

1 Prompt

Your task is to use data on the early spread of the coronavirus in 2020 in NYC to identify the channels through which the coronavirus spreads and make some policy recommendations. As a guide, refer to the paper by Almagro and Orane-Hutchinson (2020), who explore the reasons why some neighborhoods are much harder hit by the pandemic than others. Using the 2018 American Community Survey (ACS) and COVID-19 test data from March and April 2020 from the Department of Health and Mental Hygiene of New York City (DOH), they find a relationship between different occupations and the spread of the virus. As a first step you might replicate the coefficient estimates in Table 1 of Almagro and Orane-Hutchinson (2020) with the data provided.

Disparities in the incidence of testing for COVID-19 and in access to tests have been documented by Borjas (2020) and Schmitt-Grohé et. al (2020). Selection into testing which varies by neighborhood may confound the true rate of positives. You may wish to consider several different dependent variables as measures of infection rate. Almagro and Orane-Hutchinson (2020) use the fraction of positive tests to date as their measure of infection rate, but you might use the total number of positive tests per capita (which you can construct from the data provided), or deaths per capita (available to download from the DOH starting in June 2020). Could these alternative measures of infection rates provide a more compelling link between neighborhood characteristics and the spread of COVID-19?

In addition, Almagro and Orane-Hutchinson (2020) suggest that the ACS data is more likely to reflect neighborhood characteristics during the early weeks of the pandemic due to the significant economic impact of the city lockdown. How might the difference in the timing of the ACS data and these measures of exposure affect your estimates?

2 Data

For the purposes of replication, data on neighborhood characteristics at the zip code level is provided from the 2018 American Community Survey (ACS). This is combined with COVID test data from the Department of Health and Mental Hygiene of New York City (DOH). You may use the data provided or supplement it with publicly available data from the DOH, ACS or other sources to answer the prompt. The data provided in the .dta file include:

- Daily data on number of positive tests, total number of tests and fraction of positive tests (to date) ranging from April 1 to May 14, with April 2 and April 6 missing.
- Neighborhood characteristics from the ACS, such as occupations, race, age, population density, commuting patterns and income. The occupations data are further categorized into 13 groups detailed in Almagro and Orane-Hutchinson (2020).

- The variables ‘time’, ‘lat’ and ‘lng’ are provided in the dataset. The authors use spatial HAC standard errors to allow for spatial autocorrelation in the error terms for each neighborhood. If you wish to compute OLS with spatial HAC standard errors as in Almagro and Orane-Hutchinson (2020), you can use the .ado or .m files provided by Hsiang (2010). Go to <http://www.fight-entropy.com/2010/06/standard-error-adjustment-ols-for.html> (for the STATA help file), and for the .ado and .m files: <http://www.globalpolicy.science/code/> (cite Hsiang 2010 if you use this code). The authors use a distance cutoff of 2km. Why might this be more appropriate than a simple heteroskedasticity correction?

References

- [1] **Almagro, M., Orane-Hutchinson, A, 2020.** JUE Insight: “The determinants of the differential exposure to COVID-19 in New York city and their evolution over time”, Journal of Urban Economics, 103293, ISSN 0094-1190.
- [2] **Borjas, G.J. , 2020.** Demographic determinants of testing incidence and covid-19 infections in new york city neighbourhoods. COVID Econ. Vetted Real-Time Pap. (3) 12–39 .
- [3] **Hsiang, S.M., 2010.** Temperatures and cyclones strongly associated with economic production in the caribbean and central america. Proceed. Natl. Acad. Sci. 107 (35), 15367–15372.
- [4] **Schmitt-Grohé, S. , Teoh, K. , Uribe, M. , 2020.** Covid-19: testing inequality in new york city. Covid Econ. Vetted Real-Time Pap. (27) 27–43.