Search Frictions and the Design of Online Marketplaces

Andrey Fradkin*
MIT and Airbnb, Inc.
afradkin@mit.edu

http://andreyfradkin.com/assets/SearchFrictions.pdf

September 30, 2015

Abstract

Online marketplaces increasingly act as intermediaries in labor, housing, dating, and other markets where traders match with each other. These marketplaces use novel data generated by users' activities on the website to design algorithms and products that influence the search and matching process. I use internal data from Airbnb, a prominent online marketplace for housing rentals, to study the efficiency of this market and the effects of ranking algorithms. I first show that potential guests engage in limited search, are frequently rejected by hosts, and match at lower rates as a result. I then estimate a micro-founded model of search and matching and use it to show that if frictions were removed, there would be 102% more matches in the marketplace. I propose and evaluate several ranking algorithms and show that a personalized algorithm would increase the matching rate by up to 10% over the status quo. However, due to equilibrium effects, the A/B search experiments favored by internet companies can overstate the true treatment effect of an algorithm by over 100% in some cases.

*Note, this paper is currently being split into two separate parts. I thank my advisors Jon Levin, Liran Einav, Caroline Hoxby and Luigi Pistaferri for guidance and inspiration. I am grateful to Airbnb (especially the analytics and data infrastructure teams) for providing me with the data and for helping me execute every aspect of this project. I thank John Horton, Tim Bresnahan, Ali Yurukolgu, Ramesh Johari, Al Roth, Doug Bernheim, Przemek Jeziorski, Lanier Benkard, Frederic Panier, Dan Knoepfle, Itay Saporta-Eksten, Scott Baker, Pete Troyan, Adam Vogel, Dan Preston and numerous seminar participants for comments. Financial support was provided by the SIEPR Graduate Fellowship, the B.F. Haley and E.S. Shaw Fellowship for Economics, the NET Institute, and Airbnb. The paper represents the views of the author and not Airbnb, Inc.

1 Introduction

Online marketplaces have transformed how consumers search for jobs, apartments, spouses, and consumer products. These marketplaces design products such as search ranking algorithms, reputation systems, and user interfaces. However, little is known about the potential of these policies to make marketplaces more efficient and the proper way to measure the effects of these policies. I use detailed data on searches and transactions from Airbnb, a prominent online marketplace for housing and accommodations, to model the efficiency and the effects of marketplace policy. I propose three mechanisms that cause inefficiencies in these markets: that consumers cannot consider all options available, that consumers don't know which sellers are willing to transact (due to seller screening, "stale vacancies," or congestion), and that some transactions may occur inefficiently early. I build an empirical model that allows me to measure the impact of these frictions and use it to study platform changes aimed at improving matching efficiency.

I find that frictions still play an important role on the internet and that marketplace policy can reduce but not eliminate the negative consequences of those frictions. Searchers on Airbnb typically view only a subset of potential matches in the market and more than 40% of listings remain vacant for some dates. Furthermore, sellers reject proposals to transact by searchers 49% of the time, causing searchers to leave the market although there are potentially good matches remaining.[1] Frictions in the market lead to less transaction volume, consumer surplus and revenue. If searchers had information about all the options in the market and knew which sellers were willing to transact with them, there would be 102% more matches and revenue per searcher would be $117 higher.

I propose and evaluate several new ranking algorithms aimed at improving market efficiency. The preferred algorithm increases the match rate by 10% in counterfactual simulations. However, because there is test-control interference and because the partial equilibrium effect of a policy is not equivalent to the full equilibrium effect, the treatment effects of an A/B test are generally not consistent for the true effect of a policy change. I also find that ranking algorithms affect the relative contribution of searchers and sellers to the total number of matches in the market. This is especially important because, although the search technology has changed in matching markets, most models of matching do not allow for the shape of the aggregate matching function to change over time. Although I focus on ranking algorithm design in this paper, my modeling framework can be used for other difficult problems such as optimal marketing strategies, price recommendations and platform fees.[2]

[1] Horton [14] documents a similar negative effect of rejections on subsequent search effort in the context of ODesk, an online labor market platform.

[2] For example, one difficult issue for Airbnb has been to ascertain the marginal value of a host versus guest

The setting of this paper is Airbnb ("Airbed and Breakfast"), a growing online marketplace for housing rentals that has served over 9 million guests between 2008 and late 2013. Airbnb provides a platform for individuals to temporarily rent out an unused room or property to travelers.[3] Search on Airbnb begins when potential guests submit their trip details, which consist of a city, dates of stay, and number of guests, into a search engine. The search engine returns a set of listings in that city. If a listed option is appealing, the searcher sends an inquiry to the host of that listing.[4] An inquiry typically asks about the availability of the listing and sometimes asks about other listing details. The host's response is indicative of a complex decision to let a stranger stay at her place. This decision combines issues of trust, availability, hassle, and the possibility of better offers. If the host accepts, then the searcher can choose to book the place.

Airbnb is an excellent environment for studying search and matching frictions for several reasons. First, as in the labor and dating markets, both sides of the market have heterogeneous preferences towards the other side. Second, because hosts can only accommodate one trip for a set of dates, there is a potential for large congestion frictions. Burdett et al. [5] show that markets where sellers have limited capacity (such as Airbnb, the dating market, and the housing market) are characterized by different matching functions than settings where sellers have a large capacity such as the college admissions market. Lastly, the Airbnb dataset contains search, communication, and match behavior, whereas many datasets used for studying search and matching markets lack search and communication data. An especially important piece of information that I observe is whether the buyer or seller rejected the proposed transaction. Analogous data in the labor market would contain resume submissions, interview invitations, and interview outcomes at a searcher-vacancy level.

I estimate models of consideration set formation, directed search, and rejection and combine them into a simulation of market outcomes. The consideration set model determines how many pages are seen by each searcher, which filters are applied on each page and what listings are ultimately seen. This model is estimated using unique data from actual browsing and filtering behavior of searchers on Airbnb. The directed search model determines which listings from the consideration set are contacted by searchers. It is estimated using data on inquiries sent from guests to hosts. In that model, the searcher's utility for a particular

in a given market. This answer is impossible to answer well in a framework without cross-user externalities and other network effects.

[3]Similar marketplaces to Airbnb exist in industries such as contract labor (ODesk, Freelancer and Taskrabbit), dating (OkCupid, Match.com and eHarmony), craft goods (Etsy and Ebay) and personal transportation (Uber and Lyft).

[4]Airbnb has been expanding its "Instant Book" feature, which allows searchers to book a listing without communication.

listing is a function of listing characteristics such as location, size, and price and the match quality between guest and host. For example, searchers for trips with more guests prefer bigger properties and searchers who use a price filter are more price sensitive.

Upon receiving an inquiry, hosts choose whether to reject or accept a potential guest. Rejection by hosts happens in 49% of all inquiries. There are three distinct causes of rejection in search and matching markets: screening, congestion, and "stale" vacancies. Screening occurs when a host rejects a searcher because of the searcher's personal or trip characteristics. For example, some hosts may not be comfortable sharing a space with an individual who has no reviews. To capture the host's decision process, I estimate a logistic regression of hosts' screening decisions as a function of guest and host characteristics. Congestion occurs when a searcher sends an inquiry to a listing that is about to transact with another searcher. Rejections due to congestion arise endogenously in my model because transactions take time to clear. Lastly, rejections due to stale vacancies occur when a listing which is not actually available for a given week is nonetheless visible in search and contacted.[5] In my dataset, over 20% of communications are screened, over 21% are sent to stale vacancies and less than 6% are affected by congestion.[6] This result is interesting because much of the theoretical literature on directed search such as Kircher [17], Albrecht et al. [2] and Burdett et al. [5] has focused on congestion as the main cause of rejection in search markets.[7]

I combine the above models into a simulation of how the market clears over time. In the simulation, searchers enter Airbnb looking for a place to stay in a given week and market, and conduct search according to the directed search model. Inquiries sent by the searchers are either accepted or rejected by hosts according to the reasons above. If a guest chooses to book a property, the transaction takes time to clear. Additional searchers enter the market sequentially and send inquiries. The actions of searchers and hosts then generate the aggregate matching and rejection probabilities. The simulation comes within several percentage points of matching the empirical booking rates, contact rates, and rejection frequencies even though these moments were not used to determine the model parameters. This model is similar to the model of online dating in Hitsch et al. [13], which includes directed search,

[5] Stale vacancies are similar to congestion from the perspective of the searcher but are generated by a different seller behavior. Therefore, policies that reduce rejections due to congestion do not necessarily reduce rejections due to stale vacancies and vice versa.

[6] There is no good data, to my knowledge, on these rejection reasons for other markets. Analogous rejection reasons in the labor market are as follows. Screening rejections occur when a job applicant is not qualified for the job. Congestion rejections occur when a suitable applicant applies to a vacancy that is about to be filled by another applicant. Stale vacancies occur when a vacancy is listed although the employer is not hiring.

[7] However, see Chade et al. [7] for an example where college applicants do not know which college will accept them.

heterogeneous preferences, and screening. However, unlike in the setup in that paper, my setup allows me to quantify frictions[8] and to study the effect of the marketplace search engine on outcomes.

I then use the model to evaluate the effects of counterfactual search ranking algorithms on Airbnb. The baseline model I estimate uses a proxy for the ranking algorithm that was running on Airbnb during the time of the estimation sample. In the counterfactual policy experiments, I instead use one of three alternative ranking algorithms: one that shows listings that are more relevant to searchers on average, one that shows listings tailored to each searcher's preferences, and one that shows listings which maximize each searcher's transaction probabilities. I show that all policies improve outcomes and that the later two algorithms improve matching probabilities by 10%. I also simulate the results of the A/B tests favored by internet platforms to determine which policies to pursue. I show that the treatment effects estimated from A/B tests are typically different from the market level treatment effects. This occurs for two reasons, first the treatment interferes with the control, and second the partial equilibrium treatment effect is not equal to the the full equilibrium treatment effect. These forces combine to overstate the true effects of a policy by over 100% in some cases. To my knowledge, this is the first paper to quantitatively model the effect of ranking algorithms in a search and matching market.[9]

The policy space for online marketplaces encompasses search ranking, site layout, new matching mechanisms and explicit rules that affect the behavior of agents. Most marketplaces rely on experimentation to learn about the effect of a particular policy. However, experimentation must be used intelligently and in combination with other methods to learn about the effects of policies in search and matching markets. Simulations like the one in the paper can be used to quantify the potential bias of experiments, to find better experimental designs, and to look for symptoms of test-control interference in the data. Second, because the parameter space of possible marketplace policies is too large to explore with experimentation alone, simulations can be used to choose among policies to be tested in the future. Lastly, the effect of policies depends on time-varying market conditions such as the number of agents in the market and the ratio of searchers to sellers. An experiment, on the other hand, is typically run in a specific time period and set of markets. Understanding how market conditions affect outcomes can inform experimental design and can help the marketplace designer adapt policies to different situations.

[8]In Hitsch et al. [13], 68.3% of men in the sample never match with a partner. However, the paper does not examine the extent to which those rejections are inefficient.

[9]Ghose and Yang [11] and Santos and Koulayev [26] study the effect of ranking algorithms in generic hotel search engines where seller preferences are unimportant. Horton [14], Casadesus-Masanell and Halaburda [6] and Hagiu and Jullien [12] propose theoretical models where platforms have incentives to alter searchers'

The approach taken in this paper differs from the aggregate matching function approach used in many papers concerning search markets (Petrongolo and Pissarides [23]). Aggregate matching functions stipulate that the number of potential matches given a set of searchers and vacancies is determined by a matching technology which is not, in general, based on micro-foundations. The lack of micro-foundations makes aggregate functions unsuitable for studying policies, technologies, and changes in market structure which affect the matching process. For example, Lagos [20] derives a micro-founded aggregate matching function for the taxicab market and shows that its shape will be sensitive to policy. Similarly, policies in my model change the relative contribution of searchers and sellers in the matching function. Specifically, a better ranking algorithm increases the weight on hosts and decreases the weight on searchers in the matching function. This occurs because each searcher gets better search results and is more likely to match holding the quality of hosts constant.

The paper is organized as follows. Section 2 gives more detail about Airbnb. Section 3 describes the model of directed search and section 4 describes the determinants of rejection. Lastly, section 5 describes the model and shows the empirical results.

2 Data and Setting

2.1 Airbnb Background

Airbnb describes itself as a trusted community marketplace for people to list, discover, and book unique accommodations around the world − online or from a mobile phone.[10] It is a fast-growing startup that was founded in 2008 and that has more than doubled the number of guests accommodated in every year of its operation. In 2012, Airbnb accommodated over 3 million guests and listed over 180 thousand new listings. It has listings in more than 34 thousand cities worldwide and is available in more than 30 languages.

Airbnb has created a market for a previously rare transaction: the rental of an apartment or part of an apartment in a city for a short term stay by a stranger.[11] These transactions were not occurring previously because there were large costs to securely exchanging money, communicating with strangers and evaluating a stranger's trustworthiness. Airbnb was one of the first platforms to provide a set of tools and services which enabled guests and hosts to arrange stays in a relatively risk-free manner. These tools are important because hosts on

behavior through ranking algorithms.

[10] https://www.Airbnb.com/home/press

[11] Couchsurfing, a large travel based social network started in 2003, facilitates similar stays but without monetary exchange. Craigslist has listed sublets and short-term vacation rentals since the late 1990's. Vacation rentals by owners in tourist towns have also existed for a long time.

Airbnb are typically non-professionals, over 80% of whom list just a single room or property.[12]

In a typical Airbnb transaction, guests and hosts have the ability to inconvenience each other by being disruptive or by lying about the characteristics of a room or trip. Communication occurs on the platform both to screen counter-parties and to confirm the details of the trip such as the key exchange and check-in times. Importantly, hosts are allowed to deny potential guests for any reason. Guests and hosts have the opportunity to review each other after the trip. In subsection 3.2, I show that the stock of reviews on the site provides important information and affects both host and guest behavior.[13] Users of Airbnb can disclose information about themselves through photographs and textual descriptions. Airbnb also provides free professional photography services to hosts and verifies users' identity using online social networks (Facebook, LinkedIn, and Twitter) and passports or driver's licenses.[14] In the empirical section I show that this information is important in the decisions of agents on Airbnb.[15]

The payments mechanism on Airbnb is a big innovation compared to previous platforms for accommodations such as Craigslist, where payments were handled offline. The marketplace holds payments in escrow until the stay has happened in order to prevent fraud. The escrow service allows Airbnb to resolve disputes between guests and hosts and to enforce the payment of security deposits and cancellation penalties. Airbnb also offers a one million dollar insurance guarantee against major damage done to properties by guests. Airbnb generates revenue by taking a percentage fee of every transaction that takes place on the platform. Hosts typically have a 3% fee while guests have a variable fee that ranges between 6% and 12% depending on the details of the transaction.

Another important service that Airbnb provides is the management of pricing and property calendars. For each date, each listing has a binary availability that is displayed on the site. If a booking occurs, a host's calendar is automatically blocked off for the dates of the trip. Alternatively, the host can manually update the calendar to be unavailable for a set of dates. In either case, the property can no longer be booked and will not show up in search for the blocked off dates. Hosts might not update their calendars even if they are unavailable because they are uncertain about future availability, want to see potential guests or face updating costs. As a result, potential guests send inquiries to seemingly available listings which are actually off the market.

[12]In fact, there are now third party services like Guesty and Airenvy that manage an Airbnb host's listing.

[12]There is a literature that documents the importance of reviews in a variety of online platforms. For example, Pallais [22] shows that reviews are valuable and under-provided on Odesk, a labor market platform.

[13]For more details see: http://blog.airbnb.com/introducing-airbnb-verified-id/

[14]Lewis [21] shows that information voluntarily disclosed by sellers on Ebay Motors affects market prices.

2.2 The Search and Matching Process

Below is a list of steps each searcher undergoes before a match occurs. Even though the examples below are from Airbnb, each of these steps occurs in other search and matching marketplaces.

1. Using the Search Engine (Figure 1) - Searchers enter the travel dates, number of guests and location into a search engine and receive a list of results. The search can then be refined using filters and maps.

2. Investigation (Figure 2 and Figure 3) - The searcher clicks on a listing in search. The subsequent page displays additional photos, amenities, reviews, responsiveness, house rules and information about the host.

3. Communication (Figure 4) - The guest sends messages to hosts inquiring about room details and availability. A host can respond by saying that the room is unavailable, that the room is available or by asking a follow up question. The host might also not respond at all. A guest may send multiple inquiries both initially and after receiving responses.

4. Booking - If a transaction is agreed upon, the guest can click the "Book It" button. If the host accepts, the money is charged and taken in escrow by the platform. Some transactions occur without a messaging step if the host allows the "Instant Book" feature.

5. Stay - After a trip is booked, there is further communication needed to exchange keys and coordinate the details of the trip. Either side can cancel with a pre-specified penalty rate.

2.3 Data Description

My dataset contains the full set of inquiries and transactions that occurred in major US markets on Airbnb. Appropriate subsamples of the data are used throughout the paper. In this section, I narrow my data set to comprise all contacts to US markets which occurred for trips starting in 2012, while in the model sections I focus on one market and a narrower set of dates. For each communication I observe the time a message was sent, who sent the message, who received the message, and the content of the message. Each set of messages between a guest and a host is aggregated into a "thread" that contains information about trip dates, the number of guests, and the room of inquiry.

I group each set of communications in my dataset by search spell (alternatively referred to as a trip attempt). A search spell is defined as a set of inquiries and transactions pertaining to a stay by a specific guest in a specific market and in a specific week of the start of the trip. For each searcher, I observe the number of reviews, the number of pictures, demographics and other characteristics. For each property and date I observe the listed price, number of reviews, review score, location, number of pictures, and other characteristics.

There are some limitations to the dataset. I do not always observe the correct price in cases when a host changes prices for a specific date over time. The reason for this is that the data on date specific pricing is from monthly snapshots rather than from daily snapshots. This is not a major problem because most hosts do not set date specific prices and instead use a calendar wide listed price. Further, I do not observe the entire path of availability in the market. For example, a property might initially be available for a given week, then become unavailable and then become available again. In that case, I treat that property as always being available. Lastly, some of the observed characteristics such as age and gender are either missing or entered inaccurately by users.

2.4 Identifying Host Rejections

In this section I describe how to classify whether a host's response to a guest indicates that the host is interested in the transaction. For example, the response "Sorry, the place is unavailable." should be classified as a rejection whereas the response "Yes it is available. Go ahead and book it." should be classified as a non-rejection.

I use a combination of approaches to determine whether a response is a rejection. If an inquiry led to a booking or was labeled as accepted by the host then it is classified as "not rejected". If a response was labeled by the host as a rejection or if there was no response within a week then the response is classified as a rejection.

If the response does not fit into any of the above categories, a text classifier is applied to the first (or in some cases second) response by a host to a guest. I use a common technique in natural language processing called a regularized logistic regression to classify text (see Appendix A for details). I combine the text classification with cases when the response is discernible through other methods.[16]

In total, 49% of all inquiries were rejected. Of all responses classified as rejections, 37% were non-responses, 30% were classified by the host, and 32% were classified the regularized logistic regression. As a consequence, communication on Airbnb frequently fails to result in a transaction. For US Markets in 2012, just 15% of inquiries and 48% of search spells transact.

[16] Foreign language responses are removed from the analysis.

3 Search Behavior

3.1 Consideration Set Formation

Searchers in matching markets are often faced with hundreds, if not thousands, of choices. People cannot possibly consider all of these options while making a decision. The purpose of a model of consideration set formation is to specify which of the many options the searcher considers before making a choice. These models are especially important in the case of online marketplaces because search engines and recommendation algorithms affect outcomes through their effect on consideration set formation. However, consideration sets are not solely determined by the features of the marketplace. Searchers vary in the amount of pages the view and the filters they apply while searching. In this section, I describe a way to use data on search activity to model consideration set formation.

Potential guests on Airbnb search for a place to stay by entering a location, a set of dates and the number of guests into a search engine. For each query, the website returns a maximum of 21 available listings according to a ranking algorithm. Searchers can then continue to the next page of results for a given query or they can modify the search parameters by using filters or a map (seen on the left side of Figure 1). These types of refinements are frequently used on sites such as Airbnb, where the stock of listings is heterogeneous. Figure 5 displays filtering frequencies per search and searcher. Over 80% of searches and searchers use at least one filter or map action. The most commonly used browsing actions (in order of frequency) were room type filter, map, maximum price filter, neighborhood filter and the next page of results. I do not model filters that are used less than 5% of the time such as those pertaining to bedrooms, beds, and bathrooms.

Listings are ordered for each query according to their algorithmically determined scores, with highest scores first. During the sample period, each score was a function of listing (but not searcher) characteristics and the distance of the listing to the query. In a typical session, searchers do not see all listings because there are hundreds or thousands of available listings in a city.[17] Although it is theoretically possible to obtain the exact ranking algorithm from the code-base, it is infeasible in practice. The reason is that the search results are often the result of a complex set of nonlinear transformations of underlying variables. Furthermore, the variables used in the algorithm might not be logged on a high frequency basis.

It is, however, possible to estimate a proxy for the true ranking algorithm by using data on what the actual search results. The dataset for estimating the algorithm is composed of the set of listings that could have been displayed in search for a random sample non-filtered

[17] Searches in October of 2013 for stays in New York, Seattle, San Diego, Austin, Los Angeles, Chicago, Miami and many other cities resulted in more than 1000 search results.

searches in the market of study. To be in the set, listings must be available for the set of dates of the search and must be able to accommodate the number of guests selected in the search. For each such listing, h, I observe a set of observables and whether it was displayed in the first search without filters for a given searcher at time t.

I use this dataset to estimate the following equation:

$$h_{seen,hts} = 1(X'_{ht}\mu + \gamma_s + \epsilon_{hs} > 0) \tag{1}$$

where "seen" indicates whether the listing was displayed in the first search, X_{ht} are listing characteristics, γ_s is a searcher specific fixed effect and ϵ_{hs} is a logit error term. The ranking score of each listing, w_{ht}, in the simulation of market outcomes is then set to the predicted index, $X'_{ht}\hat{\mu}$, resulting from the above regression.

I then estimate a model of browsing and filtering behavior. In the model, as in the data, the first page that an individual browses displays the 21 best visible listings in the market according to the ranking scores, w_{ht}. Following the first search, the searcher draws idiosyncratic preferences for neighborhoods, prices, and room types. The frequency with which each preference is drawn is equal to the frequency of that filter being used in the data, conditional on at least one filter of that type being used. For example, searchers who use at least one room type filter, pick the "entire room" filter 83% of the time. In this example, a searcher will draw a preference for "entire room" with an 83% probability.

Following each page view, the searcher can go to the next page of results, can apply a set of filters to a new search, or can stop searching. Denote these browsing actions, A_i. I model the probability of each action occurring as a multinomial logit with the index for A_i equal to:

$$\beta_{0i} + \beta_{1i}N \tag{2}$$

N, is the chronological order of the search action for a given searcher.[18] The index for the action of doing a new search is normalized to 1. For each filter type, the probability that it is applied upon action "New Filter" is determined as a function of N. Tables 1 and 2 display the coefficients which determine the probability that each action and filter is used as a function of N. The coefficients on N are positive in the filtering models, confirming that filters are more likely to be used further along in the search process. In the simulation, the filter that is applied, i.e. room type entire or private, is determined by the randomly drawn

[18]In the summer of 2013, Airbnb launched a new interface with a map that takes up half of the screen. This paper studies the period before that feature was launched. Sorting by price was not possible during this period.

individual preferences. The full consideration set is determined by the union of the listings seen in the browsing session determined by the above process.

I abstract away from two important aspects that are present in actual consideration set formation: the effects of ranking within a page[19] and endogenous browsing behavior. I assume that the ranking of a listing within a page does not affect the consideration set formation process. The reason I make this assumption is that searchers often see the same listing at different rank positions depending on the filters applied. Therefore, I assume that searchers are at least aware of the listings that are displayed on each page. Any preference for something that is shown first as opposed to tenth on the page is captured by the error term in the utility function. Second, I assume that the consideration set formation process does not change with the ranking algorithm. This is a reasonable approximation if searchers believe that the ranking algorithm always gives the most relevant results according to Airbnb, regardless of what the actual algorithm is. However, to the extent that better ranking algorithms encourage marginal searchers to increase the amount of pages they browse, I am likely to be understating the positive effects of better ranking algorithms in my simulations.

Importantly, I am not estimating a model of the optimal choice to use a filter or browse a set of options (as in Chen and Yao [8]). Such a model would be able to identify search costs, which are not the focus of this paper. However, it would be complex and the parameters in such a model would be identified off of assumptions about the expectations of searchers. The rational expectations assumption typically used to identify search costs will not hold on Airbnb because most searchers are not experienced with city specific housing markets and have no way to know how the supply of listings on Airbnb would change if they wait. Koulayev [18] uses a rational expectations approach to estimate a model of search engine usage for a large hotel search engine. He estimates a median search cost of $10 per page of results. Given that searchers in my sample view a median of 7 pages, that estimate would imply a median incurred search costs of $70, with a significant left tail of individuals incurring hundreds of dollars of costs. These costs are unreasonably high given that the cost of booking a room is often less than $100 per night. One promising approach for future research is to model the process of learning during search as in Koulayev [19] and Santos et al. [27].

3.2 Directed Search and Preferences

The guest's choice of which property to contact from a given consideration set is determined by a random utility discrete choice model. The property characteristics visible in search

[19] There is a growing literature that examines the effect of the position of an item on advertising clicks, movie viewing decisions, and other search behavior (i.e. Ghose and Yang [11] and Jeziorski and Segal [16]).

results are price, location, number of reviews, property type and any common social connections between the guest and host. Other characteristics, which are visible upon further inspection, include additional pictures of the property, a free form textual description of the property, the text of listing reviews, a description of the host, average response time, average response rate, frequency of calendar update, security deposit, bedrooms, bathrooms and cancellation policies. The searcher evaluates these characteristics and chooses one (or more) properties to contact.

Let guest, g, enter the market at time, t. Each guest chooses a property to request to book from their consideration set. The guest receives utility from property, h, according to a linear combination of property characteristics, a property random effect, interactions with idiosyncratic preferences, and a guest specific error term according to the equation below:

$$u_{ght} = \alpha_0 + (p_{ght} + f_{ght})(FP_g'\alpha_1 + Z_g'\alpha_2) + f(X_{ht}, Z_g)'\beta_1 + \kappa_N + NF_{gh} + RF_{gh} + \gamma_h + \epsilon_{ght} \quad (3)$$

where X_{ht} is a vector of property characteristics including review quality, property type and whether the host is a property manager. Z_g is a vector of trip and guest characteristics (Nights, Age, Guests), FP_g is a set of categorical variables for ranges of the maximum price filter used by the searcher, $f(X_{ht}, Z_g)$ is a set of interactions between guest and host characteristics, p_{ht} is the nightly price of the property for the trip, f_{ght} is the platform fee, κ_N is a neighborhood fixed effect, NF_{gh} is an indicator variable for whether a listing's neighborhood was specified by a searcher's filter (or map action), RF_{gh} is an indicator variable for whether a listing's room type was specified by a searcher's filter and η_{ght} is an unobserved component of the utility which is distributed according to the type 1 Extreme Value (EV) distribution with variance 1. γ_h is a normally distributed listing level random effect. The random effect is included to account for unobserved heterogeneity at the listing level. Furthermore, this demand model allows for heterogeneity in searcher preferences about location, price and the size of the listing. Importantly, preference heterogeneity is modeled using interactions of listing characteristics with searcher characteristics and filters used.

The searcher can also choose to take the outside option and leave the online marketplace. The searcher's value of the outside option is determined by the following equation:

$$u_{got} = T_g'\mu + HP_t * FP_g + \alpha H_g + \epsilon_{got} \quad (4)$$

where T_g are guest and trip characteristics, HP_t is the average hotel price for the city on the date of check-in, FP_g is a set of categorical variables that measure the maximum price filter used by the searcher, H_g is the number of searches by the guest and ϵ_{got} is a type 1 EV error term.

The probability of making a choice between a particular property and the outside option is determined by the relative utilities of the two options, search costs, the continuation value of searching and the probability of booking a particular inquiry. Suppose that a searcher can only send an inquiry to one listing. The probability that a searcher sends an inquiry to property, h, as opposed to choosing the outside option is determined by the following equation:

$$Pr(u_{go} < b_{gh}u_{gh} + (1 - b_{gh})u_{go} - c) = Pr(u_{go} + \frac{c}{b_{gh}} < u_{gh}) \quad (5)$$

where b_{gh} is the perceived probability that searcher, g, books property, h, and c is the cost of sending an inquiry.[20] Now suppose that there are many identical options and that the searcher does not discount time. Then perceived booking probability in the searcher's optimal decision would equal the rate with which a searcher books any listing on Airbnb.

In either case, the consumer surplus in the above model depends on the value of the outside option and the magnitude of the effective search cost, $\frac{c}{b_{gh}}$. The effective search cost is difficult to identify without making strong assumptions because it is unclear how searchers' expectations of booking probabilities are formed. The reason is that searchers on Airbnb typically have little experience with using the site and rejection rates are not publicly disclosed. In robustness checks, I was unable to find a significant effect of the estimated listing specific booking probability on choices.[21]

The demand estimation procedure requires assumptions about the empirical analogue to the consideration set of each searcher. Consider the case when a searcher sees a set of listings in the first browsing session and immediately sends an inquiry. The consideration set for that scenario is composed of all the listings seen in that browsing session. In other cases a guest sends an inquiry after many days of search for a particular trip. Some properties that the guest browsed could have been booked by someone else before the decision to send an inquiry was made. Lastly, guests sometimes send an inquiry to a host without seeing a property in the search results. This might occur if the guest navigated to the property through an outside link.

I include a property in the consideration set used for estimation of the demand model if the guest saw the property and sent an inquiry to any property up to 2 days afterward. If a guest did not send an inquiry, then I include a property if it was seen in the last 2 days

[20] The cost of sending an inquiry should be thought of as the cost of writing a short message to the host or entering credit card information. This cost is distinct from the "search costs" incurred while browsing, which are the focus of Koulayev [18]

[21] In some other matching markets, such as college admissions, there is public information on option specific admissions rates which may cause searchers to behave differently.

of search. Guests who sent an inquiry more than two days after the last search, who sent an inquiry to a property that was not observed in search or who viewed fewer than 21 total listings are excluded. I make the above assumptions so that the estimation data contains only decisions that mirror the underlying model of behavior.

The variation in consideration sets provides the identification in the model. Due to dynamic nature of the market, the amount and quality of listings available in the market varies both by the week of the trip and the week of the search. Therefore, searchers who enter the market at a given time before a trip might see very different listings depending on the week of the trip.[22] This variation allows me to identify how the value of the outside option for searchers varies both by the week of the trip and the time in advance of the trip that the search occurred.

No matter how many covariates are added to the model, there will still be important aspects of the property characteristics that are difficult to observe by the econometrician but not by the guest. For example, the property may have stylish furniture in the picture. Such a listing is likely to charge a higher price than an otherwise similar property with worse furniture. I evaluate the robustness of my results to this bias in a forthcoming appendix which varies the magnitude of the coefficient on price and reruns the subsequent analysis in the paper.

Another complication with the above specification is that random utility models overstate the benefits to variety because the characteristic space expands with every product. In practice, products crowd each other out to some extent in the characteristic space. My results are robust to adding a correction term (as in Ackerberg and Rysman [1]) for the size of the consideration set.[23]

3.2.1 Demand Estimation

The estimation sample consists of 8,977 users with Airbnb accounts searching in City X between March 30, 2013 to June 25, 2013 for trips between April 2013 through July 2013. These users collectively viewed a total of 569 thousand listings while searching.[24] Searchers who only had 1 search were also excluded from the sample to reduce noise. Less than 1% of those searchers actually contact a host and these searchers are typically viewed as "non-serious" by analysts within the company. I also exclude those searchers who saw fewer than

[22] In some cases, such as the Super Bowl, searchers might strategically time entry. However, the market I study does not experience such major events.

[23] Such a term corrects for the tendency of the utility from a random utility choice to diverge as the set of options grows. The correction suffers from bias if those individuals who view larger consideration sets conditional on observed characteristics also value the outside option differently than those who view smaller consideration sets.

21 or more than 500 properties because such observations are either incomplete or likely driven by bots (as determined by Airbnb's analysts).

The estimates from the demand model for listing and outside option characteristics are displayed in Table 4 and Table 5 respectively. In both tables, Column 1 is the specification without random effects and Column 2 is the specification with random effects. Each coefficient estimate and standard error is normalized by the coefficient on price for those individuals whose highest maximum price filter during search was between \$10 and \$100.

Across specifications, guests value higher rated listings, better locations, and listings that fulfill filter criteria. Furthermore, there is large searcher heterogeneity in the price sensitivity depending on the searchers' use of price filters. Those who set no maximum price filter are 30% less sensitive to price than those that filter for prices between \$10 and \$100 per night. Trips with more guests are less price sensitive and value entire properties as opposed to private rooms. Older guests are less price sensitive, presumably because they are richer.

The value of the outside option (inclusive of effective search costs) compared to the value of a listing is important for counterfactuals because policy changes affect the set of listings that people see. If the listings shown are better, then the searcher should be less likely to choose the outside option. One very important factor determining the value of the outside option is the price of a hotel for the dates of the trip. The estimation results show that all types of searchers are less likely to choose the outside option when hotel prices are high. Furthermore, those searchers that use more stringent price filters are more sensitive to hotel prices. Lastly, those individuals that browse the most pages are least likely to pick the outside option. This term captures the fact that those individuals for whom the outside option is the worst, have the most to gain from searching intensively on Airbnb.

The standard deviation of listing random effects, which account for the listing level heterogeneity conditional on observables is \$8. The random effect is on the same order as having a maximum rating of 5 stars versus a mediocre rating of 4.5 stars. Therefore, there is still significant heterogeneity between listings conditional on observables. The standard deviation of the utility error term ranges between \$17 and \$19 per night. The unexplained component of utility is important because it often takes on values that are much larger than the values of observable listing characteristics. The error term is driven by the fact that characteristics which are observable to searcher (photo quality, amenities, etc...) are not included in the specification. Further, the error term can be inflated by unobserved preference heterogeneity and the fact the searchers may not scroll all the way through any

[24] To be included in the sample, searchers had to have fewer than 9 guests, fewer than 15 nights of stay and must have searched before the day of the check-in. I also exclude views of "Shared" rooms which comprise $< 1\%$ of all inquiries.

given set of 21 results on the search page.

3.3 Simultaneous and Sequential Search

Searchers sometimes send multiple inquiries, either initially or in sequence after being rejected. The propensity of searchers to engage in this type of search is a function of their characteristics and the amount of listings they browse that are better than the outside option. To capture this behavior, I estimate Poisson count of models of the number of inquiries sent initially and after rejection (shown in Table 5). For simultaneous search, the number of inquiries initially sent by a searcher minus 1 is regressed on trip and searcher characteristics. For sequential search, the number of subsequent inquiries is regressed on trip an searcher characteristics. The results for simultaneous search are shown in column (1) while the results for sequential search are shown in column (2). Experienced guests send more initial inquiries and more inquiries after a rejection. These results suggest that experienced guests have learned about their rejection probabilities and also value Airbnb more than new users. Furthermore, reviewed guests send fewer inquiries conditional on their experience. Lastly, those that send the most inquiries initially are more likely to send more inquiries after initial rejections. I describe how I use these estimates to simulate market outcomes in section 5.

4 Rejection and Screening

Hosts receiving inquiries from potential guests choose which of them to accept or reject. If an inquiry is suitable, the host responds with a non-rejection and waits for the guest to book. If the inquiry is not suitable, the host rejects and waits for the next inquiry. There are three reasons why hosts reject guests in my model: congestion, “stale” vacancies and screening.[25] I describe each one below.

Congestion occurs when a guest sends an inquiry to a host who is about to transact with someone else. The reason that congestion happens is that transactions take time to clear. The size of the time between inquiry and booking is determined by the time it takes for a host to respond and the time it takes for a guest to enter credit card information and confirm the transaction. Further, there is sometimes a longer conversation that occurs between guest and host about the details of the trip. I count a rejection due to congestion as one in which an inquiry is sent to a host that is subsequently booked as a results of a previous inquiry. Not all congested inquiries receive an immediate rejection. Instead, the host may tell the guest to wait until there is a response from the previous inquiry. Congestion occurs for 5.6% of inquiries in the US for 2012. Importantly, I assume that hosts evaluate each inquiry sequentially rather than waiting to receive several inquiries and picking the best. In practice, there are some cases when a host receives inquiries in parallel. For example, a host might receive several inquiries at a time if he checks Airbnb infrequently. I abstract from this scenario because many hosts are notified by text or email of an inquiry and have an incentive to respond quickly.

Another type of rejection occurs due to stale vacancies, when listings are not available to anyone for a given set of dates. There are two ways in which stale listings manifest themselves in the data. First, hosts can update their calendars to state that they are unavailable for a set of dates. For 56% of listing-weeks in the sample, hosts update their calendars to be unavailable and are not booked by anyone for that week. For these stale vacancies to matter, hosts must update their calendars after some searchers have already sent inquiries to the listing. This happens for 21% of all inquiries in the data. Some hosts with stale vacancies never update their calendar. This happens because updating the calendar is costly, because the host wants to learn information about demand, or because the host forgets. In such cases, hosts rejects all inquiries for a particular set of dates either explicitly or by not responding at all. Figure 7 shows the distribution of rejection rates (excluding congested inquiries and those for which a host updated the calendar) by hosts in a given week of check-in. Only listings that received at least 5 inquiries for the week are included. There is a wide dispersion of rejection rates and a noticeable excess mass at 1, with over 10% of host-week combinations rejecting all inquiries. Nothing in the data determines whether a particular host rejected all inquiries due to actual unavailability or due to high selectivity. However, if hosts reject each inquiry at a rate of 50%, then there should be fewer than 3% of host-week combinations that reject all inquiries. Given that that actual rate of these observations in the data is 10%, at least some of these cases must be due to stale vacancies. In total, the upper bound on inquiries rejected due to stale vacancies is 32% and the lower bound is 21%.[26]These rejections are important to model because they are so frequent on Airbnb and because they are likely to be prevalent in other matching marketplaces. However, I am not aware of any other papers that model this phenomenon.

Hosts have preferences over trips and guests, and those preferences are not displayed on the website. This leads hosts to receive inquiries from unsuitable guests and to subsequently reject those inquiries. 31% of all inquiries are potentially rejected due to this screening. Importantly, there are systematic differences in the rejection rates between different listings.

[25] Appendix C contains a static urn and ball model which demonstrates how rejections arise according to congestion, stale vacancies and screening.

[26] I've also estimated an explicit model that allows me to identify the share of inquiries rejected by screening versus stale vacancies. That model determines that 5.1% of inquiries are affected by stale vacancies for which hosts never updated their calendar.

Figure 8 displays a histogram of the mean rejection rates (excluding congested inquiries) by property for all of 2012. I exclude congested inquiries, inquiries to stale vacancies, and listings with fewer than 10 inquiries. The figure shows that many hosts reject all or almost all inquiries for a given listing while others accept almost everyone. The observed heterogeneity in rejection rates can occur because some properties are more selective, more in demand, receive different types of inquiries, or are less likely to update their calendar than others. I model these considerations in the next section.

Stale listings and screening account for most of the rejections that occur on Airbnb. These two frictions are of approximately the same magnitude and their relative importance depends on the details of market clearing. On the other hand, congestion, which has been a key focus in many theoretical models of directed search, only occurs for 5.6% of inquiries in this sample.

In total, 59% of inquiries are affected by frictions either through screening, congestion or a stale vacancy. Do these rejections matter for booking probabilities and market efficiency? The effect of a rejection depends on several factors, including the amount of listings that are substitutes to the one that rejected the guest, the cost of searching, and the beliefs of the searcher. I test whether a rejection is associated with a lower overall probability of booking by regressing the whether a guest booked a room on whether the first inquiry a guest sent was rejected. Table 3 displays the results of this regression. A first rejection is associated with a decrease in overall booking probabilities by 50%.[27] This effect is robust to week, market, and trip type characteristics. I quantify the efficiency costs of rejections in subsection 5.3.

The above framework for classifying rejections abstracts from several potentially important considerations. First, those that update their calendar to be unavailable might have accepted a particularly attractive offer. Alternatively, a calendar might have been updated because the listing was booked off of the Airbnb platform, either through another platform or in an informal transaction. This type of behavior may be important for the counterfactuals (i.e. Athey et al. [3]) and is important to understand. On Airbnb, hosts can indicate on their calendar that they were booked on another site. In total, 1.81% of inquires are sent to these multi-homing hosts. There is no good evidence on how frequently transactions initiated on Airbnb are taken off of the platform. However, this behavior is unlikely to be frequent because of the insurance, reputation and secure monetary transfer that using Airbnb offers. Another potential cause of error in the above framework is that the text classification process is not fully accurate and could have mislabeled some responses.

[27] Horton [14] finds similar sized effects for rejections on ODesk

4.1 Screening Model

Screening rejections occur on Airbnb because hosts have preferences over when and whom they host.[28] For example, a host might reject a guest because the guest is not reviewed, has a vague inquiry, or does not have enough information in his profile. Hosts also reject guests because the check-in dates of the inquiry can break up a bigger, uninterrupted time of availability for the host, preventing future inquiries. Lastly, hosts may be waiting for a better guest/trip combination or might consider a particular inquiry too much of a hassle.[29] In this section, I model the decision to reject as a function of guest, trip, and listing characteristics.

The estimating equation for the screening model is:

$$Pr(R_{gh}) = Pr(\alpha_0 + Z_h'\delta + f(X_g, Z_h)'\beta + \gamma_h + \eta_{gh} > 0) \tag{6}$$

where η_{gh} is the logit error term, R_{gh} is an indicator for whether the response is a rejection, X_g are the number of guests, guest reviews, guest gender, weekly demand, days in advance of the trip nights, guest age, searcher census tract demographics and month of check-in. Z_h are property type, property manager indicator, host age, the number of reviews and price. $f(X_g, Z_h)$ are interactions between guest and listing characteristics. $\gamma_h \sim N(0, \sigma_{sh}^2)$ represent listing specific random effects. The listing random effects account for heterogeneity in hosts' baseline propensity to reject. I account for the dynamic aspects of the host decision by controlling for the time in advance of the trip of inquiry and for the overall demand for each week of check-in. Nonetheless, the above model is incomplete because although it captures the decisions of hosts, it cannot be interpreted as a utility. In order to infer the utility of hosts, I would need to estimate a full structural model of host decision making with dynamic decisions driven by host expectations of future demand.

The dataset for estimation consists of all non-foreign language inquiries sent by guests in City X between January 2013 and July 2013. Each inquiry is then classified as a rejection according the procedure outlined in subsection 2.4. I exclude inquiries in which congestion occurred and in which the host updated her calendar for the inquiry dates after the inquiry to indicate unavailability. The final dataset consists of 18,145 observations, of which 44% were rejected.

Table 6, panels a - c, displays the results of specifications with and without listing spe-

[28]Bargaining plays a key role in other matching models with transferable utility between the two sides of the market. Bargaining is not prevalent on Airbnb. One reason might be that guests who ask for a discount seem less trustworthy and are perceived to be more of a hassle. The potential cost to a host from a non-trustworthy guest is perceived to be much greater than the potential earnings from the stay.

[29]These reasons are frequently mentioned by hosts in internal Airbnb surveys and are corroborated by the screening model estimates.

cific random effects. Panel (a) displays the coefficients on trip and guest characteristics of inquiries. Reviewed guests are less likely to be rejected, presumably because hosts find them more trustworthy. Trips with more nights are less likely to be rejected as long as they are not too long. Market conditions also matter for rejection behavior. In the specification with listing specific random effects, the standard deviation of the propensity of hosts to reject is large compared to the estimated coefficients. This confirms that some hosts are more selective than others, even conditional on observables.

Panel (b) displays the coefficients on listing characteristics in the regression. Hosts that tend to respond quickly or allow instant booking are less likely to reject. On the other hand, younger hosts, female hosts, and hosts with entire properties are more likely to reject guests. Property managers (hosts with more than 4 active listings) are less likely to reject because they care less about interacting with guests and operate more like hotels. Panel (c) displays the coefficients on interactions between guest and host characteristics. The interaction between property manager and reviewed guest is positive because property managers care less about who the guest is. Furthermore, property managers are less likely to reject last minute inquiries than non-property managers because it is easier for property managers get a listing ready for a stay on short notice. These heterogeneous preferences are important because they might generate mismatch in the search and matching process. If there is mismatch, then marketplace policy that directs searchers towards non-rejecting listings might generate large welfare gains.

5 Simulation of Equilibrium Outcomes

In this section I describe how to combine the search and screening models into a model of market equilibrium. The goal of the model is to generate market level matching outcomes from micro-foundations. My approach is similar to Roth and Xing [25], which studies congestion in the clinical psychologist market with simulation. I improve on that paper's approach by estimating preferences from the data, by modeling multiple frictions, and by comparing model outcomes to empirical moments. An alternative way to combine the model and the data would be to estimate all of the parameters jointly rather than to estimate first stage models and simulate the market outcomes. I chose the later approach because most of the calibrated parameters in the simulation have close empirical analogues in the data, because the full model has over 100 parameters, and because simulation is faster than estimation.

5.1 Simulation Setup

For a given check-in week in City X, I observe searchers, bookers, and listings. I collapse the seven days of the week into one time period and assume that each listing can only be booked once per week.[30] Every searcher who enters the market looking to stay in City X for the week of the simulation draws a consideration set, receives a mean utility from every listing in the market, and draws a random utility shock distributed according to the demand estimates from column (2) of Table 4. Similarly, each listing receives a mean probability and a random error term conditional on searcher and listing characteristics from the logit model of rejection in column (1) of Table 6. In the simulation, listings update their calendar to be unavailable at the same time that they do in the data.

The simulation requires several additional parameters related to market clearing and the intensity of sequential and simultaneous inquiries. All of the calibrated parameters are seen in Table 6. The time to transact is determined according to an exponential distribution whose mean is the empirical mean of time to transact in the data. I assume that the time between browsing and sending an inquiry is negligible. There are two parameters in the model that do not have direct analogues in the data: μ_{sim}, and μ_{seq}, which determine the extent of simultaneous (initial) and sequential search.

The amount of initial inquiries sent by a searcher is determined as follows. Each searcher draws a Poisson random variable, n_{sim} with mean equal to the product of the predicted mean from the Poisson regression in subsection 3.3 and a calibrated constant, $\mu_{sim} > 1$. The searcher then sends a number of inquiries equal to the minimum of $n_{sim} + 1$ and the number of viewed listings with utility greater than the outside option (including the costs of sending an inquiry). Sequential search is treated in the same manner. A searcher who is rejected draws a Poisson random variable, n_{seq}, according to the expected mean from the Poisson regression multiplied by $\mu_{seq} > 1$. A draw of 0 implies that the searcher does not continue search. If the draw is greater than 0, then the searcher comes back to the market 1 day later and sees the entire choice set. The searcher then sends an inquiry to the minimum of n_{seq} and the number of properties remaining which are better than the outside option. I set values of μ_{sim}, and μ_{seq} that match the rates of simultaneous and sequential inquiries in the data.

A transaction does not occur even if there is no immediate rejection 17% of the time. This event can happen because the guest does not want to book the place, the host ends up

[30] In the data, 22% of hosts who receive an at least one inquiry in a week book more than 1 trip in that week. I avoid modeling this behavior because the model fits the data and accounting for multiple bookings adds complexity to the model. If a guest is looking for a stay of more than 7 days, then only the 7 days are used for the calculation of surplus and revenue in this section. Otherwise trips that cross into other weeks would affect outcomes for the simulation week.

rejecting the guest later in the conversation, or because the initial classification was wrong. In the simulation, I assume guests leave the market with a constant probability after a non-rejection.

The calibrated parameters described above are sufficient to simulate market outcomes. However, several additional assumptions need to be made in order to calculate consumer surplus. Searchers incur costs while browsing the website and sending inquires. I calibrate these search costs by using data on the time spent searching and combining it with the shadow value of time for searchers. I assume that each inquiry takes 5 minutes to compose and that it takes .725 minutes to browse a page. The time spent browsing per page comes from calculating the median ratio of the time spent browsing and the number of pages browsed.

To calculate the value of time I assume that Airbnb users earn twice as much as the median annual income for males aged 25 - 44 in the United States.[31] If searchers work two thousand hours a year, then their shadow value of time will be $3.24 per 5 minutes. I use this search cost for the rest of the simulation exercises. Of course, there is likely to be heterogeneity in search costs, both due to differences in the shadow value of time and due to the fact that many people enjoy looking for rooms on Airbnb. However, I leave the difficult task of identifying the distribution of these search costs for future work. Lastly, I assume that searchers expect their booking probability to be the average probability of any inquiry resulting in a booking. Alternative assumptions on the expectation make little difference for consumer surplus because the costs of sending an inquiry are low relative to the benefit of booking a room.[32]

5.2 Baseline Results

The final choice situation includes 960 searchers and 1159 visible listings (56 days in advance of the check-in dates). Table 7 row (1) displays the outcomes that occurred in the data for the choice situation. 62% of searchers sent an inquiry and 37% eventually booked a room. Row (2) displays the results of the baseline simulation. The model outcomes match the data

[31] Source: U.S. Census Bureau, Current Population Survey, 2013 Annual Social and Economic Supplement.

[32] I assume that individuals don't choose whom they send inquiries to depending on their idiosyncratic probability of rejection. Most searchers do not have enough experience with the platform to know their rejection probabilities. Furthermore, there is little information on the platform about the relative likelihood that a particular host is more selective than another host.

[32] Note, these percentages are sensitive to the measured number of searchers and listings in the data. However, this measurement, especially for searchers, depends on the definition of a searcher. In this paper, I count anyone that does a search with dates and has an associated user id as a searcher. However, to the extent that some other, non-included searchers may be influenced by the policy, the exact results will change. I test how the relative ratio of searchers and hosts affects outcomes in subsection 5.6.

well considering that most of the moments in the table were not explicitly targeted by the calibration. The most significant differences between the data and simulation is that the simulation overstates the revenue of hosts by $23.[33]

5.3 Which Frictions Matter?

There are three mechanisms (excluding incurred search costs) due to which actual market outcomes differ from the solution to the social planner's problem: rejection, limited choice sets, and dynamic mis-allocation. Table 7 displays outcomes when rejection and limited choice sets are individually and jointly removed.[34] In row (7) all listings that would reject each searcher are removed from that searcher's choice set. Compared to the baseline simulation, removing rejecting hosts decreases inquiries by searchers because on average, worse quality listings are shown. However, the inquiries that are sent are more likely to result in bookings. Without rejections, the share of searchers booking increases by 38%, the average revenue increases by $46 per searcher and the average consumer surplus increases by $15 per searcher. This is a large and surprising impact given the limited supply of good listings available in the marketplace. The improvement in matching indicates that there are suitable substitutes for rejecting properties on the platform, but that those properties are not being contacted at a high enough rate.

In rows (3) - (5) of 7, I display the equilibrium effect of each rejection cause separately. Of all rejection frictions, screening has the largest effect on booking and consumer surplus. The effect of screening in equilibrium is larger than it's frequency in the data because screening properties are more likely to be included in searchers' consideration sets and are more desirable for searchers. Further, listings that screened a particular guest might have accepted another one, whereas stale listings and congested listings could not have matched with another searcher. The importance of screening suggests that Airbnb should guide guests towards hosts that are willing to transact. In fact, after the sample period in this paper, Airbnb has started explicitly personalizing search results to minimize the probability that an inquiry is rejected. Secondly, Airbnb should elicit host preferences about guests ahead of time to improve its ability to guide the guest's search.

Row (6) of Table 7 displays the simulated matching outcomes if searchers freely consider

[33]The share of searchers that books is likely to be understated in the simulation because, in the data, some listings can actually be booked two or more times in a given week. For example, there can be one two night stay and another three night stay for a given listing.

[34]In the counterfactual exercises, the searcher's expected probability of transacting given an inquiry should change. However, in the results below they do not. I will implement the correction in the future but doubt that such a correction would make much of a difference. Even if the expected booking probability per inquiry increases to 50%, the added benefit would be less than 10$

all listed properties rather than just a limited consideration set. The share of searchers that book increases by 54% and the revenue per searcher increases by $59. The effect of limited consideration sets is bigger than the cumulative effect of all rejection reasons because there are large product specific error terms. Row (8) displays the results of the simulation if the rejection friction and the partially observed choice sets were removed together. There is a 95% increase in booking probabilities from having both frictions eliminated at once, which is approximately equal to the sum of the gains from removing rejections and partial consideration sets separately.

The counterfactual with full consideration sets abstracts away from the fact that considering hundreds of properties takes significant time and effort. However, the time spent to view and consider all listings would be significant. The mean time spent in search by a searcher who sends an inquiry is 18 minutes and even those searchers do not typically consider all visible listings in the market. In order to compute the costs of browsing for these counterfactual scenarios, I assume that it takes 90 minutes to consider all available properties. Furthermore, as in the other counterfactuals, guests value each 5 minutes of time at $3.24. Therefore, the average costs of browsing all listings would be $58 per searcher. This is actually very close to the surplus gain per searcher of having full consideration sets. Column (5) of Table 7 shows the average consumer surplus including the browsing costs for each outcome. The increase in consumer surplus from seeing the full consideration set is only $5 although the increase in booking probabilities is 20 percentage points. Therefore, although the potential gains to finding the best match are high, the costs of considering all options make those gains small. However, if the platform can predict what the best match for a searcher is without the searcher engaging in costly search, the gains remain high.

The last column shows outcomes in the consumer maximizing allocation in the marketplace. Suppose that a social planner can dispatch each searcher to an appropriate host (or outside option) using information on random utility draws, characteristics and the timing of entry for every searcher and listing. The consumer surplus maximizing allocation of the 960 searchers to the 1159 properties is a solution of the assignment problem with constraints. For each assignment of a searcher to a listing, the searcher must prefer the listing to the outside option and the listing must accept the searcher. Each listing can only be assigned to one searcher and each searcher can only be assigned to one listing. As in the simulation, there is a mass of unavailable listings that are not matched and a share of searchers who do not book even if they are matched. Further, I assume that the mean number of pages browsed is 5 and that the time spent browsing is 90 minutes, as in the previous section. In the optimal scenario, the booking rate doubles, and increases by 2.4 percentage points compared to the scenario with no rejection and full consideration sets, the consumer surplus per browser increases by $41 and the hosts' revenue increases by $117 compared to the baseline. There are two reasons why the optimal allocation differs from the one in which searchers see all options except for those that would reject them. First, fewer inquiries need to be incurred for matches to occur. Second, there is a dynamic mis-allocation of listings to guests. Searchers entering the market earlier sometimes book a listing that would have provided more utility to a later entrant. This exercise shows that dynamic mis-allocation is a small factor in determining the volume of transactions compared to full consideration sets and rejection in the market.[35] However, there is potentially a large consumer surplus depending on assumptions about the browsing costs used in this scenario.

5.4 Marketplace Policy

I have shown that search frictions significantly affect consumer surplus and revenue on Airbnb. Because even small changes in conversion rates can improve profit margins, the platform should actively be aiming to reduce frictions. One way in which the platform can improve outcomes is by changing the order in which items are displayed in the marketplace. In my model, as in Dinerstein et al. [10], this corresponds to changing the ranking algorithm used for consideration set formation. [36]

In this section I consider three ranking algorithms whose goal is to improve market efficiency by showing more relevant listings to searchers. I simulate the effects of better algorithms by deriving a counterfactual listing specific weight, w_h, and using it to determine the ranking of listings in search. Let $\bar{\mu}_{gh}$ equal the deterministic part of searcher utility:

$$\bar{\mu}_{gh} = \alpha_0 + (p_{ght} + f_{ght})(1(FP \in [10, 100)) * \alpha_{1,10} + Z_g'\alpha_2) + f(X_{ht}, Z_g)'\beta_1 + \kappa_N + \gamma_h + \epsilon_{ght} \quad (7)$$

Where the price elasticity is set assuming that searcher g is the type that would set the price filter between $10 - $100.

Three weights for the algorithms are as follows:

1. $w_h = \sum_h \bar{\mu}_{gh}$ (Average Quality)

2. $w_{gh} = \bar{\mu}_{gh}$ (Personalized Quality)

3. $w_{gh} = F^{-1}(\bar{\mu}_{gh} - \bar{\mu}_{go}) * (1 - Pr(R_{gh})$ (Transaction Probability)

[35] Gavazza (2013) shows that dynamic mis-allocation is an important friction in the secondary market for airplanes.

[36] Much of the research on search and recommendation engines focuses on search for content or non-capacity constrained goods (i.e. Amazon Books). However, when the supply of a given good is capped, one buyer of a good prevents other searchers from buying that good. The preferences of searchers entering the market over time and the consideration sets they draw determine whether the good is allocated in an optimal manner.

The first ranking algorithm is calculated by averaging $\bar{\mu_{gh}}$ across searchers for each listing. This type of algorithm would be appropriate if the marketplace cannot generate personalized search results. The second algorithm generates a personalized ranking based on the mean utility of a searcher for a specific listing. The third algorithm is also personalized but shows listings that searchers are more likely to transact with. In weight (3), F^{-1} is the inverse logistic distribution, $\bar{\mu}_{go}$ is the mean value of the outside option and $Pr(R_{gh})$ is the probability that host h rejects guest g.

Table 7, rows (9) - (11) show the results of these policies. The non-personalized search algorithm increases contact rates by 13% and booking rates by 2.5% compared to the baseline and the personalized search algorithms increase the share of searchers contacting by 18% and the booking rates by 10%. All algorithms increase consumer surplus but, as expected, the personalized algorithm results in a larger surplus gain than the non-personalized algorithm. The non-personalized algorithm actually decreases revenue by steering searchers to cheaper properties. However, the personalized algorithms increase revenue by approximately $7 per searcher, a significant amount for the platform.

The demand models I use to generate the ranking algorithms do not explain a large share of searcher choice. In practice, market designers can choose how much employee time and resources to devote to generating features from data and better prediction algorithms that use those features. The increase in bookings in the counterfactual where searchers see the full choice set suggests that there are high returns to devoting resources to building better demand models and algorithms. There are few marginal costs to deploying a better algorithm while the benefit of a better algorithm is a flow that is a function of the size of the platform.

5.5 Experimentation

Internet companies typically run user level experiments to learn about the effects of policies. For example, a company may test a new feature by allowing 50% of users to see it and comparing average outcomes between the treatment and control group. These tests typically yield useful information about the effect of a policy on behavior. However, in the marketplace setting, the treatment typically affects the control. For example, if there is a finite number of sellers in the market, a treatment which improves matching, may reduce the set of options for the control. In other cases, if ranking algorithm results are correlated across searchers, adversely selected properties stay in the the search results for more searchers in the full treatment as apposed to the A/B treatment. In general, the effect of partial treatment roll-out and a full treatment roll-out in the marketplace are different.[37]

The size of the bias due to this test-control interference depends on market conditions. This is precisely the reason why structural models of market outcomes are useful for designing marketplaces policy. In this section, I simulate the difference between estimated treatment effects from an A/B test and the overall market level effect of a treatment. To do this, I assign 50% of the searchers in the simulation to an alternate ranking algorithm and compute market outcomes.

Table 8 displays three types of treatment effects for two ranking algorithms. First consider columns 1 - 3, which show outcomes for the non-personalized quality algorithm. Column (1) shows the measured treatment effects from an A/B test. Here the share of searchers booking increases by 5.5% in the treatment versus the control. Column (2) shows the difference in outcomes if everyone in the market were to be given the treatment. The overall effect on booking is an increase of 2.5%. Therefore, for this week and market, the A/B test overstates the true treatment effect of the policy by 120%. Lastly, Column (3) shows the market level effect of launching the A/B test. The increase in booking probability due to the half rollout is 6.3%, which is higher than both the A/B treatment effect and the true treatment effect. This seeming anomaly is due to the fact that the full rollout increases the share of inquiries that are rejected much more than the half roll-out does. Furthermore, the half rollout improves the outcomes of the control group because they are less likely to be rejected due to congestion.

Columns 4 - 6 display the same outcomes for the personalized quality algorithm. Here, the A/B results still overstate the true market level effects of the algorithm but the bias is much smaller (16%) for two reasons. First, the increase in rejections due to the personalized policy is only 13% while the the increase in rejections due to the non-personalized ranking algorithm is 18%. Second, because individuals in the non-personalized ranking algorithm get less relevant results, they send fewer inquiries and therefore each rejection lowers booking probabilities by a greater amount than in the personalized algorithm. As the ratio of searchers to listings increases, however, the difference in the A/B treatment effect and the market level effect increases even for the personalized algorithm. For example, if the number of searchers increases by 50%, the A/B test would overstate the market level effect of a policy by 34%.

An alternative to simulation for estimating these treatment effects is to do market level experiments. However, market level experiments might not be feasible when market definitions are ambiguous, when there are few markets, or when markets are not comparable.

[37] For a similar effect in the labor market see Crepon et al. [9] on the difference between the partial and general equilibrium effects of a job counseling program in France. Another example is Blake and Coey [4] for email advertising on Ebay.

Another option for measuring the equilibrium treatment effect is to combine models like the one in this paper with experimental results. A model of market equilibrium should have parameters that are influenced by an experiment. With experimental variation, I could estimate how an alternate ranking algorithm changes consideration set formation. The alternate consideration set model can then be used to simulate market outcomes. Importantly, such an approach does not require an exact replication of a complicated ranking algorithm within the model. What matters for outcomes is how the algorithm influences consideration set formation or other relevant features of the model. The equilibrium effects of other types of experimental interventions which change the rules of the market can be modeled in the same way.

5.6 Aggregate Matching Functions, Market Tightness and Policy

The setting of the prior simulations is one market and one week. However, market conditions vary on Airbnb both across markets and over time. In this section, I vary the ratio of searchers to hosts and the overall amount of agents in the market to study how market conditions affect matching and policy. By varying market conditions, I can obtain unbiased estimates of an aggregate matching function and to test for increasing returns to scale. This technique can also be used to study how aggregate matching functions change with marketplace policy.

I generate data to estimate a matching function by randomly sub-sampling or re-sampling agents in the market and simulating outcomes.[38] Each market condition is simulated 60 times, with the amount of listings and guests varying between 50% and 200% of the amount seen in the data.

The simulated data is then used to estimate a Cobb-Douglass matching function of the form:

$$\log(M_s) = \log(A) + \alpha \log(G_s) + \beta \log(H_s) + \epsilon_s \tag{8}$$

where M is the number of matches, G is the number of searchers and H is the number of listings in the market. Table 9 displays the estimated results of this equation, where the coefficients are allowed to vary by the search algorithm used.

The coefficient on searchers, α, is .81 while that on hosts, β, is .14. Therefore, the amount of searchers is a bigger contributor to the overall number of bookings than the amount of hosts. Unbiased matching function estimates for consider optimal advertising policies. Estimates of matching functions using cross-market and within-market data yield similar results, with $\alpha = .869$ and $\beta = .225$. However, estimates based on observational

[38] I break ties in the timing of entry of searchers by adding a random exponential noise component with a mean of an hour to each realized entry time.

data suffer from the fact that market level tightness may be a function of the market specific matching shocks. For example, in high-demand weeks, searchers may be more willing to book worse listings due to the fact that hotel prices are higher. This source of endogeneity would likely bias the estimates of α upward.

There are two forces that can cause matching on Airbnb to either exhibit increasing or decreasing returns to scale. Firstly, a market with more options should generate better matches on average because searchers can draw larger consideration sets and because those consideration sets will include better options on average. On the other hand, if consideration sets are highly correlated then thicker markets may exhibit higher levels of congestion and higher costs of adverse selection. The estimated returns to scale in the the matching function, $\alpha + \beta$, equals .95. Therefore, returns to matching in this market are slightly decreasing in scale. However, this does not imply that there are decreasing returns to scale to being a marketplace. For example, as marketplaces grow, agents collect reputation which results in better matches in the future.

Furthermore, ranking algorithms change the shape of the aggregate matching function. I show this by simulating outcomes in the market with the personalized quality algorithm at different levels of tightness. The interacted coefficients in Table 9 correspond to the matching function estimates when search rank is personalized. The improved ranking algorithm increases the share of matches due to the supply side in the market and decreases the share due to the demand side. The change in the shape of the matching function occurs because, for a given quantity and quality of listings, searchers are shown better options and are more likely to convert. However, as listings are booked, later searchers have worse options and are less likely to convert. Therefore, the quantity of listings becomes more important in determining the number of matches in the market. The above exercise demonstrates that the Cobb-Douglass matching function parameters cannot be interpreted as structural because they are changed by policy.

Some aspects of agent behavior can change with market tightness and market size but are not in my model. For example, searchers might form bigger consideration sets if the are more listings to match with, which would result in greater returns to scale if congestion remains a small issue. On the other side of the market, hosts may change pricing and rejection strategies with market conditions, although the directional impact of these adjustments on matching is unclear a priori. Lastly, reputation capital might actually be the biggest cause of returns to scale in marketplaces (i.e. Pallais [22]). The matching function estimates above do not model the effects of market conditions on the accumulation of reputation capital. For example, the size and tightness of a market may determine how difficult it is for new listings to accumulate reputation.

6 Conclusion

The rate at which heterogeneous agents successfully transact and the surplus generated by those transactions is a function of the information structure of a market. I use novel data on search and communication behavior to build a micro-founded model of matching. The underlying cause of the matching frictions in the model is that guests and hosts have heterogeneous preferences and must use communication and Airbnb's site to obtain the necessary information to transact. Airbnb provides a review system, a communication platform and many other tools to reduce the cost of searching. Even with these tools, many searches fail and frictions combine to reduce transactions on the site.

Airbnb loses profit because frictions reduce the volume of transactions and overall revenue. Airbnb's policy makers can use the data they observe about the history of both searcher and seller behavior to improve outcomes. In this paper, I simulate policies which use such data to improve the relevance of consideration sets and to reduce the chance that rejections occur. I find that better search ranking algorithms can improve transaction probabilities by 10%. These gains demonstrate that marketplace designers can generate large gains in volume, revenue and consumer surplus through better policy.

As the marketplace designer's knowledge about buyer and seller preferences approaches the full information benchmark, outcomes approach their frictionless benchmark. The ongoing reduction in the costs of storing and analyzing data, commonly referred to as the "Big Data" revolution, will likely have a profound impact on platforms like Airbnb because more and better data can improve the platform's estimates of agent preferences. Although I focus on ranking algorithm design in this paper, my modeling framework can be used for other unsolved problems in marketplace design such as optimal marketing strategies, price recommendations and platform fees. For example, in order to evaluate marketing strategies, platforms need to know the rate at which searchers and listings acquired by marketing channels cannibalize transactions from current market participants and the extent to which they extend the total volume of trade in the market. The market simulations in the previous section explicitly model these effects.

The efficiency loss due to search frictions in other search and matching markets will vary depending on the expectations of searchers, the potential benefits of searching, the ranking algorithm, the market tightness, and the details of consideration set formation. Furthermore, the industry structure of a given market also plays a role in determining the size of frictions. For example, the cost of the rejection friction for the average searcher in the accommodations market is likely to be low because hotels typically clear the market based on price and do not explicitly reject potential guests. Because of competition from hotels, Airbnb has a strong incentive to reduce the rejection friction. In the labor and housing markets, however, rejection is common to most options. Therefore, the average costs from the rejection friction are likely to be higher than in the accommodations market. Furthermore, although individual platforms in these markets still have an incentive to reduce the rejection friction, that incentive may be diminished because searchers know that rejection is likely and adjust behavior accordingly. One example of such a market is college admissions, where rejection rates are widely known by applicants.[39]

In this paper, I only model the short-run responses of agents to policy changes. In the long-run, however, policies will alter each agent's market power and perception of matching probabilities.[40] Changes in market power will cause hosts to re-optimize their pricing and rejection strategies, while changes in matching probabilities will lead guests to change how they send inquiries. Another type of adjustment might occur if the ranking algorithms can be manipulated by hosts. For example, if hosts know that the ranking algorithm favors listings with a particular amenity, they may either obtain that amenity or lie about having that amenity. The welfare and revenue implications of these endogenous adjustments are ambiguous. It will be interesting to see how online platforms deal with these long-run market dynamics.

[39]However, even in the college admissions market, incomplete information about rejection matters. Hoxby and Avery [15] show that some high school students do not use widely available information on college admissions and financial aid to make application decisions.

[40]Rochet and Tirole [24] and Weyl [28] show how pricing and other policies affect the relative gains accrued by heterogeneous agents in a marketplace.

References

[1] Ackerberg, D. A. and Rysman, M. 2005. Unobserved product differentiation in discrete-choice models: Estimating price elasticities and welfare effects. *RAND Journal of Economics 36,* 4, 771–788.

[2] Albrecht, J., Gautier, P. A., and Vroman, S. 2006. Equilibrium directed search with multiple applications. *The Review of Economic Studies 73,* 4, 869–891.

[3] Athey, S., Calvano, E., and Gans, J. 2013. The impact of the internet on advertising markets for news media. NBER Working Paper 19419, National Bureau of Economic Research, Inc.

[4] Blake, T. and Coey, D. 2014. Why marketplace experimentation is harder than it seems: The role of test-control interference. In *Proceedings of the Fifteenth ACM Conference on Economics and Computation*. EC '14. ACM, New York, NY, USA, 567–582.

[5] Burdett, K., Shi, S., and Wright, R. 2001. Pricing and matching with frictions. *Journal of Political Economy 109,* 5, 1060–1085.

[6] Casadesus-Masanell, R. and Halaburda, H. 2012. When does a platform create value by limiting choice? SSRN Scholarly Paper ID 1677624, Social Science Research Network, Rochester, NY. Nov.

[7] Chade, H., Lewis, G., and Smith, L. 2013. Student portfolios and the college admissions problem. SSRN Scholarly Paper ID 1358343, Social Science Research Network, Rochester, NY. Mar.

[8] Chen, Y. and Yao, S. 2012. Search with refinement.

[9] Crepon, B., Duflo, E., Gurgand, M., Rathelot, R., and Zamora, P. 2013. Do labor market policies have displacement effects? evidence from a clustered randomized experiment. *The Quarterly Journal of Economics 128,* 2, 531–580.

[10] Dinerstein, M., Einav, L., Levin, J., and Sundaresan, N. 2013. Consumer price search and platform design in internet commerce. *Working Paper*.

[11] Ghose, A. and Yang, S. 2009. An empirical analysis of search engine advertising: Sponsored search in electronic markets. *Management Science 55,* 10, 1605–1622.

[12] Hagiu, A. and Jullien, B. 2011. Why do intermediaries divert search? *The RAND Journal of Economics 42,* 2, 337–362.

[13] Hitsch, G. J., Hortacsu, A., and Ariely, D. 2010. Matching and sorting in online dating. *American Economic Review 100,* 1, 130–63.

[14] Horton, J. J. Misdirected search effort in a matching market: Causes, consequences and a partial solution. *Working Paper*, 2014.

[15] Hoxby, C. M. and Avery, C. 2012. The missing. Working Paper 18586, National Bureau of Economic Research. Dec.

[16] Jeziorski, P. and Segal, I. 2013. What makes them click: Empirical analysis of consumer demand for search advertising. *SSRN eLibrary 33.*

[17] Kircher, P. A. T. 2009. Efficiency of simultaneous search. *Journal of Political Economy 117,* 5, 861–913.

[18] Koulayev, S. Estimating demand in search markets: the case of online hotel bookings. *RAND Journal of Economics*.

[19] Koulayev, S. 2013. Search with dirichlet priors: Estimation and implications for consumer demand. *Journal of Business & Economic Statistics 31,* 2, 226–239.

[20] Lagos, R. 2000. An alternative approach to search frictions. *Journal of Political Economy 108,* 5, 851–873.

[21] Lewis, G. 2011. Asymmetric information, adverse selection and online disclosure: The case of eBay motors. *The American Economic Review 101,* 4, 1535–1546.

[22] Pallais, A. 2013. Inefficient hiring in entry-level labor markets. Working Paper 18917, National Bureau of Economic Research. Mar.

[23] Petrongolo, B. and Pissarides, C. A. 2001. Looking into the black box: A survey of the matching function. *Journal of Economic Literature 39,* 2, 390–431.

[24] Rochet, J.-C. and Tirole, J. 2003. Platform competition in two-sided markets. IDEI Working Paper 152, Institut d'Économie Industrielle (IDEI), Toulouse.

[25] Roth, A. E. and Xing, X. 1997. Turnaround time and bottlenecks in market clearing: Decentralized matching in the market for clinical psychologists. *Journal of Political Economy 105,* 2, 284–329.

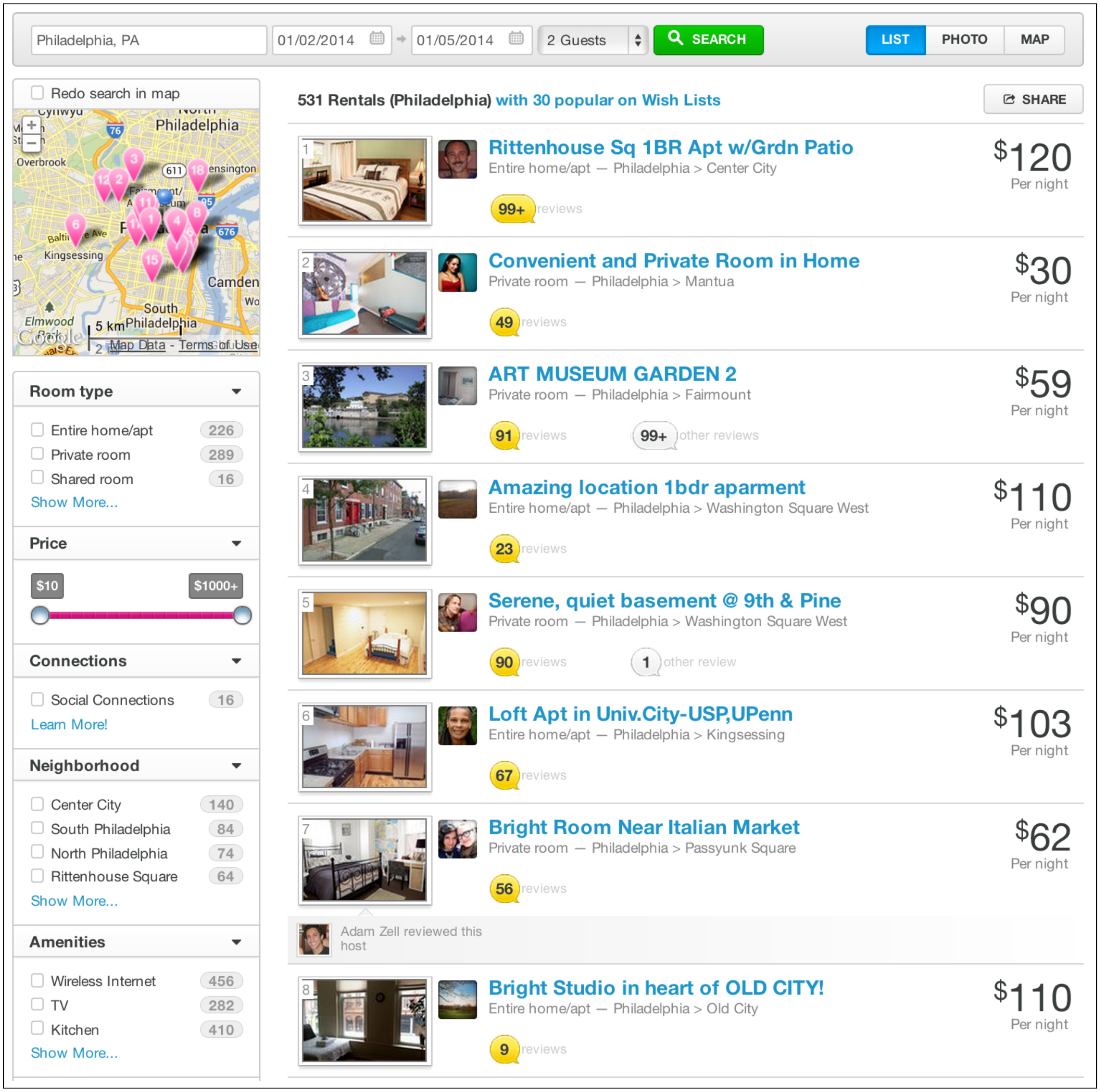
[26] Santos, B. D. l. and Koulayev, S. 2012. Optimizing click-through in online rankings for partially anonymous consumers. Working Paper 2012-04, Indiana University, Kelley School of Business, Department of Business Economics and Public Policy.

[27] Santos, D. l., Babur, Hortacsu, A., and Wildenbeest, M. R. 2013. Search with learning. SSRN Scholarly Paper ID 2163369, Social Science Research Network, Rochester, NY. Dec.

[28] Weyl, E. G. 2010. A price theory of multi-sided platforms. *The American Economic Review 100,* 4, 1642–1672.

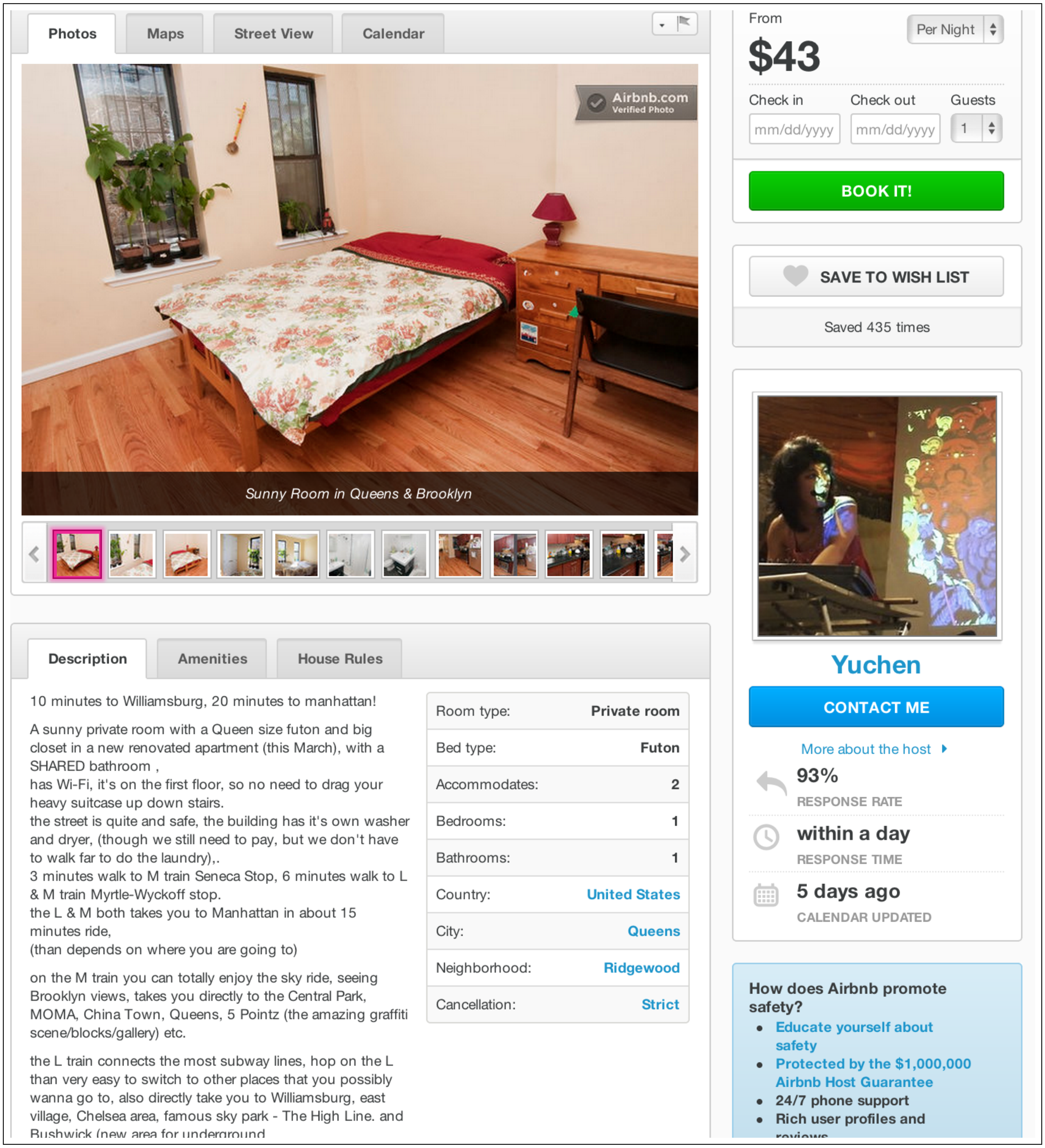
7 Figures

Figure 1: Search View



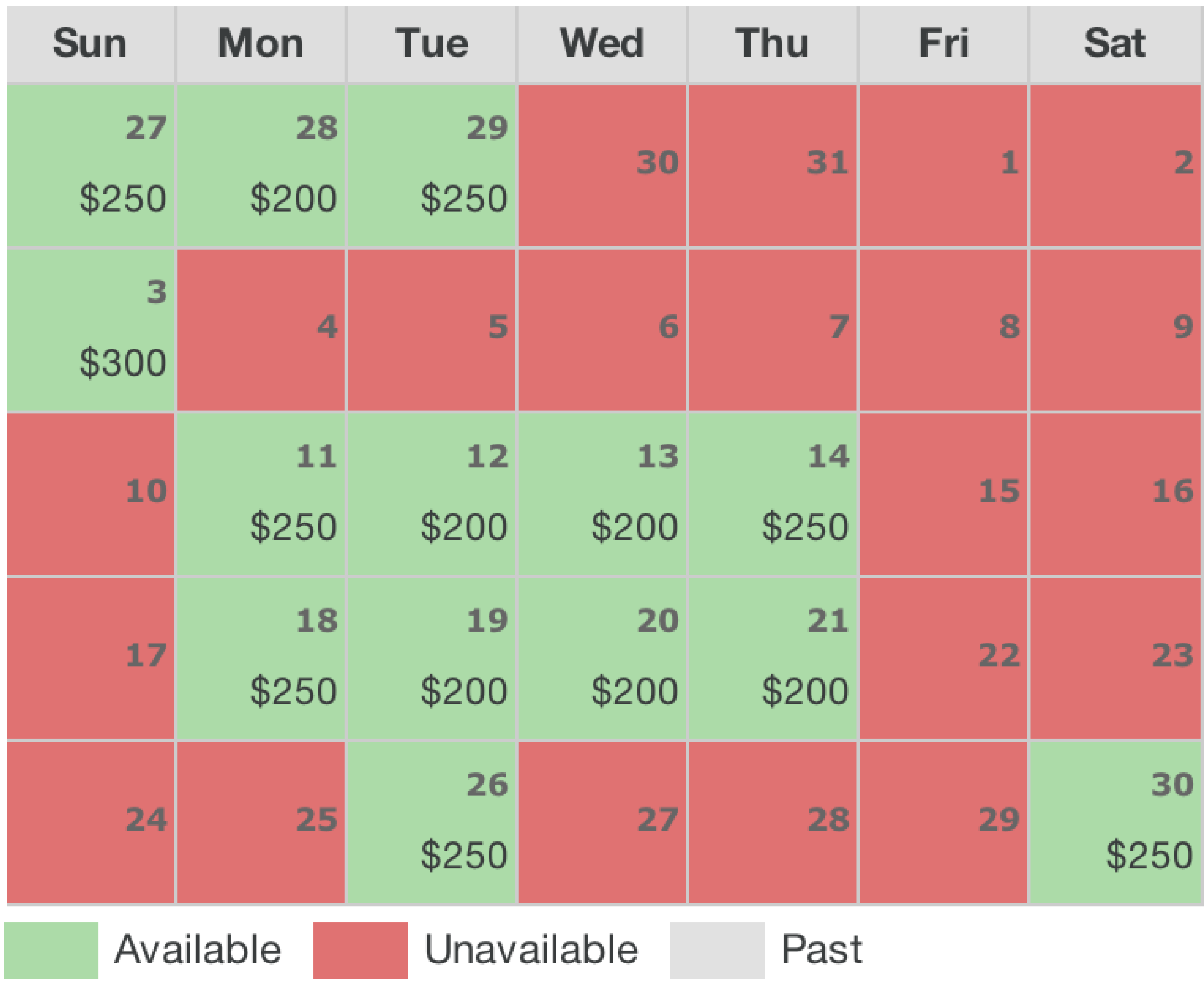
Above are the results of a search in Philadelphia for January 2, 2014 to January 5, 2014. Selecting filters or moving the map changes the set of displayed results. The searcher can scroll the page to see 21 listings before she is prompted to go to the next page of results.

Figure 2: Listing View



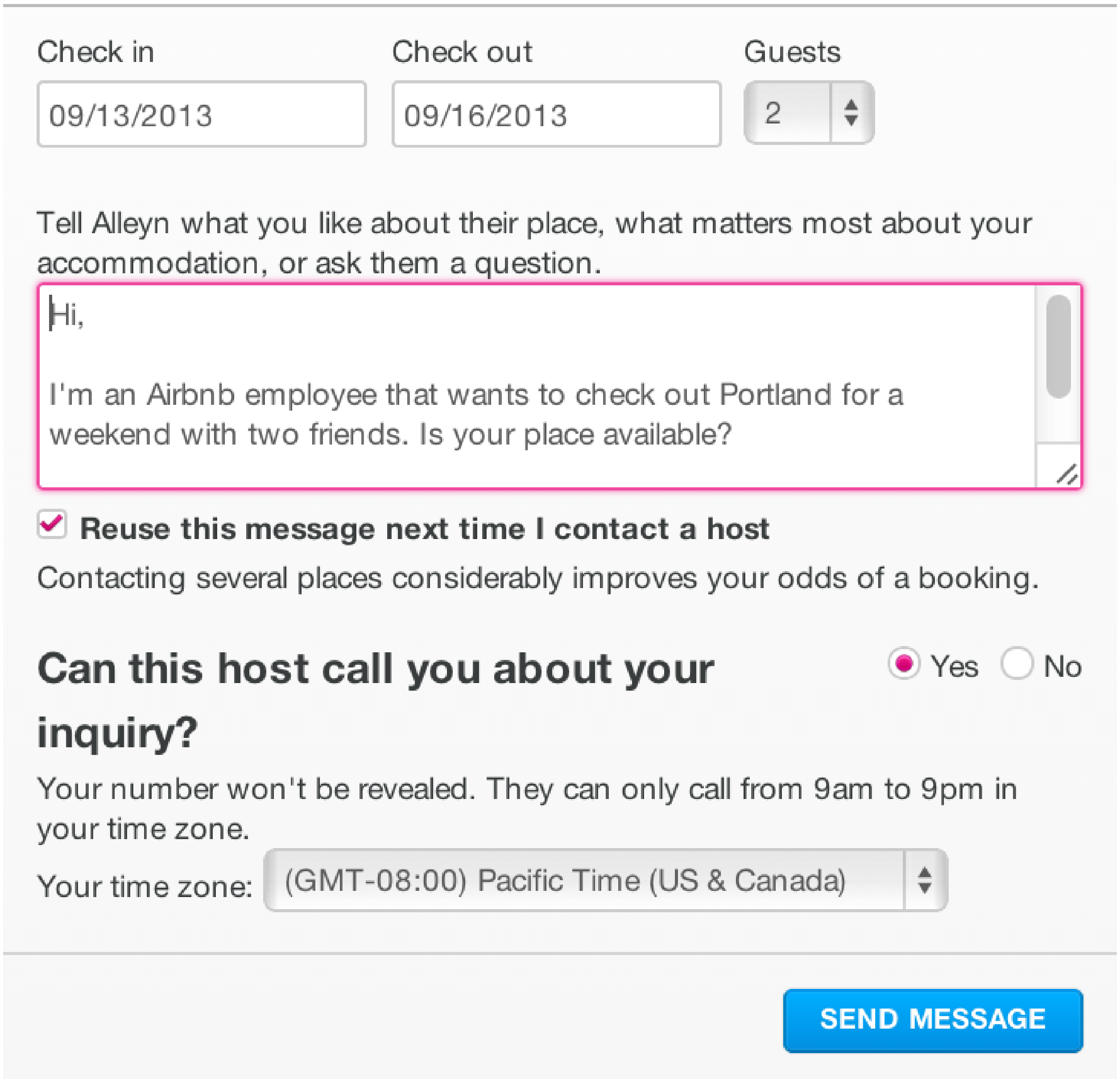
A searcher who clicks on a listing in the search results sees the following view. The ratings and text of reviews for the listing are visible lower on the page.

Figure 3: Listing Calendar



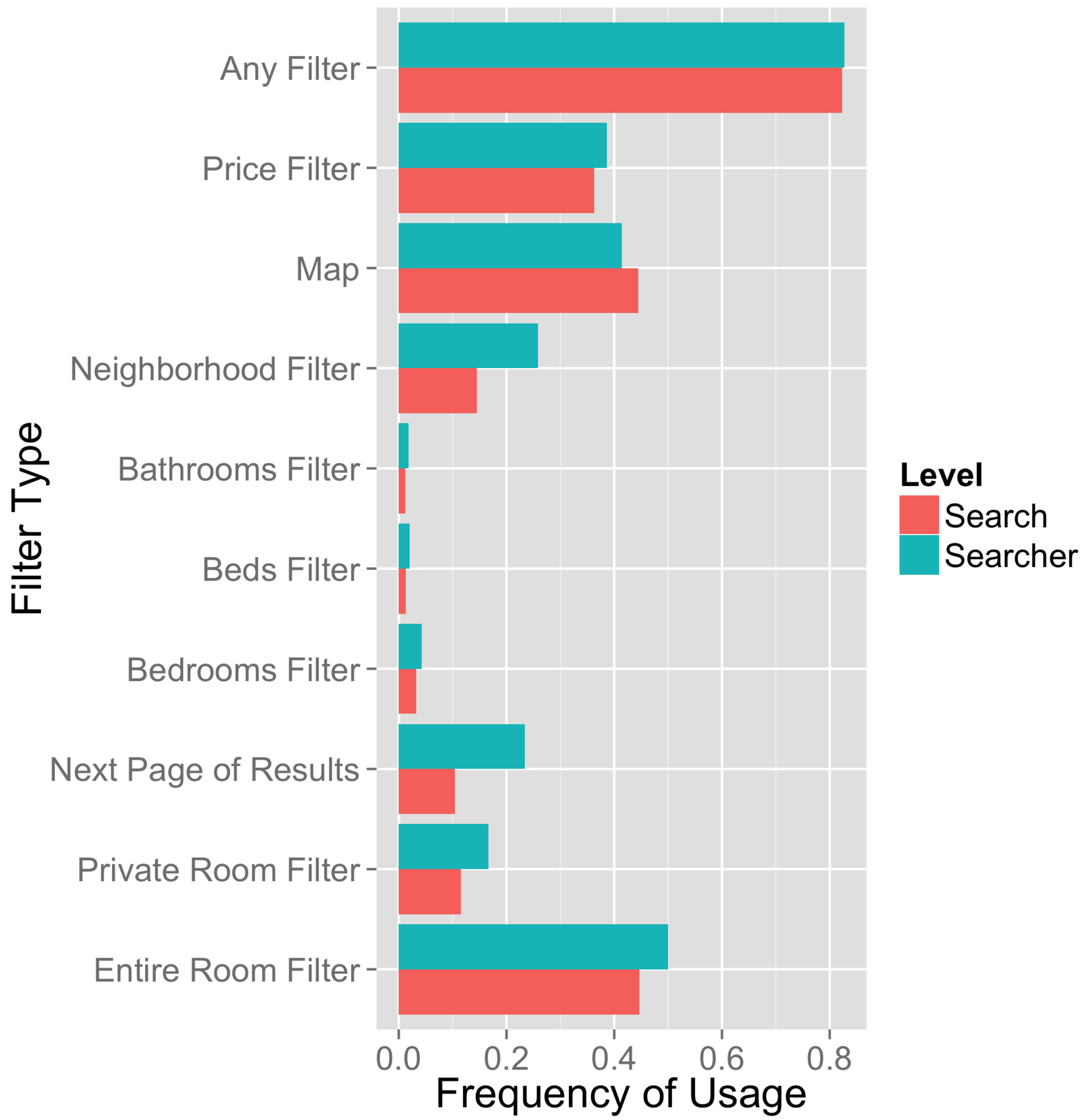
Above is the calendar tab of the listings page. Dates that have already been booked or that the host has marked unavailable are in red.

Figure 4: Inquiry Submission Form



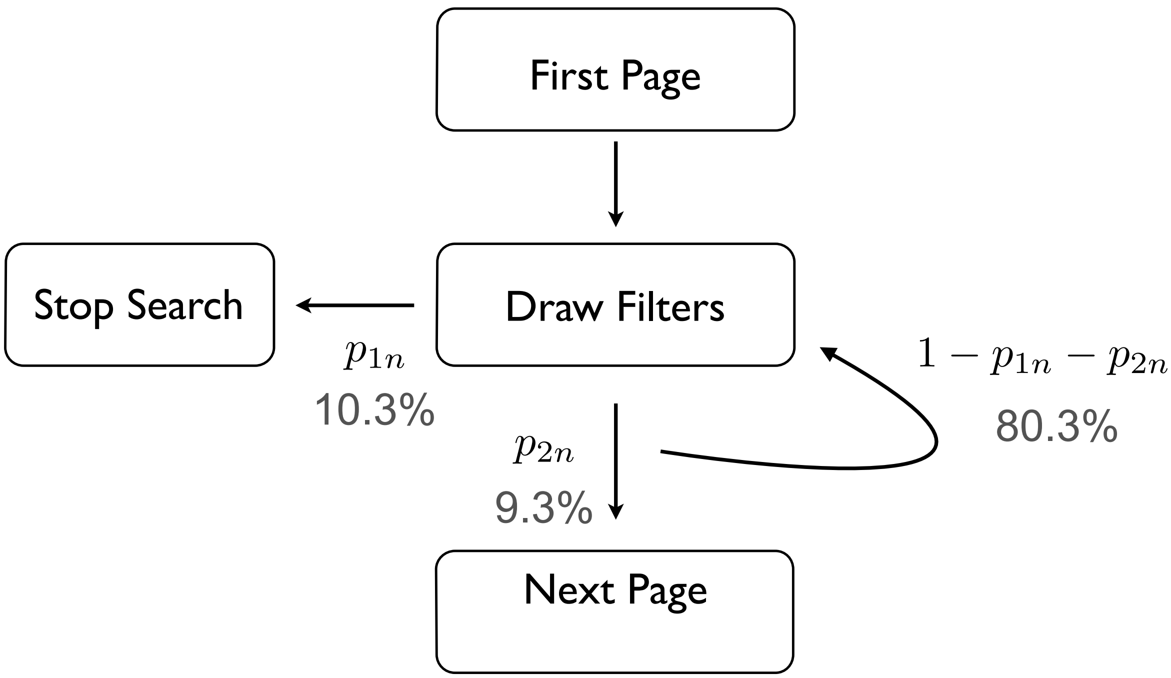
Above is the prompt that searchers see when they click the “Contact Me” button.

Figure 5: Frequency of Filter Usage



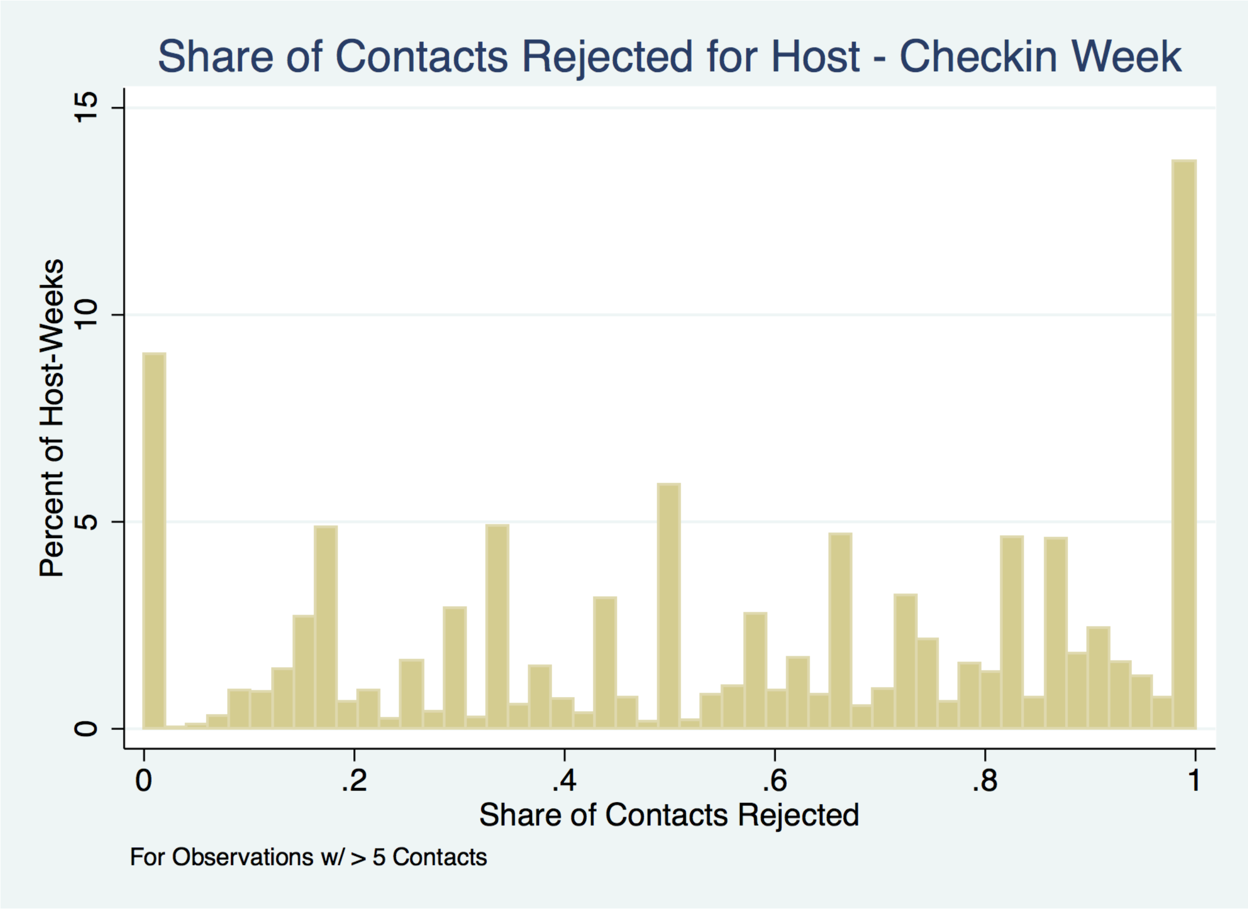
The above plot displays the share of searches and searchers that use each respective filter.

Figure 6: Consideration Set Formation Model



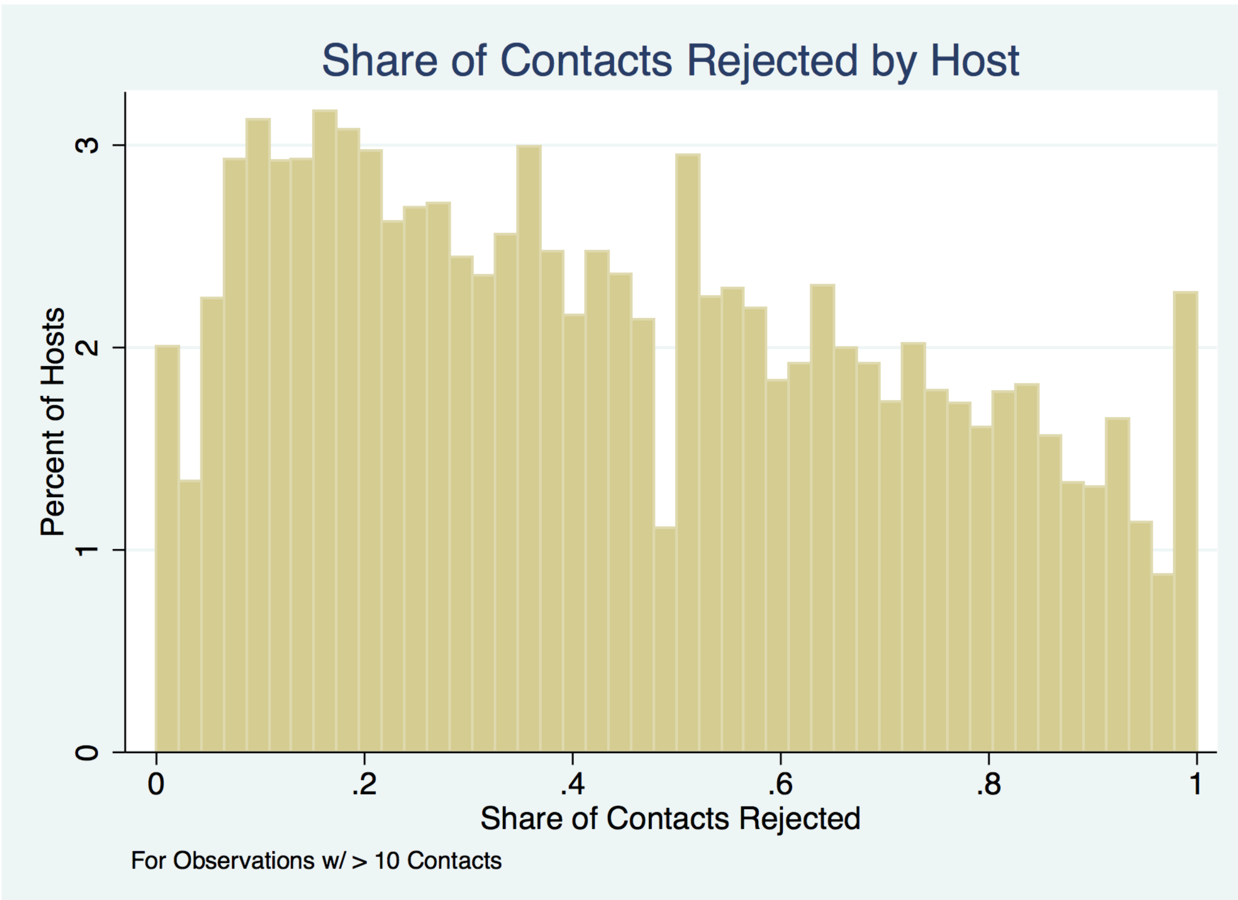
The above plot displays the possible transition paths for a seacher using the search engine. "n" denotes the sequential order of the page view and the displayed probabilities are averages across the sample.

Figure 7: Average Rejection Rates by Host-Week



Each observation in the above histogram is a listing and a week of check-in in major US markets in 2012. Only observations with more than 5 inquiries for that week are included.

Figure 8: Average Rejection Rates by Host



Each observation in the above histogram is a listing in a major US market in 2012 with at least 10 inquiries.

8 Tables

Table 1: Model of the Next Search Action

	Stop Search	Next Page of Results
Constant	-2.161	-2.747
Page View Number	-0.024	-0.017

Number of Observations: 207816

The above table displays the coefficients of a multinomial logistic regression which predicts whether the next action of a searcher is to do a new search, to stop search or to click to the next page of results. The "Page View Number" is the chronological order of the page view within the set of searches that a given searcher conducted. The normalized category in the regression is to do a new search.

Table 2: Filtering Behavior During Search

	Dependent variable:		
	Used Neighborhood Filter or Map	Used Room Type Filter	Used Price Filter
	(1)	(2)	(3)
Page View Number	0.017***	0.013***	0.007***
	(0.0002)	(0.0001)	(0.0001)
Constant	−0.095***	−0.844***	−0.101***
	(0.004)	(0.004)	(0.004)
Observations	424,884	424,884	424,884

*p<0.1; **p<0.05; ***p<0.01

The above table displays the coefficients of a logistic regression which predicts whether a given search uses a specific filter. The "Page View Number" is the chronological order of the page view within the set of searches that a given searcher conducted.

Table 3: Does Rejection Lead to Fewer Bookings?

	(1)	(2)
Rejected	-0.503***	-0.505***
	(0.00127)	(0.00128)
Week FE	No	Yes
Market FE	No	Yes
Trip Characteristics	No	Yes
Observations	547680	547680

The above table displays the results of a linear probability model that predicts whether an individual will eventually book given that their initial inquiry was rejected. Specification 2 includes week, market and trip fixed effects.

Table 4: Demand Estimates (In Dollar Terms)

	Baseline	With Listing RE
Professional Pic	0.807 (0.662)	0.839 (0.855)
Capacity	-2.455^{***} (0.311)	-2.343^{***} (0.338)
1 Room	7.393^{***} (1.671)	5.971^{***} (1.818)
Num Reviews	0.052^{***} (0.018)	0.065^{**} (0.026)
No Review	52.987^{***} (7.750)	47.775^{***} (8.808)
Avg. Rating	10.787^{***} (1.612)	10.005^{***} (1.834)
Location Quality	-1.731 (3.652)	1.010 (4.665)
Location Quality Sq.	0.426 (0.748)	-0.105 (0.964)
No Location Quality	-7.903^{***} (1.285)	-7.528^{***} (1.210)
Prop. Manager	-7.069^{***} (0.938)	-8.970^{***} (1.229)
Entire Prop.	7.961^{***} (0.982)	7.699^{***} (1.141)
Filter for Listing Neighborhood	6.156^{***} (0.725)	6.268^{***} (0.685)
Filter for Listing Room Type	13.282^{***} (0.863)	12.381^{***} (0.810)
Unusual Prop. Type	-20.527^{***} (3.849)	-12.274^{***} (4.697)
Listing Price * No Price Filter	-0.306^{***} (0.022)	-0.279^{***} (0.021)
Listing Price * Filter Price [10,100)	-1.000^{***} (0.060)	-1.000^{***} (0.060)
Listing Price * Filter Price In [100,200)	-0.514^{***} (0.023)	-0.486^{***} (0.024)
Listing Price * Filter Price In [200,300)	-0.348^{***} (0.023)	-0.317^{***} (0.023)
Listing Price * Filter Price > 300	-0.293^{***} (0.024)	-0.269^{***} (0.024)
1 Room * Num. Guests	-5.690^{***} (0.666)	-4.872^{***} (0.658)
Capacity * Num. Guests	0.293^{***} (0.096)	0.319^{***} (0.094)
Listing Price * Age NA	0.076^{***} (0.018)	0.074^{***} (0.017)
Listing Price * Age	0.002^{***} (0.0004)	0.001^{***} (0.0004)
Listing Price * Guest Reviewed	-0.022^{*} (0.011)	-0.020^{*} (0.011)
Listing Price * Nights	0.015^{***} (0.002)	0.013^{***} (0.002)
Listing Price * Guests	0.029^{***} (0.003)	0.029^{***} (0.003)
SD. Error	18.17	17.02
SD. Listing Random Effect		8.28
Num. Search Attempts	8,987	8,987
Num. Guest-Host Obs.	569,864	569,864

The above table displays the coefficients associated with the outside option in the demand model for searchers. The coefficients are normalized by the price coefficient for searchers who set a price filter between 10 and 100 dollars. The demand model is estimated on a sample of all searchers who used at least one neighborhood filters in City X from April to July of 2013. Neighborhood fixed effects are included in both specifications. "Listing Price" refers to the listing price inclusive of fees. "Prop. Manager" is an indicator that takes on the value 1 if a host manages more than 4 listings. "Professional Pic" is an indicator for whether the picture was taken by a verified Airbnb photographer. "Entire Prop." is an indicator variable that takes the value 1 when the entire property is being rented out. "Filter for Listing Neighborhood" is an indicator variable for whether the neighborhood a listing is in was filtered for. "Location Quality" refers to a locally smoothed value of the location review ratings for listings close to the viewed listing. "Capacity" refers to the number of guests a host can accommodate. "Guest Reviewed" is an indicator variable for whether the searcher has been reviewed. Standard errors are in parentheses.

Table 5: Demand Estimates (In Dollar Terms):
Outside Option Variables

	Baseline	With Listing RE
Outside Option	148.561***	140.662***
	(16.644)	(19.236)
Out. Opt. * Num. Pages	−0.332***	−0.310***
	(0.027)	(0.025)
Out. Opt. * Nights	3.191***	2.623***
	(0.352)	(0.343)
Out. Opt. * Guests	−1.085	−0.240
	(0.984)	(0.977)
Out. Opt. * Age	0.097	0.097
	(0.074)	(0.070)
Out. Opt. * Age NA	2.537	2.690
	(3.075)	(2.909)
Out. Opt. * Guest Reviewed	4.263**	4.067**
	(1.765)	(1.672)
Out. Opt. * Days In Advance	0.203***	0.211***
	(0.019)	(0.018)
Avg. Hotel Price * No Price Filter	−0.193***	−0.177***
	(0.036)	(0.034)
Avg. Hotel Price * Filter Price In [10,100)	−0.615***	−0.617***
	(0.047)	(0.046)
Avg. Hotel Price * Filter Price In [100,200)	−0.387***	−0.369***
	(0.037)	(0.036)
Avg. Hotel Price * Filter Price In [200,300)	−0.235***	−0.215***
	(0.039)	(0.037)
Avg. Hotel Price * Filter Price > 300	−0.176***	−0.164***
	(0.040)	(0.038)
SD. Error	18.17	17.02
Num. Search Attempts	8,987	8,987
Num. Guest-Host Obs.	569,864	569,864

The above table displays the coefficients associated with the outside option in the demand model for searchers. The coefficients are normalized by the price coefficient for searchers who set a price filter between 10 and 100 dollars. The demand model is estimated on a sample of all searchers who used at least one neighborhood filters in City X from April to July of 2013. Neighborhood fixed effects are included in both specifications. "Num. Pages" refers to the number of search result pages browsed by the searcher. "Avg. Hotel Price" refers to the average hotel price, "Days Ahead" refers to the number of days before the check-in at which the inquiry was sent. "Guest Reviewed" is an indicator variable for whether the searcher has been reviewed. Standard errors are in parentheses.

Table 6: The Determinants of Rejection by Host

(a) Guest Characteristics

	Baseline	Listing Random Effects
Guest Not American	-0.122^{***}	-0.161^{***}
	(0.040)	(0.045)
Guest Has Review	-0.265^{***}	-0.325^{***}
	(0.098)	(0.110)
Guest Has Prev. Stay	-0.036	-0.052
	(0.054)	(0.061)
Guest Has Desciption	-0.103^{**}	-0.066
	(0.050)	(0.057)
Guest Has Picture	-0.013	0.038
	(0.047)	(0.054)
Guest Has Rec.	0.062	0.044
	(0.105)	(0.118)
Guest Filled Female	-0.086	-0.107^{*}
	(0.055)	(0.061)
Guest Filled Male	0.063	0.031
	(0.057)	(0.064)
Days In Advance	-0.004^{***}	-0.005^{***}
	(0.001)	(0.001)
Nights	-0.205^{***}	-0.294^{***}
	(0.024)	(0.027)
Nights Sq.	0.014^{***}	0.020^{***}
	(0.002)	(0.002)
Guest Age	0.022	0.074
	(0.085)	(0.096)
Guest No Age	-0.002	-0.001
	(0.002)	(0.002)
N	17,831	17,831
Log Likelihood	-9,911.224	-8,823.851
SD. Host RE		1.298
Trip Characteristics	YES	YES
Month FE	YES	YES

The above table displays the coefficients on host characteristics in a logistic regression that predicts rejections by hosts. Column 1 displays results from a model without listing specific random effects and column 2 includes random effects. “Guest Rev.” is an indicator variable for whether the guest has been reviewed.

	Baseline	Listing Random Effects
Listing Capacity	−0.034***	−0.036
	(0.009)	(0.026)
Host Selective Instant Book	−0.434***	−0.221
	(0.075)	(0.161)
Host All Instant Book	−0.980***	−0.802***
	(0.098)	(0.191)
Social Network Instant Book	0.818***	0.793**
	(0.169)	(0.334)
No Prior Inquiries	−1.381***	−1.286**
	(0.424)	(0.577)
No Prior Responses	−1.719***	−0.127
	(0.436)	(0.603)
Response Time	0.011***	−0.001
	(0.002)	(0.004)
Response Rate	−3.036***	−1.271***
	(0.137)	(0.241)
Avg. Review Score	−0.084***	−0.017
	(0.027)	(0.042)
Price Listed	0.001***	0.001
	(0.0003)	(0.001)
No Reviews	−0.331**	0.035
	(0.136)	(0.216)
Log(Num. Room Rev. + 1)	−0.071***	−0.055
	(0.020)	(0.052)
Property Manager	−0.153*	−0.200
	(0.082)	(0.159)
Entire Property	0.433***	0.576***
	(0.080)	(0.153)
Host No Age	−0.693***	−0.711**
	(0.104)	(0.278)
Host Age	−0.023***	−0.024***
	(0.003)	(0.007)
Host Filled Female	−0.181***	−0.203
	(0.067)	(0.202)
Host Filled Male	−0.202***	−0.282
	(0.067)	(0.203)
N	17,831	17,831
SD. Host RE		1.298
Trip Characteristics	YES	YES
Month FE	YES	YES

The above table displays the coefficients on host characteristics in a logistic regression that predicts rejections by hosts. Column 1 displays results from a model without listing specific random effects and column 2 includes random effects. "Rev." refers to the number of reviews that the listing had at the time of the inquiry. "Prop. Mgr." is an indicator that takes on the value 1 if a host manages more than 4 listings. "Age" refers to the host's age. "Full Property" is an indicator variable that takes the value 1 when the entire property is being rented out.

(b) Host Characteristics

	Baseline	Random Effects
Prop. Mgr. * Rev. Guest	0.234*	0.397**
	(0.140)	(0.159)
Entire Prop. * Rev. Guest	0.224**	0.259**
	(0.101)	(0.114)
Prop. Mgr. * Days In Adv.	0.002**	0.003**
	(0.001)	(0.001)
Entire Prop. * Num. Guests	−0.118***	−0.100**
	(0.033)	(0.043)
N	17,831	17,831
Log Likelihood	-9,911.224	-8,823.851
SD. Host RE		1.298
Trip Characteristics	YES	YES
Month FE	YES	YES

The above table displays the coefficients on host characteristics in a logistic regression that predicts rejections by hosts. Column 1 displays results from a model without listing specific random effects and column 2 includes random effects. "Prop. Mgr." is an indicator that takes on the value 1 if a host manages more than 4 listings. "Full Property" is an indicator variable that takes the value 1 when the entire property is being rented out. "Days Ahead" refers to the number of days before the check-in date at which the searcher entered the market.

(c) Host - Guest Characteristics

Table 5: RE Search Outcomes

	Num. Simultaneous Contacts - 1	Num Sequential Contacts
2 Initial Con.		0.225***
		(0.027)
3 Initial Con.		0.265***
		(0.041)
4 Initial Con.		0.371***
		(0.065)
5 + Initial Con.		0.787***
		(0.044)
Nights	0.038***	0.044***
	(0.005)	(0.004)
Num. Guests	−0.011	−0.056***
	(0.009)	(0.007)
Guest Age	−0.010***	−0.011***
	(0.001)	(0.001)
Rev. Guest	−0.097**	0.004
	(0.043)	(0.036)
Exp. Guest	0.207***	0.093***
	(0.035)	(0.029)
Days Ahead	0.0001	0.003***
	(0.0003)	(0.0002)
Gender Filled Female	0.161***	0.171***
	(0.040)	(0.032)
Gender Filled Male	0.039	0.167***
	(0.042)	(0.033)
Month FE	Yes	Yes
N	16,404	10,023

Column 1 displays the results of a Poisson regression of number of inquiries after initial rejections on trip and guest characteristics. Column 2 displays the results of a regression of number of simultaneous inquiries on guest and trip characteristics. “Initial Con.” refers to the number of inquiries a searcher sent simultaneously at the start of search. “Rev.” and “Exp.” referred to whether the guest was reviewed or had a prior trip, respectively. “Days Ahead” refers to the number of days before the check-in date at which the searcher entered the market. “Filled Female” and “Filled Male” refer to cases when there was information on the gender of the searcher. The estimation sample consists of all non-foreign language inquiries sent by guests in City X between January 2012 and July 2013.

Table 6: Simulation Parameters

Calibrated Parameter	Value
Mean Hours To Book	20.24
Probability of Leaving After Non-Rejection	0.17
Sequential Search Multipler	2
Simultaneous Search Multipler	1.4

The mean hours to book and the probability of leaving after a non-rejection were calibrated to match the mean of those moments in the data. The sequential and simultaneous search multipliers were set so that the baseline simulation matches the intensity of sequential and simultaneous search in the data. Simultaneous search refers to the inquiries initially sent by a searcher. Sequential search refers to the inquiries that occur after an individual has been rejected from all the initial inquiries.

Table 7: Simulation Results

	Counterfactual	Share Contact	Share Book	Share Rejected	Avg. Revenue	Avg. Consumer Surplus
1	Data	0.621	0.374	0.473	112.754	
2	Baseline	0.628	0.353	0.438	137.507	31.516
3	Hide Screeners	0.603	0.448	0.181	171.189	42.359
4	Hide Congestion	0.625	0.362	0.420	140.105	32.720
5	Hide Unavailable	0.613	0.381	0.359	149.146	34.971
6	Full Choice Set	0.914	0.556	0.420	199.500	36.339
7	Hide Rejections	0.587	0.487	0	183.843	46.414
8	Hide Rejections and Full Choice Set	0.830	0.689	0	246.352	49.502
9	Avg. Quality Algorithm	0.708	0.362	0.517	118.166	34.270
10	Personalized Quality Algorithm	0.726	0.387	0.494	144.171	38.827
11	Transaction Probability Algorithm	0.729	0.390	0.487	144.923	40.114
12	Max. Consumer Surplus		0.713		254.869	72.926

Row 1 displays the outcomes that actually occurred for City X and the week of April 10, 2013. The remaining rows display outcomes under counterfactual scenarios. Share contacted is the share of searchers that sent at least one inquiry. Share booked is the share of searchers that booked a room. Share rejected refers to the share of all inquiries that are rejected by the host.

Table 8: A/B Test Results (Percentage Differences)

	Ranking Algorithm:					
	Avg. Quality			Personalized Quality		
	A/B Treatment Effect	Full Rollout Effect	Half Rollout Effect	A/B Treatment Effect	Full Rollout Effect	Half Rollout Effect
Share Booking	5.5	2.5	6.3	11.1	9.6	9.9
Share Contact	17.9	12.8	6.5	19.4	15.7	8.6
Initial Contacts Per Contacter	2.2	2.2	1.7	3.6	3.0	1.9
Share Rejected	24.3	17.9	2.8	21.2	12.7	1.9
Avg. Revenue	-14.1	-14.1	-3.0	5.2	4.8	5.6
Avg. Consumer Surplus	11.3	8.7	10.8	27.8	23.2	20.4

The above table displays three types of treatment effects for two of the proposed ranking algorithms: a non-personalized ranking algorithm and a personalized ranking algorithm. Column 1 shows the treatment effect (%) of seeing the non-personalized ranking algorithm in a 50% experiment at a searcher level. Column 2 shows the difference in outcomes across two identical markets where all searchers in one market saw the non-personalized ranking algorithm and all searchers in the other market saw the old ranking algorithm. Column 3 shows the difference in outcomes across two identical markets where half of all searchers in one market saw the non-personalized algorithm and all searchers in the other market saw the old ranking algorithm.

Table 9: Matching Function Estimates

	log(Contacters)	log(Inquiries)	log(Bookings)
	(1)	(2)	(3)
Personalized Rank	-0.068^{**}	-0.018	-0.200^{***}
	(0.027)	(0.042)	(0.045)
log(Listings)	0.053^{***}	0.058^{***}	0.139^{***}
	(0.003)	(0.005)	(0.005)
log(Searchers)	0.963^{***}	1.005^{***}	0.809^{***}
	(0.003)	(0.004)	(0.004)
Personalized Rank * log(Listings)	0.050^{***}	0.068^{***}	0.068^{***}
	(0.004)	(0.007)	(0.007)
Personalized Rank * log(Searchers)	-0.020^{***}	-0.034^{***}	-0.028^{***}
	(0.004)	(0.006)	(0.006)
Constant	-0.584^{***}	-0.357^{***}	-0.725^{***}
	(0.019)	(0.030)	(0.032)
Observations	1,440	1,440	1,440

Note: *p<0.1; **p<0.05; ***p<0.01

The above table displays estimates of a matching function from the simulated market outcomes where the number of guests and hosts were exogenously varied. "Personalized Rank" refers the algorithm that personalizes the results to the searcher.

A Appendix: Rejection Classification

This Appendix describes how to use a regularized logistic regression to classify the text of communications between hosts and guests. I divide each message into individual words and combinations of words (n-grams).[41] I use n-grams as features in a regularized logistic regression (RLR) that predicts whether a message a rejection or not.

The process of classification requires two steps: training and prediction. The procedure requires definitive cases when a message is either a rejection or not a rejection. A definitive case of a non-rejection occurs when a thread eventually leads to a transaction. A definitive case of a rejection occurs in a subset of messages for which hosts label that the room is "Not Available". I choose the top 30 thousand n-grams in my sample in addition to features for question marks and the number of sentences for my classifier. Prior to classification, I correct the text for common misspellings and determine the language of the text (See Appendix for details).

The RLR is estimated using 500 thousand labeled messages. The purpose of using regularization is to reduce over-fitting that might occur when so many features are included. The RLR penalizes the model for having too many non-zero coefficients. The classifier places high weight on n-grams such as "Sorry", "Unfortunately", "Is Occupied" for rejections. I test the validity of the procedure using two methods. First, I use the estimated model to classify a hold-out sample of labeled data. The classifier achieves a type 1 error of 2.6% and a type 2 error of 2.0%. That is, 2.6% of non-rejections are labeled as rejections and 2.0% of rejections are mistakenly labeled as non-rejections. The second validation I use is a manual inspection of 500 previously unlabeled messages. The classifier has a type 1 error of 3.4% and a type 2 error 8.1% in that subsample.

I combine classifications from the NLP classifier with cases when the response is discernibly through other methods. The final classification works as follows. If a contact led to a booking or if it was labeled as accepted by the host then it is classified as 'not rejected'. If a response was labeled by the host as a rejection or if there was no response within the week after the inquiry then the contact is classified as a rejection. If a response is classified as having a foreign language then it is not used in the analysis.[42]Lastly, if the contact does not fit into any of the above categories, the NLP classifier is applied to the first (or in some cases second) response by a host to a guest. In total, 49% of all inquiries were rejected. Of all contacts classified as rejections, 37% were cases in which a host did not respond, 30% were host classified rejections after an inquiry and 32% used the NLP classifier.

[41]For example, "car" is a 1-gram, "my car" is a 2-gram and "my fast car" is a 3-gram.

[42]An inquiry with 3 or more words was tested for being in a foreign language if it had more than 30% of words that were not in the English dictionary and were not common misspellings. The message text was

run through a language detection algorithm in Python called "guess-language" (http://code.google.com/p/guess-language/). If the algorithm guessed a valid non-English language then the inquiry was classified as being in a foreign language. A 50% cutoff was used for inquiries with 2 words.

B Appendix: Derivation of Sample Weights

Table B.1: Determinants of Ranking

	Baseline
Listing Age	-0.002^{***}
	(0.0001)
No Rating	-17.563
	(364.607)
Days Since Update	-0.003^{***}
	(0.001)
Price	0.0001
	(0.0001)
Avg. Rating	-0.341^{***}
	(0.048)
Num. Reviews	0.057^{***}
	(0.001)
No Reviews	11.590
	(364.607)
Entire Prop.	1.055^{***}
	(0.055)
Num. Pictures	0.034^{***}
	(0.002)
Neighborhood FE	YES
Num. Search Attempts	7,889
Num. Search-Listing Obs.	387,515

The above table displays the coefficients on listing characteristics which predict whether that listing is shown on the first page of search results. The model estimated is a conditional logistic regression.

C Appendix: An Urn and Balls Model of Matching

To see how search frictions affect the matching rate consider the simple urn and ball model of matching described in [23]. Suppose there exists a mass, G, of identical guests sending 1 inquiry each to a mass, L, of identical listings that can only transact with one guest each. The resulting number of matches produced is $L(1 - e^{-G/L})$. If an all-knowing social planner was matching guests and listings then the total amount of matches would be $\min(G, L)$.

To simplify further analysis, suppose that $L < G$. Therefore, the total inefficiency in the marketplace is a function of the failed matches, $Le^{-G/L}$. Inefficiency in this model comes from a coordination friction, where some listings reject guests because they are already booked. The welfare costs of the friction in this model are a function of the match utility of unmatched agents who could have been matched and the costs of wasted search.

Another friction that exists on Airbnb is that some listings are not actually available to anyone. Suppose that there is a mass of stale listings, U, which are not interested in transacting although they are visible to guests. The amount of matches in the marketplace becomes $L(1 - e^{\frac{-G}{L+U}})$, with the amount of inefficient search equal to: $Le^{\frac{-G}{L+U}} > Le^{-G/L}$. Stale listings increase inefficiency due to additional mismatch and wasted search costs. Stale listings could be even more important in dynamic matching markets because the likelihood of a stale listing being contacted increases as non-stale listings are booked.

Lastly, there are some hosts who are selective about which guests and trips they are willing to host. Suppose that listings are only willing to transact with a random subset of searchers. Let there be a mass of K selective sellers who are willing to transact with a random guest at a rate c. The remaining L - K non-selective sellers are willing to transact with anyone. The aggregate number of matches in the marketplace will be $(L-K)(1-e^{\frac{-G}{L+U}})+K(1-e^{\frac{-cG}{L+U}})$ and the number of inefficient inquiries is, $Le^{\frac{-G}{L}} + K(e^{\frac{-cG}{L}} - e^{\frac{-G}{L}})$, where the first term represents coordination and stale vacancies and the second term represents screening frictions. In turn, the cost of screening frictions depends on the share of listings which are selective, $\frac{K}{L}$ and on the selectivity of the listings, c.

The frictions in an urn and balls model all operate through the rejection channel. When inquiries are rejected, search effort is wasted and potential transactions do not happen. A full model, such as the one of the paper, would account for the ability of searchers to conduct sequential and simultaneous search. Furthermore, such a model would allow for some searchers to leave after being being accepted for a variety of reasons. Nonetheless, most of the intuition from an Urn and Balls model will hold in more realistic models of matching. Rejections will occur, will be costly to searchers and will slow the rate of match formation. The quantitative impact of frictions must, however, be determined within a more realistic model that accounts for heterogeneity, simultaneous search and dynamics.