$ reflects average search activity in a segment, but it does not tell us whether that activity is due to narrow local searchers or due to broader searchers who provide connection to other segments. To summarize interconnectedness, we now compare segments in terms of the inventory scanned by their typical client. The left-hand panel of Figure 5 plots the share of inventory in segment h in total Bay Area inventory (measured along the horizontal axis) against the inventory scanned by the median client of segment h (measured along the vertical axis). Every dot represents a segment, and color reflects the value on the vertical axis so the segments can be recognized in the map in the right-hand panel.

If the Bay Area were perfectly segmented, then any given segment would only have clients who scan that particular segment. As a result, all points would have to line up along the 45-

Figure 5: Scanned inventory



degree line. At the opposite extreme, if the Bay Area were perfectly integrated, then every client of every segment would scan all houses, so all points should line up along a horizontal lines at 100 percent of total inventory. Not surprisingly, the truth is in the middle: the median searcher in a segment scans multiple times more inventory than is available in the segment itself, but far less than 100 percent of the total.

Areas with a large common clientele appear in the plot as near-horizontal clusters: if any subset of segments were perfectly integrated but not connected to other segments, then it would form a horizontal line at the level of its aggregate inventory. The relevance of this effect is visible for the top cluster of pink dots. The map in the right hand panel shows that those dots represent cheaper segments in the city of San Jose. More generally, clusters of dots with high scanned inventory correspond to cheap urban areas where broad search appears to be more common.

The first column of Table 7 summarizes the distribution of inventory scanned by the median client. In the average segment, the median client scans 2 percent of the total inventory, or 45 houses. The table also clarifies that most dots in Figure 5 are clustered in the bottom left; the 75th percentile is at only at 2.7 percent of total inventory. The second column in Table 7 shows the distribution of within segment interquartile ranges for scanned inventory. The point here is that there is substantial clientele heterogeneity. Indeed, the average within-segment IQ range of inventory scanned by different searchers is, at 1.77 percent, quite similar to the across-segment IQ range of inventory scanned by the median searcher. Interestingly, clientele heterogeneity comoves strongly with overall connectedness: the correlation coefficient between the first and second columns is 63 percent. In other words, in segments that are on average more integrated with other segments, there are larger within-segment differences between the interacting narrower and broader searchers.

Table 7: Variation in scanned inventory

	by segment: total inv. scanned (in percent)		by zipcode: share of search ranges (in percent)				by city: share of search types (in percent)			
	median	IQ range	one	multiple	< one	other	one	multiple	subset	other
mean	2.00	1.77	5.5	3.8	6.2	84.5	9.7	6.7	17.8	65.8
q25	0.80	0.84	0	0	0.9	73.9	0	0	3.2	54.6
q50	1.38	1.57	1.2	0.9	2.4	93.5	4.9	1.1	13.5	66.4
q75	2.70	2.31	4.4	3.9	6.7	97.2	15.2	7.0	27.9	79.9

Search at the city and zipcode level

The right-hand columns of Table 7 ask how important detailed segment-level information is for understanding search patterns. For each zipcode and city, we first classify the share of searchers active in that zipcode or city who search entire zipcodes or cites, respectively. The categories labeled "one" and "multiple" distinguish further between searchers who scan exactly one versus exactly an integer number of zipcodes or cities. Together, they indicate the share of searchers for whom detailed segment level information is not important. The category labeled "one" collects searchers who scan only a subset of a segment, while "other" mops up other searchers for whom segment information matters because their range intersects with multiple zipcodes or cities.

The table reports mean and quartiles for the shares of each category of searcher in the cross section of segments. For example, in the average segment, only 5.5 percent of searchers select exactly the zipcode containing that segment. The distribution is highly skewed: in 75 percent of segments, the share of searchers scanning the zipcode is 4.4 percent or less. The order of magnitude of the numbers is similar at the city level and whether we consider multiple versus a single unit. We can therefore conclude that the clientele patterns at work in our data is due to searches that are not simply driven by zipcode or city. Instead, other characteristics defining a segment, in particular size and quality, play an important role.

4 Model setup

The model describes a small open economy, such as the San Francisco Bay Area. Time is continuous and the horizon is infinite. Agents live forever and discount the future using the riskless rate r.

Segments, search ranges and clienteles

The model allows for agents of different types who search across different segments of the housing market, as in our data. We use the notation introduced in Section 3.3. There is a finite set H of market segments. The measure μ^H on H counts the number of houses in each segment. We normalize the total number of houses in the economy to one:

$$\sum_{h \in H} \mu^H(h) = 1.$$

Agents have quasilinear utility over two goods: numeraire and housing services. Agents own at most one house. When an agent moves into a house, he obtains housing services $v(h) > 0$ until the house falls out of favor, which happens at the rate $\eta(h)$. After the house falls out of favor, the agent no longer receives housing services from that particular house. The agent can then put the house on the market in order to sell it and subsequently search for a new house. We assume that the search for a new house is costless, whereas putting a house on the market in segment h involves costs $c(h)$ per period.

Agent type θ is identified by a *search range*, a subset $\tilde{H}(\theta) \in H$ of market segments that he is interested in. Search ranges are part of the description of preferences – an agent will never move into a house outside $\tilde{H}(\theta)$. We use a measure μ^Θ on the set of all types Θ to count 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. The idea is that these $\bar{\mu}^\Theta - 1$ agents rent or stay in a hotel while they search for a house to buy.

The *clientele* $\tilde{\Theta}(h)$ of segment h is the set of all agents who are interested in segment h, as defined in (1). It is helpful to consider two extremes. The market is perfectly segmented if every segment is searched by a single type who is interested only in that segment. The clienteles $\tilde{\Theta}(h)$ are then disjoint sets that each contain a single type θ. In contrast, the market is fully integrated if there is only one type who searches all segments; all clienteles $\tilde{\Theta}(h)$ are then identical and contain the same type. The *inventory scanned* by type θ is $\nu^S(\theta)$.

Matching

Matching in the housing market involves searchers scanning inventory, identifying suitable properties and making contact with sellers. We capture this process by a random matching technology. We make two assumptions. First, searchers flow into segments within their search range in proportion to segment inventory. This assumption is natural if searchers are equally likely to find a favorite house anywhere in their search range. Formally, let $\tilde{\mu}^B(\theta)$ denote the number of buyers of type θ. We define the number of buyers in segment h as

$$\mu^B(h) = \sum_{\theta \in \tilde{\Theta}(h)} \frac{\mu^S(h)}{\nu^S(\theta)} \tilde{\mu}^B(\theta). \tag{4}$$

For the given segment h, buyers can belong to any type in the clientele $\tilde{\Theta}(h)$. If a type θ searches only segment h, then $\nu^S(\theta) = \mu^S(h)$ and all buyers $\tilde{\mu}^B(\theta)$ of type θ are in fact buyers in h. If segments have roughly the same inventory, searchers are equally likely to be buyers in any of the segments in their search range. More generally, the more inventory is available in h relatively to other segments in type θ's search range, the larger the share of type θ buyers who flow into h.

Our second assumption is the presence of a matching function. The match rate in segment h is given by

$$m(h) = \tilde{m}\left(\mu^B(h), \mu^S(h), h\right),$$

where $\tilde{m}$ is increasing in the number of buyers and sellers and satisfies $\tilde{m}\left(0, \mu^S, h\right) = \tilde{m}\left(\mu^B, 0, h\right) = 0$. At this point, we do not make further assumptions on the functional form of the function $\tilde{m}$. What is important is that it is allowed to depend on the segment h directly (other than through the number of buyers and inventory). For example, the process of scanning inventory could be faster in a segment because the properties are more standardized, or because more open houses are available to view properties.

Once a buyer and seller have been matched, the seller 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 $v(h)$.

Equilibrium

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 put their house on the market if and only if their house has fallen out of favor, so that the owners do not receive housing services from it, and (ii) all offers are accepted.

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 the different individual states. Let $\mu^H(h;\theta)$ denote the number of type θ agents who are homeowners in segment h, and let $\mu^S(h;\theta)$ denote the number of type θ agents whose house is listed in segment h. In steady state, all these numbers, as well as the numbers of buyers by type $\tilde{\mu}^B(\theta)$ and by segment $\mu^B(h)$, are constant. We now derive a set of equations to determine them.

The first set of equations uses the fact that $\mu^S(h)$, the number of houses for sale in segment h, is constant in steady state. As a result, the number of houses newly put on the market in segment h must equal the number of houses sold in segment h:

$$\eta(h)\left(\mu^H(h) - \mu^S(h)\right) = \tilde{m}\left(\mu^B(h), \mu^S(h), h\right). \tag{5}$$

The left-hand side shows houses coming on the market, given by the rate that houses fall out of favor multiplied by the number of houses that are not already on the market. The right-hand side shows the number of matches and thus the number of houses sold.

The second set of equations uses the fact that the rate at which houses fall out of favor in segment h is the same for all types in the clientele of h. As a result, the share of houses owned by type θ agents in h must equal the share of houses bought by type θ agents in h:

$$\frac{\mu^H(h;\theta)}{\mu^H(h)} = \frac{\mu^S(h)}{\nu^S(\theta)} \frac{\tilde{\mu}^B(\theta)}{\mu^B(h)}. \tag{6}$$

On the right-hand side, the share of type θ buyers in segment h equals the number of type θ buyers that flow to h in proportion to inventory, as in (4), divided by the total number of buyers in segment h. The equation also says that the buyer-owner ratio for any given type θ in segment h is the same and equal to the segment level buyer-owner ratio $\mu^B(h)/\mu^H(h)$.

Finally, the number of agents and the number of houses must add up to their respective totals:

$$\begin{aligned} \mu^H(h) &= \sum_{\theta \in \tilde{\Theta}(h)} \mu^H(h;\theta), \\ \mu^\Theta(\theta) &= \tilde{\mu}^B(\theta) + \sum_{h \in \tilde{H}(\theta)} \mu^H(h;\theta). \end{aligned} \tag{7}$$

Equations (5), (6) and (7) jointly determine the unknown numbers $\mu^H(h;\theta)$, $\mu^S(h)$, $\mu^B(h)$ and $\tilde{\mu}^B(\theta)$, a system of $2\#H + \#\Theta(1 + \#H)$ equations in as many unknowns.

Parameters

The model identifies three forces that determine market activity and prices in the cross section. Two forces operate at the segment level. First, the rate $\eta(h)$ at which houses fall out of favor represents differences in the supply of housing across segments. In what follows, we refer to $\eta(h)$ as a measure of *instability*: a more unstable segment is one where more houses come on the market per period. The second force is the segment-specific effect on match rates summarized by $\tilde{m}(.,.,h)$ which represents differences market frictions across segments, respectively. The third force is the demand for housing which is captured by the distribution of search ranges $\tilde{H}(\theta)$ and the number of agents $\mu^\Theta(\theta)$ of type θ.

Housing demand parameters are more complicated to study since their effect depends on the entire clientele pattern. It is nevertheless helpful to consider a summary measure at the segment level. We define the *popularity* of a segment by

$$\pi(h) = \sum_{\theta \in \tilde{\Theta}(h)} \frac{\mu^\Theta(\theta)}{\nu^H(\theta)}. \tag{8}$$

A segment is more popular if there are more agents per house who include it in their search ranges. The measure is analogous to the measure of weighted searchers (2) in that broad searchers who look at multiple segments other than h count toward the popularity of segment h. A key difference is that $\pi(h)$ is an exogenous determinant of demand for segment h, because it captures the distribution $\mu^\Theta(\theta)$ of preferences, whereas search activity $\sigma(h)$ is determined endogenously, as described below.

Observables

The observables described in Section 3.3 all have model counterparts. The *inventory share* is $I(h) = \mu^S(h)/\mu^H(h)$ and the *turnover rate* is $V(h) = m(h)/\mu^H(h)$. Search alerts represent a sample of buyers. The relative frequencies of the search ranges $\tilde{H}(\theta)$ in the model are given by $\beta(\theta) = \tilde{\mu}^B(\theta)/\left(\bar{\mu}^\Theta - 1\right)$ and are thus observable up to the constant $\bar{\mu}^\Theta - 1$. Our measure of search activity at the segment level can be written as

$$\sigma(h) = \frac{1}{\bar{\mu}^\Theta - 1} \sum_{\theta \in \tilde{\Theta}(h)} \tilde{I}(\theta) \frac{\tilde{\mu}^B(\theta)}{\nu^S(\theta)}, \tag{9}$$

where $\tilde{I}(\theta) = \nu^S(\theta)/\nu^H(\theta)$ is the inventory share measured over the search range of type θ.

Exact identification of parameters

The structure of the model implies that the supply and demand parameters — $\eta(h)$ and $\mu^\Theta(\theta)$, respectively — can be identified without taking a stand on the exact shape of the matching function. All that is required is that the dependence of the matching function on the segment is sufficiently flexible that the model can jointly match the inventory share $I(h)$, the turnover rate $V(h)$, and the relative frequencies of search ranges $\beta(\theta)$. We now derive this result from equations (5)-(7).

Dividing the market clearing condition (5) by the housing stock $\mu^H(h)$, we obtain

$$\eta(h)\,(1 - I(h)) = V(h). \tag{10}$$

The frequency of moving shocks $\eta(h)$ can thus be inferred from inventory and turnover alone. Moreover, we know from the summary statistics in Table 5 that inventory shares are small, their 90th percentile is at 3.5%. As a result, the parameter must closely track the turnover rate by segment. Intuitively, because the time a house remains on the market is much shorter than the time that it is occupied by an owner, turnover is almost entirely accounted for by the frequency of moving shocks.

The match rate for a buyer who flows to segment h is $\alpha(h) = m(h)/\mu^B(h)$. Using the definition of buyers (4), it can be expressed in terms of observables (up to a constant) as

$$\frac{1}{\alpha(h)} = \sum_{\theta \in \tilde{\Theta}(h)} \frac{I(h)}{\tilde{I}(\theta)} \frac{\beta(\theta)\left(\bar{\mu}^\Theta - 1\right)}{\nu^H(\theta)} \frac{1}{V(h)}. \tag{11}$$

Interpreting terms from the right, we have that matching is fast (at a high rate $\alpha(h)$) in segment h if turnover is high in h, if the buyer-owner ratio is high for types in the clientele of h, and if inventory is high in h relative to other segments in its clientele's search ranges.

We do not have information on the overall number of buyers $\bar{\mu}^\Theta - 1$. As an additional target moment, we set the average match rate for a buyer to 20% which is also the average match rate for inventory in our data. The average time it takes for a buyer to find a house is therefore about 5 months. This choice does not affect the relative behavior of market and search activity across segments, and therefore is not particularly important for most of our results. Formally, the average inventory-weighted match rate across types θ is the same as the average inventory-weighted match rate across segments.[8] We thus determine the constant $\bar{\mu}^{\Theta} - 1$ by setting the average of the buyer match rates $\alpha(h)$ to the average of the inventory match rate $I(h)/V(h)$.

Once the total number of buyers is determined, the number of buyers by type $\tilde{\mu}^B(\theta) = \beta(\theta)\left(\bar{\mu}^{\Theta} - 1\right)$ and by segment $\mu^B(h) = \mu^H(h)V(h)/\alpha(h)$ follow immediately. Substituting for $\mu^H(h;\theta)$ in (7) using (6) we can solve out for the type distribution $\mu^{\Theta}(\theta)$ from

$$\frac{\mu^{\Theta}(\theta) - \tilde{\mu}^B(\theta)}{\tilde{\mu}^B(\theta)} = \sum_{h \in \tilde{H}(\theta)} \frac{\mu^S(h)}{\nu^S(\theta)} \frac{\mu^H(h)}{\mu^B(h)}. \tag{12}$$

The adding up constraint says that the owner-buyer ratio for type θ agents should be the inventory weighted average of owner-buyer ratios at the segment level. We will therefore infer the presence of more types θ not only if we observe more buyers of type θ (higher $\beta(\theta)$ and hence $\tilde{\mu}^B(\theta)$), but also if type θ's search range has on average relatively more owners relative to buyers. In the latter case, more types θ agents are themselves owners, so their total number is higher.

At this point we have identified the supply and demand parameters of the model without specific assumptions on the functional form of the matching function. If we postulate such a functional form, restrictions on its parameters follows from equation (5). For example, consider the Cobb-Douglas case with a multiplicative segment-specific parameter $\bar{m}(h)$ that governs the speed of matching

$$\tilde{m}\left(\mu^B(h), \mu^S(h), h\right) = \bar{m}(h)\,\mu^B(h)^{\delta}\,\mu^S(h)^{1-\delta}.$$

For a given weight δ, the speed of matching parameter $\bar{m}(h)$ can be backed out from observables as $\bar{m}(h) = \alpha(h)^{\delta}\left(V(h)/I(h)\right)^{1-\delta}$. The speed of matching parameter is thus a geometric average of the buyer and inventory match rates.

4.1 The role of heterogeneous clienteles

We now develop some intuition for how the exogenous forces – demand, supply and frictions – drive the cross section of observables in equilibrium, and how their effects depend on the level of aggregation. The mapping from parameters to observables depends on the nature of the search patterns. For example, parameters that are "local" to segment h, such as the rate at which houses come on the market there, will matter less for local inventory and time

[8] Let $\bar{\mu}^S$ denote total inventory and consider the identity

$$\left(\bar{\mu}^S\right)^{-1} \sum_{\theta \in \Theta} \frac{\nu^S(\theta)}{\tilde{\mu}^B(\theta)} \sum_{h \in \tilde{H}(\theta)} \frac{\mu^S(h)}{\nu^S(\theta)} \frac{\tilde{\mu}^B(\theta)}{\mu^B(h)} m(h) = \left(\bar{\mu}^S\right)^{-1} \sum_{h \in \tilde{H}(\theta)} \frac{\mu^S(h)}{\mu^B(h)} m(h).$$

Here the right hand side is the average match rate across segments and the left hand side is the average match rate across types. In particular, the second sum on the left hand side is number of matches entered by type θ which depends on the relative inventory available in θ's search range as well as the share of θ in each segment's buyer pool.

on the market if segment h is more integrated with similar segments. A natural benchmark is therefore the extreme case of perfect segmentation.

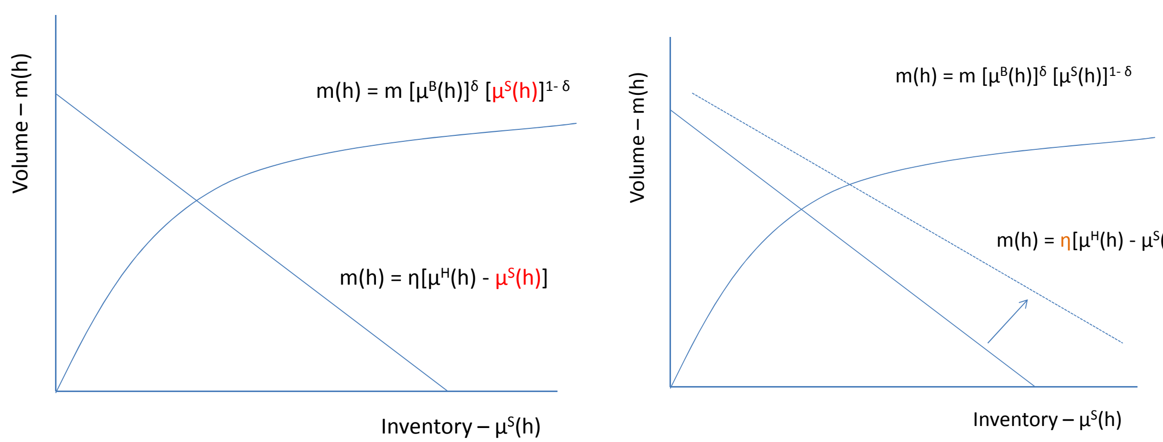
Perfect segmentation

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. The number of buyers is the difference between the number of types interested in h and the number of houses in segment h. Substituting into (5), equilibrium inventories are determined segment by segment by

$$\eta\left(h\right)\left(\mu^{H}\left(h\right)-\mu^{S}\left(h\right)\right)=\tilde{m}\left(\mu^{\Theta}\left(h\right)-\mu^{H}\left(h\right),\mu^{S}\left(h\right),h\right). \qquad (13)$$

The left panel in Figure 6 plots both sides of (13) for segment h. The left-hand side of the equation is the rate at which houses come on the market in segment h. It is a curve that 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 is the rate at which houses are sold. It is a strictly increasing curve in inventory: higher inventory means that buyers are more likely to be matched with a house. 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.

Figure 6: Equilibrium and comparative statics with perfect segmentation



Implications for the cross section of segments obtain from comparative statics. The right plot in Figure 6 makes the segment more unstable by increasing $\eta\left(h\right)$. When houses come on the market more quickly in h, the downward-sloping curve shifts to the right. In this case, volume increases together with inventory. In contrast, if the segment is more popular (higher $\pi\left(h\right)=\mu^{\Theta}\left(h\right)/\mu^{H}\left(h\right)$ as defined in (8)) so there are more buyers per house in h, the upward-sloping curve shifts up (not depicted in Figure 6). In this case, volume increases, but inventory decreases. Intuitively, an increase in either demand or supply increases volume. The difference is that an increase in supply also make the market clear more slowly (time on the market is higher) so inventory is higher.

Suppose we want to explain the stylized facts in Table 6 with a perfectly segmented model. At any level of aggregation, we observe a strong positive relationship between inventory and turnover. It follows that differences in supply – the parameter $\eta(h)$ – must be important, whereas differences in popularity must be weak enough so as not to overturn the positive relationship between inventory and turnover. At the same time, (9) says that search activity in the perfectly segmented case simply reflects the relative number of buyers scanning segment h

$$\sigma(h) = \frac{\mu^{\Theta}(h) - \mu^{H}(h)}{(\bar{\mu}^{\Theta} - 1)\,\mu^{H}(h)} = \frac{\pi(h) - 1}{\bar{\mu}^{\Theta} - 1}. \tag{14}$$

Variation in search activity across segments is thus driven only by variation in popularity. Since more popular segments have lower inventory shares, such variation in itself will always generate a downward sloping Beveridge curve. To generate instead an upward sloping Beveridge curve requires comovement of $\pi(h)$ and $\eta(h)$: if more unstable segments are also more popular, then high inventory and turnover driven by ample supply can in principle go along with more search activity driven by high demand.

Finally, consider a change in the matching technology that allows for more matches per period for given buyer and seller pools. From the equilibrium condition (13), this comparative static works like an increase in popularity: turnover increases while inventory declines. At the same time, changes in matching technology have no effect on our measure of search activity.

Partial integration and the role of broad searchers

To illustrate how heterogeneity of clienteles affects the cross section of market and search activity, we extend the example by adding one additional type: a "broad searcher" who scans all segments $h \in H$. We denote this type by 0 so $\nu^{S}(0)$ is total inventory. The buyer pool of segment h now contains narrow searchers of type h and also broad searchers who flow into segment h depending on the share of segment h inventory in total inventory (by the definition of buyers (4)). Equilibrium inventories again adjust equate the flow of houses coming on the market to the volume of sales:

$$\eta(h)\left(\mu^{H}(h) - \mu^{S}(h)\right) = \tilde{m}\left(\tilde{\mu}^{B}(h) + \frac{\mu^{S}(h)}{\nu^{S}(0)}\tilde{\mu}^{B}(0), \mu^{S}(h), h\right). \tag{15}$$

The key new feature with partial integration is that the buyer pool is endogenous and tends to move positively with inventory in equilibrium. Two effects are relevant here. On the one hand, the direct effect apparent from (15) is that a larger share of broad type 0 searchers flows to segments with higher inventory. On the other hand, competition from broad searchers implies that the number of narrow searchers looking for a house in h also increases with inventory. To see this, use the implication of (6) that buyer-owner ratios in any given segment are equated across all types in the clientele. Comparing narrow and broad buyers and owners in the clientele of h, we have

$$\frac{\tilde{\mu}^{B}(h)}{\mu^{\Theta}(h) - \tilde{\mu}^{B}(h)} = \frac{\mu^{S}(h)}{\nu^{S}(0)} \frac{\tilde{\mu}^{B}(0)}{\left[\mu^{H}(h) - \left(\mu^{\Theta}(h) - \tilde{\mu}^{B}(h)\right)\right]}, \tag{16}$$

where the second denominator on the right hand side determines broad owners in h as a residual. Holding fixed the number of houses and narrow types, a segment with a higher share of inventory must have more narrow buyers.[9]

The role of parameters for the cross section of inventory shares and turnover rates is qualitatively similar to the perfect segmentation case. In more unstable segments, inventory and volume both tend to be higher. However, the effect on inventory will typically be weaker because more searchers flow into segment h as more houses come on the market there. In other words, a higher supply endogenously gives rise to an offsetting increase in demand. An increase in the speed of matching or a larger number of types interested in h will increase turnover and decrease inventory. Of course, the interdependence of segments implies that the magnitude of effects is now more complicated to assess. For example, how the relative inventory of two segments depends on their stability now depends on the stability of other segments as well.

An important difference to the perfectly segmented case is how the presence of broad searchers alters the mapping between parameters and search activity. With partial integration, search activity (2) becomes

$$\sigma(h) = \frac{1}{\bar{\mu}^{\Theta} - 1}\left(\frac{\tilde{\mu}^{B}(h)}{\mu^{H}(h)} + \tilde{\mu}^{B}(0)\right). \tag{17}$$

Differences in search activity across segments are driven by differences in narrow buyers per house, since the contribution of broad searchers is same for all segments. It then follows from (15)-(16) that for two equally popular segments (which have identical $\mu^{\Theta}(h)$ and $\mu^{H}(h)$), the segment with more instability or slower matching must have higher inventory together with higher search activity. In other words, with partial integration, differences in stability or the speed of matching can account for an upward sloping Beveridge curve.

Hypothetical perfectly segmented benchmark

How large is the contribution of partial integration to the Beveridge curve? We ask what the Beveridge curve looks like in a economy with demand parameters changed so as to remove integration, but with all other parameters held fixed. In principle, there are many ways to construct such an economy: they differ in how broad searchers are replaced by narrow searchers of different types. A simple benchmark is a perfectly segmented economy that delivers the same inventory $I(h)$, turnover $V(h)$, and buyer match rates $\alpha(h)$. This is the economy considered by an econometrician who observes $I(h)$ and $V(h)$ as well as match rates by segment (or who knows the matching functions), but who does not have information on integration and proceeds to assume that the economy is perfectly segmented.[10]

[9]The remaining equations determining equilibrium is the definition of $\nu^{S}(0)$ as total inventory and the requirement that buyers add up to the correct total, that is,

$$\bar{\mu}^{\Theta} - 1 = \tilde{\mu}^{B}(0) + \sum_{h \in H} \tilde{\mu}^{B}(h).$$

[10]Importantly, the experiment here is a comparative static on the demand parameters designed to measure the contribution of broad searchers observed in the data to the Beveridge curve observed in the data. It

Since the hypothetical perfectly segmented economy has the same match rate and housing stock as the original economy, it also has the same number of equilibrium buyers. From (12) the type distribution for the hypothetical economy is given by $\hat{\mu}^{\Theta}(0) = 0$ and $\hat{\mu}^{\Theta}(h) = \mu^{H}(h) + \mu^{B}(h)$. Combining (14) and (17), we can write search activity in the hypothetical economy as

$$\hat{\sigma}(h) = \left(\bar{\mu}^{\Theta} - 1\right)\left(\sigma(h) + \beta(0)\left(\frac{I(h)}{\tilde{I}(0)} - 1\right)\right). \tag{18}$$

While the actual Beveridge curve consists of the locus $(\sigma(h), I(h))$ measured in the data, the hypothetical Beveridge curve adds an extra upward sloping piece that becomes more important as the share of broad searchers increases. Intuitively, removing broad searchers and replacing them by narrow searchers implies that differences in search activity must be explained by differences in popularity that are directly reflected in search activity.

5 Quantitative results

Table 8 summarizes the basic properties of demand and supply parameters. We report moments of instability $\eta(h)$ as well as popularity $\pi(h)$, our segment level summary statistic of demand. The top panel of the table provides information on the distribution of the parameters. The bottom three panels report correlations both among the parameters themselves and between parameters and observables. Here we compare variables at the segment level and with city level averages as well as variation within the city of San Francisco. We focus on San Francisco because it is the city with the largest number of segments.

The table also reports the properties of the inferred match rate $\alpha(h)$. While $\alpha(h)$ is an endogenous object rather than a parameter, it contains information about the role of matching frictions.[11] Since we do not have information to identify the shape of the matching function, we do not directly draw conclusions on matching technology. We only record what can be about the distribution of match rates from search behavior.

Instability and popularity at the segment and city level

As expected from equation (10), instability $\eta(h)$ tracks turnover almost exactly. Its moments in Table 8 are essentially the same as those reported for turnover in Table 5. In particular, unstable segments are cheaper and see more turnover and larger inventories. This is true not only across segments, but also across cities and within San Francisco.

Popularity $\pi(h)$ ranges overall between .2 and 2.4, with an IQ range between .82 and 1.17. The fact that popularity is below one for many segments is indicative of the role of partial

is also possible to construct a perfectly segmented economy that explains the data exactly by allowing the matching function to vary simultaneously.

[11]For example, with a Cobb-Douglas matching function $m(h) = \bar{m}(h)\mu^{B}(h)^{\delta}\mu^{S}(h)^{1-\delta}$, we would have

$$\log \bar{m}(h) = \delta \log \alpha(h) + (1-\delta)\log\left(V(h)/I(h)\right).$$

Table 8: Estimated parameters and hypothetical segmentation

	parameters $100 \times \eta(h)$	$\pi(h)$	$\alpha(h)$		hypothetical perfectly segmented case $\hat{\sigma}(h)$	$\hat{\sigma}(h) - \sigma(h)$	
mean	.34	1.02	0.16		1.02	0	
q25	.21	0.82	0.06		0.44	-0.17	
q50	.29	1.01	0.10		0.75	-0.04	
q75	.41	1.17	0.18		1.36	0.12	

	$\eta(h)$	$\pi(h)$	$\alpha(h)$	$I(h)$	$V(h)$	$\sigma(h)$	$\log(p(h))$
correlation across segments							
$\eta(h)$	1	$-.11$	.42	.95	1.00	$-.01$	$-.54$
$\pi(h)$		1	$-.48$	$-.12$	$-.10$	.78	.10
$\alpha(h)$			1	.43	.41	$-.44$	$-.24$
correlation across cities							
$\eta(h)$	1	$-.21$	.52	.95	1.00	$-.25$	$-.63$
$\pi(h)$		1	$-.64$	$-.24$	$-.21$	.74	.30
$\alpha(h)$			1	.56	.52	$-.66$	$-.62$
correlation within San Francisco							
$\eta(h)$	1	.35	$-.29$	.86	1.00	.50	$-.18$
$\pi(h)$		1	$-.59$	.35	.35	.85	$-.05$
$\alpha(h)$			1	$-.51$	$-.29$	$-.63$	.38

integration. Indeed, if segments were either perfectly segmented or perfectly integrated, then the number of weighted buyer would be larger than the number of houses in all segments, and popularity would have to be above one.

To see how partial integration can imply $\pi(h) < 1$, consider a simple example: assume there are two equally large and equally unstable segments 1 and 2, say, as well as an equal number of narrow searchers who scan only segment 1 and broad searchers who scan both segments. We then have $\pi(1) = 3/2$ and $\pi(2) = 1/2$. Intuitively, a segment that is considered largely by broad searchers will tend to have low popularity.

The slope of the Beveridge curve

Consider now how our model accounts for the slope of the Beveridge curve at different levels of aggregation. Popularity comoves strongly with search activity at the segment and city levels as well as within San Francisco. At the same time, popularity correlates positively with instability within San Francisco and negatively across segments and cities. It follows that part of the behavior of the Beveridge curve is explained by differences in the relationship between the exogeneous parameters. More popular cities are more stable, while more popular segments within cities are less stable.

As shown via example above, the second effect that can contribute to an upward sloping

Beveridge curve in our model is the endogenous response of a partially integrated set of clienteles to differences in instability or the matching technology, regardless of the behavior of popularity. This effect is consistent with upward slopes in cities, which tend to be more integrated. The contribution of this effect is more difficult to quantify directly, since it requires taking a derivative with respect to the entire structure of clientele patterns in the direction of less integration. We can however measure the strength of the endogenous response of broad searchers by comparing our model to a perfectly segmented benchmark; we turn to this next.

The role of partial integration

To assess the role of partial integration, we compute search activity in a hypothetical perfectly segmented economy. We construct a comparison economy along the lines introduced in our example in Section 4.1. In particular, we set the number of agents interested in segment h to the sum of the housing stock and the equilibrium buyers in h: $\hat{\mu}^{\Theta}(h) = \mu^{H}(h) + \mu^{B}(h)$. By construction, this economy matches the same pattern of market activity (inventory, turnover and match rates) as the actual economy. It also holds fixed the parameters governing supply and the matching function. However, the two economies generally differ in demand parameters and therefore in search activity, because the hypothetical economy shuts down the endogenous response of broad searchers to inventory.

Comparing the differences in search activity in the hypothetical and actual economy, $\hat{\sigma}(h)$ and $\sigma(h)$, respectively, provides a measure of the importance of endogenous responses by broad searchers at the segment level. Indeed, if $\hat{\sigma}(h)$ is, say 10% larger than $\sigma(h)$ in segment h, then market activity in segment h is driven by an endogenous buyer pool that contains many braod searchers and hence looks *as if* there were 10% more narrow searchers interested in segment h. Conversely, if $\hat{\sigma}(h)$ is lower than $\sigma(h)$ then broad searchers stay away from that segment in equilibrium and the buyer pool looks as if there are less narrow searchers interested in h.

Search activity in the hypothetical economy is compared to actual search activity in the top right of Table 8. On average, the two measures must be equal by construction, as suggested by the definition (18) of $\hat{\sigma}(h)$. Away from the mean, however, we observe substantial differences between the two measures of search activity. Those increase further in the tails: the 10th and 90th percentile of $\hat{\sigma}(h)$ are at -.40 and .40, respectively.

From (18), one would expect that differences are generated by differences in inventory. In fact, the correlation of $\hat{\sigma} - \sigma$ with I is 47% in the entire sample. It is less than one since computing correlations for the entire Bay Area averages over several partially integrated clusters, whereas the example in Section 4.1 assumes that broad searcher scan all segments. Within San Francisco, resembles more closely a perfectly integrated market, the correlation between $\hat{\sigma} - \sigma$ and I is is 78%.

6 Prices and spillovers

6.1 Equilibrium prices

Denote by $V^F(h;\theta)$ the utility of a type θ agent who obtains housing services from a house in segment h. Since sellers make take-it-or-leave offers and observe buyers' types, they charge prices equal to buyers' continuation utility. The price paid by a type θ buyer in segment h is thus $p(h,\theta) = V^F(h;\theta)$. We now show that prices are the same in all transactions in segment h. Start from the Bellman equation of a seller who puts his house on the market

$$rV^S(h;\theta) = -c(h) + \frac{m(h)}{\mu^S(h)}\left(E\left[p(h,\theta)|h\right] - V^S(h;\theta)\right)$$

where the expectation uses the equilibrium distribution of buyers $\mu^B(h;\theta)/\mu^B(h)$. It follows that the value function of the seller is therefore independent oftype. Intuitively, will be charged his continuation value as a buyer, and so cares only about the expected sale value.

Consider now the Bellman equations of an owner who does not put his house up for sale

$$rV^F(h;\theta) = v(h) + \eta(h)\left(V^S(h;\theta) - V^F(h;\theta)\right)$$

Since utility $v(h)$ and the arrival of moving shocks are also independent of type, so is $V^F(h;\theta)$. As a result, the same price $p(h)$ in all transactions in segment h.

We can combine these equation and solve for the price

$$p(h) = \frac{v(h)}{r} - \frac{\eta(h)}{r + m(h)/\mu^S(h) + \eta(h)}\frac{v(h) + c(h)}{r} \tag{19}$$

The first term is the present value of a permanent flow of housing services. 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.

6.2 Liquidity discounts

We now ask how market frictions identified by our estimation affect the dispersion of house prices across segments. The price formula (19) shows how the price is determined as the difference between a "fundamental" price $v(h)/r$ and a *liquidity discount* that capitalizes the present value of search and transaction costs. The latter are segment-specific: the popularity and instability properties of a segment derived above affect both the average time on the market (and hence search costs) as well as turnover (and hence the frequency at which transaction costs arise). What is as yet missing to evaluate the formula is a measure of fundamental value.

To estimate both fundamental value we can use the cross section of median prices together with our estimation results. In particular, for each cross section of prices and parameter vector, we can back out from (19) the vector of mean utility values $v(h)$ such that the model exactly matches the cross section of transaction prices. We postulate a real interest rate of 2% and set the transaction cost such that the average sale costs 6% of the resale value of the house, a standard number in the literature.

The results are summarized in Figure 7. The left hand panel plots median price against the liquidity discount, stated as a percentage of price. The right hand panel shows the geographic distribution of liquidity discounts. There are two notable results here. First, liquidity discounts are large – they can be up to 40% of the sales prices. Second, liquidity discounts differ widely by segment, oftentimes within the same zip code. In poor segments with high volume and high time on market, both search and transaction costs are high; as a result, prices are significantly lower than they would be in a frictionless market. In rich segments discounts are still significant, but they are considerably smaller.

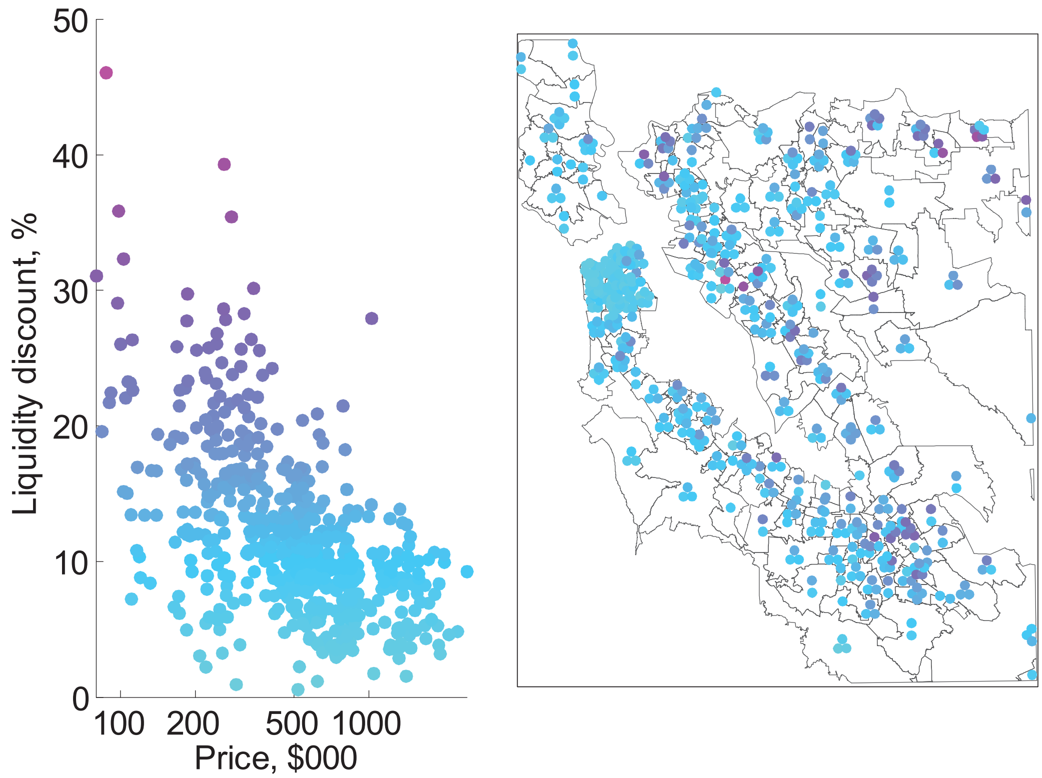


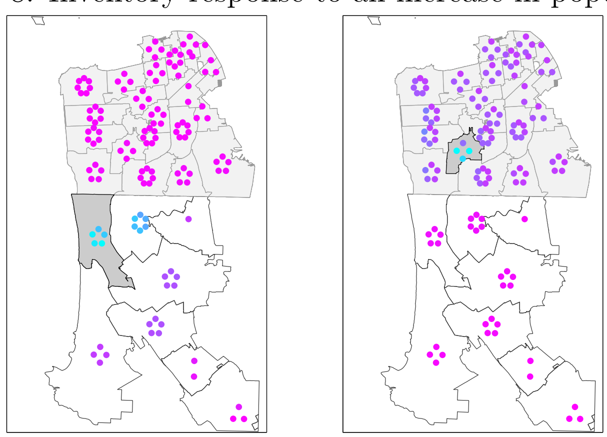
Figure 7: Liquidity Discounts. Left panel: mean zip code price vs zip code liquidity discount as a percentage of mean price; color coding reflects liquidity discount. Right panel: zip codes colored by same code as in right panel.

6.3 Comparative statics

Figure 8 shows how the steady state equilibrium changes when a particular zipcode becomes more popular. Formally, we recompute the steady state using the same parameters as above, but we increase $\mu^{Theta}(h)$ in one particular segment h. The panels are maps of only the tip of the San Francisco peninsula. The lines in the map indicate zipcode boundaries. Each zipcode contains several dots, which represents segments within zipcodes. The dots are aligned so that the cheapest segment sits at twelve o'clock and segment house prices increase clockwise. Both panels assume that the hypothetical change in popularity occurs in the gray shaded zipcode. In the left panel, the grey shaded zipcode is 94015 Daly City. In the right panel, the grey zipcode is 94127 San Francisco West Portal. The colors indicate changes in inventory in the segments, ranging from blue (drop) to pink (increase).

The result is that an increase in the popularity of 94015 lowers inventory in Daly City itself, and has spillovers on its neighboring zipcode to the east. Essentially nothing happens to the north, in the city of San Francisco itself. In contrast, an increase in popularity of 94127 lowers inventory in West Portal and has spillover effects on inventory all over San Francisco. This is because a large share of searchers scan all these segments jointly. These results show that search patterns introduce asymmetries in the transmission of shocks.

Figure 8: Inventory response to an increase in popularity.



Note: The figure shows responses in inventory to an increase in popularity of zipcode 94015 Daly City in the left panel and 94127 San Francisco West Portal in the right panel.

References

Bayer, Patrick, Fernando Ferreira, and Robert McMillan, "A unified framework for measuring preferences for schools and neighborhoods," *Journal of Political Economy*, 2007, *115* (4), 588–638.

Burnside, Craig, Martin Eichenbaum, and Sergio Rebelo, "Understanding booms and busts in housing markets," 2011.

Caplin, Andrew and John Leahy, "Trading frictions and house price dynamics," *Journal of Money, Credit and Banking*, 2011, *43* (s2), 283–303.

Favilukis, Jack, Sydney C Ludvigson, and Stijn Van Nieuwerburgh, "The macroeconomic effects of housing wealth, housing finance, and limited risk-sharing in general equilibrium," 2010.

Floetotto, Max and Johannes Stroebel, "Government Intervention in the Housing Market: Who Wins, Who Loses?," *University of Chicago*, 2012.

Garmaise, Mark J and Tobias J Moskowitz, "Confronting information asymmetries: Evidence from real estate markets," *Review of Financial Studies*, 2004, *17* (2), 405–437.

Genesove, David and Lu Han, "Search and matching in the housing market," *Journal of Urban Economics*, 2012, *72* (1), 31–45.

Glaeser, Edward L. and Joseph Gyourko, "The Impact of Building Restrictions on Housing Affordability," *FRB New York: Economic Policy Review*, 2003, *9* (2), 21–39.

Glover, Andrew, Jonathan Heathcote, Dirk Krueger, and José-Víctor Ríos-Rull, "Intergenerational redistribution in the great recession," 2011.

Goodman, Allen C and Thomas G Thibodeau, "Housing market segmentation," *Journal of housing economics*, 1998, *7* (2), 121–143.

Han, Lu and W Strange, "What is the Role of the Asking Price for a House," *Working Paper*, 2013.

Islam, Kazi Saiful and Yasushi Asami, "Housing market segmentation: A review," *Review of Urban & Regional Development Studies*, 2009, *21* (2-3), 93–109.

Krainer, John, "A theory of liquidity in residential real estate markets," *Journal of Urban Economics*, 2001, *49* (1), 32–53.

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

Leishman, Chris, "House building and product differentiation: An hedonic price approach," *Journal of Housing and the Built Environment*, 2001, *16* (2), 131–152.

Levitt, Steven D and Chad Syverson, "Market distortions when agents are better informed: The value of information in real estate transactions," *The Review of Economics and Statistics*, 2008, *90* (4), 599–611.

Manning, Alan and Barbara Petrongolo, "How Local are Labor Markets? Evidence from a Spatial Job Search Model," 2011.

Mian, Atif R and Amir Sufi, "House prices, home equity-based borrowing, and the US household leverage crisis," 2009.

National Association of Realtors, "2011 Profile of Home Buyers and Sellers," Technical Report 2011.

_ , "Digital House Hunt," Technical Report 2013.

Ngai, L Rachel and Silvana Tenreyro, "Hot and cold seasons in the housing market," 2009.

Nieuwerburgh, Stijn Van and Pierre-Olivier Weill, "Why has house price dispersion gone up?," *The Review of Economic Studies*, 2010, *77* (4), 1567–1606.

Novy-Marx, Robert, "Hot and cold markets," *Real Estate Economics*, 2009, *37* (1), 1–22.

Piazzesi, Monika and Martin Schneider, "Momentum Traders in the Housing Market: Survey Evidence and a Search Model," *The American Economic Review*, 2009, *99* (2), 406–411.

_ **and** _ , "Inflation and the price of real assets," 2012.

Poterba, James M, David N Weil, and Robert Shiller, "House price dynamics: The role of tax policy and demography," *Brookings Papers on Economic Activity*, 1991, *1991* (2), 143–203.

Stroebel, Johannes, "The impact of asymmetric information about collateral values in mortgage lending," *University of Chicago*, 2012.

Trulia, "The Largest Audience of Real Estate Consumers," Technical Report 2013.

Wheaton, William C, "Vacancy, search, and prices in a housing market matching model," *Journal of Political Economy*, 1990, pp. 1270–1292.

A Data Appendix

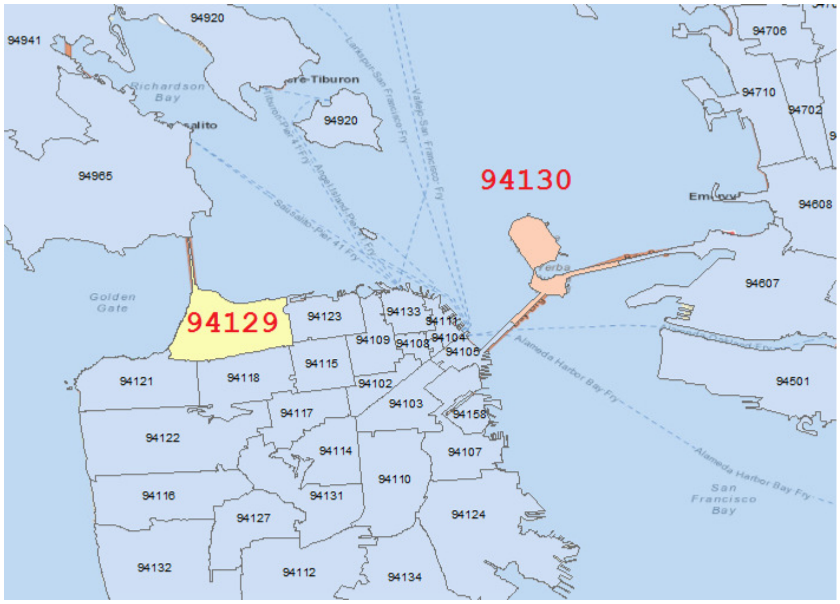
A.1 Constructing Contiguity Measures

To analyze whether all zip codes are contiguous, one challenge is provided by the San Francisco Bay. The location of this body of water means that two zip codes with non-adjacent borders should sometimes be considered as contiguous, since they are connected by a bridge such as the Golden Gate Bridge. Figure A.1 illustrates this. Zip codes 94129 and 94965 should be considered contiguous, since they can be traveled between via the Golden Gate Bridge. To take the connectivity provided by bridges into account, we manually adjust the ESRI shape files to link zip codes on either side of the Golden Gate Bridge, the Bay Bridge, the Richmond-San Rafael Bridge, the Dumbarton Bridge and the San Mateo Bridge. In addition, there is a further complication in that the bridgehead locations are sometimes in zip codes that have essentially no housing stock, and are thus never selected in search queries. For example, 94129 primarily covers the Presidio, a recreational park, that contains only 271 housing units. Similarly, 94130 covers Treasure Island in the middle of the SF Bay, again, with only a small housing stock. These zip codes are very rarely selected by search queries, which would suggest, for example, 94105 and 94607 would not be connected. This challenge is addressed by manually merging zip codes 94129 and 94130 with the Golden Gate and Bay bridge respectively. This ensure, for example, that 94118 and 94955 are connected even if 94129 was not selected.

In the following we provide examples of contiguous and non-contiguous search sets. The top left panel of Figure A.3 shows all the zip codes covered by a searcher that searched for homes in Berkeley, Fremont, Hayward, Oakland and San Leandro. This is a relatively broad set, covering most of the East Bay. The top right panel shows a contiguous set of jointly searched zip codes, with connectivity derived through the Golden Gate Bridge. The searcher queried homes in cities north of the Golden Gate Bridge (Corte Madera, Larkspur, Mill Valley, Ross, Kentfield, San Anselmo, Sausalito and Tiburon), but also added zip codes 94123 and 94115. The bottom left panel shows the zip codes covered by a searcher that selected a number of San Francisco neighborhoods. The final contiguous search set (bottom right panel) was generated by a searcher that selected a significant number of South Bay cities.[12] These are all locations with reasonable commuting distance to the tech jobs in the Silicon Valley. Notice how the addition of Newark adds zip code 94550 on the East Bay, which

[12] Atherton, Belmont, Burlingame, El Granada, Emerald Hills, Foster City, Half Moon Bay, Hillsborough, La Honda, Los Altos Hills, Los Altos, Menlo Park, Millbrae, Mountain View, Newark, Palo Alto, Portola Valley, Redwood City, San Carlos, San Mateo, Sunnyvale, Woodside.

Figure A.1: Bridge Adjustments - Contiguity Analysis



Note: This figure shows how we deal with bridges in the Bay Area for the contiguity analysis.

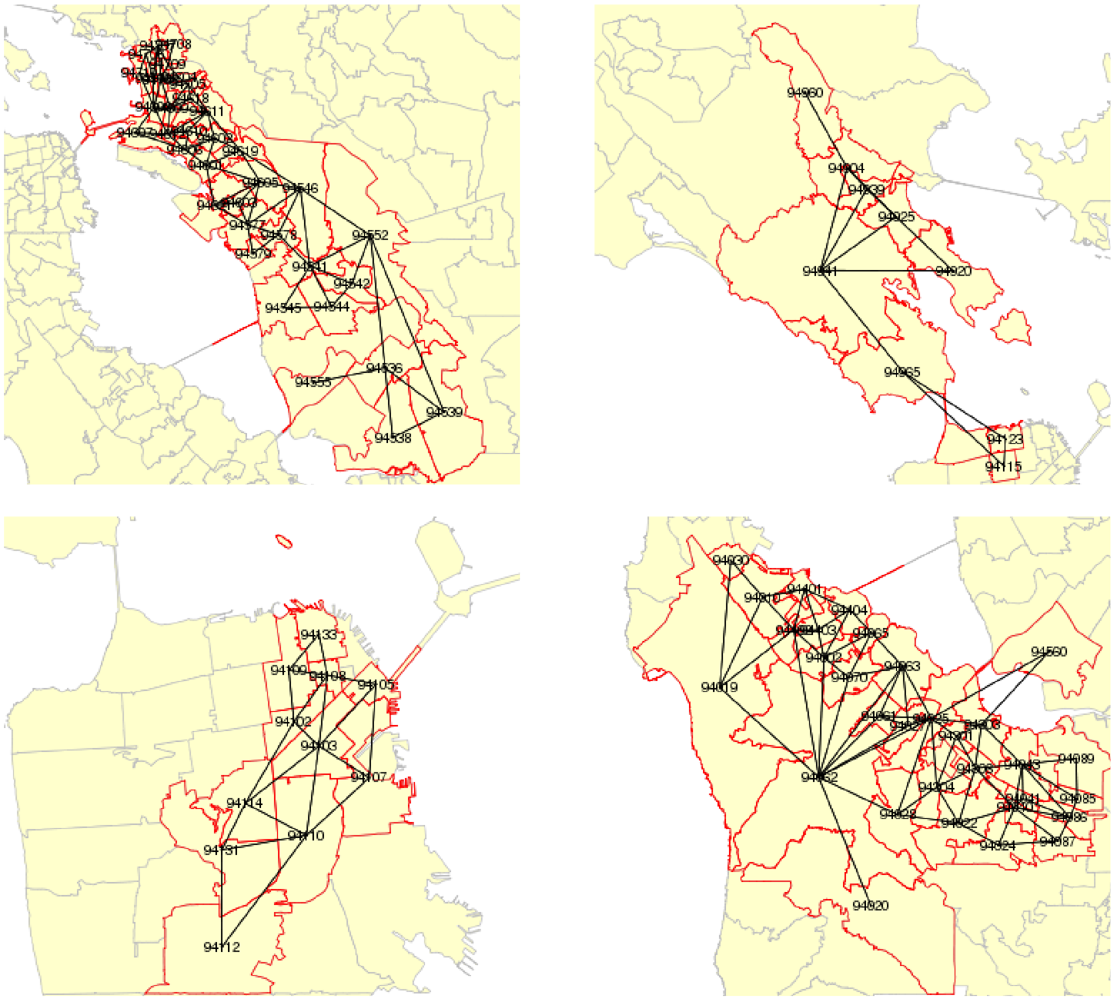
is connected to the South Bay via the Dumbarton Bridge. Not all email alerts generate sets of zip codes that are contiguous. In Figure A.3 we show four actual non-contiguous search sets. The top left panel shows the zip codes covered by a searcher that selects the cities of Cupertino, Fremont, Los Gatos, Novato, Petaluma and San Rafael. This generates three contiguous set of zip codes, rather than one large, contiguous set. The zip codes in the bottom right belong to a searcher that selected zip code 94109 and the neighborhoods Nob Hill, Noe Valley and Pacific Heights. Again, this selection generates more than one set of contiguous zip codes.

A.2 Segment Construction

This section describes the process of arriving at the set of 576 distinct housing market segments for the San Francisco Bay Area. As before, we select the geographic dimension of segments to be a zip code. Since we will compute average price, volume, time on market and inventory for each segment, we restrict ourselves to zip codes with at least 800 armslength housing transactions between 1994 and 2012. This leaves us with 191 zip codes with sufficient observations to construct these measures.

We next consider how to further split these zip codes into segments based on a qual-

Figure A.2: Sample Contiguous Queries

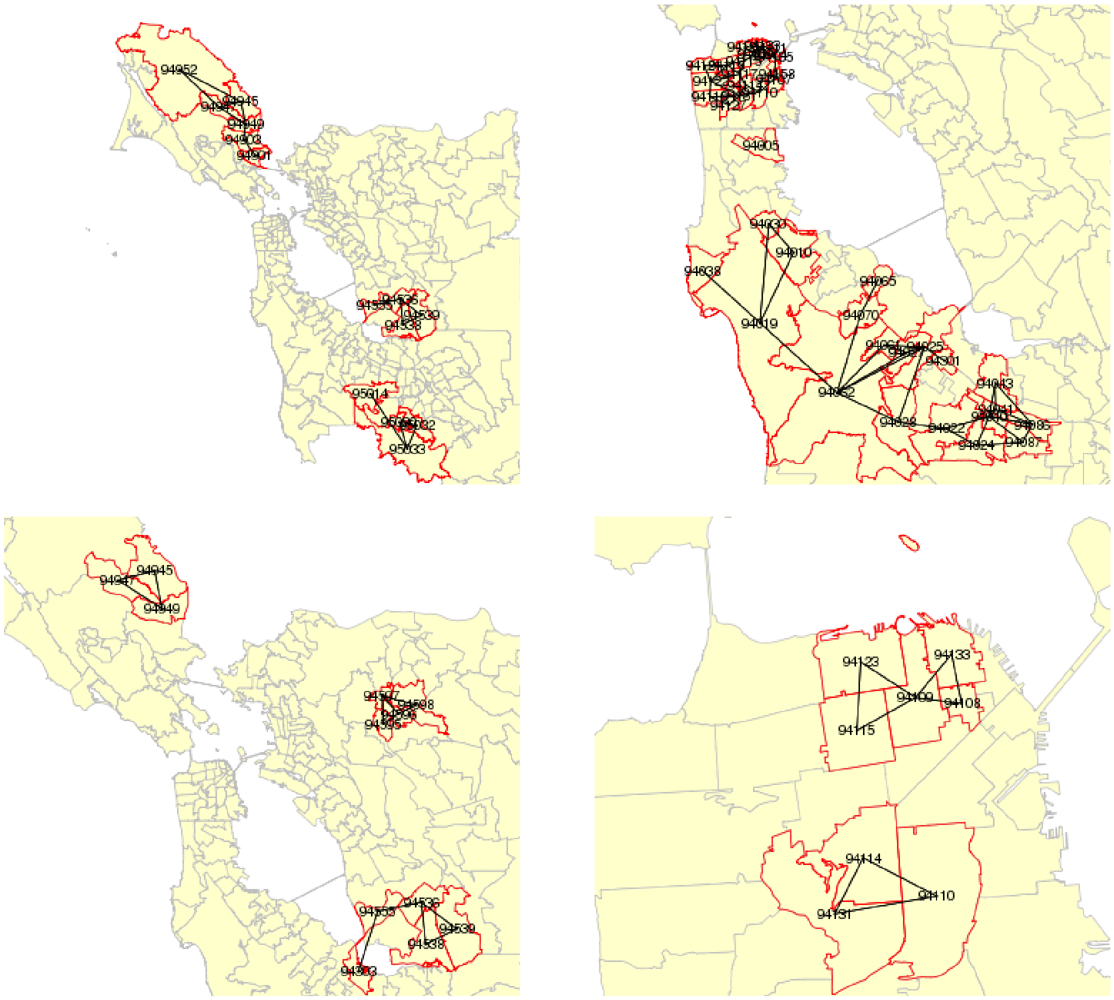


Note: This figure shows a sample of contiguous search sets. The zip codes selected by the searcher are circled in red. Zip code centroids of contiguous zip codes are connected.

ity (price) and size dimension. Importantly, we will need to observe the total housing stock in each segment in order to appropriately normalize moments such as turnover and inventory. The residential assessment records do contain information on the universe of the housing stock. However, as a result of Proposition 13, the assessed property values in California do not correspond to true market value, and it is thus not adequate to divide the total zip code housing stock into different price segments based on this assessed value.[13] To measure the housing stock in different price segments we use the U.S. Census Bureau's 2011 American Community Survey 5-year estimates, which report the total number of owner-occupied housing units per zip code for a number of price bins. We com-

[13]Allocating homes that we observe transacting into segments based on value is much easier, since this can be done on the basis of the actual transaction value, which is reported in the deeds records.

Figure A.3: Sample Non-Contiguous Queries



Note: This figure shows a sample of non-contiguous search sets. The zip codes selected by the searcher are circled in red. Zip code centroids of contiguous zip codes are connected.

bine a number of these bins to construct the total number of housing units in each of the following price bins: $< \$200k, \$200k-\$300k, \$300k-\$400k, \$400k-\$500k, \$500k-\$750k, \$750k-\$1m, > \$1m$. These bins provide the basis for selecting price cut-offs to delineate quality segments within a zip code. One complication is that the price boundaries are reported as an average for the sample years 2006-2010. Since we want segment price cut-offs to capture within zip code time-invariant quality segments, we need to adjust for average market price changes of the same-quality house over time. To do this, we adjust all prices and price boundaries to correspond to 2010 house prices.[14]

[14] This is necessary, because the Census Bureau only adjusts the reported values for multi-year survey periods by CPI inflation, not by asset price changes. This means that a \$100,000 house surveyed in 2006 will be of different quality to a \$100,000 house surveyed in 2010. We choose the price that a particular

Not all zip codes have an equal distribution of houses in each price (quality) bin. For example, Palo Alto has very few homes valued at less than $200,000, while Fremont has very few million-dollar homes. Since we want to avoid cutting a zip code into too many quality segments with essentially no housing stock to allow us measure segment-specific moments such as time on market, we next determine a set of three price cut-offs for each zip code by which to split that zip code. To determine which of the seven census price bin cut-offs should constitute segment cut-offs, we use information from the search queries. This proceeds in two steps: First we change the price parameters set in the email alerts to account for the fact that we observe queries from the entire 2006 - 2012 period. This adjusts the price parameters in each alert by the market price movements of homes in that zip code between the time the query was set and the year 2010.[15] Second, we determine which set of three ACS cutoffs is most similar to the distribution of actual price boundaries selected in search queries that cover a particular zip code. For each possible combination of three (adjusted) price cut-offs from the list of ACS cut-offs, we calculate for every email alert the minimum of the absolute distance from each of the (adjusted) search alert price restrictions to the closest cut-off.[16] We select the set of segment price cut-offs that minimizes the average of this value across all queries that cover a particular zip code. This ensures, for example, that if there are many queries that include a high limit such as $1 million, $1 million is likely to also be a segment boundary.

To determine the total housing stock in each price by zip code segment, one additional adjustment is necessary. Since the ACS reports the total number of owner-occupied housing units, while we also observe market activity for non owner-occupied units, we need to adjust the ACS-reported housing stock for each price bin by the corresponding homeownership rate. To do this, we use data from all observed armslength ownership-changing transactions

house would fetch in 2010 as our measure of that home's underlying quality. To transform the housing stock by price bin reported in the ACS into a housing stock by 2010 "quality" segment, we first construct zip code specific annual repeat sales price indices. This allows us to find the average house price changes by zip code for each year between 2006 and 2010 to the year 2010. We then calculate the average of these 5 price changes to determine the factor by which to adjust the boundaries for the price bins provided in the ACS data. Adjusting price boundaries by a zip code price index that looks at changes in median prices over time generates very similar adjustments.

[15] This ensures that the homes selected by each query correspond to our 2010 quality segment definition. Imagine that prices fell by 50% on average between 2006 and 2010. This adjustment means that a query set in 2006 that restricts price to be between $500,000 and $800,000 will search for homes in the same quality segment as a query set in 2010 that restricts price to a $250,000 - $400,000 range.

[16] For example, imagine testings how good the the boundaries 100k, 300k and 1m fit for a particular zip code. A query with an upper bound of 500k has the closest absolute distance to a cut-off of $\min\{|500 - 100|, |500 - 300|, |500 - 1000|\} = 200$. A query with an upper bound of 750k has the closest absolute distance to a cut-off of 250. A query with a lower bound of 300k and an upper bound of 600k has the closest absolute distance to a cut-off of 0. For each possible set of price cut-offs, we calculate for every query the smallest absolute distance of a query limit to a cut-off, and then find the average across all search alerts.

between 1994 and 2010 as reported in our deeds records. We first adjust the observed transaction price with the zip code level repeat sales price index, to assign each house for which we observe a transaction to one of our 2010 price (quality) bins. For each of these properties we also observe from the assessor data whether they were owner-occupied in 2010. This allows us to calculate the average homeonwerhisp rate for each price segment within a zip code, and adjust the ACS-reported stock accordingly.[17]

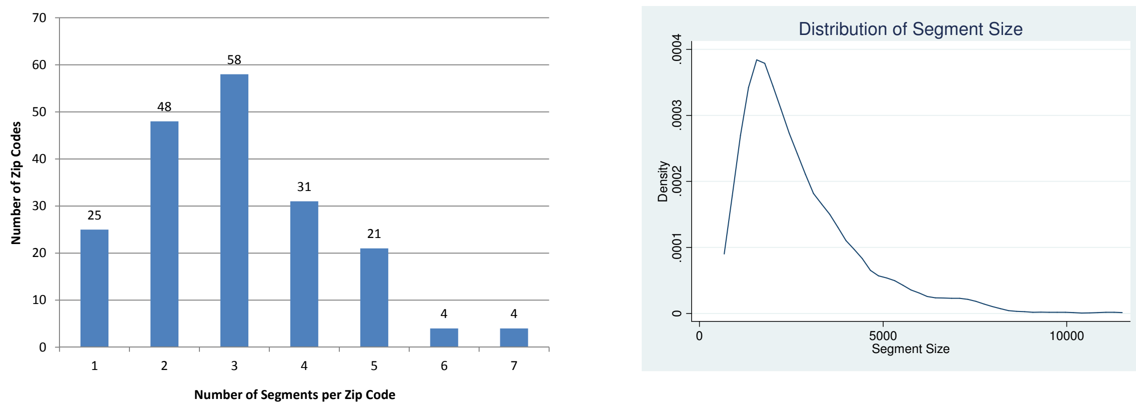
The other search dimension regularly specified in the email alerts, and that we hence wanted to incorporate in our segment definition, is the number of bathrooms as a measure of the size of a house conditional on its location and quality. Since section **??** showed that the vast majority of constraints on the number of bathrooms selected homes with either more or fewer than two bathrooms, we further divide each zip code by price bucket group into two segments: homes with less than two bathrooms, and homes with at least two bathrooms. Unfortunately the ACS does not provide a cross-tabulation of the housing stock by home value and the number of bathrooms. To split the housing stock in each price and zip code segment into the two groups by home size, we apply a similar method as above to control for homeownership rate. We use the zip code level repeat sales price index to assign each home transacted between 1994 and 2010 to a 2010 price (quality) bin. For these homes we observe the number of bathrooms from the assessor records. This allows us to calculate the average number of bathrooms for transacted homes in each zip code by price segment. We use this share to split the total housing stock in those segments into two bathroom size groups.

The approach described above splits each zip code into eight initial segments along three price cutoffs and one size cutoff. For each of these segments, we have an estimate of the total housing stock. Since we need to measure specific moments such as the average time on market with some precision, we need to ensure that each segment is sufficiently large, and has a housing stock of at least 1,000 units. If this is not the case the segment is merged with a neighboring segment until all remaining segments have a housing stock of sufficient size. For price segments where either of the two size subsegments have a stock of less than 1,000, we merge the two size segments. We then begin with the lowest price segment, see whether it has a stock of less than 1,000, and merge it with the next higher price segment. This procedure generates 576 segments. Figure A.4 shows how many segments each zip code is being split into. 25 zip codes are not split up further into segments. 48 zip codes are split into two segments, 58 zip codes are split into 3 segments. 418 segments only have a geography

[17]For example, the 2010 adjusted segment price cutoffs for zip code 94002 are $379,079, $710,775 and $947,699. This splits the zip code into 4 price buckets. The homeownerhip rate is much higher in the higher bucket (95%) than in the lowest bucket (65%). This shows the need to have a price-bucket specific adjustment for the homeownership rate to arrive at the correct segment housing stock.

and price limitation, and include homes of all sizes falling into those price categories. The right panel of figure A.4 shows the distribution of housing stock across segments. On average, segments have a stock of 2,717, with a median value of 2,271. The largest segment has a housing stock of 11,178.

Figure A.4: Segment Overview



A.3 Assigning segments to search alerts

As a next step, to analyze segments in terms of their search clientele we need to analyze each query and determine which segments are covered by that query. In section **??** we describe how we determine which zip codes are covered by each query. In this section we describe how we deal with the price and bathroom dimensions to determine the set of segments covered by each query. The challenge is that price ranges selected by queries will usually not overlap perfectly with the price cutoffs of the individual segments. For those queries that specify a price dimension, we assign a query to cover a particular segmented in one of three cases:

1. When the query completely covers the segment (that is, when the query lower bound is below the segment cutoff and the query upper bound is above the segment cutoff).
2. When the segment is open-ended (e.g. \$1 million +), and the upper bound of the query exceeds the lower bound (in this case, all queries with an upper bound in excess of \$1 million).
3. For queries that partially cover a non-open ended segment, we determine the share of the segment price range covered by the query. For example, for a segment \$300k

- $500k, the query 0-$250k covers 25%, the query $300k - $700k covers 50% of the segment. We assign all queries that cover at least 50% of the price range of a segment to cover that segment.

To deal with the bathroom dimensions, we let a query cover a segment unless it is explicitly excluded. For example, queries that want at least two bathrooms will not cover the < 2 bathroom segments and vice versa.

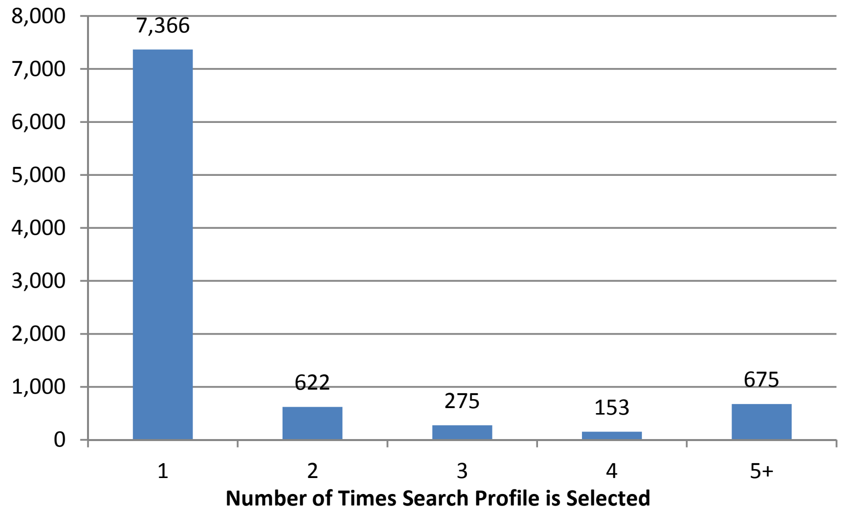
We argued above that it was hard to pool across different alerts set by the same individual, since alerts differed along a number of key dimensions including geography, home size and home quality. The housing market segments constructed above allow us to pool all segments selected by the same searcher. In particular, we begin by determining the subset of the segments that are covered by each individual search alert. This process described in more detail in Appendix A.3. After pooling all segments covered by at least one email alert set by each searcher, we arrive at a total of 9,091 unique search profiles. A total of 7,366 search profiles are selected by only a single user. Another 622 search profiles are selected twice. A total of 338 search profiles are selected more than 10 times each, with the two most commonly selected search profile showing up 1,017 and 416 times. Figure A.5 shows the distribution of how often each search profile is selected. We also analyze the total housing stock covered by each searcher. We find that the average (median) searcher covers a total stock of 57,483 (33,807) housing units.

A.4 Construction of Segment Moments

Our model links the characteristics of search patterns to segment specific moments such as price, volume, time on market and inventory. In this section we describe how we construct these moments at the segment level. We begin by identifying a set of armslength transactions, which 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 (and hence the quality) 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, for example, intra-family transfers. We drop all observations that are not a Main Deed or only transfer partial interest in a property (see Stroebel (2012) for details on this process of identifying armslength transactions).

We can then calculate the total number of transactions per segment between 2008 and

Figure A.5: Number of Searchers per Unique Profile



Note: This figure shows how often each of the 9,090 individual search profiles is selected.

2011, and and use this to construct annual volume averages. In order to allocate houses to particular segments, we adjust transaction prices for houses sold in years other than 2010 by the same price index we used to adjust listing price boundaries (see appendix A.2). We arrive at our measure of Volume Share by dividing the annual transaction volume by the segment housing stock. Inventory levels are first constructed at the monthly level. To do this we use the dataset on all home listings on Trulia.com, beginning in January 2006. We assign each listed property to a segment using its location, size and adjusted listing price. Each month we add all newly listed properties in a segment to the inventory observed in the previous month. In addition, all listings that result in a sale as observed in the deeds data get removed from the inventory. We then construct the average of these monthly inventory levels for the period 2008-2011.[18] Inventory Shares are determined by dividing inventory levels in a segment by the total housing stock in the segment. We also construct a second inventory measure, "cold inventory", which is the fraction of the housing stock that is listed, and has been on the market for more than 30 days. In constructing inventory measures,

[18]Many properties that are sold as REO resales (i.e. mortgage lenders selling properties that are acquired through a foreclosure) do not get listed through an MLS, and hence do not show up in Trulia's listing database. We thus need to construct REO resale inventory in a different way. In the deeds data we observe when a foreclosure occurs, since a foreclosure involves an ownership transfer to the bank. For those REO properties that do show up in the listings data, we calculate the median time between the foreclosure and the listing, which is 20 days. We henceforth add every foreclosed property to the inventory 20 days after we observe the foreclosure, and remove it when we observe an REO resale.

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 for which we do not observe a sale from the inventory 270 days after the initial listings (as a reference point, note that the 90th percentile of time on market for houses that do eventually get sold is about 190 days). Of course, if a house sells that was listed for more than 270 days, we record that as a sale. A second challenge in measuring inventory levels arises from the fact that Trulia's coverage of listings is not 100% (for example, there are properties that are "for sale by owner" and hence do not show up in MLS feeds), and has increased over the time period we consider. However, we do have the universe of all transactions - this allows to construct, for every segment, a measure of how many homes we observe transacting over the sample period without having previously observed a listing. We can then scale our measure of inventory by the "share sale without listing" measure for that particular segment.

To calculate the average time on market, we match home listings in the listings database with final transactions from the deeds database.[19] We find segment-specific measures of time on market by averaging the time on market across all transactions that sold between 2008 and 2011. We also calculate the average time on market conditional on the time on market exceeding 30 days, which will be used in our stock-flow model of the matching process. Finally, we calculate the share of "hot sales", i.e. transactions that are recorded within 30 days of the initial listing.

A.5 Stability of search patterns

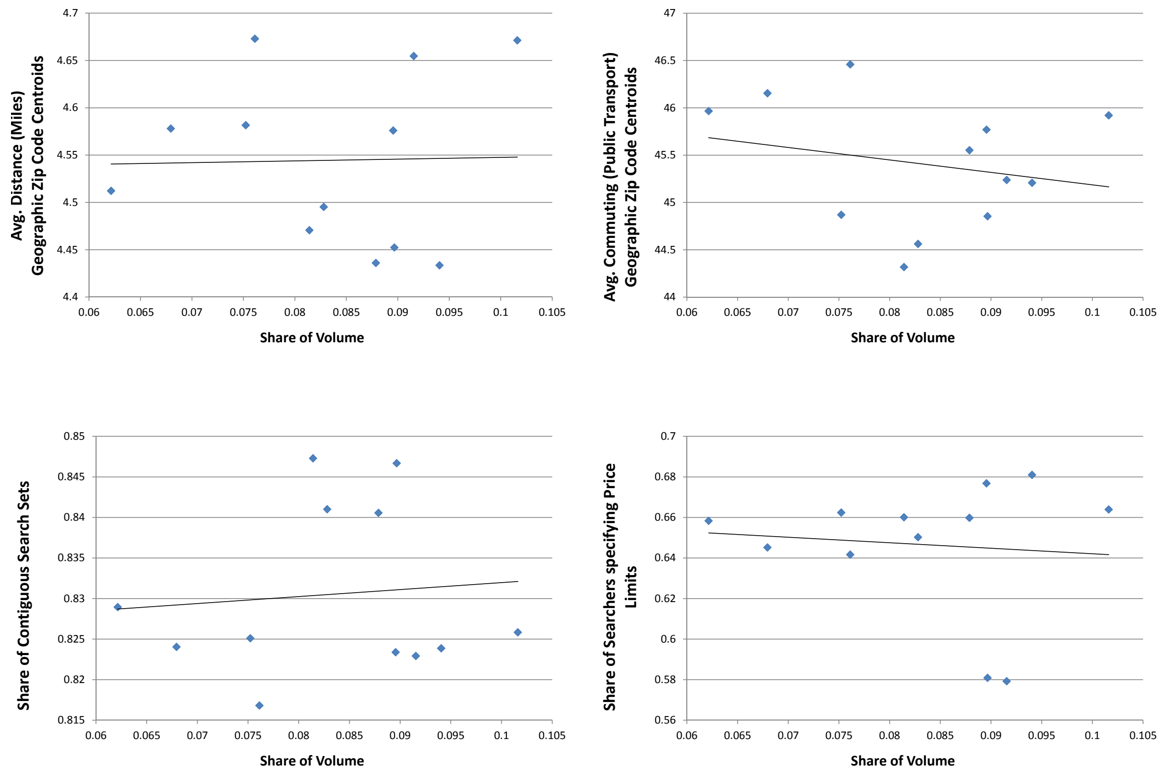
Our model below will interpret search ranges as a feature of buyer preferences. It is then interesting to ask whether ranges are invariant to changes in market conditions. In particular, do searchers change narrow the range of houses they consider when market activity is higher? To test this hypothesis, we exploit seasonal variation in housing market activity: more houses typically trade in summer as compared to winter.

Each panel of Figure A.6 shows a scatter plot of the share of total volume in a month, and a particular search dimension. We include (clockwise from top-left) the average distance between geographic zip code centroids, the average commuting time by public transport between geographic zip code centroids, the share of searches that yield contiguous search sets and the share of searches that include a price dimension. The takeaway here is that none

[19] In the very few instances when the listing price and the final sales price would suggest a different segment membership for a particular house – i.e. cases where the house is close to being at a segment boundary and sells for a price different to the listing price, we allocate the house to the segment suggested by the sales price, not the listing price.

of the search dimensions exhibit meaningful seasonality, consistent with an interpretation of search parameters as time-invariant sets driven by preferences.

Figure A.6: Non-Seasonality of Search Parameters



Note: This figure shows the correlation of search parameters with the share of annual volume in a particular month on the horizonal axis.