Welfare Estimation in Peer-to-Peer Markets with Heterogeneous Agents: the Case of Airbnb

Abdollah Farhoodi *†

November 2, 2019

Please find the latest version here

Abstract

Peer-to-peer (P2P) markets allow small suppliers with limited capital to enter the markets that were traditionally occupied by large firms. This feature provides a potential decentralized distribution of opportunities. To investigate the distribution of welfare and opportunities among agents, I study the Airbnb short-term rental market as a successful P2P marketplace. I use a daily panel of Airbnb rentals in Chicago between August 2014 to April 2017, and apply an individual-level multinomial logit model to estimate the distribution of consumer and producer surplus among differentiated agents and over time. I show that properties in less advantaged neighborhoods benefit the least from having access to the Airbnb market; even though these properties feature lower competitive pressure and lower opportunity cost of renting. My results show a disproportionate concentration of welfare in neighborhoods with higher income, house price, and in upper-class areas. However, I show evidences on higher surplus of low-income property owners specially for those who live in high demand areas.

*Email: farhood2@illinois.edu; Address: Department of Economics, University of Illinois, Urbana, IL 61801

†I am indebted to Dan Bernhardt for his invaluable advise. I am grateful for the valuable guidance and discussion from David Albouy, George Deltas, Jorge Lemus, and Guillermo Marshall.

1 Introduction

Peer-to-peer (P2P) markets are growing markets where consumers and producers trade directly usually through an online platform. In the last decade, P2P markets have been expanding exponentially. Marketplaces such as Uber, Airbnb, Lending Club, and many others, facilitate consumers' access to transportation, accommodation, and financial loans; they also let small suppliers profit from their spare capitals. This feature suggests potential large welfare gains in P2P markets for small entrepreneurs, allowing for micro-entrepreneurship for people with limited access to capital and for more "democratization of opportunities" (Sundararajan, 2016).

The contribution of this paper is to study the welfare distribution in P2P markets, while accounting for the individual-level heterogeneities among both supply and demand sides of the market. Specifically, it addresses the question of whether opportunities are equally distributed among people who rent out their properties in the Airbnb short-term rental market. On one hand, low-income suppliers are expected to value the extra income more than their higher-income counterparts, and have lower opportunity cost of supplying their properties to the rental market. On the other hand, these properties are often located in low-income neighborhoods, while there is a higher demand for properties in neighborhoods with better amenities. This demand may shift the surplus toward suppliers whose properties are already in higher-income areas. I show that this second force is dominant in the Chicago Airbnb market; even though suppliers who live in higher-income neighborhoods have higher opportunity costs, they still benefit from higher welfare gains in the market.

Properties located in high demand neighborhoods earn higher total and average surpluses. However, I show evidences which suggest higher benefit of low income suppliers with-in the same neighborhoods. I show that, on average, the total surplus of property owners who value more the marginal dollar is higher. My results show that the Airbnb market is more beneficial for the low income property owners who live in high demand neighborhoods.

The data used in this paper come from a unique dataset with detailed information regarding Airbnb listings in Chicago over a roughly two-and-a-half-year period from August 2014 to April 2017 (AirDNA, 2017). The data contain around 9 million observations and include information about available Airbnb accommodation characteristics: the location, number of bedrooms, number of bathrooms, type of accommodation (entire place, private or shared), hosts' ratings, number of photos, and other information available to a potential renter through the Airbnb website. The dataset also includes information about the time of booking and the booking status of each apartment. I use this information to estimate the demand for each property.

The findings show that the average consumer surplus is around 20% of the paid price for accommodation, and has seasonal variations with higher surplus during summer times. Similarly, average producer surplus is around 65% of hosts' revenue and is more concentrated in downtown Chicago, neighborhoods with higher income, and upper-class neighborhoods. The reason is that people are willing to pay more for the available amenities in upper-class neighborhoods. My estimations show that on average people are willing to pay up to $130 more for the same property located in downtown compared to some areas in the southern part of Chicago. Consequently, even though the opportunity cost of providing an Airbnb listing is lower for properties located in less advantaged neighborhoods of Chicago, they cannot gain as much surplus as properties located in more touristy and business-oriented areas. For instance, share of total surplus in my estimations has been roughly 22% for the central side of Chicago, while this area only includes about 1% share of the total population.

Despite the lower average and total surplus of properties located in high demand neighborhoods, I show evidences that low income property owners benefited the most from having access to the Airbnb market. I show that property owners who value the marginal money 1% more, on average, earn 0.17% more surplus. This effect is higher for properties that are located in the Central Chicago where demand is the highest. In this side Chicago, 1% higher value of money is associated with 0.32% higher total surplus.

The study of distribution of welfare and opportunities among agents in P2P markets requires accounting for the existing heterogeneities that generate differences in opportunity costs on the supply side and preferences on the demand side of the market, which I address in this paper. Measuring the welfare distribution while accounting for differences among agents is challenging. Unlike traditional markets where big firms supply masses of homogeneous goods, in P2P markets agents are usually small, they have limited capacities, and trade differentiated products in a decentralized market with various prices that change rapidly with variations in the market. This decentralized structure pose a challenge for using classic aggregated methods of supply and demand estimation for welfare analysis.

In this paper I tackle the problem of welfare analysis in P2P markets with heterogeneous agents. I use daily panel-data of transactions in Chicago for Airbnb bookings, to study the welfare of both guests and hosts in the market. I estimate time varying individual-level consumer and producer surpluses. My estimations provide better understanding of the distribution of welfare in the Airbnb market. To the best of my knowledge, this is the first study that uses this level of aggregation for both consumer and producer estimation of surplus.

I estimate the the demand for each property (listing) in Airbnb market using an individual-level multinomial logit model (McFadden, 1977). Bayer et al. (2007) estimate a similar individual-level demand for each house in the housing market. Compared to their paper, my model adds four features. First, my data includes repeated renting of listings, so I estimate the model with a panel-data and add a time dimension to the estimations. Second, to study the distribution of welfare among suppliers, I add a supply side with strategic pricing to the model. Third, around half of listings in the Airbnb market are not booked in a day, and many buyers choose a hotel or other options instead of booking an Airbnb. So, I allow for vacant units and the possibility of choosing an outside option. Fourth, listings that are booked are not available to the potential buyers, so I allow for variations in the choice sets depending on the time that guests choose to book the place. These are important features that are essential for modeling agents in Airbnb. The method that I used is applicable to housing the rental market, or with some adjustments to other P2P markets.

On the supply side, I model hosts in a static, monopolistic competition environment with heterogeneity in opportunity costs and competition in prices. The supply side model solves the relation between the producer surplus and the demand for each listing. I apply the supply side model to find the producer surplus and the opportunity cost of renting a differentiated property in the market.

My estimation method applies an instrumental variable (IV) to identify the price elasticity of utility. The instrument is BLP IV which is a commonly used IV in the literature (Bayer et al., 2007), (Berry et al., 1995). It includes the average characteristics of other listings located far from a specific property. The identification assumption is that characteristics of far enough listings do not directly affect the guests' utilities for a property. The identification assumption may be violated if unobservable demand shocks are correlated with a specific characteristic of listings such as number of bedrooms. As a robustness test, I check the results with a second IV that is created based on share of days that a listing was blocked. A listing is either available, reserved or blocked in the data, and I use the share of blocked days as an IV for price. The identification assumption is that hosts respond to demand shocks only by changing their prices, and blocking depends on personal schedule of hosts. In other words the assumption is that the personal schedule of hosts is not correlated with demand shocks. Both IVs conclude very similar estimation of parameters and support the accuracy of the identification assumptions.

There are few recent papers on welfare estimation in P2P markets. In a closely related paper, Farronato and Fradkin (2017) estimate welfare in the Airbnb market. They show that welfare gains are higher in locations and at times in which hotels are capacity constrained. For their estimation, they use the aggregate market shares of Airbnb and hotels to apply a discrete choice model of listings' demand and supply. Unlike their paper, I apply individual-level variations in prices to estimate distribution of welfare and account for heterogeneities among hosts. So, the choice set of a potential guest in my model is all available listings at the time of booking while potential guests in their paper choose between hotel and Airbnb in aggregate.

Likewise, Lam and Liu (2018) estimate a real-time demand for Uber and Lyft ride sharing. Similar to my paper, their paper accounts for heterogeneity in surpluses over time. In the home-sharing market, in addition to time, heterogeneities among hosts' characteristics and locations play important roles in shaping consumer and producer surpluses. The current study takes these differences into account. Similarly, Cohen et al.

(2016) use the Uber surge pricing algorithm to estimate the demand elasticity along several points of the aggregate demand curve, and to calculate total consumer surplus in the market. For further reading, Einav et al. (2016) provide a good review of the literature on P2P markets. My paper is the first paper that estimates the distribution welfare among differentiated agents in a P2P market.

There are papers that studied the effect of P2P markets on other markets. For instance, Fraiberger et al. (2015) by calibrating their model conclude reduction in car ownership reduces as response to growth in the P2P car rental market. Zervas et al. (2017) study the effect of Airbnb on hotel industry. Sheppard et al. (2016) and Barron et al. (2018) study the effect of Airbnb on housing market. Unlike these papers, I focus on the welfare estimation and its distribution in the Airbnb market and I do not estimate its externally on other markets.

This paper is structured as follows: Section 2 presents the data, summary statistics, and Airbnb marketplace trends. Section 3 introduces the model. Section 4 discusses the estimation and identification methodologies. Section 5 discusses the estimation results and distribution of welfare in the market. Section 6 concludes.

2 Data and Summary Statistics

This paper employs micro-level data of property listings on Airbnb platform in city of Chicago with around 9 million observations (AirDNA, 2017). The data cover the period between August 2014 and April 2017 and include an unbalanced panel of information about prices, status of listings (available, reserved, blocked), and time of bookings. The data also contain information about location, and the characteristics of listings that are available on the Airbnb website. These characteristics consist of: the number of bedrooms, number of bathrooms, number of photos, types of listings (entire place, private or shared room), amount of deposit, whether the listing is for a "super host", business ready (business ready listings provide specific accommodation standards, such as access to internet and late entry), etc. The analysis of this paper is restricted to Chicago, but it could easily be conducted in other markets where data are available.

The market includes 23,485 distinct, active properties in the time span of the data.

It covers most neighborhoods in the city. A median property is booked 65 times and earn about $6800 in 2.5 years. These properties earned a total of $195 million in the time period of the data from Aug 2014 to Apr 2017. Total revenue is not uniformly distributed in the market. Figure 1 shows that there are some block-groups that earn more than $3 million in just 2.5 year of the data while there are some other neighborhoods that has not benefited much from the Airbnb platform.

The Airbnb market is highly dynamic and its scope has has increased exponentially over time. Figure 2 shows that total monthly revenue has a strong seasonal variation and almost doubled every year. It is important to take into account this variation in the scope of market for welfare analysis. Figure 3 zooms in to the last month of the data and shows large high frequency daily variations, peaking over the weekends, in total revenue.

The market shows a seasonal variation in average prices. Figure 4 plots the trends of average price of booked and all listings in the market. Both trends follow seasonal variations with higher prices over the summer. An average booked property costs around $120 while the average overall costs $140 per day.

These trends show high variations in the market over time and across geography. They also point out on the importance of heterogeneity in listings' characteristics (such as the number of bedrooms and the listing types) for analysis of the Airbnb market. The following section shows how these heterogeneities are taken into account for welfare estimation.

3 Model

In this section, I introduce a static model of supply and demand for home sharing. On the demand side, guests face a discrete set of options and are able to choose their first best choices. Guests are price takers, and they maximize their utilities based on observed prices and listings' attributes. On the other hand, suppliers compete in a monopolistic competition environment with competition in prices. Suppliers' decisions are based on their profit function and probability that their listing is booked.

I first start with the demand side of the model and guests' decision about their choices of listing to rent in each day. The utility function is based on Berry et al. (1995) and

Bayer et al. (2007). Guests choose among a discrete set of available listings at the time of booking to maximize their indirect utility. The indirect utility of guest i from choosing listing h is as follows:

$$u_h^i = \delta_{ht} + \epsilon_h^i = \alpha P_{ht} + \xi_h + \gamma_t + \epsilon_h^i \tag{1}$$

where,

$$\xi_h = \beta X_h + \gamma_l + \nu_h,$$

δ_{ht} is the mean utility for listing h at time t, and ϵ_h^i is the individual specific utility for the listing. I follow the discrete choice literature and assume that the error term ϵ_h^i is independently and identically distributed and draws from an extreme value type-I distribution function. In the absence of more information about characteristics of guests, I assume the utility parameters are constant, and the only heterogeneity among consumers comes from ϵ_h^i.

The mean utility for each listing at a time, δ_{ht}, is divided into three parts, αP_{ht} is dis-utility from price, ξ_h is listing specific utility and γ_t is time specific utility. γ_t is a day fixed effect and captures the average utility from booking an Airbnb listing in a specific day. The listing specific utility, ξ_h, includes the utility from listing's observable characteristics X_h such as number of bedrooms, bathrooms, reviews, location fixed effect, γ_l, and a control, ν_h, for un-observable characteristics. ν_h captures the utility from un-observable characteristics such as utility from the exact location of the listing or information that is available to guests by looking at the photos of the listing but is not available to the econometrician.

Each potential guest chooses among a discrete set of available listings at the time of booking to maximize her utility. Thus, the probability of choosing a listings depends on utilities from all available options at the time of booking. Next, I normalize the utility from the outside option to zero to find the probability of choosing each listing. One can think of the outside option as the option of booking a hotel instead of Airbnb or even staying home. Using this normalization, s_{iht} is the probability that the potential guest i choose listing h at time t:

$$s_{ht}^i = \frac{\exp(\delta_{ht})}{1 + \sum_{k \in M_t^i} \exp(\delta_{kt})} \tag{2}$$

where, M_t^i is set of all available Airbnb options to the guest at the time of booking. This set has a superscript i and varies for guests depending on their time of booking. Here, I allow for variations in the choice sets depending on the guests' time of booking. The variable choice set takes into account the cases where a listing is already booked and is not in other guests' choice set anymore.

The variable choice set is an important feature of model that should be taken into account when suppliers have limited capacity. In many P2P markets and also in housing market each supplier only has one good or service to sell. For demand estimation, an already sold good should not appear in the buyers' choice set. The information about guests time of booking allows me include the variable choice in my estimations. I observe the time that each listing is booked and I exclude an early booked listing from the choice set of a guest who choose his option later.

Next, I model the profit maximization of hosts. On the supply side, hosts or suppliers compete in a static monopolistic competing environment with competition in prices. They choose their prices to maximize their expected profit as follows:

$$\max_{P_{jt}}(P_{jt} - c_{jt})Pr_{jt} \implies P_{jt} = c_{jt} - \frac{Pr_{jt}}{dPr_{jt}/dP_{jt}} \tag{3}$$

where, P_{jt} is the price of listing j at time t, Pr_{jt} is the probability that the listing is booked, and c_{jt} is the opportunity cost of renting. This cost may include the opportunity cost of not using the unit or the risk of renting a room to a stranger. It varies with both observable and un-observable characteristics of the unit, and is estimated as the residual of difference between the price and producer surplus. The opportunity costs are not observable. However, one can estimate the probability of booking and recover marginal costs and producer surplus through their relation with prices.

As is shown in equation (9), the solution to the hosts profit maximization concludes that hosts set prices equal to their opportunity cost plus a mark-up, $PS_j = -\frac{Pr_{jt}}{dPr_{jt}/dP_{jt}}$. This mark-up is hosts' surplus from renting their listing at a time. I estimate the probability of booking each listing using the demand side estimation of mean utilities and calculate this mark-up as follows:

$$Pr_{jt} = 1 - \prod_{i \in N_t} (1 - s^i_{jt}) \implies PS_{jt} = \frac{1 - \prod_{i \in N_t}(1 - s^i_{jt})}{\alpha \sum_{i \in N_t} s^i_{jt} \prod_{i \in N_t}(1 - s^i_{jt})} \tag{4}$$

where, N_t is the market size or set of all potential guests in the market at time t. The first term in equation (4) shows the relation between the probability that listing j is booked and the probabilities that each potential guest chooses the listing. The product term is the probability that nobody in the market chooses listing j. The second term shows the relation between the producer surplus and the probabilities of choosing each listing which that I estimate from the demand side of the model.

4 Estimation Methodology

Two steps estimate consumer and producer surplus. First step estimates the mean utilities for each listing and calculates the probability that each potential guest chooses each listing. The second step calculates the probability that each listing is booked, producer surplus and opportunity costs.

I follow Bayer et al. (2007) in the first step estimation of demand parameters. However, there are four important differences here compare to their estimation method. First, I add a time dimension to the model. In their paper they estimate a cross sectional model of demand for housing options, but in the Airbnb, the market is highly dynamic and varies every day. Second, I allow for choosing an outside option and vacant units. In their paper, everybody who wants to buy a house will buy a house, market clears and every seller sell her house. However, in the Airbnb market a big share of listings are not booked in a day and many potential guests choose a hotel or other option besides Airbnb. So, it is important to account for possibility of choosing an outside option in addition to having excess supply in the market.

The third difference here is that I account for variability in the choice sets based on guests' time of decision (time of booking). Guests choose the best option that maximizes their utility at the time of booking. If a listing is already booked it does not show up in the guests' choice set. This is an important feature in many peer to peer markets where there are suppliers with limited capacities and I include in my estimations. Fourth, I estimate producer surplus and allow for strategic pricing over time which is absent the in the static model of Bayer et al. (2007). All these extensions are essential in the Airbnb market where there is a high dynamic in the market size, choice sets, and pricing.

Next, I estimate the utility parameters by maximizing the likelihood of observed choices. The maximum likelihood function is as follows:

$$ll = \sum_t \sum_i \sum_h I^i_{ht} \log(s^i_{ht}) + \sum_t \sum_i I^i_{ot} \log(s^i_{ot}) \tag{5}$$

were, I^i_{ht} is one if potential guest i chooses listing h and I^i_{ot} is one if she chooses the outside option. They are zero otherwise. Here, s^i_{ht}, and s^i_{ot} are probabilities that guest i chooses listing h and the outside option, respectively. The first step is to find utility parameters $(\alpha, \xi_h, \gamma_t)$ that maximize the log-likelihood function, ll. Here, I use the first order conditions to search over ξ_h and γ_t, and apply an instrumental variable for price to find α for a given ξ_h, γ_t.

The first order derivatives of log-likelihood provide intuitive conditions to pin down the utility parameters:

$$\begin{aligned}\frac{\partial ll}{\partial \xi_h} = 0 &\implies N_h = \sum_t \sum_i s^i_{ht} \\ \frac{\partial ll}{\partial \gamma^t} = 0 &\implies N_{ot} = \sum_i s^i_{ot}\end{aligned} \tag{6}$$

where, N_h, N_{ot} are total number of times that listing h is booked and total number of potential guests who choose an outside option at time t, respectively. The first condition shows that sum of probabilities of booking a listings whenever the listing is available equals total number of times that the listing is booked. This condition pins down the property specific mean utility (ξ_h). If a listing experience more bookings over its lifetime the model predicts higher utility for it. The second condition concludes that num of probabilities of choosing outside option over all potential guests in a day equals total number of potential guests who choose an outside option. If a bigger share of people choose the outside option the model predicts lower time specific utility (γ_t) for all listings in that day.

I use these two first order maximization conditions to form the following contraction mapping:

$$\delta_{ht}^{T+1} = \delta_{ht}^{T} + \left(log(N_h) - log(\sum_t \sum_i \hat{s}_{ht}^i) \right) - \left(log(N_{ot}) - log(\sum_i \hat{s}_{ot}^i) \right). \tag{7}$$

Equation (7) is a contraction mapping and converges to a single solution. The fixed point is where both terms in parenthesis are zero. This is where both first order maximization condition in equation (6) are satisfied. For a given α, this contraction mapping converges to unique property and time mean utility fixed effects (ξ_h, γ_t). To estimate α, I apply instrumental variable regression and estimate α for given property and time fixed effects. Thus, for a given α, I estimate ξ_h and γ_t using the contraction mapping, and then re-estimate α. This process is repeated until it converges to a solution for all utility parameters $(\alpha, \xi_h, \gamma_t)$.

4.1 Instrument for Price

An instrumental variable regression approach is used to identify the slope of demand and, in turn, consumer and producer surpluses. I use two different instruments for price. The results are very similar using either of these instrument.

Estimation of coefficient of price requires using an instrument due to the possible correlation between the price and unobserved characteristics of property such as its exact location. This correlation may lead to a bias toward zero estimation of α. That is because most probably an un-observable characteristics correlated with the price and utility in the same direction. For instance, a nice apartment in a good location has provide higher utility and more demand lead the supplier to increase its prices. This potential positive correlation between utility and price conclude a less negative bias estimation of α.

The first instrument that I use is BLP instrument (Berry et al., 1995; Nevo, 2001). This is a common instrument in the literature and is used in multiple similar studies (Bayer et al., 2007). The instrument is average characteristics of competitors far from a listing. The intuition is that characteristics of other listings do not directly affect the guests' utility. So, the identification assumption is that after controlling for characteristics of other listings inside the neighborhood, characteristics of far enough properties affect the utility only through prices. Equation (1) confirms this assumption and shows that the characteristics of rival options does not appear in the guests' utility function.

The relevance is through the relation between price and utility from all options in the market. Hosts set their prices considering the demand for their property. The demand for each listing is a function of utility from all options in the market. Thus, it is a function of other listings characteristics. Equation (4) shows the relation between price and utility from other options in the market and confirms the relevance condition of the BLP instrument.

The second instruments that I use to check the robustness of results is share of days that a listing has been blocked. In the data, a listing is either available, booked, or blocked. Hosts block their property if they are not willing to rent it for a specific day. This decision depends on the personal schedule of the host. The identification assumption is that the personal schedule is not correlated with unobservable demand shocks. The intuition is that if a host expects a low demand shock, she sets a minimum price for her listing and does not respond to this shock with blocking her property. The identification assumption can be rejected if the host changes her schedule in respond to a demand shock. For instance, if she postpones blocking and using her place for personal purposes when demand is high. In the next section, I will show that the results for both instruments are very similar. Thus, I believe that personal schedule are often fixed and hosts' blocking decision is not correlated with demand shocks.

4.2 Definition of Market

The first information that is asked in the Airbnb website is destination, time of travel, and number of guests. I follow the same rule and limit the choice sets to all available listings in a day with specific number of guests in Chicago. I assume that people choose among listings in a same day. This assumption ignores the possibility of choosing among listings that are available in different days. Given that I model the decisions in a static settings, Airbnb in each day is separate market in my estimations.

I also assume that people search for listings with specific number of guests. I divide the listings to five categories of 1, 2, 3-4, 5-6, and more than 6 maximum guests. So, a property with capacity of one guest is not in the choice set of a person who is looking for a place with two or more guests. Similarly, listing with higher capacities are not in the choice set of those who are looking for places with few maximum guests. Therefore, potential guests choose among all available listings in a day with specific capacity in my estimations.

4.3 Market Size

The estimation requires information about number of potential buyers int the market (N_t). Discrete choice literature usually rely on an assumption about the size of market. For instance, Berry et al. (1995) assume every household in the U.S. is a potential buyer in the car market. In the context of housing market, Bayer et al. (2007) assume that there is no excess demand and the size of market is the same as total number of sold houses in a year. In the Airbnb market, Farronato and Fradkin (2017) assume the market size is three times total number of bookings in the corresponding month a year before. The assumption about market size affects demand estimation and the probability of that each listing is booked. As is shown in Figure 4 and discussed in the summary statistics section, there is an exponential growth and seasonal variations in the total transactions in the Airbnb market. A proper choice of market size should be able to show these variations in the market.

In the Airbnb market ideally, I need information about all potential guests who search through Airbnb website for booking in a specific day as measure of market size. In the absence of this information, I use Google searches trends for "Airbnb Chicago" as a proxy for total number of potential guests. Google reports monthly index (between 0 to 100) of total searches for a key-word. I scale this index to have an average booking rate of 60% in the first three month of the data. This scaling coefficient comes from Fradkin (2017). Using a sample data about search activities of users in an Airbnb market between September 2013 and September 2014, he shows that around 40% of users who send inquiries through Airbnb website do not book a listing and choose an outside option. I use this result to scale the index for Google search trends. Figure 5 shows the trends of calculated number of potential guest and total number of booked listings. It confirms that the scaled index captures the growth and seasonal variations in the market.

One complication is that using the scaled index of Google trends provides monthly and city-level size of the market. However, as discussed in the section 4.2 my estimation requires information about potential guests in a day looking for listings with a specific capacity. In order, find day-capacity-level size of the market, I interpolate the scaled measure of overall market size from Google trends. To do so, I multiply the scaled index by share of booked listings with specific capacity times share of booked listings in each day as is shown in Appendix B. This interpolation provides the desired level of market size.

5 Results

In this section, I start with the results of mean utility estimations. Then, I show the result of instrumental variable regression and utility for different characteristics of a listing. At the end, I estimate consumer and producer surplus, cost of renting, and their distribution in the market.

5.1 Parameter Estimates

Table 1 shows the estimation results of property and time specific mean utilities and the overall estimated mean utility in dollar terms. The mean utility estimates are all negative with an average of $-294. The negative sign is due to the large share of the outside option in the market.

Next, I find the willingness to pay for each characteristics of a listing by conducting an IV regression of mean utilities on price and the listing's characteristics. Table 2 shows the estimates of mean price coefficient (α) using the BLP-IV regressions. The coefficient of an OLS regression is -0.014. This coefficient is biased toward zero. This is due to the fact that hosts set higher prices for listings with higher utilities. The mean price coefficient is -0.026 and is higher in absolute value than the OLS estimate. I have a relatively strong instrument with first stage F-statistics (excluding the instrument) of around 14. The estimated mean price coefficient is consistent with the estimates of Farronato and

Fradkin (2017) for Airbnb market.

The IV-regression estimations with BLP-IV are robust to including characteristics of listings to the regression. BLP-IV is a widely used IV in the literature. However, the identification assumption may be violated if listings' characteristics are correlated with demand shocks. I apply a totally different IV for price based on share of days that a listings is blocked to check the robustness of the results. The second IV results are shown in table 3. The second IV is even more strong than BLP-IV with first stage F-statistics (excluding instrument) of 86. The estimates are very similar to estimations of BLP-IV regressions. Even though, one can think of scenarios to reject the identification assumption of both IVs, as discussed in section 4.1, finding consistent results with two different IVs is promising.

Table 4 shows the full regression coefficients. In this table each coefficient is the willingness to pay for a specific characteristics of the property. To find the willingness to pays in dollar values, I divide the coefficients by the coefficient of price. Using the Blocked days-IV regression, the willingness to pay for an extra bedroom and bathroom are \$32 and \$29, respectively. Similarly, Figure 6 shows the willingness to pay for different neighborhoods of Chicago. An average guest is willing to pay more than \$115 for the same property in Downtown Chicago compared to some neighborhoods in the south side. This Figure shows the importance of location in people's decision for booking an Airbnb listing. This results are intuitive since most touristic attractions, and business activities in Chicago are concentrated in Downtown. It shows that people are willing to pay less to stay in properties located in southern parts of Chicago, where the average income is lower compared to the Downtown. This result suggests a demand side source of unequal distribution of welfare across different neighborhoods in the market. I will study the distribution of welfare in more details in the next section.

5.2 Welfare Estimates

This section estimates consumer and producer surplus in terms of compensating variations. So I ask how much a guest should be compensated in order to maintain the same level of utility if she is banned from booking an Airbnb listing in Chicago.

The compensating variation in dollars for consumers is:

$$CS^i = \frac{1}{\alpha} \log \left(1 + \sum_{j \in M_t^i} \exp(\delta_j) \right). \tag{8}$$

Equation (8) is the difference between the expected maximum utility of guest i in the market and her utility when she can only use the outside option (Train, 2009). This difference is a measure of consumer surplus in the sense of compensating variation. In the above equation δ_j is the mean utility for listing j. Since the only source of heterogeneity in the consumer side is guest specific utility shock (ϵ_j^i), mean utility is assumed to be the same across all guests in the market. As discussed in section 4, my estimation method is easily applicable to a data-set with information about guests. This information could potentially provide heterogeneities across mean utilities which I do not account for in the model. However, it should be considered that M_j^i varies across guests depending on the time of making decision in the market.

Similarly, on the producer side, I ask how much a host in Chicago should be compensated in order to maintain the same level of profit if she is banned from renting her property through Airbnb. I use equation (4) to estimate producer surplus for each host in the market.

I estimate the producer surplus, assuming that agents are rational and perform strategic pricing with full information about the demand. In practice, some agents do not respond to variations in demand and do not change their prices as expected. This is a potential source of problem in most structural estimation. However, I believe that it is still reasonable to rely on average estimates of surpluses following the rationality assumption. Studying agents pricing and how it defers from the pricing of rational agents with full information, and whether there is learning in market are interesting topics for future researches.

The rationality assumption affects my estimations of producer surplus. In my estimations, around 10% of producer surplus estimates are higher than price. This means around 10% of estimates of opportunity costs are negative. To address the estimation problem, I truncate the marginal cost estimates of less than zero to zero. The overall patterns and estimations are not affected, if I winsorize the estimates or if I drop the negative estimates of opportunity cost from the data.

5.2.1 Time Variations

Total estimated consumer surplus in the time span of my data (around 2.5 years) is $21 million. This is equivalent to an average consumer surplus of $15 per day or 20% of each dollar spent. Producer surplus three times higher that consumer surplus with total of $67 million in 2.5 years. The average producer surplus is $46 per day or 65% of each dollar spent. One potential reason to have a higher share of producer surplus than consumer surplus maybe because consumers have easy access to an outside option such as Hotel. Contrarily, the outside option for hosts is not using their property for personal use. They loose money if they do not rent their property on Airbnb. This makes them to have a larger share of surplus than consumers in the market. It should be considered that hosts face the cost of renting only when someone books their property. So, I assume that properties in the Airbnb market are out of long-term rental market and I do not consider the opportunity cost of long-term rental in my estimations. A good extension to my estimations and a work in progress is adding a fixed cost to the model that captures the long-term opportunity cost of participating in Airbnb market.

Figure 7 shows the trends of estimated consumer and producer surplus as percentage of price. They both follow seasonal variations with higher share of surplus over the summer and high demand seasons. Even though, cost of entry in P2P markets is much lower than traditional markets, both seasonal and slightly upward trends of producer surplus suggest that the supply side does not perfectly adjust with the demand. They suggests that entry is not cost-less, and in contrast to a perfect competition market, hosts gain more market power when demand is high.

5.2.2 Geographic Variations

An important question is how welfare is distributed across neighborhoods and whether welfare is dis proportionally distributed across different areas. The answer to this question provides insight about who is benefiting more from the short-term rental market of Airbnb.

Figure 8 displays the distribution of total producer surplus across different neighborhoods in Chicago. This map shows how much each neighborhood has benefited from Airbnb in around 2.5 years of the data. It is apparent that total surplus is disproportionately distributed across neighborhoods. There are some neighborhoods mostly near downtown and northern parts of Chicago with surplus of more than $2 million while there are neighborhoods mostly in southern parts of Chicago with surplus of less than $50 thousand. Considering this disproportional distribution of welfare is important for any regulation of the Airbnb market.

This distribution of surplus is partly due to different numbers of bookings in different neighborhoods. There is less demand for booking an Airbnb in less touristy areas or neighborhoods further from main attractions and there are fewer active listings in these areas. Figure 10 shows the geographic distribution of average surplus per booking. Again, properties in the downtown and northern parts of Chicago earn more surplus through Airbnb per booking. This is mainly due to the more willingness to pay for these properties.

Figure 9 shows the distribution of estimated hosts' opportunity costs of renting their properties. Opportunity cost measures the risk of renting, cost of not using for personal use, and any potential loss by renting out the property. It also includes the value of money for the host. The higher hosts value the money, the lower is the estimated opportunity of renting. Higher opportunity of renting properties in downtown and northern Chicago is due to more available options for these properties. It also shows that neighborhoods in south side Chicago value more the gains from Airbnb.

Comparing the map of opportunity costs in Figure 9 and the map of total surplus in Figure 8, shows that even though the opportunity cost of providing accommodation services is lower in southern Chicago, surplus is concentrated in downtown and northern areas. This is mainly due to higher demand for more touristy and business centered areas which moves away the potential surplus from disadvantage neighborhoods towards more expensive areas.

Table 5 shows the distribution of surplus in 9 sides of Chicago. It shows how much each side would have loss if Airbnb was banned is the city. North side, followed by the west and central sides of Chicago has benefited the the most from access to Airbnb. The third and fourth columns in this table show the share of surplus and population of the side from total. As is seen, the share of producer surplus is not distributed with the share of population. For instance, the central Chicago earned 30% of total surplus while it includes only 1% of the total population.

In addition to concentration of total welfare in upper-class neighborhoods, properties in these areas earn more surplus per booking. This shows that the higher willingness to pay for accommodation in these neighborhoods dominates the larger opportunity cost of properties in these areas. Figure 10 shows the distribution of average producer surplus per booking and suggests higher market power for properties in the central Chicago. Figure 11 shows the correlation between the median income and the average surplus per booking across neighborhoods. This figure shows a strong positive correlation between the income of neighborhood and the producer surplus per booking. In fact, a 1% increase in the median income associated with around 0.2% higher producer surplus per booking.

5.2.3 Property-level Variations

This section studies the distribution of surplus across properties in the market. The results of the last section show concentration of welfare in more advantaged neighborhoods. This distribution is mostly demand driven and benefits areas where people are willing to pay more to access the existing amenities. However, it is important to find this distribution across those property owners who rent their properties in the Airbnb market.

Property owners who live in the same neighborhoods may have different characteristics and valuation of money. I show that even though the surplus is concentrated in advantaged neighborhoods, on average, property owners with higher valuation of money earn more surplus compared to others in each neighborhood. This results suggests that Airbnb is helping more the property owners with lower incomes and those with higher value for a marginal dollar. In the absence of detailed information about the hosts and their income, I use my estimations for opportunity costs as a measure for property owners' value for an extra dollar. Following, I explain the details.

My estimations allows for indirect estimation of opportunity cost of renting out properties. I estimate opportunity costs as the residual of the the difference between price and the estimated producer surplus as shown in equation (9). The following equation shows a modification of equation (9) where hosts have different valuation of money:

$$\max_{P_{jt}}(\tau_j P_{jt} - c_{jt})Pr_{jt} \implies P_{jt} = \frac{1}{\tau_j}c_{jt} - \frac{Pr_{jt}}{dPr_{jt}/dP_{jt}}. \tag{9}$$

Where, τ_j shows how much host j values an extra dollar. This parameter could be a function of income or implicit wage/time cost. The different valuations of money shows up in my estimation of opportunity costs. In fact, one should expect to find lower estimates of opportunity costs in low-income neighborhoods and where people value more the extra income (have higher τ_j). Figure 12 confirms the positive correlation of estimated opportunity costs and the median income across neighborhoods. Part of this correlation is due to correlation of opportunity cost of renting out a unit and rental price of the unit in long-term rental market. Figure 13 illustrates the strong positive correlation of opportunity costs and the average rents across neighborhoods.

To find the relationship between the property owners surplus and marginal value of dollar I regress the surplus on my estimation of opportunity cost for each property and control for neighborhood fixed effect. Controlling for the fixed effects captures the average geographic differences in the variations in the opportunity costs through different rental values. Thus, the estimated effect is most probably due to variations in how much each owner values the extra money. The first two columns of Table 6 show the negative correlation of producer surplus and opportunity costs. It shows that a 1% decrease in the opportunity cost is associated with 0.17% increase in the property owners total surplus. This results suggest that property owners with higher value for money benefit more from having access to the Airbnb.

This result is more apparent by looking at the subset of owners who are located in high demand areas or those who use Airbnb as an important source of income. Column (3) in the table 6 shows the correlation between property owners surplus and their average opportunity cost in the Central side of Chicago. In this location, a 1% decrease in the opportunity cost is associated with 0.31% increase in the property owners surpluses. Column (4) in this table looks at the share of days that a listing was blocked as a measure of importance of extra dollars for the property owner. Similar to properties owners that have lower opportunity cost, those who block more often earn lower total surplus. These findings suggest that having access to the Airbnb market is more beneficial for property owners with higher value of marginal dollar (low-income owners). Having access to the Airbnb market is specifically beneficial for the low income owners who live in high demand areas.

6 Summary

In this paper, I introduce a framework to study welfare distribution in P2P markets in which agents trade highly differentiated goods. In these markets, accounting for heterogeneity over time, location, and certain characteristics is important for welfare analysis. These factors play key roles in shaping consumer and producer surpluses. Among them are, for example, the time and location of services in ride-sharing markets the risk and amount of loans in online funding markets; and the time, location, and characteristics of listings in home-sharing markets.

This paper utilizes micro-level data of Airbnb along with individual-level multinomial logit model to account for detailed heterogeneity among agents. The findings show, on average, a consumer surplus of 20%, and a producer surplus of 65% for each dollar spent in the Chicago Airbnb market. These surpluses are not distributed homogeneously across various groups of listings, and in different times or locations.

Estimated average producer and consumer surpluses follow the seasonal variations in prices. The producer surpluses are disproportionately concentrated in high-income and upper-class neighborhoods. This concentration is mostly demand driven, and is due to higher willingness to pay for areas with more attractive amenities. My results show that people are willing to pay up to \$130 more for accommodation in the downtown areas compared to some neighborhoods mostly in southern parts of Chicago. Even though, surplus is concentrated in upper-class neighborhoods, I show evidences of higher surplus for property owners who value the marginal dollar more. My results suggests the high value of having access to the Airbnb market for low-income owners, specifically for those who live in areas with higher demand for short-term accommodation.

My estimation methodology is applicable to other P2P markets where there is a high dynamic in the market, and agents with limited capacities trade in a decentralized market. It also advances the existing methodology for demand estimation in the housing market. It allows for working with repeated sales and panel-data, and accounts for the seasonal variations in the demand, possible vacancy, and considers variable choice set of buyers in the estimation. These are important features of the housing sales or rentals markets and my estimation methodology considers these features.

References

Barron, K., Kung, E. and Proserpio, D. (2018), The sharing economy and housing affordability: Evidence from airbnb, *in* 'Proceedings of the 2018 ACM Conference on Economics and Computation', ACM, pp. 5–5.

Bayer, P., Ferreira, F. and McMillan, R. (2007), 'A unified framework for measuring preferences for schools and neighborhoods', *Journal of political economy* **115**(4), 588–638.

Berry, S., Levinsohn, J. and Pakes, A. (1995), 'Automobile prices in market equilibrium', *Econometrica: Journal of the Econometric Society* pp. 841–890.

Cohen, P., Hahn, R., Hall, J., Levitt, S. and Metcalfe, R. (2016), Using big data to estimate consumer surplus: The case of uber, Technical report, National Bureau of Economic Research.

Einav, L., Farronato, C. and Levin, J. (2016), 'Peer-to-peer markets', *Annual Review of Economics* **8**, 615–635.

Farronato, C. and Fradkin, A. (2017), The welfare effects of peer entry in the accommodation market: The case of airbnb, Technical report, Technical report, Working paper.

Fradkin, A. (2017), 'Search, matching, and the role of digital marketplace design in enabling trade: Evidence from airbnb', *Matching, and the Role of Digital Marketplace Design in Enabling Trade: Evidence from Airbnb (March 21, 2017)* .

Fraiberger, S. P., Sundararajan, A. et al. (2015), 'Peer-to-peer rental markets in the sharing economy', *NYU Stern School of Business research paper* **6**.

Lam, C. and Liu, M. (2018), 'More than taxis with an app: How ride-hailing platforms promote market efficiency'.

McFadden, D. (1977), Modelling the choice of residential location, Technical report, Cowles Foundation for Research in Economics, Yale University.

Nevo, A. (2001), 'Measuring market power in the ready-to-eat cereal industry', *Econometrica* **69**(2), 307–342.

Sheppard, S., Udell, A. et al. (2016), 'Do airbnb properties affect house prices', *Williams College Department of Economics Working Papers* **3**.

Sundararajan, A. (2016), *The sharing economy: The end of employment and the rise of crowd-based capitalism*, Mit Press.

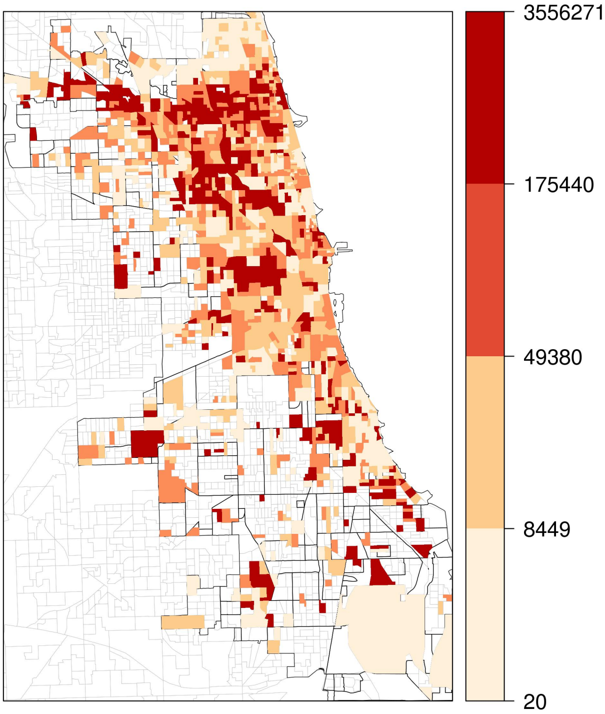
Train, K. E. (2009), *Discrete choice methods with simulation*, Cambridge university press.

Zervas, G., Proserpio, D. and Byers, J. W. (2017), 'The rise of the sharing economy: Estimating the impact of airbnb on the hotel industry', *Journal of marketing research* **54**(5), 687–705.

A Appendix

Figures

Figure 1. Block-Group Level Distribution of Total Revenue from Aug 2014 to Apr 2017



Note: Block-groups are defines in city of Chicago data portal.

Figure 2. Total Revenue Over Time

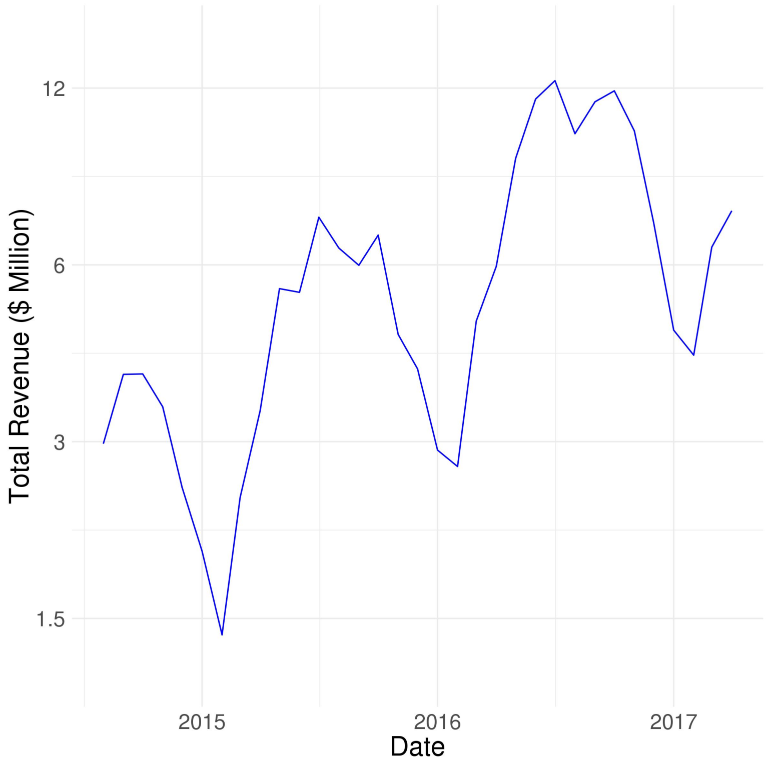


Figure 3. Total Daily Revenue in Apr-2017

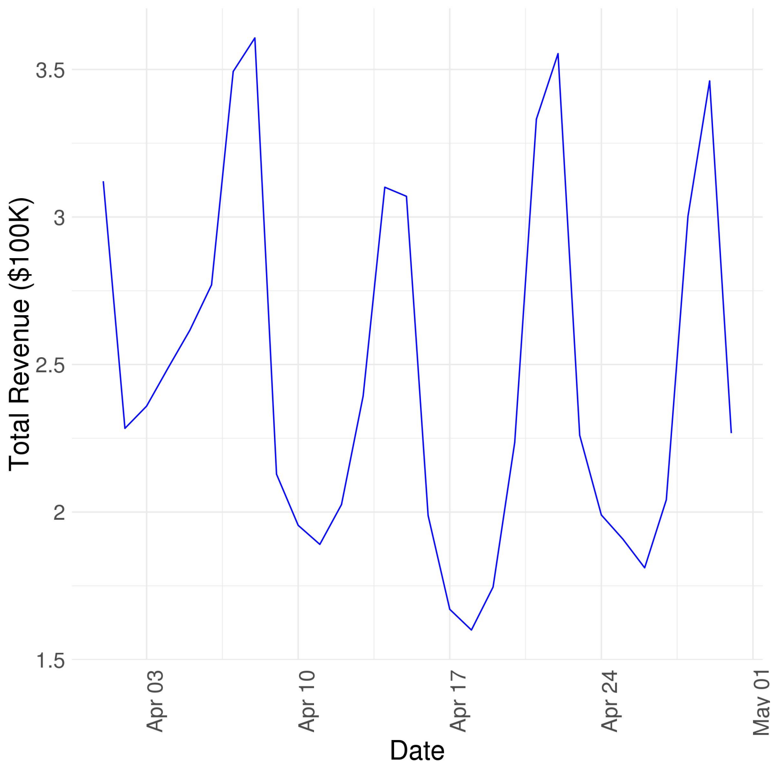


Figure 4. Average Price of All (Booked or Available) and Booked Listings Over Time

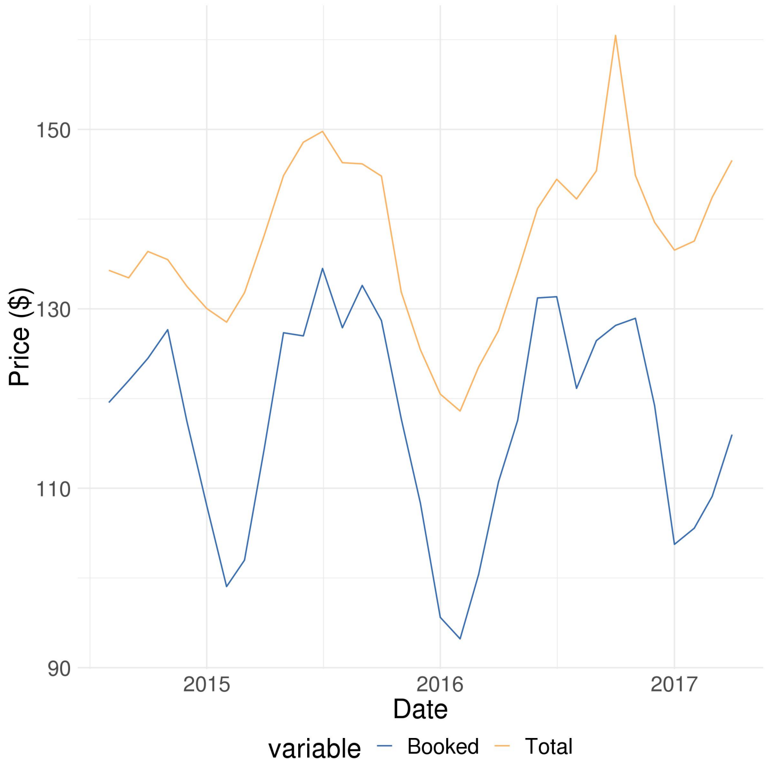


Figure 5. Monthly market size and total number of bookings

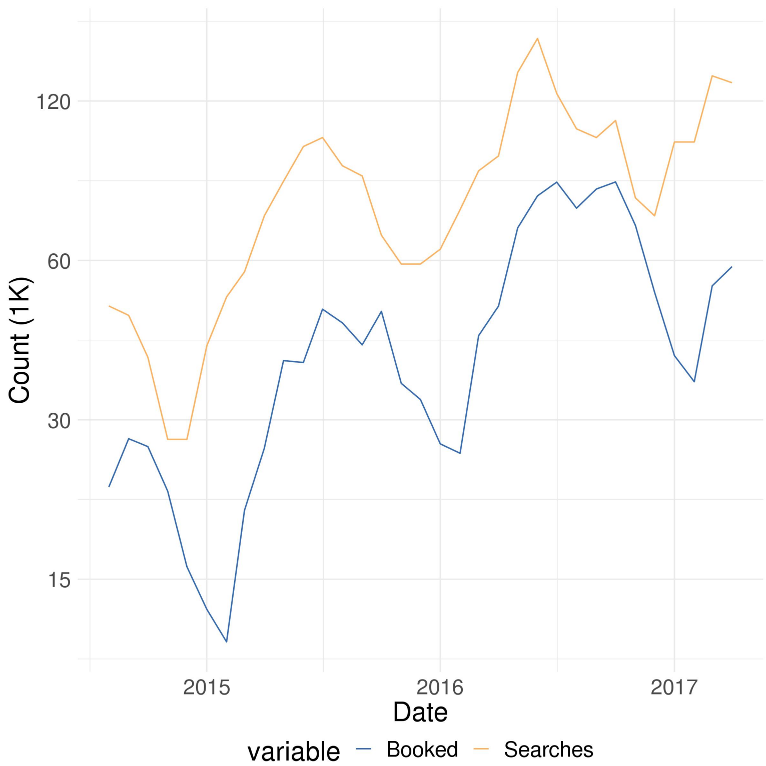
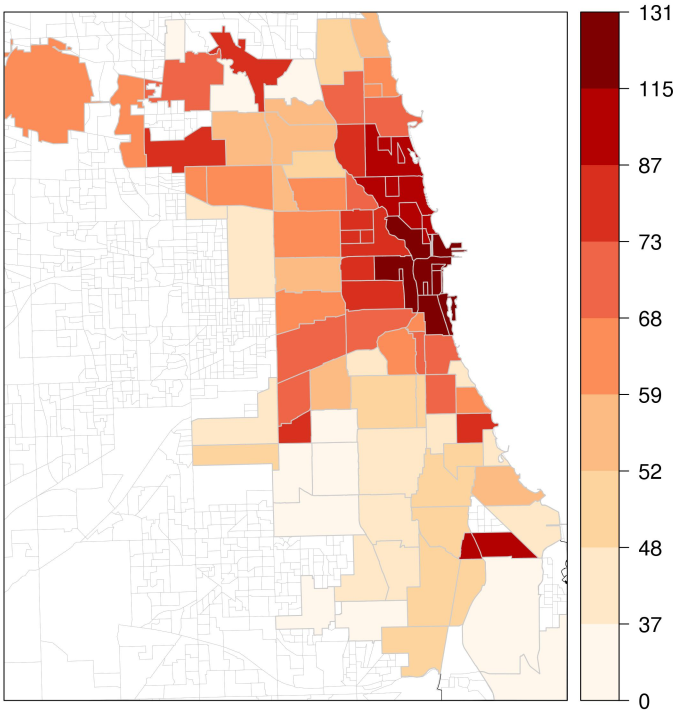


Figure 6. Average Consumers Willingness to Pay for Chicago Neighborhoods in Dollars



Note: This map is the plot of neighborhood fixed effects in the mean utility regressions. The fixed effects are adjusted with coefficient of price and presented in dollar values. The borders are community neighborhoods defines in city of Chicago data portal.

Figure 7. Average Consumer and Producer Surplus Per Dollar Spent

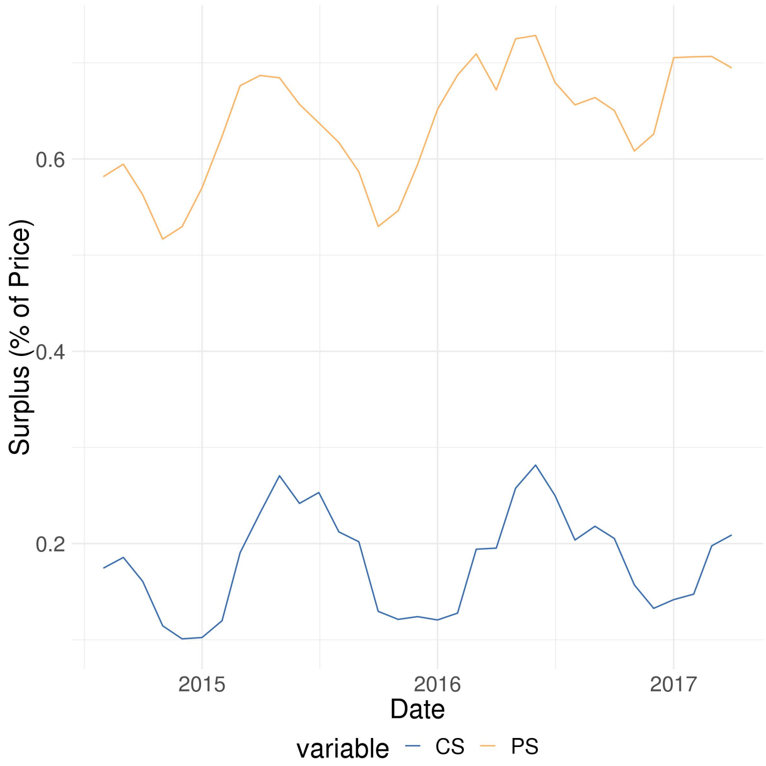
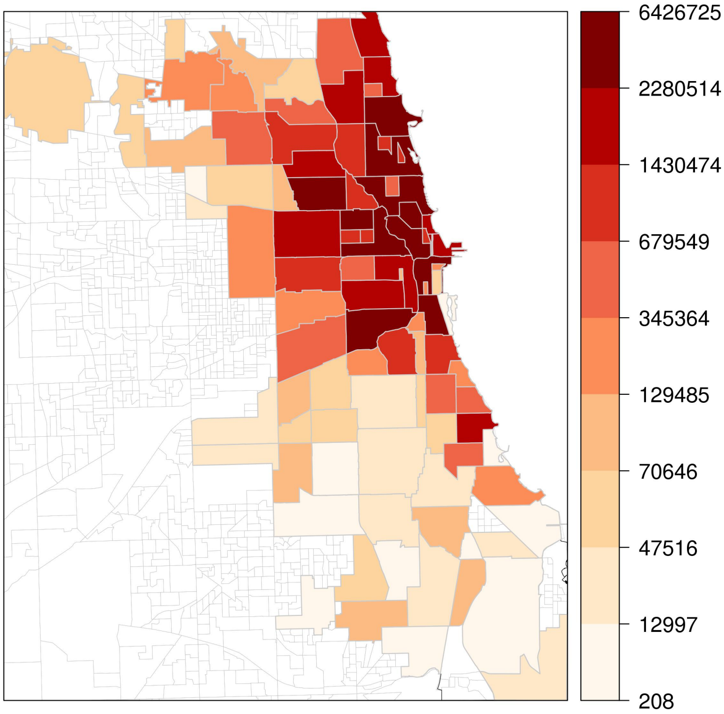
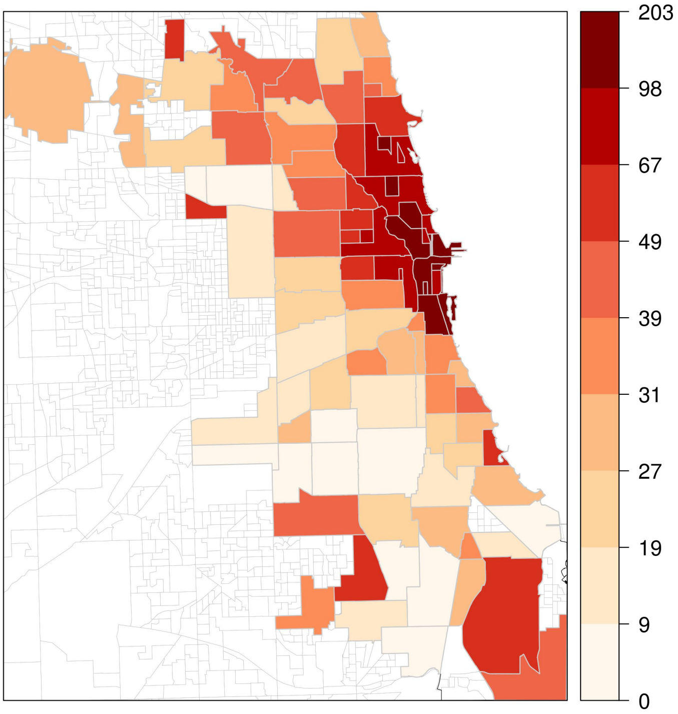


Figure 8. Geographic Distribution of Total Producer Surplus



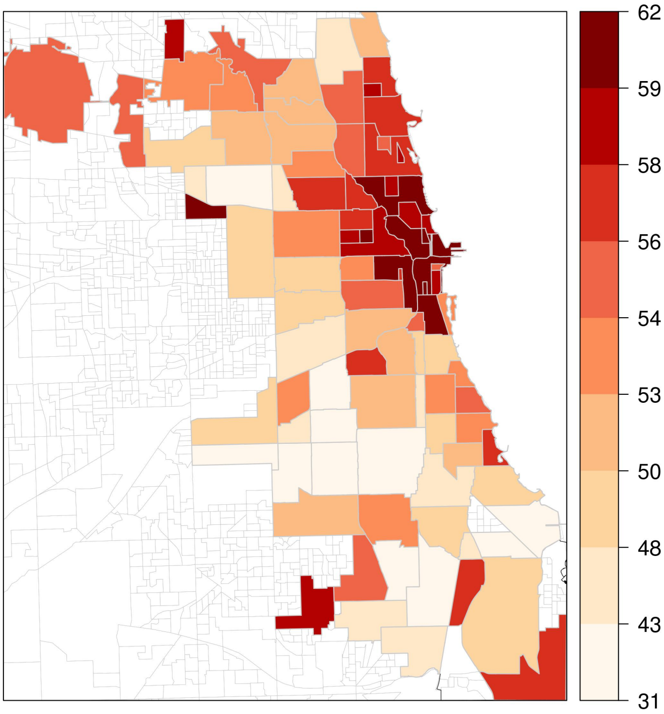
Note: The borders are community neighborhoods defines in city of Chicago data portal.

Figure 9. Geographic Distribution of Average Opportunity Cost Per Property Per Day



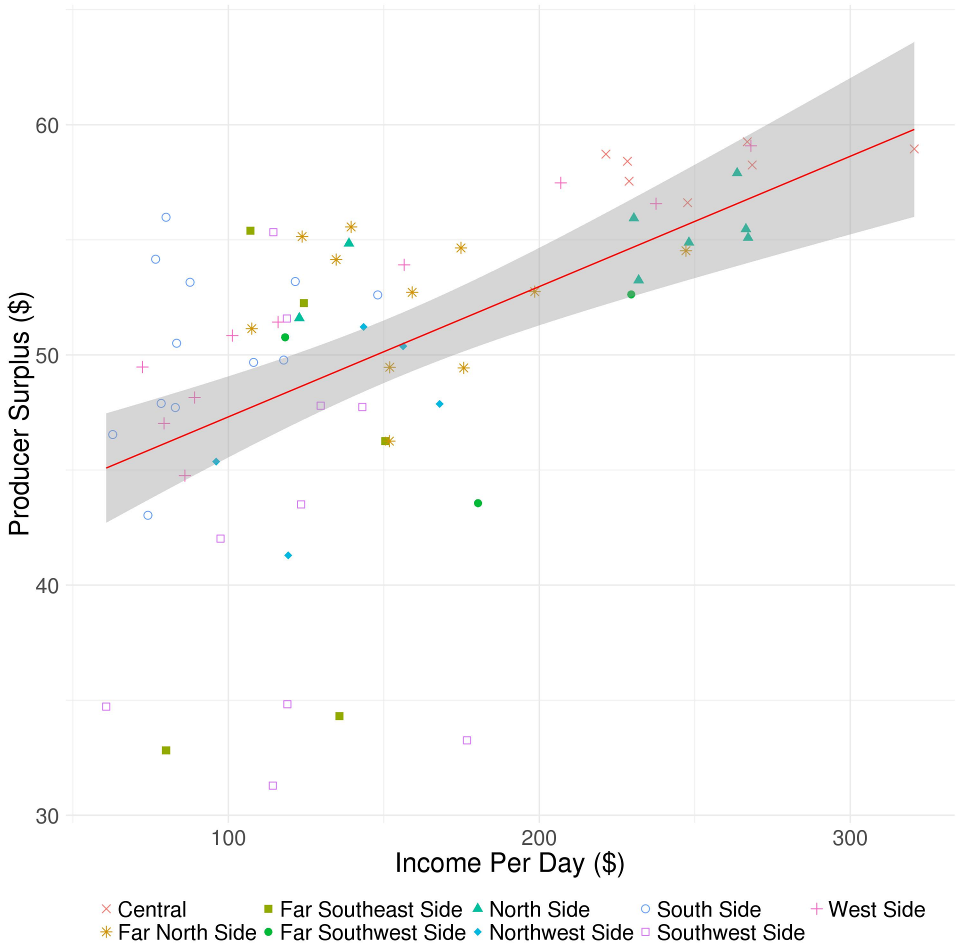
Note: The borders are community neighborhoods defines in city of Chicago data portal.

Figure 10. Geographic Distribution of Average Producer Surplus Per Property Per Day



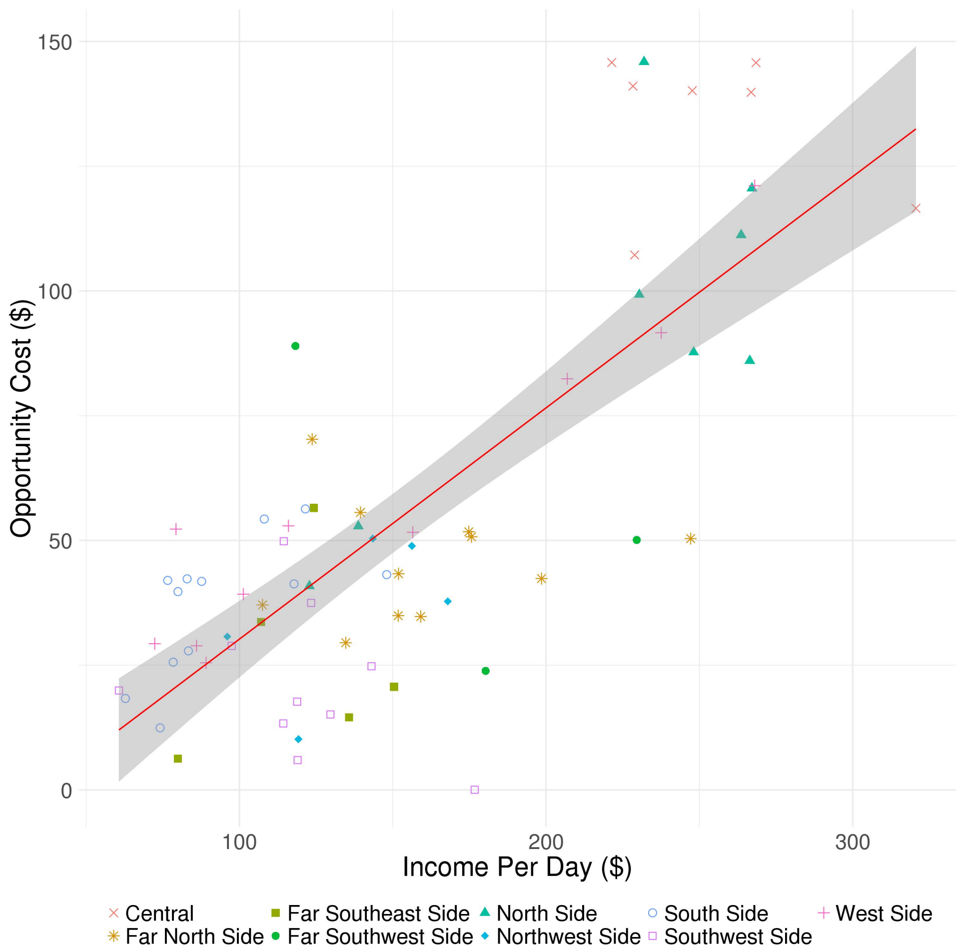
Note: The borders are community neighborhoods defines in city of Chicago data portal.

Figure 11. Correlation between average income and average producer surplus per booking across neighborhoods



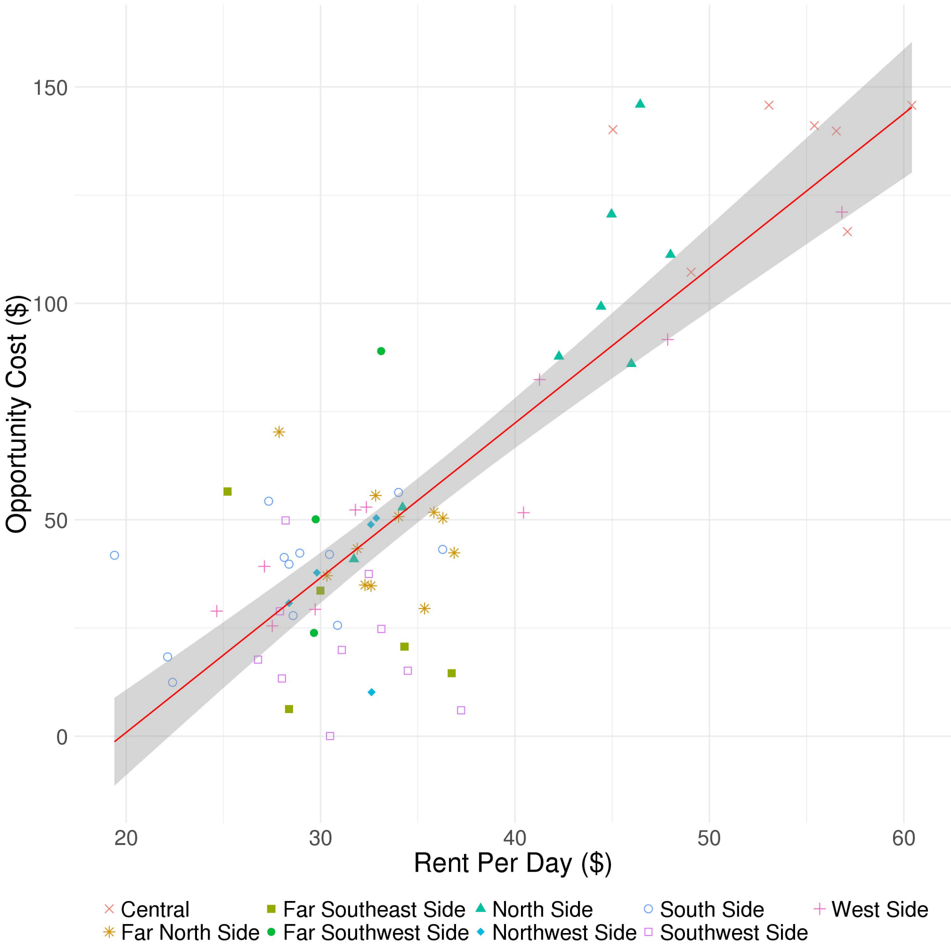
Note: Neighborhoods and Sides of Chicago are defined in the City of Chicago data portal.

Figure 12. Correlation between average income per day and the average opportunity cost of renting out listings in a day across neighborhoods



Note: Neighborhoods and Sides of Chicago are defined in the City of Chicago data portal.

Figure 13. Correlation between average rent per day and opportunity cost of renting out in a day across neighborhoods



Note: Neighborhoods and Sides of Chicago are defined in the City of Chicago data portal.

A.1 Tables

Table 1. Distribution of Mean Utility Parameters

Statistic	Pctl(25)	Mean	Pctl(75)	St. Dev.
δ_{ht}/α	−334.004	−294.465	−244.134	78.763
γ_t/α	−156.871	−120.914	−81.263	56.773
ξ_h/α	−82.545	−33.341	−1.887	84.805
Outside Option	0.311	0.457	0.592	0.157

Note: Dollar value estimation of mean utility parameters. The last row is share of outside option in the market. The negative mean utility estimations show high share of the outside option in the market.

Table 2. BLP-IV Regression of Mean Utilities

	Delta - Mean Utitlity Regression			
	(OLS)	(IV)	(IV)	(First Stage)
Price	-0.014^{***}	-0.026^{***}	-0.026^{***}	$--$
	(0.0003)	(0.005)	(0.005)	
Date FE	Yes	Yes	Yes	Yes
Neighborhood FE	Yes	No	Yes	Yes
Other Controls	Yes	No	Yes	Yes
Clustered	Yes	Yes	Yes	Yes
F-Statistics	5672.82	1526.17	2068.35	3973.53
F-Stat(excl instr.)				13.98
Observations	4,233,376	4,233,376	4,233,376	4,233,376

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

Note: This table shows the price coefficient of IV-regressions of mean utility on price and characteristics of listings. The first column is the result of an OLS regression. Second and third column are regressions with BLP-IV for price. All regressions are controlled for maximum number of guests, type of listings, and average characteristics of other options in the neighborhood. They are all clustered in the neighborhood-level. Other controls includes number of reviews, number of photos, Dummy variables for instant booking, super host, asking deposit, and business ready

Table 3. Blocked Days-IV Regression of Mean Utilities

	Delta - Mean Utility Regression			
	(OLS)	(IV)	(IV)	(First Stage)
Price	-0.015^{***}	-0.026^{***}	-0.025^{***}	
	(0.0003)	(0.003)	(0.002)	
IV				-27.327^{***}
				(2.932)
Date FE	Yes	Yes	Yes	Yes
Neighborhood FE	Yes	No	Yes	Yes
Other Controls	Yes	No	Yes	Yes
Clustered	Yes	Yes	Yes	Yes
F-Statistics	6221.01	1159	2036.18	4048.96
F-Stat(excl instr.)				86.85
Observations	4,233,376	4,233,376	4,233,376	4,233,376

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

Note: This table shows the price coefficient of IV-regressions of mean utility on price and characteristics of listings. The first column is the result of an OLS regression. Second and third column are regressions with share of blocked days as an IV for price. All regressions are controlled for maximum number of guests and type of listings. They are all clustered in the neighborhood-level. Other controls includes number of reviews, number of photos, Dummy variables for instant booking, super host, asking deposit, and business ready.

Table 4. Full Table - IV Regression of Mean Utilities

	Delta - Mean Utility Regression		
	(OLS)	(Blocked-IV)	(BLP-IV)
Max Guests	0.170***	0.250***	−0.057
	(0.018)	(0.030)	(0.117)
Entire Place	0.682***	1.070***	1.301***
	(0.034)	(0.104)	(0.220)
Bedrooms	0.502***	0.791***	0.676***
	(0.034)	(0.078)	(0.168)
Bathrooms	0.316***	0.716***	0.712***
	(0.080)	(0.131)	(0.211)
Overall Rating	0.339***	0.471***	0.511***
	(0.035)	(0.059)	(0.075)
Number of Photos	0.007***	0.010***	0.012***
	(0.002)	(0.002)	(0.003)
Business Ready	0.012	−0.039	−0.056
	(0.028)	(0.040)	(0.047)
Super-host	0.406***	0.358***	0.357***
	(0.031)	(0.042)	(0.039)
Deposit	0.062*	0.097**	0.083**
	(0.033)	(0.038)	(0.042)
Instant-book	0.316***	0.183***	0.185***
	(0.033)	(0.045)	(0.069)
Price	−0.015***	−0.025***	−0.026***
	(0.0003)	(0.002)	(0.005)
Date FE	Yes	Yes	Yes
Neighborhood FE	Yes	Yes	Yes
Neighborhood Controls	No	No	Yes
Clustered	Yes	Yes	Yes
Observations	4,233,376	4,233,376	4,233,376
Note:		*p<0.1; **p<0.05; ***p<0.01	

Note: This table shows the price coefficient of IV-regressions of mean utility on price and characteristics of listings. The first column is the result of an OLS regression. Second and third column are regressions with share of blocked days and BLP IV for price, respectively. All regressions are clustered in the neighborhood-level. Neighborhood controls includes average characteristics of rival listings in the neighborhood. Each coefficient shows the mean willingness to pay for the characteristic.

Table 5. Distribution of Producer Surplus in Different Sides of Chicago

Side	Producer Surplus (PS) ($)	Share of PS	Share of Population
North Side	22,472,189	0.276	0.190
West Side	21,858,242	0.268	0.174
Central	18,709,368	0.229	0.096
Far North Side	10,368,850	0.127	0.236
South Side	5,494,908	0.067	0.125
Northwest Side	1,657,773	0.020	0.091
Southwest Side	584,932	0.007	0.057
Far Southeast Side	266,063	0.003	0.013
Far Southwest Side	155,063	0.002	0.018

Note: First column is 9 sides Chicago defined in City of Chicago Data Portal. Second column shows total producer surplus between Aug 2014 to Apr 2017 in each side. Third column is the share of total producer surplus and the fourth column is share of population in each side.

Table 6. Property-level Correlation of Producer Surplus with Opportunity Costs and Share of Blocked Days

	Logarithm of Total Produce Surplus Per Property			
	(1)	(2)	(3)	(4)
Opportunity Cost	-0.15^{***}	-0.17^{***}	-0.31^{***}	
	(0.02)	(0.02)	(0.05)	
Share of Blocked				-1.30^{***}
				(0.05)
Neighborhood FE	No	Yes	Yes	Yes
Clustered	Yes	Yes	Yes	Yes
Subset	Total	Total	Central	Total
Controls	Yes	Yes	Yes	Yes
Observations	13,481	13,481	3,266	13,481
Adjusted R^2	0.02	0.03	0.02	0.09

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

Note: The first to third columns show the correlation between total producer surplus per property with average opportunity cost of each property. The third column is the result of regression on the subset of the Central side of Chicago. The fourth columns shows the correlation of producer surplus of each property with share of days that the property was blocked. All regressions are controlled for types of listings, number of bedrooms and number of bathrooms.

B Appendix

Estimating Market Size

Market size or number of potential guests in each day looking for propertied to rent with specific number of guests is estimates as follows:

First, I use google trend searches for “Airbnb Chicago” in each month. Google reports an index between 0-100 for number of searches. I use this index to capture seasonal variations and overall trends of market size. To have a measure of market size, I scale the index such that it concludes 60% rate of booking in the first three month of data as follows:

$$\text{Monthly market size} = \text{google trends} \times \frac{\text{total reservations in the first three month}}{\text{google trend searches in the first three month} \times 0.6}.$$

To find market-level number of potential guests, I interpolate the monthly measure of market size using share bookings with each number of maximum guests and share of bookings in each day as follows:

$$\begin{aligned} \text{Daily market size} = &\text{Monthly market size} \times \\ &\text{share of bookings with a specific maximum guests in the month} \times \\ &\text{share of bookings in a day} \end{aligned}$$

Proof of Equation (4)

$$\text{Producer surplus of listing } h \ = -\frac{Pr_{ht}}{\partial Pr_{ht}/\partial P_h}$$

$$Pr_{ht} = 1 - \prod_{i \in N_t} (1 - \frac{\exp(\delta_{ht})}{1 + \sum_{j \in M_{it}} \exp(\delta_{ht})}) = 1 - \prod_{i \in N_t} (1 - s^i_{ht})$$

$$-\frac{\partial Pr_{ht}}{\partial P_h} = \alpha \sum_{i \in N_t} \frac{\exp(\delta_{ht})}{1 + \sum_{j \in M_t} \exp(\delta_{ht})} \prod_{i \in N_t} (1 - \frac{\exp(\delta_{ht})}{1 + \sum_{j \in M_{it}} \exp(\delta_{ht})})$$

$$= \alpha \sum_{i \in N_t} s^i_{ht} \prod_{i \in N_t} (1 - s^i_{ht}) \implies$$

$$PS_{ht} = \frac{1 - \prod_{i \in N_t} (1 - s^i_{ht})}{\alpha \sum_{i \in N_t} s^i_{ht} \prod_{i \in N_t} (1 - s^i_{ht})}$$

Proof of Equation (6)

- Pr^i_{ht}: probability that guest i prefers listing h over all options at time t
- Pr_{ht}: probability that listing h is booked at time t
- $\bar{Pr}^i_{ot}$: probability that guest i prefers the outside option over un-booked listings at time t

- δ_h^t: utility of listing h at time t
- ξ_h, γ_t: listing and time mean utility fixed effects
- M_t^i : set of available listings for potential guest i in the market (Market is day - number of guests)
- N_{ot} : number of potential guests who choose the outside option at time t
- N_t : set of potential guests at time t
- B_t : set of all individual who booked a listing
- NB_t : set of all individual who did not book a listing

$$Pr_{ht} = 1 - \prod_{i \in N_t}(1 - Pr_{ht}^i) = 1 - (1 - Pr_{ht}^i)^{N_t}$$

$$\delta_h^t = \xi_h + \gamma_t$$

$$Pr_{ht}^i = \frac{\exp \delta_h^t}{1 + \sum_{j \in M_t^i} \exp \delta_j^t}, \quad Pr_{ot}^i = \frac{1}{1 + \sum_{j \in M_t^i} \exp \delta_j^t}$$

$$\frac{\partial Pr_{ht}^i}{\partial \gamma_t} = \frac{exp(\delta_h^t)(1 + \sum_{j \in M_t^i} \exp \delta_j^t) - exp(\delta_h^t) \sum_{j \in M_t^i} \exp \delta_j^t}{(1 + \sum_{j \in M_t^i} \exp \delta_j^t)^2} = Pr_{ht}^i - Pr_{ht}^i(1 - Pr_{ot}^i) = Pr_{ht}^i Pr_{ot}^i$$

$$\frac{\partial Pr_{ht}^i}{\partial \delta_h} = Pr_{ht}^i(1 - Pr_{ht}^i)$$

$$ll = \sum_t \sum_i \sum_h I_{ht}^i \log(Pr_{ht}^i) + \sum_t \sum_i I_{ot}^i \log(Pr_{ot}^i) = \sum_t \sum_i \sum_h I_{ht}^i \log(Pr_{ht}^i) + \sum_t \sum_{i \in NB} I_{ot}^i \log(Pr_{ot}^i)$$

$$\frac{\partial ll}{\partial \xi_h} = 0 \implies$$

$$\sum_t (\sum_{i=h} (1 - Pr_{ht}^i) + \sum_{i \neq h, i \in B} -Pr_{ht}^i) + \sum_t \sum_{i \in NB} -Pr_{ht}^i = 0 \implies N_h = \sum_t \sum_i Pr_{ht}^i$$

$$\frac{\partial ll}{\partial \gamma^t} = 0 \implies$$

$$\sum_{i \in B} Pr_{o_t}^i - \sum_{i \in NB} (1 - Pr_{ot}^i) = 0 \implies N_o^t = \sum_{i \in B} Pr_o^i + \sum_{i \in NB} Pr_o^i = \sum_i Pr_o^i$$