Causal Analysis of Discrimination in Online Rental Markets: A Replication Study

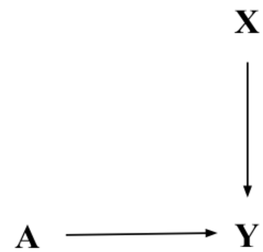
ABSTRACT

This report presents a replication analysis of discrimination patterns in AirBnB rental markets using non-parametric causal modeling techniques rather than traditional parametric regression-based approaches. Our analysis estimates both average treatment effects (ATE) and conditional average treatment effects (CATE) to understand how racial discrimination varies across different host and neighborhood characteristics. The non-parametric approach reveals an overall discrimination effect of -8.76 (95% CI: [-11.20, -6.32]) percentage points, closely aligning with the original study's findings while providing additional insights into effect heterogeneity across subgroups.

INTRODUCTION

The advent of online platform economies has raised important questions about whether digital marketplaces can reduce historical patterns of discrimination or inadvertently facilitate new forms of bias. The original experimental study of AirBnB discrimination found that applications from guests with distinctively African American names were 16 percent less likely to be accepted compared to identical applications from guests with distinctively white names.

While the original analysis employed traditional regression-based methods to estimate treatment effects, this replication study applies non-parametric causal modeling techniques to provide a more flexible approach to understanding discrimination patterns. As such, we have adopted the following causal structure (we assume no unobserved confounding):



Y = Treatment: Race-sounding name of guest (binary: African American or White)
A = Outcome: Whether the host accepts the booking request (binary: Yes or No)
X = Covariates:

- Host's race
- Host's gender
- Whether the host has multiple listings
- Whether the property is shared
- Whether the property has 10+ reviews
- The logarithm of the listing's price

EXPERIMENT

The original experiment created 20 new AirBnB accounts, identical except for guest name which was assigned according to a survey which identified names as distinctly African American or White. The 20 names were assigned as follows: 5 male African American names, 5 female African American names, 5 male White names, and 5 female White names. Because the treatment was randomly assigned, the overlap condition is satisfied. Each host was randomly assigned one of the accounts and each request was made for the same weekend. Host's responses were manually categorized and then simplified as "Yes" or "No".

METHODOLOGY

We decided to look at the same treatment effects the original authors were interested in to compare our results to the original analysis. We started by looking at the average treatment effect (ATE) to identify whether they data suggested an overall trend towards discrimination while controlling for confounders:

$$\mathbf{ATE} = \mathbb{E}[\mathbb{E}[Y|A=1,X] - \mathbb{E}[Y|A=0,X]]$$

Additionally, we wanted to determine how conditioning on certain host and neighborhood characteristics would affect our estimate of the ATE. Therefore, we calculated the conditional average treatment effect (CATE) for these variables, including host's race, whether a host has had African American guests in the past, and the diversity of the neighborhood (V represents these variables):

$$\mathbf{CATE} = \mathbb{E}[\mathbb{E}[Y|A=1,X] - \mathbb{E}[Y|A=0,X] \mid V]$$

ANALYSIS

To estimate each treatment effect, we first estimated nuisance functions using Random Forest classifiers with cross fitting. Due to random assignment of the treatment and it not being confounded by X, we estimated the propensity score with 0.5. After fitting each model, we performed sanity checks to make sure each model improved test error. We then estimated the treatment effects using the nuisance functions. To condition on continuous variables, we experimented with different methods of splitting the data into groups in order to calculate the treatment effect conditional on each group separately. These results can be found in tables in the appendix.

Overall Discrimination Effect

Our analysis revealed a substantial and statistically significant average treatment effect of -8.59 percentage points (95% CI: -10.98% to -6.20%). This indicates that guests with distinctively African American names face systematic discrimination in AirBnB acceptance rates, with nearly 9 percentage points lower probability of having their booking requests accepted compared to identical guests with white names.

Host Race and Discrimination Patterns

The analysis also reveals important differences in how Black and white hosts respond to guest race signals. White hosts demonstrate clear and statistically significant discrimination (-8.85 percentage points, 95% CI: -11.39% to -6.31%), while discrimination among Black hosts, though negative in magnitude (-7.76 percentage points) in its point estimate, is not statistically significant (95% CI: -16.59% to 1.08%). This pattern suggests that racial discrimination in the platform operates primarily through same-race bias, with white hosts systematically favoring white guests, while Black hosts show no statistically detectable preference based on guest race.

The Critical Role of Host Experience

Another finding concerns how prior experience with Black guests shapes discriminatory behavior. Hosts who have never hosted Black guests exhibit severe discrimination (-10.32 percentage points, 95% CI: -13.18% to -7.46%), while hosts with prior Black guest experience show much weaker and statistically insignificant discrimination (-4.42 percentage points, 95% CI: -9.04% to 0.20%). This 6-percentage point difference suggests that direct contact and positive experiences may substantially reduce discriminatory attitudes, supporting contact theory predictions about intergroup relations.

Complex Relationship with Guest Composition History

The relationship between hosts' historical proportion of Black guests and current discrimination reveals unexpected complexity. Hosts with no Black guest history show strong discrimination (-10.32 percentage points), consistent with the experience findings. However, among hosts with some Black guest experience, the pattern becomes non-linear and depends critically on the analytical threshold chosen.

Using the narrower threshold (0-10% Black guests), hosts with minimal Black guest experience show no significant discrimination (-0.73 percentage points, 95% CI: -6.21% to 4.74%), while those with more substantial experience (>10%) actually exhibit the strongest discrimination (-13.19 percentage points, 95% CI: -21.62% to -4.75%). This counterintuitive finding suggests that some hosts may become more discriminating after negative experiences or may represent a select group of hosts who continue hosting despite harboring strong preferences.

Using the broader threshold (0-30% Black guests) produces different results, with the middle group showing slight positive treatment (0.25 percentage points, not significant) and high-experience hosts showing moderate discrimination (-7.13 percentage points, 95% CI: -12.82% to -1.45%). This threshold sensitivity indicates that the relationship between host experience and discrimination is complex and may involve selection effects or threshold effects that require careful interpretation.

Neighborhood Context Effects

Neighborhood racial composition shows consistent but nuanced effects across different analytical approaches. The quantile-based analysis reveals that discrimination is strongest in both low-diversity (bottom third: -9.76 percentage points) and high-diversity (top third: -9.87 percentage points) neighborhoods, with somewhat weaker effects in medium-diversity areas (-6.94 percentage points). All effects are statistically significant, suggesting that neighborhood context moderates but does not eliminate discrimination.

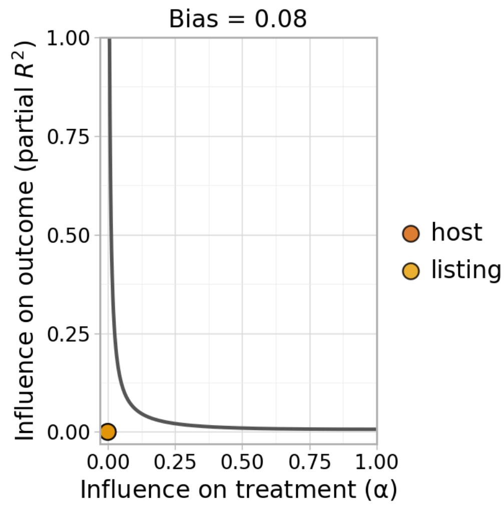
The threshold-based neighborhood analysis provides additional insight, showing strongest discrimination in moderately diverse neighborhoods (10-30% Black: -12.09 percentage points, 95% CI: -18.76% to -7.03%), moderate discrimination in low-diversity areas (<10%: -8.31 percentage points), and weaker discrimination in high-diversity neighborhoods (>30%: -6.10 percentage points, 95% CI: -12.45% to 0.25%, marginally significant). This pattern suggests that discrimination may be most pronounced in moderately integrated neighborhoods, possibly reflecting heightened racial salience in these contexts.

Sensitivity Analysis

Because this is a randomized controlled trial and the treatment is randomly assigned, there should be no unobserved confounders for the treatment. Therefore, the sensitivity analysis that addresses the issue of unobserved confounders is not necessary for this study.

We still made an Austen plot to confirm this. The observed covariates are divided into two categories: host characteristics (gender, race, whether the host has multiple listings) and listing characteristics (whether the property is shared, whether the property has 10+ reviews, the logarithm of the price). The target bias is set as 0.08, as the estimated ATE is -.0859 and this level of bias would undermine the qualitative conclusion of the study.

From the Austen plot, it is shown that both host characteristics and listing characteristics are not strong enough to cause the target bias. We believe that it is unlikely for other potential confounders to have a stronger effect that undermines the causal claim.



DISCUSSION

Comparison with Original Findings

Our non-parametric analysis strongly corroborates the original study's central finding, estimating an 8.59 percentage point discrimination effect that closely matches their 8 percentage point gap. This consistency across methodological approaches strengthens confidence in the underlying causal relationship and demonstrates that the discrimination effect is not an artifact of parametric assumptions.

However, our CATE analysis reveals important heterogeneity obscured in the original parametric framework. While the original study found "noisy" and insignificant interactions between host and guest race, our approach shows that white hosts exhibit clear discrimination (-8.85%, significant) while Black hosts show no statistically detectable discrimination. This suggests discrimination operates primarily through in-group favoritism by white hosts rather than universal prejudice.

The most striking enhancement concerns host experience with African American guests. The original study found that discrimination "drops sharply" among experienced hosts but could only reject zero difference at the 10% level. Our analysis provides much stronger evidence: hosts without prior Black guest experience show severe discrimination (-10.32 percentage points, highly significant), while experienced hosts show no significant discrimination (-4.42 percentage points, 95% CI includes zero). This 6 percentage point difference represents a substantively meaningful reduction.

Regarding neighborhood effects, where the original study found no variation with local African American population, our analysis suggests more complex patterns, though the findings are sensitive to analytical choices. While low diversity neighborhoods consistently show discrimination across different splitting methods, the results for medium and high diversity areas vary considerably depending on how neighborhoods are categorized. Some specifications suggest moderately diverse areas exhibit the strongest discrimination (-12.09 percentage points), but this pattern is not robust across all analytical approaches, indicating that neighborhood effects may be more nuanced than initially apparent.

The enhanced precision and additional insights likely reflect our non-parametric approach's ability to control for confounding variables more flexibly while avoiding restrictive functional form assumptions.

CONCLUSION

This non-parametric replication confirms substantial discrimination against African American guests in AirBnB markets while revealing important heterogeneity in discriminatory behavior. The finding that discrimination is concentrated among white hosts without prior African American guest experience provides strong empirical support for contact theory. The complex relationship between neighborhood diversity and discrimination indicates that local context matters for discriminatory behavior.

CITATIONS

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.

APPENDIX

Point Estimate	95% CI Lower	95% CI Upper
-0.0859	-0.1098	-0.0620

Table 1: Point estimate and 95% confidence interval for ATE

Host Race	Point Estimate	95% CI Lower	95% CI Upper
Black	-0.0776	-0.1659	0.0108
White	-0.0885	-0.1139	-0.0631

Table 2: Point estimate and 95% confidence interval for CATE conditioned on host's race

Host Experience	Point Estimate	95% CI Lower	95% CI Upper
Has had Black guests	-0.0442	-0.0904	0.0020
Has not had Black guests	-0.1032	-0.1318	-0.0746

Table 3: Point estimate and 95% confidence interval for CATE conditioned on whether the host has previously hosted Black guests

Host Group (by proportion of Black guests)	Proportion Range	Point Estimate	95% CI Lower	95% CI Upper
Low	0	-0.1032	-0.1318	-0.0746
Medium	(0-0.1]	-0.0073	-0.0621	0.0474
High	> 0.1	-0.1319	-0.2162	-0.0475

Table 4: Point estimate and 95% confidence interval for CATE conditioned on proportion of host's previous guests who were Black split into groups

Host Group (by proportion of Black guests)	Proportion Range	Point Estimate	95% CI Lower	95% CI Upper
Low	0	-0.1032	-0.1318	-0.0746
Medium	(0-0.3]	0.0025	-0.0755	0.0805
High	> 0.3	-0.0713	-0.1282	-0.0145

Table 5: Point estimate and 95% confidence interval for CATE conditioned on proportion of host's previous guests who were Black split into groups (alternative split)

Neighborhood Group	Proportion Range	Point Estimate	95% CI Lower	95% CI Upper
Low Diversity	[0, 0.034]	-0.0976	-0.1397	-0.0555
Medium Diversity	[0.034, 0.094]	-0.0694	-0.1118	-0.0270
High Diversity	[0.095, 0.984]	-0.0987	-0.1409	-0.0565

Table 6: Point estimate and 95% confidence interval for CATE conditioned on proportion of Black residents in AirBnB property's neighborhood (split by quantiles)

Neighborhood Group	Proportion Range	Point Estimate	95% CI Lower	95% CI Upper
Low Diversity	< 0.1	-0.0831	-0.1127	-0.0535
Medium Diversity	[0.1, 0.3]	-0.1209	-0.1876	-0.0703
High Diversity	> 0.3	-0.0610	-0.1245	0.0025

Table 7: Point estimate and 95% confidence interval for CATE conditioned on proportion of Black residents in AirBnB property's neighborhood (split by thresholds)

Neighborhood Group	Proportion Range	Point Estimate	95% CI Lower	95% CI Upper
Low Diversity	[0, 0.326]	-0.0920	-0.1182	-0.0658
Medium Diversity	[0.330, 0.651]	-0.0518	-0.1352	0.0315
High Diversity	[0.656, 0.984]	-0.0797	-0.1897	0.0303

Table 8: Point estimate and 95% confidence interval for CATE conditioned on proportion of Black residents in AirBnB property's neighborhood (split by equal-size bins)