Case 3

Mishek Thapa, Tong Wu, Sibora Seranaj and Jessie Ou

04/03/2021

1 Introduction

Stop-Question-Frisk (SQF) is a New York policing protocol, in which officers stop, question, and potentially detain a civilian suspicious of crime. The SQF strategy was introduced to New York City in 2002 and peaked in 2011 with 685,724 stops. The tactic has stirred racial profiling controversy, and many have alleged that the program unfairly targets Black and Hispanic individuals. A 2012 statistical analysis showed that "persons of African and Hispanic descent were stopped more frequently than whites, even after controlling for precinct variability and race-specific estimates of crime participation" (Gelman, Fagan, and Kiss). The SQF practice was sued on behalf of the residents and the policy was found unconstitutional after the 2013 Floyd, et al. v. City of New York lawsuit. The settling mandate required police officers to justify in detail the reason for stopping a civilian. Since 2011, the number of SQF events has fallen, however, there is still concern for ongoing racial profiling or any other disparity in treatment of minority civilians. It is incredibly important to be able to identify the racial profiling, because not only is it unjust, but also because frequent contact with the police may lead to anxiety symptoms and trauma (Geller, Fagan, Tyler, and Link).

The goal of this analysis is to determine whether there may still be evidence of racial disparities in SQF events, taking into consideration the potential spatial effects of precinct – that is, whether there is evidence of any spatial correlation. We focus on SQF events where the suspected crime was criminal possession of weapons (CPW). Using 4918 CPW events between 2017-2019, we fit a spatial error model to model the precinct-level weapon recovery rate. In terms of precincts, we hypothesize that the weapon discovery rate would be lower for precincts where the proportion of Blacks and Hispanics CPW suspects was higher.

1.1 Approach

One common approach to studying racial disparities focuses on disparate impacts – that is, the outcomes of SQF events among different racial groups. Hit rate is defined as the proportion of stops with positive outcomes (arrested, summoned, or some outcome) over the total number of stops. Hit rates are used to evaluate the efficiency of a stop, with lower hit rates corresponding to lower threshold of suspicion for stopping an individual (Ayres 2002a). We focused on SQF stops where the suspected crime was criminal possession of weapons, so our positive outcome would be defined as a weapon found on the suspected individual. This approach was adapted from a 2016 statistical analysis that modelled the ex ante probability that the detained suspect has a weapon (Goel, Rao, and Shroff). Of all kinds of disparities, we chose outcomes disparity to bypass the problem of finding an appropriate baseline to compare race-specific stop/search/frisk rates (Gould and Mastrofski 2004). Of all the positive outcomes, we chose weapon discovery in a CPW event for the reasons listed. First, weapon discovery is a direct indicator of whether a police officer's suspicion of CPW is justified, whereas a stopped suspect could be arrested or summoned for reasons other than the one he/she was stopped for. Secondly, different crime types have different hit rates, so by focusing on one type of crime (CPW), we control for varying degrees of hit rate among different crime types. Thirdly, focusing on just CPW stops also bypasses the problem of hit rate fallacy, which argues that police are biased against White people, since they have a higher arrest rate.

In order to take into consideration potential spatial effects of precincts, we aggregated all individual CPW stops in a given precinct into a proportion that represents the precinct-level weapon recovery rate.

As a disclaimer, claims about discrimination require well-designed experiments that control for all other factors to isolate for the effects of race. Our observational dataset lacks such controls, so our findings can only serve as evidence for differential treatment which produces disparate outcomes.

1.2 About Data

The data was directly taken from New York Police Department police officers' responses to form UF-250 for 36096 stop-and-frisk stops between 2017-2019. The UF-250 form has information about demographics of stopped suspect (race, sex, age, etc.), the time and location of stop (precinct and borough), the rationale for the stop, and any police-suspect interactions (search, frisk, physical force, weapon/contraband found). In addition to SQF police records, we also have population demographics data from the 2020 Census, which lists the population by race in each of NYPD's 77 precincts. The population data can be joined to police records based on precinct number. In addition, we also utilized a spatial dataset from the NYC Planning website, which includes data on the most recent police precincts boundaries used by the NYPD. Moreover, we also have access to shape files which describe the latest NYPD precinct boundaries.

1.3 Response Variable

The response, or dependent variable was precinct-level weapon recovery rate, that is, the percentage of SQF stops that yielded a positive outcome (weapon found) out of stops where the suspected crime was criminal possession of weapons.

1.4 Exploratory Data Analysis

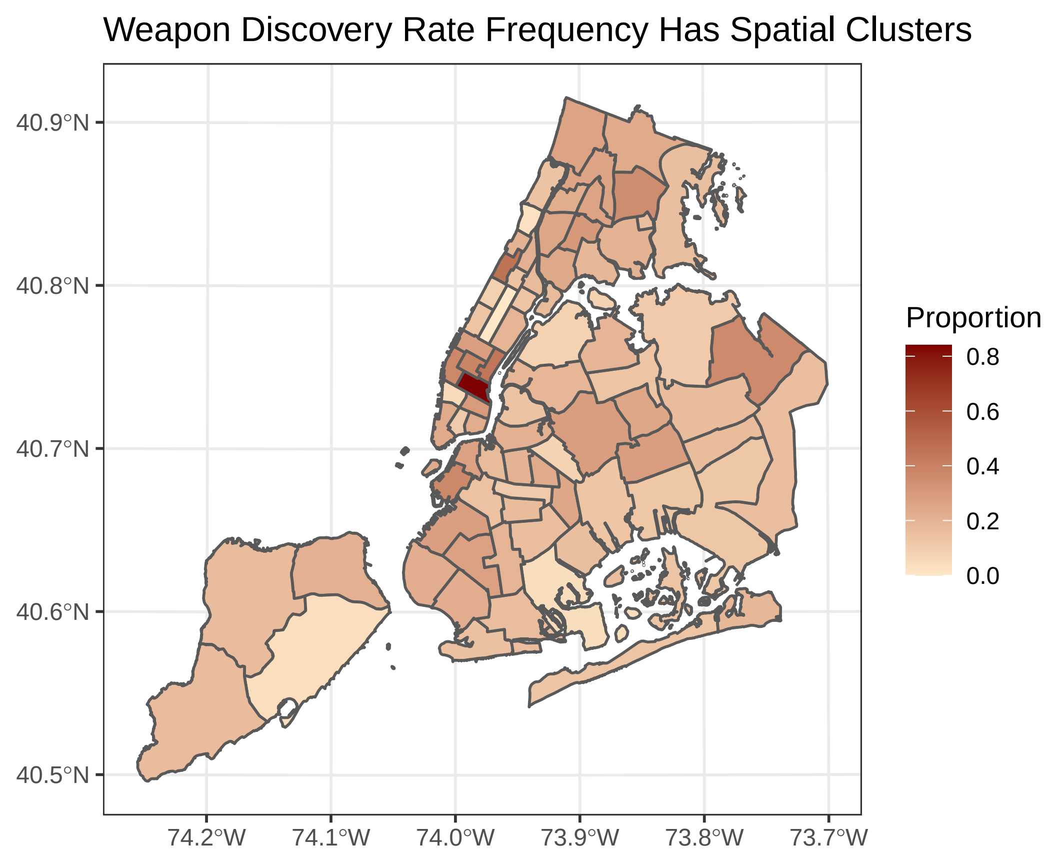


Figure 1: Choropleth map of precinct-level weapon recovery rate in New York. Precinct level weapon discovery rate has some spatial clusters. Other than a couple of outlier points, precincts with similar weapon discovery rate tend to be close together, indicating there is some spatial dependency present

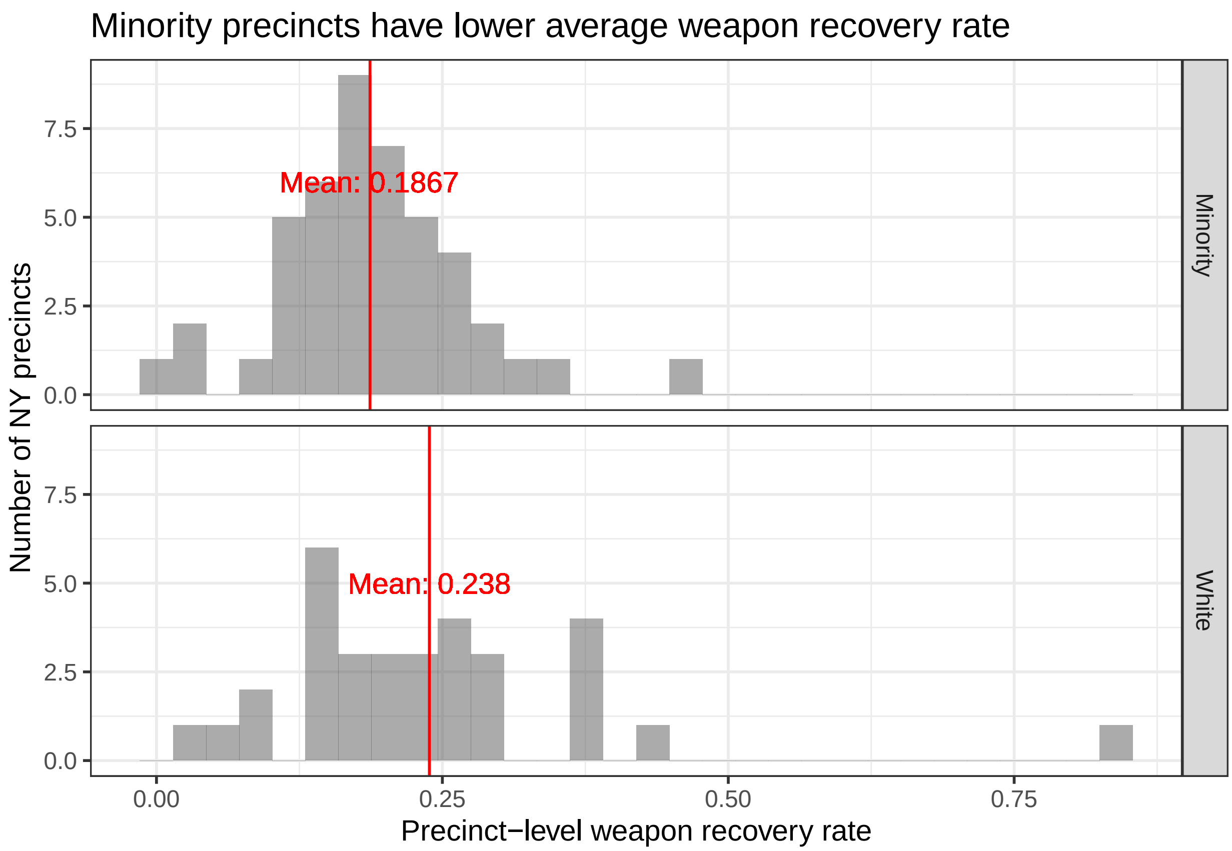


Figure 2: Precinct-level weapon recovery rate distribution, characterized by composition of precinct. A precinct is defined as White if its percentage of White residents is over 50%, otherwise it is a minority precinct. Precinct 22 has 0 weapon discovery rate, whereas precinct 13 weapon discovery rate is 0.839. The red vertical line represents the mean weapon discovery rate in each group

2 Methodology

Results from the global Moron I test for regression residuals showed there was sufficient evidence for spatial clustering among the residuals (p-value = 0.00172). To deal with spatial dependence in regression models, we fit a spatial error model with precinct-level weapon discovery rate as our outcome. In order to decide whether to fit a spatial error or a spatial lag model, we ran the Lagrange Multiplier tests based on the following hypotheses:

Spatial error model:

$$H_0 : \lambda = 0$$
$$H_1 : \lambda \neq 0$$

Spatial lag model:

$$H_0 : \rho = 0$$
$$H_1 : \rho \neq 0$$

Empirically, there was evidence that spatial dependencies exist in the linear model for error dependence and a missing spatially lagged variable (p-values are 0.01459 and 0.01202, respectively). We chose the spatial error model because we believe the residuals are correlated due to an unmeasured confounding variable and if we were to measure them, we would no longer have issues with spatial dependency. In particular, some potentially unaccounted-for confounders are historic crime rates, patrolling schedule and patrol crew. For example, there is the possibility that neighboring precincts have police officers that all live near each other, which may be associated with their policing patterns. Some police officers might have different policing strategies and make more frequent SQF stops to gather information on crime rate or to enforce police presence in a neighborhood.

However, we note that the choice of spatial error versus spatial lag model is fairly subjective. There is also a justification for the spatial lag model. As a result of potential utility in the spatial lag model, we also consider the spatial lag model as a way to measure the sensitivity of the results from the spatial error model.

The final form of the spatial error model is:

$$\begin{aligned}
\text{weapon discovery rate}_i = \beta_0 + \beta_1 \text{Prop Blacks and Hispanics stopped }_i + \beta_2 \text{Prop males stopped}_i \quad (1)\\
+ \beta_4 \text{Prop Hispanics residents}_i + \beta_5 \text{Prop Black residents}_i \quad (2)\\
+ \beta_6 \text{MajorityWhite=Yes}_i \quad (3)\\
+ \beta_7 \text{Total Population}_i + \lambda \mathbf{W} \mathbf{u} + \epsilon \quad (4)
\end{aligned}$$

where i denotes a precinct in NYC, $\mathbf{W}$ denotes the spatial weight matrix, λ denotes the coefficient on the spatially correlated errors, $\mathbf{u}$ denotes spatial errors the model is not accounting for, and ϵ denotes the idiosyncratic error term, $\epsilon \sim N(0, \sigma^2)$.

2.1 Model Diagnostics

Spatial error models relax the standard regression model assumption about the need for errors to be independent. The assumptions for spatial error models are:
Linearity between response and each of the predictors Normality in residuals Constant variance in residuals Independence in outcomes of the observations

We checked our model's normality in residuals assumptions by plotting Q-Q plots, and we examined the model residuals for non-constant variance and outliers. Furthermore, we checked for linearity assumption for all continuous predictors by generating residuals vs predictor plots. We also calculated Pearson correlation coefficient to detect multicollinearity. Evaluation of model diagnostics can be found in Appendix.

2.2 Sensitivity analysis

We also fit a spatial lag model with the same set of predictors, since empirical results from Lagrange Multiplier tests suggest there was sufficient evidence against both null hypotheses. A spatial lag model sees spatial dependence as substantively meaningful. We compare the regression coefficients with those from the spatial lag model as well as those from the OLS model.

3 Results

The p-value associated with the lambda parameter was statistically significant (p-value = 0.001), adding further evidence that the spatial error model is a better fit than the OLS model. The coefficient estimate associated with lambda is both positive and statistically significant. This means that when weapon discovery rate in neighboring precincts increases, so does the weapon discovery rate in each precinct, even after we adjust for the other explanatory variables in our model.

Variable	**Beta**	**95% CI**	**P-Value**
(Intercept)	-0.243	-0.823, 0.337	0.412
Prop. of CPW Suspects who were Blacks or Hispanics	-0.049	-0.213, 0.114	0.553
Prop. of CPW Suspects who were Male	0.590	-0.060, 1.239	0.075
Precinct Population (in 10,000s)	-0.003	-0.008, 0.003	0.309
Majority White: Yes	0.004	-0.075, 0.082	0.930
Prop. of Hispanic residents	-0.071	-0.259, 0.118	0.462
Prop. of Black residents	-0.067	-0.223, 0.090	0.402
lambda	0.408	0.165, 0.650	**<0.001**

4 Discussion

4.1 Interpretation

After accounting for spatial dependence in the errors terms of our model, we ultimately failed to find sufficient evidence supporting our hypothesis that there is a significant association between the precinct-level proportion of blacks and hispanics being stopped and the precinct-level weapon discovery rate. In other words, we failed to find sufficient evidence for racial disparity, measured by precinct-level weapon recovery rate, in our dataset. Holding all else constant, for every 0.01 increase in the precinct-level proportions of blacks and hispanics stopped, the expected weapon discovery rate decreases by 0.049, on average. The direction of association is in line with our hypothesis. However, the 95% confidence interval for this association is (-.213, .114), which includes 0, indicating that this association is not statistically significant. The wide confidence interval (and thus high p-value) associated with our variable of interest might have arisen from having both proportion of Blacks and Hispanics suspects as well as proportion of Hispanic residents as our predictors, which are moderately correlated.

In addition, we found that all the other predictors we accounted for in our model – precinct-level frequency of males stopped, total precinct population, precinct-level proportion of hispanics and blacks populations, and whether a precinct is white majority – had no significant associations with precinct-level weapon discovery rate.

4.2 Sensitivity Analysis (OLS and Spatial)

Similar to the lambda parameter in spatial error model, the coefficient estimate for rho parameter in spatial lag model is both positive and statistically significant. This adds further evidence that there exists positive spatial autocorrelation among the residuals in our model. Either spatial error or spatial lag would be a more appropriate fit than OLS model. However, similar to model output from the spatial error model, none of the predictor variables in the spatial lag model was statistically significantly associated with our response variable. Thus, we conclude that our result regarding SQF disparities is not sensitive to which of the two spatial regression models we used.

4.3 Critique on Methodology and Alternative Model Specifications

An alternative model can improve upon our model by using Census tracts or even ZIP codes, which might control for more granular location-specific details that our precinct-level data failed to capture. Compared to our precinct-level analysis which only has 77 observations, a more granular location level data could also provide larger sample size and thus reduce the chance of Type 2 error. Our model also lacks explicit information about neighborhood historical crime rate. We relied on neighborhood homogeneity as a proxy for crime rate and assumed that majority White precincts are more likely to be low-crime areas (Massey and Denton 1993). A Census tract level SQF stop data can also be joined with Census data to gather information about neighborhood's economic and demographic data.

4.4 Limitations

This study's focus on weapon discovery as the hit rate is unusual. SQF studies have used arrests (Gelman et.al., 2012), but our choice was directed to best answer whether racial skews were observed to exist in NY SQF occurrences at a precinct level. Our hope was to eliminate the bias introduced by using arrest rates since arrest rates are based on police judgement. While weapon discovery rate is more objective and is based on a clearer identifiable observation, this proxy for the police's expectations of racial groups is also imperfect because it excludes stops that may relate to other crimes such as possession of drugs. As a result, the results of this study may not be generalizable to SQF events overall.

4.5 Future Work

In addition to addressing suggestions in alternative model specifications above-mentioned, we would also want to repeat the same analysis with other types of outcome proportion for a broader picture of outcome disparity, and repeat with other types of spatial regression models to ensure our results and conclusions are technique-agnostic. For example, we can model the precinct-level drug discovery rate in drug-related stops, or the precinct-level arrest rate in all SQF stops.

5 Bibliography

Ou, J. (2021), "Case Study 0: Racial Profiling in Stop-Question-Frisk". Sta440.

Gelman A, Fagan J, Kiss A (2012), "Stop-and-frisk policy in the context of claims of racial bias." Journal of the American Statistical Association. 102(479): 813 - 823.

"Stop, Question and Frisk." Publications Reports, NYPD. . Last accessed January 20, 2021. https://www1.nyc.gov/site/nypd/stats/reports-analysis/stopfrisk.page

Goel, Sharad; Rao, Justin M.; Shroff, Ravi (2016), Precinct or prejudice? Understanding racial disparities in New York City's stop-and-frisk policy. Ann. Appl. Stat. 10, no. 1, 365–394.

Ayres, I. (2002a), "Outcome Tests of Racial Disparities in Police Practices," Justice Research and Policy, 4, 131–142. (2002b), Pervasive Prejudice: Unconventional Evidence of Race and Gender Discrimination, Chicago: University of Chicago Press.

Gould, J., and Mastrofski, S. (2004), "Suspect Searches: Assess Police Behavior Under the U.S. Constitution," Criminology and Public Policy, 3, 315–361.

Alpert, G., MacDonald, J. H., and Dunham, R. G. (2005), "Police Suspicion and Discretionary Decision Making During Citizen Stops," Criminology, 43, 407–434.

Massey, D., and Denton, N. (1993), American Apartheid: Segregation and the Making of the Underclass, Cambridge, MA: Harvard University Press.

Smith, M. R., Makarios, M., and Alpert, G. P. (2006), "Differential Suspicion: Theory Specification and Gender Effects in the Traffic Stop Context," Justice Quarterly, 23, 271–295.

6 Appendix

6.1 Additional EDA: correlation and linearity checks

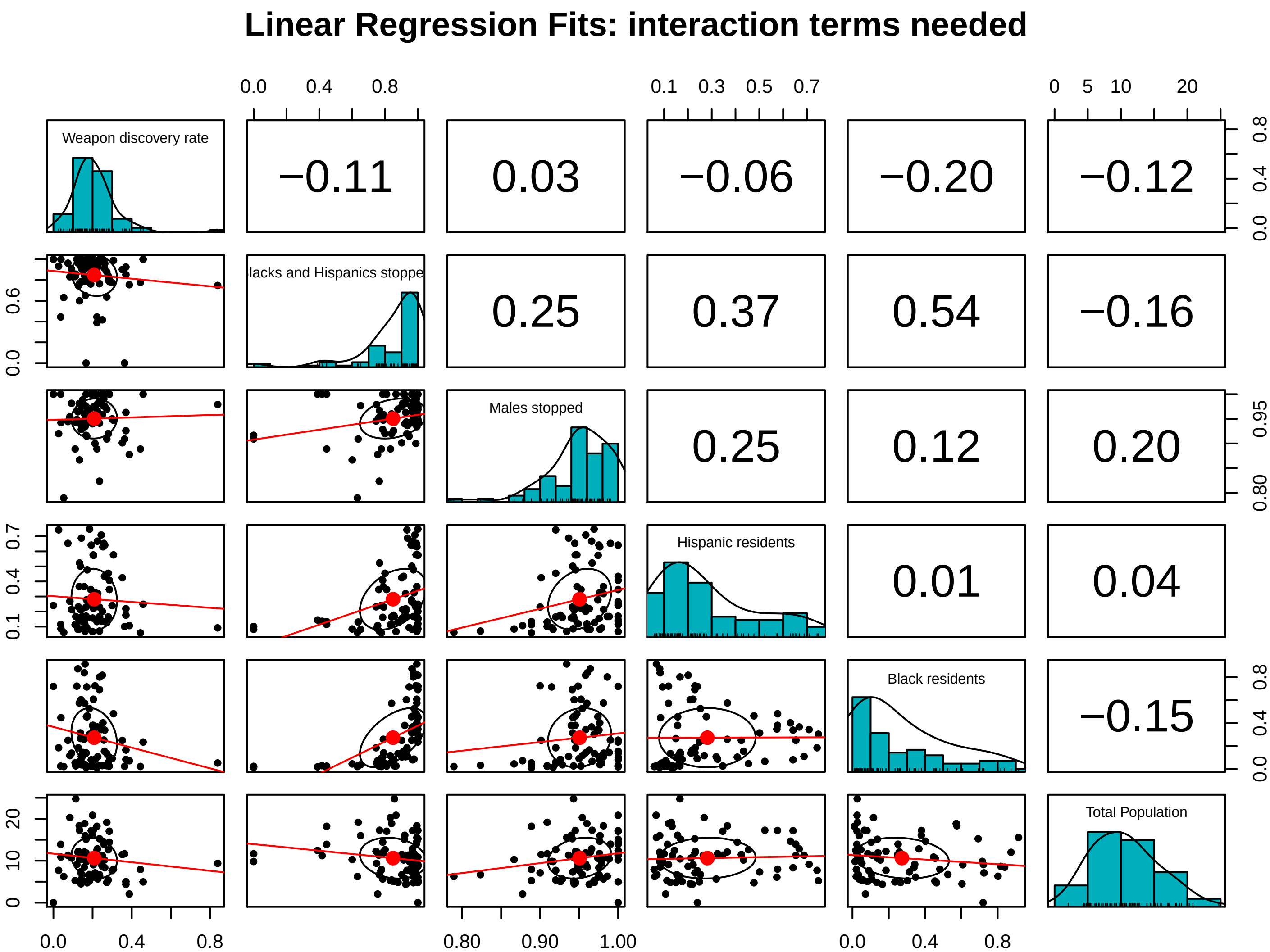
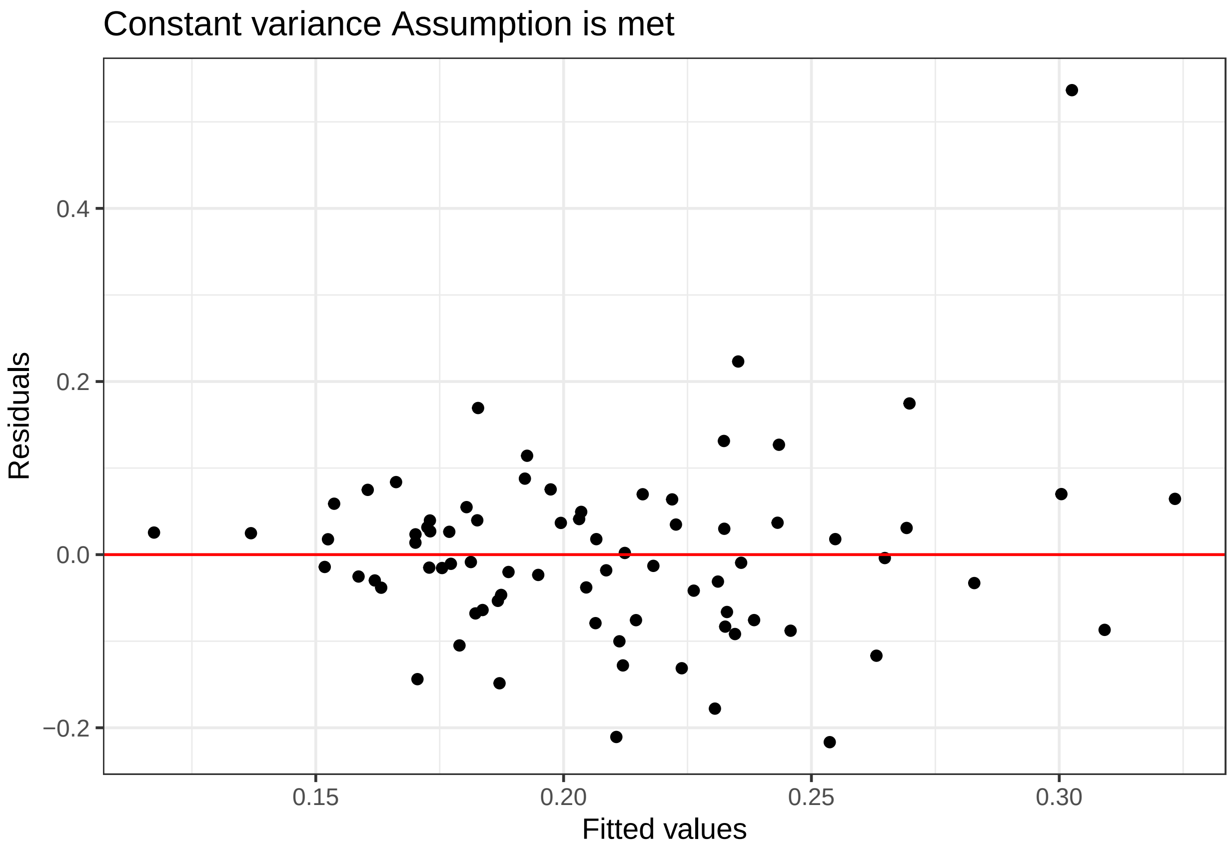


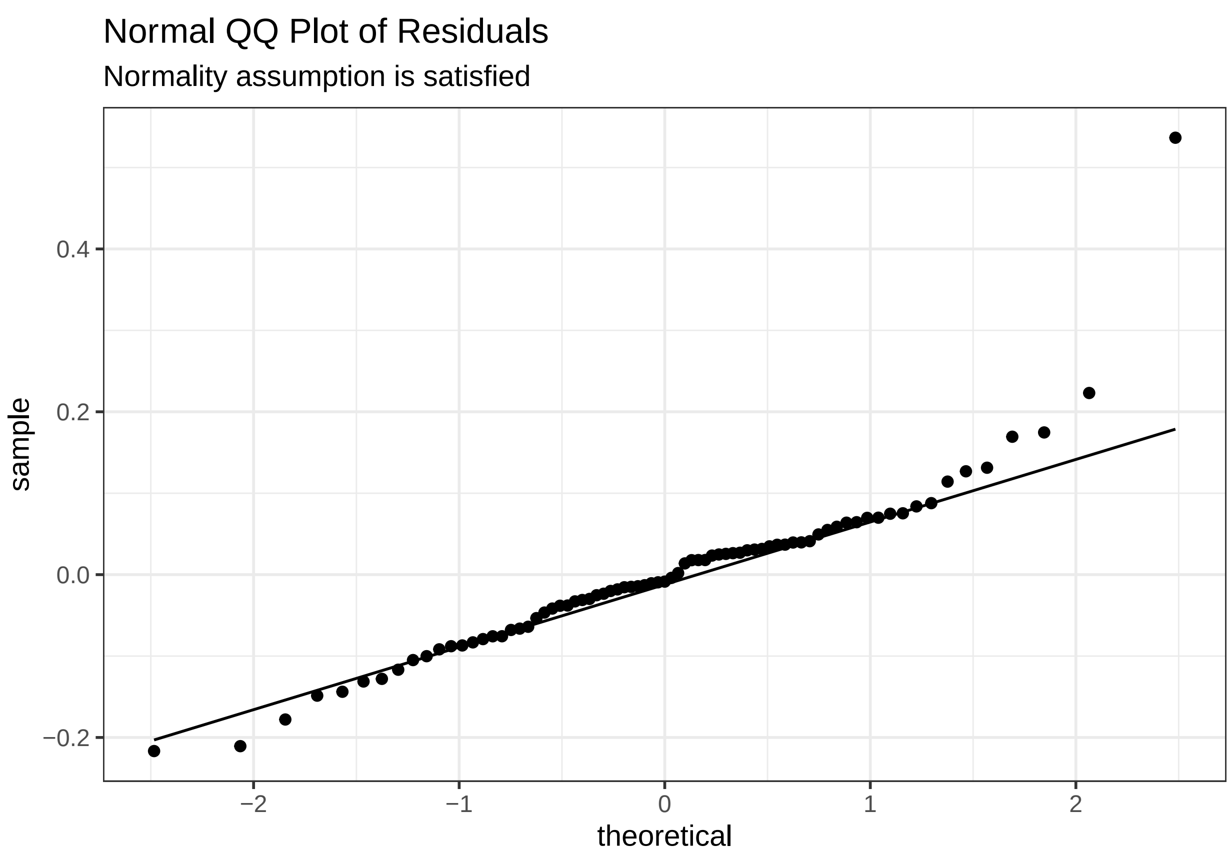
Figure 3: Numeric predictors and precinct-level weapon discovery rate linear regression Fit. Scatter plots are below the diagonal, histograms on the diagonal, and the Pearson correlation above the diagonal. All the numeric predictors except total population are proportions. There appears to be a linear relationship between proportion of Blacks stopped and proportion of Hispanics stopped, and also a linear relationship between percentage of Blacks stopped and percentage of Black residents, indicating a need for interaction terms.

6.2 Model Diagnostics for spatial error model

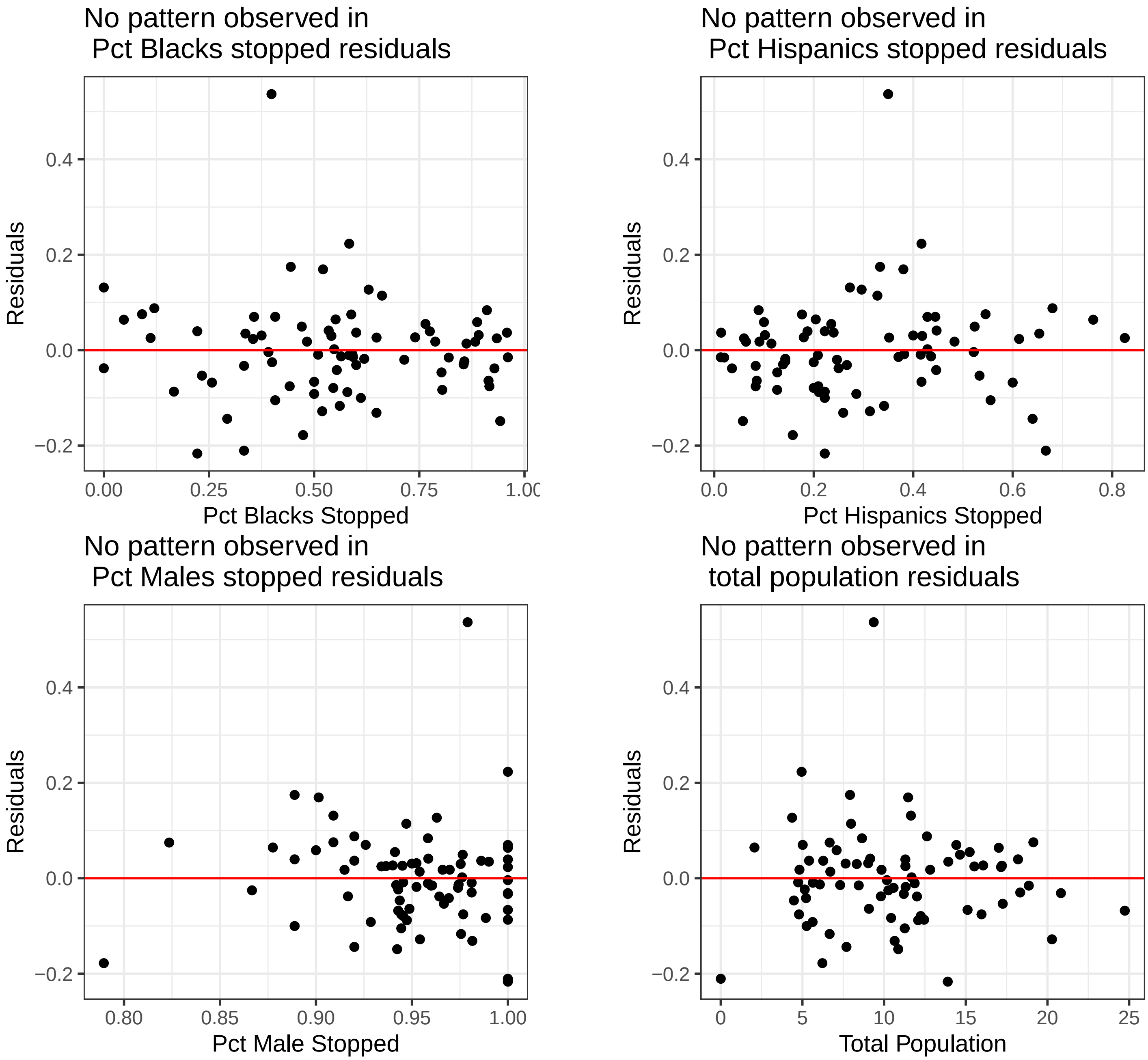
6.2.1 Constant Variance Assumption

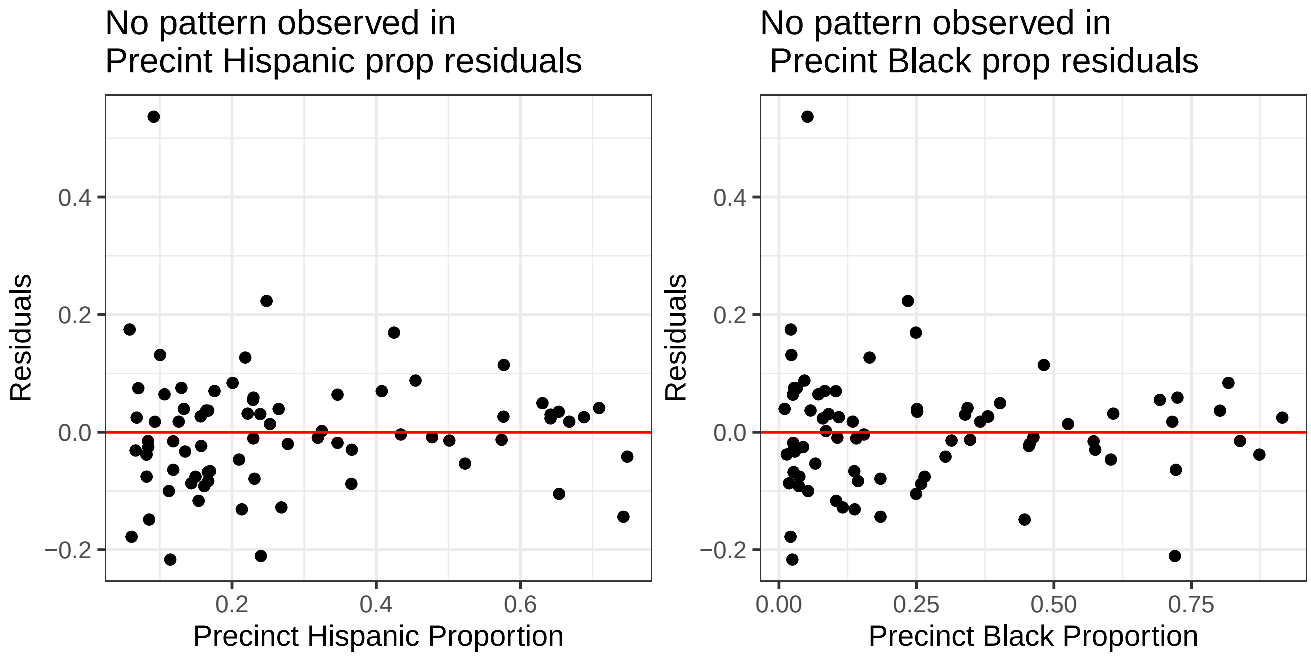


6.2.2 Normality Assumption



6.2.3 Linearity Assumption





6.3 Sensitivity Analysis

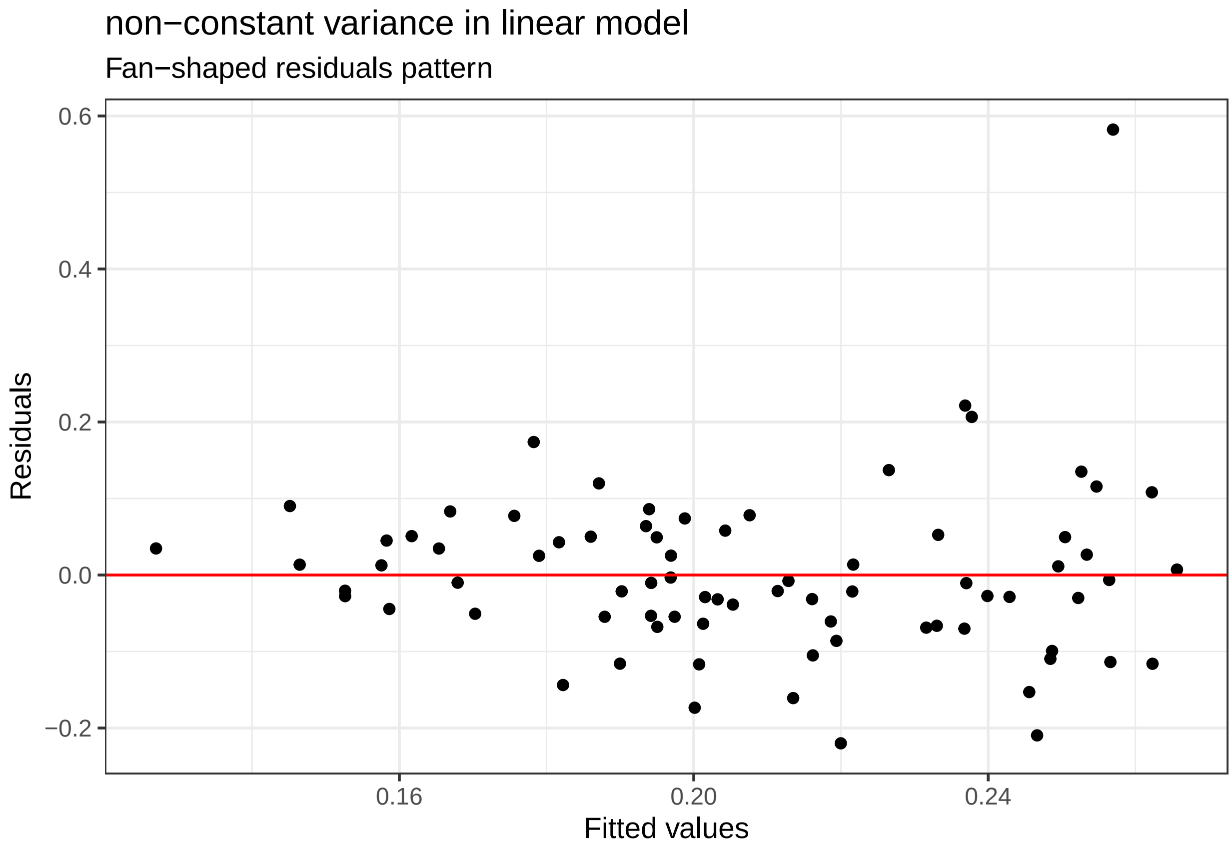
6.3.1 Spatial Lag Model

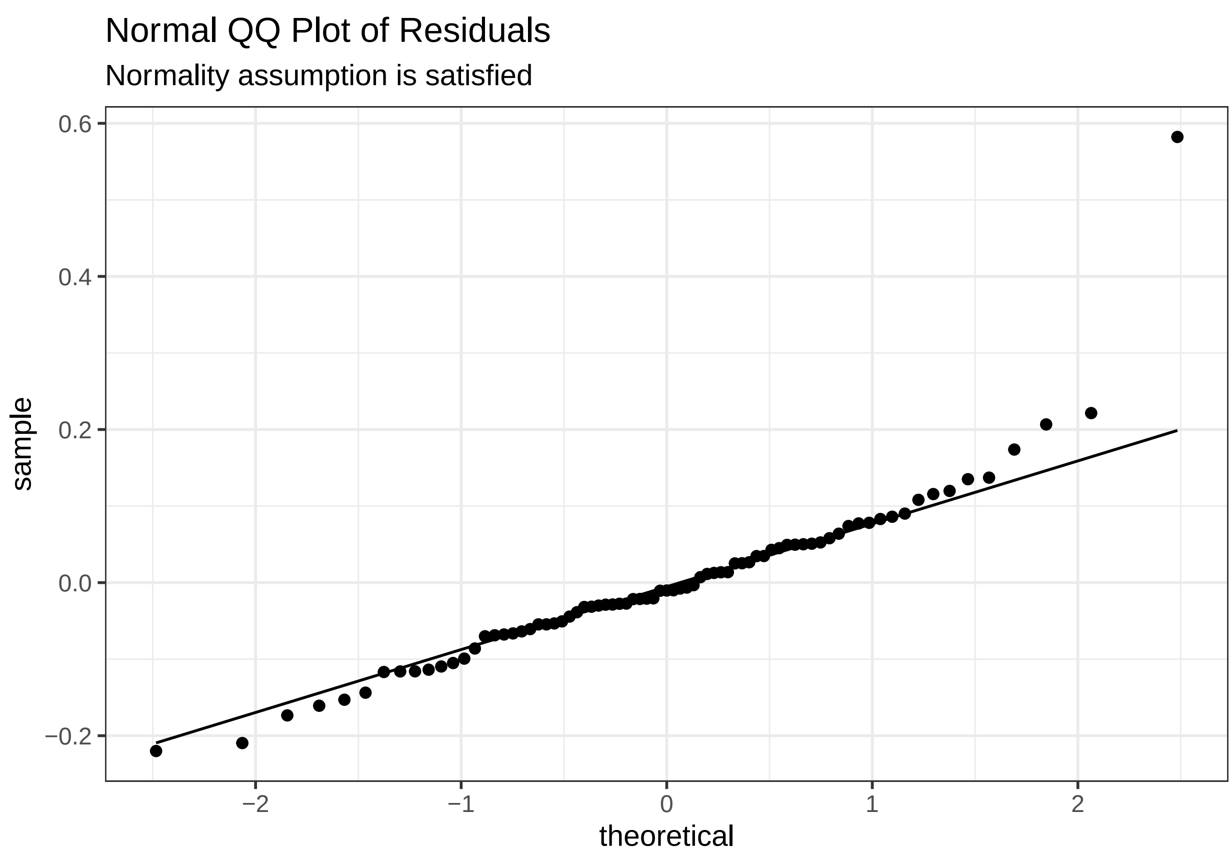
Below are the model coefficients for the spatial lag model:

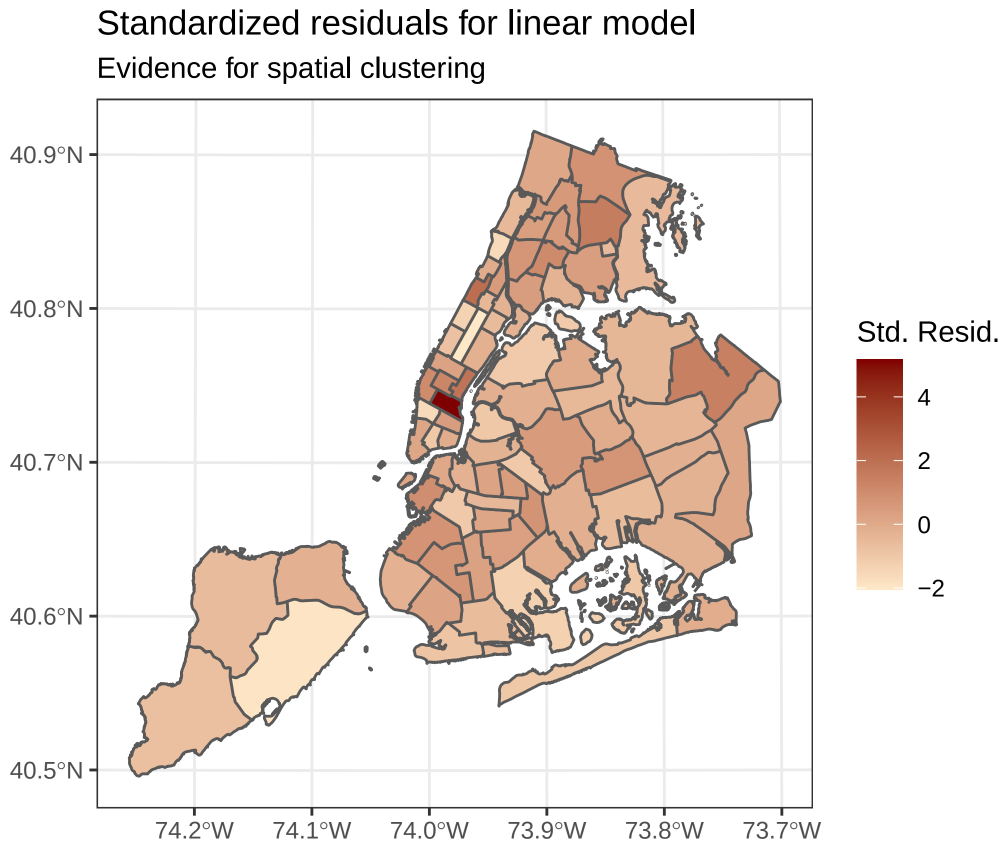
Variable	Beta	95% CI	P-Value
rho	0.374	0.128, 0.619	**0.003**
(Intercept)	-0.220	-0.788, 0.347	0.447
Prop. of CPW Suspects who were Black or Hispanic	-0.027	-0.186, 0.131	0.735
Prop. of CPW Suspects who were Male	0.451	-0.178, 1.079	0.160
Precinct Population (in 10,000s)	-0.004	-0.009, 0.002	0.180
Majority White Precinct	0.014	-0.062, 0.091	0.709
Prop. of Hispanic residents	-0.025	-0.180, 0.130	0.754
Prop. of Black residents	-0.057	-0.198, 0.085	0.430

6.4 OLS Output and Model Diagnostics

Variable	Beta	95% CI	P-Value
(Intercept)	-0.032	-0.673, 0.609	0.920
Prop. of CPW Suspects who were Black or Hispanic	0.003	-0.177, 0.183	0.972
Prop. of CPW Suspects who were Male	0.305	-0.407, 1.018	0.396
Prop. of Hispanic residents	-0.010	-0.186, 0.165	0.905
Precinct Population (in 10,000s)	-0.004	-0.010, 0.002	0.189
Prop. of Black residents	-0.074	-0.234, 0.085	0.356
Majority White Precinct			
Minority			
White	0.029	-0.057, 0.115	0.503


non−constant variance in linear model
Fan−shaped residuals pattern
Residuals
0.6
0.4
0.2
0.0
−0.2
0.16
0.20
0.24
Fitted values


Normal QQ Plot of Residuals
Normality assumption is satisfied
sample
0.6
0.4
0.2
0.0
−0.2
−2
−1
0
1
2
theoretical



	VIF Value
Percentage of Blacks/Hispanics Stopped	1.879
Male Frequency in Precinct	1.188
Proportion of Hispanics in Precinct	1.794
Total Population	1.145
Proportion of Blacks in Precinct	2.424
Majority White Precint	2.601