Consumer Search and Seller Reliability on Airbnb

Jian Jia*    Liad Wagman†

September, 2017

Abstract

We study the consequences for sellers from being perceived as unreliable. We use a theoretical model to generate predictions about how information about reliability impacts sellers, and show that the effect depends on the search process that consumers endogenously follow: purchasing after finding information that qualifies a product, versus purchasing after not finding information that disqualifies a product. We then test our predictions against data from Airbnb listings in Manhattan. We find evidence suggesting that consumers tend to use the former approach, and demonstrate significant costs sellers may incur from negative information about their reliability.

Keywords: Reputation; reliability; information acquisition; sharing economy; Airbnb

JEL Classifications: D81, D83, L14, L15

*Stuart School of Business, Illinois Institute of Technology. Email: jjia5@hawk.iit.edu.
†Stuart School of Business, Illinois Institute of Technology. Email: lwagman@stuart.iit.edu.
1 Introduction

Despite their rapid rise, still-growing platforms such as Airbnb, Uber, and Lyft have already changed travelers’ behaviors. Airbnb, a sharing economy platform in the hospitality industry, enables residents to offer their homes as short-term rental accommodations. Airbnb landlords (henceforth, ‘hosts’) can provide three different accommodation types: an entire home/apartment, a private room, and shared space. An important component of the Airbnb platform is the reciprocal reputation system it facilitates for guests and hosts — within a 14-day deadline after a guest’s stay, both guest and host (blindly) review each other. If one side does not review the other, the other’s review becomes visible after the 14-day deadline. The literature has already shown that user reviews may play an important role in decision-making and purchasing behavior,\(^1\) but one distinguishing element of Airbnb’s reputation system is that the platform also provides an automated system review, which looks similar to any other review, for listings whose hosts cancel a confirmed reservation prior to the guest’s arrival.\(^2\) These cancellation reviews signal to travelers that there may be a higher than usual probability that their lodging plans may fall through at some uncertain point prior to their arrival—a potential situation that, especially in locales that are in high demand for temporary accommodations, can be quite costly. In addition, Airbnb as a platform, recognizing the impact that cancellations may have on travelers, discourages hosts from cancelling.\(^3\) Perhaps most importantly, these automatic reviews, since they are system generated and occur only upon a confirmed cancellation by a host, are credible and non-manipulable.

In this paper, we study how consumers incorporate negative information about seller

---


\(^2\)The automated cancellation review format is: “The host canceled this reservation X days before arrival. This is an automated posting,” where \(X \geq 1\) is as stated. For same-day cancellations, guests can still post a (non-automated) review. Prior to August 2015, the format was: “The reservation was canceled X days before arrival. This is an automated posting.”

\(^3\)In addition to receiving an automated cancellation review, hosts forfeit eligibility for ‘Superhost’ status on Airbnb for a year, a status related to metrics concerning host and listing reliability, and which we will later show can have monetary value to hosts. Hosts may also incur direct monetary punishments from Airbnb in the form of a reduction in the amount of a future payout.
reliability, and the resultant costs to sellers in equilibrium. In our model, consumers assess Airbnb hosts and their listings based on information such as reviews, and then proceed with booking a reservation only if they consider a listing to be a sufficiently-likely match for their needs. To identify whether a listing is a match, we show that consumers in equilibrium follow one of two search approaches: purchasing after finding information that ‘qualifies’ a listing to be a match, versus purchasing after not finding information that would ‘disqualify’ a listing from being a match. We generate predictions, depending on the type of search, on how negative information about reliability impacts sellers in equilibrium.

There are multiple benefits to looking at system-generated cancellation reviews as a measure for negative information about seller reliability. First, they are credible, non-manipulable, and demonstrably negative. Second, while prior works that study user-generated reviews tend to focus on products such as goods, hotels or restaurants (including Mayzlin et al., 2014, and Luca and Zervas, 2016), Airbnb reviews are much more personal and rate an experience in another individual’s dwelling. As a result, reviews on Airbnb are overwhelmingly positive (Zervas et al., 2015), which may grant further weight to the negative information implied by an automated cancellation review. Third, Airbnb does not show individual guest ratings of a listing but only averages, making it less clear-cut to objectively identify negative guest reviews in a data set—a non-issue for automated cancellation reviews.

We test our predictions against data from Airbnb listings in Manhattan. Our focus on Manhattan further reinforces the weight consumers may give to automated cancellation reviews due to not only the high demand for temporary lodging (in the event a host cancels, it may be difficult to find alternate accommodations), but also the unfavorable legal environment towards short-term rentals in New York.4 We demonstrate evidence suggesting that

4In the state of New York, it is illegal to rent entire dwellings in multiunit buildings (referred to as Class A in NY Multiple Dwelling Law) for less than 30 days; renting private rooms is permitted if the host is a primary occupant of the premises. Despite their illegality, thousands of hosts offer entire-home listings as short-term rentals, and many of these are concentrated in Manhattan, where enforcement is slowly on the rise. See, for instance, http://www.crainsnewyork.com/article/20170426/REAL_ESTATE/170429915/de-blasio-ramps-up-airbnb-enforcement. Guests may perceive cancellations as correlated with enforcement actions, not only by a municipal body but also potentially by a homeowner association.
consumers tend to use the approach of searching for information that qualifies a listing as a match prior to purchasing, as opposed to purchasing after not finding information that would disqualify a listing. We then quantify costs for sellers who are associated with negative information about reliability, beyond the costs imposed by Airbnb for cancellations, and show that guests ‘punish’ listings that have cancellation reviews, with those listings charging lower rates and incurring more vacancies.

1.1 Related Literature

This paper contributes to the literature on review informativeness and reputation systems in online marketplaces (Senecal et al., 2004; Cox et al., 2009; Hu et al., 2009; Cabral and Hortacsu, 2010; Zhang and Sarvary, 2015). Works in this literature also demonstrate that reviewers can have strategic incentives to manipulate reviews, which may result in under-reporting of negative reviews, particularly when users fear retaliation on platforms with reciprocal review systems (Bolton et al., 2013; Fradkin et al., 2017). Reviewers can also suffer from selection bias, where consumers are more likely to purchase and review products and services with which they are a priori satisfied (Li and Hitt, 2008; Masterov et al., 2015). Moreover, some reviewers may in fact be businesses leaving promotional content (or damaging content for competitors) to artificially inflate their online reputations (Mayzlin et al., 2014; Luca and Zervas, 2016). Reviews that can be generated anonymously, even at a cost, may encourage manipulation (Conitzer and Wagman, 2014), and hosts on Airbnb can in fact boost the ratings of their listings—by renting nights to friends or family, or to their own alternate accounts at minimum cost.

The automated cancellation reviews we study do not suffer from these potential manipulations. Our study thus adds to this literature by focusing on what is essentially demonstrably negative system reviews, which can only be triggered by seller actions and are effectively immune from manipulation. This allows us to avoid issues concerning review authenticity, and to focus on equilibrium behavior as a function of seller reliability as driven by these
cancellation reviews. The benefit of doing so is significant, because even the sheer possibility of review manipulation may impact the beliefs and actions of both buyers and sellers, which may result in different equilibrium behavior (Dellarocas, 2006; Anderson and Simester, 2014).

This paper also contributes to the growing literature on the sharing economy and Airbnb.\textsuperscript{5} Lee et al. (2015) point out that host reputation, including the number of reviews, host responsiveness, and host tenure, can impact the price of a listing. Zervas et al. (2015) indicate that Airbnb listings have higher average ratings compared to the hotel industry. Wang and Nicolau (2017) document that host attributes are the most important price determinants of Airbnb listings. Our work complements the above by shedding some light on how consumers incorporate information about seller reliability into their search and by providing empirical evidence that suggests that perceived unreliability is associated with significant costs for sellers. These costs include lower prices, less quantity sold, and ineligibility for a ‘Superhost’ reputation badge on the platform which we show can have significant monetary benefits.

The remainder of the paper is organized as follows. Section 2 presents our theoretical model. Section 3 describes the empirical methodology and the data we use, and Section 4 reports our empirical findings. Section 5 concludes.

\section{Model}

Consider a market with sellers (hosts) and buyers (guests). Hosts advertise their listings for rent at a price per night, $p$. Listings are divided into categories (e.g., by the number of bedrooms, the number of bathrooms, neighborhood, etc, as observed on search results). Let $q$ denote the ex-ante category of a listing as it appears on the platform’s search results, and let $v(q)$ denote the value to a guest from staying at an accommodation of category $q$.

Some listings, once viewed, may be perceived as ‘riskier’ to book, due to negative in-

\textsuperscript{5}Recent works include Zervas et al. (2016), Edelman et al. (2017), Fradkin (2017), Fradkin et al. (2017), and Kim et al. (2017).
formation about the host’s reliability (i.e., whether the host will follow through with the reservation after a potential guest observes a recent cancellation in the listing’s reviews, where we use ‘riskiness’ to refer to the uncertainty implied by purchasing from a less reliable seller). Let us differentiate among listings within a given category according to their hosts’ reliability or riskiness, where this risk is modeled as a probability of additional costs for guests. That is, guests who book riskier (safer) listings are more (less) likely to incur high additional costs. We denote the type of cost outcome a guest may incur from booking a listing as either $H$ or $L$, representing high and low expected additional costs, respectively. The prior that a listing is of type $H$ is denoted by $\lambda_\theta$, where $\lambda_\theta$ is higher for riskier listings and $\theta$ represents an ascending index of riskiness. Our assumption is that once a guest views a listing, they gain some assessment of its riskiness, for instance, by looking at its most recent reviews. A guest who books an accommodation of category $q$ then incurs costs $c \in \{c_L(q), c_H(q)\}$, where $c_H(q) > c_L(q)$.

We assume that listings are more likely to generate traffic and be viewed by guests when priced lower, where $D_q(p)$ represents the potential guest traffic that views a listing of category $q$ that offers a price $p$, with $D'_q(p) < 0$. That is, we abstract from the specific optimal stopping problem an individual guest may undergo by instead looking at aggregate traffic. A listing that is viewed by a potential guest is assumed to be a match for the guest’s needs in terms of its category. Once potential guests view a listing and gain an assessment of its riskiness, they can then attempt to acquire information about the listing’s initially unobserved characteristics by, for instance, closely examining the listing’s attributes.⁶

⁶The sequential ordering of these two actions is for simplicity and it is not essential for our analysis. Specifically, if guests have a ‘base level’ of examining a listing’s attributes, proceed to read its reviews, and then more closely examine the attributes of riskier listings, then our results are unchanged.
a standard fashion, we assume that additional scrutiny (higher $\alpha$), which raises the likelihood of an informative signal, is increasingly costly. For tractability, we consider quadratic cost $\kappa(\alpha) = k \cdot \alpha^2$ for achieving search intensity $\alpha$, with the scaling parameter $k > 0$ ensuring an interior equilibrium. Thus, when a guest acquires information about a listing, he receives a signal $s$, where:

$$s = \begin{cases} 
  c & \text{with probability } \alpha \\
  \emptyset & \text{with probability } 1-\alpha 
\end{cases}$$

That is, with probability $\alpha$ the guest receives an informative signal, learning the $H$ or $L$ type that would be associated with reserving this listing, and with probability $1 - \alpha$ the guest receives an empty signal and is left with his prior belief. Different guests may associate different match values with the same listing. The implicit assumption is that all listings and their hosts can be seen as possibly being a bad match by potential guests. That is, hosts do not know whether the particular group of guest traffic attracted to their listing at any given time will perceive them as type $H$ or $L$ conditional on obtaining an informative signal — and different guests may obtain different informative signals. However, similarly to guests, hosts are also able to assess their own perceived reliability, as represented by their listing’s risk type, $\theta$, and this riskiness affects how closely guests scrutinize a listing and how much they are willing to pay for a stay. Moreover, as we will show shortly, listings will be scrutinized differently depending on how much they charge.

We normalize the variables such that $1 > c_H(q) > v(q) > c_L(q) \geq 0$. The timeline of the game is as follows: Hosts first set prices for their listings, and each listing then receives some traffic from potential guests according to its price and category. Once guests view a listing, they assess its reputation or riskiness $\theta$, and acquire information to decide whether or not to proceed with booking a reservation.\footnote{While hosts can conduct their own screening of guests after guests make their reservation requests, many hosts use a feature known as “Instant Book,” where the pool of potential guests is prescreened ex-ante according to preset parameters, and qualified guests can book a listing without host approval.}
2.1 Guest Decision

Following an informative search, a guest will choose to book a listing that is revealed to be a low-cost type and reject a listing revealed to be a high-cost type. Following an uninformative search, for $\theta \in \{r, s\}$, a guest will book a listing if its value exceeds its expected cost, i.e., if:

$$v(q) - p_{q,\theta} \geq \lambda_\theta c_H(q) + (1 - \lambda_\theta)c_L(q).$$

Equivalently, the price $p_q$ must satisfy

$$p_{q,\theta} \leq v(q) - \lambda_\theta c_H(q) - (1 - \lambda_\theta)c_L(q).$$

If the rent offered by the host satisfies (1), the guest will book the listing following an uninformative search; in other words, the guest is searching for bad news or disqualifying information about the listing, and unless such information is found, the guest will proceed with the reservation. Let $\alpha_{q,\theta}$ denote the guest’s search intensity in this case. Then the guest’s problem is given by:

$$\max_{\alpha_{q,\theta}} (1 - \lambda_\theta)(v(q) - p_{q,\theta} - c_L(q)) + \lambda_\theta(1 - \alpha_{q,\theta})(v(q) - p_{q,\theta} - c_H(q)) - k\alpha_{q,\theta}^2.$$

Solving the above yields:

$$\alpha_{q,\theta}(p_{q,\theta}) = \frac{\lambda_\theta(c_H(q) - v(q) + p_{q,\theta})}{2k}.$$

2.2 Host Decision

To simplify, we assume that the marginal cost of hosting a guest is zero. Assuming a non-zero cost does not change the qualitative nature of the results. Hosts of listings of category
\( q \) and reputation (risk type) \( \theta \) set a price \( p_{q,\theta} \) according to:

\[
P_{q,\theta}(p_{q,\theta}) = D_q(p_{q,\theta})(1 - \lambda_{q,\theta}(p_{q,\theta}))p_{q,\theta}
\]

(2)

Anticipating the choice of search intensity by guests in the following stage, hosts solve for \( p_{q,\theta} \), which gives the following first-order condition

\[
D_q'(p_{q,\theta})p_{q,\theta}[2k - \lambda_{q,\theta}^2(c_H + p_{q,\theta} - v(q))] + D_q(p_{q,\theta})[2k - \lambda_{q,\theta}^2(c_H + 2p_{q,\theta} - v(q))] = 0
\]

(3)

Rearranging gives

\[
\epsilon_{q,\theta}(p_{q,\theta})[2k - \lambda_{q,\theta}^2(c_H + p_{q,\theta} - v(q))] + 2k - \lambda_{q,\theta}^2(c_H + 2p_{q,\theta} - v(q)) = 0,
\]

(4)

where \( \epsilon_{q,\theta}(p_{q,\theta}) \) is the price elasticity of listing traffic at price point \( p_{q,\theta} \). Equilibrium price is then implicitly defined by:

\[
p_{q,\theta} = \frac{1 + \epsilon_{q,\theta}(p_{q,\theta}) \frac{2k - \lambda_{q,\theta}^2(c_H - v(q))}{2 + \epsilon_{q,\theta}(p_{q,\theta}) \lambda_{q,\theta}^2}}{2 + \epsilon_{q,\theta}(p_{q,\theta}) \lambda_{q,\theta}^2},
\]

(5)

The right-hand side of (5) is increasing in \( \epsilon_{q,\theta}(p_{q,\theta}) \). Moreover, an increase in \( \theta \) (and thus in \( \lambda_{q,\theta} \)) decreases the second term on the right-hand side and thus decreases price. However, any decrease in price is coupled with a corresponding increase in \( \epsilon_{q,\theta}(p_{q,\theta}) \), and thus an increase in the first term on the right-hand side of (5). That is, while price indeed decreases in the riskiness level \( \theta \), the increase in demand that results from the price change acts as a moderating effect.

To proceed with closed-form solutions, we henceforth make the following assumption:

**Assumption 1** Demand exhibits constant price elasticity over the price range of interest.

---

\( ^8 \)A technical condition required for the equilibrium to be well-specified is that \( \epsilon_{q,\theta}(p_{q,\theta}) \in (-1, 0] \) or \( \epsilon_{q,\theta}(p_{q,\theta}) < -2 \) holds at the equilibrium price. This is because the right-hand side of (5) does not give a well-defined price for \( \epsilon_{q,\theta}(p_{q,\theta}) \in [-2, -1] \).
Armed with Assumption 1, we have the following result.

**Proposition 1 (Bad News Search)** When demand is sufficiently elastic, an interior equilibrium exists and is well specified. In equilibrium, prices are lower for riskier listings relative to safer listings, but riskier listings exhibit lower vacancy rates, ceteris paribus.

**Proof.** Expressing search intensity as a function of price, we have:

\[
\alpha_{q,\theta} = \frac{2k(1 + \epsilon_{q,\theta}(p_{q,\theta})) + \lambda_0^2(c_H(q) - v(q))}{2k\lambda_\theta(2 + \epsilon_{q,\theta}(p_{q,\theta}))}
\]

(6)

The probability of receiving a booking for a listing of category \( q \) with riskiness \( \theta \) is:

\[
(1 - \lambda_\theta \alpha_{q,\theta})D_q(p_{q,\theta}) = \frac{2k - \lambda_0^2(c_H - v(q))}{2k(2 + \epsilon_{q,\theta}(p_{q,\theta}))}D_q(p_{q,\theta})
\]

(7)

The derivative of (7) with respect to \( \lambda_\theta \) gives:

\[
-2(2 + \epsilon_{q,\theta}(p_{q,\theta}))\lambda_\theta(c_H - v(q)) - \epsilon'_{q,\theta}(p_{q,\theta})\frac{\partial p_{q,\theta}}{\partial \lambda_\theta}(2k - \lambda_0^2(c_H - v(q)))
\]

\[
+D'_q(p_{q,\theta})(1 - \lambda_\theta \alpha_{q,\theta})\frac{\partial p_{q,\theta}}{\partial \lambda_\theta}
\]

(8)

The first expression in (8) is negative because \( \epsilon'_{q,\theta}(p_{q,\theta})\frac{\partial p_{q,\theta}}{\partial \lambda_\theta} > 0 \). The second expression in (8) is positive because \( D'_q(p_{q,\theta}) < 0 \) and \( \frac{\partial p_{q,\theta}}{\partial \lambda_\theta} < 0 \). This is intuitive. The first expression represents the decrease in vacancy due to guests screening riskier listings more intensely, while the second expression represents increased traffic due to a lower price being set for riskier listings.

Under Assumption 1, \( \epsilon'_{q,\theta}(p_{q,\theta}) = 0 \). Hence, the expression in (8) simplifies as follows:

\[
-2(2 + \epsilon_{q,\theta}(p_{q,\theta}))\lambda_\theta(c_H - v(q)) + D'_q(p_{q,\theta})(1 - \lambda_\theta \alpha_{q,\theta})\frac{\partial p_{q,\theta}}{\partial \lambda_\theta}
\]

(9)
The sign of (9) is equivalent to the sign of the following expression:

\[
-2 \frac{(2 + \epsilon_{q,\theta}(p_{q,\theta}))\lambda_{\theta}(c_H - v(q))}{2k(2 + \epsilon_{q,\theta}(p_{q,\theta}))^2} - \frac{\epsilon_{q,\theta}(p_{q,\theta})}{p_{q,\theta}}(1 - \lambda_{\theta} \alpha_{q,\theta}) \frac{4(1 + \epsilon_{q,\theta}(p_{q,\theta})k)}{\lambda_{\theta}^2(2 + \epsilon_{q,\theta}(p_{q,\theta}))^2}.
\] (10)

Substituting for price and screening intensity and rearranging, the expression in (10) is negative when \(\epsilon_{q,\theta}(p_{q,\theta}) > -\frac{\lambda_{\theta}^2(c_H - v(q))}{2k}\) and positive when \(\epsilon_{q,\theta}(p_{q,\theta}) < -\frac{\lambda_{\theta}^2(c_H - v(q))}{2k}\). It remains to verify that the price in (5) satisfies (1). This is the case provided that:

\[
v(q) - \lambda_{\theta} c_H - (1 - \lambda_{\theta})c_L - \frac{1 + \epsilon_{q,\theta}(p_{q,\theta})}{2 + \epsilon_{q,\theta}(p_{q,\theta})} \frac{2k - \lambda_{\theta}^2(c_H - v(q))}{\lambda_{\theta}^2} \geq 0.
\]

It is straightforward to see that this condition is satisfied when \(\epsilon_{q,\theta}(p_{q,\theta}) = -1\) or is in its close neighborhood, which implies the existence of a range on elasticity that satisfies (1). Rearranging and simplifying, we have that (1) is satisfied when

\[
\epsilon_{q,\theta}(p_{q,\theta}) \leq -\frac{2k - \lambda_{\theta}^2[v(q) + c_H(1 - 2\lambda_{\theta}) - 2c_L(1 - \lambda_{\theta})]}{2k - \lambda_{\theta}^2(c_H - c_L)(1 - \lambda_{\theta})}.
\]

Since this condition implies \(\epsilon_{q,\theta}(p_{q,\theta}) < -\frac{\lambda_{\theta}^2(c_H - v(q))}{2k}\), it follows that riskier listings exhibit lower vacancies under Assumption 1 when demand is sufficiently elastic. ■

The result in Proposition 1 has empirical implications. Specifically, when buyers employ the approach of searching for information that would disqualify a purchase:

**Hypothesis 1:** Riskier listings will tend to charge lower prices, ceteris paribus.

**Hypothesis 2:** Riskier listings will tend to exhibit lower vacancy rates, ceteris paribus.

Hypothesis 2 may initially seem counter intuitive. However, it is simply the result of less reputable sellers (higher \(\theta\)) having to doubly compensate buyers in terms of offering lower prices, both for the greater risk associated with booking their rentals, and for the more extensive scrutiny of their listings. In equilibrium, the lower prices offered by low-reputation
(high \( \theta \)) listings entail that they receive more traffic and overall more bookings, though their
profits will be lower. While this search approach is theoretically viable and may in fact hold
in some markets (Burke et al., 2012; Kim and Wagman, 2015), the next subsection presents
an alternate approach that may arise in equilibrium.

2.3 Searching for Qualifying Information

If a listing’s price is higher than specified by (1), then a potential guest will only proceed
with a reservation if the listing is revealed to be a low-cost type. In other words, the guest
does not allow for false positives—the guest is searching for information (‘good news’) that
would definitively qualify the listing prior to moving forward with a reservation. The guest
then chooses a search intensity as follows:

\[
\max_{\alpha_{q, \theta}} (1 - \lambda_{\theta})\alpha_{q, \theta}(v(q) - p_{q, \theta} - c_L(q)) - k\alpha_{q, \theta}^2
\]

Solving this maximization problem gives:

\[
\alpha_{q, \theta}(p_{q, \theta}) = \frac{(1 - \lambda_{\theta})(v(q) - p_{q, \theta} - c_L(q))}{2k}
\]

From the perspective of a host, when the price does not satisfy (1), expected profit is:

\[
\Pi_q^{\text{gn}}(p_{q, \theta}) = D_q(p_{q, \theta})(1 - \lambda_{\theta})\alpha_{q, \theta}(p_{q, \theta})p_{q, \theta}.
\]  \hspace{1cm} (11)

We have the following result:

**Proposition 2 (Good News Search)** When demand is sufficiently inelastic, an interior
equilibrium exists and is well specified. In equilibrium, prices are lower for riskier listings
relative to safer listings, but riskier listings exhibit higher vacancy rates, ceteris paribus.

**Proof.** Differentiating (11) with respect to \( p_{q, \theta} \), the first-order condition is given by

\[
D'(p_{q, \theta})(1 - \lambda_{\theta})\alpha_{q, \theta}(p_{q, \theta})p_{q, \theta} + D_q(p_{q, \theta})\frac{(1 - \lambda_{\theta})(v(q) - c_L - 2p_{q, \theta})}{2k} = 0
\]
which can be written as

\[(1 - \lambda_\theta)(v(q) - p_{q,\theta} - c_L(q))\epsilon_{q,\theta}(p_{q,\theta}) + v(q) - c_L - 2p_{q,\theta} = 0.\]  

(12)

Hence, price is implicitly defined by

\[p_{q,\theta} = \frac{1 + (1 - \lambda_\theta)\epsilon_{q,\theta}(p_{q,\theta})}{2 + (1 - \lambda_\theta)\epsilon_{q,\theta}(p_{q,\theta})} (v(q) - c_L),\]  

(13)

and the search intensity is given by

\[\alpha_{q,\theta}(p_{q,\theta}) = \frac{(1 - \lambda_\theta)(v(q) - c_L)}{2k(2 + (1 - \lambda_\theta)\epsilon_{q,\theta}(p_{q,\theta}))}.\]  

(14)

For \(p_{q,\theta}\) and \(\alpha_{q,\theta}(p_{q,\theta})\) to be well defined, \(\epsilon_{q,\theta}(p_{q,\theta}) \geq -\frac{1}{1 - \lambda_\theta}\) must be satisfied. Examining the fraction in (13), it is straightforward to see that it is decreasing in \(\epsilon_{q,\theta}\), and, holding \(\epsilon_{q,\theta}\) constant, it is also decreasing in \(\lambda_\theta\). Thus, as \(\theta\) (and thus \(\lambda_\theta\)) increases, price decreases, and its decrease is amplified by the corresponding increase in \(\epsilon_{q,\theta}\). That is, riskier listings increasingly lower their prices to compensate for their lower chance of finalizing a reservation, as guests only book following informative and positive signals.

The probability of receiving a reservation request can be simplified as:

\[D_q(p_{q,\theta})(1 - \lambda_\theta)\alpha_{q,\theta}(p_{q,\theta}) = D_q(p_{q,\theta}) \frac{(1 - \lambda_\theta)^2(v(q) - c_L)}{2k(2 + (1 - \lambda_\theta)\epsilon_{q,\theta}(p_{q,\theta}))}.\]  

(15)

The derivative of (15) with respect to \(\lambda_\theta\) gives

\[D'_q(p_{q,\theta}) \frac{(1 - \lambda_\theta)^2(v(q) - c_L)}{2k(2 + (1 - \lambda_\theta)\epsilon_{q,\theta}(p_{q,\theta}))} \frac{\partial p_{q,\theta}}{\partial \lambda_\theta} - \frac{(1 - \lambda_\theta)^3(v(q) - c_L)\epsilon'_{q,\theta}(p_{q,\theta})}{2k(2 + (1 - \lambda_\theta)\epsilon_{q,\theta}(p_{q,\theta}))^2} \frac{\partial p_{q,\theta}}{\partial \lambda_\theta} D_q(p_{q,\theta})\]  

(16)

\[\frac{(1 - \lambda_\theta)(v(q) - c_L)(4 - (1 - \lambda_\theta)\epsilon_{q,\theta}(p_{q,\theta}))}{2k(2 + (1 - \lambda_\theta)\epsilon_{q,\theta}(p_{q,\theta}))^2} D_q(p_{q,\theta})\]
Further simplifying, we have

$$
\alpha_{q,\theta}(p_{q,\theta}) \left[ D'_q(p_{q,\theta}) (1 - \lambda_\theta) \frac{\partial p_{q,\theta}}{\partial \lambda_\theta} - \frac{(1 - \lambda_\theta)^2 \epsilon_{q,\theta}(p_{q,\theta})}{2 + (1 - \lambda_\theta) \epsilon_{q,\theta}(p_{q,\theta})} D_q(p_{q,\theta}) \right] 
$$

(17)

$$
- \frac{(4 - (1 - \lambda_\theta) \epsilon_{q,\theta}(p_{q,\theta}))}{2 + (1 - \lambda_\theta) \epsilon_{q,\theta}(p_{q,\theta})} D_q(p_{q,\theta})
$$

With both $D'_q(p_{q,\theta})$ and $\frac{\partial p_{q,\theta}}{\partial \lambda_\theta}$ being negative, the first expression in the brackets in (17) is positive, representing the increase in traffic following a price decrease for riskier listings.

The second and third expressions are negative and represent the lower probability that potential guests obtain informative signals and proceed to book a listing. Once we apply Assumption 1 and substitute for $\frac{\partial p_{q,\theta}}{\partial \lambda_\theta}$, the middle term drops; simplifying, the sign of (17), given $\epsilon_{q,\theta}(p_{q,\theta}) \geq -\frac{1}{1 - \lambda_\theta}$, is the same as the sign of the following expression:

$$
- \frac{\epsilon^2_{q,\theta}(p_{q,\theta})(1 - \lambda_\theta)}{p_{q,\theta}} \frac{v(q) - c_L}{(2 + (1 - \lambda_\theta) \epsilon_{q,\theta}(p_{q,\theta}))^2} - \frac{4 - (1 - \lambda_\theta) \epsilon_{q,\theta}(p_{q,\theta})}{2 + (1 - \lambda_\theta) \epsilon_{q,\theta}(p_{q,\theta})} < 0,
$$

(18)

It remains to verify that the price in (13) in fact does not satisfy (1). This is the case if:

$$
v(q) - \lambda_\theta c_H - (1 - \lambda_\theta)c_L - \frac{1 + (1 - \lambda_\theta) \epsilon_{q,\theta}(p_{q,\theta})}{2 + (1 - \lambda_\theta) \epsilon_{q,\theta}(p_{q,\theta})} (v(q) - c_L) < 0.
$$

It is straightforward to see that this condition is satisfied when $\epsilon_{q,\theta}(p_{q,\theta})$ is sufficiently high (e.g., if it is close to 0). Rearranging and simplifying, we have that (1) is not satisfied when

$$
\epsilon_{q,\theta}(p_{q,\theta}) \geq -\frac{2\lambda_\theta (c_H - c_L) + c_L - v(q)}{(c_H - c_L)(1 - \lambda_\theta) \lambda_\theta}
$$

The above, combined with the requirement that $\epsilon_{q,\theta}(p_{q,\theta}) \geq -\frac{1}{1 - \lambda_\theta}$, completes the proof.

The result in Proposition 2 also has empirical implications. The first implication lines up with the former search approach, i.e., riskier listings will tend to charge lower prices, ceteris paribus. The second implication is in direct contrast to the former search approach and
provides an alternative hypothesis of what we would expect to see in the data. Specifically, when buyers employ the approach of searching for information that would qualify a purchase:

**Hypothesis 2-Alt:** Riskier listings will tend to exhibit higher vacancy rates, ceteris paribus.

The reason for the alternate hypothesis is that, despite the increase in traffic following price reductions, higher-risk hosts are significantly less likely to secure reservations. This is because buyers only book once they have qualified a listing as a match for their needs, a criterion that is more difficult for listings to satisfy, particularly those that are higher risk.

In addition to different predictions about vacancy rates, Propositions 1 and 2 also place differing constraints on the price elasticity of guest traffic (i.e., on demand). While we do not directly observe demand-side data, given that our empirical analysis is focused in a locale where short-term accommodations are in relative high demand, it is possible that demand is relatively inelastic.

### 3 Methodology and Data

We began by collecting all consumer-facing information and review content on the complete set of hosts who had advertised their listings in Manhattan on Airbnb. This dataset has monthly scrapes at slightly irregular intervals. Each listing is identified by a unique identifier and comes with time-invariant characteristics such as the host’s unique identifier, neighborhood, approximate locale (latitude and longitude positioning in a six-digit decimal format that indicates the approximate location of a listing), and property type (entire apartment, private room, or shared space; we omit the latter due to its relatively low numbers—approximately 300 listings on average—and focus on the former two listing types, which are much more densely dispersed). The listing information also contains some time-variant char-

---

9New York City comprises the second largest Airbnb market in the world. The data is publicly available and was obtained from: [http://insideairbnb.com/get-the-data.html](http://insideairbnb.com/get-the-data.html).

10There are 17 scrapes in total including Jan, Mar, April, May, Jun, Aug, Sep, Oct, Nov(two scrapes), Dec in 2015; and Jan, Feb, April, May, June, July in 2016.
acteristics such as listing price,\textsuperscript{11} the number of days during which the property is available for bookings over the next 30, 60, or 90 days, whether the host has a so-called ‘Superhost’ badge,\textsuperscript{12} number of reviews, review rating, cancellation policy, minimum nights per stay, the maximum number of guests, a measure of the host’s experience (number of days since the host’s first listing was created), review gap (number of days since the latest review), whether the listing is offered for instant booking (i.e., without requiring host approval), and the host’s average response time and response rate to guest inquiries.

We focus on short-term rentals, thus, we concentrate on listings that are offered for rent for less than 30 days, dropping the observations with minimum nights greater than 30. We also exclude observations with listing prices per night that exceed $1000 because some hosts may set their rates prohibitively high in lieu of blocking their calendars. We exclude observations before August 2015 because such observations do not contain important controls, such as cancellation policies, whether instant book was offered, and host response time and rate. We use the roughly monthly scrapes between August 2015 and July 2016, initially comprising 132,031 observations.

Next, we construct a measure of competition for each listing by using geographical mapping software (ArcGIS) to count the total number of other listings of the same type that are located in close proximity. We define close proximity by forming a geographic circle of radius 0.1 or 0.3 miles around each listing based on its approximate coordinates. This calculation is repeated for each time period, so these count measures are time-varying. We also calculate the number of days that are vacant for each listing in the period of 30-to-60 days ahead of each data time period. We focus on this time frame for vacancies for two reasons. First, some guest reservations for this time window are more likely to still be forthcoming and thus

\textsuperscript{11}Hosts may adjust prices of individual days. The listing price represents the “base price” chosen by the host for the listing, i.e., the price for days that are not specifically edited by the host. It is also the price potential guests observe when they do not enter specific dates.

\textsuperscript{12}Hosts who meet the following criteria receive a Superhost designation, which indicates high reliability: (i) Hosted at least 10 guests in the past year; (ii) maintained a high response rate and low response time; (iii) received primarily 5-star reviews; (iv) did not cancel guest reservations in the past year.
may possibly be affected by cancellations that arose since the previous data period (i.e., days 0-to-30 are more likely to have been previously booked by guests).\footnote{An alternate approach here is to focus on vacancy in days 0-to-30 and to lag the number of cancellations one period. The results are similar.} Second, 30-to-60 days is not too far in the future, making it more likely that potential reservations fall or partially fall in those dates, to make a meaningful comparison among listings).\footnote{While hosts may block certain days on their listings' calendars (e.g., when they are traveling), we believe that any such behavior is independent of the number of cancellations a listing has.}

The last step in constructing our dataset is tallying up host cancellations for each listing. To do so, we take advantage of the previously-mentioned formatting of these reviews, e.g., “The host canceled this reservation X days before arrival. This is an automated posting.” As a first step, we tally up the number of cancellation reviews for each listing in each time period. This number is one way to connect with the parameter $\theta$ in our theoretical model. In particular, listings with more cancellations correspond to riskier listings with lower seller reliability, which potential guests may associate with higher probabilities of additional costs.

Table 1 presents descriptive statistics of the whole sample, as well as separately for entire-home rentals and private-room rentals in Manhattan from August 2015 to July 2016. We separately report public listing information, the number of cancellations, and our measures of competition. The average review rating is quite high, consistent with findings by Zervas et al. (2015). The bottom two rows give the average number of same-type competitors in a 0.1-mile and 0.3-mile geographical radius. They indicate that there are, on average, 41 entire-home competitors in a 0.1-mile radius and 298 entire-home competitors in a 0.3-mile radius. Although the average number of cancellation reviews seems low, approximately 27\% of listings have at least one cancellation at the end of July 2016. Figure 1 depicts the distribution of cancellations across listings in the last time period of our sample, July 2016.

Figure 2 depicts trends of variables of interest for the two different listing types over time. Figures 2(a) and 2(e) show that the number of entire-home listings stays roughly the same overall but the number of private-room listings increases. The average prices and number of
cancellation reviews for both listing types share a similar pattern in Figures 2(b) and 2(d), respectively. The average number of vacant days in the period of 30-to-60 days ahead of each data pull is also similar for both listing types in Figure 2(c). Both listing types also have high average review ratings in Figure 2(f) of about 4.5 (in comparison, Zervas et al., 2015, report an average hotel rating of 3.8).

4 Empirical Evidence

4.1 Baseline Specification

We first test how the number of cancellation reviews affects listing vacancy and price for the entire sample, and then run subsample analysis, separating entire-home and private-room rentals. Our baseline specifications are as follows:

\[
\ln(Vacancy_{i,t}) = \alpha + \beta \text{Cancel}_{i,t} + \delta X' + \text{Neighborhood}_i + \text{Month}_t + \varepsilon_{it}. \tag{19}
\]

\[
\ln(Price_{i,t}) = \alpha + \beta \text{Cancel}_{i,t-1} + \delta X' + \text{Neighborhood}_i + \text{Month}_t + \varepsilon_{it}. \tag{20}
\]

In specification (19), the dependent variable is the logarithm of vacancy of listing \(i\) in days 30-to-60. In specification (20), the dependent variable is the logarithm of the price of listing \(i\) at time \(t\). The parameter \(\text{Cancel}_{i,t}\), the number of cancellation reviews, is our explanatory variable for both specifications. Controls in \(X\) include listing attributes such as the number of bedrooms, beds, bathrooms, reviews, as well as overall review rating, vacancy in days 0-to-30 (only for specification (20), to control for listings where immediate-term vacancy may drive price reductions), host response rate, an indicator of superhost status, an indicator of instant book, cancellation policy, the number of total listings the host has, maximum guests per stay, and minimum night per stay. \(\text{Neighborhood}_i\), indicating one of 32 neighborhoods to which the listing belongs, captures neighborhood fixed effects.\(^{15}\) We also add a time dummy

\(^{15}\)The neighborhood of a listing is either mentioned in the actual listing or obtained from its coordinates.
variable for each scrape month. We lag the parameter $\text{Cancel}_{i,t-1}$ in specification (20) in order for hosts to have sufficient time to adjust prices in the face of a new cancellation (since reservations for days 0-to-30 are likely to have already been made—this is similar to how vacancy is treated). Both specifications are estimated using ordinary least squares.

Table 2 shows the results for the vacancy regressions. Columns (1), (3), and (5) give the results without considering neighborhood and time fixed effects. The coefficient on the number of cancellation reviews is similar in all three specifications. In particular, in the entire-home subgroup considering neighborhood and time fixed effects, column (4) suggests that vacancy in days 30-to-60 will increase by 11.2% (about 3.4 days) following one standard deviation (around 2.5) increase in the number of cancellations. For private-room rentals, column (6) suggests a vacancy increase of 5.60% following one standard deviation increase in the number of cancellations. Instant book does not seem to benefit entire-home listings in comparison with private-room rentals, which exhibit a 9.36% increase in booked nights over days 30-to-60. We can also see that the number of reviews does not play a significant role in driving vacancies, despite the coefficients being significant in all specifications. However, higher review ratings and superhost eligibility do tend to lower vacancies.

Table 3 reports the results of listing price regressions. The coefficient on the number of cancellation reviews, while negative, is small. For instance, column (4) suggests that the listing price tends to decrease by 0.28% with one standard deviation increase in the (one-month lagged) number of cancellation reviews. Those negative effects may be quite small because hosts are able to adjust the prices of individual nights without needing to alter the listing’s base price. For instance, hosts who recently canceled on guests may reduce prices in the near future in order to encourage guest bookings and receive additional (non-cancellation) reviews for their listings.\footnote{Some hosts may also use Airbnb’s ‘Smart Pricing’ feature, which automatically changes the price of individual nights. We do not have data on which hosts use this feature.}

As far as a listing’s base price, Table 3 indicates that review number has a negligible

\footnote{Some hosts may also use Airbnb’s ‘Smart Pricing’ feature, which automatically changes the price of individual nights. We do not have data on which hosts use this feature.}
effect, whereas review rating has some effect, although due to high average reviews it is more likely to be a punishing effect for listings with low review ratings. Listings hosted by superhosts charge a higher base price (around 10% more for all three groups).

The results in Tables 2 and 3 on the number of cancellation reviews are consistent with our Hypothesis 1 and 2-Alt from Section 2. That is, hosts whose listings have more cancellations (i.e., listings associated with higher levels of unreliability, or risk level \( \theta \), in our theoretical model) would tend to have more vacancies and offer lower prices. The alignment with Hypothesis 2-Alt would suggest that for a given listing category, as captured by our various controls, potential guests search for information that would qualify a listing prior to booking a reservation, rather than for information that would disqualify a listing.

We also control for time-invariant listing attributes in both Tables 2 and 3. While these results are not reported, they relate to the variable \( v(q) \) in our theoretical model, which represents the value guests associate with a listing of category \( q \). In particular, it is interesting to note that the number of bedrooms and bathrooms play a major role in listing price and vacancy. For entire-home rentals in particular, listing price increases by 19.6% and vacancy decreases by 18.7% with an additional bathroom, and listing price increases by 9.4% and vacancy decreases by 11.3% with an additional bedroom.

### 4.2 Accounting for Cancellation Attributes

While the number of cancellation reviews clearly interacts with price and vacancy, it does not reflect all of the information that is presented to buyers. More specifically, these system-generated cancellation reviews also reveal how many days ahead of guests’ arrival the host canceled their reservations. Moreover, the recency of a cancellation may matter—a more recent cancellation that appears at the top of the listing’s review stack may be both more noticeable and possibly perceived as more indicative of the host’s unreliability than one that is buried beneath dozens of other reviews.

As an example, consider two listings, each with one cancellation, but one has a cancella-
tion that is at the top of its review stack and took place a day before guest arrival, while the other is on page 5 of its reviews (each review page shows 7 reviews on the desktop version of the platform) and took place 30 days before guest arrival. The listings are otherwise identical. In the previous subsection, both listings would have been treated identically, but they are unlikely to be treated as such by potential guests. One may also argue that by not accounting for these cancellation attributes, the empirical alignment with the parameter $\theta$, indicating the risk level of listings and their hosts’ unreliability in our theoretical model, may be lacking.

As a next step, we record for each cancellation review (i) its location in the listing’s review stack, and (ii) the number of days prior to guest arrival that the cancellation took place. We propose a “risk index” that would make it simple to incorporate these cancellation attributes into our specifications:

$$RiskIndex_{i,t} = \ln \left[ 1 + \frac{1}{N_{i,t}} \left( \frac{1}{\sqrt{D_1 * P_1}} + \frac{1}{\sqrt{D_2 * P_2}} + \ldots + \frac{1}{\sqrt{D_{n_{i,t}} * P_{n_{i,t}}}} \right) \right],$$

where $N_{i,t}$ is the number of reviews listing $i$ has at time $t$, $n_{i,t}$ represents the number of cancellations that listing $i$ has at time $t$, $D_j$, $j = 1, 2, \ldots, n_{i,t}$, $D_j \geq 1$, represents the “number of days before guest arrival” for each cancellation review, and $P_j$, $j = 1, 2, \ldots, n_{i,t}$, $P_j \geq 1$, gives the page number of each system cancellation review.\(^{17}\) Thus, the risk index captures both the number of days before guest arrival and the page number for each cancellation review in each time period. We normalize by the number of reviews to reduce bias, while the square roots prevent the denominators from being overspread. Table 4 shows the summary statistics of the risk index for the three different groups. Entire-home rentals exhibit a higher average than private-room listings although the magnitudes are relatively small. Figure 3 depicts risk-index averages of entire-home and private-room listings over time.

\(^{17}\)One limitation here is that fewer reviews are shown per page on the mobile app version of the platform, but the number is still proportional to the page number on the desktop version. We have alternately used the review’s position in the review stack in lieu of page number and obtained similar results.
Tables 5 and 6 report the effect of the risk index in analogous vacancy and listing price regressions. The results remain consistent with our Hypothesis 1 and 2-Alt. The coefficient on the risk index is larger for vacancies in days 30-to-60 of entire homes than of private rooms. Column (4) of table 5 indicates that vacancies in days 30-to-60 increase by 15.25% (about 4.57 days) with one standard deviation (around 10%) increase in the risk index. In contrast, column (6) suggests that private-room listings experience a 6.44% (about 1.93 days) increase in vacancies with one standard deviation increase in the risk index. Thus, these show more pronounced effects on vacancies for both entire-home and private-room listings when accounting for the attributes of cancellations.

From Table 6, the risk index also has a more negative effect on listing price compared to our previous specifications. For instance, the listing price of entire-home (private-room) rentals decreases 0.69% (0.93%) with one standard deviation increase in the risk index.\textsuperscript{18} Instant book, review number, and review rating exhibit similar effects on vacancies and price as previously. The loss of ineligibility for superhost status by those hosts who cancel reservations has significant monetary implications. Specifically, superhosts of entire-home (private-room) rentals have less vacancies under the risk-index regressions than in our previous specifications (with the number of cancellations), exhibiting 12.71% (6.31%) more occupancy in days 30-to-60.

\section*{4.3 Accounting for Competition}

Rental listings within the same neighborhood may face different degrees of competition either from nearby within-neighborhood listings or from listings that fall just outside a neighborhood’s boundaries. When a cancellation takes place, its effect on a listing may depend on the extent of nearby competition. While we are unable to account for demand-side dynamics directly, we operate under the assumption that hosts, accounting for equilibrium conditions...
considerations (including costs associated with any potential illegalities), make available listings that line up with market demand for a geographic area.

We use the measures of competition that we constructed in both our specifications in Sections 4.1 and 4.2. We use a geographical radius of 0.1 miles around the coordinates of each listing to tally up the number of competing listings of each type in each period.\(^{19}\) We hypothesize that the competitive effect from having more listings in a 0.1-mile radius would have a positive impact on vacancy and a negative impact on price. We estimate the following fixed-effect regressions:

\[
\begin{align*}
\ln(Vacancy_{i,t}) &= \alpha + \beta RiskIndex_{i,t} + \zeta Comp_{i,t} + \delta X' \\
&+ \eta RiskIndex_{i,t} \times Comp_{i,t} + Listing_i + Month_t + \varepsilon_{it} \\

\ln(Listing\ Price_{i,t}) &= \alpha + \beta RiskIndex_{i,t-1} + \zeta Comp_{i,t} + \delta X' \\
&+ \eta RiskIndex_{i,t-1} \times Comp_{i,t} + Listing_i + Month_t + \varepsilon_{it}
\end{align*}
\]

The dependent variables for both specifications are the same as previously. The variable \(Comp_{i,t}\) gives the number of competitors in a 0.1-mile radius of listing \(i\) in period \(t\). We also include interactions of the \(RiskIndex\) and \(Comp\) and center them both to better interpret the coefficients relative to average conditions, since both variables are continuous. Specifications (22) and (23) are estimated using ordinary least squares with standard errors clustered at the neighborhood level.

Tables 7 and 8 report the results. For entire-home listings, column (2) of Table 7 indicates that given the average number of competitors in a 0.1-mile radius, a listing would face a 16.35\% increase in its 30-to-60 day vacancy following a standard deviation increase in its number of cancellation reviews. Under the risk-index specification, vacancy would similarly
increase by 20.52%. On the other hand, column (6) suggests that private-room listings, under the average competitive environment, have a 7.1% increase in 30-to-60 day vacancy following a standard deviation increase in the number of cancellations, and vacancy would similarly increase by 7.5% under the risk-index specification.

Furthermore, under either specification, the effect on listing price is more negative in comparison to the same specifications but without accounting for the number of competitors. For entire-home (private-room) listings, with the average number of competing listings in a 0.1-mile radius of a listing, following a standard deviation increase in the risk index, the listing price decreases by 2.47% (1.07%). Finally, both Tables 7 and 8 suggest that vacancies (price) tend to increase (decrease) in the degree of competition, everything else held constant.

The number of competitors in a 0.1-mile radius does not appear to impact the vacancies of entire-home listings as significantly as it does private-room rentals. One possible explanation is that due to regulatory issues, the number of entire-home listings in Manhattan has remained depressed relative to demand. In contrast, the number of private rooms (which operate in less of a legal grey area in Manhattan) has steadily increased, as suggested in Figure 2(a). That is, private-room hosts face fewer legal and/or enforcement hurdles than entire-home hosts, and, as this issue has received significant attention in the press,20 would-be entire-home hosts may have been discouraged from listing their properties.

5 Conclusion

In this paper, we used automated system cancellation reviews to form a measure of seller reliability. We applied this measure to an analysis of Airbnb listings in Manhattan, and demonstrated evidence that suggests that this information may indeed be factored into the decision making of consumers, and less reliable sellers may suffer significant costs in terms of reputation (by way of being disqualified from superhost status), lower prices, and increased

vacancies. This evidence is consistent with our theoretical model.

Our study suffers from a number of limitations. First, we are only able to account for the base price of listings and not the price of individual nights. Second, we are unable to account for blocked calendar days that may arise due to, e.g., reservations from other platforms (or due to the actual cancellations—the platform automatically blocks those calendar dates). Third, we do not know whether the placement of listings in search results is affected (most likely negatively) by cancellation reviews. However, these limitations in a sense strengthen our findings, because they imply that our results represent lower bounds on the effects on listing price and vacancy.

One may also offer alternate explanations for our findings. For instance, one may argue that higher vacancies can be due to those hosts with cancellations setting higher prices, possibly because hosting incurs higher opportunity costs for them. We believe this is much less plausible, particularly since listing price (which by itself may provide some lower bound on the impact on actual nightly rates) appears to decrease in cancellations. Moreover, while our results regarding how consumers tend to search are potentially profound, our analysis is limited to Airbnb listings in Manhattan. However, these limitations suggest a number of promising directions for future work, including extending our analysis to other locales in both the U.S. and other countries (since norms, density, and the legal landscape, among other considerations, may be locale and culture specific, equilibrium behaviors could differ from what we see in Manhattan), as well as to other product platforms altogether.

References


Figure 1: Density of the distribution on the number of host cancellations across listings in Manhattan in July 2016
Figure 2: Average trend figures by time and room type

(a) Number of listing
(b) Average listing price/night
(c) Average Vacancy (days)
(d) Average No. of cancellation review
(e) Average competitors for each room type
(f) Average review rating
Figure 3: Average risk index by time and room type
Table 1: Summary Statistics

<table>
<thead>
<tr>
<th>Public Information:</th>
<th>N</th>
<th>Mean</th>
<th>S.D.</th>
<th>N</th>
<th>Mean</th>
<th>S.D.</th>
<th>N</th>
<th>Mean</th>
<th>S.D.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Listing Price ($)</td>
<td>132,031</td>
<td>174.66</td>
<td>114.76</td>
<td>86,300</td>
<td>218.68</td>
<td>120.343</td>
<td>45,731</td>
<td>100.04</td>
<td>41.75</td>
</tr>
<tr>
<td>No. Bedrooms</td>
<td>132,031</td>
<td>1.08</td>
<td>0.94</td>
<td>86,300</td>
<td>1.10</td>
<td>0.27</td>
<td>45,731</td>
<td>1.02</td>
<td>0.27</td>
</tr>
<tr>
<td>No. Bathrooms</td>
<td>132,031</td>
<td>1.09</td>
<td>0.66</td>
<td>86,300</td>
<td>1.10</td>
<td>0.344</td>
<td>45,731</td>
<td>1.09</td>
<td>0.32</td>
</tr>
<tr>
<td>No. Beds</td>
<td>132,031</td>
<td>1.52</td>
<td>0.95</td>
<td>86,300</td>
<td>1.73</td>
<td>1.053</td>
<td>45,731</td>
<td>1.14</td>
<td>0.47</td>
</tr>
<tr>
<td>Review Number</td>
<td>132,031</td>
<td>18.81</td>
<td>26.63</td>
<td>86,300</td>
<td>17.54</td>
<td>23.941</td>
<td>45,731</td>
<td>21.39</td>
<td>30.87</td>
</tr>
<tr>
<td>Review Rating</td>
<td>132,031</td>
<td>4.605</td>
<td>0.42</td>
<td>86,300</td>
<td>4.55</td>
<td>0.706</td>
<td>45,731</td>
<td>4.5</td>
<td>0.47</td>
</tr>
<tr>
<td>Vacancy Rate (next 30 days)</td>
<td>132,031</td>
<td>33.76%</td>
<td>0.30</td>
<td>86,300</td>
<td>30.42%</td>
<td>0.331</td>
<td>45,731</td>
<td>38.67%</td>
<td>0.42</td>
</tr>
<tr>
<td>Superhost Proportion</td>
<td>132,031</td>
<td>5.66%</td>
<td>0.17</td>
<td>86,300</td>
<td>5.22%</td>
<td>0.221</td>
<td>45,731</td>
<td>6.42%</td>
<td>0.17</td>
</tr>
<tr>
<td>Maximum guests per stay</td>
<td>132,031</td>
<td>2.92</td>
<td>1.68</td>
<td>86,300</td>
<td>3.48</td>
<td>1.825</td>
<td>45,731</td>
<td>1.98</td>
<td>0.89</td>
</tr>
<tr>
<td>Minimum night per stay</td>
<td>132,031</td>
<td>2.87</td>
<td>3.89</td>
<td>86,300</td>
<td>3.16</td>
<td>4.248</td>
<td>45,731</td>
<td>2.39</td>
<td>3.12</td>
</tr>
<tr>
<td>Appraisal Panel:</td>
<td>N</td>
<td></td>
<td></td>
<td>N</td>
<td></td>
<td></td>
<td>N</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No. of cancellation review</td>
<td>132,031</td>
<td>0.572</td>
<td>1.26</td>
<td>86,300</td>
<td>0.693</td>
<td>1.75</td>
<td>45,731</td>
<td>0.426</td>
<td>0.91</td>
</tr>
<tr>
<td>No. of same type competitors (0.1 mile radius)</td>
<td>132,031</td>
<td>58.59</td>
<td>39.56</td>
<td>86,300</td>
<td>41.65</td>
<td>27.97</td>
<td>45,731</td>
<td>21.12</td>
<td>13.92</td>
</tr>
<tr>
<td>No. of same type competitors (0.3 mile radius)</td>
<td>132,031</td>
<td>408.24</td>
<td>258.28</td>
<td>86,300</td>
<td>298.44</td>
<td>181.41</td>
<td>45,731</td>
<td>142.69</td>
<td>87.81</td>
</tr>
</tbody>
</table>

Note: Whole sample includes both entire-home and private-room listings. We do not report shared-space statistics because their market is quite small, comprising about 300 listings per monthly data pull on average.


Table 2: The effect of automated cancellation review on listing vacancy

<table>
<thead>
<tr>
<th>Room type</th>
<th>Minimum night</th>
<th>Room type control</th>
<th>Listing attributes Controls</th>
<th>Neighbourhood FE</th>
<th>Time FE</th>
<th>Adjusted R²</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of cancellation reviews</td>
<td>0.0034</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>0.0854</td>
</tr>
<tr>
<td>No. of reviews (in total)</td>
<td>0.031</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>0.0854</td>
</tr>
<tr>
<td>Review rating -</td>
<td>0.3254</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>0.0572</td>
</tr>
<tr>
<td>Superhost -</td>
<td>0.1268</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>0.0572</td>
</tr>
<tr>
<td>Instant book -</td>
<td>0.0936</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>0.0572</td>
</tr>
<tr>
<td>Minimum night -</td>
<td>0.0165</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>0.0572</td>
</tr>
<tr>
<td>No. of bedroom -</td>
<td>0.017</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>0.0572</td>
</tr>
<tr>
<td>No. of bathroom -</td>
<td>0.037</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>0.0572</td>
</tr>
<tr>
<td>No. of beds -</td>
<td>0.031</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>0.0572</td>
</tr>
<tr>
<td>Maximum guests per stay -</td>
<td>0.065</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>0.0572</td>
</tr>
<tr>
<td>Listing attributes not reported include cancellation policy, No. of bedrooms, No. of bathrooms, No. of beds.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: Standard errors are clustered by neighborhood and reported in parentheses. There are 32 neighborhood dummies, 11 monthly dummies. Controls for listing attributes not reported include cancellation policy, No. of bedrooms, No. of bathrooms, No. of beds. *** and ** indicate significance at the 1%, 5%, and 10% levels.
Table 3: The effect of automated cancellation review on listing price

<table>
<thead>
<tr>
<th></th>
<th>Whole Sample</th>
<th>Entire Home Rental</th>
<th>Private Room Rental</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>No. of cancellation reviews</strong></td>
<td>0</td>
<td>0.0003</td>
<td>0.0003</td>
</tr>
<tr>
<td><strong>Number of reviews</strong></td>
<td>-0.00001</td>
<td>-0.00002</td>
<td>-0.00002</td>
</tr>
<tr>
<td><strong>Review rating</strong></td>
<td>0.1367</td>
<td>0.1244</td>
<td>0.1244</td>
</tr>
<tr>
<td><strong>Superhost</strong></td>
<td>0.10</td>
<td>0.10</td>
<td>0.10</td>
</tr>
<tr>
<td><strong>Instant book</strong></td>
<td>-0.0196</td>
<td>-0.0177</td>
<td>-0.0177</td>
</tr>
<tr>
<td><strong>Minimum night</strong></td>
<td>-0.0083</td>
<td>-0.0081</td>
<td>-0.0081</td>
</tr>
<tr>
<td><strong>Room type control</strong></td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td><strong>Observations</strong></td>
<td>132,031</td>
<td>132,031</td>
<td>86,300</td>
</tr>
<tr>
<td><strong>Adjusted R^2</strong></td>
<td>0.2455</td>
<td>0.3216</td>
<td>0.5120</td>
</tr>
</tbody>
</table>

Note: Standard errors are clustered by neighborhood and reported in parentheses. There are 32 neighborhood dummies, 11 monthly dummies, controls for listing attributes not reported include: cancellation policy, No. of bedrooms, No. of bathrooms, No. of guests, maximum guests per stay, and listing count.

***, **, and * indicate significance at the 1%, 5%, and 10% levels.

Table 3: The effect of automated cancellation review on listing price.
### Table 4: Summary Statistics of Risk Index

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Std. Dev</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Whole Sample</td>
<td>1.69%</td>
<td>0.07</td>
<td>132,031</td>
</tr>
<tr>
<td>Entire Home Rental</td>
<td>1.82%</td>
<td>0.0</td>
<td>75,860</td>
</tr>
<tr>
<td>Private Room Rental</td>
<td>1.52%</td>
<td>0.0</td>
<td>64,457</td>
</tr>
</tbody>
</table>

**Note:** Whole sample includes both entire-home and private-room listings.
<table>
<thead>
<tr>
<th></th>
<th>Whole Sample</th>
<th>Entire Home Rental</th>
<th>Private Room Rental</th>
<th>Entire Home Rental</th>
<th>Private Room Rental</th>
</tr>
</thead>
<tbody>
<tr>
<td>Risk Index</td>
<td>0.9081</td>
<td>0.9968</td>
<td>1.4912</td>
<td>1.5257</td>
<td>0.6386</td>
</tr>
<tr>
<td>(0.088)</td>
<td>0.8061</td>
<td>0.8061</td>
<td>0.8061</td>
<td>0.8061</td>
<td>0.8061</td>
</tr>
<tr>
<td>Number of reviews</td>
<td>0.0036</td>
<td>0.0036</td>
<td>0.0036</td>
<td>0.0036</td>
<td>0.0036</td>
</tr>
<tr>
<td>(0.0001)</td>
<td>0.0002</td>
<td>0.0002</td>
<td>0.0002</td>
<td>0.0002</td>
<td>0.0002</td>
</tr>
<tr>
<td>Review rating</td>
<td>-0.2183</td>
<td>-0.2421</td>
<td>-0.3017</td>
<td>-0.3386</td>
<td>-0.598</td>
</tr>
<tr>
<td>(0.008)</td>
<td>0.009</td>
<td>0.012</td>
<td>0.012</td>
<td>0.014</td>
<td>0.014</td>
</tr>
<tr>
<td>Superhost</td>
<td>-0.0877</td>
<td>-0.1079</td>
<td>-0.1183</td>
<td>-0.127</td>
<td>-0.0148</td>
</tr>
<tr>
<td>(0.016)</td>
<td>0.016</td>
<td>0.016</td>
<td>0.016</td>
<td>0.020</td>
<td>0.027</td>
</tr>
<tr>
<td>Instant book</td>
<td>0.0191</td>
<td>0.0051</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(0.011)</td>
<td>0.012</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Minimum night</td>
<td>0.0027</td>
<td>0.0027</td>
<td>0.0027</td>
<td>0.0027</td>
<td>0.0027</td>
</tr>
<tr>
<td>(0.001)</td>
<td>0.001</td>
<td>0.001</td>
<td>0.001</td>
<td>0.001</td>
<td>0.001</td>
</tr>
</tbody>
</table>

Note: Standard errors are clustered by neighborhood and reported in parentheses. There are 32 neighborhood dummies, 11 monthly dummies, and listing attributes not reported include cancellation policy, No. of bedrooms, No. of bathrooms, No. of beds, maximum guests per stay, and listing counts. *** and ** indicate significance at the 1% and 10% levels, respectively. 

Table 5: The effect of constructed risk-index on listing vacancy.
Table 6: The effect of constructed risk-index on listing price

<table>
<thead>
<tr>
<th>Risk Index</th>
<th>Private Room Rental</th>
<th>Entire Home Rental</th>
<th>Whole Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>0.0001</td>
<td>0.0001</td>
<td>0.0001</td>
<td>0.0001</td>
</tr>
<tr>
<td>0.0002</td>
<td>0.0002</td>
<td>0.0002</td>
<td>0.0002</td>
</tr>
<tr>
<td>0.0003</td>
<td>0.0003</td>
<td>0.0003</td>
<td>0.0003</td>
</tr>
<tr>
<td>0.0004</td>
<td>0.0004</td>
<td>0.0004</td>
<td>0.0004</td>
</tr>
<tr>
<td>0.0005</td>
<td>0.0005</td>
<td>0.0005</td>
<td>0.0005</td>
</tr>
</tbody>
</table>

Notes: Standard errors are clustered by neighborhood and reported in parentheses. There are 32 neighborhood dummies, 11 monthly dummies, Controls for listing attributes not reported include cancellation policy, No. of bedrooms, No. of bathrooms, No. of guests and minimum night. Minimum night is included in the specifications. Significant levels: *** indicates significance at the 1% level, ** indicates significance at the 5% level, * indicates significance at the 10% level.
### Table 7: Controls for the number of competitors in listing vacancy regressions

<table>
<thead>
<tr>
<th>Controls</th>
<th>Yes</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neighborhood control</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Listing attributes Controls</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Controls</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Time FE</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Listing FE</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Neighborhood control</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Listing attributes Controls</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Controls</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Time FE</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Listing FE</td>
<td>No</td>
<td>Yes</td>
</tr>
</tbody>
</table>

#### Observations

<table>
<thead>
<tr>
<th>Observations</th>
<th>86,300</th>
<th>86,300</th>
<th>86,300</th>
<th>86,300</th>
<th>45,731</th>
<th>45,731</th>
<th>45,731</th>
<th>45,731</th>
</tr>
</thead>
</table>

#### Adjusted R^2

<table>
<thead>
<tr>
<th>Adjusted R^2</th>
<th>0.1899</th>
<th>0.2696</th>
<th>0.1896</th>
<th>0.3203</th>
<th>0.1831</th>
<th>0.2272</th>
<th>0.1819</th>
<th>0.2371</th>
</tr>
</thead>
</table>

#### Note:

- Standard errors are clustered by neighborhood and reported in parentheses.
- We also use the number of competitors in 0.3-mile radius and the results are similar.
- The risk index and the number of competitors are both mean-centered. Controls for listing attributes not reported include cancellation policy, No. of bedrooms, No. of bathrooms, No. of beds, maximum guests per stay, and listing counts.
- *** indicates significance at the 1%, 5%, and 10% levels.

---

#### Table: Controls for the number of competitors in listing vacancy regressions

<table>
<thead>
<tr>
<th>Variables</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
<th>(7)</th>
<th>(8)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cancellation reviews</td>
<td>0.0562**</td>
<td>0.0654***</td>
<td>0.0325*</td>
<td>0.0507*</td>
<td>0.007</td>
<td>0.107</td>
<td>0.0654***</td>
<td>0.0654***</td>
</tr>
<tr>
<td>Cancellation review -</td>
<td>-0.0032</td>
<td>-0.0069</td>
<td>-0.0118</td>
<td>-0.0159*</td>
<td>-0.007</td>
<td>-0.118</td>
<td>-0.0069</td>
<td>-0.0069</td>
</tr>
<tr>
<td>(Competition (0.1 mile) x Risk Index)</td>
<td>0.0009***</td>
<td>0.0005</td>
<td>0.0017**</td>
<td>0.0028***</td>
<td>0.0001</td>
<td>0.0003</td>
<td>0.0001</td>
<td>0.0001</td>
</tr>
<tr>
<td>(Competition (0.1 mile) x Risk Index)</td>
<td>1.3581**</td>
<td>2.0521***</td>
<td>0.8388**</td>
<td>0.5732***</td>
<td>0.0001</td>
<td>0.0003</td>
<td>0.0001</td>
<td>0.0001</td>
</tr>
<tr>
<td>(Competition (0.1 mile) x Risk Index)</td>
<td>1.5581***</td>
<td>2.0521***</td>
<td>0.8388**</td>
<td>0.5732***</td>
<td>0.0001</td>
<td>0.0003</td>
<td>0.0001</td>
<td>0.0001</td>
</tr>
<tr>
<td>(Competition (0.1 mile) x Risk Index)</td>
<td>0.0009***</td>
<td>0.0005</td>
<td>0.0017**</td>
<td>0.0028***</td>
<td>0.0001</td>
<td>0.0003</td>
<td>0.0001</td>
<td>0.0001</td>
</tr>
<tr>
<td>(Competition (0.1 mile) x Risk Index)</td>
<td>1.3581**</td>
<td>2.0521***</td>
<td>0.8388**</td>
<td>0.5732***</td>
<td>0.0001</td>
<td>0.0003</td>
<td>0.0001</td>
<td>0.0001</td>
</tr>
<tr>
<td>(Competition (0.1 mile) x Risk Index)</td>
<td>1.5581***</td>
<td>2.0521***</td>
<td>0.8388**</td>
<td>0.5732***</td>
<td>0.0001</td>
<td>0.0003</td>
<td>0.0001</td>
<td>0.0001</td>
</tr>
<tr>
<td>(Competition (0.1 mile) x Risk Index)</td>
<td>0.0009***</td>
<td>0.0005</td>
<td>0.0017**</td>
<td>0.0028***</td>
<td>0.0001</td>
<td>0.0003</td>
<td>0.0001</td>
<td>0.0001</td>
</tr>
<tr>
<td>(Competition (0.1 mile) x Risk Index)</td>
<td>1.3581**</td>
<td>2.0521***</td>
<td>0.8388**</td>
<td>0.5732***</td>
<td>0.0001</td>
<td>0.0003</td>
<td>0.0001</td>
<td>0.0001</td>
</tr>
<tr>
<td>(Competition (0.1 mile) x Risk Index)</td>
<td>1.5581***</td>
<td>2.0521***</td>
<td>0.8388**</td>
<td>0.5732***</td>
<td>0.0001</td>
<td>0.0003</td>
<td>0.0001</td>
<td>0.0001</td>
</tr>
<tr>
<td>(Competition (0.1 mile) x Risk Index)</td>
<td>0.0009***</td>
<td>0.0005</td>
<td>0.0017**</td>
<td>0.0028***</td>
<td>0.0001</td>
<td>0.0003</td>
<td>0.0001</td>
<td>0.0001</td>
</tr>
<tr>
<td>(Competition (0.1 mile) x Risk Index)</td>
<td>1.3581**</td>
<td>2.0521***</td>
<td>0.8388**</td>
<td>0.5732***</td>
<td>0.0001</td>
<td>0.0003</td>
<td>0.0001</td>
<td>0.0001</td>
</tr>
<tr>
<td>(Competition (0.1 mile) x Risk Index)</td>
<td>1.5581***</td>
<td>2.0521***</td>
<td>0.8388**</td>
<td>0.5732***</td>
<td>0.0001</td>
<td>0.0003</td>
<td>0.0001</td>
<td>0.0001</td>
</tr>
</tbody>
</table>

---
Table 8: Controls for the number of competitors in listing price regressions

<table>
<thead>
<tr>
<th></th>
<th>Entire home rental</th>
<th>Private room rental</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>0.003</td>
<td>0.003</td>
</tr>
<tr>
<td>(2)</td>
<td>0.3939</td>
<td>0.3939</td>
</tr>
<tr>
<td>(3)</td>
<td>0.4133</td>
<td>0.4133</td>
</tr>
<tr>
<td>(4)</td>
<td>0.4131</td>
<td>0.4131</td>
</tr>
<tr>
<td>(5)</td>
<td>0.0412</td>
<td>0.0412</td>
</tr>
<tr>
<td>(6)</td>
<td>0.0033</td>
<td>0.0033</td>
</tr>
<tr>
<td>(7)</td>
<td>0.0031</td>
<td>0.0031</td>
</tr>
<tr>
<td>(8)</td>
<td>0.0031</td>
<td>0.0031</td>
</tr>
</tbody>
</table>

Note: Standard errors are clustered by neighborhood and reported in parentheses. We also use the number of competitors in a 0.3-mile radius and the number of competitors in a 0.1-mile radius as dummy variables. The results are similar. The risk index and the number of competitors are both mean-centered. Controls for listing attributes not reported include cancellation policy, No. of bedrooms, No. of bathrooms, No. of beds, maximum guests per stay, and listing counts. ** and *** indicate significance at the 1%, 5%, and 10% levels, respectively.