Reinvestigate Racial Discrimination in the Sharing Economy: Evidence from Field Experiments on Airbnb

Biancheng Wang[1]

Abstract

Recent research has found some evidence of racial discrimination in the sharing economy. In this paper, we reinvestigate racial discrimination based on a field experiment on Airbnb with the help of Correspondence Analysis and PCA. Racial discrimination is robust across hosts with different host characteristics and location characteristics. Interestingly, African American female hosts are more likely to accept African American female guests due to homophily. However, this homophily can only be found when hosts are female. Also, we find some evidence for statistical discrimination, i.e., the discrimination is driven by imperfect information. Besides, we estimate the expected opportunity cost of discrimination by predicting the likelihood that the house is filled up using several classification methods. The low cost in the sharing economy might somehow facilitate discrimination.

Keywords: Racial Discrimination, Correspondence Analysis, PCA, Opportunity Cost

1. Introduction

Over the past half-decade, United States has always been committed to reducing discrimination against African American. Many previous researches have documented discrimination against minority groups in various traditional markets, including labor markets (Bertrand and Mullainathan, 2004) and rental markets (Ewens et al., 2014). Many regulations have been introduced to tackle with discrimination in those markets correspondingly, for example, there are kind of racial discrimination audits that officials have long used to enforce fair housing laws against traditional landlords.

With the development of sharing economy, racial discrimination has become an important issue in these new marketplaces since those "sharing economy" companies are hard to be regulated by existing policy. Some evidence of discrimination among Uber and Lyft drivers has been found (Ge et al., 2016). As a popular online short-term rental marketplace, Airbnb allows hosts to decide whether accept or reject an application after seeing the information of the guest, which may also lead to racial discrimination by hosts. On social media, some African American Airbnb users reported experiences of facing a rejection by a host, who later accepted them when they changed their profile to a white person. Edelman et al. (2017) found that requests from guests with African American-sounding names are 16% less likely to be accepted compared to identical guests with White-sounding names.

[1] Department of Statistics, UCLA (email: wangbcbill@ucla.edu)

However, some limitation can be found when replicating the analysis of Edelman et al. (2017). In their research, for a binary independent variable, only simple OLS is used in estimating the coefficients. Some results can be better represented by multivariate analysis for a better understanding.

Noticing these limitations, in this paper, we reinvestigate the existence of racial discrimination in Airbnb, figure out the factor which may influence the effect of racial discrimination and estimate the expected opportunity cost of discrimination based on the field experiments conducted by Edelman et al. (2017). The expected opportunity cost is a measure of expected loss of not doing something.

In the next section, we review the experiment design method used by Edelman et al. (2017) and describe the dataset we have. Section 3 is dedicated to the brief replication of their research and compares their results with the ones obtained by logistic regression. In Section 4, we use Correspondence Analysis to visualize the impact of race and gender on host responses, reinvestigate the influence with the help of Principal Component Analysis, and estimate expected opportunity cost of racial discrimination using some classification methods as well as feature reduction methods. Finally, we conclude and give more discussion in Section 5.

2. Experiment Design and Data[2]

Edelman et al. (2017) used a random experiment method: they selected 5 metropolitan areas, including Los Angeles, Washington, DC, Baltimore, Dallas, and St. Louis., created 20 Airbnb accounts, identical in all aspects except for guests' names, and then randomly booked available properties eight weeks in advance to observe whether host responses were affected by the race of guests. Since some hosts may have multiple listings, they contacted each host once using one account to prevent hosts from receiving multiple identical emails from almost identical accounts.

Specifically, twenty Airbnb accounts can be divided into four main treatment groups equally, African American males, African American females, white males, and white females. The list of distinctive names that signaled ethnic characteristics are drawn from Bertrand and Mullainathan (2004) which was based on the frequency of names from birth certificates of babies born between 1974 and 1979 in Massachusetts. They also conducted a survey to validate this list by asking participants classify each name as African American or white. Here, we transfer the proportion of how many people regard the name as white to a new explanatory variable capturing the probability of being white, named as "whiteness of name", which we believe can decrease the measurement error that hosts do not categorize guests' names correctly.

For hosts, they collected data from each host's profile page and employed Mechanical Turk workers to roughly categorize hosts by race, gender, and age, which might be useful in sub-sample research. Besides, with the help of face-detection techniques, they categorized the race of past guests and created a variable representing the previous experience with African American guests.

[2] The dataset is available on the website of American Economic Association (https://www.aeaweb.org/articles?id=10.1257/app.20160213)

Also, information about each listing was collected, including the price of listing, the cancellation policy, the cleaning fee, and the property type. To represent characteristics of the location, they collected some census demographic data in that census tract as well as the number of listings in the neighborhood.

About 6400 rental applications were sent to hosts during July 2015 and they tracked host responses and broke down different responses into 5 groups, from 1 to 5: “Yes”, “Conditional Yes”, “No Response”, “Conditional No”, and “No”. Edelman et al. (2017) focused on “Yes” and treated “Yes” as their response variable while used “No” for robustness test. Besides, they also tracked whether these properties, whose owners rejected the application from African American guests, remained vacant. This data will enable us to estimate the expected opportunity cost of racial discrimination.

3. Paper Replication and Extension

At the first step of analysis, we follow the logic flow of Edelman et al. (2017) but use logistic regression instead of simple OLS regression. Also, we use the probability of the name being African American, a continuous variable, instead of the dummy variable. Since the host may wrongly categorize the name to African American or white when making decisions, we believe this continuous variable can reduce this type of measurement error. Some people may concern about the situation when “Conditional Yes” is misclassified to “Yes”. Therefore, to test the robustness of our results, we also use the response variable standing for whether hosts’ responses are “No”. Besides, statistical inference here is based on clustered robust standard errors, which are clustered by the combination of guest name and city. Technically, because there are only limited variables chosen into each regression, it is easy to avoid collinearity and other violation of regression assumptions, and there is no feature selection problem with low dimensional variables.

3.1 Main Effect

Table 1 presents the main impact of race on hosts’ responses. In Column 1 and 2, we simply replicate the method Edelman et al. (2017) used while we use logistic regression in Column 3, 4, 5, and 6. Column 5 and 6 treat the Whiteness of Name as the explanatory variable which we care about, thus leading to opposite coefficients compared with the others. There are only slight differences between different models in terms of the explanatory variable we mostly care about. However, due to the theoretical advantage we mentioned before, we still use logistic regression and the continuous measure of the race of guests’ names in the following analysis.

All results in Table 1 prove the existence of racial discrimination in Airbnb and it is remarkably robust. The African American guests have a significant lower acceptance rate in this sharing economy marketplace. From the Table 1, some relationships between our control variables and response variable are quite interesting. African American hosts, female hosts, hosts with more reviews who may be an avid one, and hosts with multiple listings, who might be professional hosts, are more likely to accept the application. When the property is a shared one and has a high price, the probability of rejection will increase.

Table 1: The Impact of Race on the Decision of Hosts

Model	OLS		Logit			
Response Variable	Yes	No	Yes	No	Yes	No
Guest is African-American	-0.09^{***}	0.07^{***}	-0.37^{***}	0.38^{***}		
	(0.02)	(0.01)	(0.07)	(0.08)		
Whiteness of Name					0.41^{***}	-0.42^{***}
					(0.09)	(0.09)
Host is African American	0.09^{***}	-0.05^{*}	0.39^{***}	-0.28^{*}	0.39^{***}	-0.29^{**}
	(0.02)	(0.02)	(0.10)	(0.11)	(0.10)	(0.11)
Host is Male	-0.05^{**}	0.03^{**}	-0.20^{***}	0.16^{**}	-0.20^{***}	0.15^{**}
	(0.01)	(0.01)	(0.06)	(0.05)	(0.06)	(0.05)
Host has Multiple Listings	0.06^{***}	-0.03^{*}	0.26^{***}	-0.16^{*}	0.26^{***}	-0.16^{*}
	(0.01)	(0.01)	(0.06)	(0.06)	(0.06)	(0.06)
Shared Property	-0.07^{***}	0.01	-0.30^{***}	0.05	-0.30^{***}	0.05
	(0.02)	(0.02)	(0.07)	(0.08)	(0.07)	(0.09)
Host has 10+ Reviews	0.12^{***}	-0.04^{***}	0.50^{***}	-0.24^{***}	0.50^{***}	-0.24^{***}
	(0.01)	(0.01)	(0.05)	(0.07)	(0.06)	(0.07)
ln(Price)	-0.06^{***}	-0.02^{*}	-0.27^{***}	-0.12^{*}	-0.28^{***}	-0.11^{*}
	(0.01)	(0.01)	(0.06)	(0.05)	(0.06)	(0.05)
Constant	0.76^{***}	0.33^{***}	1.14^{***}	-0.66^{**}	0.75^{*}	-0.27
	(0.07)	(0.04)	(0.31)	(0.24)	(0.30)	(0.26)
Observations	6168	6302	6168	6302	6129	6262
Adjusted R^2	0.039	0.012				

Note: * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$. (The significance codes used in this paper are consistent.)

3.2 Effects by Host Characteristics

Since we have found the race, gender and other information about hosts may influence the likelihood of acceptance from Table 1, we analysis effects of race and gender as well as other host characteristics. Two questions can be raised here. Is racial discrimination driven by homophily in race, i.e., hots might prefer guests of the same race? Is there any distinctive racial discrimination when taking other host characteristics into consideration?

Table 2 and Table 3 present racial discrimination by race/gender and other host characteristics respectively. In Table 2, “Others” includes hosts whose picture on their profile page are hard to be classified for gender, for example, there is no human in the picture. All the results prove that discrimination is robust across host characteristics.

For the first question, if homophily were the primary factor, we might find a significant positive interaction between the race of guests and hosts. However, as we can see from Table 2, the interaction term of whole group is not significantly different from zero. When looking at male hosts and female hosts separately, we still cannot reject coefficients of zero. But it shows heterogeneity across gender since the sign of coefficients of the interaction term are opposite.

Table 2: Racial Discrimination by Race

Response Variable: Yes				
	All	Male	Female	Others
Whiteness of Name	0.36***	0.41***	0.40***	0.29*
	(0.08)	(0.11)	(0.10)	(0.13)
Host is African American	0.26*	0.76**	0.00	0.13
	(0.12)	(0.23)	(0.17)	(0.35)
Guest is African American * Host is African American	0.02	-0.41	0.39	-0.26
	(0.19)	(0.32)	(0.25)	(0.57)
Constant	-0.42***	-0.61***	-0.40***	-0.27**
	(0.06)	(0.08)	(0.07)	(0.09)
Observations	6196	1844	2324	2028

Table 3: Racial Discrimination by Other Host Characteristics

Response Variable: Yes					
Whiteness of Name	0.29**	0.36***	0.41***	0.33***	0.43***
	(0.09)	(0.09)	(0.10)	(0.08)	(0.09)
Shared Property	0.02				
	(0.06)				
Shared Property * Guest is African American	-0.12 (0.11)				
Host has Multiple Listings		0.41***			
		(0.09)			
Host has Multiple Listings * Guest is African American		-0.02 (0.11)			
Host has 10+ Reviews			0.57***		
			(0.07)		
Host has Ten+ Reviews * Guest is African American			0.04 (0.10)		
Host Looks Young				-0.10	
				(0.07)	
Host Looks Young * Guest is African American				-0.09 (0.10)	
Host has at least one review from an African American guest					0.39*** (0.06)
Host has at least one review from an African American guest * Guest is African American					0.22* (0.10)
Constant	-0.34***	-0.52***	-0.71***	-0.33***	-0.58***
	(0.07)	(0.06)	(0.08)	(0.06)	(0.07)
Observations	6196	6196	6196	6196	6196

There are mainly two types of discrimination, one is statistical discrimination and the other is taste-based discrimination (Ruomeng Cui, et al., 2016). On one hand, statistical discrimination theory posits that the differential treatment of minority groups is driven by imperfect information. In other words, in this case, if the host has a prior belief that African American guests are less reliable than white guests, they may prefer white guests. On the other hand, the taste-based discrimination theory suggests that the differential treatment of minority groups is driven purely by preferences.

Therefore, for the second question, we take the following aspects into consideration to find some evidence for taste-based discrimination: the host's proximity to the guest, the age of the host, and whether the host is professional or not. The interaction terms in the first 4 Columns of Table 3 shows that there is no significant different preference between groups with different values of characteristics above.

Also, to examine statistical discrimination, we consider the host's prior experience with African American guests. If this theory fits this case well, previous experience may somehow correct or prove the host's prior belief. Column 5 of Table 3 shows that previous experience can increase the likelihood of accepting African American guests with a significance level of 0.05. Therefore, we prefer statistical discrimination theory here, which is consistent with the finding of Ruomeng Cui, et al. (2016).

3.3 Effects by Location Characteristics

Table 4: Racial Discrimination by Location Characteristics

Response Variable: Yes			
Whiteness of Name	0.36^{***}	0.36^{***}	0.39^{***}
	(0.11)	(0.09)	(0.11)
Price Above Median	-0.24^{**}		
	(0.09)		
(Price > Median) * Guest is Black	-0.00		
	(0.12)		
Share of Black Population in Census Tract		0.20	
		(0.18)	
Share of Black Population in Census Tract * Guest is Black		0.02	
		(0.27)	
Airbnb Listings per Census Tract			-0.00
			(0.00)
Airbnb Listings per Census Tract * Guest is Black			0.00
			(0.01)
Constant	-0.27^{***}	-0.42^{***}	-0.40^{***}
	(0.07)	(0.06)	(0.08)
Observations	6196	6184	6196

The effect of racial discrimination might also be related with some location characteristics. Firstly, we separately run logistic regressions using datasets in 5 cities. As Table A1 in Appendix shows, there is no big difference among cities. Only in Baltimore, African American guest may receive a lower ratio of acceptance which is driven by the city, but this effect is not so significant. Then we introduce three factors into our test: the relative price in the same census tract, proportion of African American population, and the total listings available in the neighborhood. From Table 4, we find that discrimination does not vary based on those characteristics.

4. Analysis and Results

4.1 Correspondence Analysis

Correspondence analysis is an exploratory multivariate technique for displaying scores representing the row categories and column categories of a two-way contingency table as the coordinates of points in a low-dimensional space. Edelman et al. (2017) listed some contingency tables in their paper and claimed that they were confused about a better way to represent the data in their supplemental materials. Correspondence analysis is just the answer because it can give a better and more direct view of the relationship between guests' name and hosts' responses than a contingency table, e.g., Table 5.

Table 5: Contingency table of Guest Name and Response Category

		Response Category				
Guest ID	Guest Name	1	2	3	4	5
1	Allison Sullivan	110	44	29	58	63
2	Anne Murphy	143	51	35	55	58
3	Laurie Ryan	132	37	29	57	70
4	Kristen Sullivan	115	43	42	50	74
5	Meredith O'Brien	112	42	31	42	73
6	Brad Walsh	95	38	50	56	76
7	Greg O'Brien	113	36	37	48	74
8	Brent Baker	123	43	36	58	69
9	Brett Walsh	93	33	68	33	41
10	Todd McCarthy	97	44	65	45	58
11	Lakisha Jones	107	33	23	53	103
12	Latonya Robinson	94	28	66	51	84
13	Latoya Williams	98	45	35	53	89
14	Tamika Williams	126	37	36	46	89
15	Tanisha Jackson	81	45	44	42	88
16	Darnell Jackson	81	30	32	48	77
17	Jamal Jones	81	35	39	56	105
18	Jermaine Jones	82	29	36	58	85
19	Tyrone Robinson	68	25	64	33	57
20	Rasheed Jackson	95	30	40	62	71

We implement simple correspondence analysis on Response versus Guest ID and Response versus Guest Name. The results are plotted in the Figure 1 and Figure 2 respectively. For Guest ID, Guest IDs from 1 to 5 are names of white female, 6 to 10 are white male, 11 to 15 are African American female, and 16 to 20 are African American male. The response varies from 1 to 5 which stands for five levels from “Yes” to “No”. Figure 2 presents how those African American names and white names are separated in a clearer way.

Figure 1: Correspondence Analysis of Response vs Guest ID

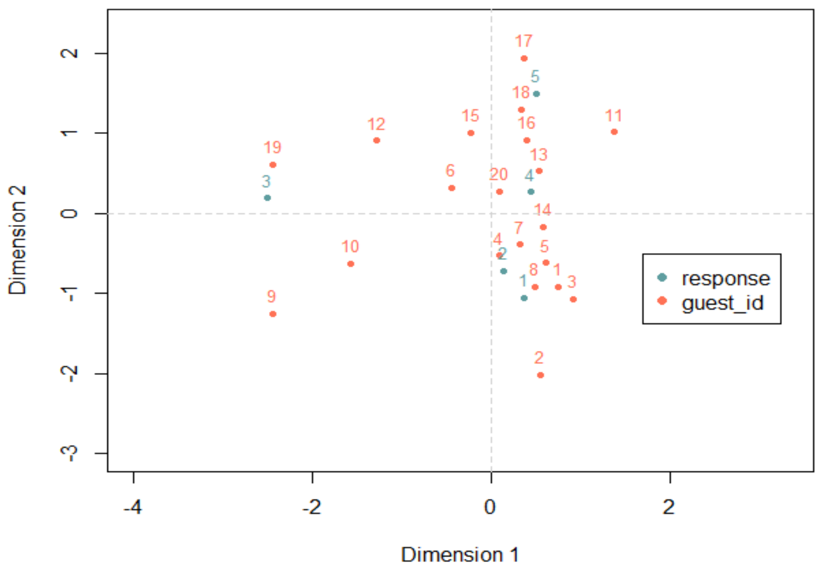
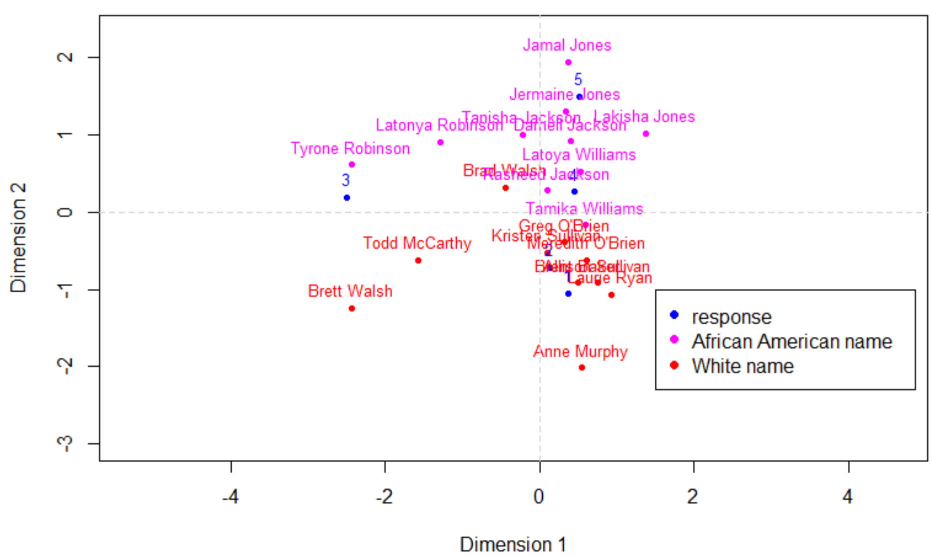


Figure 2: Correspondence Analysis of Response vs Guest Name



As we can see from figures, points representing white female are very close to “Yes” and the ones representing African American male are very close to “No” and “Conditional No” (except Tyrone Robinson who got many “No Response”), which means white female are more likely to be accepted and African American male are more likely to be rejected. For white male and African American female, these two groups of guests are scattered over the joint plot but, generally, white male are closer to “Yes” than African American female, i.e., African American female are far away from “Yes” compared with white male. Therefore, we can find some difference in the likelihood of acceptance between different genders.

4.2 PCA

Since the results of correspondence analysis shows that there might be some interesting patterns across race and gender, we try to present this by biplot using Principal Component Analysis (PCA). PCA is a dimensionality-reduction technique which can project high-dimensional data onto a lower-dimensional subspace without losing important information regarding some characteristic of the original variables. Therefore, PCA can help understand the relationship between variables and biplot can serve as a visualization method for the result of PCA.

Table 6 shows the proportion of positive responses across all race and gender groups (except undefined gender group in host). Table 7 shows the data broken down by name. We also collect ratios of acceptance across the host and location characteristics mentioned in Section 3 and present them in Appendix Table A3. Then we use these tables to conduct PCA to dig into the trend as well as the story behind the data.

Table 6: Ratio of Acceptance across all Race and Gender groups

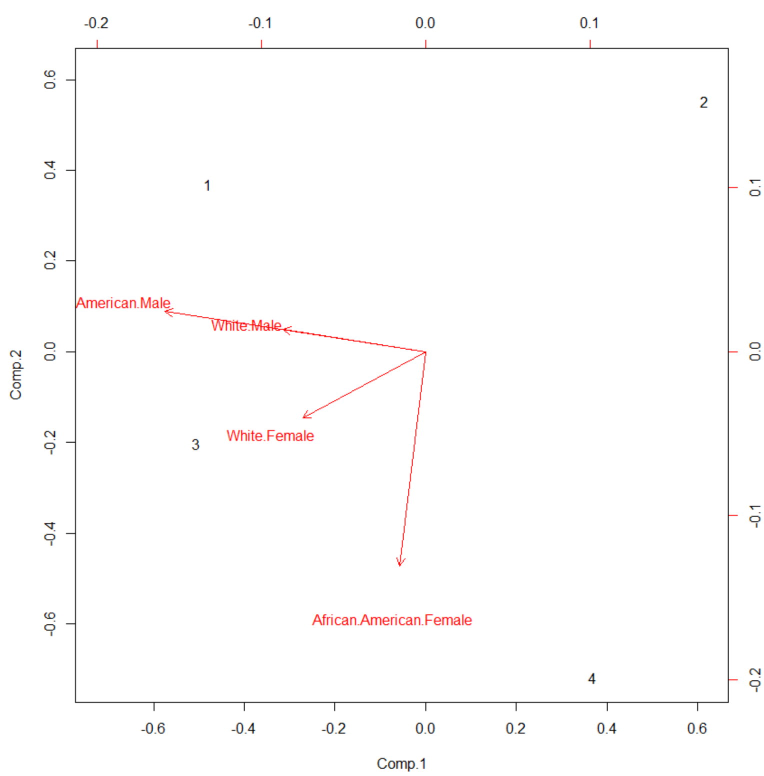
	Host Race/Gender			
Guest Race/Gender	White Male	African American Male	White Female	African American Female
White Male	0.4167	0.6429	0.4639	0.4328
African American Male	0.3540	0.4043	0.3470	0.3833
White Female	0.4882	0.5909	0.4879	0.5294
African American Female	0.3249	0.4314	0.4374	0.5942

Table 7: Ratio of Acceptance across Host Race/Gender and Guest ID

		Host Race/Gender			
Guest ID	Guest Name	White Male	African American Male	White Female	African American Female
1	Allison Sullivan	0.5082	0.7778	0.4235	0.5
2	Anne Murphy	0.5231	0.5556	0.5104	0.6
3	Laurie Ryan	0.5556	0.7273	0.5	0.5263
4	Kristen Sullivan	0.3846	0.6	0.4815	0.4615
5	Meredith O'Brien	0.4839	0	0.5161	0.6
6	Brad Walsh	0.4583	0.6	0.4304	0.5333

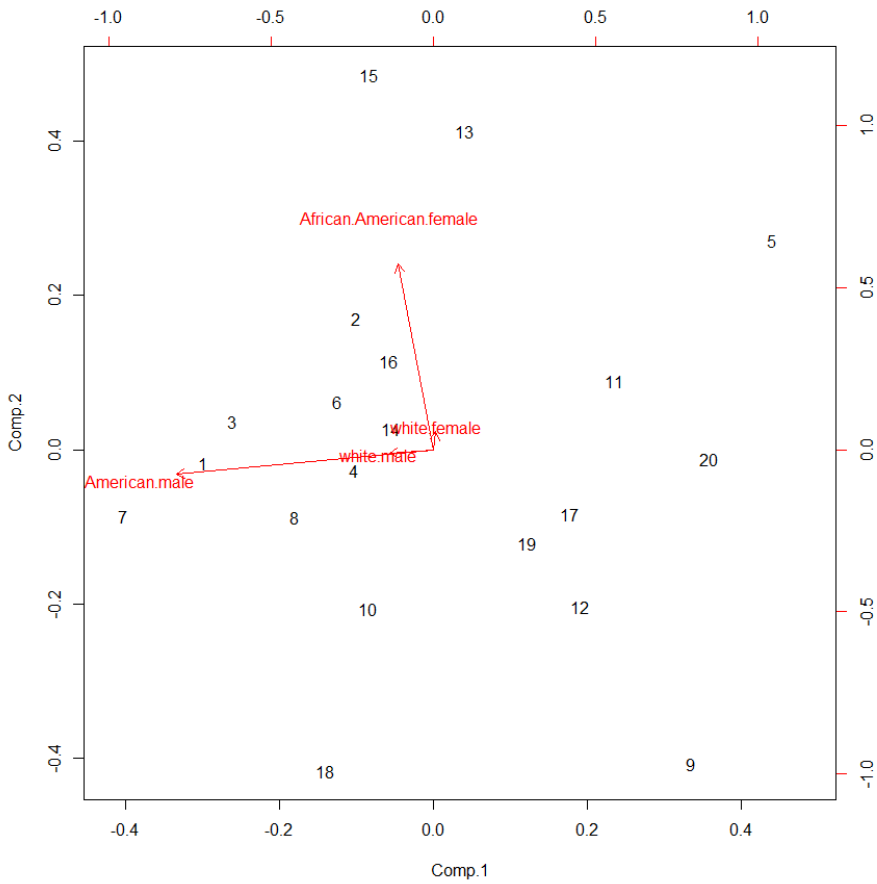
7	Greg O'Brien	0.403	0.9091	0.4524	0.4615
8	Brent Baker	0.4932	0.6667	0.4719	0.4286
9	Brett Walsh	0.3382	0.2	0.5067	0.125
10	Todd McCarthy	0.3833	0.6	0.4362	0.3333
11	Lakisha Jones	0.3538	0.25	0.4819	0.5
12	Latonya Robinson	0.2949	0.3333	0.4186	0.3
13	Latoya Williams	0.3472	0.4167	0.4316	0.7647
14	Tamika Williams	0.4237	0.5385	0.4158	0.5
15	Tanisha Jackson	0.2375	0.5556	0.4459	0.8333
16	Darnell Jackson	0.3793	0.5385	0.3433	0.5714
17	Jamal Jones	0.2895	0.3333	0.3059	0.4
18	Jermaine Jones	0.4286	0.6667	0.3188	0.2
19	Tyrone Robinson	0.4043	0.375	0.3529	0.375
20	Rasheed Jackson	0.2958	0.1429	0.4125	0.4167

Figure 3: PCA of Proportion of Positive Responses across all Race and Gender groups



In Figure 3, Component 1 can explain 63% of variance while Component 2 can explain 35%. “1”, “2”, “3”, and “4” represent white male, African American male, white female, and African American female respectively. The direction of vectors “African American Male” and vector “White Male” can best separate African American guests and white guests. That is to say, those two groups discriminate against African American. Female hosts seems to be more likely to accept female guests. We can find a strong preference between African American female hosts and guests, which can be explained by homophily. Interestingly, homophily plays an important role when hosts are female.

Figure 4: PCA of Proportion of Positive Responses by Host Race/Gender and Guest ID[3]



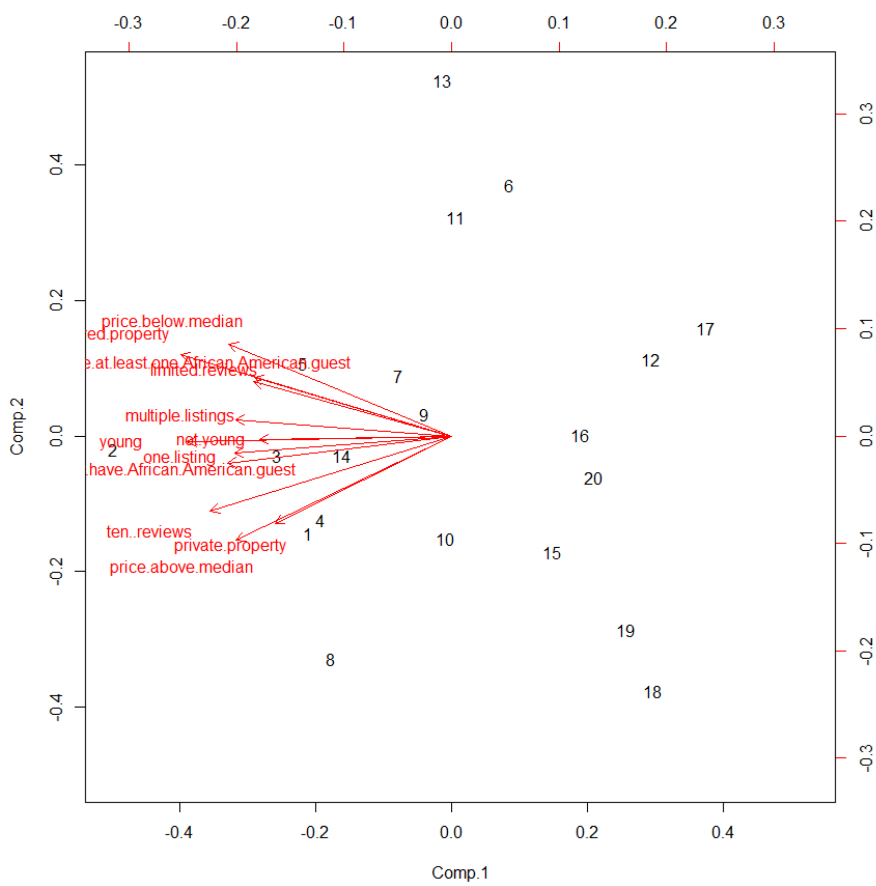
In Figure 4, we use the data from Table 7 which is broken down by name. Here, Component 1, which is mainly made by the acceptance ratio when hosts are African American male, can explain 60% of variance while Component 2, which is mainly made by the acceptance ratio when hosts are African American female, can explain 30%. As we can see, the result we find using Figure 3 remains robust.

In Figure 5, Component 1 can explain 80% of variance and the Top 4 Components can totally explain more than 95%. The direction of vectors suggests that racial discrimination is robust across different groups of hosts. However, by looking at the angles between corresponding vectors, we can find some different trends across hosts with different relative price, property setup (whether it is shared or private), number of reviews and prior experience with African American guests while the age of hosts and the number of listings are not so important. From the location of 20 Guest IDs

[3] For Guest ID, Guest IDs from 1 to 5 are names of white female, 6 to 10 are white male, 11 to 15 are African American female, and 16 to 20 are African American male.

in this biplot, it is hard to distinguish the extent of racial discrimination by those host characteristics which is consistent with the results using regression.

Figure 5: PCA of Proportion of Positive Responses by Other Host Characteristics and Guest ID



4.3 Estimate Expected Opportunity Cost

Another interesting topic is how much the expected opportunity cost of racial discrimination is. If this cost is pretty low, racial discrimination is more likely to happen since individuals are rational when making decisions which is a well-known cornerstone of economics.

To estimate this expected cost, since we have the information whether the property remained empty but rejected our application, we can use classification methods including logistic regression and linear discriminant analysis to predict the likelihood it's filled up in September, which can be denoted as p_{filled}. Airbnb charges each host a fee equal to 3 percent of the price. So the total hosts can earn from the listing is 97 percent of the price[4]. For each who rejected the application from African American guests, we can calculate the expected opportunity cost for one host by multiplying $(1 - p_{filled})$ and the money he can earn if the application were accepted.

[4] If there is cleaning fee, we add it to the listing price.

We collect all the cases that hosts rejected African American guest (Of course, not all the hosts made their decisions due to discrimination, but the cost of this action is also the cost for true racists) and separate them to two parts, 70% for raining and the other 30% for testing. After omitting the observation with missing value, there are totally 519 cases and 340 properties were not available. To predict the probability of property being filled up, both host characteristics and location characteristics are included. Thus, we have 12 categorical variables and only 5 quantitative ones.

Here, we use four classification techniques, including logistic regression, linear discriminative analysis (LDA), quadratic discriminative analysis (QDA) and random forest.

Logistic regression is one of the generalized linear regression model (GLM) and it is a widely used classification method in many areas. To avoid overfitting, we also use ridge or lasso logistic regression, which will add a penalty term on the complexity of model. Also, lasso can be regarded as a feature selection method. Subset selection is another feature selection method we usually use in linear regression, which can help us select significant features. When using logistic regression after subset selection, we find that shared properties and properties with high price are more likely to be vacant.

Both LDA and QDA are generative models. Generative models estimate class densities $P(X_i|Y_i = k)$ and priors $P(Y_i = k)$, and then choose the k maximizes $P(Y_i = k|X_i)$ by Bayes rule $P(Y_i = k|X_i) \propto P(X_i|Y_i = k)P(Y_i = k)$. LDA assumes that the class densities are normally distributed and have the same covariance matrix while the assumptions of QDA are more general. However, QDA has more parameters which may lead to overfitting problem. Here, most of our variables are two-level categorical variables. Therefore, those two methods may not be the best choice since the normality assumption fails.

Random forest is an ensemble learning method for classification. In random forest, we grow multitude of decision trees instead of just constructing one single tree. Random forest could correct the overfitting problem in a single tree method. It might be a useful approach in this case.

Table 8: Test Error of Classification and Median of Expected Opportunity Cost

Classification Methods	Test Error	Median of Expected Opportunity Cost
Logistic Regression	0.3333	44.71
Ridge Logistic Regression	0.3333	48.50
Lasso Logistic Regression	0.3462	45.99
Logistic Regression (after subset selection)	0.3526	44.67
LDA	0.3269	44.49
QDA	0.4359	21.09
Random Forest	0.3013	37.09

Table 8 shows the results of those classification models. Random forest has the lowest test error rate as we expected while the test error of QDA is the highest due to overfitting. The results of logistic regressions are not too bad. The median of expected opportunity cost of racial discrimination is roughly between 35 and 50 dollars which is much lower than 60 to 100 dollars as Edelman et al. (2017) stated. But what we calculate is just based on the revenue hosts can earn. The actual cost should be even lower since the host will spend some money on cleaning if the property were rent out. The low cost is not a good signal in this sharing economy marketplace and racial discrimination is nearly costless. This might be another reason, besides the government regulation, why we can see many racial discrimination cases in such markets.

5. Discussion

In this research, we reinvestigate racial discrimination in the sharing economy based on a field experiment on Airbnb by improving methods of Edelman et al. (2017) and using Correspondence Analysis and PCA to examine our findings. We find that racial discrimination is robust across hosts with different host characteristics and location characteristics. However, interestingly, African American female hosts are more willing to accept African American female guests. This might be related to homophily but in this case homophily only plays a role when hosts and guests are female.

We also find that the extent of racial discrimination will be decreased if hosts have some prior experience with African American guests. This may be an evidence of statistical discrimination, which means hosts reject African American guests based on their prior belief that white guests are more reliable. This finding may be helpful to Airbnb's market designers.

Robust test by changing response variable from "Yes" to "No" is conducted since we have more than two levels of hosts' responses. Fortunately, our conclusions are robust.

Besides, we predict the probability of house being filled up for those who rejected our African American guests using several classification methods, including logistic regression, LDA, QDA and random forest, and then estimate the expected opportunity cost of racial discrimination. It is quite lower than the one Edelman et al. (2017) stated. Since it is easier for hosts to find guests in Airbnb than in the traditional house rental market, the low cost is reasonable. The costless racial discrimination can be an important reason accounting for the high frequency of racial discrimination in the sharing economics markets.

More works have been done but we do not get useful results. Maybe there are other methods we can use for this case. In Section 3, we also try to use MCA (Multiple Correspondence Analysis), which is a data analysis technique for nominal categorical data, to detect and represent underlying structures in our whole dataset. However, the variation explained by each component is very close to each other. This might be caused by the fact that most of categorical variables we have are uncorrelated to others. In Section 4, we also try to use PCA and nonlinear PCA to reduce the dimension. However, it does not work well. The first two components of PCA can only explain less than 60 percent of variance while we just have 5 quantitative variables. As for nonlinear PCA, the capability of Component 1 is extremely high, but it seems to be useless for separating our observations, which is showed in Appendix Figure A1. Besides, one might notice that our prediction in Table 8 is not so accurate. But for rough estimation, we believe it is good enough.

6. Appendix

Table A1: Racial Discrimination by City

Response Variable: Yes	All cities	Baltimore	Dallas	Los Angeles	St. Louis	Washington, DC
Whiteness of Name	0.36***	0.33***	0.35***	0.45**	0.36***	0.34***
	(0.08)	(0.08)	(0.09)	(0.15)	(0.08)	(0.08)
City		0.29*	0.19	-0.00	0.08	-0.12
		(0.11)	(0.13)	(0.10)	(0.20)	(0.14)
City * Guest is African American		-0.52*	-0.09	0.12	0.09	-0.06
		(0.22)	(0.18)	(0.15)	(0.28)	(0.21)
Constant	-0.39***	-0.38***	-0.40***	-0.48***	-0.40***	-0.35***
	(0.06)	(0.06)	(0.06)	(0.11)	(0.06)	(0.05)

Table A2: Racial Discrimination by Previous Experience

Response Variable: Yes			
Whiteness of Name	0.42***	0.34***	0.43***
	(0.09)	(0.08)	(0.09)
Number of past guests who are African American	0.31***		
	(0.05)		
Number of past guests who are African American * Guest is African American	0.12		
	(0.07)		
Proportion of past guests who are African American		0.77	
		(0.41)	
Proportion of past guests who are African American * Guest is African American		-0.45	
		(0.63)	
Host has at least one review from an African American guest			0.39***
			(0.06)
Host has at least one review from an African American guest * Guest is African American			0.22*
			(0.10)
Constant	-0.57***	-0.40***	-0.58***
	(0.07)	(0.06)	(0.07)
Observations	6196	6196	6196

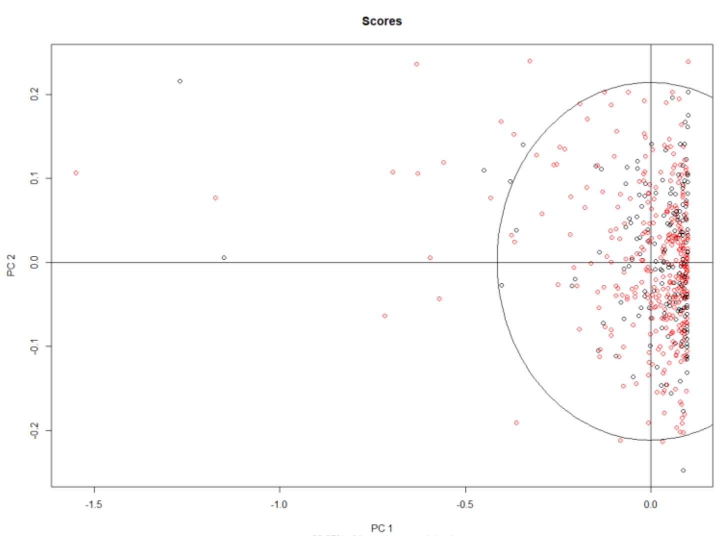
Table A3: Ratio of Acceptance among Hosts with Different Characteristics

Guest ID	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12
1	0.5	0.47	0.49	0.5	0.43	0.56	0.47	0.52	0.45	0.58	0.5	0.48
2	0.55	0.58	0.52	0.65	0.46	0.66	0.57	0.55	0.51	0.68	0.59	0.54
3	0.48	0.54	0.5	0.52	0.43	0.6	0.52	0.48	0.5	0.52	0.54	0.47
4	0.48	0.48	0.46	0.54	0.42	0.55	0.46	0.53	0.45	0.55	0.49	0.48
5	0.47	0.52	0.47	0.54	0.4	0.57	0.51	0.47	0.46	0.58	0.55	0.44
6	0.41	0.42	0.38	0.53	0.4	0.43	0.46	0.35	0.4	0.45	0.51	0.31

7	0.44	0.48	0.42	0.53	0.42	0.5	0.46	0.45	0.44	0.5	0.49	0.43
8	0.53	0.44	0.44	0.58	0.37	0.6	0.52	0.42	0.45	0.57	0.49	0.48
9	0.47	0.43	0.39	0.53	0.43	0.47	0.44	0.45	0.43	0.49	0.48	0.42
10	0.45	0.43	0.42	0.48	0.37	0.52	0.46	0.39	0.41	0.54	0.44	0.44
11	0.4	0.46	0.4	0.49	0.36	0.49	0.44	0.4	0.36	0.63	0.48	0.38
12	0.37	0.35	0.31	0.47	0.33	0.38	0.38	0.34	0.34	0.42	0.4	0.33
13	0.38	0.49	0.4	0.48	0.39	0.48	0.41	0.47	0.37	0.57	0.53	0.34
14	0.47	0.48	0.43	0.56	0.4	0.56	0.49	0.45	0.44	0.6	0.49	0.46
15	0.43	0.37	0.38	0.44	0.3	0.5	0.42	0.36	0.36	0.48	0.42	0.38
16	0.37	0.4	0.35	0.45	0.27	0.49	0.39	0.37	0.34	0.49	0.41	0.36
17	0.35	0.33	0.31	0.39	0.28	0.4	0.4	0.24	0.28	0.53	0.37	0.31
18	0.42	0.31	0.36	0.38	0.3	0.45	0.39	0.32	0.35	0.4	0.35	0.38
19	0.4	0.31	0.31	0.49	0.26	0.48	0.34	0.4	0.34	0.44	0.37	0.36
20	0.41	0.35	0.38	0.41	0.32	0.44	0.41	0.34	0.33	0.51	0.4	0.38

Notes: V1: Ratio of Acceptance for Private Property; V2: Ratio of Acceptance for Shared Property; V3: Ratio of Acceptance for Hosts with only one listing; V4: Ratio of Acceptance for Hosts with multiple listings; V5: Ratio of Acceptance for Hosts with limited reviews; V6: Ratio of Acceptance for Hosts with more than ten reviews; V7: Ratio of Acceptance for Hosts who do not look young (here researchers employed Mechanical Turk workers to roughly categorized the age of hosts); V8: Ratio of Acceptance for Hosts who looks young; V9: Ratio of Acceptance for Hosts who have never had an African American guest; V10: Ratio of Acceptance for Hosts who have at least one African American guest; V11: Ratio of Acceptance for property with a price lower than median in census tract; V12: Ratio of Acceptance for property with a price higher than median in census tract.

Figure A1: Results of Nonlinear PCA



7. References

[1] Alan Julian Izenman. 2006. Modern Multivariate Statistical Techniques: Regression, Classification, and Manifold Learning

[2] Bertrand, Marianne, and Sendhil Mullainathan. 2004. “Are Emily and Greg More Employable Than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination.” American Economic Review 94 (4): 991–1013.

[3] Edelman, Benjamin, Michael Luca, and Dan Svirsky. 2017. “Racial discrimination in the sharing economy: Evidence from a field experiment.” American Economic Journal: Applied Economics 9(2) 1-22.

[4] Ewens, Michael, Bryan Tomlin, and Liang Choon Wang. 2014. “Statistical discrimination or prejudice? a large sample field experiment.” Review of Economics and Statistics 96(1) 119-134.

[5] Ge, Yanbo, Christopher R. Knittel, Don MacKenzie, and Stephen Zoepf. 2016. “Racial and gender discrimination in transportation network companies.” Working paper, University of Washington.

[6] Ruomeng Cui, Jun Li, and Dennis J. Zhang. 2016. “Discrimination with Incomplete Information in the Sharing Economy: Evidence from Field Experiments on Airbnb.” Working paper, Available at SSRN: https://ssrn.com/abstract=2882982