Econometrics Game Preliminary Round

Shmu Padwa,* Cristopher Rosas,† Zoe Shleifer,‡ and Moses Stewart§
Harvard University
(Dated: February 14, 2024)

We utilize linear Share Positive (sp) with spatial HAC standard errors using data on the early spread of the coronavirus in 2020 in New York City (NYC) to investigate the findings from [1]. Our analysis shows that using the positive tests per-capita as the dependent variable instead of the share of positive tests does not affect the authors' conclusion that occupations are a key explanatory variable for understanding the early spread of COVID-19 in NYC. We also show that when controlling for occupation, race does not have a significant effect on the positive tests per capita. Based on these findings, we recommend that policy makers . . .

I. DATA

We use data on neighborhood characteristics at the zip code level in NYC provided from the 2018 American Community Survey (ACS) combined with COVID test data from the Department of Health and Mental Hygiene of New York City (DOH). This data was included in the *metricsgame2.dta*. We perform our analysis on two weeks in the dataset: April 4th - 10th, and May 16th - 22nd. We ignore days with missing COVID data, although ideally given more time we would repeat our analysis, supplementing the missing data in *metricsgame2.dta* by going to the source data from the ACS and DOH.

II. MODEL

As mentioned in the prompt, disparities in the incidence of testing for COVID-19 and in access to tests have been documented by both [2] and [3]. Therefore, we estimate the effect of zipcode-level features on two different dependent variables: the share of positive tests and the positive tests per capita. By repeating our analysis on both sets of dependent variables, we aim to show that our results are not volatile to differences in the incidence of testing. Ideally, we would have also repeated our analysis with deaths per capita to show that disparities in access to tests also do not influence our results.

[1] suggest that the data is more likely to reflect differences in neighborhoods early in the pandemic. Therefore, similar to [1], we repeat our analysis only using data from the week of April 4th - 10th and May 16th - 22nd to see how the roles of our dependent variables change over time. Precisely, we estimate:

$$y_{iw} = \alpha_w + \beta_w X_i + \epsilon_{iw} \tag{1}$$

For our dependent variable y_{iw} and varying sets of controls X_i. The subscript i denotes the zipcode and $w \in \{\text{April 6th}, \text{May 16th}\}$ denotes the week. To compute standard errors and p-values, we compute spatial HAC standard errors provided by [4].[5] This is more appropriate than a simple heteroscedasticity correction because the latter assumes independence between zip codes. Because these zip codes are right next to each other, features of neighboring zip codes likely affect COVID rates in nearby zip codes, leading to correlation in results and higher standard errors.

With more time, we would have liked to add borough-fixed effects to control for differences between boroughs, as well as repeat the regression with several days or weeks in the data set while maintaining zipcode-fixed effects to compute estimates on more data and plot the changes in coefficients. As well, we were interested in modeling the effect of COVID policy and local rhetoric on COVID rates. While there exists an LLM approach to this measurement of sentiment, with more time we would have implemented an IV approach, using something like registered political affiliation (ie. democrat/republican) as our instrument for rhetoric.

---

* spadwa@college.harvard.edu
† crosas@college.harvard.edu
‡ zoe_shleifer@college.harvard.edu
§ mosesstewart@college.harvard.edu

III. RESULTS

Tables 1 and 3 show the results of our regression with the dependent variable y_{iw} equal to the positive tests per capita for the week of April 6th and May 16th respectively. Tables 2 and 4 repeat this analysis with the share of positive tests as the dependent variable y_{iw} instead.

Focusing on week 1 in tables 1 and 2, we find that across the first full week of the pandemic, the significance of the effect from the share of Hispanic people in a zipcode is unchanged for both dependent variables, even controlling for job occupations. This differs slightly from the results of [1] taken over a single day. However, we reproduce the findings that the share of different job occupations, including Health, Transportation, and Technology, have a significant effect on *both* the number of positive tests per capita and the share of positive tests. Importantly, because this finding is robust against both dependent variables, this provides strong evidence that differences in incidence of testing across zipcodes are not confounding our results.

Comparing the results from week 1 in table 1 and week 2 in table 3, we find that the effect on the positive tests per capita from the share of essential tech workers, service workers, transportation, and law enforcement is smaller and less significant. This substantiates [1]'s hypothesis that the data more strongly reflects neighborhood characteristics early in the pandemic, before quarantine restrictions were put into place. Early in the pandemic, occupations in these industries played a role in the spread of COVID-19.

Another trend was that race/ethnicity markers played a greater role in the per capita regressions than the share positive regressions. For example, 1 finds that the share of Hispanic and Black residents to be statistically significant with slightly positive coefficients, while 2 does not. One understanding of this trend is that variables such as ethnicity might have a more pronounced impact on the positive rate per capita (PPC) if certain demographic groups have different probabilities of being tested within a zipcode.

IV. POLICY IMPLICATIONS

First, our finding that occupations continue to have a significant effect on COVID-19 rates even when using a measure of PPC provides stronger evidence that policymakers can target certain groups or occupations when distributing protective gear. It also provides evidence that targeted quarantine orders on individuals in at-risk occupations can help slow the spread of disease.

Because the share of minority groups has a significant effect on PPC but not SP, there is evidence that some demographic groups did not have equal access to testing within a zipcode. For policy makers to understand the spread of future diseases, it is important to have a representative sample of trends for different demographic groups. Free distribution of COVID-19 tests earlier in the pandemic may have helped improve our understanding of how the virus spread.

At a meta level, our findings emphasize the need for public health officials to adopt a multifaceted approach to their policy determination, specifically in the use of insights from various avenues of measurement to assess the different dimensions of the issue. In this case, this means using various proxies for COVID levels, which, as we have shown, can lead to more robust understanding and more holistic strategies to combat the virus.

V. REGRESSION OUTPUTS

Week 1 with y_{iw} = positives-per-capita (PPC)

Dep. Variable	- Spec. 1 Coeff -	SE 1	- Spec. 2 Coeff -	SE 2
Log Density	-0.000	(0.000)	-0.000	(0.000)
Log Commute Time	0.002	(0.003)	-0.002	(0.003)
Log Household Size	0.005***	(0.002)	0.001	(0.003)
Share Male	0.023*	(0.017)	0.032***	(0.012)
Share Hispanic	0.007***	(0.002)	0.006***	(0.002)
Share Black	-0.004*	(0.003)	0.001	(0.004)
Share Asian	0.006*	(0.004)	0.010***	(0.004)
Share 20-40	-0.201	(0.894)	-0.222	(0.929)
Share 40-60	-1.766*	(1.151)	-1.964**	(1.106)
Share >60	0.630	(0.703)	0.319	(0.912)
Log Mean Income	-0.002*	(0.001)	0.001	(0.002)
Share Pub. Trans.	-0.003	(0.003)	-0.003	(0.003)
Uninsured	-0.015*	(0.011)	-0.018**	(0.010)
Essential - Pro	.	.	-0.011	(0.013)
Nonessential - Pro	.	.	-0.010	(0.011)
Science	.	.	0.039	(0.060)
Legal	.	.	0.003	(0.030)
Health Practice	.	.	0.019	(0.023)
Health Other	.	.	0.030***	(0.014)
Firefighters	.	.	0.046	(0.043)
Law Enforcement	.	.	0.090***	(0.040)
Essential - Service	.	.	0.026*	(0.018)
Nonessential - Service	.	.	-0.027	(0.026)
Industrial	.	.	-0.008	(0.019)
Essential - Tech	.	.	-0.069**	(0.036)
Transportation	.	.	0.053***	(0.026)

FIG. 1. ***: $p < 0.05$, **: $p < 0.10$, *: $p < 0.25$

Week 1 with y_{iw} = share positives (SP)

Dep. Variable	- Spec. 1 Coeff -	SE 1	- Spec. 2 Coeff -	SE 2
Log Density	0.022***	(0.006)	0.017***	(0.007)
Log Commute Time	0.037	(0.051)	-0.062*	(0.052)
Log Household Size	0.205***	(0.047)	0.134***	(0.050)
Share Male	-0.096	(0.164)	0.167	(0.154)
Share Hispanic	-0.041*	(0.030)	-0.046*	(0.029)
Share Black	0.181***	(0.053)	0.175***	(0.054)
Share Asian	0.385***	(0.058)	0.387***	(0.056)
Share 20-40	8.281	(13.644)	-10.905	(12.688)
Share 40-60	-18.698	(19.403)	-29.458**	(16.963)
Share >60	6.619	(12.653)	-6.636	(13.573)
Log Mean Income	-0.077***	(0.017)	-0.034*	(0.027)
Share Pub. Trans.	-0.033	(0.044)	-0.030	(0.046)
Uninsured	0.111	(0.169)	0.089	(0.156)
Essential - Pro	.	.	0.230*	(0.196)
Nonessential - Pro	.	.	0.148*	(0.117)
Science	.	.	-1.369*	(1.024)
Legal	.	.	-0.638*	(0.427)
Health Practice	.	.	-0.036	(0.369)
Health Other	.	.	0.883***	(0.232)
Firefighters	.	.	2.498***	(0.696)
Law Enforcement	.	.	-0.003	(0.655)
Essential - Service	.	.	0.522**	(0.294)
Nonessential - Service	.	.	0.145	(0.446)
Industrial	.	.	0.087	(0.239)
Essential - Tech	.	.	-1.633***	(0.647)
Transportation	.	.	1.107***	(0.402)

FIG. 2. ***: $p < 0.05$, **: $p < 0.10$, *: $p < 0.25$

Week 2 with y_{iw} = positives-per-capita (PPC)

Dep. Variable	- Spec. 1 Coeff -	SE 1	- Spec. 2 Coeff -	SE 2
Log Density	-0.001**	(0.001)	-0.001**	(0.001)
Log Commute Time	0.015***	(0.006)	0.005	(0.007)
Log Household Size	0.011***	(0.005)	-0.004	(0.007)
Share Male	0.036	(0.032)	0.058***	(0.021)
Share Hispanic	0.018***	(0.004)	0.017***	(0.004)
Share Black	-0.022***	(0.007)	-0.011*	(0.008)
Share Asian	-0.009	(0.008)	0.001	(0.008)
Share 20-40	-0.751	(1.704)	-0.672	(1.768)
Share 40-60	-4.768***	(2.141)	-5.947***	(2.210)
Share >60	1.998*	(1.502)	1.367	(1.831)
Log Mean Income	-0.004**	(0.002)	0.004	(0.005)
Share Pub. Trans.	-0.003	(0.006)	-0.005	(0.006)
Uninsured	-0.002	(0.022)	0.000	(0.020)
Essential - Pro	.	.	-0.050**	(0.026)
Nonessential - Pro	.	.	-0.032*	(0.019)
Science	.	.	0.100	(0.119)
Legal	.	.	0.030	(0.054)
Health Practice	.	.	0.048	(0.048)
Health Other	.	.	0.065***	(0.029)
Firefighters	.	.	0.105	(0.093)
Law Enforcement	.	.	0.084	(0.087)
Essential - Service	.	.	0.050*	(0.041)
Nonessential - Service	.	.	-0.088*	(0.054)
Industrial	.	.	-0.037	(0.033)
Essential - Tech	.	.	-0.021	(0.082)
Transportation	.	.	0.122***	(0.049)

FIG. 3. ***: $p < 0.05$, **: $p < 0.10$, *: $p < 0.25$

Week 2 with y_{iw} = share positives (SP)

Dep. Variable	- Spec. 1 Coeff -	SE 1	- Spec. 2 Coeff -	SE 2
Log Density	0.006	(0.005)	0.005	(0.006)
Log Commute Time	0.072**	(0.038)	0.039	(0.039)
Log Household Size	0.163***	(0.033)	0.068***	(0.033)
Share Male	-0.018	(0.163)	0.005	(0.108)
Share Hispanic	0.031*	(0.024)	0.039**	(0.023)
Share Black	0.016	(0.049)	0.064*	(0.045)
Share Asian	0.116***	(0.050)	0.183***	(0.041)
Share 20-40	14.251*	(10.476)	20.267***	(8.559)
Share 40-60	17.928*	(14.507)	12.151	(11.374)
Share >60	22.373***	(9.699)	28.126***	(7.583)
Log Mean Income	-0.052***	(0.012)	0.025	(0.023)
Share Pub. Trans.	-0.022	(0.036)	-0.020	(0.035)
Uninsured	0.160*	(0.109)	0.187**	(0.107)
Essential - Pro	.	.	-0.305***	(0.137)
Nonessential - Pro	.	.	-0.321***	(0.103)
Science	.	.	-0.425	(0.631)
Legal	.	.	-0.116	(0.368)
Health Practice	.	.	-0.055	(0.268)
Health Other	.	.	0.489***	(0.180)
Firefighters	.	.	0.067	(0.433)
Law Enforcement	.	.	0.305	(0.495)
Essential - Service	.	.	0.145	(0.227)
Nonessential - Service	.	.	-0.548**	(0.300)
Industrial	.	.	-0.126	(0.185)
Essential - Tech	.	.	-0.424	(0.482)
Transportation	.	.	0.358*	(0.305)

FIG. 4. ***: $p < 0.05$, **: $p < 0.10$, *: $p < 0.25$

VI. TIME SPENT

Moses: 5 hours
Shmu: 0 hours (sick)
Zoe: 1 hours
Cristopher: 4 hours

[1] M. Almagro and A. Orane-Hutchinson, enJUE Insight: The determinants of the differential exposure to COVID-19 in New York city and their evolution over time, Journal of Urban Economics **127**, 103293 (2022).

[2] G. J. Borjas, enDemographic Determinants of Testing Incidence and COVID-19 Infections in New York City Neighborhoods (2020).

[3] S. Schmitt-Grohé, K. Teoh, and M. Uribe, enCOVID-19: Testing Inequality in New York City (2020).

[4] S. M. Hsiang, enTemperatures and cyclones strongly associated with economic production in the Caribbean and Central America, Proceedings of the National Academy of Sciences **107**, 15367 (2010), publisher: Proceedings of the National Academy of Sciences.

[5] This was our original intention, but we faced errors with the Matlab implementation with imaginary standard errors. No one in our group is fluent in MatLab, so we reimplemented in Python. We left our matlab implementation in the zip file provided.