Assessing the Racial Disparities in New York City Policing

Andrea Aveni & Jiajun Song

21 February 2021

Abstract

Several recent studies have shown evidence that police officials in NYC seem to be affected by racial bias when stopping people for random controls. However, the differences in the stopping rate could also be explained by spatial heterogeneity among different precincts or by the presence of disproportionate crime rates for different racial groups. In this study we try to assess the presence of racial disparities and to consider their evolution in time. Our study shows that even with these controls there is strong evidence for a significant bias against the African-Americans, and, to a less extent, against the Hispanics. However, we find some evidence that these discriminatory phenomena have been reducing over the course of the last years.

1 Introduction

Discrimination by public officials is a very well studied and relevant problem. We deal with this subject by considering the controversial "stop and frisk" strategy used by NYC police, consisting in unexpected stops and searches of pedestrian for a large spectrum of possible crimes. This policy has been considered by many citizens detrimental of their rights. We consider, in particular whether there is evidence for disparity in the race-specific stop rates, and whether this disparity can be explained merely by the heterogeneity of race-specific crime rates.

A key statistical problem to correctly address the disparity is taking into account the spatial variation in the intensity of policing. For instance, police is reasonably controlling more assiduously areas with higher crime rates; because of historical phenomena outside the control of the police, these areas may display a larger number of people of a certain ethnicity and this could overall give the impression that this minority is oppressed by the police. In order to deal with this problem we consider random effects for each precinct, in a particular way that allows us to take into account the correlation between near precincts.

Another central problem for interpreting the results is the choice of a baseline to compare rates. Concerning the baseline, we consider two strategies. At first, we considered the product of the hit rate and population as a relative measure of stop counts in a fair setting Alternatively,we used the number of arrests as a proxy for the race-specific crime rates, as suggested in [1]. We further compare the results obtained by using these two baseline measures and we try to provide possible explanations for their disagreement on the exact degree of discrimination.

2 Methods

2.1 Data Set

We considered different datasets. Primarily we considered the report analysis [2] provided by the New York government about the "stop and frisk" strategies performed over the years from 2017 to 2019. In this dataset, we have the individual level information about the stopped and frisked event, such as time, location, and many physical characteristics of the subjects, including their race. We also have a description of the possible law violation for which they were stopped and, for some crimes, whether they were found guilty. In particular there is a column that signals whether the police have found some illegal weapons or drugs and a column that indicates whether the suspect was arrested afterwards.

From a distinct dataset [3] we have obtained demographic information about NYC at the precinct level, including the area, the African-Americans, the Hispanics and the Whites populations. Finally from the New York Open Data [6], we obtained the number of arrests by race and precincts for each of the years considered.

2.2 Modeling Count Data

In trying to answer our research question, we want to model the number of people $S_{r,p,y}$ stopped by the police, of race r, in a certain precinct p and in a specific year y. We decided to consider only three ethnic groups, the African-Americans $r = 1$, the Hispanic or Latino $r = 2$, and the White $r = 3$. This does not comprehend the totality of the population, however the residual groups have a numerically marginal presence in the SQF dataset.[1] As for the time period, we considered only the years $y = 2017, 2018, 2019$ because the precinct division of NYC changed in 2017. Finally, as for the precincts, from a total of 77 regions, we had to remove precinct $p = 22$ corresponding to Central Park; the reason is that in this precinct there are essentially no inhabitants and so, for this precinct there cannot be demographic data.

To model the number of stops $S_{r,p,y}$ we use a Poisson regression, enhanced with random effects $\varepsilon_{r,p,y}$ for each observation to account for the possible presence of overdispersion.

$$\begin{aligned}
&S_{r,p,y} \sim \text{Pois}(A_{r,p,y} \exp(\alpha_r + \gamma_{r,y} + \beta_p + \varepsilon_{r,p,y})) \\
&\alpha \sim \mathcal{N}(0, \sigma_\alpha^2 I),\, \varepsilon \sim \mathcal{N}(0, \sigma_f^2 I) \\
&\beta \sim \mathcal{N}(0, \Sigma_\beta),\, \gamma \sim \mathcal{N}(0, \Sigma_\gamma)
\end{aligned}$$

Here $A_{r,p,y}$ is our baseline for the relative stops count we can expect to observe if there is no discrimination against any ethnic groups. In other words, assuming the absence of racial disparities, the stops count should be proportional to this baseline. This term has a crucial role in determining the interpretation of the results. Indeed, we cannot simply check whether the stop rates are heterogeneous but instead, we need to compare the number of stops with the different representation of each group in the total population, and, more importantly to the participation of each ethnic group in criminality.

With this respect we decided to use two alternative baselines. At first, we used the product of hit rate and population for each precinct and year combination. More specifically, $A_{r,p,y}$ is the population of race r in the precinct p in the year y, times the frequency at which stopped people were arrested, in the same precinct and year. This should in principle be an objective measure of criminality participation since both the hit rate and the population cannot be altered discretionarily by the police. Alternatively we used the previous year arrest count for each precinct and year combination as proposed Gelman et al. in [1]. We further discuss advantages and disadvantages of these two measures in Section 2.4.

The term α_r represents the random effect of each ethnic group on the stop count. The difference in α_r indicates the racial disparities on a log-scale among groups after being adjusted by the baseline and controlled for spatial and year variations.

The term β_p is the random effect for each precinct and it is used to control for heterogeneity among different areas of the city. In order to model the heterogeneity among precincts more realistically, we consider imposing a covariance matrix that takes into account the proximity of the neighborhoods.

The term $\gamma_{r,y}$ is the interaction among races and years. This is a crucial term that allows us to consider the evolution over time of the discrimination patterns. We impose zero correlation among different race groups but an $AR(1)$-like covariance between $\gamma_{r,y}$ and $\gamma_{r,y'}$. These assumptions are based on the fact that

[1] These three groups constitute more than 95% of the total stop and frisk events in most recent years.

discrimination patterns are a phenomenon that evolves slowly in time and therefore it is reasonable to assume there will be strong correlation between successive years.

2.3 Dealing with Spatio-Temporal Data

To construct the covariance of the spatial effects, we used a discretized Gaussian process. In particualr, using the data in [4], we were able to compute the coordinates of the barycenters (cx_p, cy_p) of each precinct p, and then, we computed the distance $d(p, q)$ between two precincts p, q as the Euclidean distance between the corresponding centers. Formally,

$$d(i, j) = \sqrt{(cx_p - cx_q)^2 + (cy_p - cy_q)^2}$$

Given this distance we imposed the covariance between β_p and β_q to be an decreasing exponential function of the distance $d(p, q)$ between the precincts. Therefore we had

$$[\Sigma_\beta]_{p,q} = \sigma_\beta^2 \exp(-\phi d(p, q))$$

Meanwhile, when dealing with temporal data, we imposed the following covariance matrix

$$[\Sigma_\gamma]_{(r,y),(r',y')} = \sigma_\gamma^2 \psi_r^{|y'-y|} \mathbf{1}_{r=r'}$$

We imposed a weakly informative Gamma prior on ϕ and a uniform $[-1, 1]$ prior for ψ_r. Note that both covariance matrices will be guaranteed to be positive-definite in this way.

2.4 Results

The model showed slow but good convergence properties as shown in Appendix Figure 5, and good model fit, shown in Appendix Figure 6. In particular the model is capable of dealing with the over-dispersion present in the data, as indicated by the standardized residuals in Figure 7.

In what follows we will define a quantity $D_{b,y}$ ($D_{h,y}$) to be the observed discrimination against Black (Hispanic) people in year y. This is defined as

$$D_{r,y} = \exp\left((\alpha_r + \gamma_{r,y}) - (\alpha_w + \gamma_{w,y})\right) \quad \text{for} \quad r = b, h$$

This quantity represents how much more likely it was in year y, for people of race r, to be stopped by the police once we have controlled out for site specific, population differential and crime-participation differentials. More in detail, this is the ratio between the multiplicative effect on count of being black over the effect of being white, adjusted for the year y.

As for the evidence of racial disparities, we have observed a significant race effect by using both our baseline measures (see the summary statistics in Table 1); in particular this suggests the presence of discrimination against the African-Americans and, to a less extent, against the Hispanics. Meanwhile, discrimination appears to be always stronger (table 1) when we used the hit rate baseline as we expected, see the next section. We also visualize the density of $D_{b,y}$ and $D_{h,y}$ when we used the arrest count as our baseline measure in Appendix Figure 1.

As for the evolution of discrimination over time, we can consider $D_{r,y}/D_{r,2017}$ for $y = 2018, 2019$ and $r = b, h$. If we look at Appendix, Figure 2 we can see that, overall, discrimination is declining and, in particular, there has been a strong significant reduction from 2017 to 2018, and a more modest reduction from 2017 to 2019.

As for the spacial random effects (Appendix Figure 1), precinct specific differences seem to be relevant to explain the variation observed in the data. We also notice (Appendix Figure 3) that the parameter ϕ linked with the correlation between near precincts is significantly different from zero, indicating that differences in stop rates changes smoothly in space, precisely as, we can imagine, the demographic data.

3 Discussion

With regards with our question of interest we found, with both baselines, that there are significant disparities against the African-Americans and the Hispanics. However, we also found that the discrimination against minority groups has been declining over the last years.

Regarding the two different baseline measures we see that, by using the hit rate times the population we obtain stronger results than by using the arrests count. We can offer two possible explanations for this phenomenon.

First of all, the different stop rates could have an indirect effect on the hit rate. In particular, we can imagine a selection bias that can lead the police to find higher hit rates among those groups that are stopped less frequently[2]. This could lead us to overestimate the white people's participation in crime, and this could in turn, result in an seemingly stronger discrimination when we compare the stop rate with the crime-participation rate. On the other hand we could have an opposite effect when we consider the arrest data. In this case, if there is discrimination against a certain group, this could lead the police to arrest this ethnicity group more frequently. Thus the crime participation of this group will be inflated and evidence for discrimination will be reduced. These considerations motivate us to focus more on the arrest baseline because this is the most averse hypothesis to our conclusion and thus showing discrimination using this measure is more conclusive.

We found that spatial random effects are important for a good model fit (Indeed we found that about 12% of the precincts had significant random effects.). And we also observe a strong correlation between near precincts (see Appendix 3). This implies that the precinct level-subdivision is sufficiently fine for our analysis, even though, it could be an interesting future research direction to imagine further heterogeneity among the precincts by directly using the coordinates of each stop.

A different further question beyond the scope of this project, could instead to try to investigate and explain a possible heterogeneity in the rapidity at which discrimination patterns are declining in different precincts. This, however, would require a larger number of yearly data.

References

[1] Andrew Gelman, Jeffrey Fagan & Alex Kiss (2007) An Analysis of the New York City Police Department's "Stop-and-Frisk" Policy in the Context of Claims of Racial Bias, Journal of the American Statistical Association,

[2] New York Police Department, Stop-Question-Frist data, `https://www1.nyc.gov/site/nypd/stats/reports-analysis/stopfrisk.page`

[3] John Keefe, Demographics data for precincts, `https://johnkeefe.net/nyc-police-precinct-and-census-data`

[4] NYC open data, Precincts Area, `https://data.cityofnewyork.us/Public-Safety/Police-Precincts/78dh-3ptz`

[5] NYC open data, ZIP code areas, `https://data.cccnewyork.org/data/map/66/median-incomes#66/50/6/107/62/a/a`

[6] NYC open data, Historic arrest counts, `https://data.cityofnewyork.us/Public-Safety/NYPD-Arrests-Data-Historic-/8h9b-rp9u`

[7] GPS Converter `https://www.expertgps.com/spcs/New-York-Long-Island-FIPS-3104-NAD27.asp`

[2] And this is indeed the case, since for instance the global hit rate for white people is larger than the one for black people, while, on the contrary, the percentage of black people arrested is larger than the percentage of white people.

Table 1: Inference for $\log D_{r.y}$; posterior mean(posterior standard error) is shown in each cell. In general, the disparities are significant but have greatly decreased compared to 2017.

Quantity	$y = 2017$	$y = 2018$	$y = 2019$
Arrest as baseline			
$\log D_{b,y}$	0.73(0.06)	0.47(0.06)	0.63(0.05)
$\log D_{h,y}$	0.29(0.06)	0.10(0.06)	0.19(0.06)
Hit rate*Pop as baseline			
$\log D_{b,y}$	2.64(0.23)	2.34(0.20)	2.52(0.20)
$\log D_{h,y}$	1.25(0.22)	1.03(0.20)	1.19(0.22)

4 Appendix

4.1 Results

Table 2: Inference for important variables. Note that to avoid identifiablility issues, we restricted α_1 and $\gamma_{r,1}$'s to be zero. The model is run with 2 chains, 5000 iterations and 500 warm-ups per chain.

	mean	se_mean	sd	2.5%	25%	50%	75%	97.5%	n_eff	Rhat
alpha[1]	0.00		0.00	0.00	0.00	0.00	0.00	0.00		
alpha[2]	-0.43	0.01	0.05	-0.54	-0.47	-0.44	-0.40	-0.33	81.33	1.01
alpha[3]	-0.73	0.01	0.06	-0.84	-0.77	-0.73	-0.69	-0.60	126.67	1.01
gamma[1,1]	0.00		0.00	0.00	0.00	0.00	0.00	0.00		
gamma[1,2]	0.01	0.00	0.05	-0.09	-0.02	0.01	0.04	0.11	118.81	1.01
gamma[1,3]	0.44	0.01	0.05	0.35	0.41	0.44	0.48	0.55	90.29	1.02
gamma[2,1]	0.00		0.00	0.00	0.00	0.00	0.00	0.00		
gamma[2,2]	0.08	0.00	0.05	-0.02	0.05	0.08	0.12	0.18	157.44	1.01
gamma[2,3]	0.44	0.00	0.05	0.34	0.40	0.44	0.47	0.55	180.01	1.02
gamma[3,1]	0.00		0.00	0.00	0.00	0.00	0.00	0.00		
gamma[3,2]	0.27	0.01	0.07	0.13	0.22	0.27	0.31	0.39	160.85	1.01
gamma[3,3]	0.54	0.01	0.06	0.40	0.50	0.54	0.58	0.66	168.63	1.02
sigmaAlpha	1.06	0.07	1.25	0.30	0.49	0.73	1.17	3.64	336.91	1.01
sigmaBeta	13.44	1.04	39.58	2.42	4.79	7.69	13.97	52.38	1458.79	1.00
sigmaGamma	0.18	0.01	0.13	0.05	0.10	0.14	0.21	0.54	703.71	1.01
phi	0.86	0.03	0.60	0.10	0.41	0.73	1.16	2.35	444.59	1.01
psi	0.57	0.01	0.28	0.04	0.36	0.60	0.81	0.98	621.27	1.00

4.2 Model Diagnostics

To check the model assumptions and how the simulated data fit the real, we examined the residual plot and the posterior predictive check.

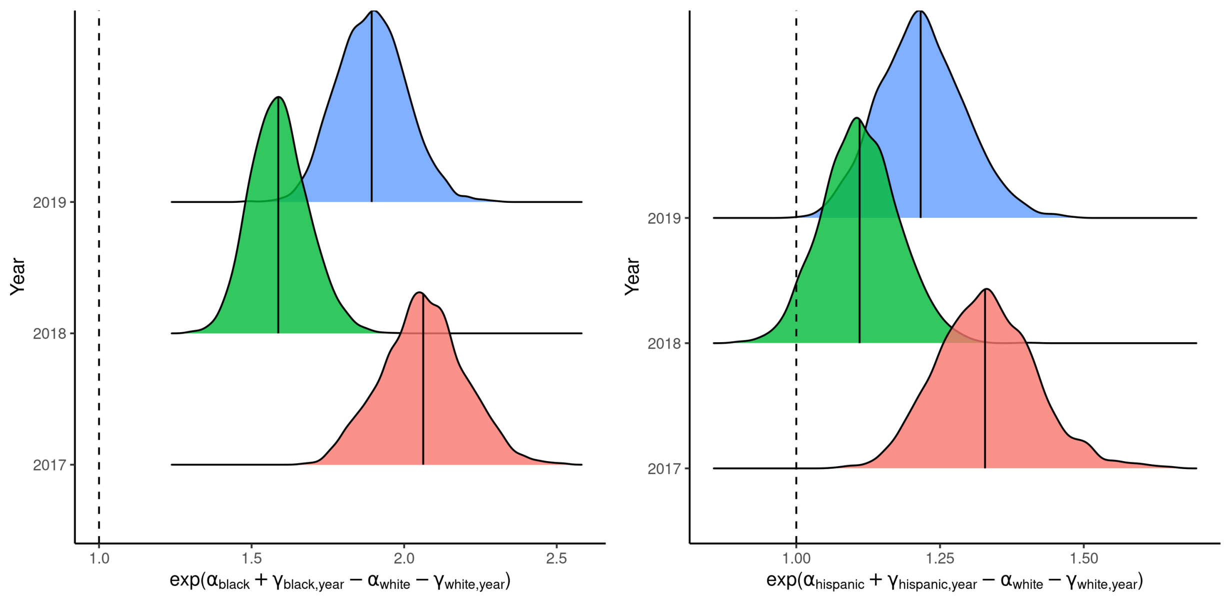

Figure 1: *Left*: This plot shows the posterior distribution of our discrimination measure for Blacks in contrast to White in the years 2017-2019. The three curves are all far away from 1, indicating that Black people are more likely to be stopped than White people having controlled for our baseline and the other fixed effects. *Right*: The same for Hispanic against White people. In this case discrimination seems lower, and in year 2018, 1 is in the hpd credibility interval.

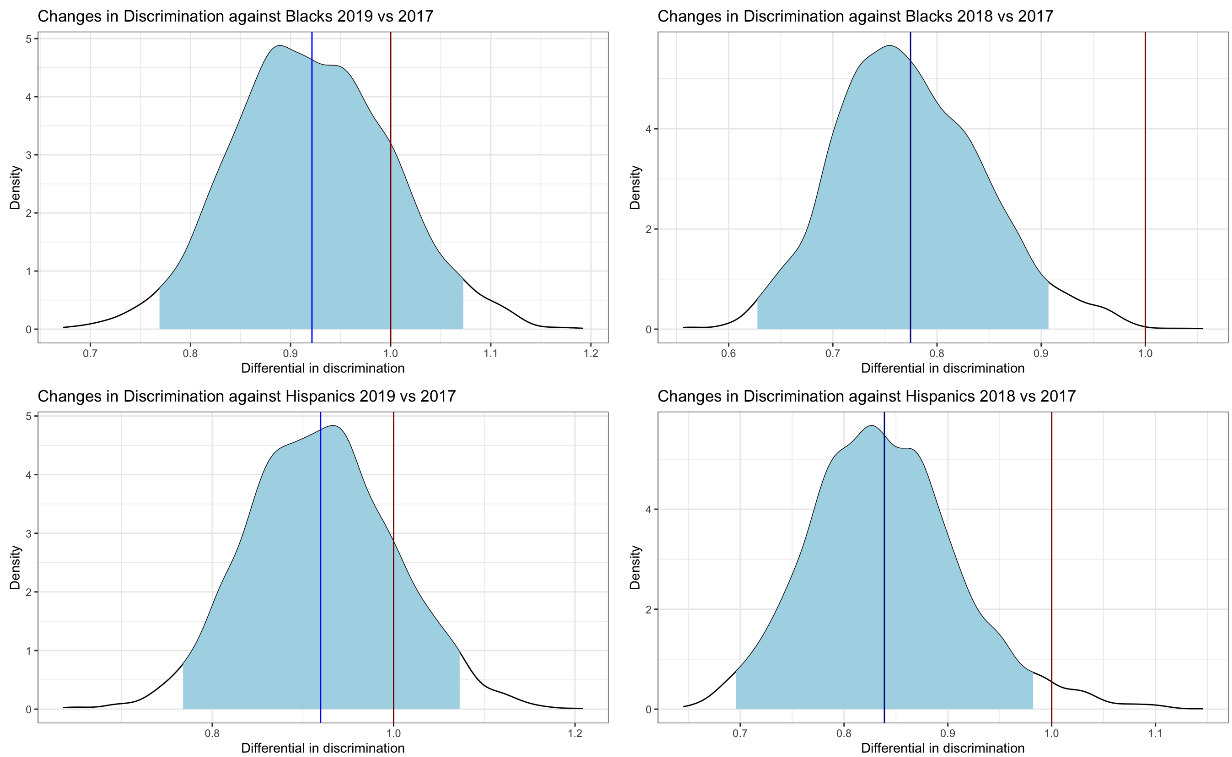

Figure 2: Here we have the density plot of the change in time of discrimination against Black people (top row) and against Hispanic people (bottom row). In particular we have the plots of $D_{r,y}/D_{r,2017}$ for $y \in \{2018, 2019\}$ and for r equal to Black and Hispanic. The decrease is significant at 95% for both races but only from 2017 to 2018. And in this case, it is stronger for Black people.

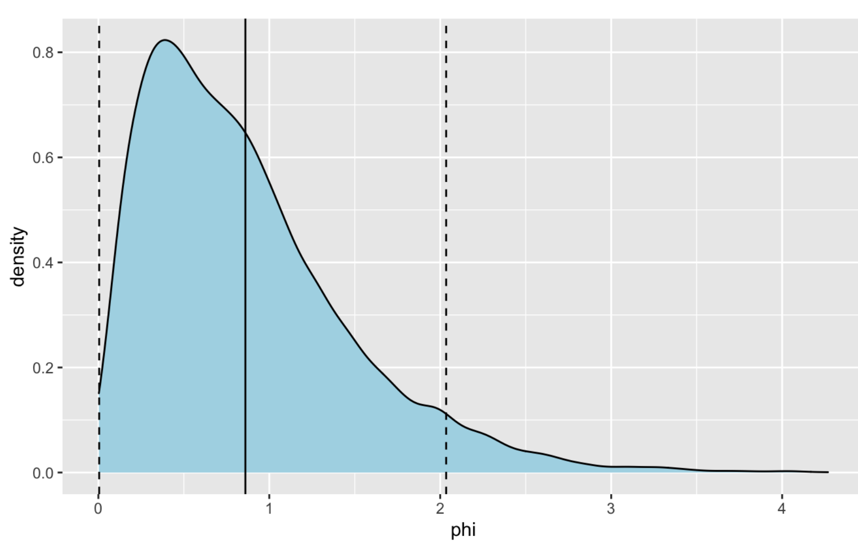

Figure 3: Posterior density of ϕ, the spatial correlation among precincts.

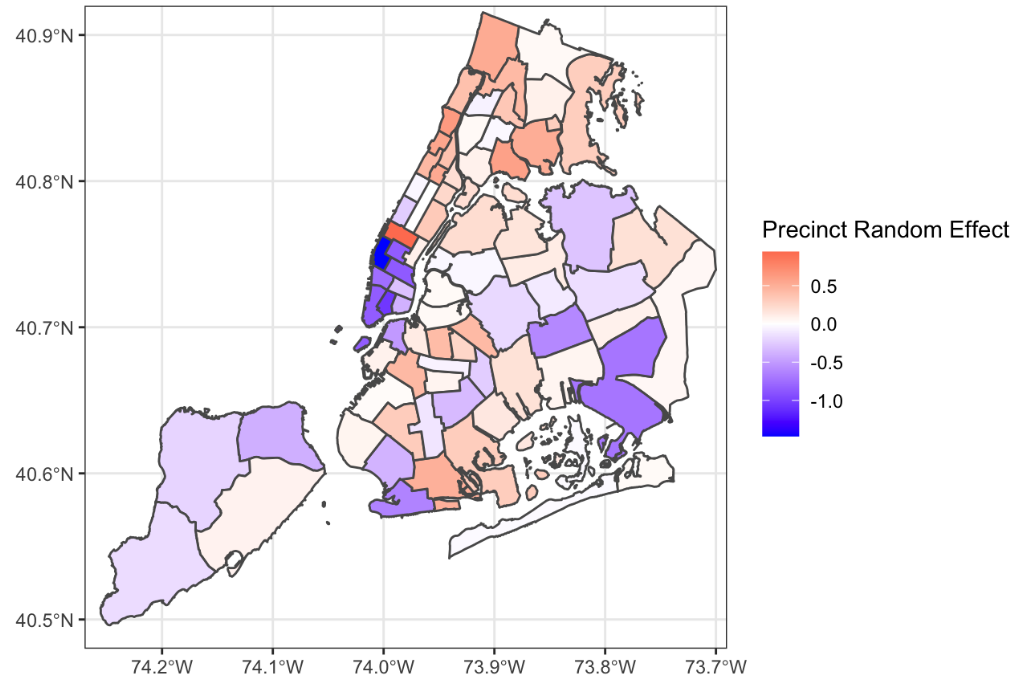

Figure 4: A map of NYC divided in precincts. Colors correspond to the value of the precinct random effect. In red zones police tends to stop people more frequently.

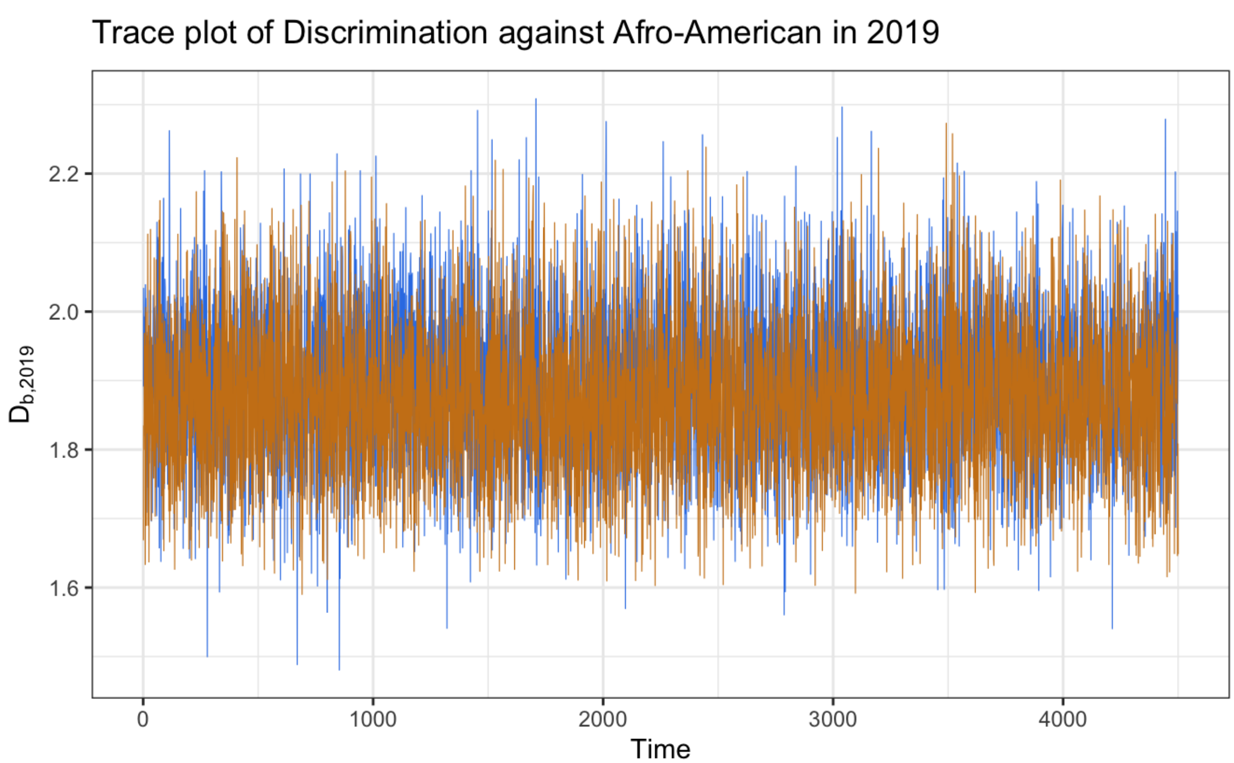

Figure 5: Trace Plot of the quantity $D_{b,2019}$ as defined in the main text. We used 5000 observations, with warm up equal to 500 and two chains. The plot indicates convergence and good mixing.

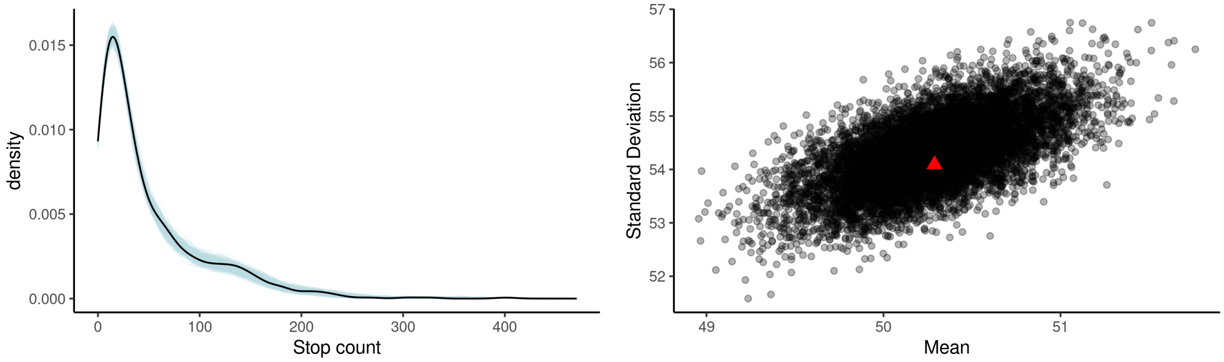

Figure 6: *Left*: Posterior predictive check. As shown in the plot, the pattern in the observed data distribution is well captured by simulations provided by the model. *Right*: Scatterplot of the standard deviation versus the average in the posterior predictive simulations. The red triangle corresponds to the real data, which lies well within the simulations.

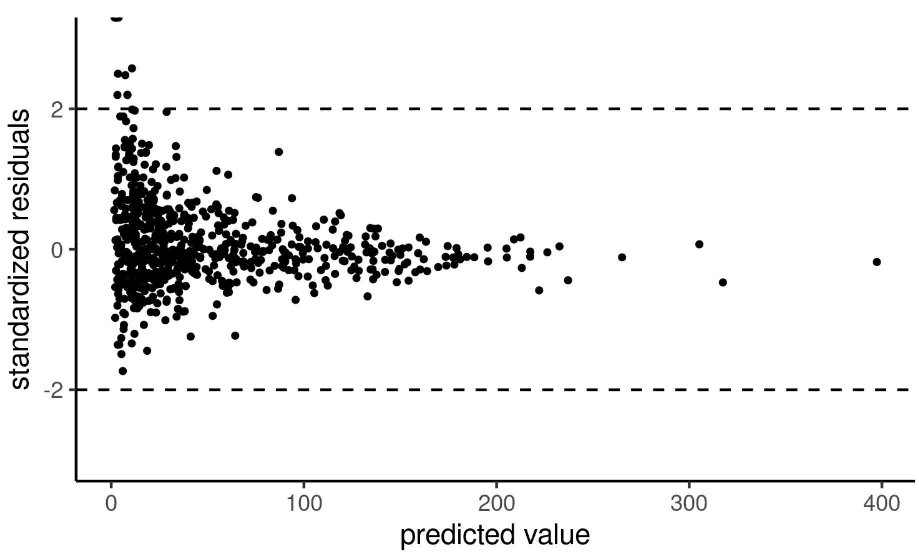

Figure 7: Standardized residual plot. $\hat{r} = (y - \hat{y})/\sqrt{\hat{y}}$. The standardized residuals should have zero mean and standard deviation one. As the plot indicates, most of the standardized residuals fall within the lines $[-2, 2]$, indicating that the variance in the data is largely explained by the model.