Segmented Housing Search*

Monika Piazzesi
Stanford & NBER

Martin Schneider
Stanford & NBER

Johannes Stroebel
NYU & NBER

Abstract

We study housing markets with multiple segments searched by heterogeneous clienteles. In the San Francisco Bay Area, search activity and inventory covary negatively across cities, but positively across market segments within cities. A quantitative search model shows how the endogenous flow of broad searchers to high-inventory segments within their search ranges induces a positive relationship between inventory and search activity across segments with a large common clientele. The prevalence of broad searchers shapes the response of housing markets to localized supply and demand shocks. Broad searchers help spread shocks across many segments and reduce their effect on local market activity.

1 Introduction

Housing markets are search markets. As a result, the typical description of a housing market includes not only price and trading volume, but also inventory available for sale – how many home sellers are searching for potential home buyers. In policy discussions, low inventory is often identified with a "housing shortage" in which housing demand outstrips housing supply. What is typically missing, however, is data on the number of potential buyers that are actually looking for a house in a given market. In addition, we usually do not observe information on the search behavior of these buyers.

Two key questions thus remain unanswered. First, is inventory really sufficient to summarize the state of the housing market? In other words, does the housing market resemble the labor market where low vacancy rates usually go along with high unemployment, as captured by a downward-sloping Beveridge curve? And second, how integrated are different housing market segments? Do homebuyers usually focus their search activity on a few neighborhoods or do they consider entire metro areas? Answers to both questions are crucial to build quantitative models of housing search and make informed policy decisions.

*This version: March 16, 2019. Author contact information: piazzesi@stanford.edu, schneidr@stanford.edu, johannes.stroebel@nyu.edu. We thank Ed Glaeser, Bob Hall, Brian Higgins, John Leahy, Giuseppe Moscarini, Robert Shimer, Vincent Sterk, Silvana Tenreyro, four anonymous referees, and conference and seminar participants at the AEA meetings, Arizona State, Berkeley, Chicago, Harvard, LBS, Mannheim, Michigan, Minneapolis Fed, Minnesota, MIT, NBER Summer Institute, Northwestern, NYU, Philadelphia Fed, SITE, SED, Stanford, Tsinghua, UCLA, USC, and Yale. We also thank Trulia for providing data. This research is supported by an NSF grant.

This paper uses a novel data set on the search behavior of home buyers to document stylized facts about housing search and to inform a quantitative housing search model. We show that the housing market in the San Francisco Bay Area is a collection of many small market segments that differ by geography and property characteristics. In every segment, there is local demand from narrow searchers who look for houses in only a handful of similar segments. At the same time, there are broader searchers who connect many segments to create integrated areas. For example, cities such as San Francisco and San Jose are broadly searched by many potential home buyers. We also document that the cross-sectional Beveridge curve depends on the level of aggregation. Across areas that do not share many common searchers, such as cities, the Beveridge curve slopes down as low-inventory cities experience high buyer search activity. In contrast, across housing market segments within cities, which share many common searchers, the Beveridge curve slopes up: segments with lower inventory draw interest from *fewer* searchers.

To understand these patterns and explore their implications, we build a model of housing search with three new features motivated by our empirical findings: (i) the housing market is a collection of many segments, (ii) agents differ in their search ranges, the set of segments they consider living in, and (iii) broad searchers flow into segments within their search range in proportion to segment inventory. The model explains how equilibrium housing market outcomes are shaped by the interaction of broad and narrow searchers.

In particular, different cities share few common searchers but differ in their popularity, the total number of agents who would consider living there. In more popular cities, higher demand means that a larger number of searchers quickly buys any house that comes on the market, so inventory is lower. As a result, the Beveridge curve slopes down across cities.

In contrast, segments within cities share many broad searchers but differ in their stability, the rate at which houses come on the market. In less stable segments, more supply means that higher inventory attracts more broad searchers who "crowd out" narrow searchers. This increases the total number of searchers per house. Less stable segments thus have more inventory and more search activity. As a result, the Beveridge curve slopes up across segments within cities, or more generally, across any set of segments with many common searchers.

The model also shows that buyer search data are essential for policy analysis. For example, to forecast market responses to new construction, it is important to avoid both too much and too little aggregation. Consider a researcher who is interested in the effect of construction on the inventory in a particular segment within a city. If she assumes incorrectly that the city is not at least partially integrated, she will miss the implications of competition between broad and narrow searchers. Indeed, building in a low-inventory segment may mostly attract broad searchers and thus neither increase inventory nor benefit narrow searchers in the segment. Similarly, researchers interested in metro area aggregates should avoid treating metro areas as fully integrated homogeneous markets. Indeed, we show that with partial integration, the overall change in metro area inventory depends on the popularity and stability of the exact segment where construction takes place.

We infer search ranges of potential home buyers from online housing search: on the popular real estate website trulia.com, home searchers can set an alert that triggers an email whenever a house with their desired characteristics comes on the market. We observe the search parameters in a large sample of such email alerts. Housing search occurs predominantly along three dimensions: geography, price, and house size as captured by the number of bathrooms (which are specified more often in searches than the number of bedrooms or the square footage.) To relate search activity to other market activity, we divide the San Francisco Bay Area into 564 distinct housing market segments along the dimensions suggested by the observed search ranges. We then express the search ranges as subsets of the set of all segments, and measure search activity at the segment level by the number of searchers per house. We also measure the cross section of turnover and inventory at the segment level from deeds records, assessment data, and "for sale" listings.

Our search model assumes random matching as well as a fixed number of houses and agents. Houses are located in one of many segments, each with its own matching function. Moving shocks induce agents to sell their current house (at a cost) and search for another house. Heterogeneous agent types are identified by their search ranges – subsets of the set of all segments that they would consider living in, as in our data. While matching is random, agents are more likely to match in those segments within their search ranges where inventory is higher. This central assumption is directly supported by the patterns in our search data. We also show that it can be derived from more primitive assumptions in a variety of settings. Prices reflect the present value of housing services less a frictional discount due to search and transaction costs.

Our quantitative analysis proceeds in two steps. We begin by focusing on quantities. We show that the distribution of preferences, moving shocks, and a measure of matching frictions can be identified from cross-sectional moments of turnover, inventory, and search activity. This identification result is independent of the details of price bargaining and the matching function. We estimate that distribution and derive summary statistics of supply and demand conditions at the segment level. We define *stable* segments as those with less frequent moving shocks, and *popular* segments as those with a larger clientele (i.e., more individuals that are potentially interested in living there), where broader searchers count less towards popularity in any segment they cover. Our estimates of segment-level popularity are higher in areas with better schools, better restaurants, and better weather. As an over-identifying restriction test of our model, we show that population flows between segments implied by our estimates are consistent with observed moves in the data.

The estimated distribution of preferences explains why Beveridge curves differ by the level of aggregation. Bay Area cities are searched by fairly distinct clienteles and differ in their popularity. Variation in city popularity generates a downward-sloping housing Beveridge curve across cities: in more popular cities, any house that comes on the market is sold more quickly, leaving less inventory in equilibrium. More popular cities are also estimated to be more stable, which contributes to the downward slope of the Beveridge curve and helps explain why inventory and turnover comove positively across cities. At the same time, segments within cities are typically integrated by broad searchers who look to buy in the entire city, or at least all of its less expensive segments. These broad searchers are attracted to unstable segments where inventory regularly comes on the market. In those market segments, they crowd out any local narrow searchers, which generates a higher number of searchers per house. In markets where broad and narrow searchers interact, variation in segment stability can therefore generate an upward-sloping Beveridge curve across segments.

To quantify the importance of the new mechanism of competition between broad and narrow searcher, we compare our estimated model to a misspecified benchmark that assumes all searchers narrowly target just one segment. This benchmark exercise follows the common approach in the literature of using data on turnover, inventory, and the time it takes to find a house to pin down parameters of a search model. In this benchmark without broad searchers, all observed search activity is attributed to local demand from narrow searchers. We show that this approach leads researchers comparing segments within a city to infer that unstable segments with high inventory are substantially more popular than they actually are, since the search activity by broad searchers attracted to the high inventory is indistinguishable from search activity by narrow searchers with a particular and targeted interest in that segment.

The second part of our quantitative analysis assumes bargaining over price as well as segment-specific Cobb-Douglas matching functions. We use the parameter estimates from the full model to study price formation and policy counterfactuals. We first infer frictional price discounts across segments, which capture the capitalized value of trading frictions faced by current and future buyers. These discounts are quantitatively large, between 10 percent and 40 percent of the frictionless house value (defined as the present discounted value of future housing services). Frictional discounts are larger in less stable segments, where houses transact more often. Frictional discounts are also larger in segments where houses take a long time to sell, for example because these segments do not attract many broad searchers. Quantitatively, the effect of search frictions on transaction prices is small compared to the effect of transaction costs.

In the final section of the paper, we explore how the response of housing markets to new construction depends on local clientele patterns. We contrast the construction of new housing in two neighborhoods that are similar in size and price, but differ in the share of broad searchers and hence their integration with the rest of the Bay Area. We find that new construction in urban San Francisco segments with many broad searchers affects many similarly priced segments across the entire city, but does not have a particularly large effect on inventory in the segments where construction actually takes place. In contrast, new construction in suburban segments close to the San Francisco city boundary – segments that mostly attract narrow searchers – increases inventory in those segments, with much smaller effects on nearby housing markets. Here, the number of narrow searchers who obtain housing surplus declines with new construction: the higher inventory makes it harder to sell houses, and increased competition from broad searchers attracted to the inventory makes it more difficult to buy. The effect of construction on aggregate Bay Area inventory also depends on the segment in which the construction takes place.

We conclude that information on clientele patterns is essential for housing policy. Too much aggregation leads to bias in predicting the effects of construction on both local and aggregate inventory. Indeed, a typical calibration to aggregate metro area moments will infer an elasticity of housing demand as if there were only broad searchers. Most concrete zoning proposals, however, focus on building in a particular area. If the metro area is not fully integrated, local clientele patterns can lead to larger or smaller responses of equilibrium inventory than one would predict using an aggregate model. Our examples suggest that these effects can be quantitatively large.

Relatedly, our results on the Beveridge curve call for a cautious and selective use of low inventory as a signal of a housing shortage. Building houses in cities with low inventory does indeed address a housing shortage: it targets cities where local excess demand is high and all new houses satisfy local demand. This reduces the time it takes households that are interested in this city to find a house, and increases steady state inventory. In contrast, building houses in low-inventory neighborhoods within a city selects segments with relatively low local excess demand. Instead, the additional inventory mostly attracts more broad searchers who have no specific preference for the location. The inflow of broad searchers who now compete with the narrow searchers for the additional inventory reduces the effect that the new construction has on the time it takes these narrow households to find a home.

Our paper contributes to a growing body of empirical work that analyzes housing market activity in the cross section. The typical approach is to sort houses within geographic units (such as cities) based on similarity along property characteristics, and to then compare measures of market activity such as turnover and time on market across these segments (see Goodman and Thibodeau, 1998; Leishman, 2001; Islam and Asami, 2009). There is, however, little direct evidence on housing search behavior. An exception is Genesove and Han (2012), who build a time-series of search activity at the city level from survey data on buyers' house hunting experiences. Our paper offers a new source of demand-side information and uses the distribution of online searchers' criteria to define segments for which market activity is then measured.[1] Our approach emphasizes the heterogeneity of market and search activity within cities and across potential buyers.[2]

We build on a large literature on housing search models (see Han and Strange, 2015, for a

[1]Our paper therefore contributes to an emerging literature that shows how the increasing availability of data from online services such as eBay, Facebook, Trulia, and others allows researchers to overcome important measurement challenges across the social sciences (see, for example, Einav et al., 2015; Bailey et al., 2017, 2018a,b).

[2]As such, our paper also contributes to the literature that considers house valuation with heterogeneous regions and buyers. For example, Poterba (1991) consider the role of demographics for prices, Bayer, Ferreira and McMillan (2007) look at school quality, and Hurst, Guerrieri and Hartley (2013) study the effects of gentrification. Stroebel (2016) and Kurlat and Stroebel (2015) investigate the market impact of asymmetric information about property and neighborhood characteristics, respectively. Landvoigt, Piazzesi and Schneider (2015) study the effect of credit constraints on prices in an assignment model with many quality segments. They consider competitive equilibria of a model with homogeneous preferences. In contrast, this paper emphasizes frictional discounts due to search and transaction costs, as well as heterogeneity in search preferences within a metro area. Davis and Van Nieuwerburgh (2015) and Piazzesi and Schneider (2017) survey the broader literature on business cycles, asset prices, and housing, including studies that do not rely on search frictions.

review). Recent studies have turned to quantitative evaluations using micro data (see Diaz and Perez, 2013; Head et al., 2014; Ngai and Tenreyro, 2014; Guren and McQuade, 2014; Anenberg and Bayer, 2014; Halket and Pignatti, 2015; Burnside et al., 2016; Guren, 2017). Most of these authors are interested in how search modifies the time-series dynamics of house prices and market activity in homogeneous housing markets. In contrast, our focus is on steady states of markets with rich cross sections of houses and buyers. Our theoretical model is based on earlier random matching models such as Wheaton (1990), Krainer (2001), Albrecht, Anderson, Smith and Vroman (2007), Novy-Marx (2009), and Piazzesi and Schneider (2009); the new element is that we allow for matching by heterogeneous buyers across multiple interconnected segments.

The paper in the labor search literature that is closest to our work is by Manning and Petrongolo (2017), who estimate a search and matching model for local labor markets. They divide space into small areas that searchers can search jointly, thus generating spillover effects across regions. Their demand-side data consist of job seekers' home addresses and the vacancies they apply to. They infer a distribution of preferences that includes a parameter for how fast utility declines with commuting time. Our detailed search data allows us to put less structure on utility, and to use both spatial and quality information to define the commodity space.

The rest of the paper is structured as follows. Section 2 describes our housing search data, and presents key patterns of housing search behavior. Section 3 establishes stylized facts on housing market and search activity at the segment level, exploring how the slope of the housing Beveridge curve varies with the level of aggregation. Section 4 presents a reduced-form model of a single segment that highlights the key economic forces arising from the interaction of broad and narrow searchers in housing markets. Section 5 estimates a fully fledged housing search model with many segments that quantifies the importance of these forces. Section 6 uses estimates from this model to infer frictional discounts and to explore the response of prices and quantities in different market segments to local housing market shocks. The final section concludes by discussing the importance of understanding the segmentation of search clienteles across a number of other important search markets, such as over-the-counter financial markets, labor markets, and dating markets.

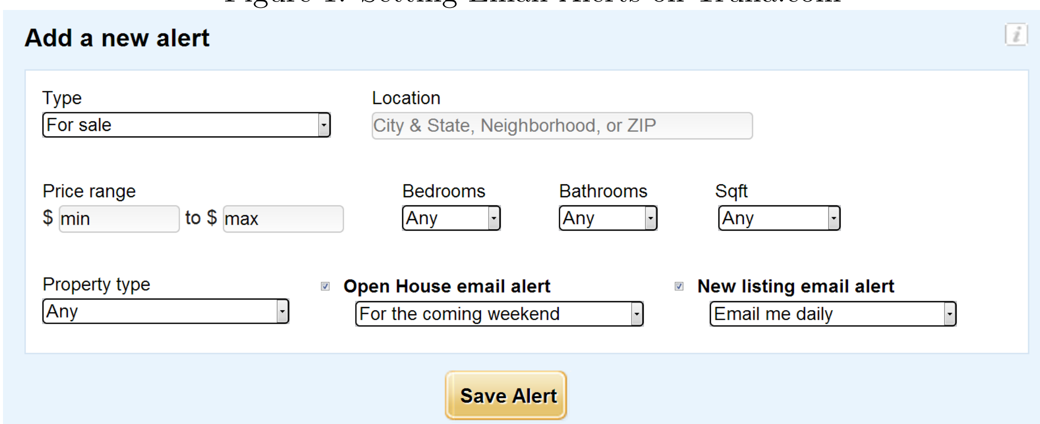
2 Understanding Housing Search Behavior

We document housing search behavior in the San Francisco Bay Area using email alerts set on the popular real estate website trulia.com. The San Francisco Bay Area is a major urban agglomeration in Northern California that includes San Francisco, San Jose, Oakland, as well as a number of other cities. We analyze data from Alameda, Contra Costa, Marin, San Benito, San Francisco, San Mateo, and Santa Clara counties. In the 2010 Census, these counties had a population of about 6 million people living in 2.2 million housing units. This section first describes the email alert data and then highlights the key findings from Appendix A, where we provide a detailed analysis of the raw search patterns.

2.1 Search Data From 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 list of zip codes, neighborhoods, or cities. Neighborhoods are geographic units commonly listed on realtor maps that are often aligned with zip codes. When users fill out the form, an auto-complete function suggests names of neighborhoods or cities.

Figure 1: Setting Email Alerts on Trulia.com



Note: Figure shows the web form for setting email alerts on trulia.com.

The second row in the form provides the option of specifying property characteristics beyond geography. Price ranges may be set by providing a lower bound, an upper bound, or both. For bedrooms, bathrooms, and house size, there is the option to set a lower bound or an upper bound. In the third row, "Property type" allows narrowing the search to "Single family home," "Condo," and several smaller categories. The remaining fields govern how emails are processed: for the "New listing email alerts" relevant to our study, the options are to receive a daily or weekly email.[3]

We observe 40,525 "For sale" email alerts set for Bay Area properties between March 2006 and April 2012. Those alerts were set by 23,597 unique home searchers, identified by their (scrambled) email addresses. Almost 70 percent of searchers set only one alert, and more than 90 percent of searchers set three or fewer alerts. Since we are interested in search ranges rather than individual alerts, we pool alerts set by the same searcher, as described in Appendix A.

Representativeness of Search Behavior

We do not observe demographic information on the home searchers in our sample. Thus, we cannot provide direct evidence that searchers on trulia.com are representative of the overall pool of home searchers. However, surveys conducted by the National Association of Realtors (2013) during

[3]Trulia also provides a second way for potential home buyers to set an email alert. After looking at results from regular searches on their website, generally along the same dimensions as those in Figure 1, users can press a single button: "Send me an email whenever houses with these characteristics come on the market."

our sample period suggest that the internet is the most important tool in the modern home-buying process. Indeed, over 90 percent of home buyers rely on the internet in their search. The fraction of people who deemed real estate websites "very important" as a source of information was 76 percent, substantially larger than the 68 percent who found real estate agents "very important."

Internet use for home search is also not concentrated among younger or richer buyers: 86 percent of home buyers between the ages of 45 and 65 go online to search for a home. The median age of home buyers using the internet is 42, and their 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 only slightly wealthier (the median income of all home buyers is $80,900). These statistics suggest that we can learn from online home search about overall home search behavior. Moreover, trulia.com, with approximately 24 million unique monthly visitors during our sample period (71 percent of whom report planning to purchase in the next 6 months), has similar user demographics to those of the overall online home search audience (Trulia, 2013).

2.2 Dimensions of Housing Search

To compare the geographic dimension of individuals' search ranges, we express them in terms of the zip codes that are covered by the pooled email alerts. We observe wide heterogeneity in the geographic breadth considered by various home searchers. While 25% of searchers are narrowly interested in a single zip code, among those individuals who select more than one zip code, the 10-90 percentile range of the maximum geographic distance between zip codes selected by the same searcher is 2.3 miles to 21.1 miles. Appendix A.2 provides further details.

Roughly two-thirds of the email alerts include search parameters in addition to geography. The other fields that are specified regularly are listing price (two-thirds of all alerts) and the number of bathrooms (one-third of all alerts). On the price dimension, among those searchers who set both an upper and a lower bound, the 10th percentile selects a price range of $100k, the median selects a price range of $300k, and the 90th percentile selects a price range of $1.1m. Among the same searchers, the median person specifies a price range of $\pm$ 27% around the mid-point of the range. At the 10th percentile of the distribution, this figure is $\pm$ 12.5% around the mid-point, and at the 90th percentile it is $\pm$ 58%. Among those searchers who specify the number of bathrooms, almost all select a lower bound of "2".

We want to develop a model that captures the heterogeneity of these search ranges. One possible approach to summarize their geographic dimensions would be to use contiguous and/or circular subsets of a plane. This approach does not work well for the Bay Area with its complicated topology. Moreover, many searchers look for houses in zip codes that are not necessarily adjacent to each other. Our approach in the next section is therefore to define a discrete grid of market segments using zip codes as the basic geographic unit, which we further subdivide along quality and size dimensions. A search range can then be represented as a subset of market segments, allowing us to accommodate the observed non-contiguous and non-circular search patterns.

3 Segment-Level Housing Search and Market Activity

We now describe how we divide the San Francisco Bay Area into a finite number of housing market segments, motivated by the search ranges inferred from our email alert data. We then establish stylized facts on market and search activity at the segment level.

The finest partition of the Bay Area housing stock into different market segments that could be motivated by our search data is obtained by the join of all search ranges in our sample. The preferences of each searcher could then be expressed exactly as a subset of these 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 inventory and time on market.

Our approach, therefore, is to get as close as possible to the finest partition, but subject to the constraint that segments must be sufficiently large in terms of volume and housing stock. We start with zip codes as the level of geography, and then subdivide zip codes along price and size boundaries that are common in email alerts that cover a particular zip code. We then merge small but similar segments within a zip code until all remaining segments have at least 1,200 housing units. This process leads us to a final set $\mathcal{H}$ of 564 segments that are sufficiently large to accurately measure housing market activity. These segments contain houses within a zip code that are of similar quality (based on price) and size (based on the number of bathrooms), with cutoffs that are close to cutoffs regularly used in email alerts. Appendix B.1 provides details on the algorithm.

We then express each search range as a subset of these segments. Here we start from the raw search ranges, specified along the dimensions geography, quality, and size; we ignore the other dimensions that are rarely specified. We then determine which segments are (approximately) covered by each range, using an algorithm described in Appendix B.2. This produces a set Θ of 11,503 distinct search ranges that is each represented as a subset of $\mathcal{H}$.

3.1 Segment-level Housing Market Activity

To measure housing market activity at the segment level, we combine three main data sets. We start with the universe of ownership-changing deeds in the San Francisco Bay Area between 1994 and April 2012. From the deeds data, we obtain the property address, transaction date, transaction price, and the type of deed (e.g. Intra-Family Transfer Deed, Warranty Deed). We use the type of deed to identify arms-length transactions (see Appendix B.3 for details). We combine these transaction deeds with the universe of tax assessment records for the year 2009. This data set includes information on property characteristics such as construction year, size, and the number of bedrooms and bathrooms. Finally, we use data on all property listings on trulia.com between October 2005 and December 2011. The key variables from this data set are listing date, listing price, and the listing address. The latter can be used to match listing data to deeds data. We can then construct a measure of time on market for each property that eventually sells, as well as the inventory that is for sale in a market segment at each point in time.

Throughout our analysis, we pool observations for the period 2008-2011, a time period for which we observe information on both housing search and housing market activity. 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 by ensuring that there are sufficiently many observations to measure market activity in segments with few listings and low housing turnover. In Appendix C, we show that segment-level market and search activity are quite stable over time within our sample period. To make prices comparable across years, we convert all prices to 2010 dollars using zip code-level repeat sales price indices.

Segment-Level Facts: Notation

We next present segment-level facts about market and search activity. These facts reveal a number of interesting patterns that motivate the subsequent quantitative exercise. The following notation, summarized in Table 1, is useful to organize facts reported at the segment level. As before, $\mathcal{H}$ denotes the set of all 564 segments. The housing stock in segment h is given by $H(h)$. We normalize the total Bay Area housing stock to a unit mass: $\sum_{h \in \mathcal{H}} H(h) = 1$.

The average monthly turnover rate $V(h)$ in segment h is defined as the number of arms-length transactions in that month $m(h)$ divided by the housing stock $H(h)$. The mean time on market $T(h)$ in segment h is defined as months between listing and sales date, less one month for the typical escrow period. Our measure of inventory in segment h is $L(h) := T(h)\, m(h)$.[4] We also define the inventory share $I(h) = L(h)/H(h)$ as the share of the housing stock that is for sale.

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

$$\tilde{H}(\theta) = \sum_{h \in \tilde{\mathcal{H}}(\theta)} H(h).$$

Similarly, we define the total inventory considered by searcher θ as $\tilde{L}(\theta) = \sum_{h \in \tilde{\mathcal{H}}(\theta)} L(h)$, which is the sum over all inventory for sale in segments in θ's search range $\tilde{\mathcal{H}}(\theta)$. The *clientele* of segment h consists of all searchers who consider segment h as part of their search range, that is,

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

The pattern of clienteles reflects the interconnectedness of segments. One polar case is a perfectly segmented market, in which $|\mathcal{H}|$ types have search ranges that each consist of a single segment, and

[4] This measure of inventory conditions on houses that are eventually sold, since time on market $T(h)$ is based on actual sales. Alternatively, one could construct measures of inventory directly from listings data. The resulting data series are noisy because of the incomplete coverage of Trulia listings data, and the need to make assumptions on when the few listings that do not sell are removed. We discuss the trade-offs involved in the choice of how to measure inventory in Appendix B.3.

Table 1: Main Notation in the Paper

Symbol	Definition	Additional Details
Segment-Level Notation		
$\mathcal{H}$	Set of all segments h	$h \in \mathcal{H}$
$H(h)$	Housing stock in h	Normalization: $\sum_{h\in\mathcal{H}} H(h) = 1$
$m(h)$	Transactions in h	
$V(h)$	Monthly turnover rate in h	$V(h) = m(h)/H(h)$
$T(h)$	Average time on market in h	Measured in months
$L(h)$	Inventory in h	Steady state: $L(h) := T(h)m(h)$
$I(h)$	Inventory share in h	$I(h) = L(h)/H(h)$
$v(h)$	Flow of housing services in h	
$p(h)$	Transaction price in h	See equation E.2 for details
$\sigma(h)$	Weighted searchers per house in h	$\sigma(h) = \frac{1}{H(h)} \sum_{\theta\in\tilde{\Theta}(h)} \frac{\tilde{B}(\theta)}{\bar{B}} \frac{H(h)}{\tilde{H}(\theta)}$
$\tilde{\Theta}(h)$	Clientele of h	$\tilde{\Theta}(h) = \{\theta \in \Theta : h \in \tilde{\mathcal{H}}(\theta)\}$
$B(h)$	Buyers in h	$B(h) = \sum_{\theta\in\tilde{\Theta}(h)} \tilde{B}(\theta) \frac{L(h)}{\tilde{L}(\theta)}$
$\tilde{m}(B(h), L(h), h)$	Matching function in h	
$\eta(h)$	Instability of h	
$\alpha(h)$	House finding rate in h	$\alpha(h) = \frac{m(h)}{B(h)}$
$\pi(h)$	Popularity of h	$\pi(h) := \frac{1}{H(h)} \sum_{\theta\in\tilde{\Theta}(h)} \mu(\theta) \frac{H(h)}{\tilde{H}(\theta)}$
Searcher-Level Notation		
Θ	Set of search ranges θ	$\theta \in \Theta$
$\mu(\theta)$	Measure of θ	$\sum_{\theta\in\Theta} \mu(\theta) = \bar{\mu} > 1$
$\tilde{\mathcal{H}}(\theta)$	Segments scanned by θ	$\tilde{\mathcal{H}}(\theta) \subset H$
$\tilde{H}(\theta)$	Housing stock of interest to θ	$\tilde{H}(\theta) = \sum_{h\in\tilde{\mathcal{H}}(\theta)} H(h)$
$\tilde{L}(\theta)$	Inventory considered by θ	$\tilde{L}(\theta) = \sum_{h\in\tilde{\mathcal{H}}(\theta)} L(h)$
$\tilde{B}(\theta)$	Buyers of type θ	$\sum_{\theta\in\Theta} \tilde{B}(\theta) = \bar{B}$
Other Notation		
c	Proportional transaction cost	

each segment has a homogeneous clientele of one type who searches only in that segment. Another polar case is a perfectly integrated market, where a single type searches across all segments 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 $\tilde{B}(\theta)$ denote the number of buyers with search range θ. The total number of buyers is $\sum_{\theta\in\Theta} \tilde{B}(\theta) = \bar{B}$. The distribution of searchers interested in segment h is then obtained by integrating the distribution $\tilde{B}(\theta)/\bar{B}$ over $\tilde{\Theta}(h)$. We want a summary statistic of search activity that is comparable across segments that differ in the number of broad and narrow searchers. We compute a measure of *weighted searchers per house* in segment h that weights searchers according to their search breadths:

$$\sigma(h) = \frac{1}{H(h)} \sum_{\theta \in \tilde{\Theta}(h)} \frac{\tilde{B}(\theta)}{\bar{B}} \frac{H(h)}{\tilde{H}(\theta)} = \sum_{\theta \in \tilde{\Theta}(h)} \frac{\tilde{B}(\theta)}{\bar{B}\,\tilde{H}(\theta)}. \tag{2}$$

Weighting the contribution of each searcher type by $H(h)/\tilde{H}(\theta)$ makes broader searchers count less toward search activity in segment h. If every searcher was looking at only one segment, then $\sigma(h)$ would simply reflect the relative number of searchers per house in h, since in that case the housing stock $\tilde{H}(\theta)$ of a searcher θ interested in h would equal $H(h)$. More generally, our measure of search activity $\sigma(h)$ is an index which is normalized to one for the entire Bay Area.[5] So far, all summary statistics have been defined at the segment level. 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 the housing stock, aggregation uses the housing stock as weights. For example, the turnover rate over some subset $G \subset \mathcal{H}$, such as a zip code or city, is computed as

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

Segment-Level Facts: Summary Statistics

Table 2 presents summary statistics on market and search activity for the San Francisco Bay Area as a whole, as well as by segment. The housing market is relatively illiquid, which is consistent with our sample period 2008-2011 covering the housing bust when housing transactions were unusually low. On average, only 1.1 percent of the Bay Area housing stock is for sale at any point in time. The average monthly turnover rate is 0.24 percent, so that the typical house turns over once every $100/(0.24 \times 12) = 35$ years. The cross-sectional variation in market activity at the segment level is substantial. For example, the 75th percentile of inventory share is 1.51 percent, over twice as much as the 0.61 percent inventory share at the 25th percentile.

The distribution of $\sigma(h)$, our measure of search activity, is positively skewed across segments: most segments have less than one weighted searcher per house. The minimum of 0.05 is achieved by a segment in the Sacramento Delta, which is only considered by a few broad searchers. Other segments have substantially more search activity, all the way to a maximum of 7.05 in a segment in central San Francisco, which attracts many narrow searchers in addition to broad searchers.

[5] The sum over all $\sigma(h)$'s in the Bay Area, weighted by housing stock, is:

$$\sum_{h \in \mathcal{H}} \frac{H(h)}{1} \sigma(h) = \sum_{h \in \mathcal{H}} \sum_{\theta \in \tilde{\Theta}(h)} \frac{\tilde{B}(\theta)}{\bar{B}} \frac{H(h)}{\tilde{H}(\theta)} = \sum_{\theta \in \Theta} \sum_{h \in \tilde{\mathcal{H}}(\theta)} \frac{\tilde{B}(\theta)}{\bar{B}} \frac{H(h)}{\tilde{H}(\theta)} = \sum_{\theta \in \Theta} \frac{\tilde{B}(\theta)}{\bar{B}\,\tilde{H}(\theta)} \sum_{h \in \tilde{\mathcal{H}}(\theta)} H(h) = \sum_{\theta \in \Theta} \frac{\tilde{B}(\theta)}{\bar{B}} = 1.$$

Table 2: Summary Statistics of Market and Search Activity

	Inventory Share I (in percent)	Turnover Rate V (in percent)	Search Activity σ	Mean Price (in k$)	Housing Stock
Bay Area	1.14	0.24	1.00	649	2,216,021
Across-Segment					
Min	0.00	0.00	0.05	72	1,221
P25	0.61	0.16	0.51	306	2,333
P50	0.94	0.21	0.81	518	3,298
P75	1.51	0.29	1.31	790	4,858
Max	6.66	0.97	7.05	2,491	13,167

Note: Table shows summary statistics for market and search activity, both for the entire San Francisco Bay Area as well as across the 564 housing market segments. We consider inventory share I, turnover rate V, search activity σ as defined in equation (2), mean price, and total housing stock.

Segment-Level Housing Market Activity: Within and Across Submarket Correlation

Table 3 reports cross-sectional correlations of observables for "submarkets" at different levels of aggregation. These submarkets are the 564 segments, the 191 zip codes, and the 96 cities in our data. The left panels show volatilities and correlation coefficients *across* submarkets. The right panels consider segment-level variation *within* submarkets.

A comparison of volatilities shows substantial variation across segments within the same zip code. Indeed, the zip code-level movements account for only 46%, 44%, and 64% of the across-segment variance in inventory share I, turnover rate V, and search activity σ, respectively.

The comovement of search activity (σ) and market activity (V and I) depends crucially on the level of aggregation. While it is close to zero at the segment level, it turns negative when analyzed across zip codes, and even more negative when analyzed across cities. In addition, more expensive zip codes and cities are searched more. In contrast, search activity comoves positively with inventory and turnover across segments within zip codes or cities. Within cities and zip codes, more expensive segments are also searched less.

The relationship between inventory and measures of search activity is reminiscent of the "Beveridge curve" in studies of the labor market. The Beveridge curve is a relationship between vacant job positions and unemployed workers searching for jobs, usually presented in the *time series*. Here we have a relationship between vacant homes and individuals searching for homes, but presented in the *cross section*. The Beveridge curve is *downward sloping* across broad units of aggregation such as cities, while it is on average *upward sloping* across small segments within broad units. Panel A of Figure 2 shows the Beveridge curve relationship across Bay Area cities, and Panel C across all segments within the city of San Francisco. The San Francisco patterns are not unusual. Indeed, the Beveridge curve is upward sloping in 64 out of the 74 cities with at least two segments, representing 84% of the total Bay Area housing stock. This fact is also not primarily driven by small cities: the within-city across-segment Beveridge curve slopes up for 23 of the largest 25 cities

Table 3: Cross-Sectional Variation in Market and Search Activity

Variation Across Cities

	I	V	σ	$\log(p)$
vol	0.42	0.07	0.48	0.46
corr	1	0.93	-0.42	-0.74
		1	-0.34	-0.59
			1	0.53
				1

Avg. Variation Within Cities

	I	V	σ	$\log(p)$
vol	0.54	0.09	0.40	0.42
corr	1	0.88	0.51	-0.67
		1	0.42	-0.52
			1	-0.51
				1

Variation Across Zip Codes

	I	V	σ	$\log(p)$
vol	0.50	0.08	0.65	0.50
corr	1	0.91	-0.18	-0.67
		1	-0.15	-0.52
			1	0.41
				1

Avg. Variation Within Zip Codes

	I	V	σ	$\log(p)$
vol	0.60	0.09	0.55	0.41
corr	1	0.83	0.58	-0.76
		1	0.41	-0.55
			1	-0.54
				1

Variation Across Segments

	I	V	σ	$\log(p)$
vol	0.74	0.12	0.81	0.64
corr	1	0.93	0.03	-0.63
		1	0.01	-0.48
			1	0.08
				1

Note: Table shows the cross-sectional variation in market and search activity at different levels of aggregation: 564 segments, 191 zip codes, and 96 cities. The left column presents statistics across these units, the right column presents statistics across segments within these units. We present inventory share I, turnover rate V, search activity σ, and log(price). We present both volatilities (standard deviations) as well as correlations within and across units.

by housing stock, with an average correlation of 0.44, and a 25th percentile correlation of 0.23.

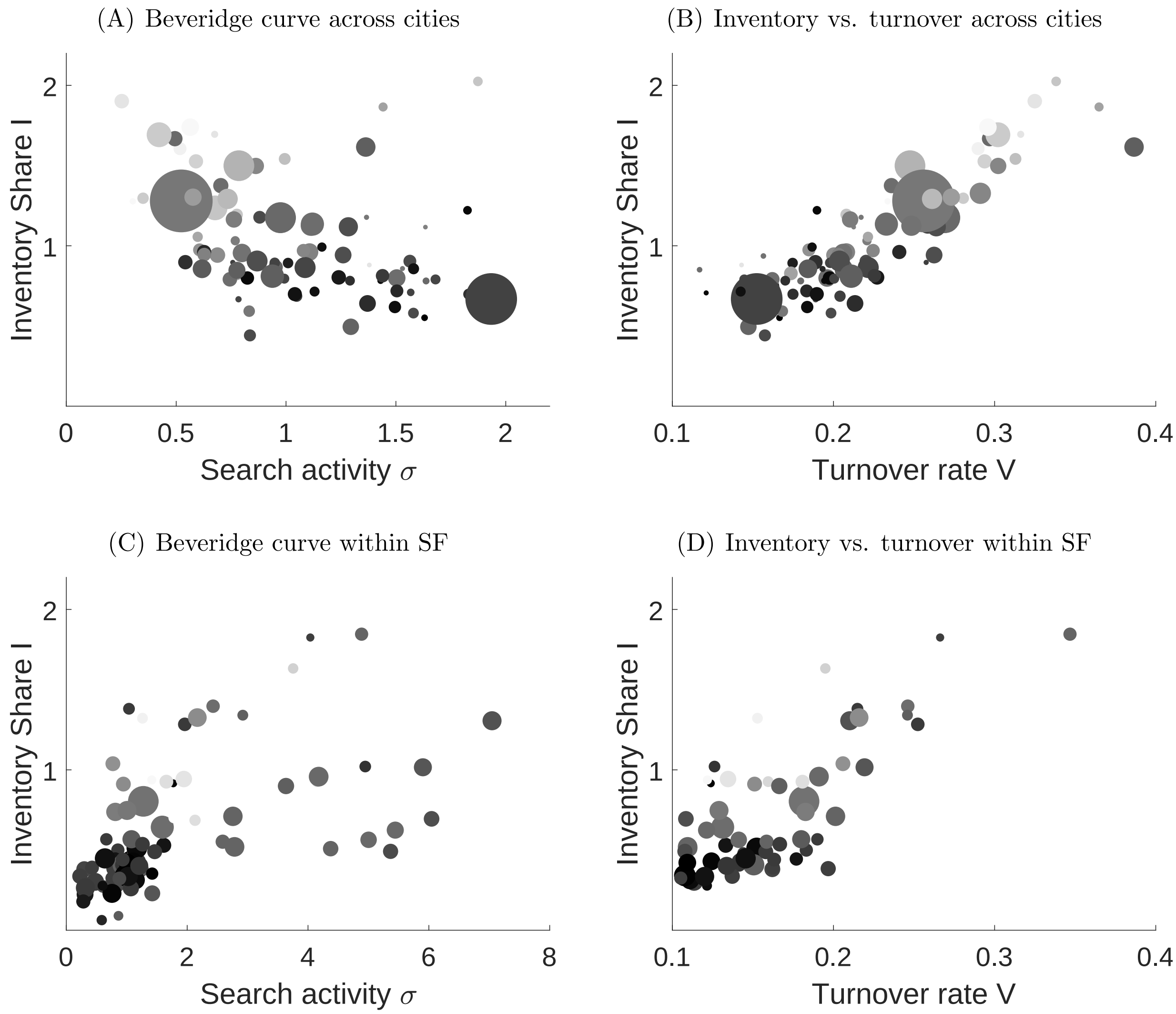
For our market activity indicators and prices, the nature of covariation across and within cities is very similar, both qualitatively and quantitatively: the inventory share and the turnover rate comove positively, and both are negatively correlated with price.[6] To illustrate this, Panel B of Figure 2 shows the positive correlation between inventory shares and turnover rates across cities in the San Francisco Bay Area. Panel D shows the positive correlation between inventory shares and turnover rates across segments within the city of San Francisco.

Segment-Level Facts: Search-Breadth

The measure $\sigma(h)$ reflects the average search activity in a segment, but it does not tell us whether that activity is due to narrow searchers or due to broader searchers who provide connection to other segments. Indeed, the same $\sigma(h)$ could arise when there are a few narrow searchers who

[6]The high-turnover low-price segments could correspond to segments with many starter homes (see Ortalo-Magne and Rady, 2006).

Figure 2: Inventory, Search Activity, and Volume

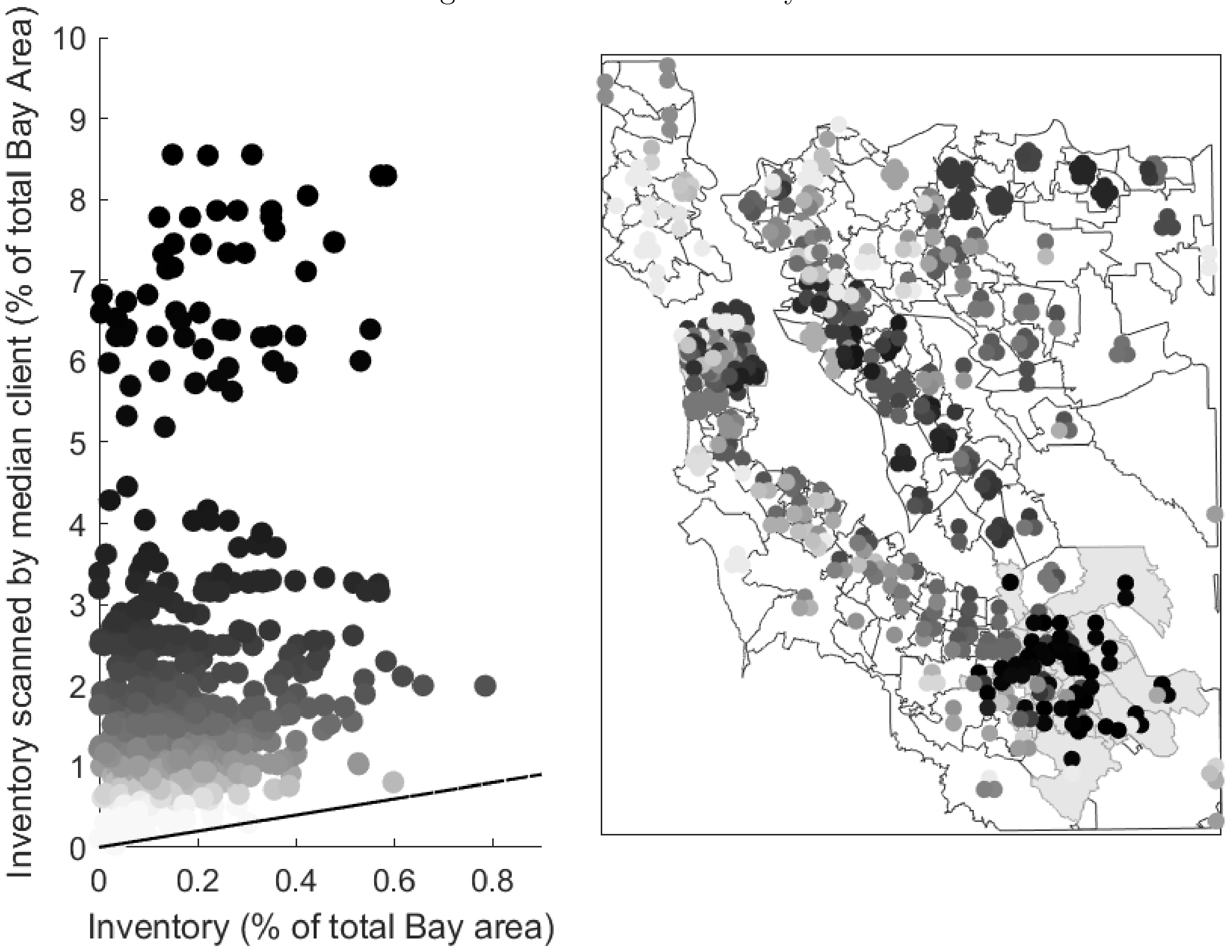


Note: Figure shows the inventory share I, search activity σ, and the turnover rate V, both across and within cities. Panel A shows inventory shares and search activity (the Beveridge curve) across cities. Panel B shows inventory shares and turnover rates across cities. Panel C shows the Beveridge curve across segments within the city of San Francisco. Panel D shows inventory shares and turnover rates across segments within San Francisco. The size of the dots reflects the size of the house stock (from small to large). The color of the dots reflects housing prices from cheap (light) to expensive (dark).

fully target their search effort to a given segment, or when there are relatively more broad searchers whose search effort in a given segment is diluted because they also consider other segments. To summarize interconnectedness, we compare segments in terms of the inventory scanned by their median client. The left panel of Figure 3 plots the share of inventory in segment h in total Bay Area inventory against the share of Bay Area inventory scanned by the median client of segment h. Every dot represents a segment, and colors reflect the value on the vertical axis so that segments can be recognized in the map in the right panel.

If the Bay Area were perfectly segmented, then any given segment would only be searched

Figure 3: Scanned Inventory



Note: Figure shows scanned inventory by median searcher in each segment, presented as a share of total Bay Area inventory. In both panels, each dot corresponds to one segment. In the left panel, the horizontal axis shows the inventory in that segment as a fraction of total Bay Area inventory. The vertical axis shows the fraction of total Bay Area inventory scanned by the median searcher in that segment. The straight line is the 45-degree line. The right panel shows the geographic distribution of these segments. Dots for segments within the same zip code are arranged clockwise by price with the lowest priced segment at noon.

by individuals who scan that particular segment. As a result, all dots would line up along the 45-degree line (which is the straight line in the left panel of Figure 3). At the opposite extreme, if the Bay Area were perfectly integrated, then every client of every segment would scan all houses, so all dots should line up along a horizontal line at 100 percent of total inventory (located north of our current figure). The reality 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 the Bay Area total.

Appendix B.4 provides summary statistics on this measure of search breadth. The median searcher in the average segment scans 2.1 percent of the total Bay Area inventory; in segments at the 25th percentile of the distribution, the median searcher scans 0.94 percent of Bay Area inventory, while in segments at the 75th percentile of the distribution, the median searcher scans 2.53 percent of Bay Area inventory. Remarkably, the average within-segment inter-quartile range of inventory scanned by different searchers looking in the same segment is similarly large, at 1.75 percent of the Bay Area inventory. This highlights that while segments differ in the average search breadth of their clientele, most segments feature competition between narrow and broad searchers. The interaction between these broad and narrow searchers will be a central force in our model.

Areas with a large common clientele appear in the left panel 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. This effect is visible for the top cluster of dark black dots. The map in the right panel shows that those dots represent cheaper segments in the city of San Jose, which is marked grey. More generally, clusters of dots with high scanned inventory correspond to cheap urban areas, where broad search appears to be more common.

3.2 Search Intensity within Search Range

How do searchers choose among the inventory within their search range? Are searchers equally interested in all properties, or do they prefer properties in some segments in their search range to those in others? To address this question, we exploit data on property views by individual home searchers on trulia.com. After defining a search range on trulia.com, a user is presented with a list of properties that are included in that search range (see Appendix Figure A.7 for a screen shot of the interface). This list provides basic information on each property, such as its location, the listing price, a picture, and the first lines of a description of the property. Home searchers then actively click on those houses that attract their particular interest to view additional property information and potentially contact the realtor representing the seller.

We have obtained data on such detailed property views from trulia.com for a random subset of users visiting the site in April 2012. These data contain the set of listings viewed within a "session," defined as all views by the same user (represented by their IP address) within one day. We interpret viewing a property's listing details as an expression of particular interest in that property. In Appendix A.5, we analyze how this signal of particular interest is distributed across listings in the various segments searched by an individual. We find that the rate at which particular interest is expressed for properties in different segments is directly related to the share of total inventory made up by those segments in the individuals' overall search ranges. This finding suggest that, conditional on the search range, the probability of finding a favorite house in any one particular segment is proportional to the inventory in that segment. This observation motivates one of our key modeling assumptions in the next section.

4 A Stylized Model of a Single Segment

In this section, we consider a simple reduced-from model of a single segment. We show how the cross section of observables introduced in Section 3.1 – turnover, inventory, and search activity – is shaped by supply and demand forces. The equations of the reduced-form model describe flows of buyers and houses in a particular segment. They hold in steady state equilibrium for a large class of continuous-time search models. We specify and estimate one such model in Section 5. The purpose of the discussion below is to illustrate the key mechanisms that shape the housing Beveridge curve independently of model details such as price formation and how broad searchers select houses within their search ranges. Most importantly, we show that the interaction of broad and narrow searchers explains the key stylized fact about the Beveridge curve within and across cities from Figure 2. Moreover, we show how policy analysis at the wrong level of aggregation can lead to misleading conclusions.

Consider a segment with mass H of houses. If L houses are listed for sale and B agents are looking to buy, transactions occur at the rate $m(B, L)$, where the matching function m is increasing in both arguments and exhibits constant returns to scale. Agents own at most one house. The $H - L = H(1 - I)$ homeowners who do not already list their house for sale receive moving shocks at a constant rate η. Upon receiving a shock, they list their house, but stay in it until they sell; only then do they look for a new house. Individual agents thus cycle across three states: owning but not listing, owning and listing, and looking for a new house. In steady state, the number of agents in each state is constant. The share of houses coming on the market must thus be the same as the turnover rate at which houses are sold, or $\eta(1 - I) = V$. We now characterize the distribution of agents under different assumptions about buyer behavior.[7]

Narrow Searchers and the Cross Section of Cities

Suppose first that all agents are narrow types who only want to live in the segment under study. If the total number of narrow-type agents is $N > H$, the number of agents looking for a house is $B^N = N - H$. All houses are owned by $N - B^N$ narrow types. In steady state, the number of houses those owners put up for sale must equal the number of houses sold:

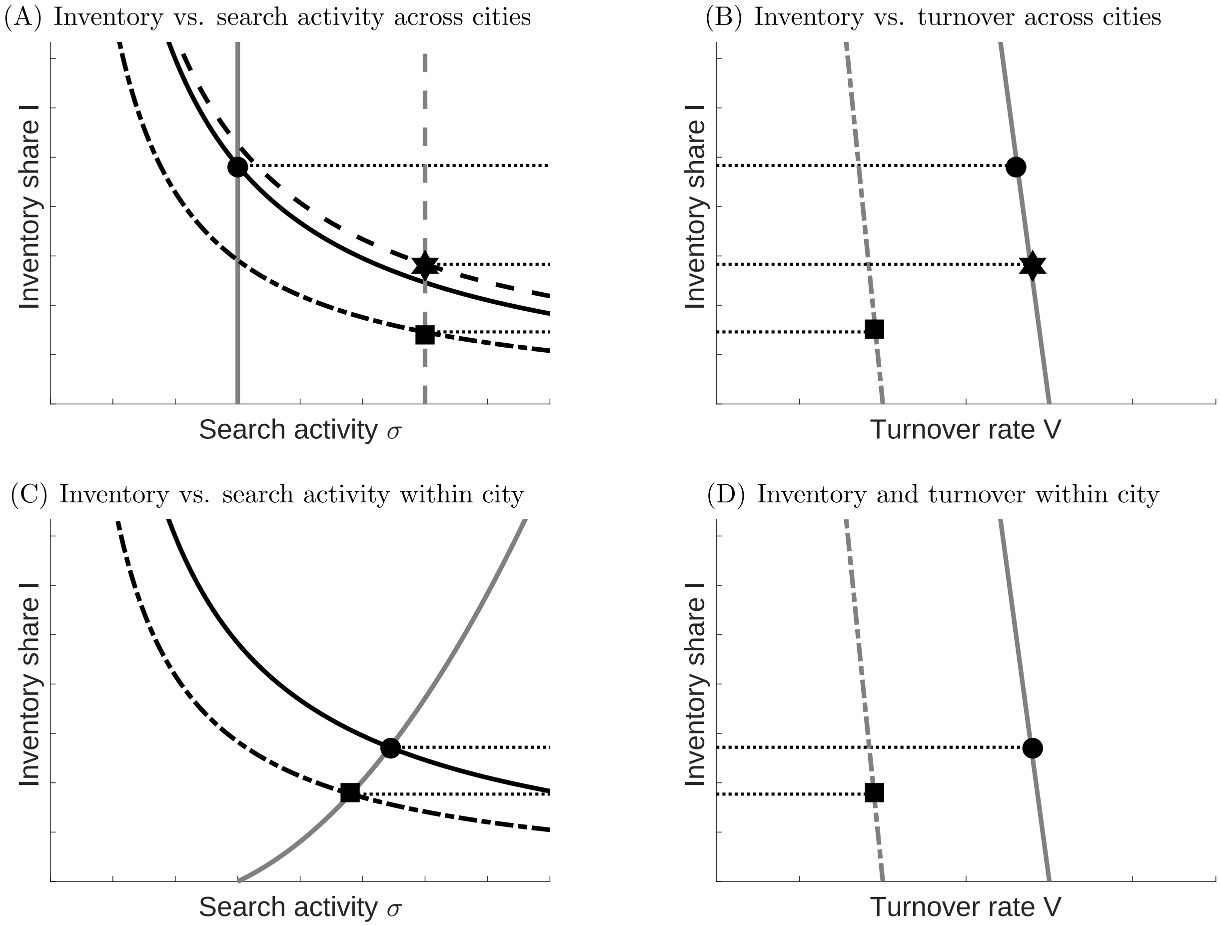
$$\eta (1 - I) \left(N - B^N\right) = m\left(B^N, I\left(N - B^N\right)\right). \tag{3}$$

The equation implies an equilibrium relationship between the number of narrow searchers and the inventory share: if a larger share of houses is for sale, then it is easier for narrow types to find a house so the number of narrow searchers is smaller.

The model with narrow types is useful to interpret our scatter plots on the cross section of cities. The top row panels of Figure 4 are designed to explain the variation in the scatter plots in the top row panels of Figure 2. The left panel relates inventory share and search activity $\sigma = B^N/H$ for two segments of the same size H, indicated by different line styles. For each segment, there is a black downward sloping curve described by (3) as well as well as a gray vertical bar to indicate the exogenous number of narrow searchers. Solid lines describe a baseline segment and the baseline equilibrium inventory share and search activity are intersections marked by circles. The right panel relates inventory share and turnover: the grey downward sloping curve is the condition

[7] Formally, an individual agent's state evolves according to a stationary Markov chain. In steady state, the ergodic distribution of this chain is the same as the distribution of agents in the population.

Figure 4: Inventory, Search activity, and Volume in Stylized Single-Segment Model



Note: Figure shows inventory I, search activity σ, and turnover rate V, both across and within cities for the stylized model of one segment. In the left panels, the solid downward-sloping curves describe equilibrium flows among narrow types (3). The other solid curves capture the interaction (4) between narrow and broad searchers. The right panels describe the flow relationship $\eta(1-I) = V$ between inventory and turnover. More narrow types (higher N) creates the dashed lines, while less instability (lower η) generates the dash-dotted lines.

$\eta(1-I) = V$: since houses come on the market at a constant rate, higher inventory means lower turnover. Equilibrium turnover is at the point on the curve picked out by the inventory share determined in the left panel.

In the data, cities with lower inventory have more searchers and less turnover. In the absence of broad types, these patterns require the co-movement of two forces. First, some cities must be more "popular", in the sense that more (narrow) agents N are interested in living there. This force is essential to generate the observed variation in the number of searchers per house, $\sigma = (N-H)/H$. Higher N also affects the equilibrium condition (3) – if there are more narrow types, it becomes harder to find a house at a given inventory, so the number of narrow searchers increases. Both curves in the top left panel of Figure 4 thus shift to the right and are now dashed; the equilibrium of a city with higher N is marked by stars, with lower inventory and more search activity.

Variation in popularity – measured by N – is thus consistent with cities lining up along a downward-sloping Beveridge curve. By itself, however, it is not consistent with the positive co-movement of inventory and turnover that we see in the data. Indeed, higher N does not affect the relationship between inventory and turnover $\eta(1-I)=V$ in the right panel. Reading across from the new star equilibrium in the left panel to the right we obtain that more popular cities see lower inventory and more turnover. Intuitively, a larger pool of agents looking for houses not only reduces inventory but also allows for faster matching.

To account for the co-movement of inventory and turnover in the data, we need a second force: more popular cities must also be "more stable", which means that houses come on the market at a slower rate η. The dash-dotted lines show the effect of a lower η, and the new equilibrium is marked by squares. If agents become unhappy at a slower rate, then for a given pool of searchers we have less inventory (the square in the left panel) and lower turnover (the square in the right panel). The higher stability in more popular cities thus contributes to the downward-sloping Beveridge curve, while helping to explain the positive co-movement of inventory and turnover.

Broad Searchers and Competition within Cities

To study the cross section of segments in a city, we now allow for "broad types" who are also interested in other segments. However, at any instant, broad types can try to buy in only one segment. The number of broad types who want to buy in the segment under study is determined by a non-negative and increasing function of inventory, $B^B(L)$. This function captures the sensitivity of broad searchers to local conditions: a segment with higher inventory attracts more broad types wanting to buy there, consistent with the evidence in Section 3.2. The idea is that, all else equal, the more houses are available for sale in a segment, the more likely it is that a broad searcher will find her favorite house in that segment.[8]

How can we characterize the equilibrium with broad searchers? The key new feature is that the number of narrow searchers B^N is no longer exogenously given – it is instead determined by the equilibrium interaction between broad and narrow types. As before, the steady state distribution of agents across individual states must be constant, but it now includes the number of broad types in each state. Equilibrium flows must thus be consistent with broad types trading only with broad types, and narrow types only trading with narrow types, as described by condition (3), which thus continues to hold. Moreover, the ratio of total buyers B to owners H must be the same as the buyer-owner ratio for narrow types only:

[8]In the simple model here, we assume that $B^B(L)$ does not depend directly on η, but only responds to inventory. Our quantitative model below has the same feature – it assumes that broad searchers flow to segments in their search range in proportion to inventory. In principle, it is also possible to allow $B^B(.)$ to depend on η such that a segment may be less attractive to broad searchers if houses there fall out of favor more quickly. For example, a common class of models in urban economics assumes that broad searchers are indifferent in equilibrium between buying in the segment under study and receiving fixed utility elsewhere. This would imply a function $B^B(L,\eta)=q(\eta)L$, where q is the tightness of the market – the ratio of buyers to sellers. Tightness is decreasing in η: there are fewer buyers per seller if home ownership yields utility for a shorter amount of time. Appendix D shows that under plausible conditions on parameters, our results extend to models in which buyer flows also depend on η.

$$\frac{B}{H} = \frac{B^N + B^B(L)}{H} = \frac{B^N}{N - B^N}. \tag{4}$$

We thus obtain a positive relationship between L (and thus I) and B^N that reflects *competition between broad and narrow types.* Higher inventory L attracts more broad searchers who crowd out ownership from narrow searchers. This increases the equilibrium number of narrow searchers B^N. The latter continues to map directly into our measure of search activity σ, since broad searchers do not affect σ differentially across segments.[9]

The bottom row of Figure 4 illustrates the impact of broad searchers. In the left panel, equilibrium inventory and search activity are determined by the intersection of equations (3) and (4). The right panel again shows the flow condition $\eta(1 - I) = V$. Section 3.1 documents that within most Bay Area cities, the Beveridge curve is *upward sloping*: segments with more inventory have more search activity. In addition, those segments have more turnover. Our setup explains this pattern with the attraction of broad searchers to segments with higher inventory. The bottom row in Figure 4 holds the number of narrow searchers fixed and only varies the stability parameter η. The downward-sloping dash-dotted line in the bottom left panel is thus the same as in the top left panel; it shows flows among narrow types in a more stable segment (lower η).

Regardless of the clientele structure, more stability reduces inventory – in this regard the top and bottom left panels agree. The new element in the bottom left panel is that lower inventory attracts fewer broad searchers who crowd out narrow searchers. Mechanically, we move down along the upward-sloping curve described by equation (4) to the equilibrium marked by a square. In a more stable segment, more narrow types own a house and overall search activity is lower. The more stable segment thus has both less inventory and less search activity, generating an upward-sloping Beveridge curve. The bottom right panel in Figure 4 shows that the more stable segment also has less turnover (also marked by a square), consistent with the pattern in the data.

The Effects of Housing Policy and the Importance of Search Data

We next illustrate the importance of search data for predicting the effects of local housing policies. Consider a researcher who tries to forecast the effects of building houses in a metro area on inventory and the pool of homeowners there. In the absence of data on housing search ranges, the typical approach is to assume that the area of interest is populated by homogeneous agents who only search in that area. With this identifying assumption, the model parameters can be inferred from observable moments such as the housing stock H, the turnover rate V, inventory share I, and the average time it takes to find a house, which, in steady state, equals $B/(VH)$. Specifically, from our discussion of a segment with only narrow types, we have $\eta = V/(1 - I)$,

[9] We have defined search activity σ as the (weighted) number of searchers that include a segment in their search range. Broad searchers who include many segments in their search range therefore have the same direct impact on σ across all segments in their range. Any differences in σ across segments are thus due to differences in the number of narrow searchers, B^N.

$N = (B/(VH))VH + H$ and $B = N - H$. It is also possible to estimate the shape of the matching function. We now ask what happens if the identifying assumption of homogeneity is incorrect because it imposes either too much or too little aggregation.

It is helpful to first calculate the effects of construction in a single segment with a fixed number of narrow searchers. Equilibrium inventory is characterized by the flow condition (3) with $B^N = N - H$. For each additional house built, the number of houses that ends up in inventory is

$$\frac{dL}{dH} = \frac{L}{B}\frac{B + \varepsilon_B H(1-I)}{L + \varepsilon_L H(1-I)}, \tag{5}$$

where ε_B and ε_L correspond to the elasticities of the matching function $m(B, L)$ with respects to buyers and listings, respectively.[10]

Inventory always increases after new construction, and more so if the matching function is relatively more responsive to the number of buyers than to the number of listings. We can simplify further if we follow the typical calibration strategy and assume a symmetric matching function as well as numbers of buyers and listings that are small relative to the housing stock. The latter property usually obtains because inventory shares are small in the data, and average buyer search time $B/(VH)$ is similar in magnitude to average seller time on market, I/V. With these simplifications, both elasticities are equal to one half and the second fraction is close to one.

We thus obtain a very simple approximate formula for the impact of new construction: for each new house built, L/B houses end up in inventory. If inventory in the initial equilibrium is exactly equal to the number of buyers, then new construction does not lead to more occupied houses – it simply leads to more inventory. If $L > B$, inventory increases by more than the housing stock: adding too many houses "clogs" the market and makes it more difficult for sellers to find buyers. Only if $L < B$ does not all of the new construction end up in inventory. So far in this section, we have kept the model at a reduced form level without introducing preferences and welfare. However, under additional assumptions, we can use the number of houses not on the market $H - L$ as an index of consumer welfare. Indeed, suppose that agents have quasilinear preferences over housing and a numeraire good, and that they receive utility from their house only before they receive a moving shock, as in our quantitative model below. Pareto optima then maximize the sum of utilities across agents less construction costs. The sum of utilities is proportional to $H - L$, so construction improves welfare only if $L < B$.[11]

---

[10] The flow equation $m(N-H, L) = \eta(H-L)$ determines the number of houses in inventory L as a function of the housing stock H. By applying the implicit function theorem, we get

$$\frac{dL}{dH} = \frac{\eta + m_1}{\eta + m_2} = \frac{\eta + \frac{m_1(N-H)}{m}\frac{m}{N-H}}{\eta + \frac{m_2 L}{m}\frac{m}{L}} = \frac{1 + \varepsilon_B \frac{H-L}{B}}{1 + \varepsilon_L \frac{H-L}{L}} = \frac{L}{B}\frac{B + \varepsilon_B (H-L)}{L + \varepsilon_L (H-L)},$$

where the elasticities are $\varepsilon_B = m_1(N-H)/m$ and $\varepsilon_L = m_2 L/m$.

[11] If welfare $W = H - L$, then $dW/dH = 1 - dL/dH$, which, according to equation (5), is only positive if $dL/dH \approx L/B < 1$.

Consider now a researcher who assumes too much aggregation. For a stark thought experiment, suppose that a metro area is actually a collection of many unconnected segments, but the researcher incorrectly assumes that it is perfectly integrated, that is, he assumes that all segments are jointly searched by one broad type. Using this identifying assumption, the researcher can estimate the aggregate listings-buyers ratio L/B for the entire metro area. He can then predict the effects on inventory and welfare based on that aggregate ratio. However, imperfect integration of the area implies that the actual effect on aggregate inventory and welfare may depend on where exactly the new houses are built. For example, the researcher may measure $L/B = 1$ and conclude that there is no housing shortage in the metro area, while it may well be that some segments have $L/B < 1$, so that construction there may be welfare improving if it were sufficiently cheap.

For a second thought experiment, consider the opposite situation: an area is actually fully integrated, and a researcher focuses on one segment assuming that it is not connected to any other segment. Using this identifying assumption, the researcher estimates a segment-specific ratio L/B and uses this ratio to infer a number of narrow searchers N. However, we are actually in a version of the model above with $N = 0$ and only broad searchers who flow to segments according to $B^B(L)$. As a result, if the segment studied by the researcher is small relative to the rest of the city, we would expect the inventory share and turnover rate to be entirely unaffected by the construction of new homes. The new houses are bought by broad searchers without affecting local conditions. Formally, with only broad searchers, inventory is determined by $\eta (H - L) = m \left(B^B (L), L\right)$. Since the segment under study is small, changes in its scale do not affect the inventory share, but only the number of broad searchers who enter.

This thought experiment illustrates the pitfalls of focusing on inventory to guide housing policy. It is tempting to view low inventory in some segment as a "housing shortage" that makes building there particularly important (e.g., Forbes, 2016; USA Today, 2016). With narrow searchers and a downward-sloping Beveridge curve, low inventory indeed flags a large number of agents who desire local housing. With broad searchers and an upward-sloping Beveridge curve, however, low inventory may simply reflect that few houses come on the market. There is no obvious reason to build in low-inventory segments since broad searchers have no particular preference to live there.

Direct data on the pool of searchers helps to distinguish the two cases and can hence improve policy decisions. Indeed, a sample of email alerts would reveal the relative share of narrow searchers B^N/B. Based on local conditions – given by equation (3) – this additional information identifies N and hence puts restrictions on the function $B^B(L)$.[12] In the absence of such data, the quantitative results below suggest that a model with only narrow (broad) searchers will provide a reasonable approximation of the searcher pool across (within) cities.

[12]For example, if $B^B(.)$ is the same for all segments, then it can be traced out from the cross section of $I, V, V/B$ and B^N/B. In fact, in this case the supply and demand parameters η, $B^B(.)$, and N can be identified without taking a stand on the shape of the matching function.

5 A Quantitative Model with Multiple Segments

In this section, we quantify the effects discussed above. To this end, we specify a fully fledged housing search model with multiple segments. At the segment level, equilibrium actions in this model imply the same flow equations as in the simple single-segment model from Section 4 for particular buyer flow functions $B^B(L)$ that accommodate the rich clientele structure in our data. The model here also makes explicit how transactions occur and how prices are determined. Appendix D shows the robustness of the inference to alternative modeling assumptions of the search behavior, the flow of broad buyers, and the price setting mechanism.

Segments and Preferences

We continue to use the notation we introduced to present the facts in Section 3.1, and which we summarized in Table 1. There are $\mathcal{H}$ market segments with mass $H(h)$ houses in segment h; the mass of houses in the entire Bay Area is normalized to one. Agent type θ is identified by search range $\tilde{\mathcal{H}}(\theta) \subset \mathcal{H}$. Search ranges are part of preferences, and type-θ agents never enjoy a house outside of $\tilde{\mathcal{H}}(\theta)$. We use the measure μ on the set of types Θ to count the number of agents of each type. The total number of agents is $\bar{\mu} = \sum_{\theta \in \Theta} \mu(\theta) > 1$. The clientele $\tilde{\Theta}(h)$ of segment h is the set of all agents who are interested in segment h, as defined in equation (1). The inventory and housing stock scanned by type θ are $\tilde{L}(\theta)$ and $\tilde{H}(\theta)$, respectively.

Agents live forever, discount the future at rate r, and receive quasilinear utility from a numeraire good and housing services. Agents only obtain housing services from a "favorite" house. After an agent moves into his favorite house in segment h, he obtains housing services $v(h) > 0$. Houses fall out of favor at rate $\eta(h)$ and then no longer provide housing services. Agents can put a house up for sale at no cost. Once the house is sold, agents search for a new house, again at no cost. Sellers incur a proportional cost c upon sale of a house. Matching in segment h occurs at the rate $\tilde{m}(B(h), L(h), h)$, where $\tilde{m}$ exhibits constant returns to scale in its first two arguments.[13]

How do agents decide on a favorite house within their search range $\tilde{\mathcal{H}}(\theta)$? Our approach is guided by the evidence in Section 3.2: interest in individual segments within a search range is proportional to segment inventory. We thus assume that agents are equally likely to "fall in love" with any house for sale in $\tilde{\mathcal{H}}(\theta)$. The number of buyers $B(h)$ in segment h is:

$$B(h) = \sum_{\theta \in \tilde{\Theta}(h)} \tilde{B}(\theta) \frac{L(h)}{\tilde{L}(\theta)}, \tag{6}$$

where $\tilde{B}(\theta)$ is the number of type-θ buyers. For a narrow type θ who considers only segment h, we have $L(h) = \tilde{L}(\theta)$, so this buyer contributes one-for-one to $B(h)$. To determine the contribution

[13] The matching function is increasing in the number of buyers and sellers and satisfies $\tilde{m}(0, L, h) = \tilde{m}(B, 0, h) = 0$. It is also 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 some segments because the properties there are more standardized, or because more open houses are available to view properties.

to $B(h)$ of a broad type θ, we multiply the number of these buyers by $L(h)/\tilde{L}(\theta)$, which represents the share of total inventory scanned by type θ that is in segment h.

Expression (6) is the counterpart to equation (4) in the one-segment reduced-form model from Section 4. In that model, the spirit of our buyer-flow assumption would be captured by the functional form $B^B(L) = qL$ for some exogenous constant q. In the current model, segments differ in their clientele structure. As a result, there are many segment-specific measures of the sensitivity of a segment's buyer pool to segment inventory, $q(h) = \sum_{\theta \in \tilde{\Theta}(h)} \tilde{B}(\theta)/\tilde{L}(\theta)$, that are all jointly determined in equilibrium.

Matching, Bargaining, and Equilibrium

Once a buyer and seller have 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, she pays the offer price. The seller receives the offer price net of the proportional cost c which goes to a real estate agent. The seller then starts to search, whereas the buyer moves into the house and begins to receive utility $v(h)$.

An equilibrium is a collection of agent choices such that each agent chooses optimally given the distribution of others' choices. 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. 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 no longer receive housing services from it, and (ii) all offers are accepted.

The model endogenously determines inventory shares $I(h)$ and turnover rates $V(h)$ for each segment. It also determines the number of searchers $\tilde{B}(\theta)$ of each type, which allow us to calculate segment-level search activity $\sigma(h)$ as defined by equation (2). The cross section of these observables is shaped by three distinct forces. Supply is represented by the rate $\eta(h)$ at which houses fall out of favor. Demand is captured by the clientele structure, the distribution of types $\mu(\theta)$. Finally, the segment-specific effect on match rates $\tilde{m}(.,.,h)$ represents differences in market frictions across segments. The next section describes how data on the observable endogenous variables $I(h), V(h)$, and $\tilde{B}(\theta)$ allow us to quantify these three forces.

5.1 Housing Demand and the Cross Section of Housing Markets

Our quantitative analysis proceeds in two steps. First, we show how the cross section of inventory, turnover, and search activity in the Bay Area reflects housing demand – in particular, the presence of broad searchers – as well as the other two exogenous forces capturing supply and market frictions. We provide additional evidence that validates and helps interpret the demand estimates. Section 6 then studies price formation and conducts counterfactuals.

Identification

The analysis in this section requires only the supply and demand parameters; it does not depend on the details of bargaining or the shape of the matching function. The intuition is as in Section 4. With a fixed number of houses and agents, the steady state distribution of agent states (searching for a house, listing one for sale, and owning without listing) is pinned down by house and agent flows, regardless of pricing. Moreover, matching frictions matter for agent flows only via the rates at which buyers find houses in a given segment, defined as $\alpha(h) = m(h)/B(h)$.

More formally, our identification result in Appendix F establishes a one-to-one mapping between two sets of numbers. The first set consists of the supply and demand parameters, $\eta(h)$ and $\mu(\theta)$, as well as the vector of house finding rates $\alpha(h)$. The second set consists of objects we observe in the data: the inventory share $I(h)$, the turnover rate $V(h)$, the relative frequencies of search ranges $\tilde{B}(\theta)$, and the *average* time it takes for a buyer to find a house.[14] We use this mapping to back out the vectors of $\eta(h)$, $\mu(\theta)$, and $\alpha(h)$.

We refer to the supply parameter $\eta(h)$ as the *instability* of the segment: in a more unstable segment, houses come on the market at a faster rate. To measure the *popularity* of a segment, we define the weighted number of agents who are interested in a house there:

$$\pi(h) := \frac{1}{H(h)} \sum_{\theta \in \tilde{\Theta}(h)} \mu(\theta) \frac{H(h)}{\tilde{H}(\theta)} = \sum_{\theta \in \tilde{\Theta}(h)} \frac{\mu(\theta)}{\tilde{H}(\theta)}. \tag{7}$$

Note that this definition weights types of agents by their share in the overall population (which is not directly observable), while search activity $\sigma(h)$ in equation (2) weights them by their share among searchers (which we could observe directly in the email alert data). Popularity is thus an exogenous determinant of demand for segment h; it only depends on the distribution of types $\mu(\theta)$. In contrast, $\sigma(h)$ is determined endogenously and depends on the equilibrium share of agents of each type that are searching at any point in time. We continue to count individuals who are potentially interested in many segments less towards the popularity of segment h than individuals who only like that particular segment.

Parameter Estimates

Table 4 summarizes our estimates of instability $\eta(h)$, popularity $\pi(h)$, and the house finding rate $\alpha(h)$. Panel A provides information on the distribution of the estimates across segments. Panel B reports correlations both among the estimates themselves as well as between estimates and observables. Here we compare variables across all segments, across all cities, and across

[14] While the equilibrium depends on the total number of buyers of each type, $\tilde{B}(\theta)$, we only observe them up to a constant: the email alert data allow us to infer the relative number of each type, $\tilde{B}(\theta)/\bar{B}$, but we have no information on the overall number of buyers, $\bar{B}$. As an additional target moment, we thus set the average match rate for buyers to 20 percent per month, corresponding to 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 is thus not important for most of our results.

segments within the city of San Francisco.

Instability $\eta(h)$, which captures the frequency of moving shocks in a segment, tracks turnover almost exactly, and its moments in Table 4 are essentially the same as those reported for the turnover rate in Table 2. The result follows from the flow condition $\eta(h)(1-I(h)) = V(h)$ for each segment, together with the summary statistics in Table 2. Inventory shares are so small — their 75th percentile is at 1.5% — that $\eta(h) \approx V(h)$. Intuitively, because a house remains on the market for a much shorter time than it is occupied by an owner, turnover is almost entirely accounted for by the frequency of moving shocks.

Popularity at the segment level, $\pi(h)$, ranges between 0.2 and 2.4, with an inter-quartile range of 0.80 to 1.15. The fact that popularity is below one for many segments is indicative of the importance of partial integration. If segments were either perfectly segmented or perfectly integrated, the number of (weighted) people interested in each segment would be larger than the number of houses, and popularity would have to be above one. However, segments that are only considered by relatively broad searchers can have a popularity less than one, because nobody that lives there has a particular preference for that location. Popularity also comoves strongly with search activity at all levels of aggregation: in equilibrium, more popular segments generally have more (weighted) searchers per house.

The Cross Section of Cities

The middle part of Table 4's Panel B shows how the forces from Figure 4 quantitatively account for market activity in the cross section of cities. Instability generates comovement of inventory and turnover – in fact, it is almost perfectly correlated with both. At the same time, more popular segments see more searchers per house and lower inventory. Differences in popularity are thus a key force behind the downward-sloping Beveridge curve across cities. The positive impact of popularity on turnover is muted by the fact that more popular cities are also more stable.

In addition to the supply and demand effects that were already present in the simple model in Section 4, the quantitative model also identifies significant differences in matching frictions across cities. Indeed, we find that in cities with low inventory and turnover but many searchers, matching is particularly slow (that is, $\alpha(h)$ is low).[15] Intuitively, cities with many searchers must be popular. Slow matching explains why this popularity does not translate into high turnover.

The Quantitative Importance of Partial Integration on Estimates of Segment Popularity

To develop a summary statistic for the quantitative effect of broad searchers, we compare the actual economy – which exhibits partial integration – with a hypothetical perfectly segmented benchmark economy. We construct the latter by changing demand parameters so as to remove integration, holding all other parameters fixed. At the same time, we require that the bench-

[15] To illustrate how the matching technology affects $\alpha(h)$, consider a Cobb-Douglas matching function $m(h) = \bar{m}(h)\, B(h)^{\delta}\, L(h)^{1-\delta}$, which implies $\log \bar{m}(h) = \delta \log \alpha(h) + (1-\delta) \log (V(h)/I(h))$.

Table 4: Quantitative Results from Estimating the Model

Panel A: Estimated parameters vs. hypothetical segmentation

	Estimated parameters			Hypothetical perfectly segmented benchmark	
	$100 \times \eta(h)$	$\pi(h)$	$\alpha(h)$	$\hat{\sigma}(h) - \sigma(h)$	$\hat{\pi}(h) - \pi(h)$
Mean	0.24	1.00	0.19	0.00	0.02
Q25	0.16	0.80	0.07	−0.17	−0.13
Q50	0.21	1.00	0.13	−0.04	0.03
Q75	0.29	1.15	0.23	0.12	0.21

Panel B: Correlation between estimated parameters and data moments

	$\eta(h)$	$\pi(h)$	$\alpha(h)$	$I(h)$	$V(h)$	$\sigma(h)$	$\log(p(h))$
Across segments							
$\eta(h)$	1	−0.01	0.24	0.93	1.00	0.13	−0.48
$\pi(h)$		1	−0.53	−0.03	−0.01	0.73	0.07
$\alpha(h)$			1	0.27	0.23	−0.43	−0.22
$\hat{\sigma}(h) - \sigma(h)$	0.44	0.10	−0.04	0.46	0.44	0.28	−0.16
Across cities							
$\eta(h)$	1	−0.23	0.52	0.93	1.00	−0.24	−0.55
$\pi(h)$		1	−0.62	−0.27	−0.23	0.69	0.36
$\alpha(h)$			1	0.57	0.52	−0.70	−0.65
$\hat{\sigma}(h) - \sigma(h)$	0.37	−0.01	−0.00	0.39	0.37	0.09	−0.11
Within San Francisco							
$\eta(h)$	1	0.16	−0.01	0.89	1.00	0.26	−0.35
$\pi(h)$		1	−0.59	0.18	0.16	0.86	−0.04
$\alpha(h)$			1	−0.17	−0.01	−0.57	0.07
$\hat{\sigma}(h) - \sigma(h)$	0.67	0.16	−0.12	0.67	0.67	0.32	−0.14

Note: The estimated parameters from the model exactly match three moments from the data: the inventory share $I(h)$, the turnover rate $V(h)$ and the relative frequency of search ranges $\beta(\theta)$. The total number of buyers $\bar{\mu} - 1$ is set to match a 5 month average buyer search time.

mark economy still matches the same observed inventory shares and turnover rates as the actual economy. This pins down a unique vector of hypothetical demand parameters $\hat{\mu}$.

The benchmark economy can be viewed as the model estimated by an econometrician who observes the inventory share $I(h)$, turnover rate $V(h)$, and buyer match rate $\alpha(h)$ by segment, but who does not have information on integration and proceeds by assuming that the economy is perfectly segmented. This misspecification will lead the econometrician to infer the wrong demand parameters, and hence incorrect measures of popularity $\hat{\pi}(h)$. We thus use the distribution of the difference $\hat{\pi}(h) - \pi(h)$ to assess the size of the specification error from ignoring partial integration.[16]

Table 4 shows that the error incurred by an econometrician who ignores partial integration is large. Quantiles for the difference in popularities $\hat{\pi}(h) - \pi(h)$ across segments are reported in the rightmost column in Panel A. The interquartile range of this difference is of the same order of magnitude as the interquartile range of the estimated parameter $\pi(h)$ itself. The reason is that the range of estimated popularities is much narrower for the hypothetical fully-segmented benchmark economy: the 25th percentile for $\hat{\pi}(h)$ is at 1.01, while the 75th percentile is at 1.03.

Intuitively, the econometrician infers too little dispersion in popularity across segments because he ignores the endogenous response of broad searchers to market conditions. A model with narrow searchers must explain observed market and search activity entirely via narrow searchers' demand, as opposed to broad searchers chasing inventory. In the actual economy, some unpopular and unstable segments attract very few narrow searchers, but nevertheless see substantial search activity due to attention from broad searchers who are drawn to the segment's high inventory. Ignoring partial integration therefore overstates the popularity of such unstable high-inventory segments. As discussed in Section 4, overestimating the popularity of a segment generates misleading predictions on the quantity-effects of additional construction, changes in zoning regulation, public transit investments, and other place-based policies. We discuss below how it will also bias estimates of the effects of such policies on equilibrium transaction prices.

Broad Searchers and the Beveridge Curve within Cities

We next quantify the contribution of broad searchers to the shape of the Beveridge curve within cities. In particular, the perfectly segmented benchmark economy in the previous section predicts a different cross section of search activity, $\hat{\sigma}(h)$, and hence a different Beveridge curve. We can therefore use the difference in search activity $\hat{\sigma}(h) - \sigma(h)$ to summarize the effect of partial integration on the Beveridge curve. If integration was not important, then the actual and hypothetical Beveridge curves would coincide.

The key difference between the benchmark and actual Beveridge curves is that the slope of the latter partially reflects the response of broad searchers to inventory. In high inventory segments,

[16] The other parameter estimates are not materially affected by misspecification. We have seen that the parameter $\eta(h)$ closely tracks turnover – this result follows from the flow conditions regardless of the clientele patterns. Moreover, given the equilibrium conditions and the fact that the matching function remains unchanged, the benchmark economy predicts the same number of buyers and buyer match rates $\alpha(h)$ as the actual economy.

broad searchers crowd out narrow searchers which increases the overall number of searchers per house. As a result, the actual Beveridge curve is steeper than the benchmark curve in (σ, I)-plane and the difference $\hat{\sigma}(h) - \sigma(h)$ should be positively correlated with inventory in integrated areas.

Table 4 shows that the contribution of broad searchers to the slope of the Beveridge curve within San Francisco is large. The correlation coefficient of 0.67 between $I(h)$ and $\hat{\sigma}(h) - \sigma(h)$ implies that a one standard-deviation increase in inventory increases $\hat{\sigma}(h) - \sigma(h)$ by 0.45 standard deviations of the search activity measure $\sigma(h)$ itself. Within San Francisco, a perfectly segmented benchmark that ignores broad searchers thus generates a Beveridge curve that is much flatter then the true Beveridge curve we measure. Across cities, in contrast, broad searchers matter less: ignoring the presence of broad searchers adds only about 10 percent worth of search activity per unit of inventory. This finding is consistent with the results in Section 3.1 that showed less integration across cities.

What exogenous forces generate the differences in inventory that attract the broad searchers? In principle, differences in instability, popularity, or matching frictions could all generate differences in inventory across segments. Quantitatively, however, the flow of broad searchers is directed to a large extent by differences in instability. Indeed, the correlation of $\hat{\sigma}(h) - \sigma(h)$ with $\eta(h)$ is much larger than that with $\pi(h)$ or $\alpha(h)$. We can therefore sum up the key mechanism as follows: in more unstable segments, more houses come on the market. If these segments are part of an integrated area, then broad searchers are attracted to the higher inventory and crowd out narrow searchers, generating an upward-sloping Beveridge curve.

San Francisco Bay Area vs. Other Metro Areas

The key intuition described above is that the endogenous flow of broad searchers to high-inventory segments within their search ranges can induce an upward-sloping Beveridge curve across market segments with a large common clientele. Our search data reveal that most housing search in the Bay Area is along city lines: 61% of search queries specify a city as the finest geographic unit. This implies a large common search clientele within Bay Area cities, and rationalizes the upward-sloping Beveridge curve across segments within most cities. However, in other parts of the United States, the geographic and political units that are jointly searched can potentially differ. For example, in Massachusetts, "cities" are much smaller political units, and different cities such as Cambridge and Sommerville are probably regularly searched jointly. This generates the potential for an upward sloping Beveridge curve across segments in those units. On the other hand, in New York City, searchers are unlikely to consider the entire city, and might only search parts of different boroughs jointly. This suggests we are less likely to see an upward sloping Beveridge curve across all New York City submarkets than we are to see it across submarkets within the same borough.

5.2 Further Evidence on Housing Demand

Our estimation infers housing demand from data on market and search activity. To help with interpreting these demand estimates, we now relate our estimates of popularity – our segment-level summary statistic for demand – to observable characteristics of segments. We also relate search breadth – our key individual-level statistic – to searcher demographics. Finally, we compare model-implied household flows between segments to data on actual moves, providing an over-identifying restrictions test on the structure of our model.

Popularity and Segment-Level Observables

What makes a segment popular? And how can policy makers and researchers identify popular segments in the absence of detailed search data? One characteristic associated with popular cities appears to be the service flow from properties in the segment, which is a proxy for their quality. In particular, Table 4 shows a 36% correlation between log price and popularity across cities. More expensive cities thus have a larger average clientele interested in living there.

When comparing segments within cities, the correlation between price and popularity is much weaker. To understand which other characteristics are correlated with segment popularity, we construct the housing-stock-weighted average popularity of all segments in a zip code. We then use a simple regression to relate this zip code-level popularity measure to characteristics observable at the zip code level: school quality, the availability of restaurants and bars, crime levels, and weather conditions.[17]

The results in Table 5 show that the availability of bars and restaurants is the most important correlate of popularity at the zip code level: a one-standard-deviation increase in the number of restaurants and bars is associated with an increase in popularity of 0.14, or 0.53 standard deviations, and the number of bars and restaurants explains 25.3% of the across-zip code heterogeneity in popularity. The strong correlation between the number of restaurants and popularity is true both unconditionally, as well as when comparing zip codes within cities. Other zip code characteristics also matter, but they are quantitatively less important. A one-standard-deviation increase in school quality increases popularity by 0.042, whereas a one-standard-deviation increase in violent crime reduces popularity by 0.038. The weather in a zip code is also correlated with popularity: more rain, more extreme hot weather, and more extreme cold weather all reduce popularity, with more extreme hot weather having the largest effect. These correlations suggest that using zip code-level characteristics such as the number of restaurants and bars might allow policy

[17] We measure school quality as the student-weighted average Academic Performance Index (API) across all schools in a zip code, as reported by the California Department of Education. To measure the availability of restaurants, we divide the number of establishments with SIC code 58 (Eating & Drinking Places) by the number of housing units. Crime levels are measured on a scale of 0-100, as provided by Bestplaces.net. To measure weather, we calculate the total number of inches of rain, the total number of cooling degree days, and the total number of heating degree days, as reported by Melissa Data. Heating Degree Days (HDD) and Cooling Degree Days (HDD) are measures of how far, and for how long, temperatures deviate from 70 Fahrenheit. For example, every day the temperature is at 65, counts as 5 HDD.

Table 5: Drivers of Zip Code Level Popularity

	Coefficient	Effect of 1 s.d. increase on zip code popularity	R^2 in univariate regression
Restaurants & Bars (Per 100 housing units)	0.147***	0.141	25.3%
	(0.027)		
School Quality (Average API / 100)	0.052**	0.042	2.6%
	(0.026)		
Violent Crime (Scale 0-100)	-0.002*	-0.038	0.4%
	(0.001)		
Rain (Inches)	-0.001**	-0.018	0.0%
	(0.000)		
Cooling Degree Days (k days)	-0.108**	-0.033	2.3%
	(0.050)		
Heating Degree Days (k days)	-0.054*	-0.025	0.2%
	(0.026)		

Note: The first column shows the coefficient estimates from a multivariate OLS regression of zip code-level popularity on observable zip code characteristics. N = 183. Robust standard errors in parentheses. Significance levels: * (p<0.10), ** (p<0.05), *** (p<0.01). The multivariate regression has an R^2 of 32.6%. The second column shows the implied effects of a one-standard-deviation increase in the characteristics on zip code popularity. The across-segment standard deviation of popularity is 0.263. The final column shows the R^2 from univariate regressions of popularity on each of the zip code level characteristics.

makers who do not observe housing search data to better target localized housing policies (such as additional construction) to particularly popular segments.

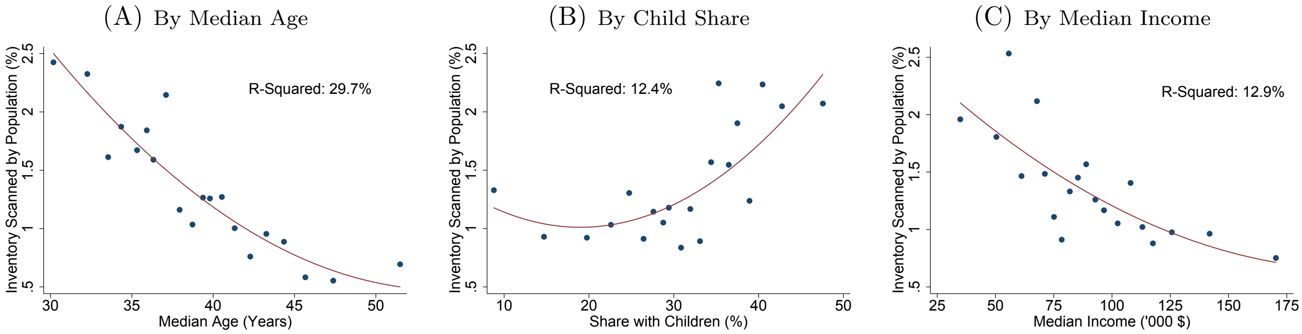
Search Breadth and Demographics

Who are the broad searchers? While we do not directly observe demographics for the individuals setting email alerts in our data, our model estimates imply a distribution of search breadths for households living in equilibrium in each segment. We can thus compare the average search breadth of people living in a zip code, measured by the share of Bay Area inventory scanned, with demographic information from the 5-year estimates of the 2012 American Community Survey.[18]

We focus on residents' age, income, and the presence of children. Figure 5 presents binscatter plots to show the relationship between these demographic measures and search breadth. In Panel A, we group our 191 zip codes by the median age of the inhabitants. There is a strong negative relationship between age and search breadth. People living in zip codes with a median age of 30 search five times as much inventory, on average, as people living in zip codes with a median age of 50, and differences in median age explain almost one-third of the across-zip code heterogeneity in the average search breadth of residents. Panel B shows that the average search breadth of households living in zip codes with many children is higher, and Panel C shows that people living in zip codes with higher median incomes have smaller search ranges. All these relationships are

[18]The average search breadth of people *living* in a zip code in equilibrium can be different from that of people *searching* in a zip code, which we explored in Figure 3.

Figure 5: Search Breadth and Demographics



Note: Figure shows binscatter plots of the average search breadth of people living in a zip code in equilibrium (measured as the average share of Bay Area inventory scanned), separately by zip code demographics. We show search breadth by median age (Panel A), by the share of population with children (Panel B), and by median zip code income (Panel C).

true both unconditionally, and conditional on the other two demographic measures.

While our paper takes individuals' search ranges as the primitive representation of their preferences rather than deriving them as part of an optimization problem, there are a number of ways to rationalize the observed correlations between search breadth and searcher characteristics. For instance, to the extent that older and richer people perceive a higher marginal value of time (for example, because they earn a higher wage), this would reduce the distance that these agents are willing to commute to work, and naturally lead to a narrower optimal search range. We view it as a promising area for future work to endogenize buyer search ranges.

Medium-Run Endogeneity of Search Ranges and Segment Characteristics

The focus of our paper is to explore the steady-state relationships between housing search patterns and segment-level equilibrium housing market outcomes. This allows us to study how, at any point in time, the distribution of housing market activity across segments is shaped by the interaction of heterogeneous search clienteles, and how these housing market outcomes might respond to small shocks to local supply or demand conditions. Our set-up is not designed to study a potential endogenous medium-run feedback between changes in segment-level searcher clientele, segment characteristics, and housing market activity, that might arise, for example, due to the gentrification of neighborhoods. We think that the study of such medium-run transitions between different steady states is an interesting area for further research.

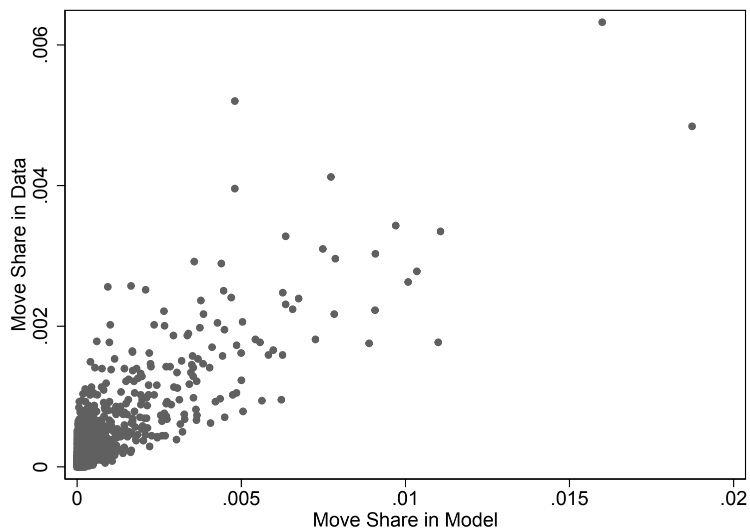
Moving Patterns in Model vs. Data

While our estimation targets only moments of the cross section of housing market activity, the model also has implications for the flow of households between segments. We now confront those implications with data on actual household moves. Such an evaluation of model predictions that were not targeted in the estimation provides a joint test of our model assumptions and the quality of our search data. To measure the flow of households between segments, we use a sample of all individuals who moved to a new Bay Area address between May 2012 and October 2012. The data come from Acxiom, a marketing analytics company that compiles this list of movers from "Change of Address" notices filed with the United States Postal Service.

Our sample contains 96,170 individuals moving to a new address in one of the 191 zip codes in our sample; for these individuals, we have information on their new and previous addresses. We focus on movers who had previously also lived in one of the 191 Bay Area zip codes, about 75% of the sample. We then compute the shares of movers between each of the 191 $\times$ 191 (directed) zip code pairs in the data and compare them to the shares of movers predicted by the model, which we obtain by aggregating model-implied segment-to-segment flows to the zip code level.

Figure 6 shows a scatter plot of the mover share in the data and the model at the (directed) zip code-pair level. The correlation coefficient is 82%. The high correlation is not only driven by households moving within the same zip code: when we exclude such moves, the correlation drops only slightly, to 75%. We conclude that moving patterns provide additional support for our quantitative account of Bay Area housing market dynamics.

Figure 6: Moving Activity: Model vs. Data



Note: Figure shows the share of total moves predicted by the model (horizontal axis) vs. the moving share in the data (vertical axis) for each (directed) pair of zip codes.

6 Prices and Spillovers

In this final section, we explore how the forces in our model relate to equilibrium prices across segments. We also explore how the clientele structure in the Bay Area shapes the responses of different segments to housing market shocks, such as the influx of new narrow searchers as a result of the gentrification of neighborhoods.

6.1 Equilibrium Prices and Frictional Discounts

Our model captures two distinct housing market frictions. The first is search: owners whose house falls out of favor spend time first looking for a buyer and then for a new house. During this time they forgo the flow utility of living in their favorite house. The second friction is the transaction cost paid upon sale. In equilibrium, both of these costs are capitalized and reduce the house price relative to a frictionless model: every buyer takes into account that both he and all potential future buyers may have to sell and hence search and pay transaction costs.

Appendix E derives a convenient approximate formula for the equilibrium price in a segment, $p(h)$, which highlights how the resulting *frictional discount* reflects both frictions:

$$p\left(h\right) \approx \frac{v(h)}{r}\left(1 - I\left(h\right)\right)\frac{r}{r + cV\left(h\right)}. \tag{8}$$

In a frictionless market, matching is instantaneous, so that there is no outstanding inventory ($I\left(h\right) = 0$), and there are no transaction costs ($c = 0$). As a result, the price simply reflects the present value of future housing services $v\left(h\right)/r$.

Search and transaction costs modify the frictionless price $v(h)/r$ by first reducing housing services proportionately by $I(h)$ and then increasing the discount rate to $r + cV(h)$. The inventory share measures the price discount due to search frictions, and captures the fact that no household obtains housing services while a property is listed for sale. This discount is zero when matching is instantaneous. From Table 2, the size of the search discount is typically a few percentage points. The interest rate does not matter for its size (at least approximately) because time on market is fairly short. The second fraction, $r/(r + cV)$, measures the present value of transaction costs: it is zero if there is no turnover or if selling houses is costless. Here the interest rate is important: if future transaction costs are discounted at a lower rate, then the discount is larger.

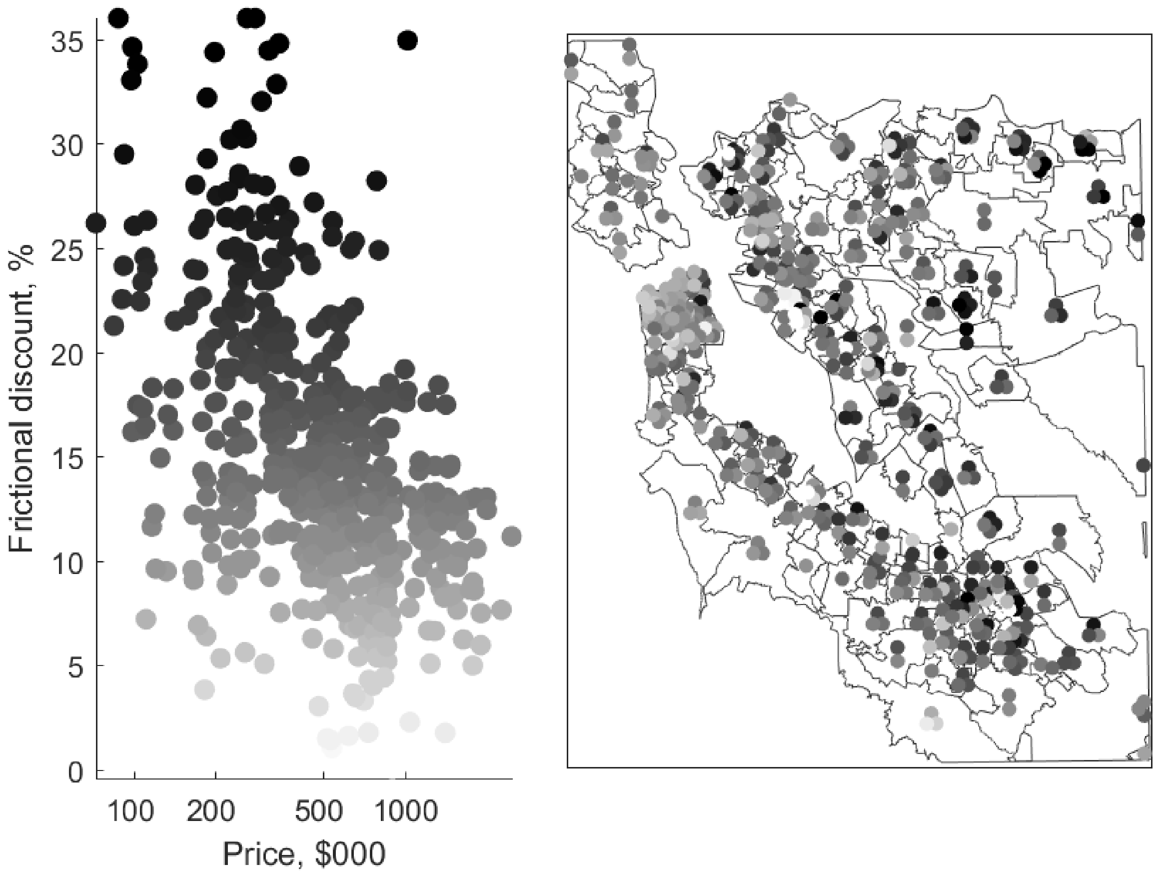
Frictional Discounts by Segment

We now ask how market frictions quantitatively affect house prices across segments. To compute the frictional discount, we need an estimate of the present value of future housing services in a segment. If we assume a real interest rate of 1 percent and a transaction cost equal to 6 percent of the resale value of the house, we can use equation (8) to back out the utility values $v(h)$ such that the model exactly matches the cross section of transaction prices, inventory shares, and turnover.

Figure 7 shows the results. The left panel plots transaction prices by segment against the frictional discounts, stated as a percentage of frictionless price. The right panel shows the geographic distribution of the frictional discounts.

There are two notable results. First, frictional discounts are large. The median discount is 14 percent and the 90th percentile discount is 24 percent. From Table 2 and the approximating formula above, both search and transaction costs contribute to this result, usually in the same

Figure 7: Frictional Discounts



Note: Figure shows frictional discounts across segments, stated as a percentage of the frictionless price $v(h)/r$. The left panel shows the mean segment price versus the segment frictional discount in percent of frictionless price. Color coding reflects the size of the frictional discount. The right panel shows a map with dots for each segment in same color as in left panel. Dots for segments within the same zip code are arranged clockwise by price with the lowest-priced segment at noon.

direction since inventory shares and turnover rates are highly correlated. However, transaction costs are quantitatively more important. While search costs generate frictional discounts up to 6 percent, the capitalized value of transaction costs is what leads to double-digit discounts. Intuitively, the relatively small frictional discount due to search costs is because inventory shares are relatively low, and houses are therefore occupied most of the time.

The second result is that frictional discounts differ widely by segment, often within the same zip code. Table 3 shows that inventory share and turnover rate exhibit about the same amount of variation within and across zip codes. The search and transaction costs inherit these properties, respectively. In poor (low-price) segments with high turnover and high inventory, both search and transaction costs are high. As a consequence, prices are significantly lower than they would be in a frictionless market. In rich (high-price) segments with low turnover and low inventory, frictional discounts are still significant, but they are considerably smaller.

6.2 Comparative Statics

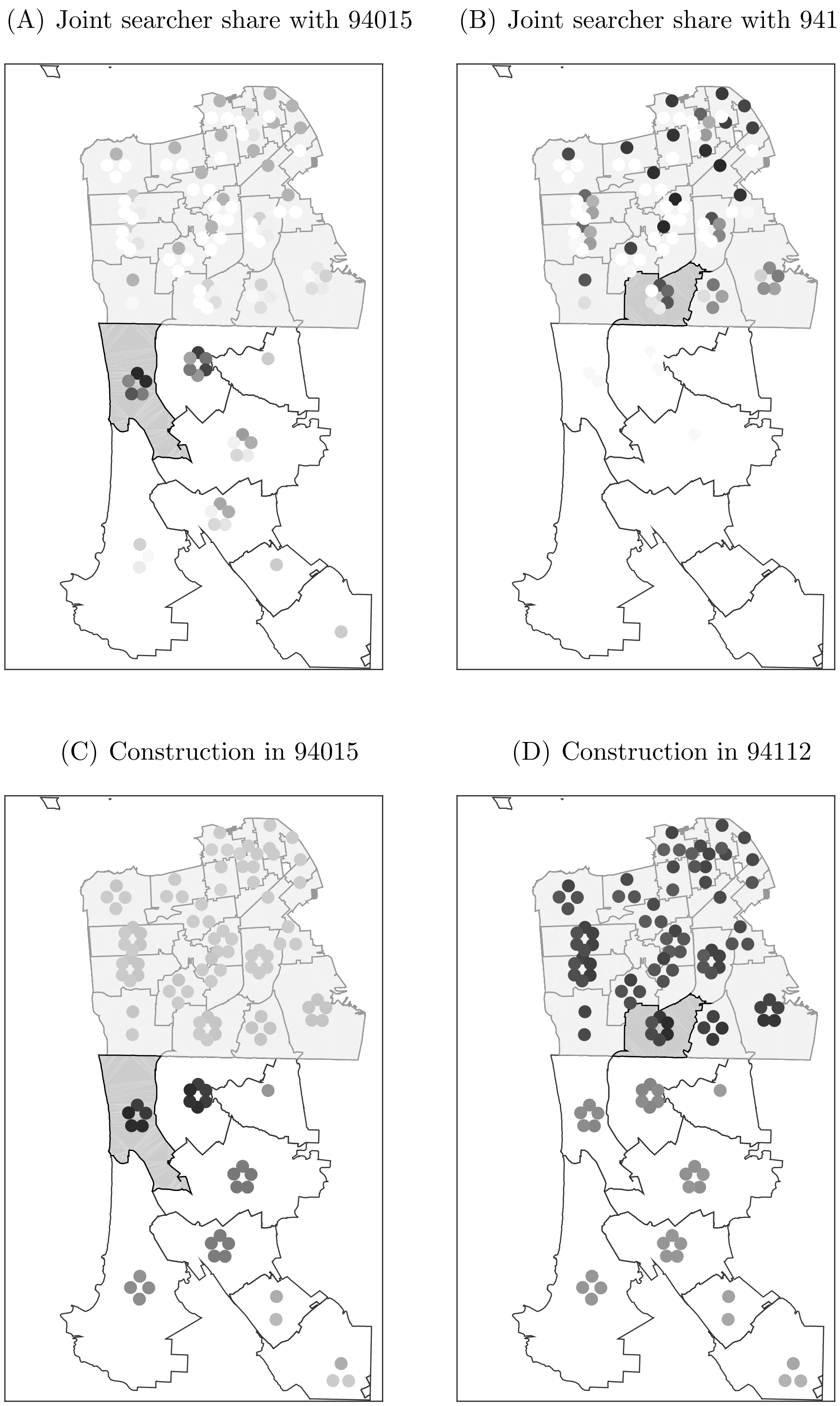
In this section, we perform comparative static exercises to show how clientele patterns matter for the transmission of localized shocks across housing market segments. Motivated by recent debates about housing shortages in parts of the San Francisco Bay Area, we ask what happens when new construction adds houses in different neighborhoods. We compare two neighborhoods that are similar in size and price, but differ in their clientele patterns. For each of these neighborhoods, we recompute the steady state equilibrium under the assumption that 1,000 additional houses are added in that neighborhood. Since both neighborhoods consist of several segments, we allocate new houses to segments in proportion to their housing stock.

The first neighborhood is zip code 94015 in Daly City, a suburb right outside the San Francisco city limits. It contains about 11,000 houses with an average value of $480K. The average inventory is 105 houses; at our estimated parameters, this inventory is considered by 233 active searchers. We choose Daly City because narrow buyers are prevalent there. Among agents interested in Daly City, the share of narrow types – defined here as those who search in five segments or less – is 28%. In contrast, the share of broad types – defined as those who search in 20 segments or more – is only 15%. The second neighborhood is San Francisco's Outer Mission, zip code 94112. For comparability with 94015, we select only the cheapest three segments in this zip code; we thereby obtain about the same total housing stock and average price. However, the population of searchers is quite different. There are 1,053 searchers looking at an average inventory of 84 houses. Moreover, among agents interested in the Outer Mission, the share of narrow types as defined above is only 3%, which is tiny compared to the 74% share of broad types.

Figure 8 illustrates how broad searchers integrate the two neighborhoods with the rest of the San Francisco Peninsula. Panel A shows the *joint searcher share* of each segmented with the Daly City zip code 94015, that is, the share of 94015 searchers who also search that segment. Darker colors indicate more integration between 94015 and the respective segment. Zip code 94015, which is the area shaded in dark grey, is most integrated with its adjacent zip code 94014. There are also weaker connections to towns to the south as well as the city of San Francisco to the north, shaded in light gray. The San Francisco segments that are searched jointly with Daly City tend to be the less expensive ones within a zip code, indicated by the top twelve o'clock dots of the circle for that zip code. Panel B of Figure 8 maps joint searcher shares with the Outer Mission. There is strong integration with cheap segments within all of San Francisco, whereas the connection to more expensive segments as well to segments outside the city limits is much weaker.

Panels C and D of Figure 8 show the effects of new construction on inventory shares across segments. As expected from our simple model in Section 4, construction increases inventory. At the same time, turnover rates remain essentially unchanged, since houses come on the market at the same rates $\eta(h)$ as before. The time to sell a house thus moves proportionally with inventory. Moreover, price effects follow from equation (8): while changes in the transaction cost discount

Figure 8: Joint Searcher Share and Inventory Response to Construction



Note: Panels A and B show the joint searcher share with 94014 and 94112, respectively. Panels C and D show the percent change in inventory share in response to the construction of 1,000 houses in zip code 94015 (Daly City) and the three cheapest segments of zip code 94112 (San Francisco Outer Mission).

are negligible, changes in inventory shares move the search discount. The bottom panels of Figure 8 thus also show the distribution of increases in time on market and declines in price. Darker dots indicate stronger responses in inventory shares (as well as time on market and price changes).

For both zip codes, we see the strongest responses in the zip codes where construction takes place. The main result is that a shock to a less integrated neighborhood has locally larger effects that spread less widely to other segments. In particular, the effects of construction in Daly City 94015 are disproportionately felt in 94015 itself, as well in the adjacent zip code 94014. Here, inventory shares increase by more than 30 percent. The effect on the typical segment in the city of San Francisco to the north, shaded in light gray, is much smaller at around 8 percent. In contrast, construction in the Outer Mission has sizeable spillover effects throughout the entire city of San Francisco. In the segments where construction takes place, the average increase in inventory is only about 18 percent. However, in all San Francisco neighborhoods, the increase is 13 percent or more. Noticeably, there are similar-sized increases even in the more expensive segments in San Francisco, which do not share many direct broad searchers with the Outer Mission (see Panel B). These are generated by chains of connections through other searchers who integrate cheaper and more expensive segments within a zip code. Since changes in inventory translate directly into frictional discounts, price responses are also smaller and more diluted for shocks that hit the Outer Mission, but larger and more concentrated for shocks to Daly City.

Differences in integration are also reflected in the surplus from housing that accrues to different types of agents. Indeed, construction in Daly City *lowers* the number of narrow types who are happy homeowners. While it has become easier to find a house, it is also harder to sell it, which may lower the overall speed at which unhappy narrow agents can again become happy owners. The effect is already present if there are only narrow searchers, as in the perfectly segmented toy example of Section 4; here it is compounded by an inflow of broad searchers from neighboring zip codes. After construction in Daly City, the number of narrow types who are "unhappy" — that is, narrow types who do not receive surplus from a house — increases by 8 percent. At the same time, broad types benefit: the number of unhappy broad searchers interested in Daly City declines by 4 percent. For the Outer Mission, in contrast, the numbers of unhappy agents declines by 7 percent for both narrow and broad types interested in living there.

Information on clientele patterns is also critical for assessing the impact of construction on Bay Area aggregates. Consider a researcher who is interested only in Bay Area aggregates, and hence calibrates a model with one segment to the entire Bay Area. If our model is correct, and the researcher calibrates the number of agents to match average buyer search time, he finds the same number of buyers as we do, which is approximately the same as the number of houses in inventory. In the absence of information about clientele patterns, his predicted response to the construction of 1,000 houses is then that 1,074 more houses flow to inventory, and that welfare would decline as a result of the construction. On aggregate, a similar buyer search time and seller time on market imply that there is no shortage of housing. In contrast, our model says that building in

Daly City or the Outer Mission increases aggregate Bay Area inventory by only 1,017 and 847 houses, respectively.[19] This is because those areas have local housing shortages that allow the absorption of more houses; as a result, depending on the construction costs, welfare may increase from construction in these areas.

The experiments also illustrate the dangers of treating disaggregated units such as segments and zip codes as independent. Indeed, as highlighted above, a perfectly segmented economy would not be able to distinguish the spillover effects of more versus less integrated neighborhoods. It would also provide misleading summary information about the demand for housing. For example, at our estimated parameters, mean popularity is 1.03 for Daly City and 0.75 for the Outer Mission, which features more broad searchers. In contrast, in the perfectly segmented benchmark economy introduced in Section 5, mean popularities are 1.008 and 1.014 for Daly City and the Outer Mission, respectively. An econometrician using a perfectly segmented economy to assess the effects of new construction would therefore incorrectly expect similarly sized increases in local inventory in response to construction in the Outer Mission and Daly City. In addition, the econometrician would predict no spillovers to other segments for construction in either market.

7 Conclusions and Segmented Search in Other Markets

Most search markets feature competition between broad and narrow searchers. We show that observing the structure of these search clienteles is important for understanding the forces behind equilibrium market outcomes such as the shape of the Beveridge curve, as well as the response of different market segments to shocks. We also demonstrate how data from online search behavior can allow researchers to overcome the challenge of measuring clientele patterns. We expect that similar data from online services such as Facebook, LinkedIn, Tinder, ZipRecruiter, and Indeed will allow researchers to measure the clientele structure in other search markets, from dating to job search (e.g., see Bailey et al., 2017, 2018a,b). This will improve our understanding of both the cross sectional patterns across submarkets, as well as the response of these markets to shocks.

While our analysis highlights the importance of understanding the interaction of broad and narrow searchers in the housing market, our insights are likely to also be important in other search markets. For example, Treasury securities are sold in over-the-counter search markets. In these markets, some investors might be particularly interested in purchasing Treasuries of certain maturities (e.g., pensions funds engaging in duration matching might only buy long-duration Treasuries), but there might be other buyers, such as hedge funds, that consider a broader range of maturities. Understanding the segmentation of the buyer clientele, and the interaction of broad and narrow investors at different maturities, is important for determining the optimal maturity

[19] Interestingly, the local increase in inventory is quite small: within San Francisco, construction in Daly City or the Outer Mission increases inventory by 76 and 148 houses, respectively. Much of the overall increase is accounted for by small increases of one or two houses per segment across most of the Bay Area. This occurs even in the absence of broad searchers who search the entire area, via chains of connections by more narrow searchers.

structure of newly issued debt. For example, monetary policy interventions, such as the maturity extension program ("Operation Twist"), which aimed at flattening the yield curve by selling short-maturity Treasuries and buying long-maturity Treasuries, will be less effective in the presence of "broad" investors who are indifferent between buying Treasuries with a wide range of maturities (see Swanson, 2011; Greenwood and Vayanos, 2014).

Similarly, the degree of segmentation in labor market search is substantial and time-varying. For example, recent research has documented that job seekers in areas with depressed housing markets apply for fewer jobs that require relocation, because of the difficulties with selling underwater homes (see Brown and Matsa, 2016). Our model allows researchers to understand the extent to which the resulting increase in regional segmentation of labor search contributes to a decline in the ability of labor flows to facilitate regional risk sharing. In addition, labor market segmentation across industry and occupational groups has increased as a result of the specialization of human capital, with many vacancies only attracting applications from a small set of highly specialized job seekers. The mechanisms highlighted in this paper show how such changes in segmentation influence the labor market effects of immigration, and affect the efficacy of policies such as targeted job training programs (e.g., Peri and Sparber, 2009).

The dating market provides a further setting in which some searchers with very narrow preferences interact with other searchers who are less particular about the characteristics of their preferred match. More recently, the rise of online dating services has increased the ability of narrow searchers to target their search to their particular preferences. An interesting question is whether the resulting increase in segmentation of the dating market has contributed to longer "times on market" and the increase in the age of marriage.

References

Albrecht, James, Axel Anderson, Eric Smith, and Susan Vroman, "Opportunistic matching in the Housing Market," *International Economic Review*, 2007, *48*, 641–664.

Anenberg, Elliot and Patrick Bayer, "Endogenous Sources of volatility in housing markets: the joint buyer-seller problem," *National Bureach of Economic Research*, 2014, *18980.*

Bailey, Michael, Eduardo Dávila, Theresa Kuchler, and Johannes Stroebel, "House price beliefs and mortgage leverage choice," Technical Report, National Bureau of Economic Research 2017.

_ , Rachel Cao, Theresa Kuchler, Johannes Stroebel, and Arlene Wong, "Social Connectedness: Measurement, Determinants, and Effects," *Journal of Economic Perspectives*, 2018, *32* (3), 259–80.

_ , Ruiqing Cao, Theresa Kuchler, and Johannes Stroebel, "The economic effects of social networks: Evidence from the housing market," *Journal of Political Economy*, 2018, *126* (6), 2224–2276.

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.

Brown, Jennifer and David A Matsa, "Locked in by Leverage: Job Search during the Housing Crisis," Technical Report, National Bureau of Economic Research 2016.

Burnside, Craig, Martin Eichenbaum, and Sergio Rebelo, "Understanding booms and busts in housing markets," *Journal of Political Economy*, 2016, *124* (4), 1088–1147.

Davis, Morris and Stijn Van Nieuwerburgh, "Housing, finance and the macroeconomy," *Handbook of Urban and Regional Economics, Edited by Gilles Duranton, J. Vernon Henderson and William C. Strange*, 2015, *5*, 753–811.

Diaz, Antonia and Belen Perez, "House prices, sales, and time-on-market: a search-theoretic framework," *Journal of International Economics*, 2013, *53*, 837–872.

Einav, Liran, Theresa Kuchler, Jonathan Levin, and Neel Sundaresan, "Assessing sale strategies in online markets using matched listings," *American Economic Journal: Microeconomics*, 2015, *7* (2), 215–47.

Forbes, "Four More Years? No Quick End In Sight For The U.S. Housing Shortage," 2016.

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

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

Greenwood, Robin and Dimitri Vayanos, "Bond supply and excess bond returns," *Review of Financial Studies*, 2014, *27* (3), 663–713.

Guren, Adam and Tim McQuade, "How do foreclosures exacerbate housing downturns?," 2014.

Guren, Adam M, "House Price Momentum and Strategic Complementarity," *Forthcoming Journal of Political Economy*, 2017.

Halket, Jonathan and Matteo Pignatti, "Homeownership and the scarcity of rentals," *Journal of Monetary Economics*, 2015, *76*, 107–123.

Han, Lu and Will Strange, "The microstructure of housing markets: search, bargaining, and brokerage," *Handbook of Regional and Urban Economics, Edited by Duranton, G., Henderson J.V., Strange, W.C.*, 2015, *5*, 813–886.

Head, Allen, Huw Lloyd-Ellis, and Hongfei Sun, "Search and the dynamics of house prices and construction," *American Economic Review*, 2014, *104*, 1172–1210.

Hurst, Erik, Veronica Guerrieri, and Dan Hartley, "Endogenous gentrification and housing price dynamics," *Journal of Public Economics*, 2013, *100* (C), 45–60.

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.

Kurlat, Pablo and Johannes Stroebel, "Testing for information asymmetries in real estate markets," *Review of Financial Studies*, 2015, *28* (8), 2429–2461.

Landvoigt, Tim, Monika Piazzesi, and Martin Schneider, "The Housing Market(s) of San Diego," *American Economic Review*, 2015, *105* (4), 1371–1407.

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

Manning, Alan and Barbara Petrongolo, "How local are labor markets? Evidence from a spatial job search model," *Forthcoming American Economic Review*, 2017.

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," *American Economic Review*, 2014, *104* (12), 3991–4026.

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

Ortalo-Magne, Francois and Sven Rady, "Housing market dynamics: On the contribution of income shocks and credit constraints," *The Review of Economic Studies*, 2006, *73* (2), 459–485.

Peri, Giovanni and Chad Sparber, "Task specialization, immigration, and wages," *American Economic Journal: Applied Economics*, 2009, *1* (3), 135–169.

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** _ , "Housing and macroeconomics," *Handbook of Macroeconomics, Edited by John B. Taylor and Harald Uhlig*, 2017, *2*, 1547–1640.

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

Stroebel, Johannes, "Asymmetric information about collateral values," *Journal of Finance*, 2016, *LXXI* (3), 1071 – 1111.

Swanson, Eric T, "Let's twist again: a high-frequency event-study analysis of operation twist and its implications for QE2," *Brookings Papers on Economic Activity*, 2011, *2011* (1), 151–188.

Trulia, "The largest audience of real estate consumers," Technical Report 2013.

USA Today, "Housing shortage eases in some markets," 2016.

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

Segmented Housing Search

Online Appendix

Monika Piazzesi Martin Schneider Johannes Stroebel

A Description of Search Behavior

In this Appendix, we provide additional descriptive statistics of the housing search behavior inferred from our email alerts data. In Appendix A.1, we analyze the frequency with which the major dimensions of search (geography, price, and the number of bathrooms) are selected. In Appendix A.2, we establish stylized facts on the geographic breadth of housing search. In Appendix A.3, we discuss the key characteristics of housing search along the price and size dimensions. In Appendix A.4, we provide information about how the size and price dimension correlate with the geographic breadth of the search range. For example, we document that searchers that are more specific about the price range and home size cover larger geographic areas. In Section A.5, we explore how searchers flow to different segments within their search ranges.

A.1 Major Dimensions of Search

As discussed in the paper, all email alerts require information on the geographic dimension of the potential homebuyer's search range. Roughly a third of the alerts do not specify any restrictions in addition to geography. The other fields that are used regularly include listing price and the number of bathrooms. Table A.1 shows the distribution of the dimensions that are specified across the email alerts in our sample.

Table A.1: Distribution of Email 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: Table shows the distribution of parameters that trulia.com users specify in addition to geography.

Just under a third of email alerts specify criteria for both price and number of bathrooms, while another third only specify a price criterion. The remaining five percent of email alerts specify just a bathroom criterion in addition to geography. Other fields in Figure 1 are used much less. For example, only 1.3 percent of email alerts specify square footage while 2.7 percent of alerts specify the number of bedrooms. While the latter two fields are alternative measures of size, the minimum number of bathrooms is the most commonly used filter to place restrictions on home size.

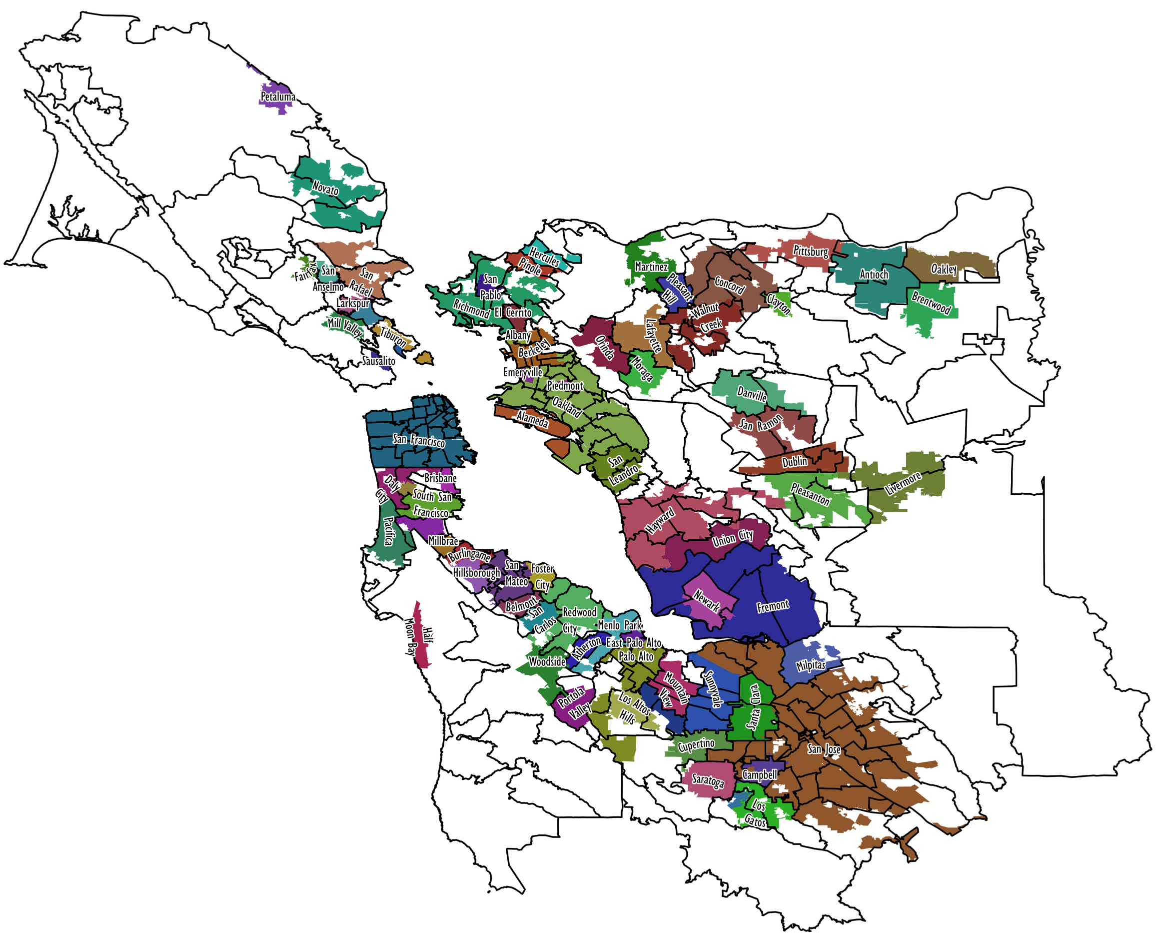
A.2 Search by Geography

We next describe the geographic dimensions of housing search inferred from the email alerts. We begin by outlining how we deal with alerts that specify the geographic criterion at different levels of aggregation. We then provide summary statistics on the distribution of distances covered by the email alerts in our sample.

A.2.1 Assigning Zip Codes to Email Alerts

Each email alert defines the geographic dimension of housing search by selecting one or more city, zip code, or neighborhood. About 61 percent of alerts define the finest geographic unit in terms of cities, 18 percent in terms of zip codes, and the remaining 21 percent in terms of neighborhoods. Some searchers include geographies in terms of cities, zip codes, and neighborhoods in the same alert. Figure A.1 shows how zip codes and cities overlap in the San Francisco Bay Area. Many cities cover multiple zip codes. Those parts of zip codes that are not covered by cities are usually sparsely populated.

Figure A.1: San Francisco Bay Area – Cities and Zip Codes



Note: This figure shows the geographic distribution of zip codes and cities in the San Francisco Bay Area. The base map are zip codes, the colored regions correspond to cities.

In order to compare email alerts that specify geography at different levels of aggregation, we translate every alert into the set of zip codes that are (approximately) covered. This requires dealing with alerts that specify geography at a level that might not perfectly overlap with zip codes. For alerts that select listings at the city level, we include all zip codes that are at least partially within the range of that city (i.e., for a searcher who is looking in Mountain View, we assign the alert to cover the zip codes 94040, 94041, and 94043). Neighborhoods and zip codes also do not line up perfectly, and so for each neighborhood we again consider all zip codes that are at least partially within that neighborhood (i.e., for a searcher who is looking in San Francisco's Mission District, we assign the alert to cover zip codes 94103, and 94110). This provides a list of zip codes that are covered by each email alert. The alerts cover 191 unique Bay Area zip codes.

A.2.2 Distance

To summarize how search ranges reflect geographic considerations, we construct various measures of size of the area considered. Since the unit of observation we are interested in is the searcher, not the email alert, we pool all zip codes that are covered in at least one email alert by a particular searcher. About 26 percent of searchers consider only a single zip code. For the remaining searchers, we measure the average and maximum of the geographic distances and travel times between all zip codes contained in their search ranges. We focus on distances between population-weighted zip code centroids. Population weighting is useful, since we are interested in the distance between agglomerations within zip codes that might reflect searchers' commutes.

Table A.2: Distribution of Distances Across 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
Mean Car Travel Time	3.8	8.9	13.1	19.7	132.5	14.0
Max Public Transport Time	10.5	40.5	79.0	375.0	573.5	140.1
Mean Public Transport Time	9.3	27.3	48.0	120.0	375.0	169.9

Note: Table shows summary statistics of geographic distance and travel time between the population-weighted centroids of all zip codes selected by a searcher. We focus on searchers who select more than one zip code. Travel times are measured in minutes, distances are measured in miles.

Table A.2 reports these measures of the geographic breadth of housing search. Geographic distance is measured in miles, and corresponds to direct "as the crow flies" distance. For the average searcher, the maximum distance between two zip codes included in a search range is 9.7 miles. There is significant heterogeneity in the geographic breadth of the search ranges. Indeed, searchers at the 90th percentile of the distribution have a maximum distance between covered zip codes of

21.1 miles, while searchers at the 10th percentile have a maximum geographic distance of 2.3 miles. These would usually be searchers that select two neighboring zip codes.

We also report the maximum and average travel times by car or public transport between the population-weighted zip code centroids. Travel times are calculated using Google Maps, and are measured as of 8am on Wednesday, March 20, 2013.[20] The size of the typical search range is consistent with reasonable commuting times guiding geographic selections. For example, the median search range includes zip codes with a maximum travel time by car of about 20.5 minutes; again, there is sizable heterogeneity in this measure; the across-searcher 10-90 percentile range of maximum travel times is 9.5 minutes to 38.5 minutes for travel by car, and 40.5 minutes to 375 minutes for travel by public transport.

In Appendix C we show that these geographic patterns of housing search are constant across the years in our sample, as well as across the seasonality of the housing market, suggesting that they represent measures of household preferences that are time-invariant throughout our sample.

A.2.3 Contiguity

To guide our modeling of clientele heterogeneity, we next explore whether there is a simple and parsimonious organizing principle for observed geographic 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 between any two zip code centroids in the range without ever leaving the range. We begin by describing how we construct measures of contiguity, before providing summary statistics on how many search queries cover contiguous geographies.

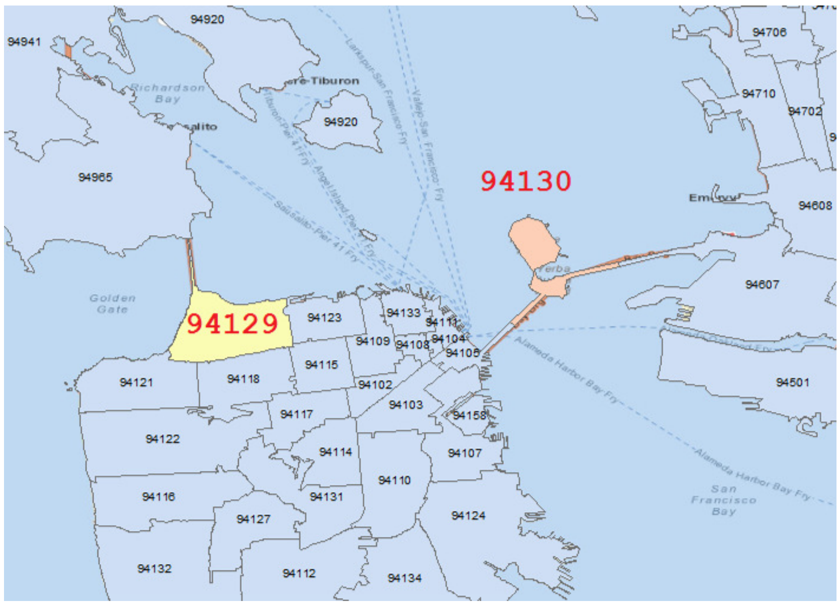
Dealing with the San Francicso Bay

To analyze whether all zip codes covered by a particular search query 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.2 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

[20] 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 transportation network.

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 email alerts, which would suggest, for example, that 94105 and 94607 are not connected. This challenge is addressed by manually merging zip codes 94129 and 94130 with the Golden Gate Bridge and Bay Bridge respectively. This ensures, for example, that 94118 and 94965 are connected even if 94129 was not selected.

Figure A.2: Bridge Adjustments - Contiguity Analysis



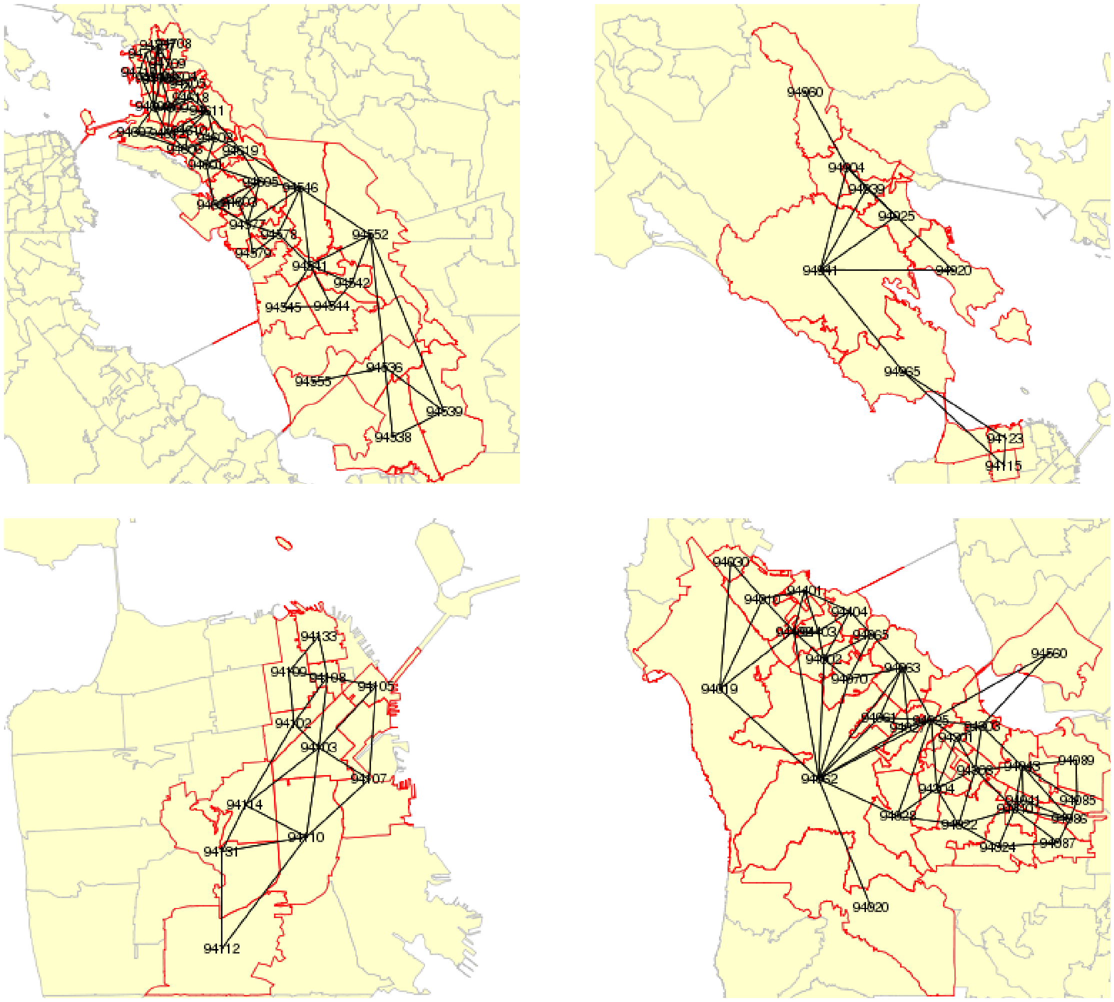
Note: This figure shows how we deal with bridges in the Bay Area for the contiguity analysis.

Examples of Contiguous and Non-Contiguous Search Sets

In the following, we provide examples of contiguous and non-contiguous search sets. In Figure A.3, we show four actual contiguous search sets from our data. The top left panel 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.[21] These are all locations with reasonable commuting distance to the tech jobs in Silicon Valley.

[21] The selected cities are 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, and Woodside.

Figure A.3: 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.

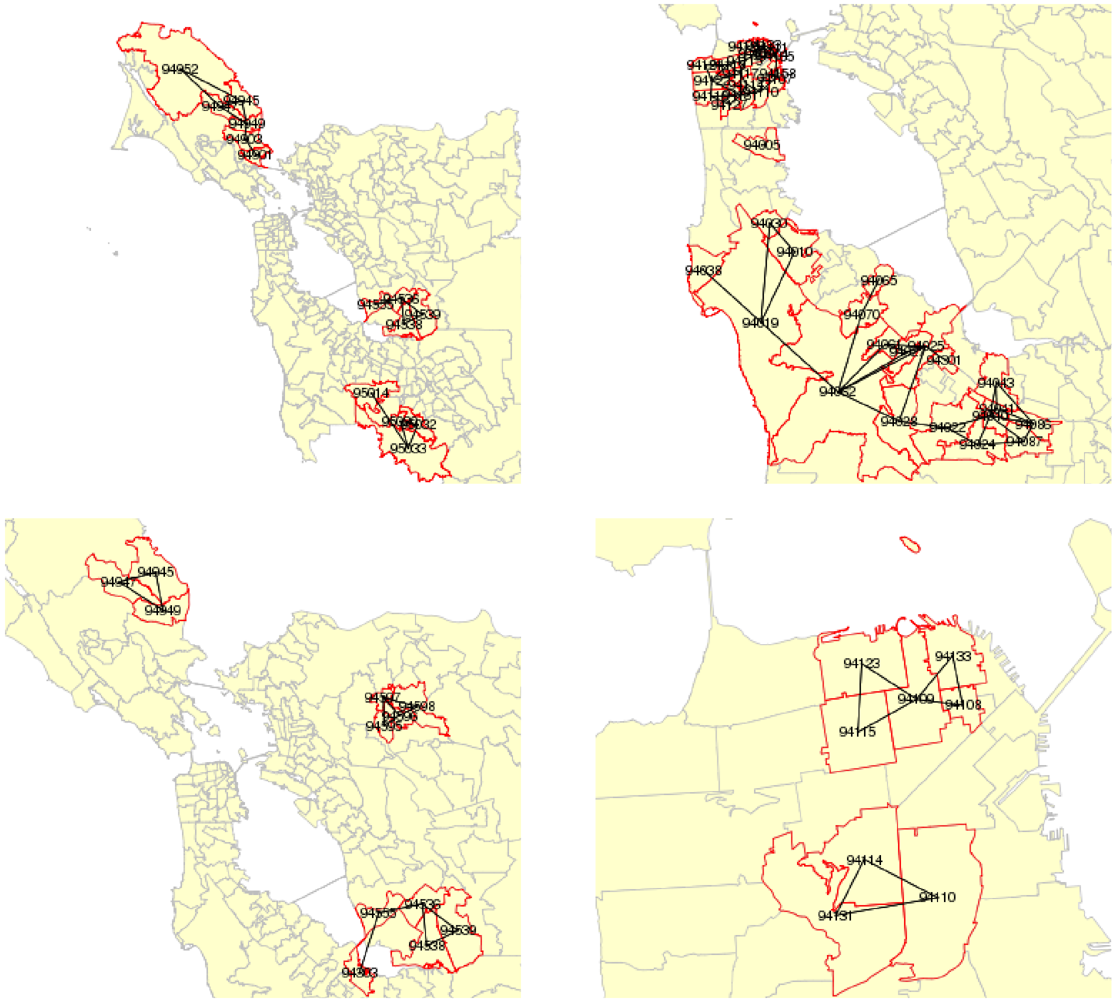
Notice how the addition of Newark adds zip code 94560 in the East Bay, which is connected to the South Bay via the Dumbarton Bridge.

In Figure A.4, 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 sets 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.

Summary Statistics on Contiguity

Table A.3 shows summary statistics of our measure of contiguity by the number of zip codes included in the search range. The second column reports the share of email alerts that select

Figure A.4: 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.

contiguous geographies. While only 18 percent of searchers have non-contiguous search ranges, they tend to come from broad searchers who consider more than five distinct zip codes, and hence provide market integration across neighborhood and city boundaries. The third and fourth columns report the mean and maximum number of contiguous areas covered by an email alert. 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. Overall, it is clear that a model of housing search that parameterizes search areas to be contiguous might provide a good approximation for some applications. However, in our setting this approximation would miss an important role played by searchers that integrate geographically distant housing markets.

Table A.3: Contiguity Analysis – Summary Statistics

Number of Zips Covered	Share Contiguous	Contiguous Areas Mean	Contiguous Areas Max	Total
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: Table shows summary statistics for contiguity measures across searchers that select different numbers of zip codes.

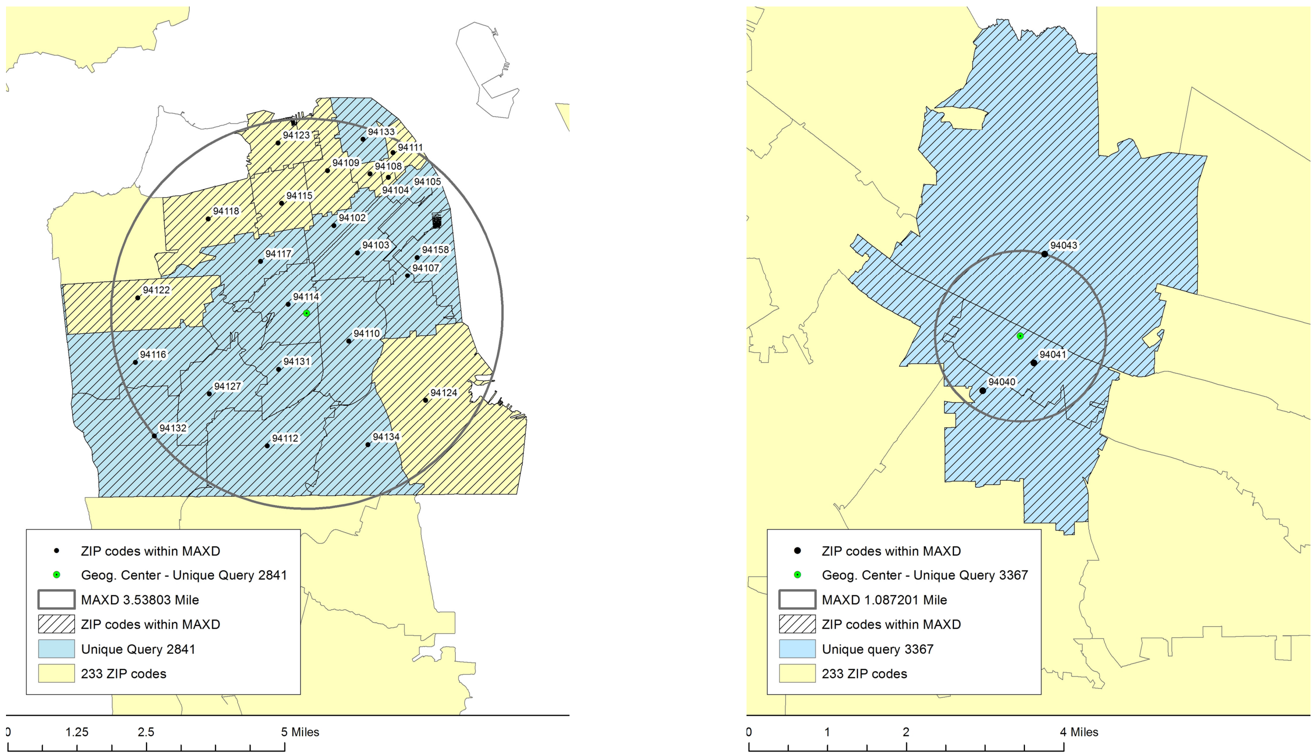
A.2.4 Circularity

A stylized model of geographic search might view a search range as being circular around a central point such as a job or a school. We ask whether the observed search ranges can be suitably approximated by such a model. To do this, we compute, for each searcher, the geographic center of the search range as the average longitude and latitude of all zip code 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 next compute the number of zip code centroids (not necessarily contained in the search range) that are within the 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 A.5 illustrates this procedure.

About 47 percent of all searchers that cover more than one zip code 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. For non-contiguous ranges, the share of zip codes covered falls to 33 percent.

Overall, we conclude that real-world housing search behavior cannot be well approximated using a simple and parsimonious search specification, either in terms of selecting contiguous geographies, or in terms of taking a "circular" search approach. This conclusion motivates our modeling approach in the paper, which is highly flexible and allows us to capture non-contiguous and non-circular search patterns.

Figure A.5: Explanation of Circularity Test



Note: Figure provides examples of the circularity measure. 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. 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. The left panel shows a non-circular search set, the right panel a circular search set.

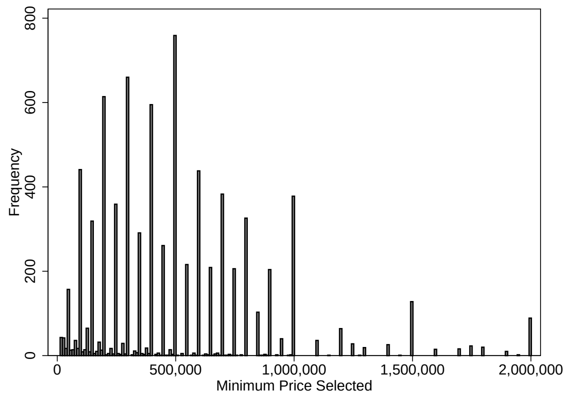
A.3 Search by Price and Size

Out of the 63 percent of email alerts that specify a price criterion, 50 percent specify both an upper bound and a lower bound, whereas 48 percent specify only an upper bound; only 2 percent select just a lower bound. Panels A and B of Figure A.6 show 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 $\$100,000$.

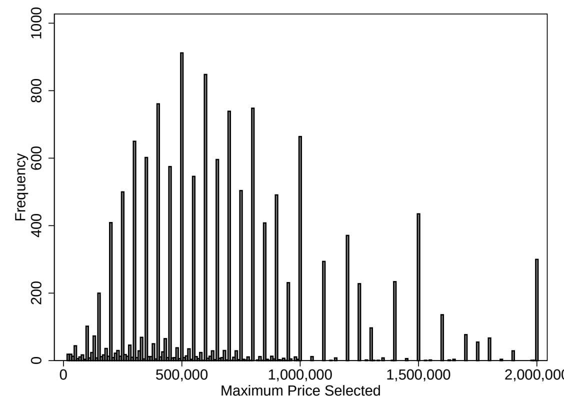
There is significant heterogeneity in the breadth of the price ranges selected by different home searchers. Among those searchers who set both an upper and a lower bound, the 10th percentile selects a price range of \$100,000, the median a price range of \$300,000, and the 90th percentile a price range of \$1.13 million. Among those searchers, the median person selects a price range of $\pm$ 27% around the mid-point of the range. At the 10th percentile of the distribution, this figure is $\pm$ 12.5% around the mid-point, and at the 90th percentile it is $\pm$ 58%. Panel C of Figure A.6 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.

Figure A.6: Price and Size Criteria of Housing Search

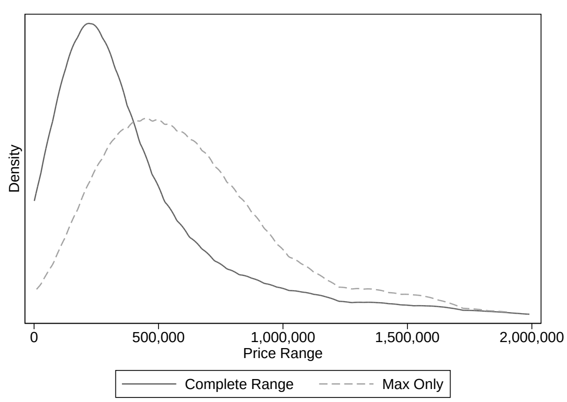
(A) Minimum House Price



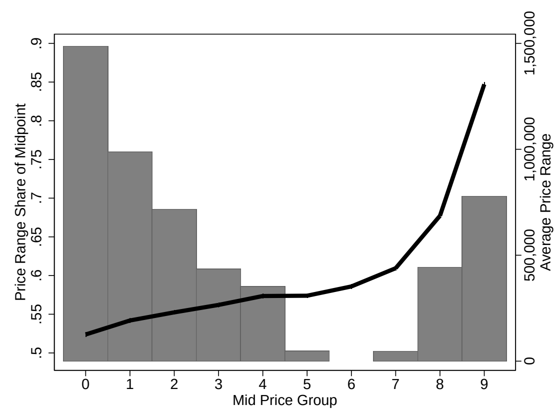
(B) Maximum House Price



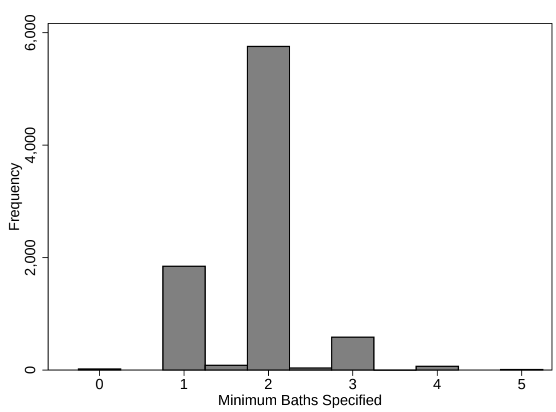
(C) Price Range



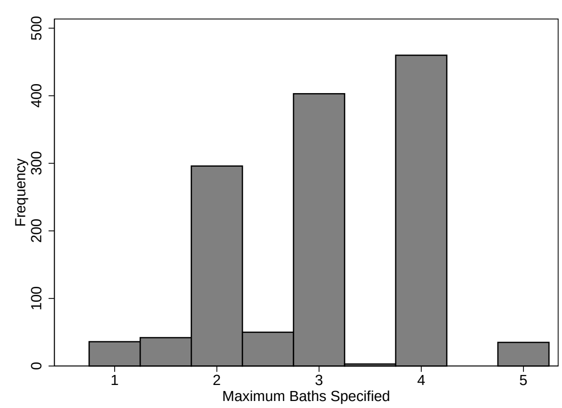
(D) Price Range by Midprice



(E) Minimum Number of Bathrooms



(F) Maximum Number of Bathrooms



Note: Panels A and B show histograms in steps of $10,000 of the minimum and maximum listing price parameters selected in email alerts. Panel C shows the distribution of price ranges across queries both for queries that only select a price upper bound (dashed line), as well as for those queries that select an upper bound and a lower bound (solid line). Panel D 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. Panels E and F show histograms in steps of 0.5 of the minimum and maximum bathroom selected, respectively.

Panel D shows that searchers who consider more expensive houses specify wider price ranges. We bin the midprice of price ranges into 10 groups. The solid 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 by 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.

In addition to geography and price, the third dimension that is regularly populated in the email alerts is a constraint on the number of bathrooms. Panels E and F of Figure A.6 show the distribution of bathroom cutoffs selected for the Bay Area. 68% of all bathroom limits are set at a value of 2, most of them as a lower bound. This setting primarily excludes studios, 1 bedroom apartments, and very small houses.

A.4 Tradeoffs between Search Dimensions

The three major search dimensions we have identified (geography, price, and size) 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 A.4 provides evidence on how different search dimensions interact. It shows that searchers who are more specific on the price or home size dimensions 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 10.6 miles, while searchers who do not specify a price range cover only 7.3 zip codes with an average maximum distance of 7.9 miles.

Table A.4: Geography, Price and Bath Parameter Interaction

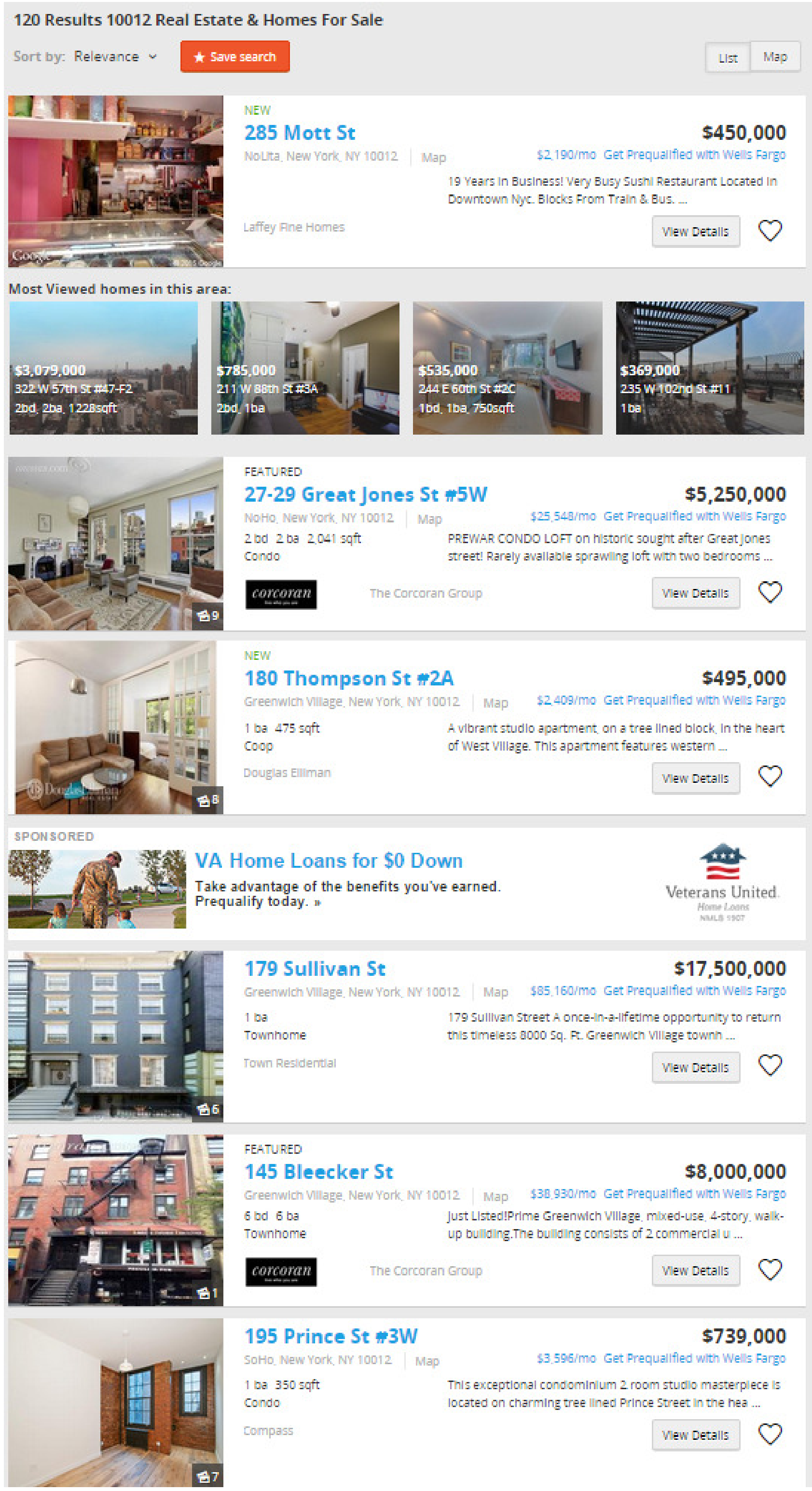
	No Price	Price	No Bath	Bath
# Zips Covered	7.3	10.3	8.8	10.0
Max Distance (Miles)	7.9	10.6	8.9	11.1
Max Time Car (Min)	20.8	24.5	22.5	24.7
Max Time Public Transport (Min)	75.8	92.4	82.1	95.9
Is Contiguous	54%	62%	59%	60%

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

A.5 Search within Search Range

In Section 3.2, we discussed the evidence for our key modeling assumption that the probability that a searcher would find her favorite home in any one of her considered segments is proportional to the share of that segment's inventory in total inventory across all considered segments. In this Appendix, we provide some additional discussion of this assumption.

Figure A.7: Example of Search Return List



Note: Example of list of properties for sale returned by trulia.com for a search query looking at all houses in zip code 10012.

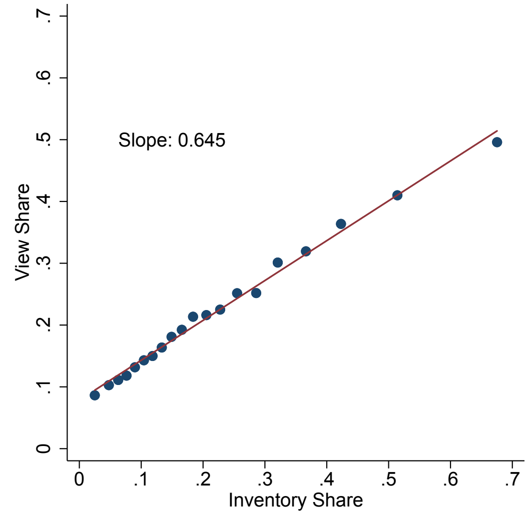
Unfortunately, we cannot map individual home searchers to their final purchases. Therefore, to address this question, we exploit data on property views by individual home searchers on trulia.com. After defining a search range on trulia.com, a user is presented with a list of properties that are included in that search range (see Appendix Figure A.7). This list provides basic information on each property, such as a picture, its location, the listing price, and the first lines of a description of the property. Home searchers then actively click on those houses that attract their particular interest to view additional property information.

We have obtained these individual property-view data from trulia.com for a random subset of users visiting the site in April 2012. These data contain the set of listings viewed within a "session," defined as all views by the same user within one day. We use these data to test whether users express interest by clicking on properties across different segments in the same proportion as those segments' inventory in the searcher's overall range. While we are unable to observe a user's search range in these particular data, we focus on those sessions that contain views of properties located in at least two of our 576 Bay Area segments. We observe information from 6,242 such sessions. About 25% of sessions include two property views, 19% include three property views, and 15% include more than 10 property views. In about half of these sessions, users view properties in just two segments, in 20% of sessions users view properties in 3 segments, and in 5% of sessions users view properties in more than 7 segments. For each session and segment, we calculate the share of views of houses in that segment, relative to the total number of houses viewed in the session ("view share"). We also calculate the share of inventory in that segment, relative to the total inventory in all segments in which houses were viewed in that session ("inventory share").

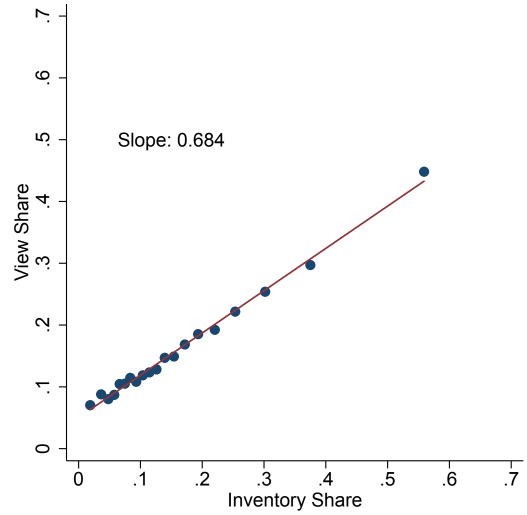
Appendix Figure A.8 shows binned scatter plots documenting the relationship between view share and inventory share. The unit of observation is a session-segment. In Panel A we focus on the 2,704 sessions where the user viewed at least five individual properties. We split the resulting 12,225 session-segments into 20 equally sized buckets ordered by inventory share. On the horizontal axis we plot, for each bucket, the average inventory share, and on the vertical axis we plot the average view share. There is a strong positive relationship. The rate at which particular interest is expressed for properties in different segments is increasing in the share of total inventory made up by those segments in the overall search range. For those segments with the lowest inventory share, we find view shares to be somewhat above inventory shares, while for segments with a very high inventory share we find view shares to be somewhat below inventory shares. This effect is mechanical, and a result of observing relatively few property views per session on average, combined with the fact that we infer a searcher's search range from the properties viewed. To see this, take a session in which 5 properties were viewed. For us to know that a segment was included in that session's search range, we need to observe at least one property from that segment being viewed. Therefore, any segment that we know is included will at least have 20% view share, even if its inventory share was only 5%. As we investigate sessions with more properties viewed, this bias gets smaller, as shown in Panels B to D of Appendix Figure A.8.

Figure A.8: “View Share” vs. “Inventory Share”

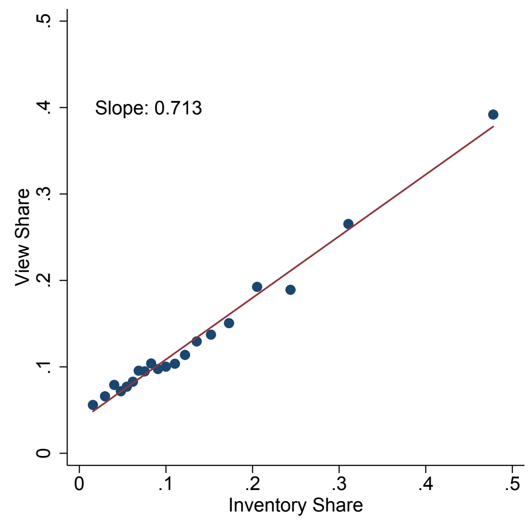
(A) At least 5 property views per session (N = 2,704)



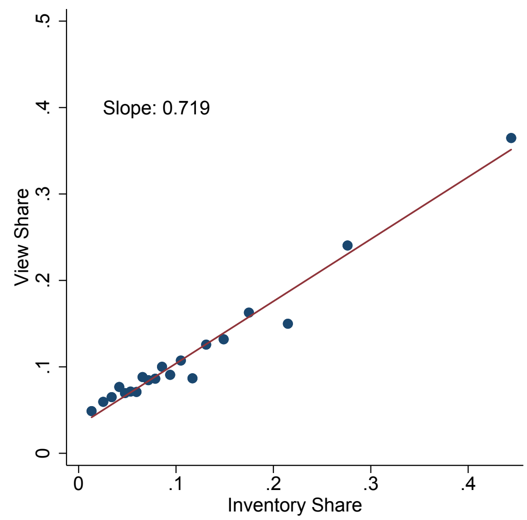
(B) At least 10 property views per session (N = 1,043)



(C) At least 15 property views per session (N = 500)



(D) At least 20 property views per session (N = 279)



Note: Figure shows binscatter plots at the segment-session level; different panels vary the minimum number of property views required for a session to be included. On the horizontal axis is the share of inventory of a segment, relative to the total inventory in all segments viewed in that session. On the vertical axis is the share of properties in that segment viewed, relative to the total number of properties viewed in that session.

B Construction of Segments, and Segment-Level Activity

In this Appendix, we discuss how we construct the housing market segments based on the email alerts in our data. We also describe how we assign segments to email alerts, and how we measure segment-level housing market activity.

B.1 Segment Construction

This section describes the process of arriving at the 564 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 arms-length housing transactions between 1994 and 2012. This leaves us with 191 zip codes with sufficient observations to construct these measures.

We next describe how we further split these zip codes into segments based on a quality (price) and size dimension. Importantly, we 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 contain information on the universe of the housing stock. However, as a result of Proposition 13, assessed property values in California do not correspond to true market values, and it is thus not appropriate to divide the total zip code housing stock into different price segments based on this assessed value.[22] To measure the housing stock in different price segments, we thus use data from the 5-year estimates from the 2011 American Community Survey, which reports the total number of owner-occupied housing units per zip code for a number of price bins. We use these data to construct the total number of owner-occupied 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 adjust all prices and price boundaries to 2010 house prices.[23]

Not all zip codes have an equal distribution of houses in each price (quality) bin. For example, Palo Alto has few homes valued at less than \$200,000, while Fremont has few million-dollar homes. Since we want to avoid cutting a zip code into too many quality segments with essentially

[22] 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.

[23] 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 house would fetch in 2010 as our measure of that home's underlying quality. To transform the housing stock in each price bin reported in the ACS into a housing stock for different 2010-"quality" segments, we first construct zip code-specific annual repeat sales house 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.

no housing stock to 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 ACS price bin cut-offs should constitute segment cut-offs, we use information from the email alerts. This proceeds in two steps: First we change the price parameters set in the email alerts to account for the fact that we observe alerts 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 alert was set and 2010.[24] Second, we determine which set of three ACS cutoffs is most similar to the distribution of actual price cut-offs selected in search queries that cover a particular zip code. In particular, 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) email alert price restrictions to the closest cut-off.[25] We select the set of segment-specific price cut-offs that minimizes the average of this value across all alerts that cover a particular zip code. This ensures, for example, that if many email alerts covering a zip code include a high cut-off such as $1 million (either as an upper bound, or as a lower bound), $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 rental 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 arms-length ownership-changing transactions between 1994 and 2010, as reported in our deeds records. We first adjust the observed transaction prices 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 homeownership rate for each price segment within a zip code, and adjust the ACS-reported stock accordingly.[26] To assess the quality of the resulting adjustment, note that the total resulting housing stock across our segments is approximately 2.2 million, very close to the total Bay Area housing stock in the 2010 census.

The other search dimension regularly specified in the email alerts is the number of bathrooms

[24] 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 an alert set in 2006 that restricts price to be between $500,000 and $800,000 will search for homes in the same quality segment as an alert set in 2010 that restricts price to a $250,000 - $400,000 range.

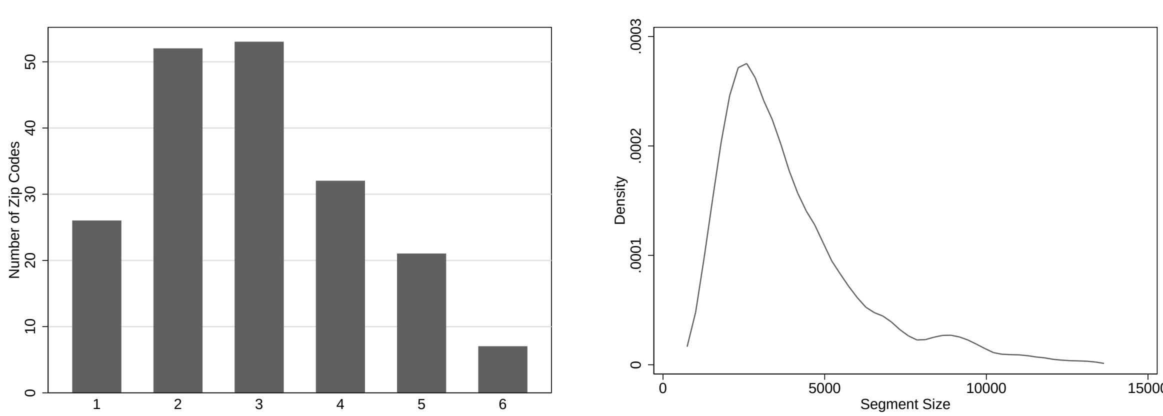
[25] For example, imagine testing how good the boundaries $100k, $300k and $1m fit for a particular zip code. An alert with an upper bound of $500k has the closest absolute distance to a cut-off of $\min\{|500-100|, |500-300|, |500-1000|\} = 200$. An alert with an upper bound of $750k has the closest absolute distance to a cut-off of 250. An alert 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 alert the smallest absolute distance of an alert limit to a cut-off, and then find the average across all email alerts.

[26] 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 homeownership rate is much higher in the highest 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.

as a measure of the size of a house conditional on its location and quality. Since Section A.3 shows 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 fewer 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 zip code by price segment into groups by bathrooms, 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 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 segment-specific moments such as the average time on market with some precision, we want to ensure that each segment has a housing stock of at least 1,200 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 magnitude. For price segments where either of the two size subsegments have a stock of less than 1,200, we merge the two size segments. We then begin with the lowest price segment, see whether it has a stock of less than 1,200, and if not merge it with the next higher price segment. This procedure generates 564 segments.

Figure B.1: Segment Overview



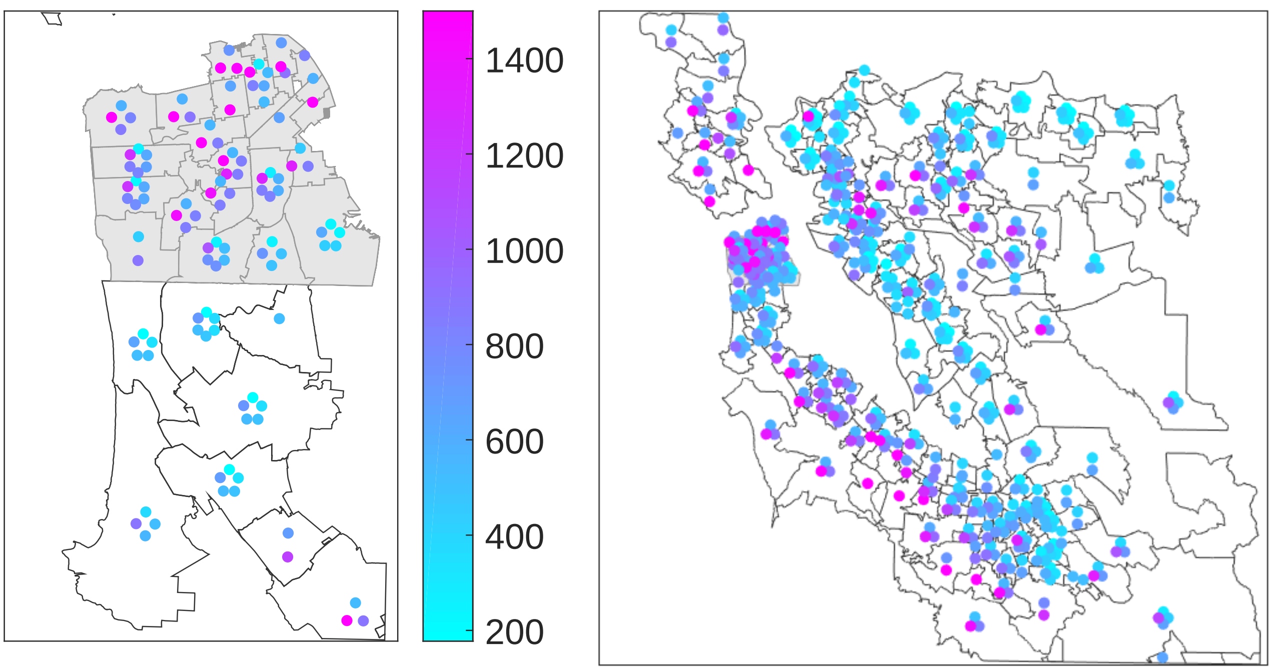
Note: The left panel shows the number of segments that the 191 zip codes are split into. The right panel shows the distribution of the number of housing units across segments.

Figure B.1 shows how many segments each zip code is split into. 26 zip codes are not split up further into segments. 52 zip codes are split into two segments, 53 zip codes are split into 3 segments. The right panel of Figure B.1 shows the distribution of housing stock across segments. On average, segments have a housing stock of 3,929, with a median value of 3,298. The largest segment has a housing stock of 13,167.

A First Look at the Segments

The left panel of Figure B.2 shows a map of the city of San Francisco in gray in addition to the area south of the city. The black lines delineate zip codes. The white areas without boundaries are water. Within each zip code, there are up to six dots that represent segments. The segments are aligned clockwise starting with the cheapest segment at twelve o'clock. The colors correspond to the average house price in the segment, with the USD amounts in thousands indicated by the legend. The map shows the substantial heterogeneity of house prices in a city like San Francisco. There is also large heterogeneity within zip codes: indeed, the variance of log prices across zip codes captures only 60 percent of the variance at the (more disaggregated) segment level.

Figure B.2: Segments in San Francisco and the Bay Area



Note: The left panel shows a map of downtown San Francisco as the shaded area in addition to areas south of downtown. The right panel shows a map of the entire Bay Area. The color bar indicates the price of the segment in thousands of USD.

The map in the right panel of Figure B.2 shows the entire Bay Area. The busy agglomeration of segments in the upper-left quadrant of the Bay Area map is San Francisco. The mostly light blue segments in the lower-right quadrant are in the cheaper city of San Jose. The pink segments between these two cities are Silicon Valley cities like Palo Alto and Atherton, while the light blue segments across the water are Oakland and other cheaper East Bay cities. The light blue dots in the upper-right corner belong to the Sacramento Delta.

B.2 Assigning Segments to Email Alerts

We next describe how we assign which segments are covered by each email alert. In Appendix A.2, we already discussed how we determine which zip codes are covered by each alert, and how we deal with alerts that specify geography at a different level of aggregation. We next discuss how we incorporate the price and size dimensions of housing search to determine which segments in a zip code are covered by each alert. The challenge is that price ranges selected will usually not overlap perfectly with the price cutoffs of the individual segments. For those alerts that specify a price dimension, we assign an alert to cover a particular segment in one of three cases:

1. When the alert completely covers the segment (that is, when the alert lower bound is below the segment cutoff and the alert upper bound is above the segment cutoff).

2. When the segment is open-ended (e.g. \$1 million +), and the upper bound of the alert exceeds the lower bound (in this case, all alerts with an upper bound in excess of \$1 million).

3. For alerts that partially cover a non-open ended segment, we determine the share of the segment price range covered by the alert. For example, an alert with price range \$200k - \$500k covers 20% of the segment 0-\$250k, and 50% of the segment \$250k - \$750k. We assign all alerts that include at least 50% of the price range of a segment to cover that segment.

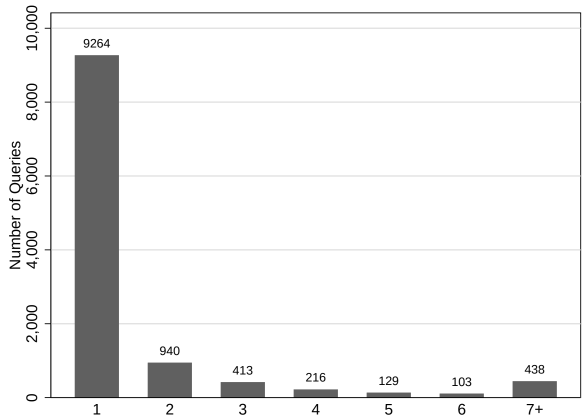
To incorporate the bathroom dimensions, we let an alert cover a segment unless it is explicitly excluded. For example, alerts that require at least two bathrooms will not cover the <2 bathroom segments and vice versa.

The housing market segments constructed above allow us to pool across all email alerts set by the same individual. In particular, we add all segments that are covered by at least one email alert of an individual to that individual's search set. After pooling all segments covered by the same searcher in this way, we arrive at a total of 11,503 unique search profiles, set by the 23,597 unique users in our data. Figure B.3 shows the distribution of how many different searchers are represented by the different search profiles. A total of 9,264 search profiles represent only a single searcher, 940 search profiles represent two searchers. 438 search profiles represent at least seven searchers, while the most common search profile represents 788 unique searchers.

B.3 Construction of Segment-Level Market Activity

Segment Price. Our model links the characteristics of search patterns to segment-specific measures of market activity such as price, turnover, 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 arms-length 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. This excludes, for example, intra-family transfers. We drop all

Figure B.3: Number of Searchers per Unique Profile



Note: Figure shows how many unique individuals are represented by each of the 11,503 individual search profiles.

observations that are not a Main Deed or only transfer partial interest in a property (see Stroebel, 2016, for details on this process of identifying arms-length transactions).

Turnover Rate. We then calculate the total number of transactions per segment between 2008 and 2011, and use this to construct annual volume averages. To allocate houses to particular segments, we adjust transaction prices for houses sold in years other than 2010 by the same house price index we used to adjust listing price boundaries (see Appendix B.1). We then measure "turnover rate" by dividing the annual transaction volume by the segment housing stock.

Time on Market. To calculate the average time on market, we use the data set on all home listings on trulia.com, beginning in January 2006, and match those home listings with final transactions from the deeds database.[27] This match is done via the (standardized) addresses across the two data sets. Panel A of Figure B.4 shows the time series of the share of listings that we are unable to match to deeds data, starting in 2008, the first year of our estimation sample. On average, for properties listed between January 2008 and July 2011, we can match between 85% and 90% of all listings to subsequent transactions. This number is relatively constant across segments. There are three reasons for why we may not be able to match all listings to transactions:

1. The listed property sells, but due to a different formatting of the address in the listing and the deed (or an incomplete address in the listing), we cannot match listing and sale.

2. The property got withdrawn from the market without being sold.

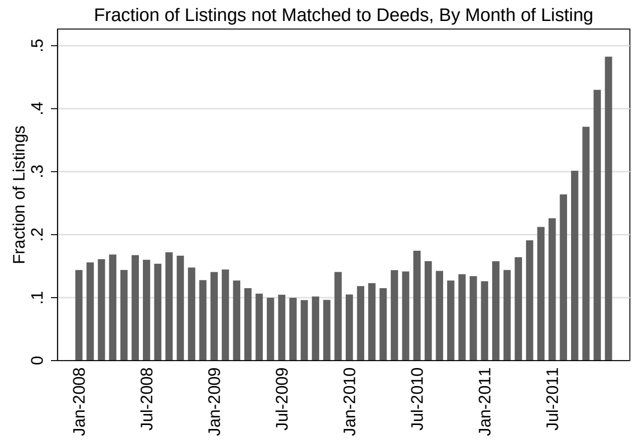
3. The property is still for sale by the end of our transaction sample (April 2012).

[27] We measure time on market as the period between the first listing of the property and the final transaction; this combines across listing spells of properties that are repeatedly listed and delisted by real estate agents in order to avoid the appearance of a stale listing. We subtract one month to allow for the typical escrow period.

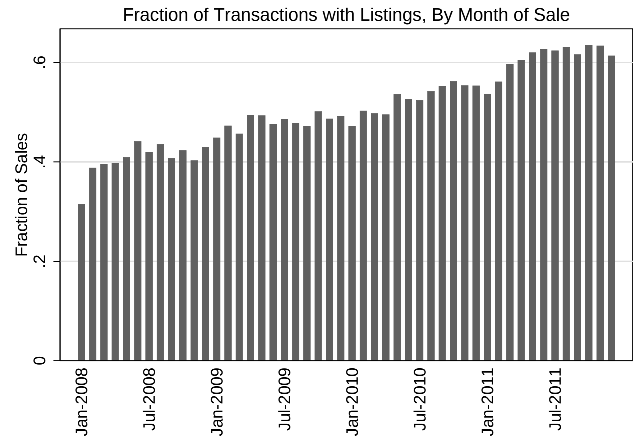
The increase in the share of listings without sales from the middle of 2011 onward is likely due to the last reason. Across all listings that we can match to a final transaction, the 90th percentile of time on market is 343 days, the 95th percentile is 502 days (the median is 84 days, and the mean is 144 days). This means that a significant number of properties listed toward the end of our sample will not sell by the end of our deeds data window (April 2012).

Figure B.4: Listings-Transaction Match Rates

(A) Listings without sales



(B) Sales with Listings



Note: Panel A shows the number of listings for which we do not eventually find a deed to match, by month of listing. Panel B shows the share of transactions for which we also observe a listing, by month of sale. Both Panels cover the period 2008-2011, for which we observe Trulia email alerts.

We find segment-level measures of time on market by averaging the time on market across all transactions between 2008 and 2011.[28] We also constructed an alternative measure where we restricted the sample to be the time on market for all properties that were listed between 2008 and July 2011; for this calculation we excluded the last 6 months in order to avoid the problem of the censoring of time on market numbers for properties that are listed towards the end of our sample. Both measures provide very similar measures of segment-level time on market, and a nearly identical ranking of segments across this dimension.

Inventory Share. As discussed in the paper, in our baseline estimates we construct the inventory share with the steady-state equation $I = T \times V$ instead of calculating it using the actual listings we observe. There is a trade-off in this choice. In particular, as discussed in the paper, the downside of our approach is that it does not include in our measure of inventory properties that are listed and subsequently delisted. However, as we show above, this number is relatively small: it is bounded above by 10% to 15% of all listings, and manual examination of un-matched listings

[28] 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 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. In addition, in our baseline estimates we exclude the few properties where we observe a time of more than 900 days between listing and sale; this does not have a significant effect on the final measures of segment-level time on market.

suggests that the vast majority of our inability to match listings to sales is due to incomplete addresses in listings. In addition, as discussed above, the share of listings without sales is relatively similar across segments; therefore, abstracting from withdrawn listings is unlikely to affect the cross-sectional patterns that we focus on in our analysis.

We also tried an alternative approach to constructing the inventory share based on actual listings that we observe in the data. Under this approach, a property gets added to inventory the first time a listing appears, and gets removed once the property sells. For every month, we then have a set of properties that are on the market in that month, and we can form segment-level averages by calculating the share of the housing stock that is for sale across all the months between January 2008 and December 2011. Since we observe listings starting in October 2005, we do not require a "burn in period" at the beginning of the sample. One issue with this approach is that the coverage of Trulia listings data during our time period is not complete. Indeed, Panel B of Figure B.4 shows the number of transactions that can be matched to a previous listing is increasing throughout our sample, from about 40% at the beginning, to about 60% towards the end of our sample. Much of this improvement is due to the increasing coverage of the Trulia data. To not significantly bias the results under this alternative approach, one has to make two adjustments to the data. First, one has to decide how to treat properties that never sell; in our baseline approach, we removed them from the sample if they stayed on the market for more than 270 days. Second, we have to adjust for the incomplete coverage of the Trulia deeds data, which might differ by segment. To do this, we scale up the actual inventory share by the fraction of deeds with no listings in that segment.

While the measures using both approaches produce similar cross-sectional patterns in the data, and similar cross-sectional rankings across segments, we prefer calculating the inventory share from the more precisely measured time on market; it has less measurement error, and, for the reasons discussed above, any bias introduced by ignoring delistings is likely to be small.

B.4 Segment-Level Search Breadth

In this Appendix, we present more details on the search-breadth of the various searchers covering each segment. The first column of Table B.1 summarizes the distribution of inventory scanned by the median searcher. In the average segment, the median searcher scans 2.1 percent of the total Bay Area inventory. The table also clarifies that most dots in the left panel of Figure 3 are clustered in the bottom left; the 75th percentile of the distribution is at only at 2.5 percent of total inventory. The second column in Table B.1 shows the distribution of the within-segment interquartile range for scanned inventory. There is substantial within-segment heterogeneity in the clientele's breadth of housing search. Indeed, the average within-segment IQ range of inventory scanned by different searchers is, at 1.75 percent, larger than 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 65 percent. In other words, in segments that are, on average, more integrated with other segments, there are larger within-segment differences between the interacting narrow and broad searchers.

Table B.1: Variation in Scanned Inventory by Clienteles

	SEGMENT: total inv. scanned (in percent)		ZIP CODE: share of search ranges (in percent)				CITY: share of search types (in percent)			
	median	IQ range	one	many	subset	other	one	many	subset	other
Mean	2.10	1.75	5.9	4.4	5.8	84.5	10.7	8.1	16.8	64.4
Q25	0.94	1.04	0	0	0.8	72.3	0	0	3.1	53.2
Q50	1.55	1.56	1.2	1.2	2.4	93.0	8.0	1.7	12.4	66.6
Q75	2.53	2.17	6.5	4.7	6.5	97.1	17.0	6.3	25.4	77.2

Note: Table provides information on the inventory scanned by clienteles at different levels of aggregation. The second column measures the inventory scanned by the median searcher in a segment. The third column is the interquartile range of inventory scanned across all searchers in a segment. The other columns report mean and quartiles for the share of different searcher categories across segments. For each of the two geographies, zip code and city, a searcher is a category "one" searcher if the reason he searches the segment is because he uniquely selected that geography. The searcher is a category "many" searcher, if he only selected on that geography, but included more than one unit. A "subset" searcher covers the segment, but only selected a subset of one zip code or city to be included. "Other" searchers cover subsets of multiple zip codes or cities.

Search at the City and Zip Code Level

The right columns of Table B.1 demonstrate the importance of detailed segment-level information for understanding search patterns. For each zip code and city, we consider all searchers who are active in that zip code or city. We then separate these searchers into how they searched in that geography. We first classify the share of searchers who scan exactly one zip code or city in its entirety (column "one"). The category "many" captures searchers that scan only based on the particular geography, but consider more than one zip code or city. Together, they indicate the share of searchers for whom detailed segment-level information beyond geography is not important. The category labeled "subset" collects searchers who scan only a subset of the geography (i.e., they scan less than one zip code, or less than one city), for example because they select that zip code in addition to a price cutoff. The final category "other" collects searchers for whom segment information matters because their range intersects with multiple zip codes or cities, without covering them in their entirety.

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 5.9 percent of searchers select exactly the zip code containing that segment. An additional 4.4 percent of searchers specify their search query only in terms of zip codes, but specify more than just one zip code. The distribution is highly skewed: in 75 percent of segments, the share of searchers scanning exactly the zip code is 6.5 percent or less. The magnitude of the numbers is larger at the city level, but still relatively small. We therefore conclude that the clientele patterns at work in our data are not simply driven by searches selecting a single zip code or city. Instead, other characteristics defining a segment, in particular size and quality, play an important role.

C Time-Series Stability of Patterns

Our empirical analysis focuses on the years 2008 to 2011, a period for which we are able to observe both search behavior and housing market activity (while we observe some email alerts from before 2008, the vast majority come from after that period). One natural question is thus the extent to which our conclusions generalize beyond the time period under investigation. In the following sections, we show that, as far as is possible to say with our data, both the search patterns as well as the ranking of segments in terms of market-level activity are relatively stable over time. This increases our confidence that the patterns we investigate in this paper are not just an artifact of the particular period studied.

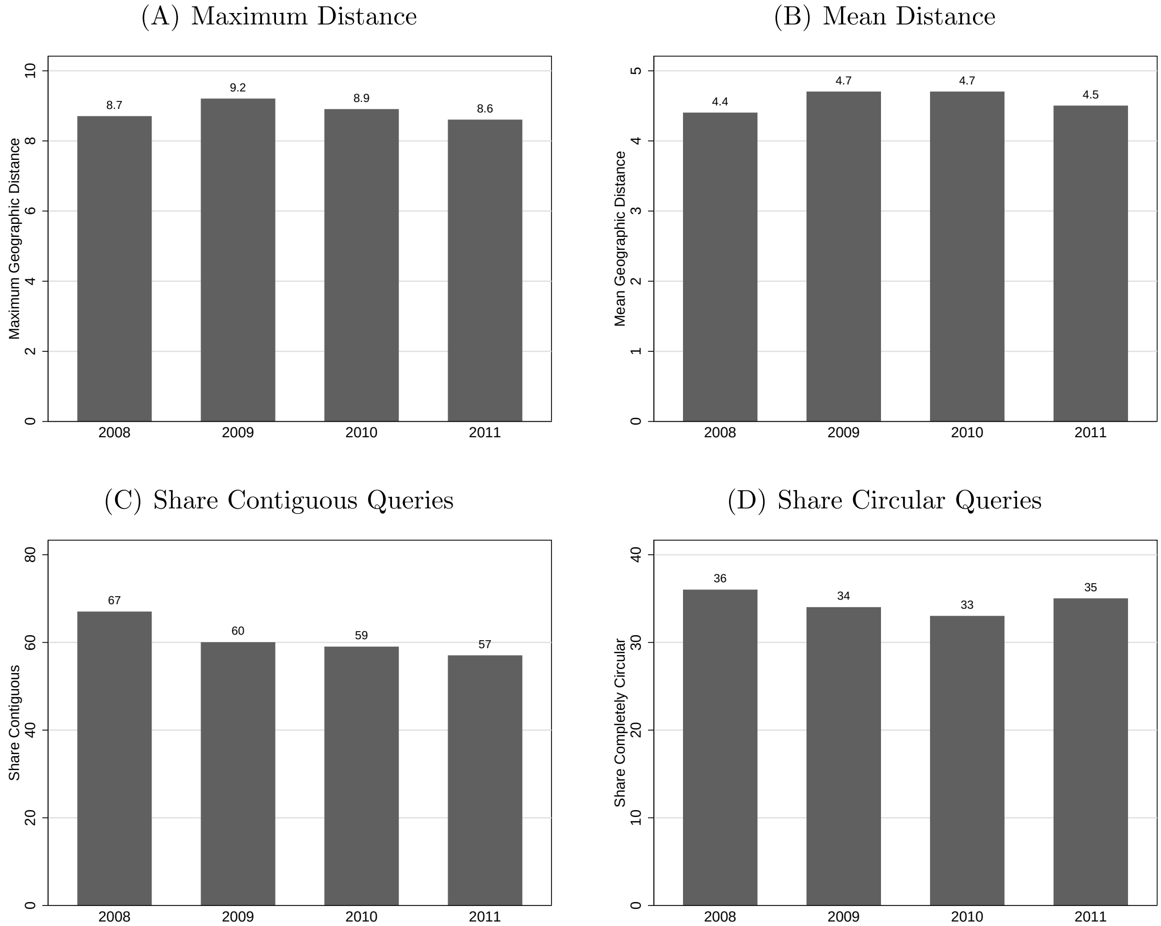
C.1 Stability of Search Patterns

Our model interprets the observed search ranges as a time-invariant feature of buyer preferences. For example, our assumption in the comparative statics exercises in Section 6.2 is that the search breadth would not adjust based on the amount of inventory within the range (see also our discussion in Section 5.2). It is then valuable to analyze empirically whether the search ranges are indeed invariant to changes in market conditions. In particular, do searchers narrow the range of houses they consider when market activity is higher, and there is more inventory in each segment? We provide two tests that show no evidence that the parameters of housing search vary with housing market activity.

In a first test, we explore important summary statistics on the geographic breadth of each email alert between 2008 and 2011, split by the year in which the email alert was set. The results are presented in Figure C.1. We consider the maximum distance between geographic zip code centroids (Panel A), the average distance between geographic zip code centroids (Panel B), the share of searches that yield contiguous search sets (Panel C), and the share of "circular" queries as defined in Appendix A.2.4 (Panel D). We find that these important parameters of search activity are very stable across the years in our sample. This suggests that they do indeed capture time-invariant preference parameters of households over the Bay Area housing stock, that does not respond to the relative supply available across the search range.

A second test exploits seasonal variation in housing market activity: more houses typically trade in the summer as compared to the winter. This can be seen in Panel A of Figure C.2, which shows the average share of total annual transaction volume over our sample in each month. Market activity is twice as high in June than it is in January. Panels B to E of Figure C.2 show averages of the same summary statistics on the search parameters as Figure C.1, split by the month of the year when the email alert was set. As before, none of the search dimensions exhibit meaningful seasonality, consistent with an interpretation of search parameters as time-invariant measures of preferences that do not vary with market activity.

Figure C.1: Non-Cyclicality of Search Parameters



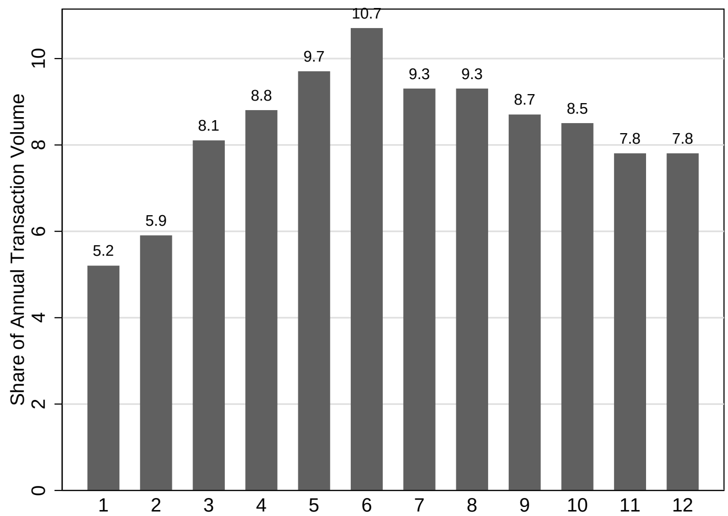
Note: Figure shows average values of search parameters by the year when the email alert was set. We report the maximum distance between zip code centroids (Panel A), the mean distance between zip code centroids (Panel B), the share of contiguous queries (Panel C) and the share of circular queries (Panel D).

C.2 Stability of Market Activity

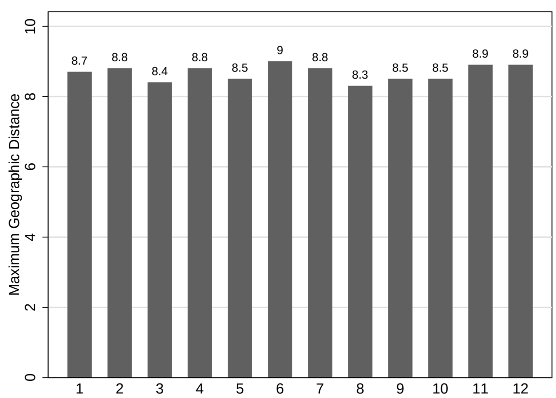
A related question is whether the market-level outcomes are particular to our period of study. Unfortunately, we do not have the data to analyze the key patterns outside of our sample period; in particular, the listings data necessary to construct inventory shares and time on market are only reliably observed during that period. However, to provide some evidence that our results are not just a feature of a period of declining house prices, we exploit the fact that our sample period includes two rather distinct housing market episodes: a period of declining house prices, between January 2008 and March 2009, and a period of relatively flat or even increasing house prices between April 2009 and December 2011 (see Panel A of Figure C.3). This allows us to test whether the across-segment patterns we observe are similar in these two episodes.

Figure C.2: Non-Seasonality of Search Parameters

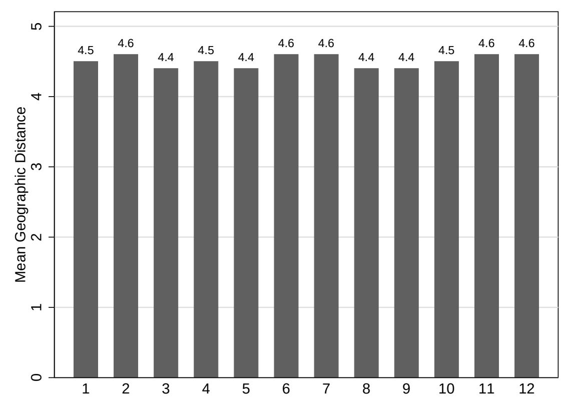
(A) Share of Annual Transactions



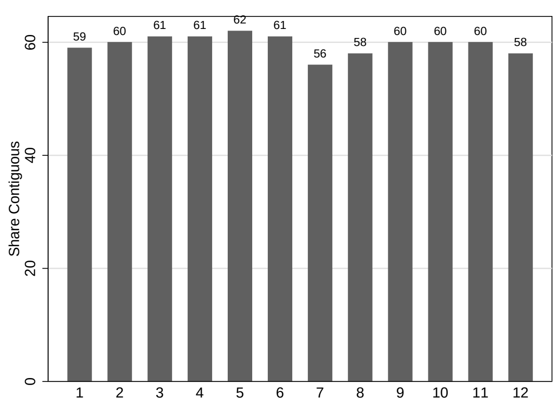
(B) Maximum Distance



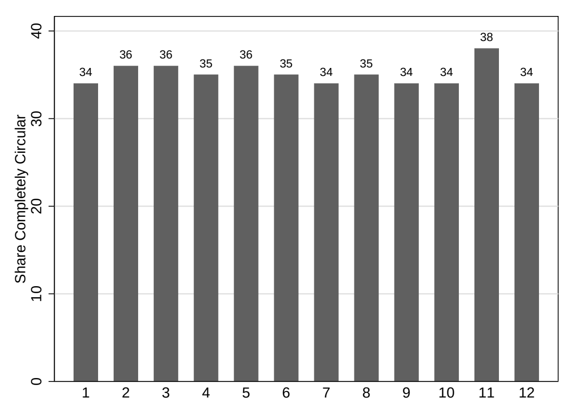
(C) Mean Distance



(D) Share Contiguous Queries

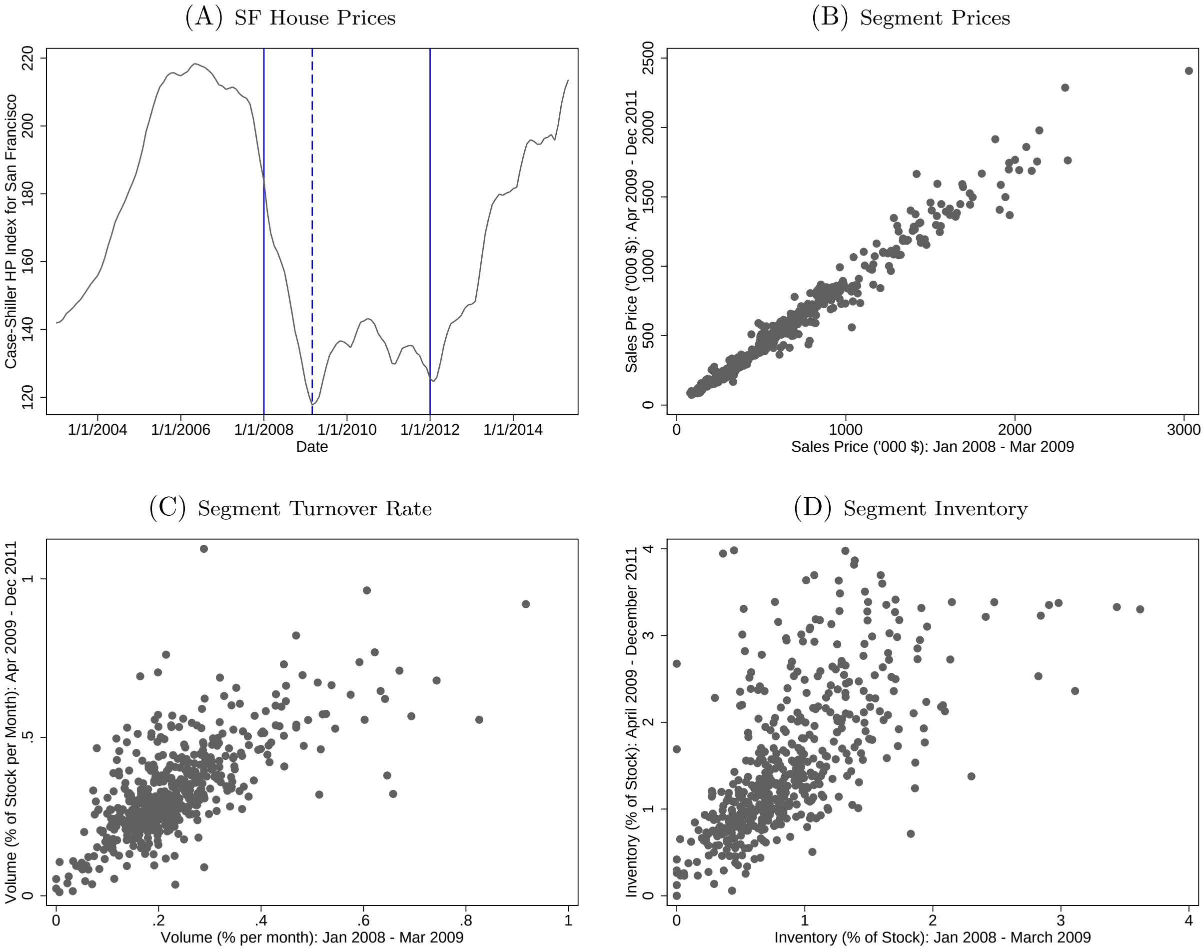


(E) Share Circular Queries



Note: Panel A shows the share of total annual transaction volume in each month. Panels B - E show average values of search parameters by month of search query. We report the maximum distance between zip code centroids (Panel B), the mean distance between zip code centroids (Panel C), the share of contiguous queries (Panel D) and the share of circular queries (Panel E). Months increase from January to December on the horizontal axis.

Figure C.3: Stability of Segment Moments



Note: Panel A shows the Case-Shiller House Price Index for San Francisco. Our sample period is delineated with vertical solid lines. It is split into a period of declining house prices and stable house prices by a dashed vertical line. Panels B, C, and D show segment-level scatter plots of key housing market moments (prices, turnover rate, and inventory, respectively) across the periods of declining and stable house prices.

Panel B, C, and D of Figure C.3 show the across-episode correlation of segment-level house prices, turnover rates, and inventory shares. There is a high correlation between these measures of market activity across the "declining market" period and the "stable market" period: zip codes with high volume and high inventory during the 2008 bust also have high volume and high inventory during the subsequent stable price period. Indeed, the Spearman's rank correlation coefficient for these two variables is 0.70 and 0.75, respectively. It is much higher, at 0.96, for the more-precisely measured average house price per segment.

We can also analyze the correlation between different moments across segments in both the bust period and the stable period, and compare it to the pooled correlation presented in Table 3 of the paper. For example, inventory and volume had an across-segment correlation of 0.93 in the pooled period. This measure was 0.81 and 0.83 in the bust and the stable period, respectively. Similarly, the correlation between inventory and price levels in the pooled sample was -0.63, while it was -0.29 and -0.43 in the bust and the stable period, respectively. The lower correlation in either of the sub-periods highlights the additional noise introduced by splitting the sample, and reinforces our choice to analyze the pooled sample in our baseline analysis.

Therefore, while we cannot rule out that the observed relationship between inventory, volume, and search behavior might look different during a housing boom period, it is reassuring that there are no significant differences in the cross sectional relationship between these variables during periods of strongly declining prices and periods of stable prices.

D Robustness: Search Models and the Beveridge Curve

In this Appendix, we show that the effects derived in the single-segment reduced-form model described in Section 4 obtain in many fully fledged search models. In particular, they hold under alternative assumptions on the broad buyer flow $B^B(L)$. They are also consistent with different setups for equilibrium search and pricing. We provide regularity conditions under which there exists an equilibrium in which broad and narrow searchers interact. Varying the instability parameter η then gives rise to an upward-sloping Beveridge curve. These regularity conditions essentially require that broad searchers do not value the segment under study too differently from other segments in their search range.

In what follows, we first describe a set of common assumptions on preferences and perform some calculations that are helpful in all setups that we study. Appendix D.1 then studies two possible specification of a random matching model. For each setup, we describe when the random matching model gives rise to the flow equations in Section 4. We also provide conditions under which variation in the instability parameter η implies an upward-sloping Beveridge curve. In Appendix D.2, we instead consider competitive search (that is, directed search with price posting) and provide analogous results.

Basic Setup

Throughout this analysis, we make the same standard assumptions on preferences as in the main paper. Agents have quasilinear utility over two goods: numeraire and housing services. They can own at most one house. When an agent moves into a house, he obtains housing services v until he becomes unhappy with the house, which happens at the rate η. Once an agent is unhappy with his house, he 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 segment under study is "small" relative to the number of segments considered by broad searchers. This implies that a broad searcher who leaves the segment assigns probability zero to matching in that segment again. As a result, his continuation utility upon selling his house is independent of local conditions in the segment under study. In contrast, narrow searchers know that they will never leave the segment; their continuation utility is endogenous and varies with local housing market conditions.

In all models we consider, agents who own a home decide whether to put it on the market and contribute to inventory L. Agents who do not own decide whether or not to search and contribute to the buyer pool B. Moreover, these decisions are straightforward: owners put their house on the market if and only if they are unhappy and all non-owners search. These properties follow because utility is increasing in housing services and search is costless. They are not affected by the nature of matching or the outside option for broad searchers.

Alternative Assumptions on Search

In the following sections, we compare two popular formulations for matching and price determination: random search, where prices are determined by Nash bargaining between potential buyers and sellers after a match has occurred, and directed search, where sellers post specific prices, and buyer flows potentially respond to these prices.

Under *random search*, matches occur at the rate $m(B, L)$, where B and L represent the *total* number of buyers and sellers in the segment. The matching function is increasing in both arguments and has constant returns to scale. Transactions occur when the match-surplus is positive, and prices in each match are determined by ex-post Nash bargaining over that surplus.

Within random search models, we analyze two specifications for how broad searchers flow to different segments within their search range. The first specification was introduced in Section 4: broad searchers flow to segments *in proportion to segment inventory* L. The idea is that broad searchers scan available inventory and determine their favorite house, which is the only house that would yield them utility. They are therefore not indifferent to living in any other segment, and their outside option is to go back to the buyer pool and scan inventory again. The probability that a broad searcher finds her favorite house in the segment under study is qL for some constant q.

The second specification for buyer flows in the random search model is that broad searchers must be *indifferent* in equilibrium between trying to buy a house in all segments in their range. It is based on the idea that selection within an agent's search range responds only to overall market conditions in the segment. The function B^B is then determined from that indifference condition, and does not necessarily have to be increasing in inventory.

In contrast, under *competitive search*, sellers post prices and buyers direct their search effort to a segment with a particular price. It is useful to think of a submarket identified by a price, so buyers choose the submarket to visit. If $L(p)$ sellers post the price p and $B(p)$ buyers visit submarket p, matches there occur at the rate $m(B(p), L(p))$. We will study an equilibrium with directed search in which borrowers are again indifferent between directing their attention to any particular submarket within their search range.

Characterizing equilibrium

The key property of equilibrium that implies the flow equations in the text is that agents of different types cycle across states (buyer, seller and happy owner) at the same rates. It holds in a random matching setup if all matches result in a transaction: we then have common transition rates $m(B, L)/B$, $m(B, L)/L$ and η out of the buyer, seller and happy owner states, respectively. The same transition rates obtain in a competitive search setup if all buyers visit the same submarket. Below we establish the existence of equilibria with this property.

To emphasize common elements across equilibrium concepts, we now collect equations that always hold in the equilibria we study. We first restate the flow equations (3) and (4) using the accounting identity $I = L/H$ as:

$$\begin{aligned} \eta (1 - L/H) \left(N - B^N\right) &= m\left(B^N, L/H\left(N - B^N\right)\right), \\ \frac{B}{H} &= \frac{B^N}{N - B^N}. \end{aligned} \tag{D.1}$$

The first equation equates houses put up for sale by narrow searchers to new matches by narrow searchers. The second equation says that the buyer-owner ratio has to be the same for all types as well for the aggregate number of agents active in the segment.

We need notation to describe equilibrium values and prices. We index broad and narrow types by $j = B, N$, respectively, and denote the values of a type-j happy owner, seller, and buyer by U_H^j, U_S^j, and U_B^j, respectively. We further write U_E^j for the value of type j after he has sold his house in the segment under study. Transaction prices can, in general, depend on both the seller and the buyer type: we denote by $p(j, k)$ the price at which a type-j seller sells to a type-k buyer. We further write $E_B^j[p(k, j)]$ and $E_S^j[p(j, k)]$ for type j's expected transaction price when he is a buyer or seller, respectively, where the probabilities are inferred from the buyer and seller pools.

Consider next the steady state Bellman equations at the optimal actions. If types cycle across states at the same rates, type j's values are related by

$$\begin{aligned} rU_H^j &= v + \eta\left(U_S^j - U_H^j\right), \\ rU_S^j &= \frac{m(B, L)}{L}\left(E_S^j[p(j,k)] + U_E^j - U_S^j\right), \\ rU_B^j &= \frac{m(B, L)}{B}\left(U_H^j - U_B^j - E_B^j[p(k,j)]\right). \end{aligned} \tag{D.2}$$

In addition, for narrow buyers we have $U_E^N = U_B^N$: a narrow buyer who has sold again searches in the segment under study. In contrast, for broad searchers the utility after sale U_E^B is given exogenously.

For each equilibrium concept, we need to find the 13 numbers: 6 values (U_H^j, U_S^j, U_B^j for $j = N, B$), 4 prices ($p(j, k)$ for $j = N, B$ and $k = N, B$) and the 3 quantities L, B, and B^N. We have 6 Bellman equations in (D.2) and the two flow equations in (D.1). For each equilibrium concept, there will be 4 additional equations that determine prices. The final equation comes from the assumption on the behavior of broad searchers. With indifference, we have $U_B^B = U_E^B$. With proportional flows, we have instead the condition $B = qL + B^N$.

D.1 Equilibrium with random matching and bargaining

In an equilibrium with random matching and bargaining, buyers and sellers meet at random. Conditional on a match, a transaction occurs if match surplus is positive. Otherwise, the buyer and seller revert back to their respective pools. The outside options of a type-k buyer and a type-j

seller are U_B^k and U_S^j, respectively. Surplus in a match is the sum of buyer and seller surplus

$$\left(U_H^k - p\left(j,k\right) - U_B^k\right) + \left(U_B^j + p\left(j,k\right) - U_S^j\right).$$

We look for equilibria in which all matches lead to a transaction. Let θ denote the bargaining weight of the buyer. If a transaction occurs, the price is set so that the seller's surplus is a share $1-\theta$ of total surplus (Nash bargaining), so that:

$$p\left(j,k\right) = \left(1-\theta\right)\left(U_H^k - U_B^k\right) + \theta\left(U_S^j - U_B^j\right). \tag{D.3}$$

This formula delivers four equations for price formation that characterize equilibrium together with the flow and Bellman equations, (D.1) and (D.2). We now distinguish between specifications of buyer flows based on indifference and proportional flows – which contribute the remaining equation – and work out the equilibrium in each case.

D.1.1 Proportional flows

As discussed above, a first possible assumption about how buyers flow to segments within their search range is that they flow to segments in proportion of inventory in those segments. Based on the empirical evidence presented in the main body of the paper, we choose this assumption in the set-up of our single-segment model in Section 4, as well as in the quantitative model.

With the proportional-flow assumption, the remaining equation to fully define the equilibrium is $B = qL + B^N$. It follows that the system of equations can be solved in two blocks. First, the flow equations alone determine a unique solution for inventory and the number of searchers. Indeed, in the $\left(B^N, I\right)$ plane, the equation $\frac{B^N}{N-B^N} = qIH + B^N$ describes a continuous upward-sloping schedule that converges to infinity as $B^N \to N$. The first equation in (D.1) describes a continuous monotonically decreasing schedule with $I = 1$ for $B^N = 0$. With these two relationships, there is a unique solution $\left(B^N, I\right)$, as drawn in the lower left panel of Figure 4. Given B^N and $L = IH$, the Bellman and price equations imply values for each type as well as prices for each buyer-seller pair.

To show existence of equilibrium in which all matches lead to transactions, it remains to check that surplus is positive in all transactions. We impose a restriction on broad types continuation utility: broad types' utility from buying in the segment under study has to be sufficiently similar to utility from buying in other segments. We thus make sure that sellers do not prefer to skip trades with a certain type of buyer in order to wait for a buyer who is willing to pay more. Similarly, we want buyers not to skip trades with a seller of a certain type in order to wait for a seller who is willing to accept a lower price.

A simple sufficient condition for existence of equilibrium is

$$U_E^B = \frac{v}{r}\frac{\theta m(B,L)/B}{(r+\eta+(1-\theta)\,m(B,L)/L)}, \tag{D.4}$$

where B and L follow from the flow equations. The right hand side is the value of a buyer in a hypothetical market in which all types are narrow, but with B and L reflecting actual buyer and seller flows, including broad types. The condition is satisfied in particular *if the seller has all the bargaining power*, as we assume in our quantitative exercise in Section 5. Indeed, if $\theta = 0$, the seller makes a take-it-or-leave-offer and $U_E^B = 0$, so buyers have no power in other segments, then (D.4) holds exactly.

Condition (D.4) implies that there is an equilibrium in which all valuations are the same across types and surplus is positive in all transactions. Indeed, suppose valuations are the same. By (D.3), prices are then the same in all transactions. It follows from (D.2) that the value of a narrow searcher is exactly equal to the right hand side. The condition says that broad types' expected continuation utility U_E^B matches that value. Working through (D.2), the result then follows.

Condition (D.4) is not necessary: even if it is not satisfied, there can be an equilibrium in which all matches lead to transactions, so the flow equations from the text continue to hold. Indeed, the surplus in the benchmark equilibrium

$$\frac{v}{r+\eta+(1-\theta)\,m(B,L)/L}$$

is strictly positive. Small enough changes to parameters will therefore not alter equilibrium flows. Of course, we may have different prices in different buyer-seller meetings. Since the equilibrium flows separate into blocks, however, this does not affect flows.

Upward-sloping Beveridge curve

With proportional flows, an increase in the instability parameter η always increases both inventory and search activity. In other words, variation in η generates an upward-sloping Beveridge curve. This follows directly from the bottom left panel in Figure 4. Indeed, with proportional flows the upward-sloping curve is independent of η, whereas the downward-sloping schedule shifts up with η. Intuitively, an increase in η means that inventory must increase, which attracts more broad searchers and hence leaves more narrow types without a house, thus increasing search activity.

D.1.2 Indifference

In a directed-search equilibrium with indifference of buyers across all segments in their search range, the remaining equation is $U_B^B = U_E^B$. In contrast to the case of proportional flows, we no longer have two separate blocks for flows as well as values and prices. Instead, we need to jointly solve (D.2) and (D.1) for values, prices, and the equilibrium flows. As before, values depend on the queue length $q = B/L$ that governs matching probabilities. What is new now is that values feed back to the queue length since the number of broad searchers B^B is chosen to ensure indifference.

If θ and U_B^E are either both zero, or if they are both positive and N is sufficiently large, then there exists an equilibrium in which all values are equated and all matches lead to transactions. Indeed, assuming equal values and hence equal prices in all transactions, we can find values from (D.4) for a given q. In particular, the value of a narrow buyer equals the right hand side of (D.4). To verify that buyer values are indeed equal, it must be possible to choose q so as to satisfy (D.4). If this is the case, all other values will be equal also. We thus study the existence of a solution q to

$$rU_B^E\left(r+\eta+(1-\theta)\,m\left(q,1\right)\right)=v\theta m\left(1,q^{-1}\right)$$

If $U_B^E = \theta = 0$ the condition is clearly satisfied. Consider the case where both are positive. Since the matching function is homogeneous of degree one, the left hand side is increasing in q and the right hand side is decreasing. We assume that $m\left(1,q^{-1}\right)$ goes to infinity for $q \to 0$, which ensures existence of a unique solution for q. Inventory follows as $L = H\eta/\left(\eta+m\left(q,1\right)\right)$, the total number of searchers is $B = qH\eta/\left(\eta+m\left(q,1\right)\right)$ and the number of narrow searchers is $B^N = qH\eta/\left(\left(q+1\right)\eta+m\left(q,1\right)\right)$. We require that N is sufficiently large so $B^N < N$.

Upward-sloping Beveridge curve

When can variation in the separation rate η generate an upward-sloping Beveridge curve in this setting? The following result shows how inventory and search activity respond locally to η.

Proposition 1. If the initial equilibrium satisfies

$$\frac{(L)^2}{\left(H-L\right)^2}<1-\theta+\frac{r}{m\left(B,L\right)/L}\varepsilon_L,$$

then a small increase in η leads to higher inventory, a larger number of total searchers B and a larger number of narrow searchers B^N.

The proof is provided below. The proposition says that faster separation increases inventory and crowding out if the initial equilibrium inventory share I is small enough relative to (i) the interest rate, (ii) the elasticity of the matching function relative to inventory, (iii) the bargaining weight of the sellers, and (iv) the ratio of volume to inventory m/L. In particular, given the small inventory shares observed in our data, on the order of 0.05, most bargaining weights for the seller will guarantee the result. Even for $\theta = 1$, the fact that I/V is typically larger than one will guarantee the result unless the interest rate is very small.[29]

[29]The proposition provides a condition in terms of endogenous variables, as opposed to primitives of the model. We choose this type of statement since it relates naturally to our observables. The key inference is that if we were to quantify a model such that it matches low inventory shares as in the data, then an increase in η will move the

Why is a condition needed? An increase in the separation rate gives rise to two counteracting effects. On the one hand, more houses come on the market so inventory increases, more broad searchers flow in and crowd out narrow searchers. This is the dominant effect discussed in the text and also illustrated in the proportional-flows case above. On the other hand, an increase in η also lowers the surplus of a match, given by

$$U_H - U_S = \frac{v}{r + \eta + (1-\theta)\, m(B,L)/L}$$

This is because any buyers knows that he will become unhappy more quickly. The decrease in the value of a match implies that the queue length must become shorter to keep broad searchers indifferent between the segment and their outside option. Per unit of inventory, then, there will be fewer searchers overall. The condition tells us that under plausible conditions on housing markets, this second effect is weak.

Proof of Proposition 1. B and L are determined from

$$\begin{aligned} \eta(H-L) &= m(B,L), \\ B\left(r/m(B,L) + 1/(H-L) + (1-\theta)/L\right) &= \frac{\theta v}{rU_B}, \end{aligned}$$

The first equation describes a downward-sloping schedule in the (B,L) plane that shifts up if η increases (that is, L increases for given B). The second equation describes a schedule that is independent of η. If it is locally increasing, then a change in η locally increases B and L. The condition ensures that this is the case. Indeed, the LHS of the second equation is increasing in B. It is also decreasing in L if

$$-\frac{rm_2(B,L)}{m(B,L)^2} + \frac{1}{(H-L)^2} - \frac{1-\theta}{L^2} < 0.$$

Rearranging delivers the condition in the proposition. ■

D.2 Competitive search

In a competitive search equilibrium, owners decide as before to put their house on the market and non-owners decide whether to search. The new feature is that putting a house on the market requires posting a price. There are many submarkets identified by price and searchers direct their search to one submarket. Broad searchers may also direct their search to other segments and earn the outside option U_E^B. The matching function says how many matches there are in the submarket given the number of buyers and sellers. An equilibrium consists of listing and search decisions, together with a price such that all decisions are optimal given others' choices.

equilibrium up along an upward-sloping Beveridge curve.

We further impose a common "subgame perfection" requirement to restrict sellers' beliefs about how many customers they can attract by posting a particular price. A seller who posts price $\tilde{p}$ expects to attract a queue $\tilde{q}$ such that broad buyers are indifferent between his posted price and the best price available elsewhere in the market. In steady state equilibrium, utility at the best price is captured by the equilibrium values that satisfy the Bellman equations. We focus on equilibria in which valuations are equated across types in equilibrium – they will exist under similar conditions as for the case of random matching.

To derive the key pricing equation, consider the price posting choice of an individual seller. He chooses a price $\tilde{p}$ and a queue length $\tilde{q}$ to solve

$$\max_{\tilde{p},\tilde{q}} m(\tilde{q},1)(\tilde{p}+U_B-U_S)$$
$$\text{s.t. } rU_B = \frac{m(\tilde{q},1)}{\tilde{q}}(U_H-U_B-\tilde{p})$$

The constraint captures indifference of buyers between their best option in the market U_B and the utility from visiting the submarket with price $\tilde{p}$.[30]

We can solve the constraint for the price $\tilde{p}$ and substitute into the objective function. The first-order condition for $\tilde{q}$ then delivers the optimal queue length from $m_1(\tilde{q},1)(U_H-U_S)=rU_B$. In equilibrium, only one submarket is open in each segment, with equilibrium price p and equilibrium queue $q=B/I$. Substituting into the first-order condition and using the Bellman equation for buyers, we obtain

$$p = U_H - U_B + \frac{(B/I)\, m_1(B/I,1)}{m(B/I,1)}(U_H-U_S).$$

This equation takes the same form as (D.3), with the bargaining weight of the seller replaced by the elasticity of turnover relative to inventory.

With a Cobb-Douglas matching function, an equilibrium exists if U_E^B is positive and N is large enough. The argument builds on that for existence with random matching and indifference. Indeed, in the Cobb-Douglas case, the power on B in the matching function takes the spot of his bargaining weight in the pricing equation. The earlier argument for positive θ and U_B^E thus goes through unchanged.

Upward-sloping Beveridge Curve

[30]Intuitively, the constraint works like a demand function,

$$\tilde{p} = U_H - U_B - \frac{\tilde{q}}{m(\tilde{q},1)} rU_B$$

For example, with Cobb-Douglas matching $m(b,s)=\bar{m}b^\delta s^{1-\delta}$, we have $\tilde{p}=a-b\tilde{q}^{1-\delta}$, so demand is "more elastic" if δ is higher. A seller who undercuts other sellers by charging a lower price then attracts a longer queue. As a result, his probability of selling goes up, which is good for profits. The individual seller looks for the sweet spot where the change in profit from changing the queue offsets the change in profit from changing the price.

With a Cobb-Douglas matching function, the proposition of the previous subsection applies directly and we have again that variation in the separation rate generates an upward-sloping Beveridge curve. More generally, the sellers' share of surplus may also change with η. Since the proposition is about small local changes, it continues to hold for matching functions that are close to Cobb-Douglas. More general conditions could be derived for other functions; we do not pursue this extension here.

E Prices and Frictional Discounts

In this appendix, we study price formation. We first illustrate the forces driving prices in the context of the simple model described in Appendix D. We focus on equilibria in which all matches lead to a transaction and valuations are equal across types, and decompose the price into a frictionless price as well as an adjustment for search and transactions costs. We then consider the setup of our quantitative model in Section 5 and derive the approximate price formula (8) used in the text to interpret our quantitative results.

Prices in the Simple Model

With equal values, the Bellman equations (D.2) imply that the steady state price is the same in all transactions and satisfies

$$p = \frac{v}{r} - \frac{v}{r}\frac{\eta + \theta\left[r + \left(m\left(B,L\right)/L\right)\,\left(L/B\right)\right]}{r + \eta + \left(1-\theta\right)m\left(B,L\right)/L} - \frac{c\,p}{r}\frac{m\left(B,L\right)}{L}\frac{\eta - \theta\left(\eta + \left(r+\eta\right)L/B\right)}{r + \eta + \left(1-\theta\right)m\left(B,L\right)/L}, \tag{E.1}$$

where θ is either the bargaining weight of the buyer (with random search) or the power on buyers in a Cobb-Douglas matching function (with competitive search). In a completely frictionless market, we have no transaction costs ($c = 0$) and instantaneous matching ($m\left(B,L\right)/L \to \infty$ and $L/B \to 0$) and the price is given by the first term v/r, the present value of the housing service flow.

More generally, the second and third terms represent frictions in the housing market. The second term is a discount for search. It is zero if matching is instantaneous and it is increasing in the instability parameter η: in less stable segments search becomes necessary more often. The search discount is also larger if the buyer has a larger bargaining weight. Intuitively, more powerful buyers can force the seller to bear more of the search cost via a lower sales price.

The third term in (E.1) reflects the presence of transactions costs which is zero if $c = 0$. As matching becomes instantaneous, it converges to the present value of transaction costs cV/r. Indeed, as matching becomes faster, the fraction (E.1) converges to $cp\eta/r$, and we also have $I \to 0$ and hence $V \to \eta$. The basic force that higher volume segments have lower prices due to the capitalization of transaction costs is thus present even if there are no search frictions.

Prices in the Quantitative Model

We now derive a convenient formula for prices in the quantitative model. It resembles (E.1) for $\theta = 0$. Indeed, in the quantitative model prices are the same across all transactions since the seller has all the bargaining power. Formally, let $U_F\left(h;\theta\right)$ denote the utility of a type-θ agent who obtains housing services from a house in segment h. Since sellers make take-it-or-leave-it 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 $p\left(h,\theta\right) = U_F\left(h;\theta\right).$

We now show that prices are the same in all transactions in segment h. We start from the

Bellman equation of a seller who puts his house on the market

$$rU_S(h;\theta) = \frac{V(h)}{I(h)}\left(E\left[p(h,\theta)(1-c)|h\right] - U_S(h;\theta)\right),$$

where the expectation uses the equilibrium distribution of buyers of type θ. It follows that the value function of the seller is independent of type. Intuitively, a seller knows that once he becomes a buyer, his continuation value is zero. Seller utility thus derives only from the expected sale price, about which all seller types care equally.

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

$$rU_F(h;\theta) = v(h) + \eta(h)\left(U_S(h;\theta) - U_F(h;\theta)\right).$$

Since utility $v(h)$ and the arrival of moving shocks are also independent of type, so is $U_F(h;\theta)$. As a result, the same price $p(h)$ is paid in all transactions in segment h. We can combine these equations and determine the price from:

$$p(h) = \frac{v(h)}{r} - \frac{\eta(h)}{r + V(h)/I(h) + \eta(h)}\left(\frac{v(h)}{r} + \frac{cp(h)}{r}\frac{V(h)}{I(h)}\right). \tag{E.2}$$

In a given equilibrium, this formula relates the segment price $p(h)$ to the service flow $v(h)$ as well as parameters and observables fit by the model. In particular, it implies a one-to-one relationship between service flow and price – except for knife-edge situations which do not occur in our exercise, segments with different service flow will see different prices. Through the lens of the model, search ranges defined in terms of price can thus be viewed as reflecting differences in the service flow – in the paper, we refer to this as the "quality" of the segment.

We thus obtain a useful shortcut to interpret the numerical results below: solving out, we write the price as

$$p(h) \approx \frac{v(h)(1-I(h))}{r + cV(h)} = \frac{v(h)}{r}(1-I(h))\frac{r}{r+cV(h)}.$$

which is the approximate price formula (8). In our quantitative model in Section 5, the approximation is very good. Indeed, the maximal approximation error across all segments is 15 basis points.

F Identification in Quantitative Model

This appendix states a system of equations that characterizes steady state equilibrium in the quantitative model of Section 5, and then shows how the parameters of that model can be identified from data on search and housing market activity.

Characterization of Equilibrium

We derive a system of equations that determines the steady state distribution of agent states (that is, searching for a house, listing one for sale, or owning without listing). Since there are fixed numbers of agents and houses, that distribution can be studied independently of prices and value functions. We need notation for the number of agents in each state. Let $H(h;\theta)$ denote the number of type-θ agents who are homeowners in segment h, and let $L(h;\theta)$ denote the number of type-θ agents whose house is listed in segment h. In steady state, all those numbers, as well as the numbers of buyers by type $\tilde{B}(\theta)$ and by segment $B(h)$, are constant.

The first set of equations uses the fact that the number of houses for sale $L(h)$ 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:

$$\eta(h)(H(h) - L(h)) = \tilde{m}(B(h), L(h), h). \tag{F.1}$$

The left-hand side shows the number of houses coming on the market, given by the rate at which houses fall out of favor multiplied by the number of houses that are not already on the market. The right-hand side shows the 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:

$$\frac{H(h;\theta)}{H(h)} = \frac{L(h)}{\tilde{L}(\theta)} \frac{\tilde{B}(\theta)}{B(h)}. \tag{F.2}$$

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 (6), 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 $B(h)/H(h)$.

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

$$\begin{aligned} H(h) &= \sum_{\theta \in \tilde{\Theta}(h)} H(h;\theta), \\ \mu(\theta) &= \tilde{B}(\theta) + \sum_{h \in \tilde{\mathcal{H}}(\theta)} H(h;\theta). \end{aligned} \tag{F.3}$$

Equations (F.1), (F.2) and (F.3) jointly determine the unknown objects $L(h)$, $B(h)$, $H(h;\theta)$, and $\tilde{B}(\theta)/\bar{B}$, a system of $2|\mathcal{H}| + |\Theta| \times |\mathcal{H}| + (|\Theta| - 1)$ equations in as many unknowns.

Identification

Our model implies a one-to-one mapping between two sets of numbers. The first set consists of the parameters $\eta(h)$ and $\mu(\theta)$ as well as the vector of rates at which buyers find houses in a given segment, defined as $\alpha(h) = m(h)/B(h)$. The second set consists of listings $L(h)$ (or, equivalently, the inventory share $I(h) = L(h)/H(h)$), the turnover rate $V(h)$, the relative frequencies of search ranges $\tilde{B}(\theta)/\bar{B}$, and the *average* time it takes for a buyer to find a house.

Dividing (F.1) by $H(h)$, the frequency of moving shocks $\eta(h)$ can be written directly as a function of inventory and turnover:

$$\eta(h)(1 - I(h)) = V(h). \tag{F.4}$$

Using the definition of buyers (6), the match rate $\alpha(h)$ for a buyer who flows to segment h can be expressed in terms of observables (up to a constant) as

$$\frac{1}{\alpha(h)} = \frac{B(h)}{m(h)} = \sum_{\theta \in \tilde{\Theta}(h)} \frac{I(h)}{\tilde{I}(\theta)} \frac{\tilde{B}(\theta)}{\tilde{H}(\theta)} \frac{1}{V(h)}, \tag{F.5}$$

where $\tilde{I}(\theta) = \tilde{L}(\theta)/\tilde{H}(\theta)$. Interpreting terms from the right, we have that matching is fast – at a high rate $\alpha(h)$ – in segment h if the turnover rate is high in h, if the buyer-owner ratio is low for types in the clientele of h, and if the inventory share is low in h relative to other segments in its clientele's search ranges.

It remains to identify the distribution of searcher types μ. We determine the constant $\bar{B}$ by setting the average of the buyer match rates $\alpha(h)$ to the average of the inventory match rates $m(h)/L(h) = V(h)/I(h)$. The number of buyers by type $\tilde{B}(\theta)$ and by segment $B(h) = m(h)/\alpha(h)$ follow immediately. Substituting for $H(h;\theta)$ in (F.3) using (F.2) we can solve out for the type distribution $\mu(\theta)$ from

$$\frac{\mu(\theta) - \tilde{B}(\theta)}{\tilde{B}(\theta)} = \sum_{h \in \tilde{H}(\theta)} \frac{L(h)}{\tilde{L}(\theta)} \frac{H(h)}{B(h)}. \tag{F.6}$$

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, 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 follow from equation (F.1). For example, consider the Cobb-Douglas case with a multiplicative segment-specific parameter $\bar{m}(h)$ that governs the speed of matching

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

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