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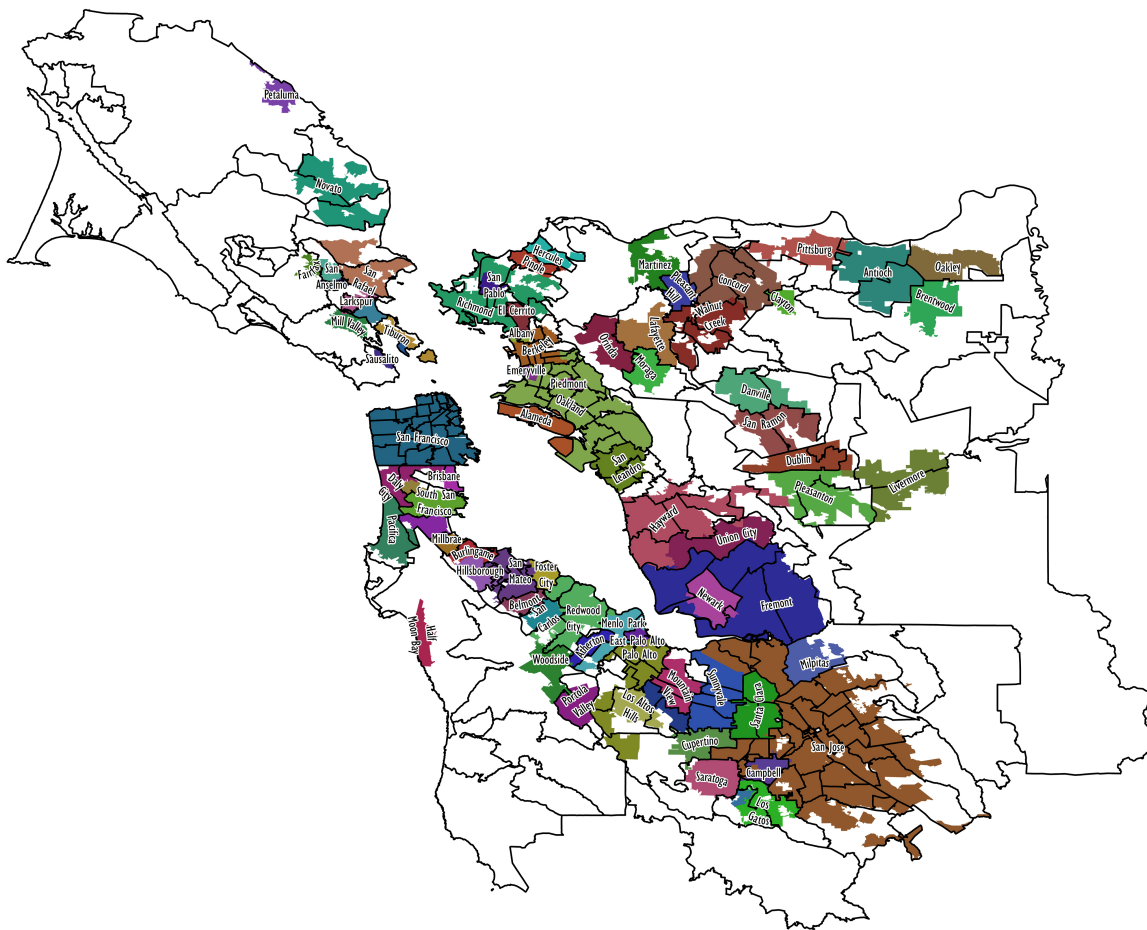
Note: This figure shows the geographic distribution of zip codes and cities in the San Francisco Bay Area. The base map are zip codes, the colored regions correspond to cities.

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	69.9

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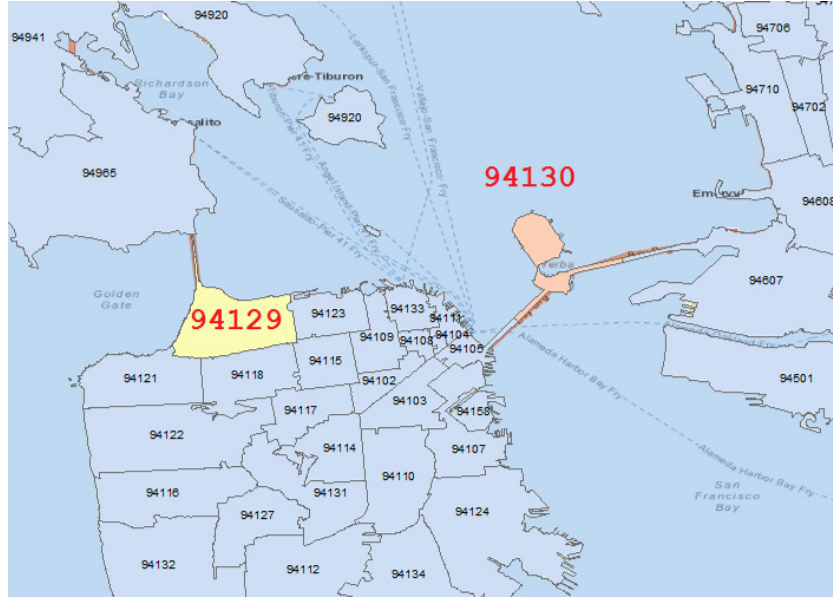
Note: This figure shows how we deal with bridges in the Bay Area for the contiguity analysis.

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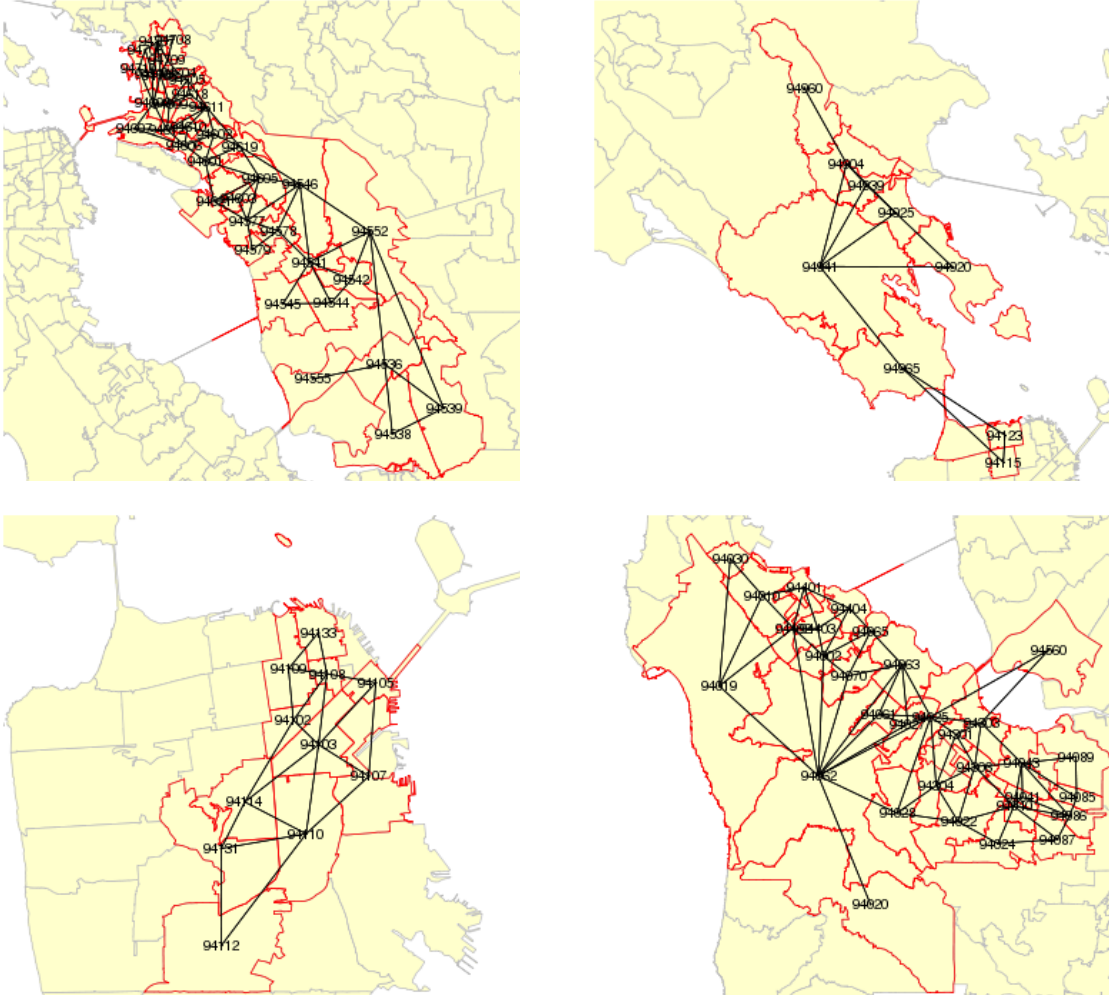
Note: This figure shows a sample of contiguous search sets. The zip codes selected by the searcher are circled in red. Zip code centroids of contiguous zip codes are connected.

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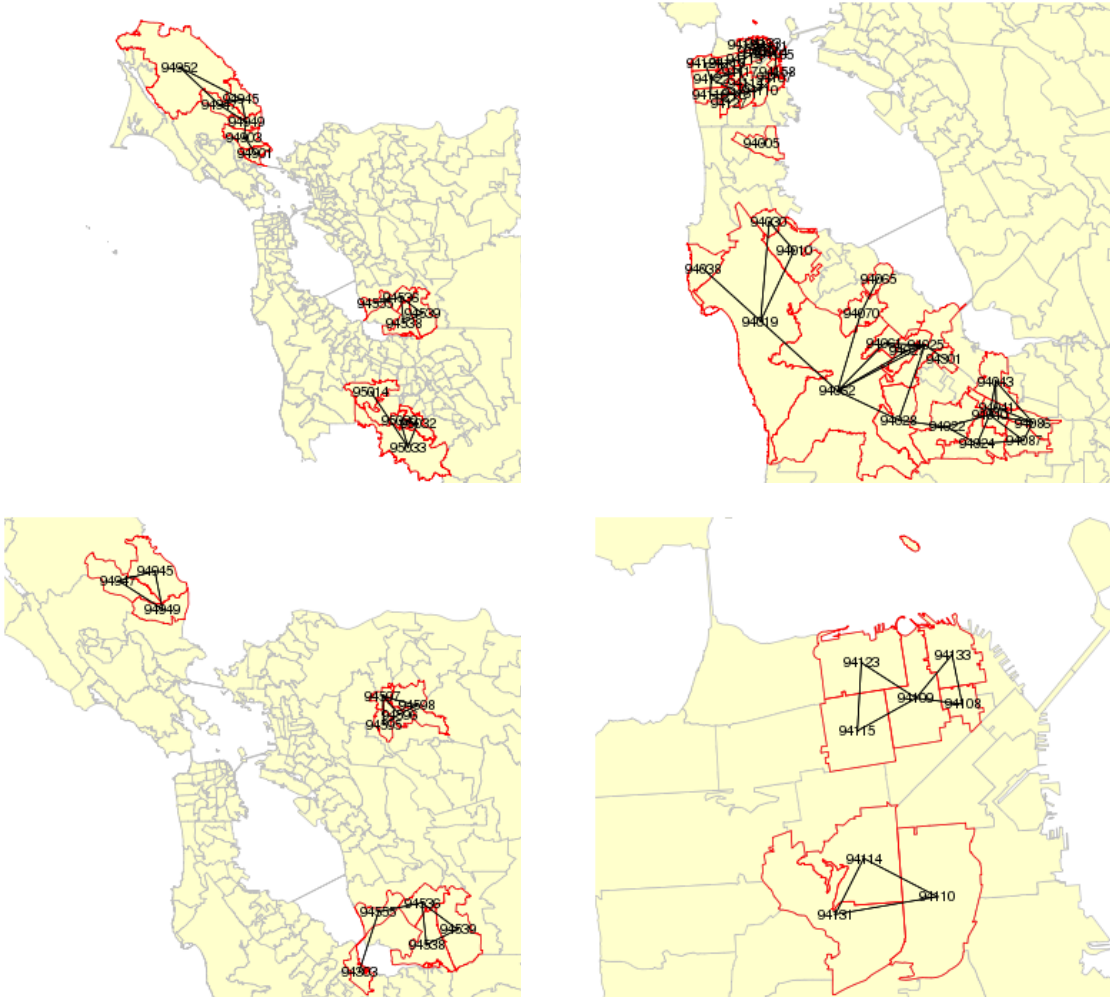
Note: This figure shows a sample of non-contiguous search sets. The zip codes selected by the searcher are circled in red. Zip code centroids of contiguous zip codes are connected.

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

| | | CONTIGUOUS AREAS | | |
Number of Zips Covered	Share Contiguous	Mean	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
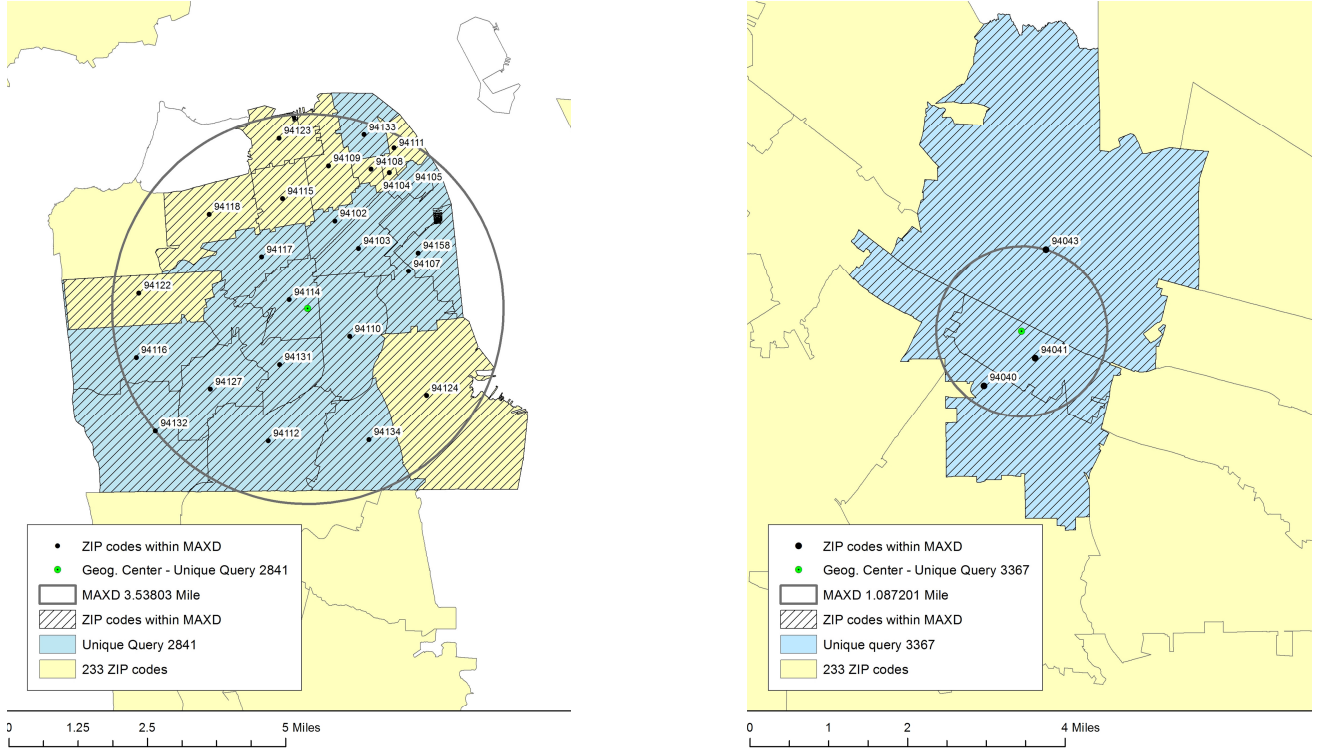
Note: Figure provides examples of the circularity measure. All zip codes that are part of the search set are shown in blue. The geographic center of each search set is given in green. The circle is centered around this geographic center and has radius equal to the furthest distance of any zip code centroid 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
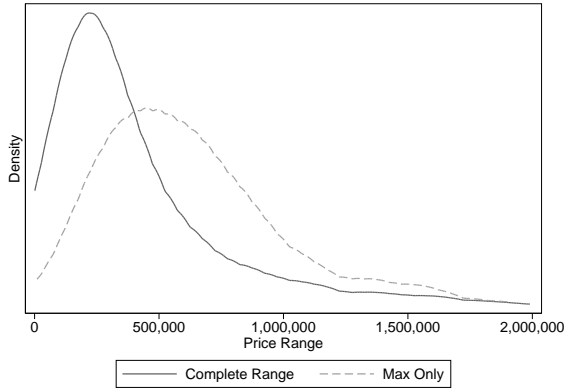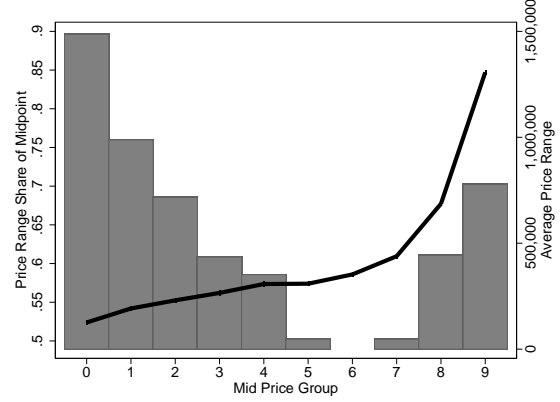
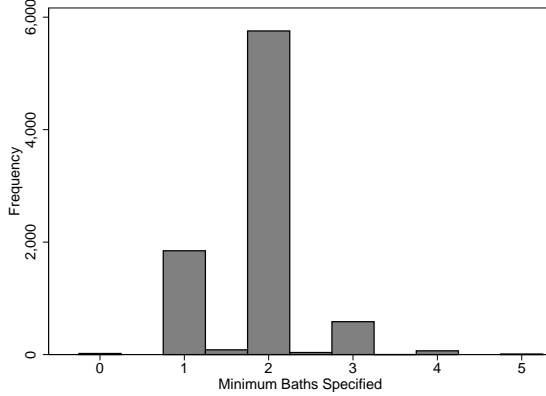(A) Minimum House Price
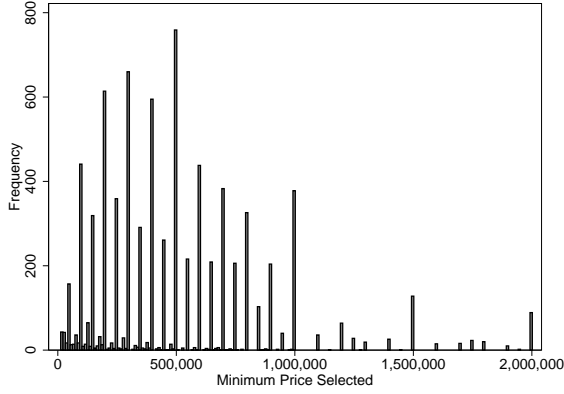
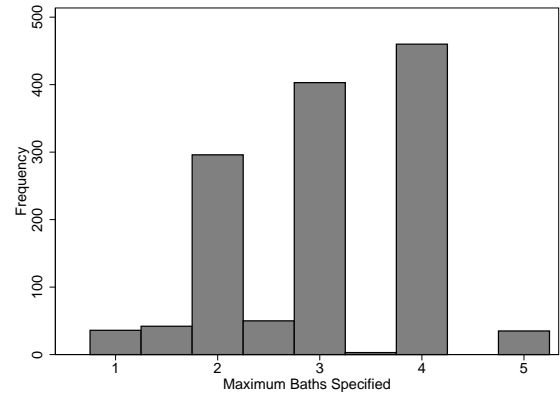(B) Maximum House Price
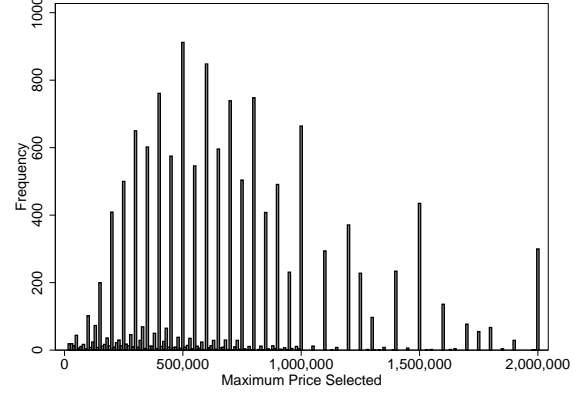
(C) Price Range

(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 II.B, 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.
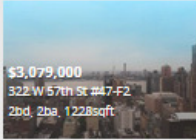
Figure A.7: Example of Search Return List

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)

Slope: 0.645

Slope: 0.684
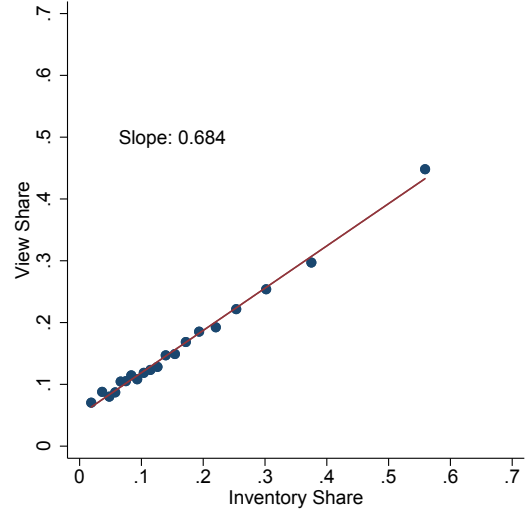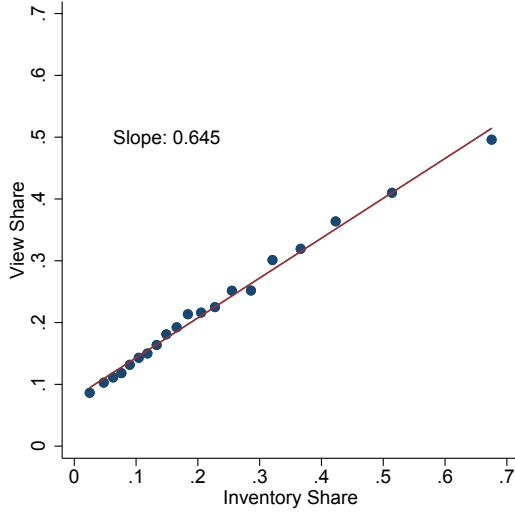
(C) At least 15 property views per session (N = 500) (D) At least 20 property views per session (N = 279)

Slope: 0.713

Slope: 0.719
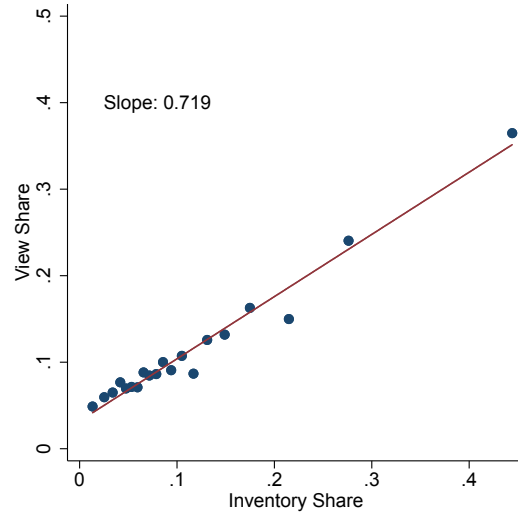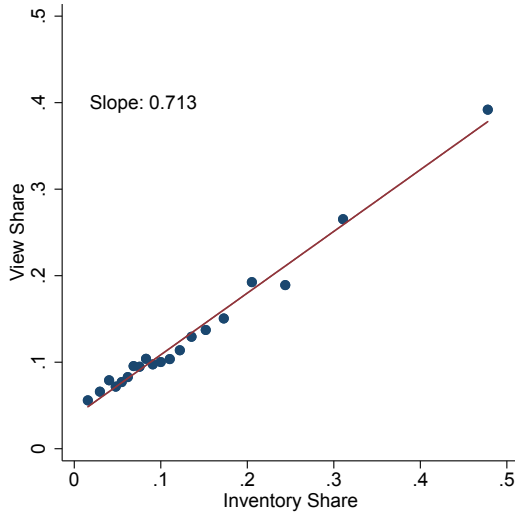
Note: Figure shows binscatter plots at the segment-session level; different panels vary the minimum number of property views required for a session to be included. On the horizontal axis is the share of inventory of a segment, relative to the total inventory in all segments viewed in that session. On 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 Appendix 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 1,500 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 1,500, we merge the two size segments. We then begin with the lowest price segment, see whether it has a stock of less than 1,500, and if not merge it with the next higher price segment. This procedure generates 564 segments. Only one of these segments, zip code 94111, has fewer than 1,500 housing units.

Figure B.1: Segment Overview
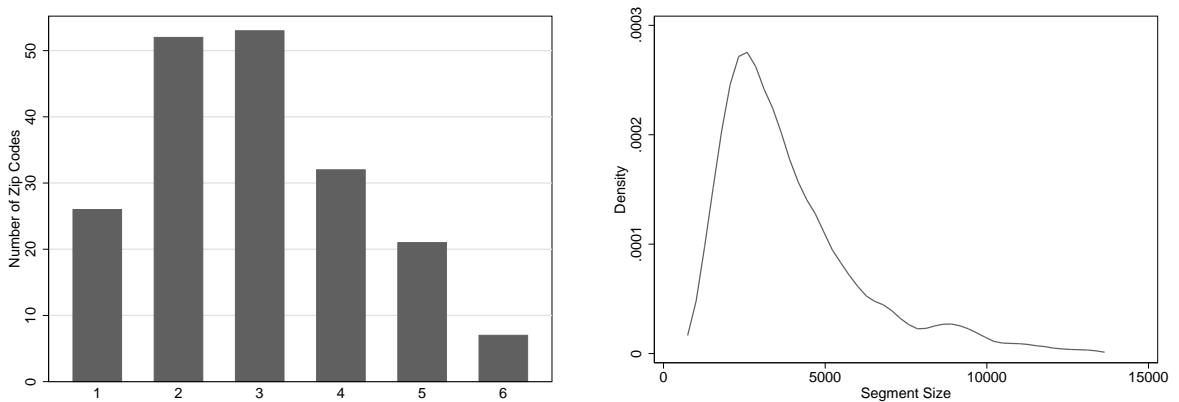
Note: The left panel shows the number of segments that the 191 zip codes are split into. The right panel shows the distribution of the number of housing units across segments.

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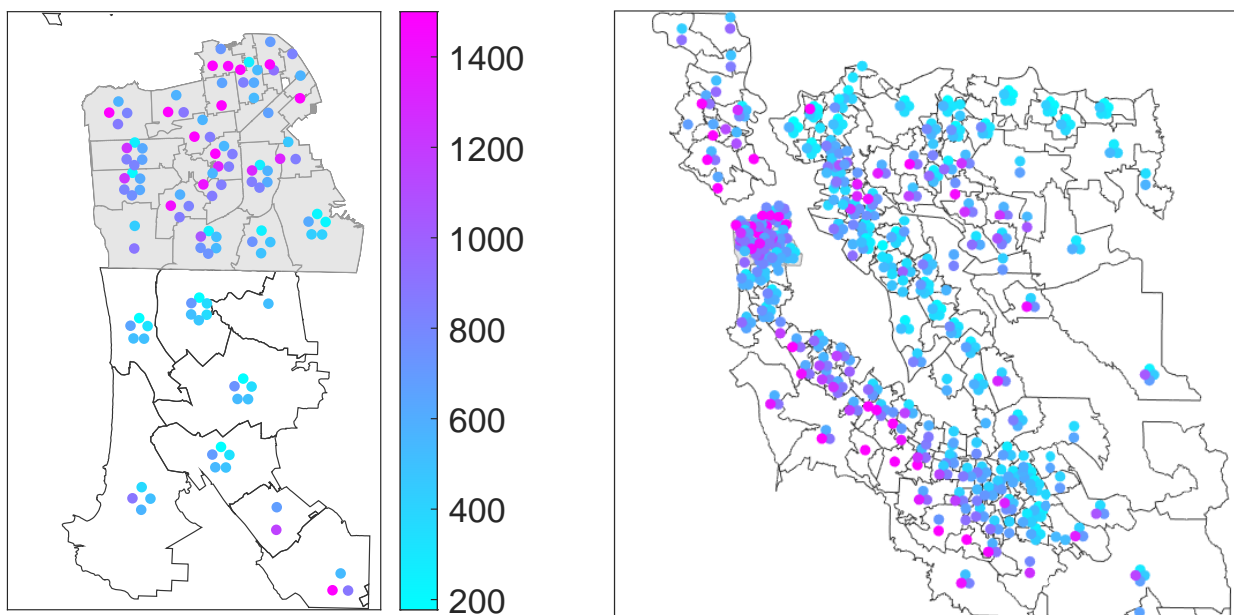
Note: The left panel shows a map of downtown San Francisco as the shaded area in addition to areas south of downtown. The right panel shows a map of the entire Bay Area. The color bar indicates the price of the segment in thousands of USD.

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 alert 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 9,008 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 7,287 search profiles represent only a single searcher, 630 search profiles represent two searchers. 482 search profiles represent at least seven searchers, while the most common search profile represents 1,017 unique searchers.

Our working assumption is that searchers are interested in all houses that their email alert covers. For searchers who specify a price range, this is a natural assumption, since those searchers have spent time to think about and specify a suitable range to query. The intent of searchers who set very broad alerts – e.g., by specifying only a very broad geography – is less obvious. For example, searchers whose query only specifies the entire city of San Jose may expect to be able to scan inventory quickly according to additional criteria that we do not observe. While our data

19

Figure B.3: Number of Searchers per Unique Profile
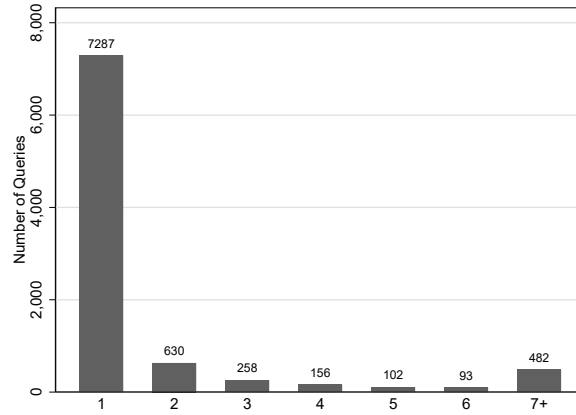
Note: Figure shows how many unique individuals are represented by each of the 9,008 individual search profiles.

do not allow us to determine whether searchers with broad ranges would truly be interested in all properties, we can make sure that our results are not driven by the existence of very broad alerts.

We thus introduce a screen for "questionable" alerts that cover a price dispersion that is unusually high in light of our evidence on explicitly set price ranges.[27] We have run our analysis both with all searches included as well as with the questionable searches removed from the sample. The results are very similar across the approaches. A plausible reason for this is that a city like San Jose is already integrated by broad searchers who set reasonable price ranges, so that adding very broad searchers has little additional effect. Since we emphasize the role of broad searchers in this paper, our baseline calibration follows a conservative approach that leaves out questionable queries, leaving us with a set of 4,956 unique queries set by 18,679 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

---

[27]Concretely, we first explore the empirical distribution of price ranges for those alerts who explicitly specify a price range. We do this separately for ten bins of searchers divided by the minimum price of their range: the smallest bin covers prices up to $100K and we move up in steps of $100K to the top bin with prices above $900K. For each bin, we measure the 95th percentile of the price range, and call it the "cutoff range" for that bin. We then turn to the universe of all alerts. For each alert – whether or not a price range is specified – we can measure the highest and lowest median segment price covered by the search. We place alerts into the above bins by their lowest median segment price. An alert is the identifed as "questionable" if the difference between the highest and lowest median segment price is above the cutoff range for its bin. For example, very broad alerts that specify the entire city of San Jose are questionable because they cover dispersion of prices that is much larger than the typical range with a minimum price in the cheapest San Jose segment.

hence the quality) of the property. This excludes, for example, intra-family transfers. We drop all observations that are not a Main Deed or only transfer partial interest in a property (see Stroebel, 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.[28] 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).

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).

We find segment-level measures of time on market by averaging the time on market across all transactions between 2008 and 2011.[29] 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

---

[28]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.

[29]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.

Figure B.4: Listings-Transaction Match Rates

(A) Listings without sales	(B) Sales with Listings
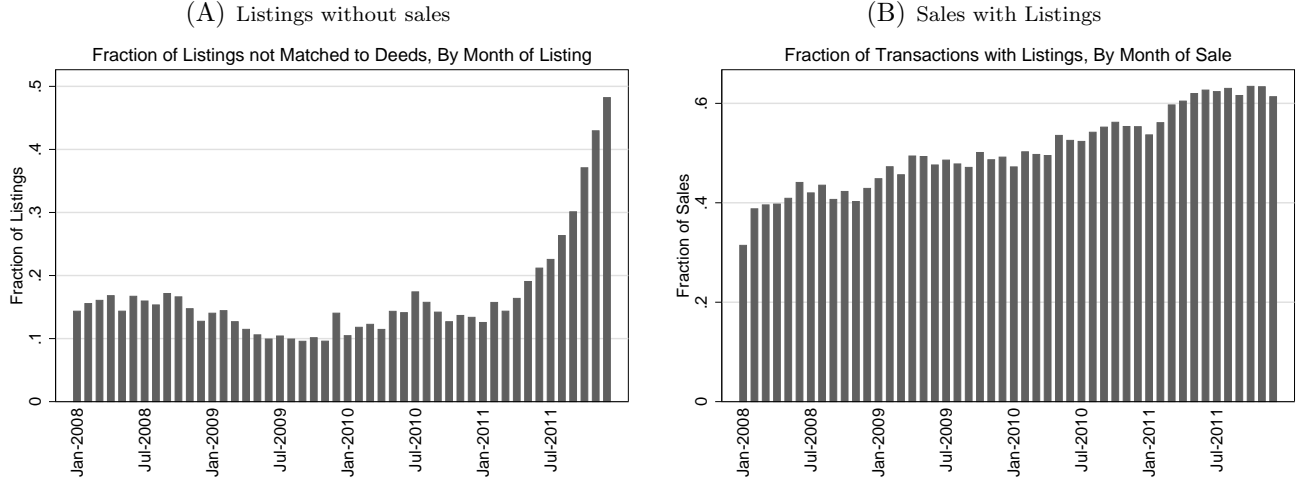
Note: Panel A shows the number of listings for which we do not eventually find a deed to match, by month of listing. Panel B shows the share of transactions for which we also observe a listing, by month of sale. Both Panels cover the period 2008-2011, for which we observe Trulia email alerts.

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 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

22

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.

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 seg-

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	5.5	3.7	6.3	84.5	9.7	7.2	17.8	65.3
Q25	0.94	1.04	0.0	0.0	0.8	72.5	0.0	0.0	3.1	54.5
Q50	1.56	1.57	1.2	1.3	2.5	93.7	5.2	2.0	12.7	67.5
Q75	2.53	2.17	4.6	3.7	6.7	97.1	15.9	6.3	26.4	79.6

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.

ment 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 5.5 percent of searchers select exactly the zip code containing that segment. An additional 3.7 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 4.6 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 V.B is that the search breadth would not adjust based on the amount of inventory within the range (see also our discussion in Section IV.B). 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

(A) Maximum Distance

(B) Mean Distance

(C) Share Contiguous Queries

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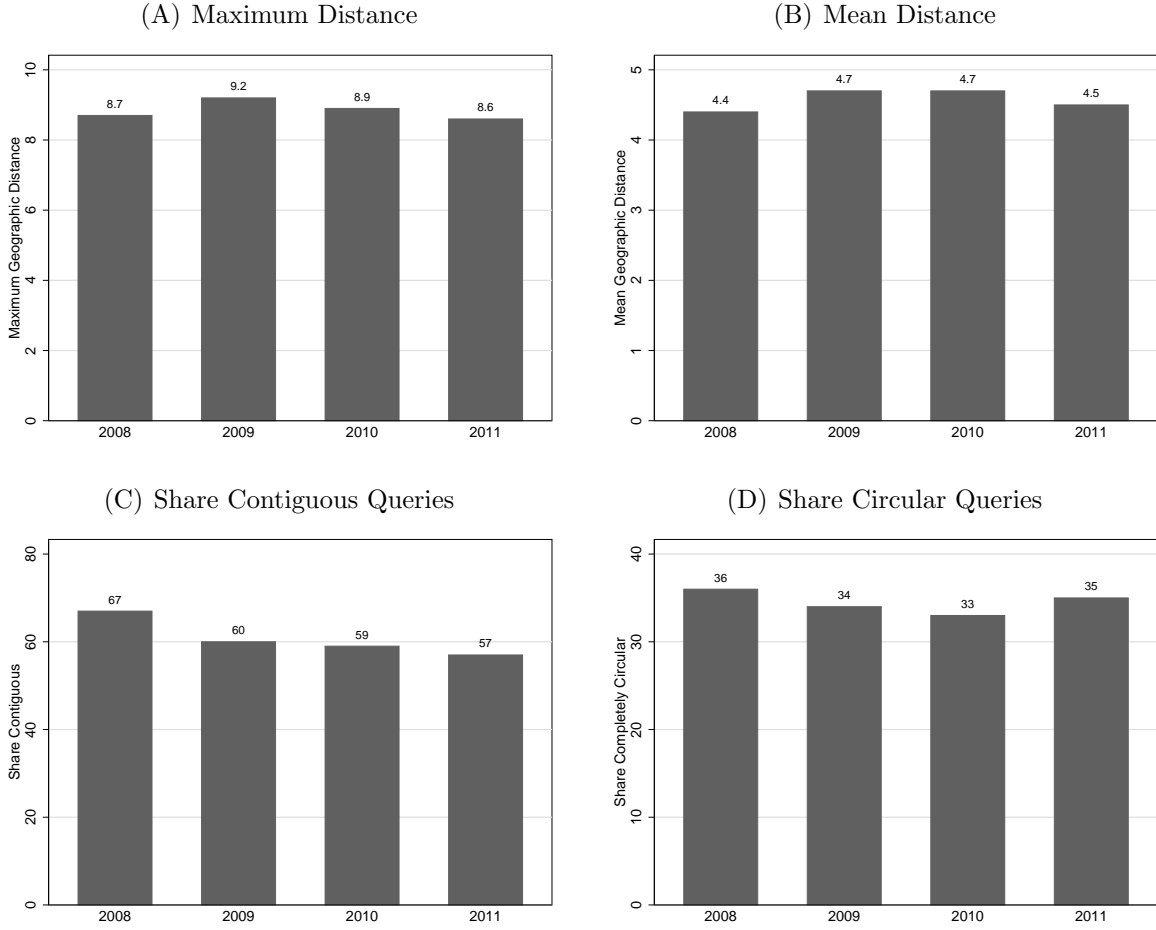
Note: Figure shows average values of search parameters by the year when the email alert was set. We report the maximum distance between zip code centroids (Panel A), the mean distance between zip code centroids (Panel B), the share of contiguous queries (Panel C) and the share of circular queries (Panel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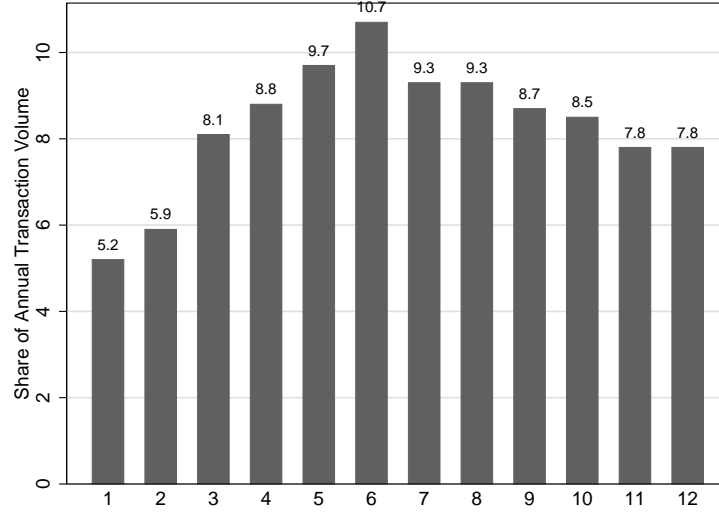
(B) Maximum Distance

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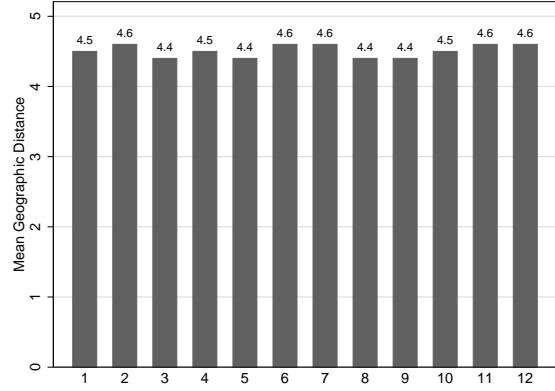
(D) Share Contiguous Queries
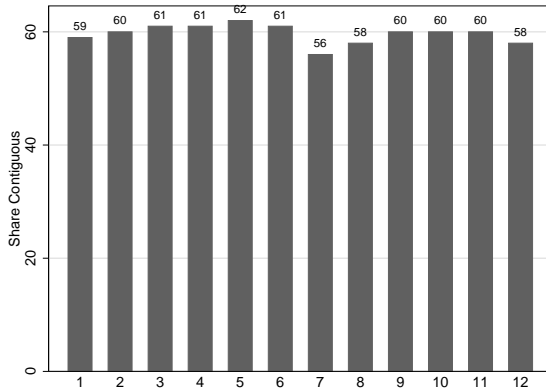
(E) Share Circular Queries
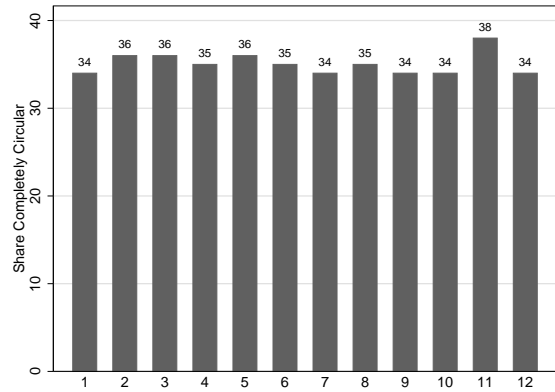
Note: Panel A shows the share of total annual transaction volume in each month. Panels B - E show average values of search parameters by month of search query. We report the maximum distance between zip code centroids (Panel B), the mean distance between zip code centroids (Panel C), the share of contiguous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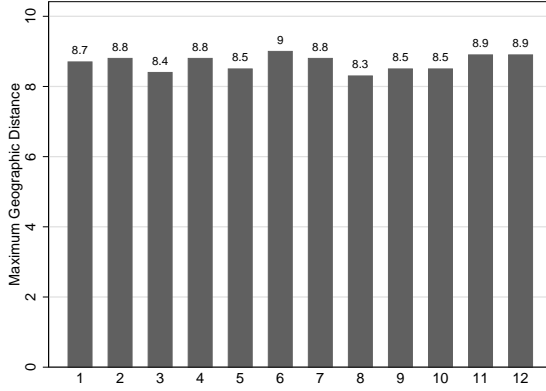ies (Panel D) and the share of circular queries (Panel E). Months increase from January to December on the horizontal axis.

Figure C.3: Stability of Segment Moments

(A) SF House Prices   (B) Segment Prices

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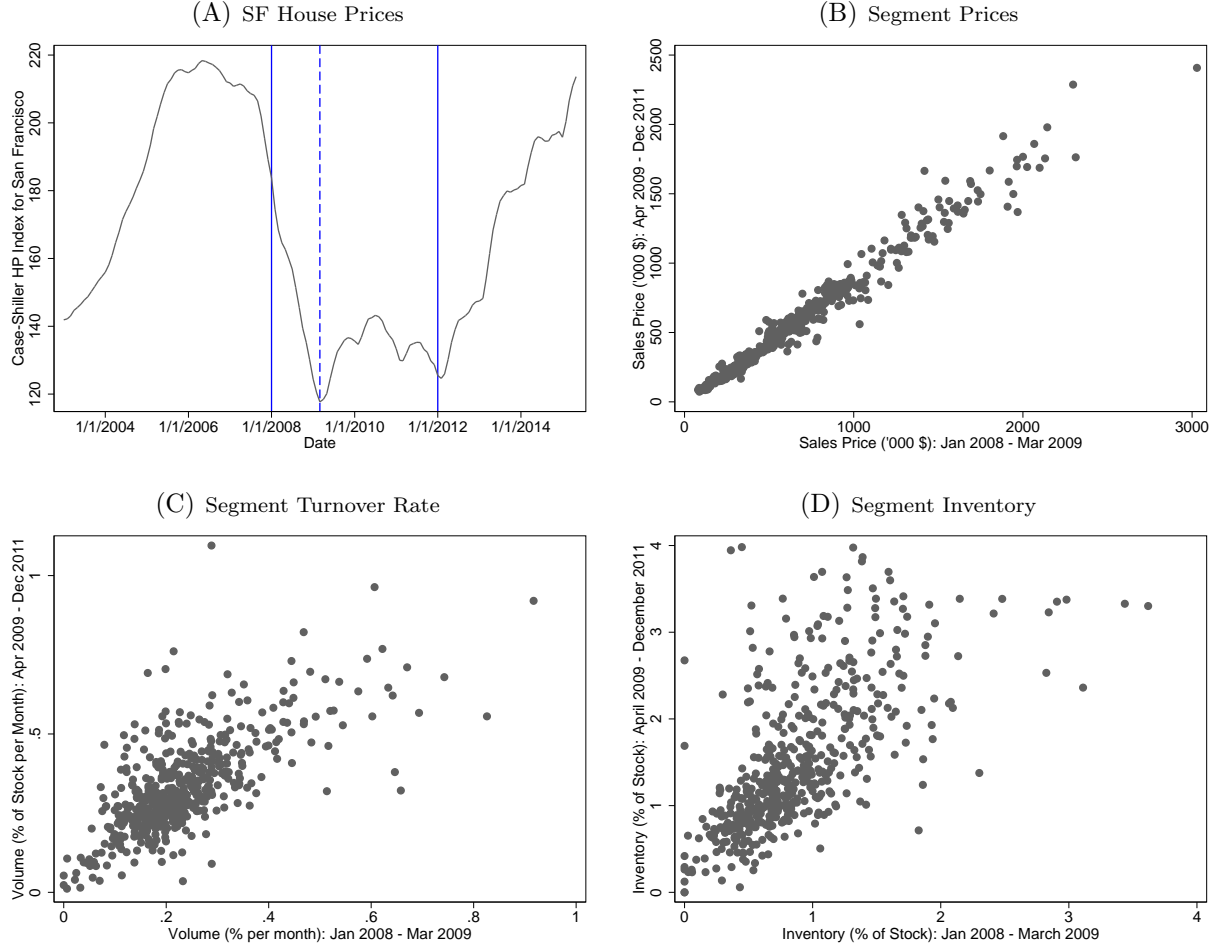
Note: Panel A shows the Case-Shiller House Price Index for San Francisco. Our sample period is delineated with vertical solid lines. It is split into a period of declining house prices and stable house prices by a dashed vertical line. Panels B, C, and D show segment-level scatter plots of key housing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

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

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 $p\left(j, k\right)$ the price at which a type-j seller sells to a type-k buyer. We further write $E_B^j\left[p\left(k, j\right)\right]$ and $E_S^j\left[p\left(j, k\right)\right]$ 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

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

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 ($p\left(j, k\right)$ 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 III, 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

$$U_E^B = \frac{v}{r} \frac{\theta m\left(B, L\right)/B}{\left(r + \eta + \left(1 - \theta\right) m\left(B, L\right)/L\right)},$$ (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 IV. 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

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

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

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

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 $m(1, q^{-1})$ goes to infinity for $q \to 0$, which ensures existence of a unique solution for q. Inventory follows as $L = H\eta/(\eta + m(q, 1))$, the total number of searchers is $B = qH\eta/(\eta + m(q, 1))$ and the number of narrow searchers is $B^N = qH\eta/((q + 1)\eta + m(q, 1))$. 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

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

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.[30]

---

[30]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

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

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

$$-\frac{r m_2 (B, L)}{m (B, L)^2} + \frac{1}{(H - L)^2} - \frac{1 - \theta}{L^2} < 0.$$

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 $\tilde{p}$.[31]

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

---

[31]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 IV 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

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

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 IV, 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 IV, 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:

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

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.