Racial Discrimination in the Sharing Economy:

An Economic Analysis of New York City Airbnb Listings

By

The present study tests the presence of racial discrimination by use of the hedonic pricing approach. On Airbnb, guests are effectively shopping for temporary homes. Properties differ in size, layout, location, and included amenities, and are priced accordingly. The hedonic approach determines the price of each of these individual characteristics that are included in a home by use of a linear regression that predicts the total price (O'Sullivan 284). One of the earliest examples of the hedonic pricing approach in practice is John Kain and John Quigley's (1970) study of housing prices in St. Louis. They regressed market prices of both owner and renter occupied units on 39 characteristics that represented the quality of each housing bundle. These 39 variables included seven measures of the quality of households (i.e. condition of floors, walls, windows, etc.), seven measures of the quality of the structure (i.e. condition of driveways and walkways, landscaping, etc.), eight measures of the quality of surrounding properties (i.e. condition of structures, parcels, etc.) and 17 measures of the quality of the block (i.e. condition aaono=n

theory, discrimination can impact prices if the parties involved hold tastes for discrimination. The present study takes the factor of host race into account in an effort to determine the economic costs of racial discrimination on Airbnb, by conducting a similar analysis to that seen in Kain and Quigley's (1970) study.

Various studies have examined the impact of discrimination in the sharing economy, particularly in the context of Airbnb. The recent study by Lee, Hyun, Ryu, Lee, Rhee and Suh (2015) examined the impact of features associated with the sale of Airbnb accommodations. The study included data from 4,178 rooms across five major cities in the United States: New York City, Chicago, Los Angeles, San Francisco and Seattle. In order to measure the number of sales of each unit over the two-month period of data collection, they used the change in number of reviews on each unit ("review delta") as a proxy for the minimum number of reservations over the time period. They collected data on August 1 and October 1 of 2014 in an effort to capture this change in reviews. Since sale data is not public, and reviews can only be written after an accommodation is booked, this data point appears to be an acceptable proxy for sales.

The model presented in their study includes a linear regression involving a multitude of

p hi20 y d s 4 che nhe er of ry d chd Ohe nesesne of led e meudago,factoriccricmwfjc cud

reviews on the listing, and the membership seniority of the host. Other significant predictors included whether or not the accommodation included a TV, air conditioner, shampoo, essentials, cleaning fees and a minimum stay requirement. They did not include the results for the price variable in the study. Although they included a wide variety of social factors, Lee et. al. neglected to include information about race of each respective host, so they could not test for racial discrimination with their data set.

Ert, Fleischer and Magan (2016) further examined the impact of social features and their impact on Airbnb listings in their recent study that assessed the role of personal photos on Airbnb. The study aimed at answering the question as to whether or not consumers infer sellers' trustworthiness from their personal photos, a process that they describe as "virtual-based trust", as well as the sellers' perceived attractiveness. In turn, they hypothesized that this visual-based trust and attractiveness impacts consumers' decision making as to whether or not to book an accommodation. In order to conduct this analysis, they collected similar photographs of 70 amateur actors (35 females and 35 males) and constructed mock Airbnb listings for each one. In an effort to assess the perceived trustworthiness and attractiveness of each host and listing, they employed a group of 31 undergraduate students who rated the 70 actors based on attractiveness and apparent trustworthiness, and 21 undergraduate students who evaluated the photographs of 39 rooms based on whether or not they would rent each accommodation. Ert. et. al. (2016) then gathered 566 Israeli participants from an online panel of 120,000 volunteers who selected preferred accommodations from sets of two of the mock listings.

The results of Ert et. al.'s (2016) mixed logit analysis, which estimated the effect of the visual-based trustworthiness and attractiveness of the hosts on the probability that their listings will be selected, confirmed their hypothesis that visual-based trust affects listing choice. The

each host, they employed workers on Amazon Mechanical Turk to examine the photo of each host included in the study. The workers coded the race of each host into one of the following categories: White, Black, Hispanic, Asian, Unclear but Non-White, Multiple Races, Not

Asian a 1, and hosts they perceived as White a 0. They omitted any hosts that did not appear to fit in to either of these categories, as well as any hosts for which race was uncertain.

Wang et. al. (2015) confirmed Edelman and Luca's (2014) results, which found that minority hosts face discrimination and therefore charge lower prices than White hosts. However,

wide variety of social factors, they do not include race as a predictor. Since the previously mentioned studies indicated that guests take the race of the hosts into consideration when booking accommodations, I include the race of the host as my variable of interest.

III. Data

In my research, I used a data set of New York City Airbnb listings from October 2016 provided by Airdna (Airdna.com 2016). Airdna is a company based in the United States that provides Airbnb data and analytics to vacation rental entrepreneurs and investors. They track the daily performance of over 2,000,000 listings across roughly 5,000 cities around the globe. This data set provided me with the occupancy rates that I needed to properly conduct my analyses. The original data set provided by Airdna included information on 118,530 listings in the New

friendly workspace, an iron, hangers, a hairdryer, a TV, shampoo, heating, essentials, air conditioning and whether or not the listing is accepting of pets, families and events. The inclusion of these amenities or lack thereof is public on every listing, but the data set provided by Airdna did not include them. In order to include them, I needed to access each individual listing and determine whether or not each amenity was provided.

whom race was not clear, were removed from the study. Considering they deemed this method adequate, I carried out the same process. I sorted through each listing in my data set and labeled them as having a White or non-White host, and skipped over any listings for which the race of the host was ambiguous to me, with the goal of reaching 500 total listings. Furthermore, any listings that did not have a picture of the actual host, as well as those depicting multiple individuals of different ethnicities, were also left out from the data set. This resulted in a total of relationship with other variables in the analysis. Therefore, I found it unnecessary to include the results of my correlation matrix in the study.

IV. Model

The present study includes three linear regression models. The first is a simple hedonic pricing model matching to the best extent possible that of the Wang et. al. (2015) study. The model predicts the price of each listing, and includes the following:

$$lprice = \beta + \beta\ sqrtlbedrooms + \beta\ whiteq + \beta\ lmaxguests + \beta\ lbathroomssq + \varepsilon$$

Variable transformations, descriptions and statistics are provided in Table 1. While this model may provide some indication as to whether or not guests select their accommodations based on the race of the host, it includes a very limited number of variables. In turn, the model presents the potential for omitted variable bias. Omitted variable bias occurs when a predictor variable that is correlated with other repressors and partially determines the dependent variable is left out of the analysis. By leaving these predictors out, the model provides biased results of the coefficient on the included variables (Stock and Watson 2007). Since each Airbnb accommodation includes a diverse basket of characteristics and amenities, I felt as though the model listed above did not present a comprehensive prediction of price. In order to address the potential omitted variable bias involved in the first model, I created a second one that includes a wide variety of new variables that may influence the price of a listing, such as ratings and reviews, neighborhoods, property types and included amenities. This model contains the following:

$$lprice = \beta + \beta\ sqrtlbedrooms + \beta\ whitesq + \beta\ lmaxguests + \beta\ lbathroomssq + \beta\ lOccupancy + \beta\ CreatedDate + \beta\ OverallRating + \beta\ NumberofReviews + \beta\ ResponseRate + \beta\ Superhost + \beta\ SecurityDeposit + \beta\ CleaningFee + \beta\ ExtraPeopleFee +$$

$$\boldsymbol{\beta}\ \mathbf{MinimumStay} + \boldsymbol{\beta}\ \mathbf{NumberofPhotos} + \boldsymbol{\beta}\ \mathbf{Instabook} + \boldsymbol{\beta}\ \mathbf{White} + \boldsymbol{\beta}\ \mathbf{FreeParking} +$$

$$\boldsymbol{\beta}\ \mathbf{Elevator} + \boldsymbol{\beta}\ \mathbf{Pets} + \boldsymbol{\beta}\ \mathbf{HrCheckin} + \boldsymbol{\beta}\ \mathbf{FamilyFriendly}\ 2$$

this model is not perfect. It includes strange variable transformations that I saw as unnecessary for the analysis. The model also provides little detail on the physical characteristics of each listing, as well as certain social factors; therefore, I believe it may represent omitted variable bias. In order to remove this bias and improve the test of racial discrimination, I constructed a stronger model including far more social and physical features of each listing that predicts the price of each listing.

The results of my second regression are shown in Table 3. The new model presents an R-squared of 0.7977, implying that the regression predicts 79.77% of the variability in listing prices. Despite adding many new variables to the prior model, some of the results held. In this regression, whitesq and lmaxguests again present positive coefficients that are significant at the 99% confidence level. According to the model, White hosts charge 7.21% higher prices than non-White hosts for listings with similar characteristics. This figure is still positive and significant, and the extent to which White hosts charge more than non-White hosts has increased from the prior model. This suggests that the first model did indeed present omitted variable bias.

belongings, and hosts price these accommodations accordingly. However, guests are willing to pay the highest price for their own, private dwelling.

While most of these results were in line with expectations, others provided surprising results. Airbnb's that provide a laptop friendly workspace were priced 8.88% lower than those that did not. I cannot understand why this is the case, as having LaptopFriendly was not correlated with any other variables in the study. Similarly, I found it surprising that a one percent increase in occupancy rate led to a 45.15% decrease in price. While these low prices might be attracting guests, one would expect these hosts to raise their prices, as guests may view their listings as underpriced.

The results of my third regression are displayed in Table 4. In this model, I change the dependent variable from lprice to OccupancyRateLTM, as I believe the occupancy rate of a listing over the previous 12 months will provide greater insight into the possibility of racial discrimination on Airbnb than the prices of listings. The adjusted R-squared for my model was 0.3272, implying that the included independent variables explained 32.72% of the variation in occupancy rates. The variable of interest in my model, White, was statistically significant and positive, implying that guests take the race of the hosts into account when booking accommodations. Specifically, White hosts received an occupancy rate 6.18% higher than non-White hosts over the previous 12 months. Therefore, I conclude that racial discrimination is present on Airbnb. Other statistically significant positive coefficients included CreatedDate, OverallRating, NumberofReviews, ResponseRate, Minimum Stay, Instabook, Pool, and Neighborhood3, PropertyType5 and PropertyType6, while lprice, ListingTypes2, ListingTypes3, ExtraPeopleFee, WirelessInternet, and SuitableforEvents were all significant and negative.

the type of vehicle necessary, thus preventing users from selecting drivers based upon their ethnicity. If Airbnb were to follow this model, users could input specific accommodation factors that they find necessary, such as a certain number of rooms or access to a gym, as well as a

hosts also achieve higher annual occupancy rates than non-White hosts, despite charging higher prices. In the future, Airbnb consider changing their business model, in an effort to prevent racial discrimination and to provide equal opportunities for all users.

References:

Becker, Gary S.. The Economics of Discrimination. Chicago, US: University of Chicago Press, 1971. ProQuest ebrary. Web. 7 December 2016.

Edelman, Benjamin G., and Michael Luca. "Digital Discrimination: The Case of Airbnb.com." (2014). Harvard Business School, 10 Jan. 2014. Web. 25 Mar. 2016.

Edelman, Benjamin, Michael Luca, and Dan Svirsky. “Racial Discrimination in the Sharing Economy: Evidence from a Field Experiment.” Harvard Business School, 6 Jan. 2016. Web. 20 Mar. 2016.

Ert, Eyal, Aliza Fleischer, and Nathan Magen. "Trust and Reputation in The Sharing Economy: The Role of Personal Photos on Airbnb." Tourism Management 55 (2016): 62-73. SSRN Online Journals. Web. 15 Oct. 2016.

Kain, John F., and John M. Quigley. "Measuring the Value of Housing Quality." Journal of the American Statistical Association 65.330 (1970): 532-48. JSTOR. Web. 5 Dec. 2016.

Lee, Donghun, Woochang Hyun, Jeongwoo Ryu, Woo Jung Lee, Wonjong Rhee, and Bongwon Suh. "An Analysis of Social Features Associated with Room Sales of Airbnb." Proceedings of The 18th ACM Conference Companion on Computer Supported Cooperative Work & Social Computing (2015): 219-22. ACM Digital Library. Web. 20 Feb. 2016.

"New York City Annual Property Report October 11, 2016." Airdna. N.p., n.d. Web. 11 Oct. 2016. <https://www.airdna.co/>.

O'Sullivan, Arthur. Urban Economics. 6th ed. Boston: McGraw-Hill/Irwin, 2007. Print.

Stock, James H., and Mark W. Watson. Introduction to Econometrics. 2nd ed. Boston: Pearson/Addison Wesley, 2007. Print.

Table 1: Variables and Descriptive Statistics

Table 2: Regression 1 Results

Number of Observations = 470
F(4, 465) = 74.91
Prob > F = 0.0000
R-squared = 0.3809
Root MSE = .43658
Robust Standard Errors

lprice	Coefficient	Std. Err.	t	P>t

Table 4: Regression 3 Results

Number of Observations = 476
F(59, 415) = .
Prob > F = .
R-squared = 0.4122
Root MSE = .19957
Robust Standard Errors

OccupancyRateLTM	Coefficient	Std. Err.	t	P>t
White	0.0618387***	0.0217837	2.84	0.005
lprice	-0.2516029***	0.0370762	-6.79	0.000
NumberofReviews	0.0021339***	0.0003182	6.71	0.000

Heating	**0.0752769**	**0.0478945**	**1.57**	**0.117**
Essentials	**0.0145394**	**0.0286715**	**0.51**	**0.612**
AC	**-0.0496289**	**0.0358701**	**-1.38**	**0.167**
Cancellation1	**-0.0088888**	**0.0313522**	**-0.28**	**0.777**
Cancellation2	**0.0000000**	**(omitted)**		
Cancellation3	**0.0267752**	**0.0234947**	**1.14**	**0.255**
ListingTypes1	**0.0000000**	**(omitted)**		

Table 5: Variable Groups and F-Test Results

Base Amenities	Luxury Amenities
$F(18, 415) = 0.89$, $p = 0.5928$	$F(12, 415) = 2.71$, $p = 0.0015$