Investigating Racial Profiling
MCMC Modeling For the NYC Stop-and-Frisk Policy

Belle Xu and Flora Shi

April 25, 2021

1 Introduction

Stop-and-frisk, or a Terry Stop (derived from the *Terry v. Ohio* case), is a policing practice which allows a police officer to detain and question any person as well as to give a pat-down on the detained person's outer clothing for weapons under reasonable suspicion; should suspicions escalate after a initial series of questions, this may lead to a search or arrest[1][2]. It is a common practice in not only major cities in the United States such as Chicago, Los Angeles, and Philadelphia, but also in countries including Canada and the United Kingdom [3][4][5].

Notably, the stop-and-frisk policy was vastly used in New York City under Mayor Rudy Giuliani and Mayor Michael Bloomberg, and reached its height in 2002 to 2013 [6]. Throughout those 12 years, there were approximately 423,500 stops on average per year as recorded by the New York Police Department[6]. However, the concern of potential racial profiling rose to controversy. According to an analysis conducted by the Center of Constitutional Rights, those stops were largely targeted at Black and Latino New Yorkers: they made up nearly 85% of all stops in 2011. There have been multiple class action lawsuits (*Schoolcraft v. City of New York*, *Floyd, et al. v. City of New York, et al.*, *Davis V. City Of New York*, etc.) which challenged the targeted and aggressive stop-and-frisk policy adopted by the New York Police Department [7][8][9][10].

While many statistical research and analysis have been conducted on this issue and that the stop-and-frisk policy has become more and more mild, the concern of racial profiling in policing practices still remains in question today. In this project, we wish to investigate the elements that act behind the decision-making process of the stop-and-frisk policy, especially focusing on the effects of ethnicity, population, and the number of arrests in the past year on the number of stops within a precinct to determine whether a racial disparity truly exists. To accomplish this, we initially propose 3 negative binomial regression models and compare them using Bayesian model selection metrics, such as the Bayes' Factor and LOOIC, to find a base model with the best predictive power. We then incorporate interactions of interest into our base model and verify that the model with interactions has better predictive performance. After interpreting the coefficients of main effects as well as covariates associated with ethnicity, we find that after controlling for the effects of population proportion, crime type, and the number of past arrests, there still exists a racial disparity since the average number of stops for blacks and Hispanics are approximately 5 and 3 times higher than that of whites, respectively.

2 Data

We decide to use data from a modeling example *Social science modeling: police stops by ethnic group with variation across precincts* presented by Andrew Gelman and Jennifer Hill in their book *Data Analysis Using Regression and Multilevel/Hierarchical Models*[11]. The original dataset comes from the stop-and-frisk records kept by the New York City Police Department (NYPD) during the 1990s and includes around 175,000 records of stops over a 15-month period in 1998-1999 [12]. It is worth noting that while the police did not need to keep a record for every stop, there were certain conditions under which the police were required to keep records[12]. These required records account for 72% of the total stops recorded in the original dataset[12].

The data to be used for this project is a summarized or shrunken version of the original dataset. Instead of recording a stop-and-frisk event, each entry in our dataset represents the number of stops for a certain crime type for members of the same ethnic group in a particular precinct. There are 6 variables:

- **stop**: the number of stops of members of that ethnic group in that precinct
- **pop**: the population of that ethnic group in that precinct
- **past.arrests**: the number of arrests by people of that ethnic group in that precinct in the previous year as recorded by the Department of Criminal Justice Services (DCJS)
- **precinct**: an indicator variable that specifies the precinct (numbered from 1 to 75)
- **eth**: an indicator variable that specifies the ethnic group (1: black, 2: hispanic, 3: white)
- **crime**: an indicator variable that specifies the suspected charges of the person being stopped (1: violent crime, 2: weapon offense, 3: property crime, 4: drug crime)

To capture the characteristics of each precinct, we add the following variables:

- **pop.prop**: the population proportion of an ethnic group in a precinct, calculated by the population of an ethnic group in a precinct divided by the total population in that precinct
- **crime.prop**: the proportion of past arrests of an ethnic group for a particular crime in a precinct, calculated by the number of past arrests of an ethnic group for a particular crime in a precinct divided by the total number of past arrests for all crimes in the same precinct

Note that in precinct 48, the number of past arrests for property crime for white people is 0. As one of our proposed models (where the number of past arrests acts as an offset) does not allow us to have a 0 value for *past.arrests*, we decide to modify this value from 0 to 1. We choose not to omit this observation because we do not want to lose information for white people and rule out the probability that they could be arrested for property crimes in this precinct. It is also safe to assume that at least 1 white person has been arrested for a property crime in the past and this assumption offers us a better model interpretation.

3 Exploratory Data Analysis (EDA)

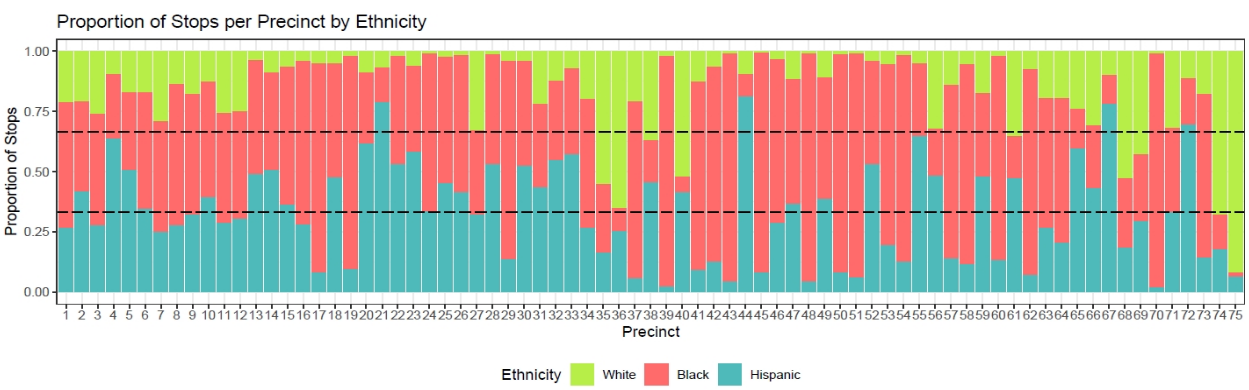

Figure 1: Proportion of stops by ethnicity in each precinct.

In Figures 1 and 2, we observe that blacks and Hispanics account for a much higher proportion of stops and past arrests than whites in most precincts, even in precincts where whites make up the majority of the population (see precincts 1-12, 34-38, 40-42, 61-69). Within a precinct, the distribution of stops is similar to the distribution of past arrests: if an ethnic group accounts for a high proportion of past arrests, it also accounts for a high proportion of stops. It seems possible that the past arrest proportion of an ethnic group is an important factor in determining whether the police stop members of that ethnic group.

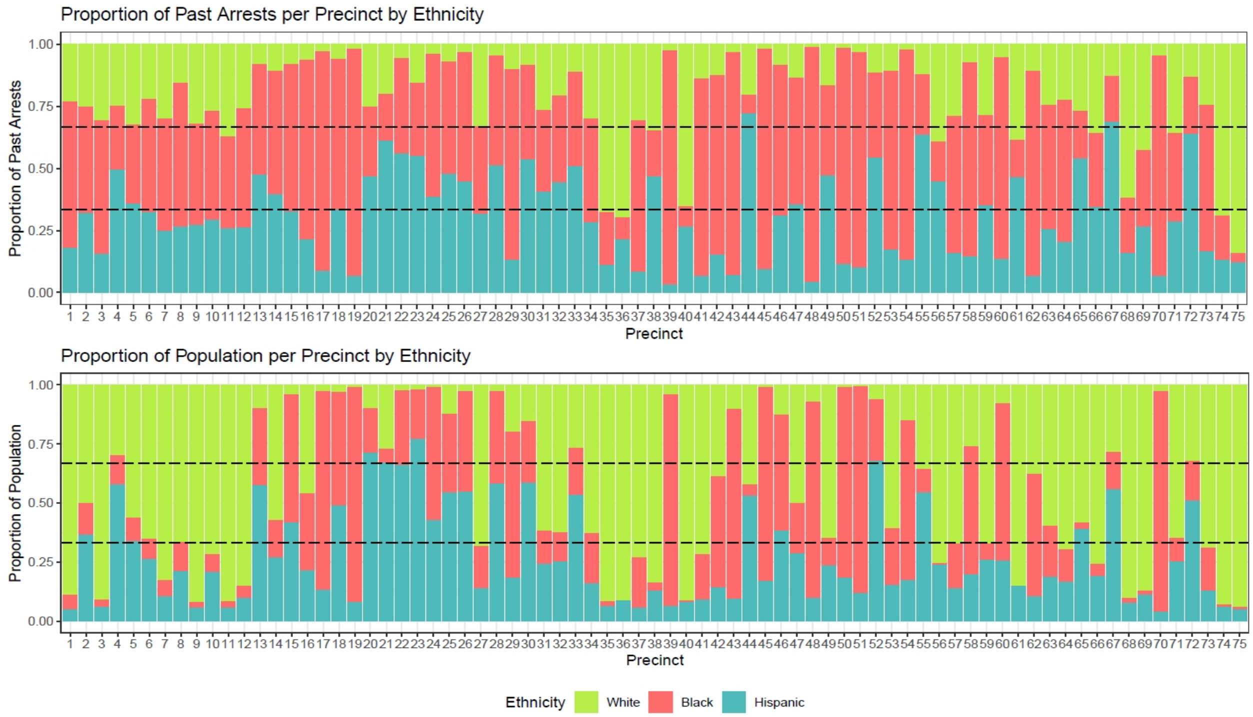

Figure 2: Proportion of past arrests (top) and population (bottom) by ethnicity in each precinct.

However, if we account for different types of crime, we observe that a high proportion of past arrests for a crime does not necessarily imply a high proportion of stops for that crime: for example, while blacks were arrested mostly for drug crime, they were stopped by police mostly for violent crime and weapon offenses (see Figure 3). A similar mismatch can be seen in other ethnic groups as well (see Appendix).

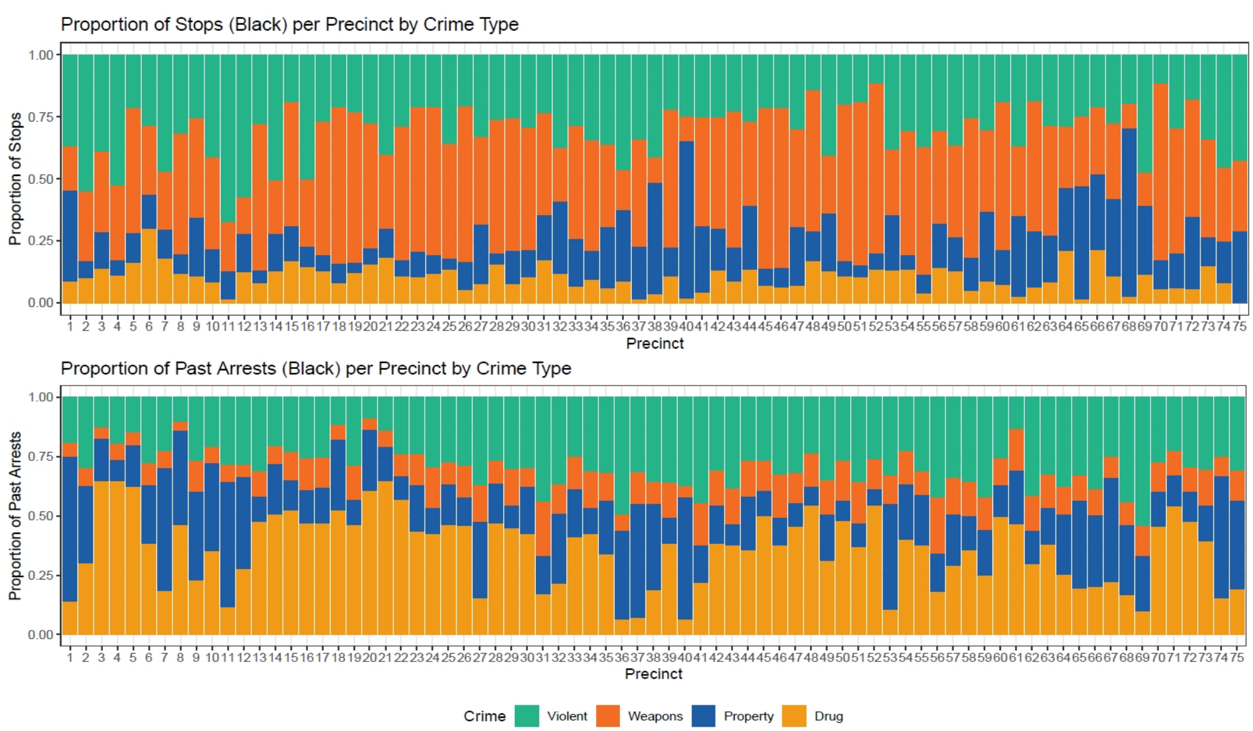

Figure 3: Proportion of stops (top) and past arrests (bottom) in each precinct for each crime for blacks.

In general, we observe from Figures 4 and 5 that as the population proportion for an ethnic group increases, the past arrest proportion and stop proportion for that ethnic group also increases. However, the past arrest proportion and stop proportion for blacks and Hispanics increase at a significantly faster rate than whites.

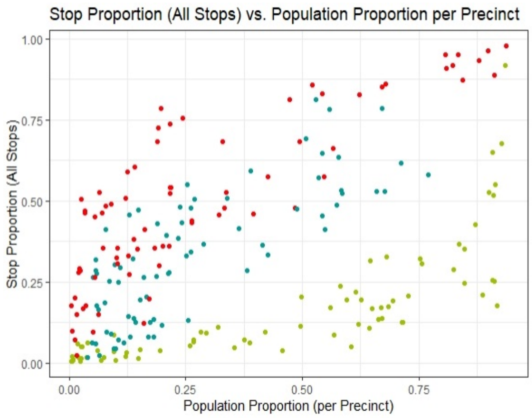

Figure 4: Proportion of stops vs. population proportion within a precinct for each ethnicity.

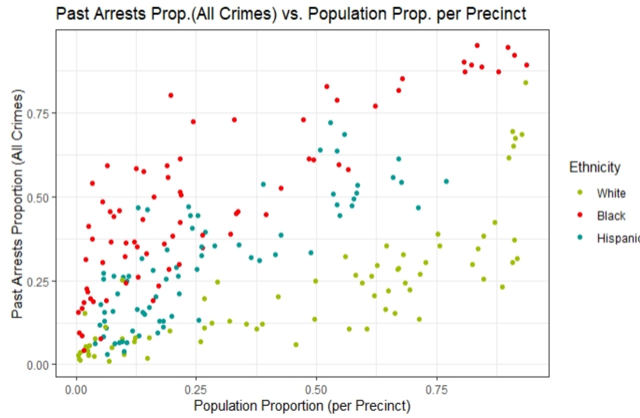

Figure 5: Proportion of past arrests (crime types are undifferentiated) vs. population proportion within a precinct for each ethnicity.

Given these observations, we decide to investigate the effects of ethnicity, population proportion, the number of past arrests, crime type, and their pairwise interaction on the number of stops. Since we model on count data (the number of stops), we initially propose a Poisson regression model. However, the mean of the number of stops in our dataset is 146.0222 while the variance is 47254.93. Since the variance is clearly exponentially larger than the mean, our data does not satisfy the assumption of a Poisson distribution due to its overdispersion. This implies that a negative binomial model— an alternative to the Poisson model for count data when there is overdispersion — might be a better sampling model for the number of stops.

4 Modeling

Motivated by the EDA, we will use a negative binomial regression model to determine the effects of each covariate. In equation, this gives us $y_i \sim$ Negative-Binomial(mean $= \mu_i$, overdispersion $= \phi$) such that $\mathbb{E}(y_i) = \mu_i$, $Var(y_i) = \mu_i + \mu_i^2/\phi$, and $\log(\mu_i) = X_i^T \boldsymbol{\beta}$ [13]. The regression coefficients $\boldsymbol{\beta}$ summarize the associations between the covariates and μ_i by the expression: $\mu_i = exp(ln(t_i) + \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + ... + \beta_k x_{ki})$, where t_i is the offset variable if there is one, and each x_{ji} is a covariate [14].

We initially proposed three models:

- **Model 1**: $\log(\text{mean(stops)}) = \log(\text{past arrests}) + \beta_0 + \beta_{1-2} * \text{ethnicity} + \beta_3 * \text{population prop.} + \beta_{4-6} * \text{crime type} + \beta_{7-12} * \text{ethnicity} * \text{crime type} + \beta_{13-15} * \text{crime type} * \text{population prop.} + \beta_{16-17} * \text{ethnicity} * \text{population prop.} = X^T \boldsymbol{\beta}$

- **Model 2**: $\log(\text{mean(stops)}) = \beta_0 + \beta_{1-2} * \text{ethnicity} + \beta_3 * \text{population prop.} + \beta_4 * \text{crime prop.} + \beta_{5-7} * \text{crime type} + \beta_{8-9} * \text{ethnicity} * \text{population prop.} + \beta_{10-11} * \text{ethnicity} * \text{crime prop.} + \beta_{12} * \text{crime prop.} * \text{population prop.} + \beta_{13-18} * \text{ethnicity} * \text{crime type} + \beta_{19-21} * \text{crime type} * \text{population prop.} + \beta_{22-24} * \text{crime type} * \text{crime prop.} = X^T \boldsymbol{\beta}$

- **Model 3**: $\log(\text{mean(stops)}) = \beta_0 + \beta_1 * \log(\text{past.arrests}) + \beta_{2-3} * \text{eth} + \beta_4 * \text{pop.prop} + \beta_{5-7} * \text{crime} + \beta_{8-13} * \text{eth} * \text{crime} + \beta_{14-16} * \text{crime} * \text{pop.prop} + \beta_{17-18} * \text{eth} * \text{pop.prop} = X^T \boldsymbol{\beta}$

Model 1 implies that $\frac{\text{mean(stops)}}{\text{past.arrests}} = X^T\boldsymbol{\beta}$. Note that the logarithm of the number of past arrests is the offset of this model and the regression coefficients $\boldsymbol{\beta}$ summarize the relationship between covariates and the rate of stops per arrest in the past year. Models 2 and 3 imply that mean(stops) $= X^T\boldsymbol{\beta}$, where the regression coefficients $\boldsymbol{\beta}$ summarize the relationship between covariates and the number of stops.

For all models, we propose a weakly informative prior as we have little prior information on the distribution of coefficients for each covariate. As suggested by Gelman, we choose to use a Cauchy prior with a location parameter of 0 and a scale of 2.5 for $\boldsymbol{\beta}$ [15]. This symmetric prior allows the coefficients to take on both positive and negative values just like a normal prior, but also allows a higher probability for them to take on more extreme values on the tails of the distribution than a normal prior with the same parameters.

Note that for ethnicity, we encode white people as the baseline since we are interested in the number of stops of blacks and Hispanics compared to whites. All modeling is automated by the *stan_glm*() function with proper inputs (see GitHub attached in the Appendix section for full code) [16].

In general, we use the leave-one-out cross-validation information criterion (LOOIC) to compare models that are not nested within each other and use Bayes' Factor to compare nested models. The Bayes' Factor diagnostic allows us to test under which model the observed data are more probable by comparing the marginal likelihood of the two models[17][18]. However, this metric could be ambiguous due to the Lindley's Paradox, in which the Bayes' Factor would often prefer the model with less covariates when we have a weakly informative prior that does not strongly favor one model over the other [19]. Since Bayes' Factor may not always be consistent when comparing non-nested models, we use LOOIC as a second metric that measures predictive accuracy of a model. LOOIC omits one observation from the original dataset in each iteration and tests the predictive performance of the input models on the omitted observation [20].

LOOIC indicates that Model 3 has stronger predictive power than Model 1 and Model 2. Since we are also interested in whether the effect of the logarithm number of past arrests would change based on crime type, ethnicity, and population proportion based on our observations in the EDA, we add these potential interactions to Model 3 and obtain a full model. Note that we do not add these interactions to Model 3 initially because we want to use Model 3 to test whether the logarithm number of past arrests should be used as the offset. Since both the Bayes' Factor and LOOIC indicate that the full model has better predictive performance and is more probable given our data, we accept it as our full model. Based on the ACF (autocorrelation function) plots and traceplots (see Appendix), all 4000 posterior samples generated from the *stan_glm*() function converge and mix well.

From Figure 6 below, we can see that the point representing the data mean and standard deviation lies approximately in the middle of the points representing the mean and standard deviation of our posterior predictive samples generated from our model[21][22]. This indicates that our model accurately captures the key features of our dataset. (More diagnostic plots and metrics are attached in the Appendix.)

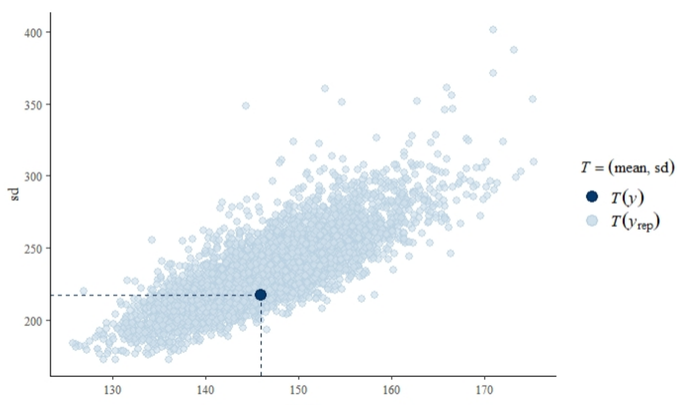

Figure 6: Posterior predictive graphical check which plots the mean and standard deviation of the original dataset against the ones generated from our posterior predictive model. Closeness of the dark blue dot (original data) to the middle of the cluster indicates better fit.

5 Results

	Coefficient Estimate (Median)	95% Credible Interval
Intercept	-1.291	(-2.045, -0.538)
log(past.arrests)	0.959	(0.751, 1.165)
eth1(Black)	1.671	(0.911, 2.437)
eth2(Hispanic)	1.000	(0.295, 1.755)
pop.prop	2.525	(1.339, 3.809)
crime2 (Weapon)	1.346	(0.659, 2.056)
crime3 (Property)	2.473	(1.709, 3.232)
crime4 (Drug)	0.703	(-0.034, 1.432)
eth1:pop.prop	-0.507	(-1.208, 0.243)
eth2:pop.prop	-0.077	(-0.843, 0.749)
eth1:crime2	-0.501	(-1.051, 0.013)
eth2:crime2	-0.275	(-0.744, 0.205)
eth1:crime3	-0.426	(-0.920, 0.054)
eth2:crime3	0.089	(-0.347, 0.511)
eth1:crime4	-0.070	(-0.532, 0.414)
eth2:crime4	0.257	(-0.179, 0.679)
log(past.arrests):eth1	-0.069	(-0.234, 0.089)
log(past.arrests):eth2	-0.035	(-0.199, 0.127)

Table 1: Select coefficient estimates (median) and 95% credible interval values associated with intercept, main effects, and interactions involving ethnicity from the final posterior model.

By analyzing the results of our posterior distribution, we wish to investigate and gain insight on whether a racial disparity truly exists in the mean number of stops. Since we only focus on the main effects as well as effects related to ethnicity, irrelevant coefficients are omitted here (see Appendix for full model coefficients).

Based on the table above, the 95% credible intervals for all coefficients of interactions associated with ethnicity contain 0, which indicates that these interactions are not statistically significant. Thus, we do not have enough evidence to suggest that the effect of population proportion, logarithm of the number of past arrests, and the crime type will change for blacks and Hispanics compared to whites.

According to our regression coefficients, regardless of ethnic group, for every 1 percent increase in the population proportion, the average number of stops for violent crime is expected to multiply by a factor of $exp(0.01 \cdot 2.525) \approx 1.026$, holding all else constant. This result is reasonable: even if the police stop people uniformly at random, the number of stops for an ethnic group will increase because the police will have a higher chance of meeting members of this ethnic group on the streets when the population proportion for this ethnic group increases.

Moreover, regardless of ethnic group, when the logarithm of the number of past arrests increase by 1, the average number of stops for violent crime is expected to multiply by a factor of $exp(0.959) \approx 2.609$, holding all else constant. The number of past arrests is a reasonable indicator of the crime rate of a particular crime in a precinct. On the one hand, if the crime rate of a precinct is high, a police could have a high chance of seeing people committing or about to commit this crime. On the other hand, if the crime rate is a factor of the police's active decision-making process of stopping a civilian, then bearing a high past crime rate in mind, a police might consciously stop civilians to prevent crimes even under mild suspicion. In either case, as the number of past arrests increases, the average number of stops is expected to increase.

Note that though we only interpret the case for violent crime, the effect of population proportion and logarithm of the number of past arrests on the number of stops is still positive for other crimes. Thus, we can draw analogous analyses for weapon crimes, property crimes, and drug crimes.

Though the effects of population proportion, logarithm of the number of past arrests, and crime type are similar for all ethnic groups, we see that the coefficient of the indicator variable for ethnicity associated with blacks is 1.671 while the coefficient associated with Hispanics is 1.000. These coefficients can be interpreted as follows:

- For black people, the average number of stops for violent crimes will expected to be $exp(1.671) \approx 5.317$ times the average number of stops for violent crimes for white people, holding all else constant.
- For Hispanic people, the average number of stops for violent crimes will expected to be $exp(1.000) \approx 2.718$ times the average number of stops for violent crimes for white people, holding all else constant.

Clearly, the coefficient of ethnicity reveals the fundamental differences between the number of stops for each ethnic group. After controlling for the effects of population proportion, crime type, and the number of past arrests, the average number of stops for blacks is more than 5 times higher than that of whites, while the average number of stops for Hispanics is close to 3 times higher than that of whites.

To understand this result further, we can consider it from another perspective. Let us assume that the population proportion for blacks and whites in a particular precinct is the same. Based on our negative binomial model, to reach the same average number of stops for violent crimes (represented by μ), the logarithm of the number of past arrests for whites is much higher than that for blacks or Hispanics as derived below. Note that c indicates the sum of all other terms in the model which we are holding constant:

For white people:

$$\begin{aligned}\mu &= exp(-1.291 + 0.959 \cdot log(past.arrests) + c)\\ \frac{\mu}{exp(-1.291 + c)} &= exp(0.959 \cdot log(past.arrests))\\ \frac{log(\frac{\mu}{exp(-1.291+c)})}{0.959} &= log(past.arrests)\end{aligned}$$

For black people:

$$\begin{aligned}\mu &= exp(-1.291 + 1.671 + 0.959 \cdot log(past.arrests) + c)\\ \frac{\mu}{exp(-1.291 + 1.671 + c)} &= exp(0.959 \cdot log(past.arrests))\\ \frac{log(\frac{\mu}{exp(-1.291+1.671+c)})}{0.959} &= log(past.arrests)\end{aligned}$$

Since $exp(-1.291) \approx 0.275$ is smaller than $exp(-1.291 + 1.671) \approx 1.46$, white people would need a significantly larger amount of past arrests than blacks or Hispanics to induce the same average number of stops. If the number of past arrests is truly a part of the polices' active decision-making process when stopping a civilian, this interpretation implies more tolerance for whites when considering past arrests.

6 Conclusion and Discussion

Based on our results above, we can reasonably conclude that during the period for which we have stop-and-frisk data, a racial disparity in the average number of stops does exist even after we control for the effects of population proportion of different ethnic groups, the number of past arrests, and crime types. Regardless of crime type, the average number of stops for blacks and Hispanics is approximately 5.317 and 2.718 times that for whites, respectively.

Some limitations of our study lies within both the modeling approach as well as the conclusions that can be safely drawn from our model. In terms of the modeling approach, since we use Bayesian regression methods, our analysis could be improved if we have a more accurate prior model. Although there exists substantial literature on stop-and-frisk policy in New York City, it is difficult to incorporate them into our prior belief: our data comes from stop-and-frisk records over a 15-month period in 1998-1999 while most past studies (which we only had time to briefly read on) used annual stop-and-frisk data after 2000. Given little information on the similarities of stop-and-frisk records across different time periods, we choose not to incorporate results from these studies and thus propose a weak Cauchy prior to account for the limited prior information. If we can find more relevant previous analysis and incorporate them as prior beliefs, our model's predictive accuracy as well as its related inferences could be potentially improved. Furthermore, we only run 4,000 samples in our Markov Chain Monte Carlo modeling method for each of the coefficient estimates due to hardware constraints (each model takes approximately thirty minutes to run.) As the effective sample sizes for each chain are verified to be sufficiently large (see Appendix), the estimates in our final model are still feasible. However, the precision of these estimates can be further improved if we could increase the number of iterations to ideally greater than 40,000 samples.

While our model suggests that a racial disparity in the average number of stops is substantial, we should be cautious to conclude that there exists discrimination against ethnic minorities in NYPD's stop-and-frisk policy. It is worth noting that our dataset is a summarized version of the stop-and-frisk records rather than the true records themselves. While we extract as much information as we can from our dataset to investigate the existence of racial disparities, it is impossible to truly infer the police's decision-making process and the situation when the stop event happens. For example, we cannot infer whether a civilian was stopped under strong suspicion of committing a crime (i.e., seen actively or about to actively commit a crime) or when he/she was just taking a peaceful walk. Moreover, we also lack information on whether a stop has led to a real arrest or even a conviction. If the person being stopped gets arrested or convicted, then it is more evident that the stop is effective and reasonable. To know these missing information requires a substantially more detailed and comprehensive record of the stop events or an even more developed monitoring system for the police's practices as it is possible for false or inaccurate records to exist. Therefore, our results only suggest a statistically significant racial disparity in the average number of stops rather than racial bias in the stop-and-frisk policy.

Racial profiling in policing practices is a controversial topic that inquires studies and efforts from various disciplines. While our analysis is limited, it is still useful gain some quantitative insight and realize that a racial disparity does exist statistically and can affect the lives of ethnic minority communities. Bearing this in mind, the NYPD and policy makers could then proceed to reflect upon and improve the policing practices to ensure that every stop has an accurate and justified reason, regardless of the ethnicity of the person being stopped.

7 References

[1] *Stop and frisk.* URL: https://www.law.cornell.edu/wex/stop%5C_and%5C_frisk.

[2] David S. Kemp. *Terry v. Ohio, 392 U.S. 1 (1968).* URL: https://supreme.justia.com/cases/federal/us/392/1/.

[3] Becca James. *Stop and frisk in 4 cities: The importance of open police data.* Mar. 2015. URL: https://sunlightfoundation.com/2015/03/02/stop-and-frisk-in-4-cities-the-importance-of-open-police-data-2/.

[4] *Here's What You Need to Know About Carding.* 2017. URL: https://www.cbc.ca/firsthand/m%5C_features/heres-what-you-need-to-know-about-carding.

[5] *Police powers to stop and search: your rights.* Feb. 2017. URL: https://www.gov.uk/police-powers-to-stop-and-search-your-rights.

[6] Bridge Initiative Team. *Factsheet: NYPD STOP AND FRISK POLICY.* June 2020. URL: https://bridge.georgetown.edu/research/factsheet-nypd-stop-and-frisk-policy/.

[7] Center for Constitutional Rights. *Stop and Frisk - The Human Impact.* July 2012. URL: https://ccrjustice.org/sites/default/files/attach/2015/08/the-human-impact-report.pdf.

[8] *Schoolcraft v. City of N.Y.* May 2015. URL: https://casetext.com/case/schoolcraft-v-city-of-ny-9.

[9] Center for Constitutional Rights. *Floyd, et al. v. City of New York, et al.* Jan. 2008. URL: https://ccrjustice.org/home/what-we-do/our-cases/floyd-et-al-v-city-new-york-et-al.

[10] *Case: Davis v. City of New York.* 2010. URL: https://www.naacpldf.org/case-issue/davis-v-city-new-york/.

[11] Andrew Gelman and Jennifer Hill. “Social science modeling: police stops by ethnic group with variation across precincts”. In: *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press, 2006.

[12] Andrew Gelman, Jeff Fagan, and Alex Kiss. “An Analysis of the New York City Police Department’s “Stop-and-Frisk” Policy in the Context of Claims of Racial Bias”. In: *Journal of the American Statistical Association* 102.479 (Sept. 2007), pp. 813–823. DOI: 10.1198/016214506000001040. URL: http://www.stat.columbia.edu/~gelman/research/published/frisk9.pdf.

[13] Jeffery B. Arnold. *Negative Binomial: Estimating Homicides in Census Tracks*. May 2018. URL: https://jrnold.github.io/bugs-examples-in-stan/index.html.

[14] Jerry L. Hintze. “Ch326: Negative Binomial Regression”. In: *Number cruncher statistical system: statistical software*. NCSS, LLC., 1992, 326–1-326–2. URL: https://ncss-wpengine.netdna-ssl.com/wp-content/themes/ncss/pdf/Procedures/NCSS/Negative%5C_Binomial%5C_Regression.pdf.

[15] Andrew Gelman et al. “A Weakly Informative Default Prior Distribution For Logistic and Other Regression Models”. In: *Annals of Applied Statistics*. 1360-1383 2.4 (2008), pp. 1360–1383. DOI: 10.1214/08-AOAS191. URL: http://www.stat.columbia.edu/~gelman/research/published/priors11.pdf.

[16] Jonah Gabry and Ben Goodrich. *How to Use the rstanarm Package*. July 2020. URL: https://mc-stan.org/rstanarm/articles/rstanarm.html.

[17] Dominique Makowski, Mattan S. Ben-Shachar, and Daniel Lüdecke. “bayestestR: Describing Effects and their Uncertainty, Existence and Significance within the Bayesian Framework”. In: *Journal of Open Source Software* (Aug. 2019). DOI: https://doi.org/10.21105/joss.01541. URL: https://joss.theoj.org/papers/10.21105/joss.01541.

[18] Dominique Makowski et al. *Indices of Effect Existence and Significance in the Bayesian Framework*. Dec. 2019. URL: https://www.frontiersin.org/articles/10.3389/fpsyg.2019.02767/full.

[19] Christian P. Robert. *On the Jeffreys–Lindley’s Paradox*. Nov. 2013. URL: https://arxiv.org/pdf/1303.5973.pdf.

[20] Aki Vehtari, Andrew Gelman, and Jonah Gabry. *Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC*. Sept. 2016. URL: https://arxiv.org/pdf/1507.04544.pdf.

[21] Jonah Gabry and Ben Goodrich. *Graphical posterior predictive checks - pp_check.stanreg*. 2020. URL: http://mc-stan.org/rstanarm/reference/pp%5C_check.stanreg.html.

[22] Jonah Gabry. *Graphical Posterior Predictive Checks using the Bayesplot Package*. Jan. 2021. URL: https://cran.r-project.org/web/packages/bayesplot/vignettes/graphical-ppcs.html.

8 Appendix

8.1 GitHub Code

GitHub Code: https://github.com/cshi56/360-Final-Project

8.2 Exploratory Data Analysis

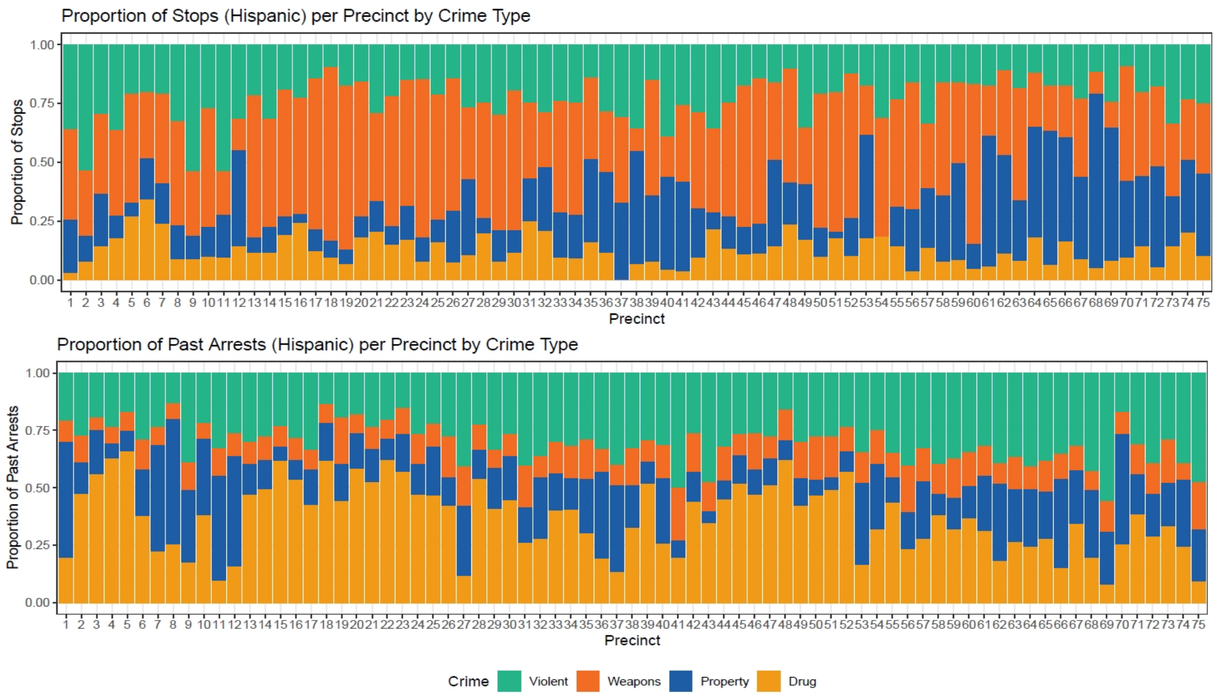

Figure 7: Proportion of stops and past arrests in each precinct for each crime for Hispanics. While the past arrests for Hispanics largely consists of drug crimes, they are stopped mostly for weapon crimes.

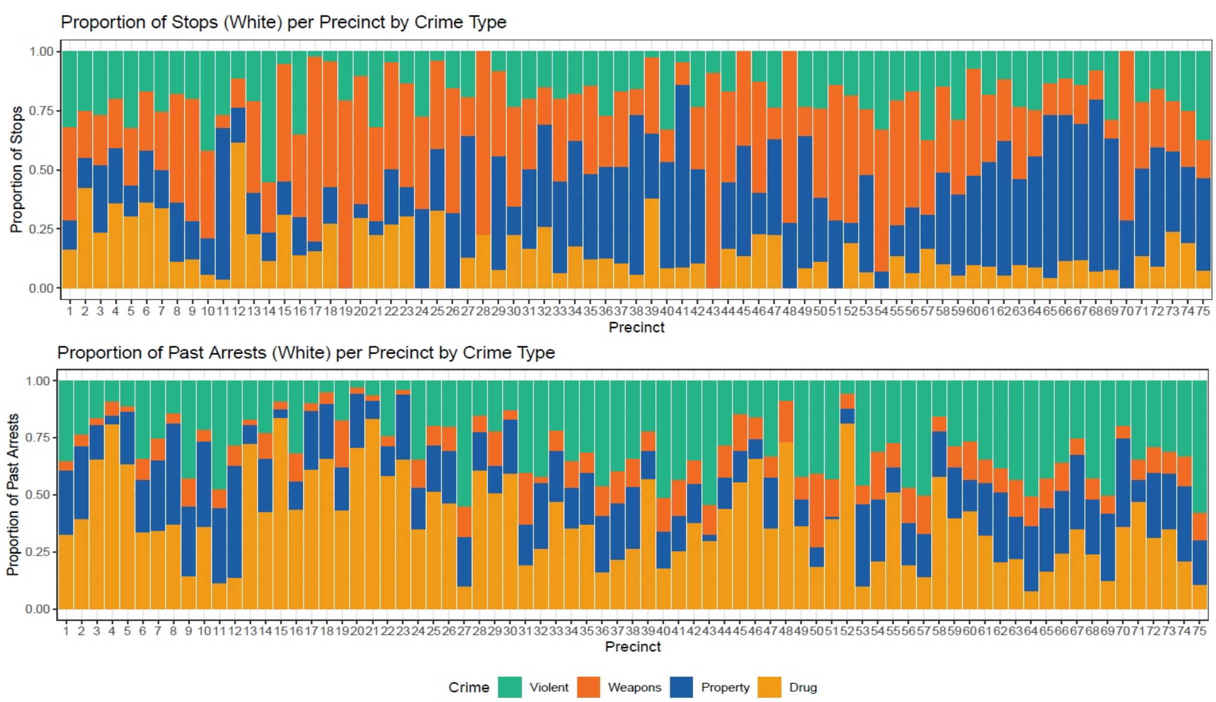

Figure 8: Proportion of stops and past arrests in each precinct for each crime for whites. While the past arrests for whites largely consists of drug crimes, they are stopped mostly for weapon or property crimes.

8.3 Modeling

ACF plots and traceplots: For all ACF plots and traceplots (52 in total), please see the pdf file *Diagnostic-Plots* on GitHub. The ACF (autocorrelation function) and traceplots wact as a measure for MCMC chain convergence and mixture. In general, the ACF plots for each coefficient estimate showed that all chains converged as all ACF values approached 0; and the traceplots for each coefficient estimate showed that all chains converged and mixed well as the traceplots are uncluttered (does not linger in one area) and explores the region evenly (the trace does not visit one area significantly less than another).

Other diagnostics:

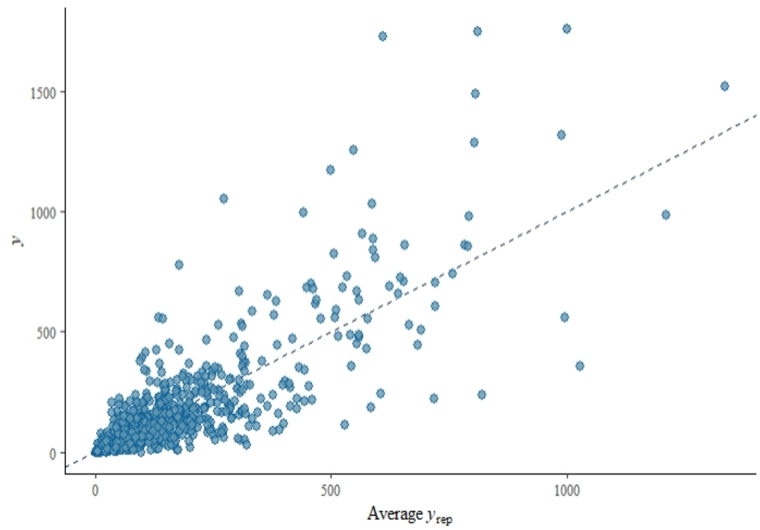

Figure 9: Posterior predictive graphical check which plots the mean of the samples generated from our posterior predictive model. The closer the dots are to the diagonal line (indicating the true data mean), the fitted our model is. We can see that our model is better fit for lower values of stops, but has a lot of variance when predicting higher values.

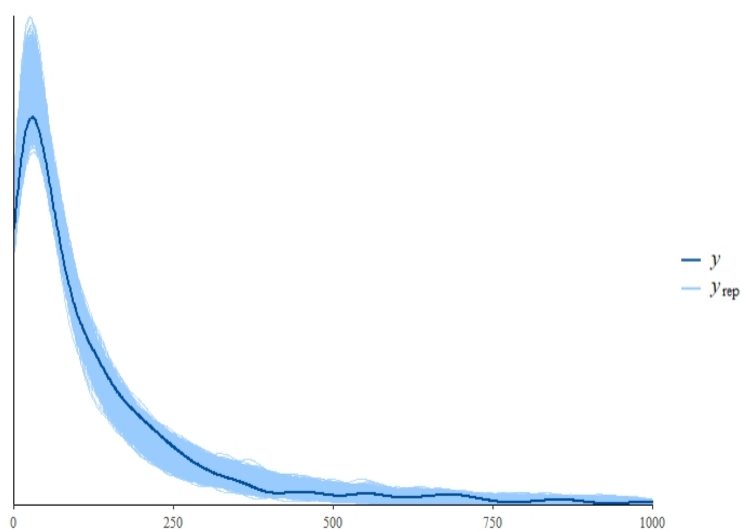

Figure 10: Posterior predictive graphical check which plots the density of the samples generated from our posterior predictive model against the true data density. A model correctly captures the data if the data density line lies approximately within the middle of the sample densities. Our model achieves this goal.

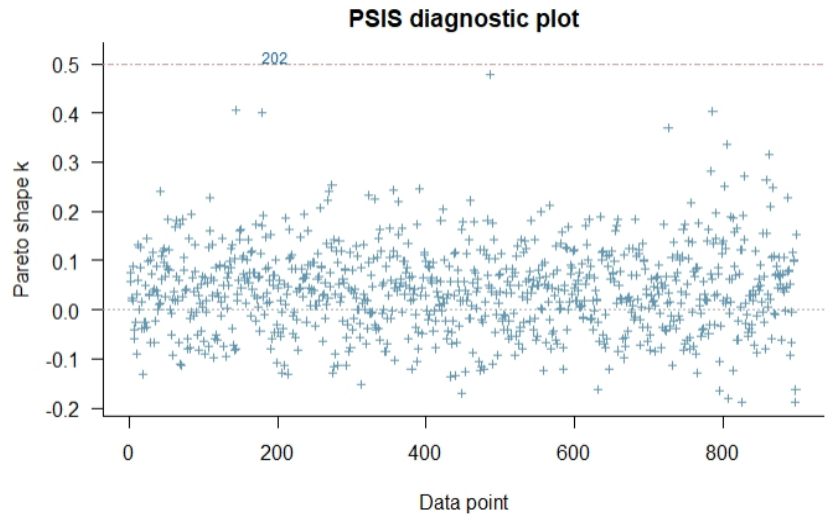

Figure 11: Graphical check for the Pareto-K diagnostic used in looIC computations. This graph shows if there are any outlier observations in the given dataset that could greatly skew our posterior model fit. As we can see, only one outlier exists and it does not deviate greatly in magnitude (Pareto-K = 0.5), therefore our model would not be greatly skewed due to outliers.

	mcse	Rhat	n_eff
Intercept	0.0	1.0	1527
log(past.arrests)	0.0	1.0	1300
eth1(Black)	0.0	1.0	1816
eth2(Hispanic)	0.0	1.0	1935
pop.prop	0.0	1.0	1628
crime2 (Weapon)	0.0	1.0	1621
crime3 (Property)	0.0	1.0	1611
crime4 (Drug)	0.0	1.0	1791
eth1:pop.prop	0.0	1.0	2113
eth2:pop.prop	0.0	1.0	2585
eth1:crime2	0.0	1.0	1641
eth2:crime2	0.0	1.0	1597
eth1:crime3	0.0	1.0	1491
eth2:crime3	0.0	1.0	1595
eth1:crime4	0.0	1.0	1649
eth2:crime4	0.0	1.0	1880
pop.prop:crime2	0.0	1.0	1449
pop.prop:crime3	0.0	1.0	1366
pop.prop:crime4	0.0	1.0	1556
log(past.arrests):eth1	0.0	1.0	1821
log(past.arrests):eth2	0.0	1.0	1999
log(past.arrests):pop.prop	0.0	1.0	1950
log(past.arrests):crime2	0.0	1.0	1363
log(past.arrests):crime3	0.0	1.0	1194
log(past.arrests):crime4	0.0	1.0	1253
reciprocal dispersion	0.0	1.0	5549
mean_PPD	0.1	1.0	3912
log-posterior	0.1	1.0	1487

Table 2: MCMC diagnostic values for our final fitted model. The interpretation can be directly taken from the R output: for each parameter, mcse is Monte Carlo standard error, n_eff is a crude measure of effective sample size, and R-hat is the potential scale reduction factor on split chains (at convergence R-hat=1). Clearly all chains converged since all R-hat values are 1. All effective sample sizes are large enough to obtain an accurate chain (no error output from the stan_glm() function).

8.4 Results

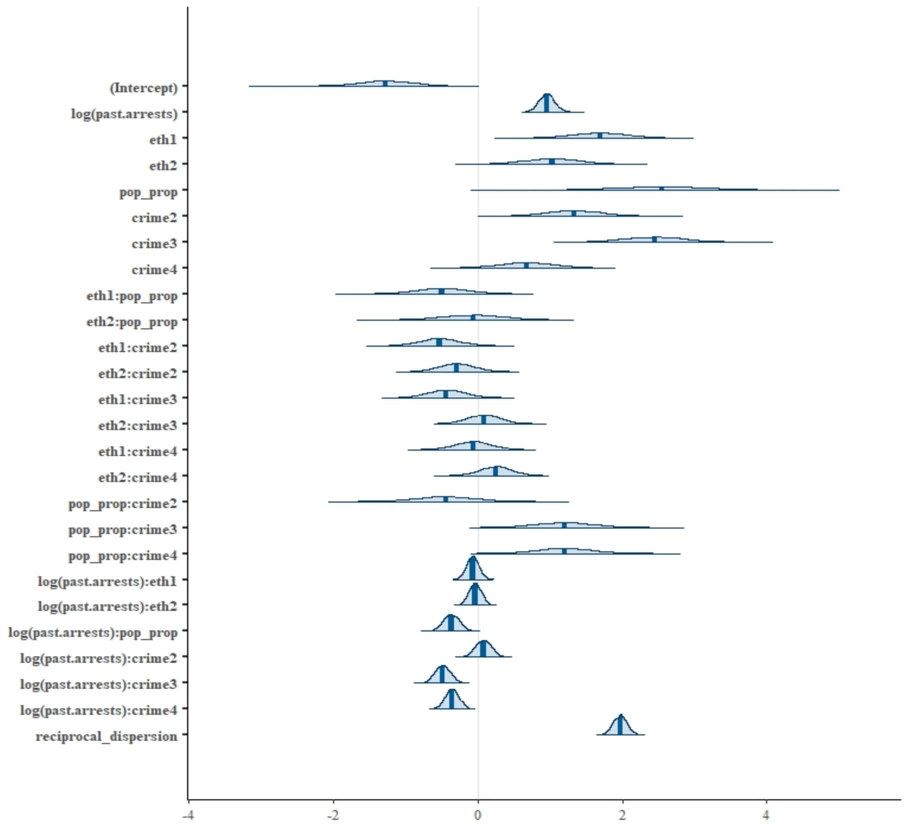

Figure 12: Posterior distributions of coefficients in the final model.

	Coefficient Estimate (Median)	95% Credible Interval
Intercept	-1.291	(-2.045, -0.538)
log(past.arrests)	0.959	(0.751, 1.165)
eth1(Black)	1.671	(0.911, 2.437)
eth2(Hispanic)	1.000	(0.295, 1.755)
pop.prop	2.525	(1.339, 3.809)
crime2 (Weapon)	1.346	(0.659, 2.056)
crime3 (Property)	2.473	(1.709, 3.232)
crime4 (Drug)	0.703	(-0.034, 1.432)
eth1:pop.prop	-0.507	(-1.208, 0.243)
eth2:pop.prop	-0.077	(-0.843, 0.749)
eth1:crime2	-0.501	(-1.051, 0.013)
eth2:crime2	-0.275	(-0.744, 0.205)
eth1:crime3	-0.426	(-0.920, 0.054)
eth2:crime3	0.089	(-0.347, 0.511)
eth1:crime4	-0.070	(-0.532, 0.414)
eth2:crime4	0.257	(-0.179, 0.679)
pop.prop:crime2	-0.447	(-1.238, 0.372)
pop.prop:crime3	1.199	(0.446, 1.979)
pop.prop:crime4	1.203	(0.446, 2.004)
log(past.arrests):eth1	-0.069	(-0.234, 0.089)
log(past.arrests):eth2	-0.035	(-0.199, 0.127)
log(past.arrests):pop.prop	-0.368	(-0.580, -0.150)
log(past.arrests):crime2	0.076	(-0.141, 0.294)
log(past.arrests):crime3	-0.489	(-0.697, -0.285)
log(past.arrests):crime4	-0.357	(-0.548, -0.170)
reciprocal dispersion	N/A	(1.779, 2.160)

Table 3: Full model coefficient estimates (median) and 95% credible interval values.