Willingness To Test Across Different Ethnicities: Evidence From Lockdown in New York City, 2020

Xianglin Li, Junkai Sun

February 2024

Abstract

Many studies have confirmed that there is a huge health inequality across different ethnic groups especially for Black people during COVID-19 and one possible reason is people of certain races may have preferences regarding taking tests. To validate the hypothesis we use distance to black churches as an instrumental variable of racial composition and uncover the causal relationship between racial composition and test participation rate. The main finding is that increasing the share of Black people within a borough would lead to a decrease in the test participation rate when other things are equal, which means people from certain ethnicities are less willing to take the test compared to their white peers. The policy implication from this is that the government could identify coronavirus cases from an early stage by promoting awareness of test-taking or making it compulsory to lower the infection rate within the city and reduce the unequal health outcomes of different ethnic groups during COVID-19.

1 Introduction

According to the report published by the New York City government, the COVID-19 hospitalization rate was far higher than their white peers from NYC-Health (2019). A similar pattern can also be found in the graph on the percentage of patients testing positive for COVID-19 by zip code in New York City the area where the main residents are black or ethnic minorities have the highest percentage of positive testing from Almagro Orane-Hutchinson (2020). The infection rate of Coronavirus within the city not only depends on the percentage of positive test outcomes but also depends on the percentage of people taking the test. For a metropolitan city

like New York City, the test participation rate may differ across ethnicities. Existing literature on this health inequality outcome including Schmitt-Grohé . (2020) focuses more on the former that there is a disparity in the total number of test provisions that the government of New York City may be inclined to set up more test centers in certain boroughs compared to others and our research aims to explore whether it is instead the attitudes towards coronavirus testings within different ethnic groups that caused this health disparity.

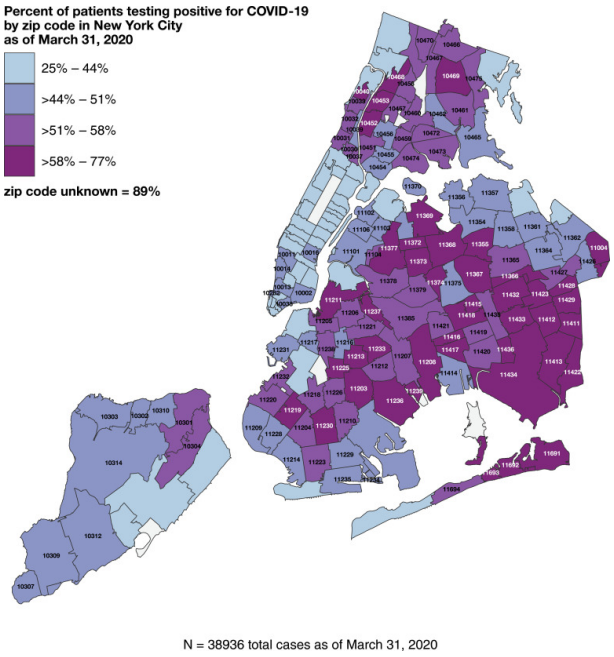

Figure 1: Rate of Positive Tests ($\frac{Positive\ tests}{Total\ tests}$) by Zip Code as of March 31, 2020 Almagro Orane-Hutchinson (2020)

2 Data

The dataset we use is a combination of the 2018 American Community Survey (ACS) and the COVID test data from the Department of Health and Mental Hygiene of New York City (DOH) during the period from April 1_{st}, 2020 to June 9_{th} when the government of the New York City implemented a full lockdown during the period, precisely until the eve of the reopening of the New York City. The former could show the demographic features of each borough at the ZIP level and

the latter could demonstrate the severity of coronavirus spread. We complement our datasets with the church archive on the black presence in the Episcopal Diocese of New York(Churches archive,2019) to understand black people's local history in New York City.

3 Methodology

A simple regression of test participation rate on racial composition controlling for other time-fixed effects and demographic features could not reveal a causal relationship between race ratio and test participation rate that people from certain racial backgrounds are more likely to hold certain preferences regarding taking tests. This is because the racial composition may depend on other factors like educational attainments which may also influence the test participation rate that people with higher educational achievements may be more likely to take tests. To eliminate the endogeneity problem, we use an instrumental variable to filter out the part of the variation in racial composition that is unrelated to other variables that may influence the test participation rate.

We use the church archive on the black presence in the Episcopal Diocese of New York to select all black churches with particular historical meanings to black people and record their corresponding latitude and longitude. We create a variable called distance to the black church by calculating the distance of each ZIP to the nearest Black church in the neighborhood based on comparing the respective latitude and longitude and use it as the instrumental variable for the racial percentage variable within each ZIP because based on the existing literature Wikipedia (2024) black people in the New York City have a local culture of attending the activities organized by the black churches in the neighborhood together and the distance to the black church in each ZIP is an approximate measure of the proportion of black people within that area. To validate the literature, we run a first-stage regression of racial composition on the distance to the black church and find some degree of relevance. Besides, the distance to the black churches is almost as random as assigned, and apart from racial composition, there is no other channel that would significantly influence the positive test rate and test participation rate.

$$y_{it} = y_{it-1} + \beta_{it} \times x_i + \textit{Control variables}\epsilon_{it}$$

Our main regression is in equation above, where y_{it} is the participation rate of test in i_{th} borough at time t; x_i is the share of African American people in i_{th} borough.

To test the hypothesis that the different racial composition would cause a disparity in the test participation rate that people in certain ethnic groups are less willing to take the test even if other things are equal, We adopt a 2sls to regress out the outcome of interest on the race variable instrumented by the distance to black churches and control the time effects and the percentage of positive tests by yesterday.

The total controls can be decomposed into three different parts in this spatial panel data. The first one is the time-fixed effect, which means that time has the same influence on every region. The second is the spatial fixed effect, which means that every borough has some time-invariant features. The last is the interaction effects between the time and location and this is because the time trend of each borough may differ due to the different demographic features of the boroughs. To capture these three different effects, both time-related and demographic-related variables should be included as controls.

Using two-way fixed effects is not a feasible solution because they also absorb the difference in the composition of race of different boroughs, which are our variables of interest. Instead, we include one-way time fixed effects and add demographic features such as the share of different races, mean wage, and mean income as controls, complemented by the positive rate of yesterday's test to account for the fact that residents in the borough will notice the positive rate of their neighborhood and are more likely to go to take tests when they notice a high positive rate.

The traditional HAC standard error measurement considers spatial autocorrelation but fails to account for the settlement patterns of ethnic groups that people from the same race group may tend to live in a compact community Hsiang (2010). Instead, We use the K-means clustering to split boroughs based on the distance and the composition of races to capture the settlement patterns and split the regions into smaller pieces. It turns out that Manhattan is divided into 4 pieces, and the numbers for Bronx, Brooklyn, Queens, and Staten Island are 2,2,3 and 2 respectively.

One potential limitation of our instrumental variable is that we do not have full access to all the black churches in New York City but only those with historical significance and this inadequacy in the number of black churches identified may prevent us from establishing a very strong first stage regression result between the distance to black churches to racial composition.

One alternative to the instrumental variable approach is utilizing Propensity Score Matching (PSM) to draw a causal inference, which can stimulate randomized controlled experiments by matching samples based on similar demographic features. However, the matching mechanism of the PSM depends on the discontinuity of the outcome of interest which could not be applied to our datasets. Besides, PSM could

not control time trends within each borough at the same time, and would not allow the time trend to vary within each borough. Most importantly, PSM could not deal with the problem of endogeneity. Therefore, we stick to the instrumental variable approach compared to the PSM due to these concerns.

4 Result

Table 1 shows that increasing the share of African Americans will lead to a negative effect on the participation rate of the COVID-19 test, which is both statistically significant and economically meaningful that one p.p. increase in the share of African American people in a borough will decrease the participation rate by 0.27 p.p.at 10% significant level.

Due to time constraints, we can not gather adequate data on all the black churches in New York City to generate a strong instrument variable. Considering this limit, we do not adopt the STATA commands ivreg2 and ivregress to do the 2sls, which raises a weak IV problem partially due to our strict choice of standard error. Instead, we did a first-stage regression by hand and specified a normal robust standard, which rejected the existence of a weak IV and over-identification problems. Finally, the demographic features presented in the 2018 American Community survey may not correspond to the scenario in 2020 and most of the columns are not predetermined, which may lead to the problem of bad controls.

Table 1: Main Result

First-Stage		Second-Stage	
Share_black		Participation_rate	
Dis	-0.2714	L_participation_rate	3.4432
	(-1.13)		(1.48)
		IV_Share_black	-0.2113
			(-1.27)
		Mean_age	0.0403
			(0.94)
Constant	22.5409	Constant	6.2041
	(6.17)		(1.26)

5 Conclusion

Based on the empirical results, the main policy implication we can get is that the New York City government can reduce the spread of coronavirus by promoting awareness of test-taking behaviors among certain ethnic groups or making it compulsory to take tests.

Our team consists of 2 members and spends about 11 hours on the project.

References

almagro2020Almagro, M. Orane-Hutchinson, A. 2020. The determinants of the differential exposure to COVID-19 in New York city and their evolution over time The determinants of the differential exposure to covid-19 in new york city and their evolution over time. Journal of Urban Economics103293.

HsingHsiang, SM. 2010. Temperatures and cyclones strongly associated with economic production in the Caribbean and Central America Temperatures and cyclones strongly associated with economic production in the caribbean and central america. Proceedings of the National Academy of Sciences1073515367-15372. https://www.pnas.org/doi/abs/10.1073/pnas.1009510107 10.1073/pnas.1009510107

covid19dataNYC-Health. 2019. COVID-19: Data Archive. Covid-19: Data archive.

schmittgroehe2020Schmitt-Grohé, S., Teoh, K. Uribe, M. 2020. Covid-19: testing inequality in new york city Covid-19: testing inequality in new york city. Covid Econ. Vetted Real-Time Pap.2727–43.

blackchurch2024Wikipedia. 2024. Black Church. Black church. https://en.wikipedia.org/wiki/Black$_c$*hurch*