# MACSS 30000: Final Paper

Ari Boyarsky

aboyarsky@uchicago.edu

December 6, 2017

## 1 Introduction

Online marketplaces have become highly pervasive in the past decade. Websites such as eBay, Alibaba, and others represent excellent sources for economist to study otherwise latent phenomena. Studies of these platforms allow researchers to get at a variety of market phenomena such as auctions, ecommerce and more (Kauffman et al. 2010; Xiong 2010). Airbnb, an online market that allows users to rent homes for short stays, offers researchers similar benefits. Discrimination has been a systemic issue in the labor market. While economist and social scientists have attempted to address the issue with a variety of strategies creating effective policy is difficult due to the scarcity of data available. However, Airbnbs system of renting from other individuals, many of whom must identify themselves and provide a photo allows researchers a unique opportunity to gather and analyze data to objectively identify the extent of racial discrimination. Additionally, having access to this data allows researchers to better propose effective solutions to this issue. Lucas Edelmans and his colleagues studies in 2014 and 2017 attempt to utilize data from Airbnb to shed light on these phenomena. This paper will analyze Edelmans research designs to showcase how digitally enhanced research can provide a clearer understanding of otherwise opaque phenomena. We will use the theoretical framework set out by Matthew Salganik (2017) to systematically evaluate these research choices. Additionally, we propose another type of digitally enhanced research design using a survey that can further these findings.

## 2 Edelman and Luca (2014)

Edelmans and Lucas 2014 paper on discrimination on Airbnb shiwcases the power of digitally enhanced research designs. The paper uses observational data to find that a significant difference in rental price between black and non-black hosts. This result implies a significant disparity in demand markets for black and non-black hosts.

## 2.1 Research Design

Edelman and Luca conducted this study utilizing a data set of photos of New York landlords on Airbnb and the respective rental prices for their units. Additionally, the collected data on the rental unit itself. To adjust for quality of unit, the authors sent the unit data to Amazon Mechanical Turk where Turkers were asked to rate the quality of each rental property. Thus, the researchers were able to compare rental price across similar level of quality. Additionally, to identify the race of each landlord, Mechanical Turk users were also asked to categorize the race of each landlord photo. Finally, based on this data researchers were able to determine that black landlords were charging 12% less for a similar unit. This implies that they were also earning 12% less.

This research design utilizes several techniques to make it digitally enhanced. First, the authors are using Mechanical Turk (MTurk) to enhance their data. MTurk is a digital platform that allows researchers to pay MTurk users, turkers, to assist in projects by completing tasks. These may be experimental in nature, survey based, or they may be data generating processes such as this. In fact, Turkers here are being used as a classification task. Next, the researchers are also using some of the big data and digitally enhanced characteristics of observational data as laid out by Salganik (2017). While the data used by Edelman and Luca may not be big, it certainly has some characteristics of big data. First, it is non-reactive as Airbnb users are not aware of the experiment they cannot change their behavior. This is key when studying a socially biased and controversial topic such as racial discrimination. Additionally, it is always on, at least up until the data is collected. This means that the data is longitudinal so if Edelman and Luca want they can track how prices change over time as well. This is important since rental prices will change with certain dates and times. Additionally, as with other digital data sources there are also characteristics that may be considered drawbacks. First, the data is dirty, hence it had to be sent to MTurk to clean (i.e. identify race and unit quality). It may be inaccessible since it requires working with Airbnb or scrapping the site. Since the data is based on New York it may not be representative of a pervasive trend. Finally, the data is sensitive since it requires further publicizing some landlords data such as rental unit information and photo. Thus, Edelman and Luca employ Slaganiks perspectives on observational data to digitally enhance their research design.

## 2.2 Efficacy of Design

Based on our discussion of research design we cane easily pinpoint the costs and benefits of this research design. Clearly, the benefits are the fact that this data is always on and non-reactive. Non-reactive is especially key since if this were not true landlords, Airbnb, or other stakeholder would be able to react to the survey to hide their bias, in this case discriminatory preferences. Thus, the non-reactivity of this data is vital to this experiment. The Always on characteristic is also important as it means we have longitudinal data. This allows the researchers to use more advanced statistical methodology in order to make casual claims as to price effects. That is, that discrimination against black landlords causes a significant decrease in revenue for those same landlords.

However, there also downsides to this research design. First, the data in incomplete as it is not easily verifiable that the data set includes all New York landlords on Airbnb. Also since not all fields on Airbnb are required it is possible critical information is lost in the rating on MTurk. Also, the data is dirty in the sense that it requires techniques to improve its quality. One such adjustment is the use of MTurk for enhancing the data also increases a possibility of bias or somewhat inaccurate measurements that could harm the results. Additionally, the data is not easily accessible an requires a fair amount of technical know how to replicate. The data may also drift over time, as prices increase and decrease over dates (i.e. holidays). Finally, the data is sensitive as it requires that the researchers use a fair bit of personal information such as landlord rental property info and photos to conduct the experiment. All this means that a fair bit of assumptions was made in arriving at the result. Thus, there are drawbacks to the design.

Overall, I do think the design is effective as evidenced by the statistically significant results. However, I would be hesitant to go so far as to say that there is a clear causal link. As there may be other variables that explain the variance in price effects.

# 3 Edelman et al. (2017)

Following the work of the 2014 paper, Edelman et al. (2017) again study the impact of discrimination on the Airbnb marketplace. However, while the 2014 paper studied the effect of discrimination on the Airbnb demand market. That is, how discrimination against black landlords decreased the price of black landlord rental units. This paper focuses on supply side discrimination. That is, since Airbnb landlords must accept the application of possible short-term tenants, how discrimination on the part of landlords impacts which applications they accept. Indeed, Edelman et al. find that applications from users with distinctly African American names are 16% less likely to be accepted.

## 3.1 Research Design

In this study, Edelman et al. scraped data from Airbnb in 5 major cities. This included rental unit data and landlord data. They then used computational techniques such as MTurk and an image classification tool Face++ to categorize hosts into age, gender, and racial categories. They then created 20 Airbnb accounts with names that were based on the conditional probability that the individual was a certain race or gender given the name. Using this technique, they created 10 black 10 non-black. And, 5 men and 5 women in each group. Finally, over a period 3 week they sent 6,400 messages to hosts inquiring about rental availability. The results showed that there was evidence for discrimination. However, the authors also suggest the type of Airbnb also matters. That is, whether the host is renting a private room, shared room, or entire (private) unit.

Overall the design employs several computational research designs. Primarily, the researchers are collecting bigish data using data scraping and an experimental design. They

have about 6,000 observations which is large though perhaps short of the 100,000s of observations characteristic of many big data sets. Additionally, they collect this data using computational techniques such as web craping. They also use MTurk to enhance this data by identifying race. In addition to MTurk they also utilize Face++ which uses machine learning classification techniques to identify the race, gender, and age of an individual given a face. Finally, they are using an experimental design in which they leverage a digital platform (Airbnb) to send individuals messages and analyze the resulting data. This is very much in line with Salganik (2017) argument that some digitally enhanced experiments may utilize already existing environments. In this case the researchers are utilizing the online environment created by Airbnb.

## 3.2   Efficacy of Design

The power of this design comes from its ability to exercise a randomized control trial (RCT). This is of course considered the gold standard for testing causal effect. Indeed, the researchers are able to randomly assign Airbnb hoists to treatment groups in which they must choose to accept or decline a individual given their name. Of course, the name is varied specifically so as to give the impression that the individual asking to rent the property is black, non-black, male or female. This allows the researchers to test the effect of each name on acceptance rate. This is furthered since the name assignment is random and thus independent of any other characteristics.

However, there are some downsides to this research strategy. As per Salganiks advice one major issue with experimental designs are their ethicality. Salganik argues that researchers should replace invasive methods if possible, refine treatment to make it harmless and reduce the number of participants. The first ethical challenge that arose in this study was that Airbnb was blocking their scraping efforts. Wisely, Edelman et al. choose to stop after they collected 5 cities worth of data instead of the planned twenty. This shows restraint for sensitive possibly proprietary data. Additionally, it shows an effort tor reduce any possible harm. Of course, they did not replace the experimental methods and they did have to look at some sensitive data such as landlord rental information and profile information. Thus, it is possible that there was an ethical issue here. Luckily, Edelman et al. did attain IRB approval as well. Thus, it does seem that most of the ethical issues were addressed. The actual content of the experiment seems strong. However, one possible source of error may be that they validated their list of black and non0-blakc names using a survey where participants only 3 seconds to answer. While, tis was done to encourage an instinctive response based on biases, it may also not be enough time to register the name nd that could have led to some errors. Also, Airbnb users have more than 3 seconds to process a name and so that disparity could have some effect. Overall, the experimental design does appear to be effective and allow Edelman et al. to make a causal claim that there is discrimination in the Airbnb marketplace.

# 4 Value Added of Observational and Experimental Design

Both studies conducted here by Edelman provide quite a bit of information for readers. Hence, there is a significant value added in both research projects. First, the initial 2014 research design using observation data is target towards detecting demand side discriminatory practices that have a supply side price effect. While the alter 2017 experimental design is targeted at detecting supply side discriminatory practices that impact potential tenants. That is, that the 2014 paper explores how tenants discriminate while the 2017 paper explores how Airbnb hosts discriminate.

However, there are other benefits to the implementation of both research designs. First, while the 2014 paper does find significant statistical evidence supporting discrimination in the Airbnb marketplace the use of observational data means that there may be some omitted variable bias or other confounding effects that are responsible for the correlation. While, we can test this using causal inference techniques such as IV or RDD methods, we could also attempt to design a randomized control trial to further assess the evidence of racial discrimination in the Airbnb marketplace. Hence, the 2017 paper does exactly this. The paper utilizes an RCT to evaluate the impact racialized name on Airbnb host acceptance rates. Thus, it allows us to say with greater certainty that racial discrimination is at play in Airbnb. We also see the effect of names and pictures which is key since Airbnb does not neccarily require pictures. Additionally, this study (2017) tests other locations not including New York which means that we have more evidence that this trend is pervasive and not location based.

Furthermore, both studies provide us with slightly different data. This is very important if we want to study the implications of racial discrimination and possibly provide solutions to these problems. Specifically, the combination of experiments provides us with supply-side and demand-side data so that we can tackle these challenges in both spheres. Thus, the experiments give us stronger evidence in combination and allow us to assess the effect from two different market perspectives.

# 5 Digital Survey Based Alternative

A further research design that could be employed to study this phenomena in more depth is a survey based method. A survey approach would strive to show that there exists racial discrimination in the Airbnb marketplace. Indeed, we would be looking for some difference that implies the effect of racial discrimination between two groups, black and non-black.

The survey we approach we use mixes survey based research with experimental research. Our approach may be used to approach both supply-side and demand-side effects. We run this experiment on MTurk as we can use it for quick results for a relatively low cost. Additionally, while MTurk users may be baised one way or another we can use statistical adjustment such as post stratification techniques to adjust for this. Indeed, research has

shown that MTurk surveys that are paired with statistical adjustment are generally very representative (Berinsky et al. 2012; Kennedy et al. 2016). First, we present MTurkers a page asking them for demographic date (race, age, education level, etc.). Next we show them instructions that tell them we are studying what makes people choose one Airbnb over another (we do not mention race here). To ensure they read the instruction we have a multiple-choice question below where they are meant to answer correctly only if they read the instructions (thi information will be buried in the instructions):

Instructions. Please Answer the Below Question with C regardless of the correct answer.

1. 2+2 = ?   A. 1   B. 4   C. 5   D. 6

Next, we show them a fake Airbnb listing. They must decide if they will accept it or not. We do not show them any pictures. But we do give them a previous review and some other data. We also give them the name of the Airbnb host and some very basic data (age, etc.). We do not tell the host race. But, we do randomly vary the name based on our list of male and female black and non-black names from the 2017 paper. Finally, we present them a page which asks them to rank the factors that led to their decision. We also ask them to write one-two sentences in an open-ended saying why they made their decision. Additionally, we ask if they use Airbnb in real life or not. This may be useful is we see a major deviation in the non-Airbnb vs Airbnb group.

An example of the Airbnb listing may look like this:

Host Name: John Smith
Location: Manhattan (Greenwich Village), New York, New York
Bedrooms: 1 (Private, 1 bed)
Sample Review: I loved my stay here.

The above is just an example more data could be included based on previous research on Airbnb. Notice this evaluates demand-side biases, to do the same for supply-side we simply give them a profile of an Airbnb user varying the name based on random assignment of race and gender.

The last page and instructions act as a throw-off so that survey participants do not know we are testing for racial bias. Finally, we collect this data, perhaps 1000 or so samples. We weight the responses based on post-stratification using the US census for the basis and we study the effects. We can also use the response on the last page to evaluate some of the decisions, and perhaps control or specific details. The open ended may also help get at latent trends in the decision-making process.

This design avoids some of the sensitivity of the initial two designs as we use fake Airbnb profile, so our data is not as sensitive. Additionally, we are able to survey a greater number of users and by asking for the open-ended question and other influential factors we can also attempt to better find confounding factors and control for them. Additionally, we can easily

switch between demand-side and supply-side perspectives by simply changing the phrasing of the questions and data presented. Thus, we can take advantage of some of Slaganiks positive characteristics of surveys such as non-probability surveys and linking survey data to design an effective experiment to compliment the Edelman 2014 and 2017 studies.

Of course, the biggest concern with the survey is that participants guess our purpose in studying racial bias through varying names. Thus, we will limit each MTurker to only one hit and we will try to throw them off with the instructions and the final page which asks other characteristics. Additionally, since this is outside the Airbnb environment there may be other problems with how users react. Finally, since we cannot verify the data users provide us we must take it with some skepticism. Still, if this technique results in positive effects, then in conjunction with the previous observational and experimental works we would have a strong case for racial discrimination in the Airbnb marketplace.

# References

[1] Alessandro Acquisti, Curtis R. Taylor, and Liad Wagman. The economics of privacy. 2016.

[2] Mengti Xiong. Analysis of chinese e-commerce development. 2010.

[3] Robert J. Kauffman, Hsiangchu Lai, and Chao-Tsung Ho. Incentive mechanisms, fairness and participation in online group-buying auctions. *Electronic Commerce Research and Applications*, 9(3):249–262, 2010.

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

[5] Adam J. Berinsky, Gregory A. Huber, and Gabriel S. Lenz. Evaluating online labor markets for experimental research: Amazon. com's mechanical turk. *Political Analysis*, 20(3):351–368, 2012.

[6] Courtney Kennedy, Andrew Mercer, Scott Keeter, and Nick Hatley. Evaluating online nonprobability surveys. 2016.

[7] Matthew J. Salganik. *Bit by Bit: Social Research in the Digital Age.* Princeton University Press, 2017.