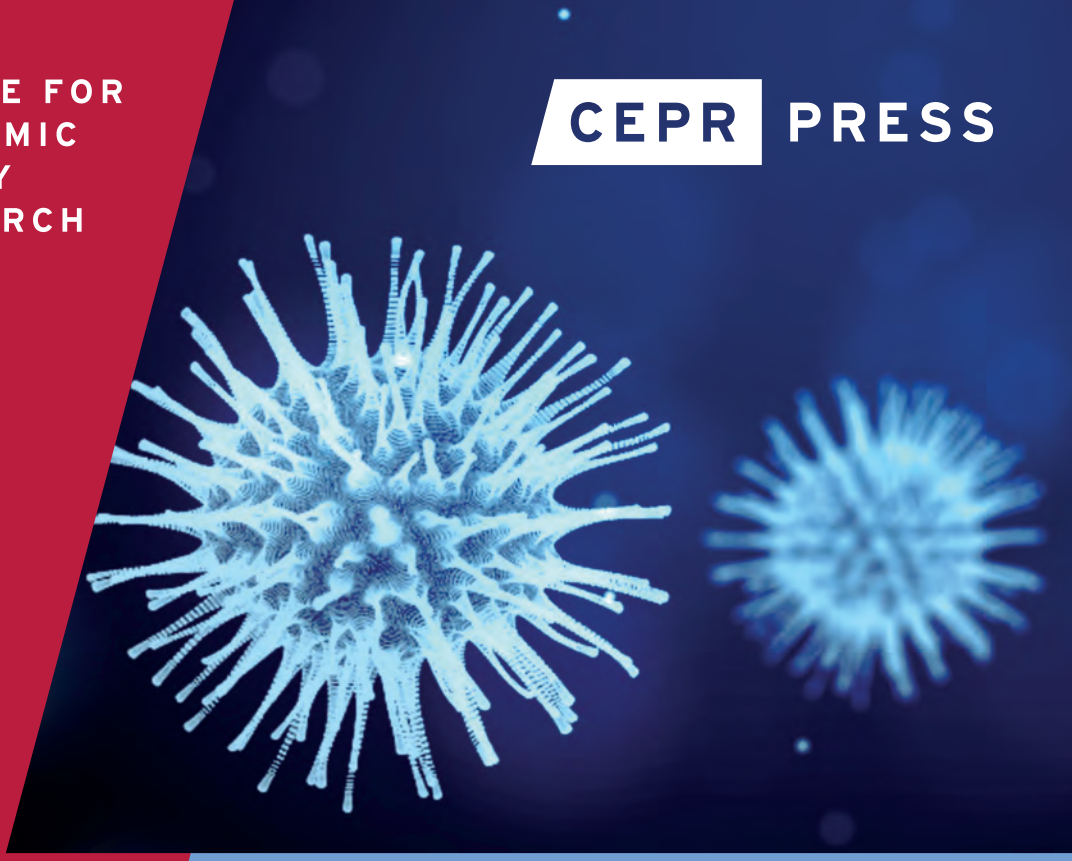
CENTRE FOR ECONOMIC POLICY RESEARCH

CEPR PRESS



COVID ECONOMICS

VETTED AND REAL-TIME PAPERS

ISSUE 3
10 APRIL 2020

Covid Economics
Vetted and Real-Time Papers

Covid Economics, Vetted and Real-Time Papers, from CEPR, brings together formal investigations on the economic issues emanating from the Covid outbreak, based on explicit theory and/or empirical evidence, to improve the knowledge base.

Founder: Beatrice Weder di Mauro, President of CEPR
Editor: Charles Wyplosz, Graduate Institute Geneva and CEPR

Contact: Submissions should be made at https://portal.cepr.org/call-papers-covid-economics-real-time-journal-cej. Other queries should be sent to covidecon@cepr.org.

Ethics

Covid Economics will publish high quality analyses of economic aspects of the health crisis. However, the pandemic also raises a number of complex ethical issues. Economists tend to think about trade-offs, in this case lives vs. costs, patient selection at a time of scarcity, and more. In the spirit of academic freedom, neither the Editors of *Covid Economics* nor CEPR take a stand on these issues and therefore do not bear any responsibility for views expressed in the journal's articles.

Editorial Board

Covid Economics
Vetted and Real-Time Papers

Issue 3, 10 April 2020

Contents

CEPR PRESS

Covid Economics **Issue 3, 10 April 2020**

Data gaps and the policy response to the novel coronavirus

James H. Stock[1]

Date submitted: 4 April 2020; **Date accepted:** 5 April 2020

This note lays out the basic Susceptible-Infected-Recovered (SIR) epidemiological model of contagion, with a target audience of economists who want a framework for understanding the effects of social distancing and containment policies on the evolution of contagion and interactions with the economy. A key parameter, the asymptomatic rate (the fraction of the infected that are not tested under current guidelines), is not well estimated in the literature because tests for the coronavirus have been targeted at the sick and vulnerable, however it could be estimated by random sampling of the population. In this simple model, different policies that yield the same transmission rate β have the same health outcomes but can have very different economic costs. Thus, one way to frame the economics of shutdown policy is as finding the most efficient policies to achieve a given β, then determining the path of β that trades off the economic cost against the cost of excess lives lost by overwhelming the health care system.

Covid Economics 3, 10 April 2020: 1-11

1 Professor of Political Economy, Harvard University. I thank Andy Atkeson, Oleg Itshoki, Erin Lake, and Arash Nekoei for helpful comments.

Covid Economics 3, 10 April 2020: 1-11

This note lays out the basic Susceptible-Infected-Recovered (SIR) epidemiological model of contagion, with a target audience of economists who want a framework for understanding the effects of social distancing and containment policies on the evolution of contagion and interactions with the economy.[1] The model is calibrated to the most recent data. Its simple nature abstracts from many important considerations, and its output is not intended to supersede estimates from more sophisticated epidemiological models.

This note makes four main points. First, the effect of social distancing and business shutdowns on epidemic dynamics enters the model through a single parameter, the case transmission rate β. For a specified case transmission rate, the policy design question is how to achieve that case transmission rate while minimizing economic cost. A second economic question is, what is the optimal case path for β, trading off the economic cost of that path against the costs in deaths. The model serves to focus attention on these questions of central importance in the interaction between health policy and economic policy.

Second, the parameters of the model are not well estimated in the literature on the coronavirus because of the lack of available data. Data on prevalence, for example, is obtained from positive rates of testing for the coronavirus, however so far tests have been rationed and almost entirely have been administered to a selected population, those at risk and showing symptoms. Thus, the fraction of tests that are positive do not estimate the population rate of infection.

Third, using Bayes Law, it is possible to re-express the model in terms of β and the asymptomatic rate, which is the fraction of the infected who show sufficiently mild, or no, symptoms so that they are not tested under current testing guidelines. The virtue of re-expressing the model this way is that it makes use of the positive testing rate, on which there is good data. The COVID-19 asymptomatic rate is unidentified in our model and recent point estimates in the epidemiological literature range from 0.18 to 0.86 (wider if sampling uncertainty is incorporated). However, the asymptomatic rate could be estimated accurately and quickly by testing a random sample of the overall population.

[1] Subsequent to writing this note, two very good and closely related papers have come out as NBER working papers, Atkeson (2020) and Eichenbaum, Rebelo, and Trabant (2020). Atkeson (2020) works through the SIR model and provides simulations for a calibrated version under different isolation scenarios. Eichenbaum, Rebelo, and Trabant (2020) merge the SIR model with a representative agent macro model to track macroeconomic outcomes. Relative to those papers, the contribution here is to illustrate the dependence of the economic outcomes on some parameters that are very poorly determined in the current literature, and to show how the model can be calibrated using data on testing from a sample that is selected under historical testing guidelines.

Covid Economics 3, 10 April 2020: 1-11

Fourth, the policy response and its economic consequences hinge on the value of the asymptomatic rate. We illustrate this for three policy paths for β representing various levels of shutdown and follow-up measures: two "flatten-the-curve" paths and a virus containment path. These different paths would have very different economic consequences, although those are not modeled. The health outcomes on the flatten-the-curve measures depend strongly on the asymptomatic rate.

A simple calibrated epidemiological model

Under the simplifying assumptions that the population mixes homogeneously, that the asymptomatic are as infectious as the symptomatic (possibly not true, see Li et. al. (2020)), that the population is equally susceptible to contagion, and that those who have been previously infected are no longer susceptible, the infection rate follows the so-called SIR model (see for example Allen (2017)). The simple SIR model used here abstracts from mortality. The discrete-time version of the SIR model at the weekly time scale is:

$$\Delta S_t = -\beta I_{t-1} \frac{S_{t-1}}{N} \tag{1}$$

$$\Delta R_t = \gamma I_{t-1}, \tag{2}$$

$$\Delta I_t = \beta I_{t-1} \frac{S_{t-1}}{N} - \gamma I_{t-1} \tag{3}$$

where S_t is the number of susceptible, I_t is the number of infected, R_t is the number of recovered, and N is the (constant) total population. Assuming that everyone in the population is initially susceptible, then $N = S_t + I_t + R_t$. The coefficient β is the transmission rate and γ is the recovery rate.

Equation (1) says that the number of newly infected is the number of prior infected times the transmission rate times the fraction of the population that is susceptible; the number of susceptibles is reduced one-for-one by the number of newly infected. Equation (2) says that the number of recovered increases by the number recovered in the current period. Equation (3) says that the current number of infections increases by the number of new infections, minus the number of recoveries, which follows from the identity $N = S_t + I_t + R_t$.

A common summary of disease transmission is the basic reproduction number, R_0, which is $R_0 = \beta/\gamma$. R_0 is the total number of cases produced by contagion from a single case, when the entire population is susceptible and β and γ are at their no-policy values.

Covid Economics 3, 10 April 2020: 1-11

In this model, policy operates by manipulating the values of the parameters. The baseline values of β and γ can be considered no-policy values. Self-quarantine, social distancing, and school and business closures act to reduce the transmission rate β. Health interventions, such as medical treatment or drugs (should they become available) could serve to increase the recovery rate γ. Policies that decrease β and/or increase γ serve to reduce R_0.

Parameter values and observable implications

The model has two unknown parameters, β and γ. For the coronavirus, surprisingly little data exists to estimate β and γ because testing has been limited and the testing that has been done has largely been targeted to the sick, especially the sick who are either the most vulnerable or who might benefit the most from hospitalization. That is, testing has largely been of the symptomatic. Such testing guidelines miss cases that are variously referred to as asymptomatic or undetected, which different concepts although I treat them here as synonymous. A case can be undetected because the individual has no symptoms, because she has sufficiently mild symptoms (cold or allergy symptoms) that she did not think to report the case, or because she reported her symptoms to a medical professional but did not meet strict guidelines for receiving the test.

Estimation of the model in a conventional sense, that is, fitting (1) - (3) using time series data, is not possible because there are no data on I_t and R_t with which to fit the model. Obtaining estimates of I_t would require ongoing random testing of the population, which has not happened. Similarly, estimates of R_t could be deduced from I_t (given γ), or alternatively could be obtained by ongoing random sampling of tests for serum antibodies in response to the coronavirus, however such tests are not yet widely available and have not been deployed on random samples in the United States.

The absence of random testing of the population poses an additional problem. Although the recovery rate γ for the seriously ill can be estimated from data on those whose disease progression has been tracked, it is not estimated among the asymptomatic. The recovery rate plausibly differs among the symptomatic and asymptomatic, complicating direct estimation of γ from medical case data.

Although there are no data on I_t and R_t, there are widely available data on the results of testing (e.g., Roser, Ritchie, and Ortiz-Ospina (2020)). Because testing in the United States has largely focused on the symptomatic (putting aside small nonrepresentative asymptomatic groups like NBA players), it is

Covid Economics 3, 10 April 2020: 1-11

plausible that the positive testing rate estimates the rate of infection among the symptomatic. Using Bayes law, the model can be augmented to take advantage of time series data on the positive testing rate.

The positive testing rate can be used to calibrate the SIR model as follows. Dividing both sides of (1) - (3) by N expresses all quantities as population rates or, at the individual level, probabilities. Under the simplifying assumption that only the symptomatic are tested, we can use Bayes law to express the positive testing rate in terms of the symptomatic rate (the fraction of infections that are symptomatic):

$$\Pr\left[I_t \middle| Symptomatic_t\right] = \frac{\Pr\left[Symptomatic_t \middle| I_t\right]\Pr\left[I_t\right]}{\Pr\left[Symptomatic_t\right]} = \left(1-\pi_a\right)\frac{\Pr\left[I_t\right]}{\Pr\left[Symptomatic_t\right]} \tag{4}$$

where I_t and S_t refer to the infected and susceptible as above and where $\pi_a = \Pr\left[Asymptomatic_t \middle| I_t\right]$ $= 1 - \Pr\left[Symptomatic_t \middle| I_t\right]$ is the asymptomatic rate (the undetected infection rate).

The fraction of the population that is symptomatic (the denominator in (4)) is,

$$\begin{aligned}\Pr\left[Symptomatic_t\right] &= \Pr\left[Symptomatic_t \middle| I_t\right]\Pr\left[I_t\right] + \Pr\left[Symptomatic_t \middle| S_t\right]\Pr\left[S_t\right] \\ &\quad + \Pr\left[Symptomatic_t \middle| R_t\right]\Pr\left[R_t\right] \\ &= \left(1-\pi_a\right)\Pr\left[I_t\right] + s_0\left(\Pr\left[S_t\right] + \Pr\left[R_t\right]\right),\end{aligned} \tag{5}$$

where s_0 is the baseline rate of symptoms among the susceptible and recovered (normal colds and allergies).

Assuming that testing has been random among the symptomatic, the fraction of tests that are positive estimates $\Pr\left[I_t \middle| Symptomatic_t\right]$. The expanded system (1) - (5) has five equations and four parameters: β, γ, π_a, and s_0.

I do not explore estimation of the model here using time series data on the positive testing rate, although that should be possible. Instead, I illustrate its use and the policy importance of the key parameter π_a, the asymptomatic rate in a calibrated simulation.

Covid Economics 3, 10 April 2020: 1-11

Model calibration and simulation.

I now turn to a numerical illustration of the model and policy interventions. For γ, I assume that the half-life of an infection is 6 days ($\gamma = 0.55$). I set $s_0 = 0.02$. For the week of March 21, 2020, the positive testing rate in the United States was approximately 10% (Roser, Ritchie, and Ortiz-Ospina (2020)). as initial conditions, I assume that there were 50 (unknown) cases in the US in the week ending January 4, 2020. Even with these calibrations, the model (1) - (5) is underidentified by one parameter. I therefore fix the asymptomatic rate π_a at a predetermined value and solve for β such that the positive testing rate is 10% for the week ending March 21, 2020.

The limited available evidence on the asymptomatic rate π_a has been reviewed in Qui (March 20, 2020). Mizumoto et al's (March 12, 2020) estimate of the asymptomatic rate suggested it could be as low as 18%, however that study used data from the *Diamond Princess* which is heavily skewed towards elderly tourists and is thus unlikely to be representative. Other estimates are higher, including 31% (Nishiura et. al. (February 13, 2020)) for 565 Japanese nationals evacuated from Wuhan, however asymptomatic was defined as showing no symptoms, not simply falling short of US testing guidelines. Early data from Iceland suggest an asymptomatic rate of roughly one-half, however that uses a narrow definition of asymptomatic; a looser definition closer to US testing guidelines suggests a much higher asymptomatic rate[2]. Li et. al. (March 16, 2020) estimate a much higher rate of 86% for undetected cases for China. None of these studies are for representative random samples in the United States. Based on this limited evidence, I adopt two values of the asymptomatic rate, 0.30 (for example, used in Pueyo (March 19, 2020) and 0.86.

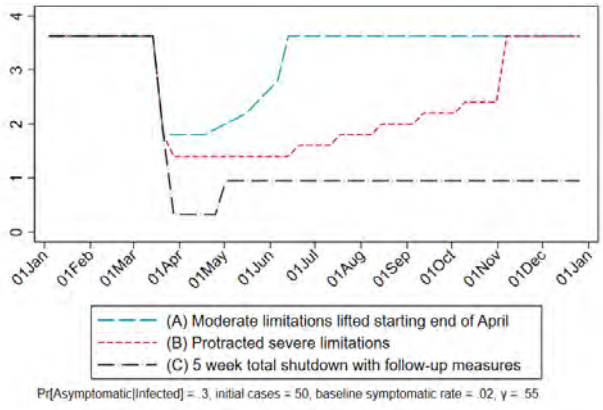
Policy paths. Shutdown policy operates through β. I consider three social distancing/economic shutdown cases. None are optimized and numerical values should not be taken literally. Rather, the point is to illustrate the sensitivity of the outcomes to the asymptomatic rate or, equivalently, to illustrate how different the paths for β (or equivalently, $R_0 = \beta/\gamma$) need to be to achieve a given infection rate under different values of the asymptomatic rate.

[2] Iceland started a large-scale voluntary testing program in the second week of March. A March 15 report on initial results states "Of those samples which have thus far been taken, 700 have been tested. Kári [CEO of deCODE Genetics, the company running the testing] says that about half of those who tested positive have shown no symptoms, and the other half show symptoms have having a regular cold." (The Reykjuavik Grapevine, March 15, 2020 at https://grapevine.is/news/2020/03/15/first-results-of-general-population-screening-about-1-of-icelanders-with-coronavirus/, also see https://www.government.is/news/article/2020/03/15/Large-scale-testing-of-general-population-in-Iceland-underway/.) Under current US testing guidelines, both categories of cases would largely be untested and thus undetected.

Covid Economics 3, 10 April 2020: 1-11

The three shutdown paths of R_0 are shown in Figure 1.[3] Paths (A) and (B) represent two different "flatten the curve" strategies, in which the infection rate is managed so as not to overwhelm the health system but eventually the population achieves herd immunity. In contrast, path (C) aims to suppress the virus until a vaccine is developed. Specifically, path (A) posits moderate shutdowns that continue through the end of April, which are then slowly lifted, with complete lifting in the final week of June. Path (B) posits more severe shutdowns continuing for three months, which are then slowly lifted over the next five months. Path (C) shows what is in effect a total shutdown lasting for three months, with subsequent monitoring, ongoing testing, ongoing social distancing, and contact tracing with quarantining. Under path (C), R_0 is 0.32 for 5 weeks, a value taken from Wang et. al.'s (March 6, 2020) estimate of R_0 in Wuhan after total shutdown measures were imposed (they estimate a pre-policy R_0 of 3.86).

Figure 1. Three policy-induced paths of R_0



The epidemiological outcomes under the three paths are shown in Figures 2-4 for policy paths (A), (B), and (C) respectively. In each figure, the left panel shows the rates of those currently infected and symptomatic (in the notation above, Pr[$Symptomatic_t$, I_t]) and of the ever-infected (currently infected + recovered) for the low value of the asymptomatic rate (0.3), and the right panel shows these paths for the high value (0.86).

The outcomes for the two "flattening the curve" scenarios depend strongly on the (unknown) asymptomatic rate. The least restrictive policy (A) flattens the curve effectively if the asymptomatic rate

[3] The no-policy values of R_0 in these figures is computed for $\pi_a = 0.3$.

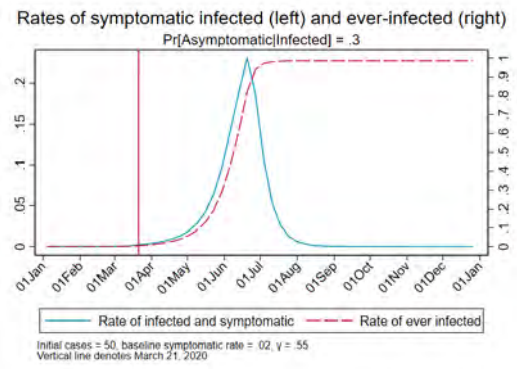
Covid Economics 3, 10 April 2020: 1-11

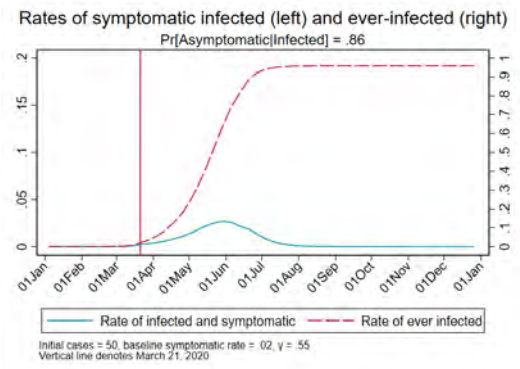
is high, and herd immunity is achieved my mid-summer. If the asymptomatic rate is low, the least restrictive policy results in very high infected symptomatic rates that would overwhelm the health system. The protracted severe limitations policy (B) flattens the curve under the lower asymptomatic rate, although the infected symptomatic rate reaches 5% in late summer. Although I do not model the economic costs, the economic costs of paths (A) and (B) are likely to be very different, with (B) resulting in high costs and a longer deeper recession.

The left panel of Figure 2 and the right panel of Figure 3 represent two high-cost scenarios, the former costly in deaths, the latter costly in economic outcomes; avoiding either requires knowledge of the asymptomatic rate and the rates of infections and recoveries.

The alternative path (C), essentially a total 5-week shutdown of face-to-face interaction as was done in Wuhan, would have the most severe immediate economic costs. Under the parameters here, this suffices to suppress the virus. If R_0 is kept below 1 until a vaccine is developed, then the total rate of ever-infected in the population remains low.

Figure 2. Policy path (A) with low (left) and high (right) asymptomatic rates





Covid Economics 3, 10 April 2020: 1-11

Figure 3. Policy path (B) with low (left) and high (right) asymptomatic rates

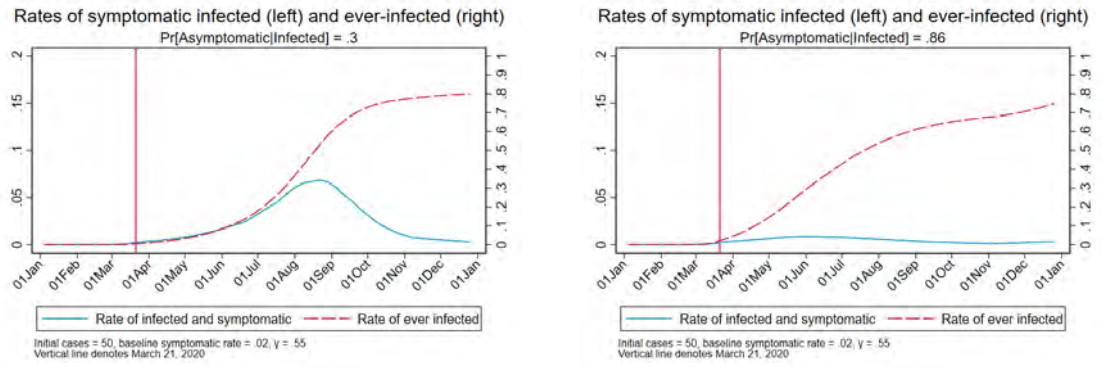
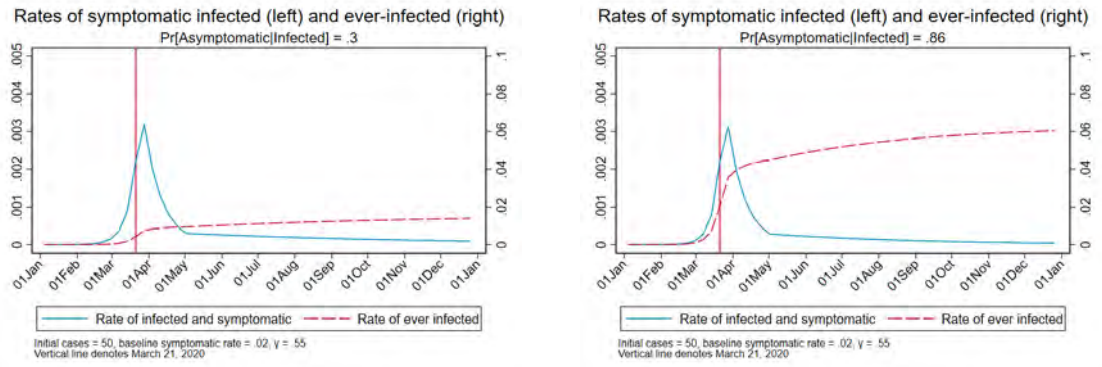


Figure 4. Policy path (C) with low (left) and high (right) asymptomatic rates

Note: Axis values differ from Figures 2 and 3.



Summary

Policy outcomes hinge critically on a key unknown parameter, the fraction of infected who are asymptomatic, and on the current rates of infected and recovered in the population. Evidence on the asymptomatic rate is scanty, however it could readily be estimated by randomized testing.

From an economic policy design perspective, this simplified model has the virtue of summarizing the epidemiological effect of shutdown policies in a single parameter, the contagion parameter β. In this simple model, different policies that yield the same β will have the same health outcomes. However,

different policies might have very different economic costs. Thus, one way to frame the economics of shutdown policies is as finding the most efficient policies to achieve a given β, then solving for the optimal path of β that trades off the economic cost against the cost of excess lives lost by overwhelming the health care system.

References

Allen, L.J.S. (2017). "A primer on stochastic epidemic models: Formulation, numerical simulation, and analysis." *Infectious Disease Modeling* 2(2), 128-142.

Atkeson, A. (2020), "What will be the economic impact of COVID-19 in the US? Rough estimates of disease scenarios," NBER Working Paper 26867, March 2020.

Eichenbaum, M., S. Rebelo, and M. Trabant (2020), "The macroeconomics of epidemics," NBER Working Paper 26882, March 2020.

Li, R. et. al. (2020)., "Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV2)," *Science*, published online March 16, 2020 DOI: 10.1126/science.abb3221 at https://science.sciencemag.org/content/early/2020/03/13/science.abb3221.

Mizumoto, K. et. al., "Estimating the asymptomatic proportion of coronavirus disease 2019 (COVID-19) cases on board the Diamond Princess cruise ship, Yokohama, Japan, 2020," *Eurosurveillance* 25(10), March 12, 2020 at https://www.eurosurveillance.org/content/10.2807/1560-7917.ES.2020.25.10.2000180.

Nishiura, H. et. al. (2020), "Estimation of the asymptomatic ratio of novel coronavirus infections (COVID-19)," forthcoming, *International Journal of Infectious Disease*, posted online February 13, 2020 at https://doi.org/10.1016/j.ijid.2020.03.020.

Pueyo, T. (2020), "Coronavirus: The Hammer and the Dance," *Medium* March 19, 2020 at https://medium.com/@tomaspueyo/coronavirus-the-hammer-and-the-dance-be9337092b56

Covid Economics 3, 10 April 2020: 1-11

Qui, J. (2020). "Covert coronavirus infections could be seeding new outbreaks," *Nature* News, posted online March 20,2020 at https://www.nature.com/articles/d41586-020-00822-x.

Roser, M., H. Ritchie, and E. Ortiz-Ospina (2020). "Coronavirus Disease (COVID-19) Statistics and Research," *Our World in Data* at https://ourworldindata.org/coronavirus.

Wang, C. et. al. (2020), "Evolving Epidemiology and Impact of Non-pharmaceutical Interventions on the Outbreak of Coronavirus Disease 2019 in Wuhan, China," March 6, 2020 at https://www.medrxiv.org/content/10.1101/2020.03.03.20030593v1.

Covid Economics 3, 10 April 2020: 1-11

CEPR PRESS

Covid Economics **Issue 3, 10 April 2020**

Demographic determinants of testing incidence and Covid-19 infections in New York City neighbourhoods

George J. Borjas[1]

Date submitted: 8 April 2020; **Date accepted:** 8 April 2020

New York City is the hot spot of the Covid-19 pandemic in the United States. This paper merges information on the number of tests and the number of infections at the New York City zip code level with demographic and socioeconomic information from the decennial census and the American Community Surveys. People residing in poor or immigrant neighbourhoods were less likely to be tested; but the likelihood that a test was positive was larger in those neighbourhoods, as well as in neighbourhoods with larger households or predominantly black populations. The rate of infection in the population depends on both the frequency of tests and on the fraction of positive tests among those tested. The non-randomness in testing across New York City neighbourhoods indicates that the observed correlation between the rate of infection and the socioeconomic characteristics of a community tells an incomplete story of how the pandemic evolved in a congested urban setting.

1 Robert W. Scrivner Professor of Economics and Social Policy, Harvard Kennedy School. I am grateful to Hugh Cassidy, Daniel Hamermesh, and Gordon Hanson for helpful comments and suggestions.

COVID ECONOMICS
VETTED AND REAL-TIME PAPERS

CEPR PRESS

Demographic Determinants of Testing Incidence and COVID-19 Infections in New York City Neighborhoods

George J. Borjas*

1. Introduction

The apparent difficulty in managing the COVID-19 pandemic raises a multitude of research questions that need to be answered in order to minimize both the human suffering and the economic costs. Why did hot spots arise in some cities or countries, but not in others? Is the pandemic disproportionately affecting particular demographic or socioeconomic groups? Can the destructive path of the pandemic be effectively controlled by a more careful targeting of the scarce testing resources? What is the net cost of the restrictions on geographic mobility, work arrangements, and social gatherings that are now common in countries affected by the pandemic? Addressing all of these questions, however, requires the availability of detailed data that would allow a researcher to search for systematic empirical patterns that might suggest an effective path towards a reduction in future costs.[1]

The New York metropolitan area is currently a hot spot of the COVID-19 pandemic in the United States. The first COVID-19 test in the city was conducted on January 29, 2020, with the first positive result not confirmed until February 23, 2020. Nevertheless, by April 6, 2020, New York City had already identified 68,776 persons infected with the virus, and 2,738 persons had succumbed to the infection.

* Robert W. Scrivner Professor of Economics and Social Policy, Harvard Kennedy School; Research Associate, National Bureau of Economic Research; and Program Coordinator, IZA Program on Labor Mobility, Institute for the Study of Labor. I am grateful to Hugh Cassidy, Daniel Hamermesh, and Gordon Hanson for helpful comments and suggestions.

[1] Economists have already begun to analyze many of the questions raised by the pandemic; see, for example, Baker et al (2020), Bergen, Herkenhoff, and Mongey (2020), Harris (2020, and Lang, Wang, and Yang (2020).

Covid Economics 3, 10 April 2020: 12-39

Beginning on March 30, 2020, the NYC Department of Health and Mental Hygiene began to release detailed information on the number of tests administered and the number of positive results for persons residing in each of 177 zip codes, which I will roughly interpret as urban "neighborhoods."[2] Note, however, that the average population in a zip code in New York City is around 46,000 persons (so that the typical New York neighborhood actually has the population of a small city). This paper exploits the data to identify the demographic characteristics that correlate with the very different patterns in COVID-19 testing and infections across neighborhoods.

Specifically, I merged the neighborhood-level counts of tests administered and positive results with data from both the 2010 decennial census and the 2010-2014 American Community Surveys (ACS). The merging allows me to paint a detailed picture of the neighborhoods where COVID-19 testing was more prevalent and where the number of positive cases was relatively high.

The evidence suggests that a small number of neighborhood characteristics help explain a relatively large fraction of the variance (over 70 percent) in positive test results (conditional on being tested) across New York City neighborhoods. In particular, the conditional probability of a positive test result is far greater for persons living in poor neighborhoods, in neighborhoods where large numbers of people reside together, and in neighborhoods with a large black or immigrant population. At the same time, however, persons residing in poor or immigrant neighborhoods were less likely to be tested.

Much of the discussion over the spread and impact of the COVID-19 epidemic focuses on the evolution of a single statistic: the number of infections per 100,000 persons in the

[2] The data file, however, does not report the number of fatalities resulting from the pandemic.

Covid Economics 3, 10 April 2020: 12-39

population. The analysis of the New York City data starkly illustrates the problem with focusing on this single number. The rate of infection in the population depends on two distinct factors: the frequency of tests in a particular area and the fraction of positive tests among those tested.

Because the incidence of COVID-19 testing was not random across New York City neighborhoods, it is crucial to examine how socioeconomic characteristics correlate with each of the two determinants of the rate of infection. It turns out that some characteristics—for instance, household income—are correlated in opposite directions with the incidence of testing and with the likelihood that a test leads to a positive result. In the end, we might find little correlation between household income and infection rates, but this "zero correlation" masks the fact that household income is related to *both* the rate of testing and the likelihood that a test yields a positive result. Persons residing in poorer neighborhoods were less likely to be tested, but once the test was administered, were more likely to be afflicted with the virus. In short, the finding that the rate of infection in the population is uncorrelated with a particular socioeconomic characteristic may not provide the information that is required to understand the evolution of the pandemic.

Covid Economics 3, 10 April 2020: 12-39

2. Data

The key data file analyzed in this paper reports the cumulative number of COVID-19 tests administered as well as the cumulative number of positive results for each zip code in the City of New York.[3] These data were first released on March 30, 2020 by the NYC Department of Health and Mental Hygiene (and have been updated since). Unless otherwise noted, the analysis in this paper uses the cumulative counts as of April 5, 2020. The data consist of 177

[3] The New York City zip codes are numbered from 10001 to 11697.

identifiable zip codes, with some persons being allocated to a "non-identifiable" category that is excluded from the empirical exercise that follows.

It is important to emphasize that the zip code in the publicly available data refers to the *zip code of residence* for the person who received the test. For expositional convenience, I will use the terms "zip code" and "neighborhood" interchangeably.

Figure 1, prepared by the New York City Department of Health, shows the geographic spread in infection across neighborhoods. Even a superficial look at the map suggests that many neighborhoods in Manhattan have had relatively few COVID-19 cases, while some neighborhoods in the Bronx, Brooklyn, or Queens have seen a very high case load. Some of the initial media reports of these data provide a detailed discussion of the pandemic in specific neighborhoods (Honan, 2020; Buchanan et al, 2020; and Marsh 2020).

The first two rows of Table 1 summarize key characteristics of the data. The cumulative number of persons who had been tested for COVID-19 infection in the average zip code prior to April 5, 2020 was 598.2, but there was a very large variation across neighborhoods in the incidence of testing. Around 219 persons had been tested in the neighborhood at the 10^{th} percentile, and almost 5 times as many (1013) had been tested in the neighborhood at the 90^{th} percentile. The table also shows a similarly large geographic dispersion in the number of persons who tested positive for the virus, from 101 persons in the neighborhood at the 10^{th} percentile to 624 persons in the neighborhood at the 90^{th} percentile.

I merged the data released by the NYC Department of Health with information from other sources, specifically the 2010 decennial census and the 2010-2014 American Community Surveys (ACS). This merging allows me to obtain a relatively simple description of the demographic and socioeconomic characteristics for each of the 177 neighborhoods.

Covid Economics 3, 10 April 2020: 12-39

The demographic data from the 2010 decennial census at the zip code level is publicly available in the data archives maintained by the Census Bureau.[4] The characteristics that I downloaded from the decennial census are: the population of the neighborhood, the fraction of the population that is male, the fraction of the population that is over age 60, the fraction of the population that is minority (i.e. Hispanic, non-Hispanic black, or non-Hispanic Asian), and the number of persons in the average household. Table 1 shows that all of these demographic characteristics vary substantially across NYC neighborhoods. For instance, the mean household size is 2.6 persons, but the range between the 10th and 90th percentile neighborhood goes from 1.7 to 3.2 persons. Similarly, the range for the fraction of the neighborhood's population that is black goes from 1.3 to 52.1 percent, with a mean of 16.1 percent.

I use the neighborhood's population to normalize the data on COVID-19 testing and infections. Specifically, Table 1 also reports both the number of tests and the number of infections per 100,000 persons. In the 10th percentile neighborhood, the incidence of testing was 902.9 persons, while the incidence was twice as large, or 1840.6, in the 90th percentile neighborhood. These large geographic differences in the incidence of testing will play a role in the analysis that follows.

The table also reports the rate of infection per 100,000 persons, calculated simply as the number of persons infected with COVID-19 divided by population (times 100,000). This number also differs substantially across neighborhoods, from 442.2 in the 10th percentile neighborhood to 1107.4 in the 90th percentile neighborhood.

[4] The data are available at census.data.gov.

Covid Economics 3, 10 April 2020: 12-39

Finally, the table reports the percent of tests that confirmed an infection. The fraction of tests that had positive results was 54.0 percent in the average neighborhood, but it ranged from 39.4 in the 10th percentile neighborhood to 64.3 percent in the 90th percentile neighborhood.

Unfortunately, the zip code-level data available from the decennial census does not contain information on the income level of the neighborhood or on other socioeconomic variables that might be relevant in understanding the spread of the pandemic. I used data from the merged 2010-2014 ACS to obtain this type of additional information. In particular, I used the ACS to calculate the mean household income of the neighborhood and the fraction of the neighborhood's population that is foreign-born.[5]

Table 1 also reports the summary statistics for these two additional variables. As with the data drawn from the decennial census, there is a great deal of dispersion across neighborhoods in these characteristics. For example, the range for mean household income goes from $46,800 in the 10th percentile to $164,000 in the 90th percentile, and the range for the fraction of the neighborhood's population that is foreign-born goes from 19.3 to 52.5 percent.

Let V_i be the number of persons in neighborhood i who tested positive for COVID-19, and let P_i be the population of the neighborhood. The fraction of persons in a neighborhood who tested positive (which can be easily translated into the number of persons infected with the virus per 100,000 people) can be written as:

$$\frac{V_i}{P_i} = \left(\frac{T_i}{P_i}\right) \times \left(\frac{V_i}{T_i}\right), \tag{1}$$

[5] These data are downloadable from the usa.com website; http://www.usa.com/rank/.

Covid Economics 3, 10 April 2020: 12-39

where T_i gives the number of residents in the neighborhood who have been tested.

Equation (1) shows that the infection rate in the population (V_i/P_i) depends on two distinct ratios: (1) the rate at which persons were tested; and (2) the rate at which those who were tested, in fact, tested positive for the virus. Testing in the United States was very sporadic in the early stages of the pandemic. Even in a city as large as New York City, fewer than 1,000 persons had been tested by March 10, 2020, and only 110,606 persons (out of a population of 8.6 million) had been tested by April 5, 2020.[6]

Equation (1) also shows that the infection rate in the population depends on the fraction of tests that turn out to be positive. This variable, too, is likely to be correlated with specific socioeconomic characteristics (for example, there may be more infections among people living in neighborhoods where large families cluster together).

It is unclear at the outset whether a given socioeconomic variable x should be similarly correlated (even in sign) with each of the two ratios in the right-hand-side of equation (1). As a result, the correlation between the population infection rate (V_i/P_i) and x may confound the two very different roles that x might be playing in the pandemic.

Let me conclude with an important caveat. The demographic and socioeconomic characteristics chosen to describe the neighborhood were selected not only because they tend to be the usual suspects used in many studies in the social science literature, but also because it is unlikely that they affected the evolution of test frequency and testing results during the month of March 2020.

[6] The cumulative number of tests for March 10 was reported in the testing.csv file prepared by the NYC Department of Health and available at https://github.com/nychealth/coronavirus-data. The file was downloaded on April 2, 2020, but had disappeared from the website by April 5.

Covid Economics 3, 10 April 2020: 12-39

It is not difficult to find neighborhood characteristics that likely affected the allocation of testing resources as the pandemic took hold in the city. By the middle of March 2020, for instance, it became well known that the large Hassidic community in Brooklyn was particularly hard hit by the virus (Stack and Schweber, 2020; Sales, 2020). The Hassidic population in New York City is concentrated in a small number of neighborhoods (Gallagher, 2009).[7] After learning of this specific problem, NYC health officials might have responded to the outbreak by reallocating resources to increase the incidence of testing in the affected neighborhoods. Such a behavioral response would lead to an increased number of tests in Hassidic neighborhoods, with a likelihood that many of those tests were positive.

A positive correlation between, say, the fraction of a neighborhood's population that is Hassidic and the fraction of positive COVID-19 tests would then measure a response to an actual outbreak rather than a pre-existing "predisposition." The socioeconomic variables chosen for the analysis and summarized in Table 1 were, in part, chosen because they tend to avoid this endogeneity problem.[8]

3. Descriptive evidence

Before proceeding to the regression analysis, it is instructive to illustrate graphically the simple correlation between some of the neighborhood characteristics and both the incidence of testing and the likelihood of a positive test result. Figure 2, for example, presents scatter diagrams that show how differences in the mean household income of the neighborhood's

[7] The main zip codes are 11211, 11218, 11219, 11204, and 11230, which include the Williamsburg and Borough Park sections of Brooklyn.

[8] One potential exception is the variable that gives the fraction of the population that is over age 60. As the pandemic progressed, it became apparent that the elderly were severely affected by the illness and testing resources may have been reallocated to neighborhoods that had a disproportionately large number of older persons.

Covid Economics 3, 10 April 2020: 12-39

population relate to: (a) the incidence of testing per 100,000 persons ($V_i/P_i \times 100{,}000$); (b) the percent of tests that confirm an infection ($V_i/T_i \times 100$); and (c) the number of infections per 100,000 persons ($V_i/P_i \times 100{,}000$).

Panel A of the figure shows the relationship between the incidence of testing and mean household income (in logs). It is obvious that there is one outlying zip code where residents were far more likely to be tested. The neighborhood with the largest (normalized) number of tests, with about 4,400 tests per 100,000 persons, has zip code 10018. This zip code roughly encompasses the west side of midtown Manhattan between 34th and 42nd streets, from 5th Avenue to the Hudson River.[9]

The graph suggests a positive relation between the incidence of testing and household income in the neighborhood. In other words, persons residing in wealthier neighborhoods were more likely to be tested than persons residing in poorer neighborhoods.

The middle panel of Figure 2 shows the relation between the probability that a test yields a positive result and mean household income. It is obvious that this probability is far higher in poorer neighborhoods. The figure suggests that the fraction of tests that reveal an existing infection is around 60 percent in neighborhoods at the 10th percentile of the income distribution (where the log of household income is 10.8), but the statistic falls to about 45 percent in neighborhoods that are in the 90th percentile (where the log of household income is 12.0).

[9] It is difficult to understand why the incidence of testing was so high in the west midtown area. There are no media reports suggesting the existence of a notable COVID-19 outbreak in this neighborhood. The second highest incidence of testing, with 2,537 tests per 100,000 persons, has zip code 11370, and it encompasses the Elmhurst area of Queens. That area, however, became notorious in mid-March because of the large number of virus-related admissions and fatalities at the Elmhurst Hospital Center (Rothfeld et al, 2020; Russell, 2020). It seems plausible that health officials might have responded to the crisis by increasing testing in that neighborhood in the last half of March, "artificially" inflating the testing incidence.

Covid Economics 3, 10 April 2020: 12-39

The bottom panel of the figure shows the net impact of these two conflicting correlations by linking the incidence of COVID-19 infection in the population and household income. Although the resulting correlation is negative, it is visually evident that it is relatively weak simply because persons residing in poorer neighborhoods were less likely to be tested and hence have a smaller chance of showing up in the infection counts.

Figure 3 provides an analogous summary of the correlation between the average household size in the neighborhood and the various outcomes. Panel A reveals that the incidence of testing was essentially independent of average family size. This allocation of testing resources might seem questionable given the very strong positive correlation between household size and the likelihood that a test yields a positive result illustrated in Panel B of the figure. In neighborhoods where small households are prevalent, fewer than 50 percent of the tests came back positive. In neighborhoods with larger households, over 60 percent of the tests came back positive.

The combination of these two effects attenuates the relationship illustrated in Panel C of the figure. There is a weaker positive correlation between the incidence of infection in the population and household size simply because testing was not particularly common in the neighborhoods where large families cluster.

Panel A of Figure 4 shows the relationship between the incidence of testing and the minority composition of the neighborhood, defined as the fraction of the neighborhood's population that is either Hispanic, non-Hispanic black, or non-Hispanic Asian. The graph suggests a negative relation between the incidence of testing and the percent of the neighborhood's population that is minority. In other words, persons residing in predominantly

Covid Economics 3, 10 April 2020: 12-39

minority neighborhoods were less likely to be tested than persons residing in mainly white neighborhoods.[10]

The middle panel of Figure 4 shows the relation between the probability that a test yields a positive result and the minority composition of the neighborhood. It is obvious that this probability is far higher in minority neighborhoods. The figure suggests that the fraction of tests that reveal an existing infection is around 45 percent in neighborhoods where 20 percent of the population is minority and increases to nearly 60 percent in neighborhoods that are over 60 percent minority.

The bottom panel of the figure shows the net impact of these two conflicting correlations by linking the incidence of COVID-19 infection in the population (per 100,000 persons) and the percent minority variable. Although the resulting correlation is positive, it is relatively weak simply because persons residing in minority neighborhoods were less likely to be tested and hence have a smaller chance of showing up in the infection counts.

Finally, Figure 5 illustrates the relationship between the outcome variables and the percent of the neighborhood's population that is foreign-born. The top panel of the figure reveals that persons residing in immigrant neighborhoods were less likely to be tested. The middle panel, however, shows a strong positive relation between the probability that a test is positive and percent immigrant. The combination of these two correlations summarized in Panel C leads to a weaker positive relation between the "immigrant-ness" of a neighborhood and the incidence of COVID-19.

[10] One important caveat: The regression analysis presented below conducts a more detailed examination of this correlation and shows that the negative correlation between the incidence of testing and the racial composition of the neighborhood is driven by Asian neighborhoods.

Covid Economics 3, 10 April 2020: 12-39

As this graphical description of the data suggest, it is crucial to examine the distinct channels through which any socioeconomic or demographic variable can affect the incidence of testing and the likelihood that any given test yields a positive result. At least in the early stage of the pandemic in New York City, testing resources were not randomly allocated across neighborhoods, contaminating the correlation between any socioeconomic characteristic and the (observed) population incidence of COVID-19 infection. The next section will estimate regression models that confirm many of the implications of these scatter diagrams even after controlling for the role played by other neighborhood characteristics.

Covid Economics 3, 10 April 2020: 12-39

4. Regression Results

Table 2 reports the main regressions that relate each of the three key outcome variables to the set of socioeconomic characteristics introduced earlier. The regressions are estimated using a grouped logit estimator. For expositional convenience, Table 2 reports the marginal effect of each of the regressors, where the marginal effect is defined as the percentage change in the number of "successes" in the dependent variable.[11]

The regressions confirm the insight emphasized in the previous section—that focusing solely on the incidence of COVID-19 in the population masks a lot of what is actually going on in New York City. Consider, for example, the impact of mean household income (introduced in logs in the regressions). The last column of the table shows that although the relative number of infections in the population is positively related to household income, after holding constant other neighborhood characteristics, the coefficient is not very significant. This finding would

[11] Specifically, the regressions are estimated using the *glogit* command in STATA, and the marginal effects are estimated using the *eydx* option.

lead to the inference that COVID-19 infections affected poor and wealthy NYC neighborhoods in roughly the same way.

However, the regression reported in the first column of the table documents that persons residing in wealthy neighborhoods were tested at much *higher* rates than persons residing in poor neighborhoods. The 90-10 gap in log household income is about 1.2 log points. This implies that the number of tests administered to persons residing in the wealthier neighborhoods was about 21.6 percent greater than the number of tests administered to persons residing in the poorer neighborhoods (for a given population size).

At the same time, however, the persons who were tested and who resided in the wealthy neighborhoods were *less likely* to test positive for the virus.[12] The number of positive test results declines by about 5.5 percent when comparing the 90th percentile neighborhood with the 10th percentile neighborhood (for a given number of tests administered). In short, the analysis of the distinct effects of household wealth on the incidence of testing and on the conditional probability of testing positive (holding constant other neighborhood characteristics) provides a different picture than the one implied by the weaker positive correlation between household income and the incidence of COVID-19 infections in the population.

It is important to note that the correlation between household income and the various outcome variables reported in Table 2 can be interpreted in different ways, and that the currently available data do not allow us to distinguish among alternative explanations. For instance, the regressions indicate that residents from wealthier NYC neighborhoods were tested relatively more often than their counterparts from poorer neighborhoods. But *why* did this difference arise?

[12] It is worth noting that the R-squared of the grouped logit regression on the conditional probability of a positive test result is quite high, exceeding 0.7. A small number of neighborhood characteristics do a remarkably good job of explaining the geographic dispersion in this probability.

Covid Economics 3, 10 April 2020: 12-39

It is possible that the scarce testing resources were disproportionately allocated to wealthier neighborhoods. But it is also possible that persons residing in the wealthier neighborhoods might have had better information networks or could more easily afford to obtain tests elsewhere in the city. It would be important to resolve this puzzle, but the publicly available data released by the NYC Department of Health reports only the zip code of residence for the persons tested and does not report the zip code where the test was administered.

The regression in Table 2 shows that household size is an important determinant of the various outcome variables, and in this case the effects on the two components of the infection rate work in the same direction, so that the correlation between the incidence of infection in the population and household size is very strong. In particular, persons residing in neighborhoods where large families were cluster were far more likely to be tested. Similarly, given that a test was administered, infections were much more likely to be detected in those neighborhoods. This is not surprising, of course, as the grouping together of a larger number of people raises the risk of exposure to COVID-19. The fact that the two effects work in the same direction leads to a very strong correlation between household size and the incidence of infections in the population: One additional person in the household increases the number of infections by about 46.4 percent.[13]

Table 2 also shows that predominantly male neighborhoods were more likely to be affected by COVID-19 infections. This finding, however, is solely by the fact that those neighborhoods were tested more intensively. The relative number of positive cases among persons who were tested is independent of the gender composition of the neighborhood's

[13] Although this may seem implausibly large, it is consistent with the raw data illustrated in Panel C of Figure 3. The figure suggests that an increase in household size from 1.5 to 2.5 increases the number of infections per 100,000 persons from about 500 to about 750, a 50 percent increase.

Covid Economics 3, 10 April 2020: 12-39

population. There is a similar pattern in the correlation between infections and the percent of the neighborhood's population that is over age 60. Persons residing in these neighborhoods were also far more likely to be tested, but the outcome of the test was independent of the age composition of the neighborhood.

The regressions examine the link between minority neighborhoods and COVID-19 infections in more detail than the graphical analysis presented in the previous section. In particular, the regression introduces a vector of variables measuring the percent of the neighborhood's population that belongs to each of the three main ethnic/racial groups. The analysis reveals striking differences across the groups.

Persons residing in neighborhoods with a predominantly black or, to a lesser extent, Hispanic population were more likely to be tested. In the case of black neighborhoods, these tests were also more likely to result in the detection of an infection. As a result, there is a strong positive correlation between the incidence of COVID-19 infections and the percent black (and, to a lesser extent, percent Hispanic) variables.

The regression also shows, however, that the sign of these correlations is reversed for Asian neighborhoods. Persons residing in those neighborhoods were less likely to be tested. Once tested, the tests were less likely to come back positive. As a result, the incidence of COVID-19 infections among persons residing in predominantly Asian neighborhoods was far lower than for persons residing in other types of neighborhoods.

Given the geographic origin of the COVID-19 pandemic in Wuhan, China, the results regarding the incidence of testing and frequency of positive test results in Asian neighborhoods are unexpected. Note, however, that 60 percent of Asian immigrants in the city were not born in China. Nevertheless, it would be interesting to uncover why the spread of the virus in

Covid Economics 3, 10 April 2020: 12-39

neighborhoods with a large Asian population behaved in different ways than the spread in black or Hispanic (or white) neighborhoods.

Finally, New York City has long been an urban area that attracts very large numbers of immigrants. The regression reported in the last column of Table 2 indicates that the "immigrant-ness" of the neighborhood's population is uncorrelated with the incidence of an infection in the population (holding the other neighborhood characteristics constant). In other words, the number of infections per 100,000 persons was essentially the same for persons living in immigrant or in "native" neighborhoods. However, this result arises because persons in immigrant neighborhoods were less likely to be tested, and far more likely to test positive once the test was administered. The evidence regarding the correlation between immigrant neighborhoods and CUVID-19 infections in New York City, therefore, again reveals how the nonrandom allocation of testing resources in the city affects the perception of which groups are disproportionately affected by the virus.

I should note that the regression analysis summarized in Table 2 only examines the role of a small number of variables that hope to capture key aspects of how the COVID-19 pandemic affected the various neighborhoods of New York City. By focusing on a small number of regressors, the analysis avoids the problem of multicollinearity that a kitchen-sink approach would introduce, particularly given the limited nature of the sample (a single cross-section of outcomes summarizing the cumulative impact of the pandemic on NYC as of April 5, 2020) and the relatively small number of observations.

Nevertheless, the regressions ignored the obvious fact that the various neighborhoods are located in one of the five distinct boroughs that make up the City of New York (Manhattan, the Bronx, Brooklyn, Queens, and Staten Island). Figure 1 suggests that there may be noticeable

Covid Economics 3, 10 April 2020: 12-39

differences across the five boroughs in the incidence of COVID-19 infections per 100,000 persons (which, of course, partly reflect the different kinds of neighborhoods that compose the various boroughs).

Table 3 re-estimated the regressions after adding a set of borough fixed effects. The results tend to resemble those reported in Table 2, although many of the coefficients are not statistically significant. For example, the correlation between household income and the various outcome variables is weaker once the regression controls for borough of residence. This is not surprising, as it is well known that Manhattan has a far higher level of household income than the other boroughs, so that the Manhattan fixed effect attenuates the impact of household income on the incidence of testing and on the likelihood of positive test results.[14] Nevertheless, the regressions in Table 3 suggest that—even *within* boroughs—the population in the neighborhoods that were predominantly black or Hispanic, or had larger households, or had a large immigrant population, are more likely to test positive once tested. Similarly, the population in the neighborhoods where large households cluster or are predominantly male—even within a borough—are still more likely to be tested.[15]

6. Summary

This paper documents the characteristics of New York City neighborhoods that were most affected by the COVID-19 pandemic. It uses data compiled by the New York City

[14] Manhattan's average household income is $125,900, as compared to between $48,700 and $90,300 for the other four boroughs.

[15] The coefficients of the borough fixed effects in Table 3 suggest the presence of borough-wide impacts that would be interesting to pursue in further research, as they may provide insight into the evolution of the pandemic. For example, persons residing in Brooklyn had a relatively lower incidence of testing than persons from other boroughs, while persons residing in the Bronx had a relatively higher incidence. In addition, the likelihood a test yielded a positive result was far higher in Brooklyn and Queens than in the other boroughs.

Covid Economics 3, 10 April 2020: 12-39

Department of Health and Mental Hygiene that reports the number of COVID-19 tests administered and the number of persons infected at the level of a zip code. I merged these administrative counts with data from the 2010 decennial census and the 2010-2014 American Community Surveys to describe the neighborhoods where COVID-19 testing was relatively more common and where the number of positive test results was relatively high.

The probability of a positive test result (conditional on testing) is larger in poorer neighborhoods, in neighborhoods where large numbers of people reside together, and in neighborhoods with a large black or immigrant population. At the same time, however, persons residing in poorer or immigrant neighborhoods were less likely to be tested.

The rate of infection in a given population (i.e., the number of infected persons divided by the size of the population) depends on two separate factors: the frequency of tests and the fraction of positive tests among those tested. The New York City experience suggests that the net correlation between socioeconomic characteristics and the rate of infection sometimes captures the net impact of perhaps two conflicting forces. As a result, an understanding of which types of neighborhoods are disproportionately affected by the pandemic requires an examination of how socioeconomic characteristics correlate with each of the two determinants of the rate of infection in the population.

Covid Economics 3, 10 April 2020: 12-39

References

Baker, Scott R. R.A. Farrokhnia, Steffen Meyer, Michaela Pagel, and Constantine Yannelis, "How Does Household Spending Respond to an Epidemic Consumption During the 2020 COVID-19 Pandemic," NBER Working Paper No. 26949, April 2020.

Berger. David W., Kyle F. Herkenhoff, and Simon Mongey, "An SEIR Infectious Disease Model with Testing and Conditional Quarantine," NBER Working Paper No. 26901, March 2020.

Buchanan, Larry, Jugal K. Patel, Brian M. Rosenthal, and Anjali Singhvi, "A Month of Coronavirus in New York City: See the Hardest-Hit Areas," *New York Times*, April 1, 2020.

Gallagher, Patrick. "Identification and Analysis of Orthodox Jewish Enclaves in Brooklyn, New York: A GIS Based Approach," *Middle States Geographer*, 2009, 42: 83-89
Russell, David. "Elmhurst deals with coronavirus crisis," Queens Chronicle, April 2, 2020. https://www.qchron.com/editions/central/elmhurst-deals-with-coronavirus-crisis/article_c4e30040-cc13-5762-bfa1-4521d1daafc7.html.

Harris, Jeffrey E. "The Coronavirus Epidemic Curve is Already Flattening in New York City," NBER Working Paper No. 26917, April 2020.

Honan, Katie. "The New York Neighborhoods with the Most Coronavirus Cases," *Wall Street Journal*, April 1, 2020.

Lang. Hanming, Long Wang, Yang Yang, "Human Mobility Restrictions and the Spread of the Novel Coronavirus (2019-nCoV) in China," NBER Working Paper No. 26906, April 2020.

Marsh, Julia. "NYC Map Shows Total Cases Testing Positive for Coronavirus by ZIP Code," *New York Post*, April 1, 2020.

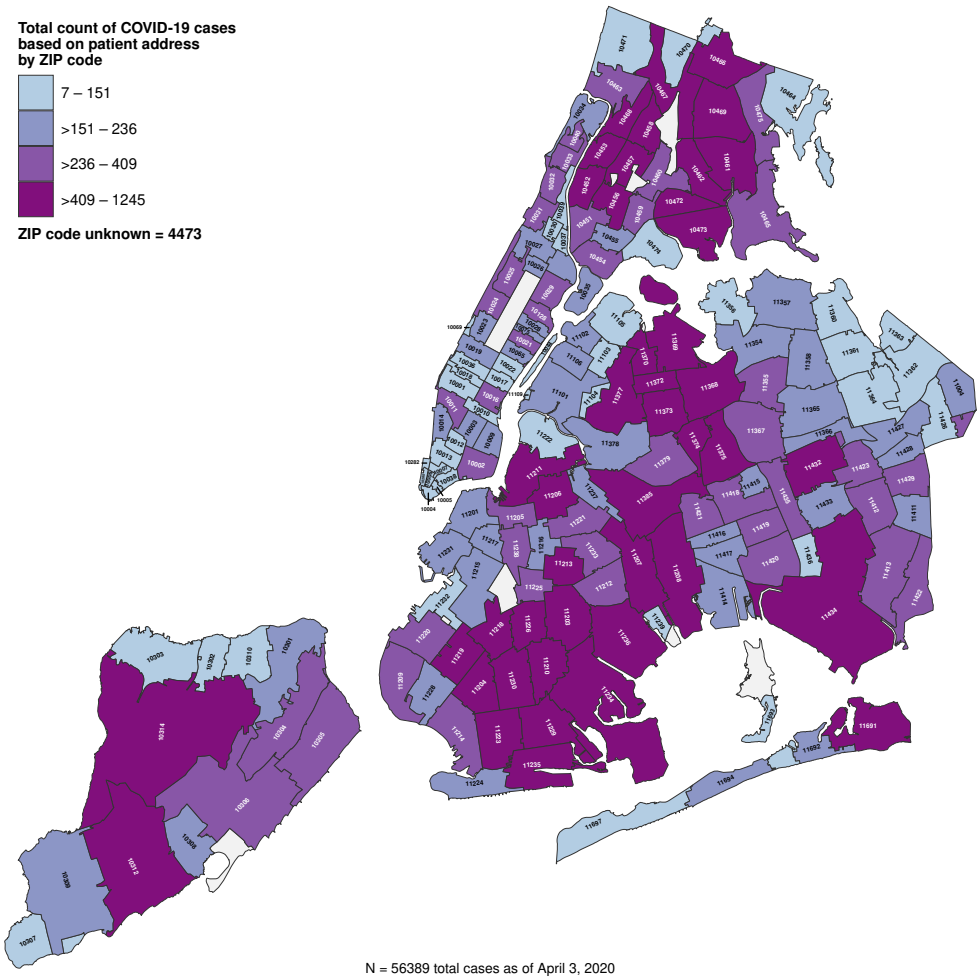
Rothfeld, Michael, Somini Sengupta, Joseph Goldstein, and Brian M. Rosenthal. **"**13 Deaths in a Day: An 'Apocalyptic' Coronavirus Surge at an N.Y.C. Hospital," *New York Times*, March 25, 2020.

Sales, Ben. "Brooklyn's Orthodox neighborhoods have especially high rates of the coronavirus," Jewish Telegraphic Agency, April 2, 2020. https://www.jta.org/2020/04/02/united-states/brooklyns-orthodox-neighborhoods-have-especially-high-rates-of-coronavirus.

Stack, Liam and Nate Schweber, "Coronavirus: 'Huge Spike' in Brooklyn Hasidic Community," New York Times, March 18, 2020. https://www.nytimes.com/2020/03/18/nyregion/Coronavirus-brooklyn-hasidic-jews.html.

Covid Economics 3, 10 April 2020: 12-39

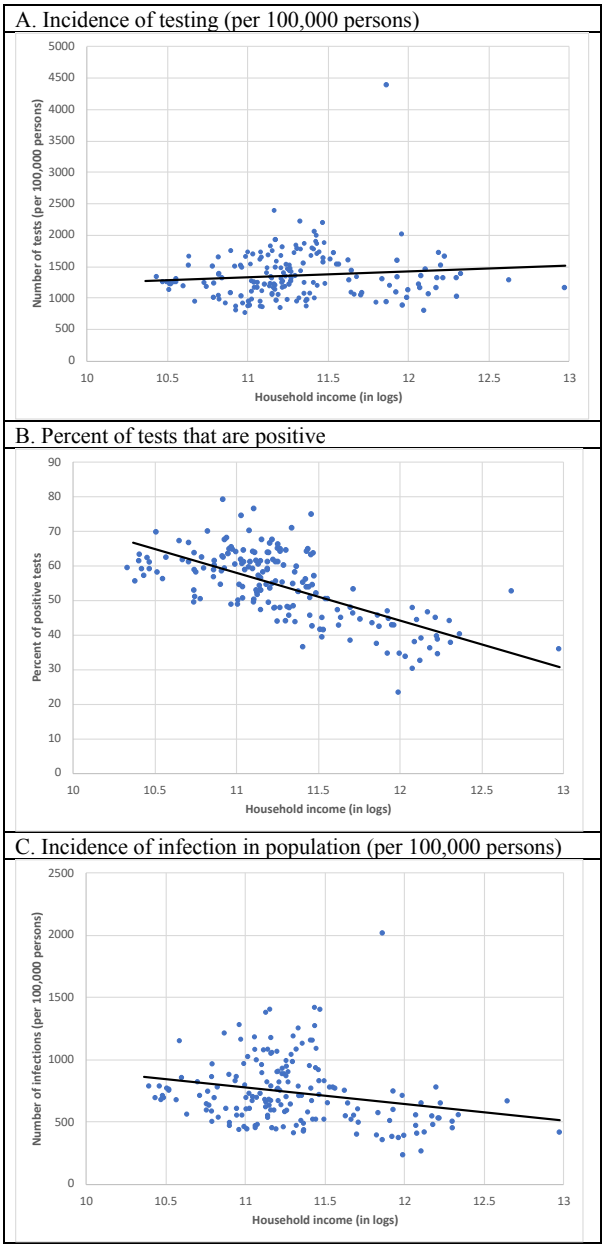
Figure 1. Covid-19 cases in New York City, by zip code (as of April 3, 2020)



Source: Map prepared by the NYC Department of Health; https://www1.nyc.gov/assets/doh/downloads/pdf/imm/covid-19-cases-by-zip-04032020-1.pdf.

Covid Economics 3, 10 April 2020: 12-39

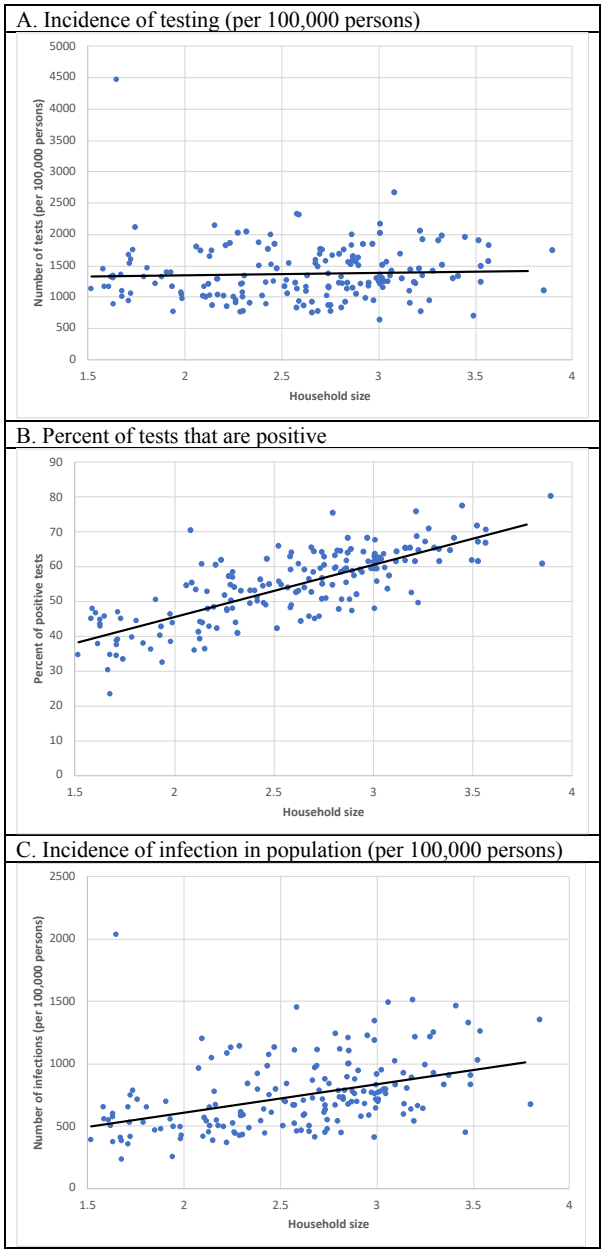
Figure 2. Household income and COVID-19



Notes: The income variable gives the household income for the average household in the zip code. See the text for details on the construction of the variables.

Covid Economics 3, 10 April 2020: 12-39

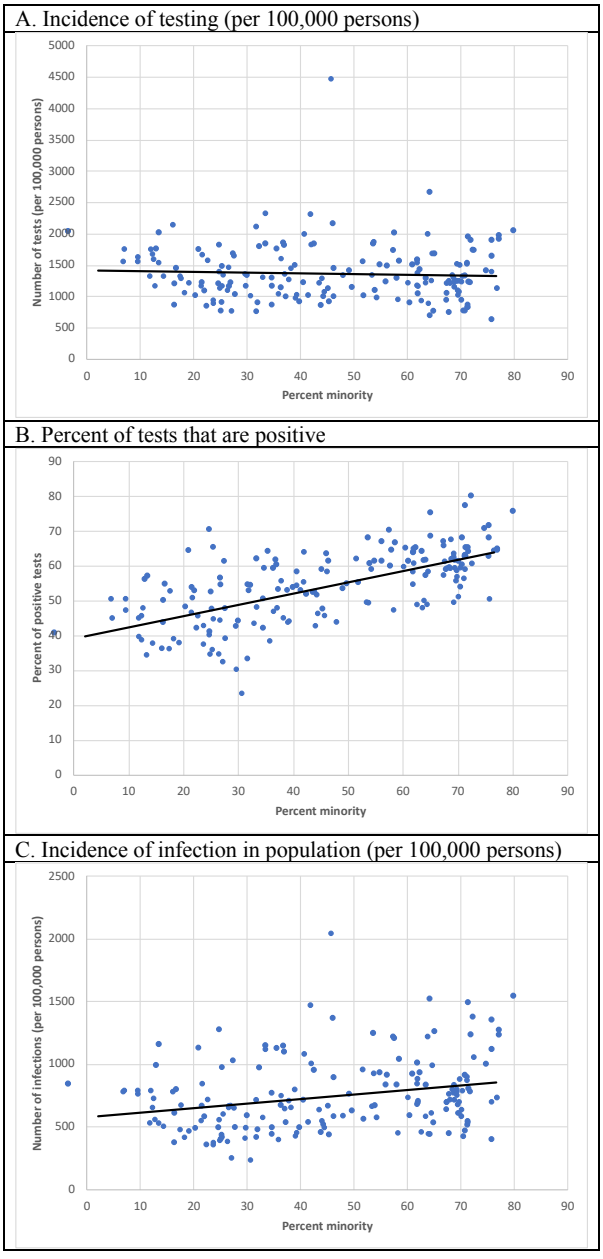
Figure 3. Household size and COVID-19



Notes: The household size variable gives the number of persons in the average household. See the text for details on the construction of the variables.

Covid Economics 3, 10 April 2020: 12-39

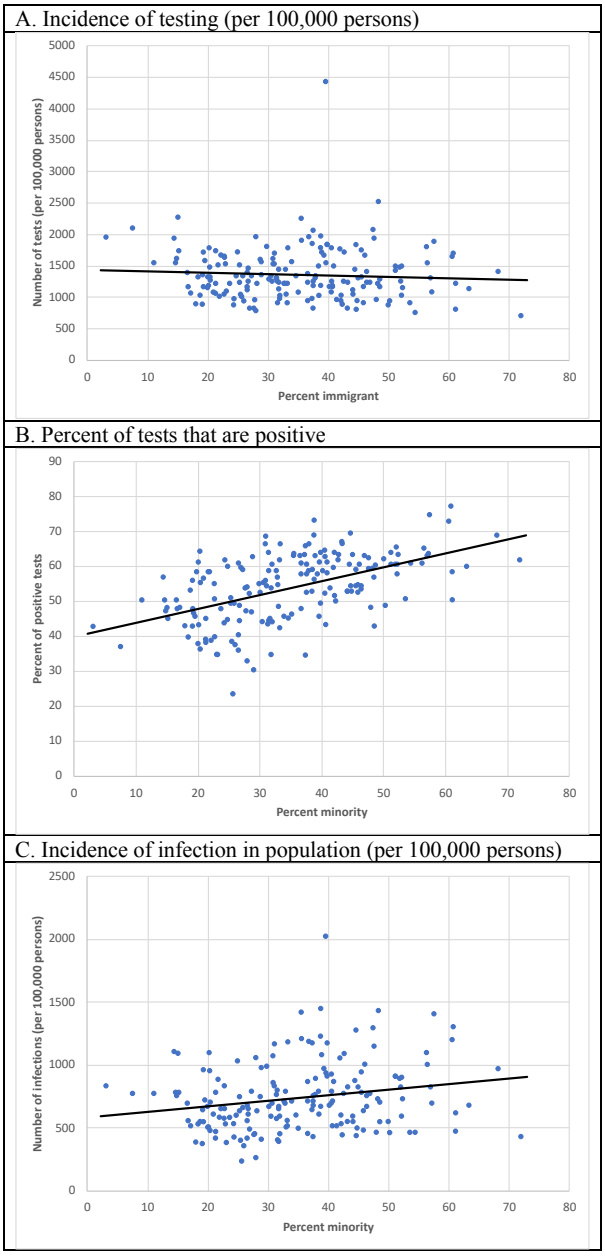
Figure 4. Minority neighborhoods and COVID-19



Notes: The percent minority variable gives the fraction of the zip code's population that is either Hispanic, non-Hispanic black, or non-Hispanic Asian. See the text for details on the construction of the variables.

Covid Economics 3, 10 April 2020: 12-39

Figure 5. Immigrant neighborhoods and COVID-19



Notes: The percent immigrant variable gives the fraction of the zip code's population that is foreign-born. See the text for details on the construction of the variables.

Covid Economics 3, 10 April 2020: 12-39

Table 1. Summary statistics

Variable:	Mean	10th percentile	Median	90th percentile
Number of tests	598.2	219.0	515.0	1013.0
Number of positive tests	337.9	101.0	274.0	624.0
Population of zip code (in 1000s)	46.1	14.0	41.0	85.9
Number of tests per 100,000 persons	1362.1	902.9	1296.2	1840.6
No. of positive tests per 100,000 persons	740.3	442.2	692.7	1107.4
Percent of tests that are positive	54.0	39.4	54.9	64.3
Average household income (in 1000s)	90.4	46.8	74.0	164.0
Size of household	2.6	1.7	2.7	3.2
Percent male	47.7	44.7	47.5	50.3
Percent age ≥ 60	17.5	11.4	16.8	25.3
Percent black	16.1	1.3	6.4	52.1
Percent Hispanic	18.8	5.6	13.1	44.3
Percent Asian	10.5	1.1	6.9	25.1
Percent immigrant	35.1	19.3	34.0	52.5

Notes: The sample has 177 observations. The data on the number of COVID-19 tests and the number of positive results were compiled by the New York City Department of Health and Mental Hygiene. The counts are cumulative to April 5, 2020. The data for the zip code's population, the average size of the household, the percent male, and the percent in the various ethnic/racial groups are drawn from the 2010 decennial census files. The variables measuring household income and the percent of the population that is immigrant are drawn from the pooled 2010-2014 American Community Surveys.

Covid Economics 3, 10 April 2020: 12-39

Table 2. Correlates of COVID-19 testing incidence and infections

	Number of tests relative to the population	Number of infections relative to number of tests conducted	Number of infections relative to the population
Log mean household income	0.182	-0.046	0.107
	(0.066)	(0.028)	(0.077)
Average household size	0.284	0.179	0.464
	(0.055)	(0.023)	(0.064)
Percent male	0.038	0.002	0.043
	(0.010)	(0.004)	(0.012)
Percent age ≥ 60	0.029	-0.000	0.031
	(0.005)	(0.002)	(0.006)
Percent black	0.002	0.002	0.004
	(0.001)	(0.001)	(0.002)
Percent Hispanic	0.002	0.000	0.003
	(0.002)	(0.001)	(0.002)
Percent Asian	-0.008	-0.001	-0.009
	(0.003)	(0.001)	(0.003)
Percent immigrant	-0.003	0.004	-0.000
	(0.002)	(0.001)	(0.002)
R-squared	0.331	0.724	0.448

Notes: Standard errors in parentheses. All regressions have 177 observations and are estimated using the command *glogit* in STATA. This procedure estimates grouped logit regression models using weighted least squares. The coefficients report the marginal effect of the independent variable in percentage terms, specifically giving the percent change in the number of "successful" cases (i.e., the numerator of the dependent variable) resulting from a one-unit change in the independent variable.

Covid Economics 3, 10 April 2020: 12-39

Table 3. Correlates of COVID-19 testing incidence and infections, including borough fixed effects

	Number of tests relative to the population	Number of infections relative to number of tests conducted	Number of infections relative to the population
Log mean household income	0.130	-0.040	0.061
	(0.069)	(0.030)	(0.085)
Average household size	0.254	0.154	0.390
	(0.058)	(0.025)	(0.070)
Percent male	0.043	-0.002	0.045
	(0.009)	(0.004)	(0.011)
Percent age ≥ 60	0.021	-0.000	0.023
	(0.005)	(0.002)	(0.006)
Percent black	0.001	0.002	0.003
	(0.001)	(0.001)	(0.002)
Percent Hispanic	-0.005	0.002	-0.003
	(0.002)	(0.001)	(0.003)
Percent Asian	-0.012	-0.001	-0.013
	(0.002)	(0.001)	(0.003)
Percent immigrant	-0.000	0.002	0.001
	(0.002)	(0.001)	(0.002)
Manhattan			---
Bronx	0.228	0.005	0.243
	(0.064)	(0.0302	(0.079)
Queens	0.060	0.121	0.188
	(0.062)	(0.027)	(0.076)
Brooklyn	-0.178	0.097	-0.079
	(0.066)	(0.030)	(0.082)
Staten Island	-0.035	-0.014	-0.024
	(0.088)	(0.041)	(0.111)
R-squared	0.517	0.778	0.567

Notes: Standard errors in parentheses. All regressions have 177 observations and are estimated using the command *glogit* in STATA. This procedure estimates grouped logit regression models using weighted least squares. The coefficients report the marginal effect of the independent variable in percentage terms, specifically giving the percent change in the number of "successful" cases (i.e., the numerator of the dependent variable) resulting from a one-unit change in the independent variable.

Covid Economics 3, 10 April 2020: 12-39

CEPR PRESS

Covid Economics **Issue 3, 10 April 2020**

Economic policy responses to a pandemic: Developing the Covid-19 economic stimulus index

Ceyhun Elgin,[1] Gokce Basbug[2] and Abdullah Yalaman[3,4]

Date submitted: 1 April 2020; **Date accepted:** 3 April 2020

In this paper, we conduct a comprehensive review of different economic policy measures adopted by 166 countries as a response to the COVID-19 pandemic and create a large database including fiscal, monetary, and exchange rate measures. Furthermore, using principle component analysis (PCA), we construct a COVID-19 Economic Stimulus Index (CESI) that combines all adopted policy measures. This index standardises economic responses taken by governments and allows us to study cross-country differences in policies. Finally, using simple cross-country OLS regressions we report that the median age of the population, the number of hospital beds per-capita, GDP per-capita, and the number of total cases are all significantly associated with the extent of countries' economic policy responses.

Covid Economics 3, 10 April 2020: 40-53

1 Lecturer in Discipline, Columbia University.
2 Assistant Professor, Sungkyunkwan University.
3 Professor at Eskişehir Osmangazi University
4 Our dataset will be regularly updated every week and the latest version is available at www.ceyhunelgin.com.

Covid Economics 3, 10 April 2020: 40-53

1 Introduction

The coronavirus (COVID-19) outbreak emerged in Wuhan, China in December of 2019 and still persists globally. The COVID-19 pandemic has spread to 199 countries and territories causing 777,798 cases and 37,272 deaths as of March 31, 2020. (Roser, Ritchie, and Ortiz-Ospina, 2020). In addition to human suffering and loss of lives, the outbreak generated a major global economic downturn. The world's largest economies (G7 and China) are among the ones that have been most affected by the pandemic (Baldwin and Weder di Mauro, 2020).

The COVID-19 pandemic has direct negative effects on the economy in several different ways. To name a few, infected workers who are isolated or hospitalized cannot join the workforce, which has several demand and supply-side implications. Furthermore, the psychological effect of the pandemic leads to withdrawal from economic activity by agents who prefers to adopt "wait and see" approach.

To decrease the transmission rate of COVID-19 and to reduce burden on healthcare systems, governments have adopted a wide range of stringent public health measures including school and factory closures, travel restrictions, and city lockdowns (Atkeson, 2020). These measures have been effective in slowing down the growth of new infections, as seen in the cases of Singapore and Hong Kong (Anderson et al. 2020). However, these non-pharmaceutical measures also distort economic activity by limiting human mobility and business operations (Eichenbaum, Rebelo, and Trabandt, 2020). Specifically, the COVID-19 pandemic and associated public health controls have disrupted supply chains and diminished activity in manufacturing and service sectors, which in turn led to increased layoffs. The stock markets crashed worldwide and the number of unemployment claims rose to unprecedented levels.

To mitigate the negative effects of public health controls on the economy and to sustain public welfare, governments adopted economic packages including fiscal, monetary, and financial policy measures (Gourinchas, 2020). These economic measures targeting households, firms, health systems and banks vary across countries in breadth and scope (Weder di Mauro, 2020).

Monetary polices adopted by countries usually consist of liquidity support to banks (IMF, 2020). Typical fiscal policies include transfers to households and businesses, extension of social safety benefits, and funds for the healthcare system. For example, South Korea introduced cash transfers for quarantined individuals, consumption coupons for low-income households, and wage and rent support for small businesses. Germany expanded access to short-term work subsidy, increased childcare benefits for low-income parents, and provided grants to small business owners and self-employed persons who were affected by the outbreak. United Kingdom provided funding for the National Health Service, introduced measures to support businesses including property tax holidays, direct grants for small firms, and

Covid Economics 3, 10 April 2020: 40-53

compensation for sick pay leave, and strengthened the social safety net to support vulnerable people.

In this paper, we conduct a comprehensive review of different economic policy measures adopted by 166 countries as a response to COVID-19 pandemic and create a large database including fiscal, monetary and exchange rate measures. Next, using the principle component analysis (PCA), we construct a COVID-19 Economic Stimulus Index (CESI) that combines all adopted fiscal, monetary, and exchange rate measures. This index standardizes the economic responses taken by governments, thus allows to study cross-country differences in policies. We further investigate to what extent countries' economic responses are shaped by several country characteristics, pandemic-related variables and public health measures (Correia, Luck, and Verner, 2020). Our findings show that the median age of the population, the number of hospital beds per-capita, GDP per-capita and the number of total cases are significantly associated with the extent of countries' economics policy responses.

The rest of the paper is organized as follows: The next section includes a description of our data sources as well as a short characterization of the PCA. The third section presents our results. Finally, in the last section we conclude.

2 Data and Methodology

2.1 Data Sources

To construct a comprehensive database of countries' policy measures, we used the information provided by the International Monetary Fund (IMF COVID-19 Policy Tracker, 2020). To improve data validity, we cross-checked this information using different sources. When the reported information was not up-to-date, we replaced it with most recent information gathered from various sources such as news channels and government websites. The current version of our dataset uses all available information by March 31st, 2020.

The economic policy package database we created includes six policy variables classified under three categories. These categories are, fiscal policy, monetary policy and balance of payment/ exchange rate policy. Fiscal policy package includes all the adopted fiscal measures and is coded as a percentage of GDP. The monetary policy category includes three different variables: 1) Interest rate cut[1] by the monetary policy authority (coded as a percentage of the ongoing rate on February 1st, 2020), 2) The size of the macro-financial package (coded as a percentage of GDP), and 3) Other monetary policy measures (coded as a dummy variable taking the value of 1 if there are such measures and 0, otherwise). Finally, the balance of payment (BoP) and exchange rate policy category includes two variables. The first one reports

[1]Whenever possible we used the rate cut in the policy rate. When there are multiple rate cuts, we calculated an arithmetic average of all rate cuts.

Covid Economics 3, 10 April 2020: 40-53

specific BoP measures coded as a percentage of GDP and the second one is a dummy variable taking the value of 1 if there are other reported measures and 0, otherwise.

In addition to economic policy measures, we gathered data on up-to-date public health measures and pandemic-related variables using different sources on the Internet. This data include countries' 2019 median age and COVID-19 infection rate (defined by the ratio of total COVID-19 cases to population). Moreover, data on hospital beds (per 1,000 people) and current health expenditures (as a percentage of GDP) are obtained from the World Bank. Finally, we use the recently reported government response stringency index of Hale and Webster (2020) as an additional explanatory variable in our analyses.

2.2 Developing the COVID-19 Economic Stimulus Index (CESI)

There are several different methodologies used for index development and each has different advantages and disadvantages. In this paper, we use the PCA[2] that is one of the most frequently used method for index development. Specifically, the PCA helps reduce the number of variables in an analysis by describing a series of uncorrelated linear combinations of the variables that contain most of the variance. Moreover, the eigenvectors associated with the PCA give significant information about the different variables used to create the index. We report the principle components as well as the eigenvalues and the proportion of the variance explained in Table A.4 in the appendix[3]. Depending on the PCA, we simply use all six policy variables in our dataset to create a composite index as a predicted variable with the fiscal policy stimulus and interest rate cut having the largest weights in the overall index.

3 Results

Table 1 reports descriptive summary statistics of all economic policy variables as well as the COVID-19 Economic Stimulus Index (CESI)constructed using the PCA. The whole data series are reported in Tables A.1 to A.3 in the Appendix. In Figure 1, we illustrate the histogram of the CESI and the associated fitted kernel and normal distributions. Accordingly, the index has a right-skewed distribution, which is also apparent in Figures A.1 and A.2. This is largely because there are several countries with significant interest rate cuts and fiscal policy packages increasing the mean; however, at the same time a large number of countries also have not yet

[2]PCA originated from the works of Pearson (1901) and Hotelling (1933). For many different applications, see Rencherand Christensen, 2012; Li et. al., 2019; Kumar and Anbanandam, 2019; Deutsch and Beinker, 2019; Bala et. al., 2019; Obeng-Ahenkora and Danso, 2020.

[3]The index that comes out of the PCA analysis is satisfactory in explaining the overall variance with more than two component and also satisfies other desirable criteria needed in a PCA analysis. Nevertheless, later, we are planning to use several other methods such as the structural equation modelling or factor analysis to supplement the PCA.

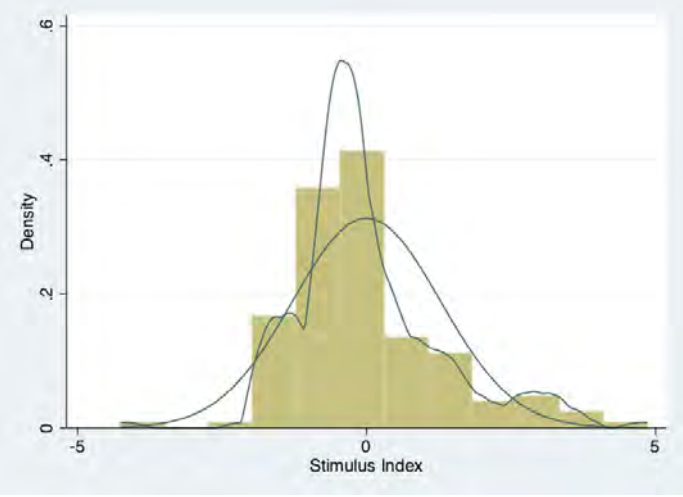
Covid Economics 3, 10 April 2020: 40-53

Table 1: Summary Statistics of the Dataset

	Mean	Median	Std. Dev.	Min	Max
COVID-19 Economic Stimulus Index(CESI)	0.00	-0.31	1.28	-4.25	4.85
Fiscal Policy Stimulus (%)	2.09	0.48	3.60	-7.20	17.80
Interest Rate Cut (%)	11.63	0.00	21.47	-29.73	100.00
Macro-Financial Package (% of GDP)	1.87	0.00	4.02	0.00	26.00
Other Monetary Measures(0-1 dummy)	0.85	1.00	0.36	0.00	1.00
BoP Measures (% GDP)	0.10	0.00	0.58	0.00	6.00
Other BoP Measure (0-1 dummy)	0.19	0.00	0.40	0.00	1.00

implemented any stimulus packages. This is also apparent in the level of standard deviations, which exhibits a sizable amount of variation across countries.

Figure 1: The COVID-19 Economic Stimulus Index (CESI): Histogram and Cumulative Distribution



Next, we conduct some simple cross-country regressions with our stimulus index as the dependent variable and country characteristics, several public health measures as well as the real GDP per-capita as independent variables. We report the results of six regression analyses in Table 2.

In the first regression, we regress the CESI score on the median age of population. The results show that the median age has a significant positive relationship with the economic responses, indicating that countries with older populations introduced larger stimulus pack-

Covid Economics 3, 10 April 2020: 40-53

Table 2: Cross-Country OLS Regressions

Dep. Var.	CESI	CESI	CESI	CESI	CESI	CESI	CESI
Median Age	0.07*	0.10*	0.09*	0.09*	0.06*	0.05*	0.05*
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Hospital Beds (per-capita)		-0.15*	-0.13*	-0.12**	-0.11*	-0.11*	-0.11*
		(0.04)	(0.04)	(0.04)	(0.06)	(0.04)	(0.04)
Infection Rate (%)			546.25*	224.83	-69.56	-149.34	-151.30
			(211.49)	(237.44)	(196.50)	(220.96)	(225.97)
Stringency Index				0.004			
				(0.006)			
GDP per-capita (000 USD)					0.03*	0.03*	0.03*
					(0.001)	(0.001)	(0.001)
Total Cases						0.007**	0.008**
						(0.003)	(0.004)
Health Expenditures (% GDP)							-0.03
							(0.04)
R-squared	0.27	0.31	0.34	0.30	0.43	0.43	0.43
Observations	146	146	143	69	140	140	139
F-Test	0.00	0.00	0.00	0.00	0.00	0.00	0.00

All regressions include a constant. Robust standard errors are reported in parentheses. *, **, *** denote 1, 5 and 10% confidence levels, respectively.

ages. It is important to note that the median age is significantly and positively associated with the level of economic response in a consistent manner throughout all regression models.

In the second regression, we included the number of hospital beds per-capita into the model, which is significantly and negatively associated with the size of the economic stimulus. After controlling for other variables in the following regressions, the number of beds per capita consistently remains negatively associated with the size of the economic stimulus. This result implies that countries where the number of beds per capita is lower, more stringent economic stimulus is adopted.

In the third regression model, we add the infection rate which refers to the ratio of total positive cases to population. After controlling for median age and hospital beds, the infection rate is positively associated with the economic response, indicating that countries with higher infection rates adopted stronger economic measures.

Then, we regress the Stringency Index on the CESI score. The Stringency Index consists of public health controls adopted by governments in response to the pandemic. The analysis show that after controlling for median age, the number of beds, and the infection rate, the

Stringency index does not predict the economic stimulus package. It is important to note that due to data availability, we lose half of the sample when we introduce the Stringency Index.

Finally, in the remaining models, we introduce GDP per capita, the number of total cases, and health expenditure. The first two of these variables are significantly associated with the CESI score, indicating that countries with higher GDP per capita and a higher number of cases, announced larger economic stimulus packages.

4 Conclusion

In this paper, we first introduced a large database where we quantified the economic policies adopted by national governments throughout the global COVID-19 pandemic. Second, using PCA methodology, we developed the COVID-19 Economic Stimulus Index (CESI) which allowed us to aggregate and standardize varying economic responses across countries. Finally, we presented some preliminary results on the predictors of governments' economic responses. Our findings, without establishing any causality, show some significant correlations of population characteristics, public health-related, and economic variables (e.g., GDP per capita, health expenditures) with economic stimulus packages announced by governments. Specifically, we find that in countries where the median age is higher (which is highly relevant in the case of the COVID-19, as it disproportionally affects older patients), the number of hospital beds per-capita is lower and GDP per-capita is higher, the stimulus is more pronounced.

In our analyses, the Stringency Index which measures countries' public health controls such as school closures and travel restrictions did not predict the level of economic responses. Although we lose a significant number of cases when we introduce the Stringency Index, this non-significant finding implies that governments' economic responses are more motivated by reacting to the pandemic (i.e., infection rate), rather than mitigating the negative economic implications of public health controls.

Although our study has some limitations (e.g. endogeneity, omitted variable bias), we believe that it contributes to our understanding of the economics of the COVID-19 pandemic mainly in two ways. First, we believe that the economic stimulus package database which will be updated on a daily basis and the index will be helpful to other researchers while studying the outcomes of economic responses. Secondly, our study sheds some light on the predictors of the economic responses adopted by governments.

Covid Economics 3, 10 April 2020: 40-53

Covid Economics 3, 10 April 2020: 40-53

References

1. Anderson, R. M., Heesterbeek, H., Klinkenberg, D., Hollingsworth, T. D. (2020). How will country-based mitigation measures influence the course of the COVID-19 epidemic?. The Lancet 395 (10228), 931-934.
2. Atkeson, A. (2020). What Will Be the Economic Impact of COVID-19 in the US? Rough Estimates of Disease Scenarios. No. w26867. NBER
3. Baldwin, R., Weder di Mauro, B. (2020), Introduction. In "Economics in the time of COVID-19", Baldwin Weder di Mauro (eds). CEPR Press, London, UK.
4. Correia, S., Luck, S., Verner, E. (2020). Pandemics Depress the Economy, Public Health Interventions Do Not: Evidence from the 1918 Flu. Available at SSRN: http://dx.doi.org/10.2139/ssrn.3561560
5. Deutsch, H. P., Beinker, M. W. (2019). Principal Component Analysis. In Derivatives and Internal Models (pp. 793-804). Palgrave Macmillan, Cham.
6. Eichenbaum, M. S., Rebelo, S., Trabandt, M. (2020). The Macroeconomics of Epidemics. No. w26882. NBER.
7. Gourinchas, P-O. (2020). Flattening the Pandemic and Recession Curves. Mitigating the COVID Economic Crisis: Act Fast and Do Whatever. 31. CEPR Press, London, UK.
8. Hale, T., Webster, S. (2020). Oxford COVID-19 Government Response Tracker. Available at: https://www.bsg.ox.ac.uk/research/research-projects/oxford-covid-19-government-response-tracker
9. Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. Journal of Educational Psychology 24: 417-441.
10. International Monetary Fund 2002 Retrieved from https://www.imf.org/en/Topics/imf-and-covid19/Policy-Responses-to-COVID-19#U on May 31, 2019.
11. Kumar, A., Anbanandam, R. (2019). Development of social sustainability index for freight transportation system. Journal of cleaner production, 210, 77-92.
12. Li, X., Jiang, H., Xiong, X., Shao, H. (2019). Rolling bearing health prognosis using a modified health index based hierarchical gated recurrent unit network. Mechanism and Machine Theory, 133, 229-249.
13. Obeng-Ahenkora, N. K., Danso, H. (2020). Principal component analysis of factors influencing pricing decisions of building materials in Ghana. International Journal of Construction Management, 20(2), 122-129.
14. Pearson, K. 1901. On lines and planes of closest fit to systems of points in space. Philosophical Magazine, Series 6, 2: 559-572.

15. Roser, M. Ritchie, H., Ortiz-Ospina, E. (2020) - "Coronavirus Disease (COVID-19) ? Statistics and Research". Published online at OurWorldInData.org. Retrieved from https://ourworldindata.org/coronavirus on May 31, 2019.

16. Rencher, A. C., Christensen, W. F. (2012). Methods of Multivariate Analysis. 3rd ed. Hoboken, NJ: Wiley.

17. Weder di Mauro, B. (2020) Macroeconomics of the flu. In "Economics in the time of COVID-19", Baldwin Weder di Mauro (eds). CEPR Press, London, UK.

Covid Economics 3, 10 April 2020: 40-53

Covid Economics 3, 10 April 2020: 40-53

A Appendix

In three tables below, we report the comprehensive database we constructed. It includes six economic policy variables as well as the Economic Stimulus Index. Our dataset will be regularly updated every week and the latest version is available at www.ceyhunelgin.com.

Figure A.1: Fiscal Stimulus Packages: Histogram and Cumulative Distribution

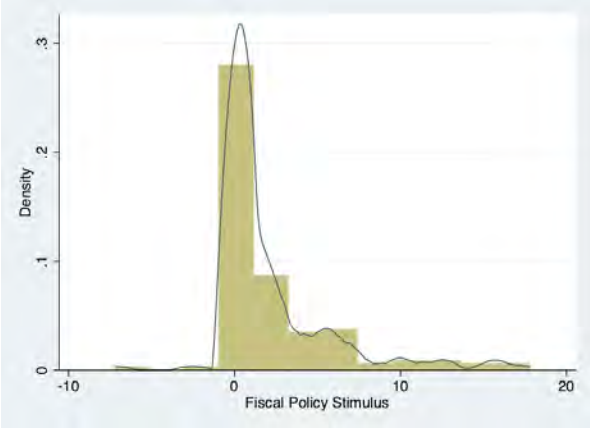
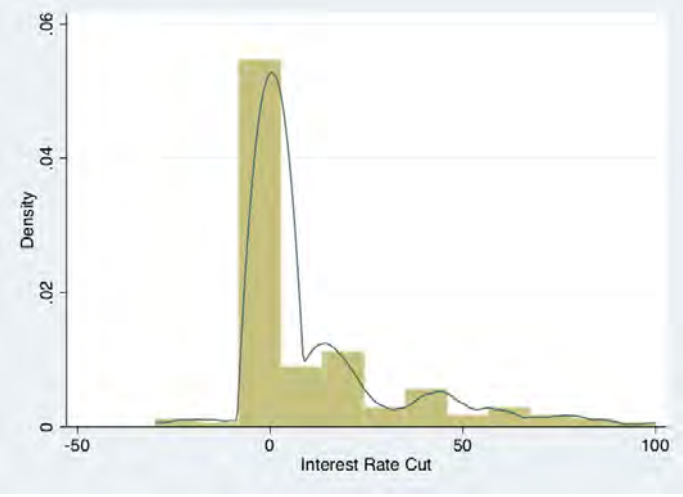


Figure A.2: Interest Rate Cuts: Histogram and Cumulative Distribution



Covid Economics 3, 10 April 2020: 40-53

Table A.1: Economic Policy Packages and the CESI

Country	Fiscal (% GDP)	Rate Cut (%)	Macro-Financial (% GDP)	Other Monetary	BoP (% GDP)	Other BoP	Stimulus Index
Afghan	0.13	0.00	0.00	0.00	0.00	0.00	-1.60
Albania	1.30	50.00	0.00	1.00	0.00	0.00	0.58
Algeria	-7.20	13.57	0.00	0.00	6.00	0.00	-4.25
Angola	0.00	0.00	0.00	0.00	0.00	0.00	-1.62
Argentina	1.00	5.00	0.00	1.00	0.22	0.00	-0.35
Armenia	0.00	4.55	0.00	0.00	0.00	0.00	-1.53
Australia	9.70	0.67	4.71	1.00	0.00	0.00	1.69
Austria	17.8	0.00	7.31	1.00	0.00	0.00	3.34
Azerbaijan	1.01	11.36	0.00	1.00	0.00	1.00	-0.57
Bahamas	0.20	0.00	0.00	1.00	0.00	0.00	-0.50
Bahrain	5.30	52.27	26.00	1.00	0.00	0.00	4.85
Bangladesh	0.01	7.27	0.00	1.00	0.00	0.00	-0.40
Barbados	1.40	0.00	0.00	1.00	0.00	0.00	-0.31
Belarus	0.00	0.00	0.00	1.00	0.00	1.00	-0.94
Belgium	12.3	0.00	7.31	1.00	0.00	0.00	2.46
Belize	1.00	0.00	0.00	1.00	0.00	0.00	-0.37
Benin	0.10	0.00	0.03	1.00	0.00	0.00	-0.51
Bhutan	0.00	0.00	0.00	0.00	0.00	0.00	-1.62
Bolivia	0.46	0.00	1.18	1.00	0.00	0.00	-0.30
Bosnia	3.25	0.00	0.00	1.00	0.00	0.00	-0.01
Botswana	0.28	0.00	0.00	0.00	0.00	0.00	-1.57
Brazil	3.50	28.46	3.17	1.00	1.69	1.00	0.09
Brunei	0.21	0.00	0.00	1.00	0.00	0.00	-0.50
Bulgaria	2.00	0.00	8.60	0.00	0.00	0.00	-0.11
Burkina Faso	0.00	0.00	0.00	1.00	0.00	0.00	-0.53
Burundi	0.50	0.00	0.00	0.00	0.00	0.00	-1.54
Cabo Verde	0.00	0.00	0.00	0.00	0.00	0.00	-1.62
Cambodia	1.34	0.00	0.00	1.00	0.00	0.00	-0.32
Cameroon	0.10	0.00	0.00	1.00	0.00	0.00	-0.52
Canada	6.00	57.14	2.60	1.00	0.00	0.00	1.82
CAR	1.90	0.00	0.00	1.00	0.00	0.00	-0.23
Chad	0.23	0.00	0.00	1.00	0.00	0.00	-0.50
Chile	4.70	42.86	1.36	1.00	0.00	1.00	0.78
China	1.20	0.00	14.14	1.00	0.00	1.00	1.20
Hong Kong	5.30	57.00	0.00	1.00	0.00	0.00	1.35
Colombia	0.40	0.00	1.17	1.00	0.43	1.00	-0.84
Congo, DR	0.30	16.67	0.00	1.00	0.00	0.00	-0.18
Congo, R	0.32	0.00	0.00	1.00	0.00	0.00	-0.48
Costa Rica	0.00	44.44	0.00	1.00	0.00	1.00	-0.14
Cote Ivory	0.30	0.00	0.00	1.00	0.00	0.00	-0.49
Croatia	0.30	75.00	1.25	1.00	2.94	0.00	0.20
Cyprus	3.30	0.00	7.77	1.00	0.00	0.00	1.07
Czech	2.00	22.22	0.00	1.00	0.00	0.00	0.19
Denmark	5.30	-20.00	0.00	1.00	0.00	0.00	-0.04
Djibouti	0.00	0.00	0.00	1.00	0.00	0.00	-0.53
Ecuador	1.00	0.00	0.00	1.00	0.00	0.00	-0.37
Egypt	2.00	23.53	0.00	1.00	0.00	0.00	0.21
El Salvador	1.34	0.00	0.00	1.00	0.00	0.00	-0.32
Equit. Guinea	0.07	0.00	0.00	1.00	0.00	0.00	-0.52
Eritrea	0.00	0.00	0.00	0.00	0.00	0.00	-1.62
Estonia	7.00	0.00	7.71	1.00	0.00	0.00	1.66
Eswatini	0.14	17.88	0.00	0.00	0.00	0.00	-1.27

Fiscal stands for the fiscal policy package as a percent of GDP, Rate cut is the interest rate cut as a percent of the pre-crisis level, Macro-Financial is the monetary stimulus package as a percent of GDP, other monetary is a dummy variable taking the value of 1 if there are other accompanying monetary measures, BoP is the monetary intervention to control the balance of payments and the exchange rate as a percent of GDP and finally, Other BoP is a dummy variable taking the value of 1 if there are other accompanying measures towards stabilizing BoP and exchange rate.

Covid Economics 3, 10 April 2020: 40-53

Table A.2: Economic Policy Packages and the CESI

Country	Fiscal (% GDP)	Rate Cut (%)	Macro-Financial (% GDP)	Other Monetary	BoP (% GDP)	Other BoP	Stimulus Index
Ethiopia	0.15	0.00	0.00	0.00	0.00	0.00	-1.59
Fiji	7.24	50.00	0.00	0.00	0.00	0.00	0.45
Finland	1.00	0.00	7.31	1.00	0.00	0.00	0.64
France	15.30	0.00	7.31	1.00	0.00	0.00	2.94
Gabon	0.25	0.00	0.00	1.00	0.00	0.00	-0.49
Gambia	0.60	-8.00	0.00	1.00	0.00	0.00	-0.58
Georgia	2.00	0.000	0.00	1.00	0.63	0.00	-0.39
Germany	4.80	0.00	12.49	1.00	0.00	0.00	1.96
Ghana	0.15	9.38	0.00	1.00	0.00	0.00	-0.34
Greece	5.00	0.00	7.31	1.00	0.00	0.00	1.28
Guatemala	1.57	18.18	0.00	1.00	0.00	0.00	0.05
Guinea	0.00	0.00	0.00	0.00	0.00	0.00	-1.62
Guinea Bissau	0.00	0.00	0.00	1.00	0.00	0.00	-0.53
Guyana	0.00	0.00	0.00	1.00	0.00	0.00	-0.53
Haiti	0.00	0.00	0.00	1.00	0.00	0.00	-0.53
Honduras	2.20	14.29	2.80	1.00	0.00	0.00	0.46
Hungary	0.39	0.00	1.00	1.00	0.00	0.00	-0.33
Iceland	7.80	43.18	1.00	1.00	0.30	0.00	1.55
India	0.20	0.00	1.10	1.00	0.00	1.00	-0.76
Indonesia	0.20	10.00	0.00	1.00	0.00	1.00	-0.73
Iran	0.56	0.00	0.06	1.00	0.33	1.00	-0.94
Iraq	0.01	0.00	1.00	0.00	0.00	0.00	-1.48
Ireland	2.58	0.00	7.31	1.00	0.00	0.00	0.89
Israel	1.10	0.00	3.5	1.00	0.00	1.00	-0.28
Italy	1.70	0.00	7.31	1.00	0.00	0.00	0.75
Jamaica	1.40	0.00	0.00	1.00	0.00	0.00	-0.31
Japan	4.90	0.00	0.30	1.00	0.00	0.00	0.40
Jordan	0.00	37.50	1.83	1.00	0.00	0.00	0.40
Kazakhstan	3.45	-29.73	0.00	1.00	0.00	1.00	-0.92
Kenya	0.00	15.58	0.00	1.00	0.00	0.00	-0.25
Korea	0.80	40.00	0.34	1.00	0.00	1.00	-0.04
Kosovo	0.08	0.00	0.00	1.00	0.00	0.00	-0.52
Kuwait	1.40	45.45	0.02	1.00	0.00	0.00	0.52
Kyrgyz Rep.	0.10	-17.65	0.00	1.00	0.00	1.00	-1.24
Laos	0.01	0.00	0.00	1.00	0.00	1.00	-0.94
Latvia	3.30	0.00	7.31	1.00	0.00	0.00	1.01
Lebanon	0.00	0.00	0.00	1.00	0.00	0.00	-0.53
Lesotho	0.00	16.00	0.00	1.00	0.00	0.00	-0.25
Liberia	0.00	0.00	0.00	0.00	0.00	0.00	-1.62
Lithuania	5.30	0.00	7.31	1.00	0.00	0.00	1.33
Luxemburg	15.6	0.00	7.31	1.00	0.00	0.00	2.99
Madagascar	0.03	0.00	0.30	1.00	0.00	1.00	-0.89
Malawi	0.25	0.00	0.00	0.00	0.00	0.00	-1.58
Malaysia	16.22	9.09	0.20	1.00	0.00	0.00	2.27
Maldives	2.80	0.00	0.00	1.00	0.00	1.00	-0.49
Mali	0.06	0.00	0.00	1.00	0.00	0.00	-0.52
Malta	12.3	0.00	13.31	1.00	0.00	0.00	3.28
Mauritania	0.13	25.43	0.00	1.00	0.00	0.00	-0.05
Mauritius	1.38	14.93	2.00	1.00	0.00	1.00	-0.17
Mexico	0.70	7.14	0.60	1.00	0.00	1.00	-0.62
Moldova	0.00	40.91	0.00	1.00	0.00	0.00	0.21
Mongolia	0.03	9.09	0.80	1.00	0.00	0.00	-0.25
Montenegro	0.02	0.00	0.00	1.00	0.00	1.00	-0.94
Morocco	0.84	12.5	0.00	1.00	0.00	1.00	-0.58
Mozambique	0.10	0.00	0.00	1.00	0.00	0.00	-0.52

Fiscal stands for the fiscal policy package as a percent of GDP, Rate cut is the interest rate cut as a percent of the pre-crisis level, Macro-Financial is the monetary stimulus package as a percent of GDP, other monetary is a dummy variable taking the value of 1 if there are other accompanying monetary measures, BoP is the monetary intervention to control the balance of payments and the exchange rate as a percent of GDP and finally, Other BoP is a dummy variable taking the value of 1 if there are other accompanying measures towards stabilizing BoP and exchange rate.

Covid Economics 3, 10 April 2020: 40-53

Table A.3: Economic Policy Packages and the CESI

Country	Fiscal (% GDP)	Rate Cut (%)	Macro-Financial (% GDP)	Other Monetary	BoP (% GDP)	Other BoP	Stimulus Index
Myanmar	0.10	18.75	0.00	1.00	0.00	0.00	-0.18
Namibia	0.00	16.00	0.00	0.00	0.00	0.00	-1.33
Nepal	0.00	0.00	0.00	1.00	0.00	0.00	-0.53
Netherlands	2.30	0.00	7.31	1.00	0.00	0.00	0.84
New Zealand	5.40	75.00	8.86	1.00	0.00	0.00	2.91
Nicaragua	0.00	7.50	0.00	1.00	0.00	0.00	-0.40
Niger	0.02	0.00	0.00	1.00	0.00	0.00	-0.53
Nigeria	0.01	0.00	2.40	1.00	0.00	1.00	-0.61
N. Macedonia	0.20	12.50	0.00	1.00	0.00	1.00	-0.68
Norway	2.20	83.33	0.00	1.00	0.00	0.00	1.33
Oman	-2.50	60.00	25.09	1.00	0.00	0.00	3.61
Pakistan	2.54	16.98	0.00	1.00	0.00	0.00	0.18
Papua N. Guiena	0.05	0.00	0.00	0.00	0.00	0.00	-1.61
Paraguay	6.50	18.75	0.00	0.00	0.00	0.00	-0.23
Peru	0.78	44.44	0.00	1.00	0.90	0.00	0.14
Philippines	0.15	18.75	1.60	1.00	0.00	0.00	0.05
Poland	6.50	33.33	0.00	1.00	0.00	0.00	1.12
Portugal	4.70	0.00	7.31	1.00	0.00	0.00	1.23
Qatar	13.00	43.73	1.43	1.00	0.00	0.00	2.55
Panama	3.25	0.00	2.00	1.00	0.00	0.00	0.26
Romania	3.00	20.00	0.00	1.00	0.00	0.00	0.31
Russia	0.30	0.00	0.39	1.00	0.00	0.00	-0.43
Rwanda	1.50	0.00	0.00	1.00	0.00	0.00	-0.29
San Marino	0.00	0.00	0.00	1.00	0.00	0.00	-0.53
Saudi Arabia	0.80	63.49	1.90	1.00	0.00	0.00	1.00
Senegal	7.00	0.00	0.00	1.00	0.00	0.00	0.59
Serbia	1.00	22.22	0.00	1.00	0.00	0.00	0.04
Seychelles	0.00	20.00	2.27	1.00	0.00	0.00	0.14
Sierra Leone	0.00	9.09	0.00	1.00	0.00	1.00	-0.78
Singapore	10.50	0.00	0.00	1.00	0.00	0.00	1.16
Slovak Rep.	0.30	0.00	7.31	1.00	0.00	0.00	0.52
Slovenia	6.60	0.00	7.31	1.00	0.00	0.00	1.54
S. Africa	0.20	16.00	0.00	1.00	0.00	0.00	-0.21
Spain	1.00	0.00	7.31	1.00	0.00	0.00	0.64
SriLanka	0.11	3.59	0.00	1.00	0.00	1.00	-0.86
Sudan	0.00	0.00	0.00	0.00	0.00	0.00	-1.62
Suriname	0.01	0.00	0.00	0.00	0.00	0.00	-1.62
Sweden	9.20	73.33	9.45	1.00	0.00	0.00	3.58
Switzerland	6.00	0.00	0.51	1.00	2.9 0	0.00	-0.33
Tajikistan	0.00	-4.08	0.00	1.00	0.00	1.00	-1.01
Tanzania	0.00	0.00	0.00	0.00	0.00	0.00	-1.62
Thailand	3.00	40.00	0.58	1.00	0.00	1.00	0.35
Togo	2.00	0.00	0.00	1.00	0.00	0.00	-0.21
Tonga	0.00	0.00	0.00	1.00	0.00	0.00	-0.53
Trinidad Tobago	3.25	30.00	0.00	1.00	0.00	1.00	0.13
Tunisia	2.00	12.90	0.99	1.00	0.00	0.00	0.16
Turkey	2.00	9.30	0.00	1.00	0.00	1.00	-0.45
Turkmenistan	0.00	0.00	0.00	0.00	0.00	1.00	-2.02
Uganda	0.02	0.00	0.00	1.00	0.00	0.00	-0.53
Ukraine	0.00	0.00	0.00	1.00	0.00	0.00	-0.53
UAE	1.80	62.50	6.70	1.00	0.00	0.00	1.81
UK	2.50	86.67	9.09	1.00	0.00	0.00	2.69
United States	10.50	100	0.00	1.00	0.00	0.00	2.97
Uruguay	0.00	0.00	1.00	1.00	0.00	1.00	-0.80
Uzbekistan	1.50	0.00	0.00	1.00	0.00	0.00	-0.30
Vietnam	0.33	14.48	3.30	1.00	0.00	0.00	0.24
Yemen	0.00	0.00	0.00	0.00	0.00	0.00	-1.62
Zambia	0.02	0.00	0.00	1.00	0.00	0.00	-0.53
Zimbabwe	0.21	28.57	0.03	0.00	0.00	1.00	-0.39

Fiscal stands for the fiscal policy package as a percent of GDP, Rate cut is the interest rate cut as a percent of the pre-crisis level, Macro-Financial is the monetary stimulus package as a percent of GDP, other monetary is a dummy variable taking the value of 1 if there are other accompanying monetary measures, BoP is the monetary intervention to control the balance of payments and the exchange rate as a percent of GDP and finally, Other BoP is a dummy variable taking the value of 1 if there are other accompanying measures towards stabilizing BoP and exchange rate.

Covid Economics 3, 10 April 2020: 40-53

Table A.4: Principle Component Analysis

Component	Eigen Value	Difference	Proportion	Cumulative		
Component 1	1.63	0.44	0.27	0.27		
Component 2	1.19	0.10	0.20	0.47		
Component 3	1.08	0.33	0.18	0.65		
Component 4	0.75	0.04	0.13	0.78		
Component 5	0.71	0.07	0.12	0.89		
Component 6	0.64		0.11	1.00		
Variable	Comp 1	Comp 2	Comp 3	Comp 4	Comp 5	Comp 6
Fiscal	0.58	-0.04	-0.22	0.15	-0.22	0.73
Rate Cut	0.39	-0.21	0.54	-0.67	0.2	0.05
Macro-Financial	0.55	-0.12	-0.01	0.49	0.53	-0.39
Other Monetary	0.39	0.54	0.25	0.03	-0.59	-0.39
BoP	-0.17	-0.43	0.67	0.50	-0.28	0.12
Other BoP	-0.16	0.68	0.38	0.17	0.44	0.38

CEPR PRESS

Covid Economics **Issue 3, 10 April 2020**

Flight to safety: 2020 Democratic primary election results and Covid-19

James Bisbee[1] and Dan Honig[2]

Date submitted: 2 April 2020; **Date accepted:** 3 April 2020

What is the impact of anxiety on vote choice? Building on a well-documented phenomenon in finance, we posit that voters will exhibit a "flight to safety" by turning toward establishment candidates. We test this theory in the context of the Democratic primary election of 2020 by examining changes in the vote shares of Bernie Sanders, a candidate promising disruptive change. We use the outbreak of the novel coronavirus across both space and time to identify a causal effect of the outbreak on voting. By comparing counties with and without reported cases in their local media market, before and after the outbreak of the virus, we show that COVID-19 resulted in diminished support for Sanders as compared to his support in the 2016 election, and interpret this to be the result of COVID-induced anxiety altering vote choice. We test alternative mechanisms, such as differential changes in turnout by age groups more and less supportive of Sanders, selection effects in which areas less supportive of Sanders were more exposed, and the coincident timing of the outbreak with the Democratic party rallying around Biden. We find little support for these alternative pathways, bolstering our claim that the results are consistent with a political flight to safety. Our findings suggest an as-yet underappreciated preference for "safe" candidates in times of social anxiety.

1 Postdoctoral Research Associate, Niehaus Center for Globalization and Governance, Princeton University.
2 Assistant Professor of International Development, Johns Hopkins School of Advanced International Studies.

Covid Economics 3, 10 April 2020: 54-84

Covid Economics 3, 10 April 2020: 54-84

1 Introduction

As COVID-19 began to dominate the headlines of US Newspapers in March 2020, it displaced coverage of the Democratic primary election. In that campaign, one of the two leading candidates had been running on a platform centered on universal health care. One might reasonably imagine that a growing pandemic would lead to a surge in support for that candidate. It did not.

Bernie Sanders did not merely fail to surge as the novel coronavirus came to dominate the headlines in 2020 – if anything, his campaign faded. We explore whether COVID-19 had anything to do with Sanders' decline in support.

The novel coronavirus appeared in the middle of the Democratic primary season. On Super Tuesday, COVID-19 cases were in the news, but public concern in the US was modest. There were a little over 100 total cases in the US, and the bulk of TV and print news content was focused on the elections. And why shouldn't it be? President Trump, after all, had said just a few days prior that COVID-19 was soon "going to disappear" (Leonhardt, 2020).

President Trump's projection was, unfortunately, incorrect. Just 14 days later, when voters in 3 states cast their ballots in March 17th Democratic primaries, President Trump had already declared a national state of emergency; most of the nation's schools were closed; the stock market had lost over 20% of its value; and many people had begun staying home to practice "social distancing".

In this paper, we ask whether the novel coronavirus hurt the electoral prospects of Bernie Sanders, the more anti-establishment candidate. We find that it does – with COVID-19's appearance associated with as much as a 7 percentage point decline for Sanders in areas where COVID-19 appears prior to voting as compared to similar counties where COVID-19 only appears following the casting of votes.

Our interrogation is motivated by a well-documented financial phenomenon that has, as yet, not been applied to voting behavior – namely, a "flight to safety". We pre-specified this hypothesis and our empirical specifications, registering a preanalysis plan prior to analyzing any data, and prior to the primaries of March 17th.[1] We explore whether anxiety generated by the unexpected outbreak of COVID-19 impacted voting decisions. Empirically, we compare how counties voted before and after the virus was widespread, in areas where the virus was relatively prevalent and where it was not. We show that where the virus emerged prior to the primary election, vote shares for Sanders fell.

Examining a primary election between two challengers allows us to avoid confounding our theorized flight to safety with a purely retrospective evaluation of the incumbent. Precisely because neither Biden nor Sanders held power during the outbreak, we are able to isolate the mainstream versus anti-establishment distinction between the Biden and Sanders campaigns, allowing for a tighter test of the motivating theory. Of course, there are other

[1]This registration is at https://egap.org/content/fear-and-flight-safety-covid-19-and-2020-democratic-primary, however the design itself is temporarily gated. We are happy to provide the plan itself on request.

Covid Economics 3, 10 April 2020: 54-84

dimensions along which the Biden and Sanders campaigns differ besides the distinction between their mainstream and anti-establishment qualities. We consider and reject plausible alternative mechanisms, providing suggestive evidence that our interpretation of the findings – a "flight to safety" – is most consistent with the empirical evidence.

That this result obtains despite what should be a policy platform whose appeal increases with the pandemic leads us to conclude that the power of the "flight to safety" in the context of voting is an important, but as yet unaccounted for, phenomenon in the voting literature.

2 Theory & Context

Scholars of financial markets and market analysts often discuss markets' "flight to safety" (e.g. Adrian, Crump and Vogt 2019; Inghelbrecht et al. 2013). As market outcomes become more uncertain, risk appetite falls. Anxiety drives market players to reduce their level of risk. In the context of investing, this behavior typically involves shifting assets towards more liquid and Government-insured assets, which are perceived as safer.

While there is a literature on voters' response to terrorism, (e.g. Getmansky and Zeitzoff 2014; Montalvo 2011) there is little research on the effects of anxiety more broadly and whether anxiety shifts votes shifting towards candidates perceived as less risky.[2] Existing studies of crisis voting largely focus on the retrospective evaluation of incumbents in the context of adverse shocks, be they security-related (Gutiérrez, 2014), economy-related (e.g. Nezi 2012; Remmer 1991; Abramson et al. 2007), or broadly about the performance of incumbents in crises (e.g. Smith 1998).

Typically, political scientists rely on either rational actor models or cognitive frameworks to predict vote choice (i.e., Canes-Wrone, Herron and Shotts 2001; Green and Palmquist 1994; Maskin and Tirole 2004). A rational actor model might predict that Bernie Sanders' platform emphasizing universal healthcare access should win the day by appealing to a timely concern of voters. The "flight to safety" perspective generates the opposite prediction – namely that the Sanders campaign would suffer from the increased anxiety generated by the outbreak of the novel coronavirus. By examining the effect of anxiety on the choice between two aspirants for President not part of the administration in power at the time of the anxiety-inducing crisis, this paper provides insight on whether a more general "flight to safety" occurs in voting independent of any attribution of responsibility to the candidates for the crisis itself.

In the case of the 2020 democratic primary, Joe Biden represented safety and Bernie Sanders a disruption of "political as usual". Biden portrayed himself as representing continuity and the security of the known – an "Obama-Biden Democrat", as Biden himself put

[2]One notable exception is Campante, Depetris-Chauvin and Durante 2020, who examine candidates' strategic manipulation of Ebola-induced fear of immigrants in the 2014 US midterm elections and find results complementary to this paper's.

Covid Economics 3, 10 April 2020: 54-84

it in a campaign speech (Fegenheimer and Glueck, 2020). Sanders, in contrast, promised to "change the power the structure in America" (Stewart 2020), portraying himself as a candidate who (in the words of his 2020 campaign spokesman) pushed against "the limits of politics as usual" (Eilperin 2020). Voters apparently understood these divergent appeals, with exit polls in a number of states indicating that Sanders won a majority of those voters for whom the most important quality in a candidate was "Can Bring Needed Change", while Biden was preferred by those for who most valued "Can Unite the Country".[3]

We hypothesize that growing anxiety due to the outbreak of the novel coronavirus reduces the appeal of a disruptive outsider like Sanders. We predict that a political flight to safety will manifest in decreased votes for Sanders where voting occurs after a COVID-19 infection is identified in a Designated Market Area (DMA), all else equal. Those living in places where positive COVID-19 tests occurred are likely to have experienced more anxiety than those for whom infection was a more distant possibility, at least during the period we examine.[4] We use these twin sources of variation in anxiety induced by the disease – i.e., cross-sectional variation due to differences in exposure and temporal variation in the timing of the outbreak – to empirically estimate the effect of COVID-19 on Democratic primary vote choice.

We emphasize that if the anxiety mechanism we describe does *not* obtain, voters might be *more* supportive of Sanders due to his policy platform, making this a particularly hard test for the theory. That is, Sanders' emphasis on universal healthcare should appeal to voters who are exposed to the novel coronavirus and face a more acute need for care. Similarly, Sanders' more expansive protections for working class voters should grow more appealing as the spectre of recession and job losses grew. Given that Sanders' policy platform should be more attractive following COVID-19's emergence, we believe our empirics constitute a hard test of the motivating theory.

3 Data and Methods

We combine several data sources to measure our outcome variable, explanatory variable, and controls.

[3]See exit polls as reported by CNN, https://edition.cnn.com/election/2020/entrance-and-exit-polls/STATE NAME/democratic, e.g. those from Michigan and Washington. In some states – e.g. California – the candidates won a plurality, but not the majority, of those who felt the most important quality was change and unity respectively.

[4]As national media coverage of the outbreak became ubiquitous, our ability to leverage geographic variation declines. We discuss and test these SUTVA assumptions below and in our Supporting Information.

Covid Economics 3, 10 April 2020: 54-84

Outcome Variable

Our outcome variable is the change in the county-level vote share for Bernie Sanders between 2016 and 2020. The 2016 data was obtained from https://www.nytimes.com/elections/2016/results/primaries/[STATE]. The 2020 data was obtained from the "State Results" tab on the https://www.nytimes.com/interactive/2020/03/17/us/elections/results-primary-elections-florida-illinois-arizona.html page at noon on March 18th. At the time of writing, over 97% of counties had 100% reporting.

Throughout our paper, we refer to the "start-date" of the outbreak as either after March 1st, after March 3rd, or after March 10th. These dates are chosen such that the three waves of primary elections in March fall into either treatment or control, as defined in Table 1. We further exploit the timing of elections for robustness checks and placebo tests in our Supporting Information.

Start Date	Control	Treatment
March 1st	Feb	ST, March 10th, & March 17th
March 3rd	Feb & ST	March 10th & March 17th
March 10th	Feb, ST, & March 10th	March 17th

Table 1: Treatment and control elections by outbreak "start date". February (Feb) primaries include IA, NV, and SC. Super Tuesday (ST) primaries include AL, AR, CA, CO, ME, MN, NC, OK, TN, TX, UT, and VA. March 10th primaries include ID, MI, MS, ND, and WA. March 17th primaries include AZ, FL, and IL. (Ohio's was postponed due to the outbreak.) MA, VT, and NH are excluded as they do not aggregate votes by county in reporting totals.

Explanatory Variable

We use data from two separate sources for county-level COVID-19 infection data. The first is the github account for Johns Hopkins University CSSE Coronavirus Resource Center https://github.com/CSSEGISandData/COVID-19/blob/master/csse_covid_19_data/csse_covid_19_time_series/time_series_19-covid-Confirmed.csv. The second is from a non-profit website developed by a variety of academics and professionals called 1Point3Acres https://coronavirus.1point3acres.com/#stat. We scraped these data in the evening of March 16th, 2020 using the `rSelenium` package for `R`. At the time of writing, the JHU data coverage only extends through March 9th while our March 16th scrape of 1Point3Acres is, to the best of our understanding, accurate for that date. Maps of the geographic distribution of the outbreak by DMA on March 2nd (the eve of Super Tuesday), March 9th (the eve of the second round of multiple state primaries), and March 16th (the eve of Arizona, Illinois, and Florida) are presented in Figure 1.

Covid Economics 3, 10 April 2020: 54-84

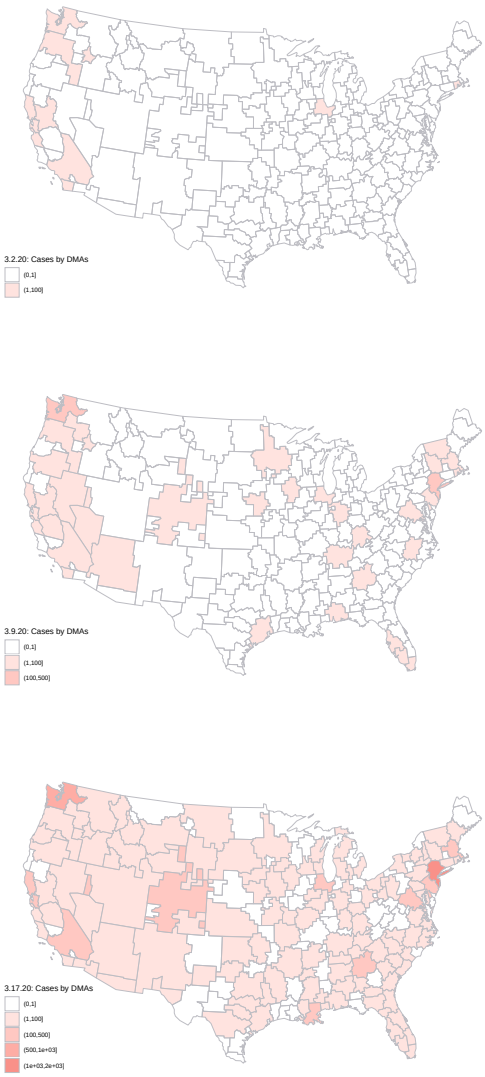


Figure 1: Maps of the geographic and temporal variation in the spread of the virus between March 2nd and March 17th. March 2nd and March 9th data are from Johns Hopkins University. March 16th data are from 1Point3Acres.

Controls

We obtain a rich set of pre-treatment county-level controls from the five year averages of the American Community Survey (2018). These county-level controls are:

Covid Economics 3, 10 April 2020: 54-84

- total population
- % of the population that is rural
- % of the population that is white
- % of the population with a bachelor's degree
- the county's old-age dependency ratio (retirees to workers)
- Share of households that are headed by a woman without a husband present
- % of the population that speaks only English
- % of population between 18 and 64 that is below the poverty level
- % of the population employed in manufacturing
- the county-level unemployment rate
- the county-level labor force participation rate
- the median household income

Methods

We are interested in identifying the causal effect of exposure to the novel coronavirus on Democratic primary voters' decisions. While the outbreak of COVID-19 was an exogenous shock to voter anxiety, it is confounded in three ways. First, the timing of treatment is colinear with other explanations for changing electoral fortunes, such as the decision by several primary candidates to drop out (Staff, 2020), signaling a consolidation of party support behind Biden (Yglesias and Beauchamp, 2020). A simple before-after comparison of election returns would be unable to disentagle our "flight to safety" theory from a coincidental shift in electoral momentum.

Second, we might expect that older voters are more dissuaded from appearing at the polls following the appearance of COVID-19 due to the increased risks of exposure. Insofar as younger voters are relatively more supportive of Sanders, this would bias our results in a conservative direction, making it harder to identify a negative relationship between exposure and Sanders' vote share.[5]

Third, if areas that were already more anti-Sanders were also those most exposed to the outbreak, our results would pick up a spurious selection effect.

[5]There is also the possibility of a selection bias which would obtain of exposed counties were more anti-Sanders to begin with. We predict Sanders' 2016 voteshare as a function of exposure and find, if anything, these counties are *more* pro-Sanders. To the degree that there is selection bias, it works against our results.

Covid Economics 3, 10 April 2020: 54-84

We posit that anxiety due to the disease is a function of both temporal and geographical variation, allowing us to address these confounds. We define exposure as binary variable taking on the value of 1 if a county c resides in a designated market area (DMA) with confirmed cases of COVID-19 on the eve of their primary election date, and 0 otherwise, denoted with $COVID_c$. We estimate the difference in the change in Bernie Sanders' vote share from 2016 to 2020, between exposed and unexposed counties, or $E[\Delta Y_1 - \Delta Y_0]$, where ΔY is the county-level change in Sanders' primary vote share, and the 1 and 0 subscripts represented treated and control counties, respectively.[6] We also include fixed effects for the date of the election, thus restricting comparisons to differences in a given election "wave" and thus partially controlling for anything that may have varied between election dates (e.g. number of candidates in the race). This helps ensure that our findings are indeed picking up on growing anxiety around COVID-19 associated with local knowledge (via media coverage) of cases. Our simplest specification, as per our PAP, takes the form:

$$\Delta Y_c = \beta_0 + \beta_1 COVID_c + \gamma \mathbf{X} + \lambda + \epsilon_c \tag{1}$$

where $\mathbf{X}$ is the vector of county-level controls summarized above, and λ are date of election fixed effects.

However, since counties with earlier exposure to COVID-19 are disproportionately more densely populated coastal areas, this specification risks dissimilarities between treatment and control counties. To address this bias, we also leverage the staggered timing of both vote date and exposure to estimate a pseudo difference-in-differences (DID) specification where we compare the difference in the outcome between treated and control groups prior to the outbreak to the difference in these groups following the outbreak.

This design is complicated by the fact that, unlike standard DID settings, we do not observe outcomes in the pre and post period for every unit, precluding our ability to measure $E[Y_{i,t=1} - Y_{i,t=0}]$ at the county level. Instead, we must assume that those counties who voted in the pre period but would go on to be exposed to COVID-19 are valid counter-factuals for those counties that were exposed to COVID-19 and voted in the post period. Similarly, we must assume that the control counties that voted in the pre period (i.e., those that did not experience the COVID-19 outbreak in the post period) are valid counterfactuals for the control counties that voted in the post period.

We augment our conditional independence assumption (CIA) with matching and balancing strategies to ensure we are comparing otherwise similar counties who differ only in the timing of their exposure to COVID-19. We obtain good balance on a rich set of pre-treatment

[6] By defining the outbreak as happening after Super Tuesday when the field narrowed to a contest between Sanders and Biden the 2020 Sanders vote share is more directly comparable to the 2016 contest, effectively a two-candidate contest between Sanders and Hillary Clinton. To the extent that this comparison fails we expect it will bias results downwards rather than increase the chance of spurious statistically significant findings; when the field narrowed after Super Tuesday that increased the Sanders vote share and thus biases against us finding decreased support for Sanders, all else equal. Furthermore we also acknowledge that several states (including CO, ME, MN, UT, and WA) switched from caucuses to primaries between 2016 and 2020. While Bernie does better on average in caucus states, all five of these are included in the pre-period in our main specification, ensuring that they do not drive our results.

Covid Economics 3, 10 April 2020: 54-84

covariates using either nearest neighbor matching (based on minimized Mahalanobis distance), or covariate balanced propensity score weights (CBPS).[7] Our causal interpretation hinges on our claim that exposure is as-if randomly assigned to counties conditional on the observables we control for, match on, and balance over.

One final concern that we believe grows more problematic as the virus spreads is the Stable Unit Treatment Value Assumption, or SUTVA. Substantively, this assumption requires that our control counties are not affected by treatment spillovers from treated counties. Our treatment exposure is defined at the DMA-level, based on the assumption that the salience of the disease is elevated via local media markets which report on more geographically proximate cases. We believe this is sensible for the beginning of March, when the virus was just beginning to spread across the United States. However, by the time of the March 17th elections, national media outlets (e.g. cable news, newspapers, news websites, and online social media such as Facebook (Roose and Dance, 2020)) had shifted coverage to focus almost exclusively on the outbreak as the crisis worsened. Thus many of our notionally "control" counties experienced substantial levels of anxiety despite not residing in a DMA with confirmed cases of the virus, with "control" counties becoming decreasingly valid counterfactuals for counties in a DMA where a COVID-19 case had been diagnosed with each passing week. We include an exhaustive series of pairwise comparisons in which we define one primary election as treated, and another as control, in our Supporting Information.

4 Results

Our main results are summarized in Table 3, in which treatment is defined at the DMA as all confirmed cases of COVID-19 on March 9th, 2020 as reported in the Johns Hopkins University data as of March 21st, 2020. The first two columns present the coefficients on a binary measure of exposure (1 if any cases were recorded in the DMA, 0 otherwise), and a continuous count of the number of confirmed cases as of March 9th, 2020. Clustered standard errors at the DMA-election are presented in parentheses. The coefficients in the odd-numbered columns represent the standard deviation change in support for Bernie Sanders between 2020 and 2016 (approximately 12 percentage points) due to the impact of exposure to the virus. The coefficients in the even-numbered columns represent the same standard deviation change in Sanders' support due to the impact of a one standard deviation increase in the number of confirmed cases of COVID-19 (approximately 12.5 cases).

The results indicate that counties that voted after Super Tuesday (March 3rd) and which were exposed to the novel coronavirus were less likely to support Sanders as compared to counties that voted prior to March 10th and counties that voted on or after March 10th but did not reside in a DMA with any reported cases. According to Columns 1 and 2, being exposed to the virus corresponds to an estimated 0.36 standard deviation decline in support for Sanders as compared to his 2016 vote share, over and above the decline in Sanders vote share in matched counties in the control group. This corresponds to an expected change

[7] Balance results are included in our Supporting Information.

Covid Economics 3, 10 April 2020: 54-84

of approximately 4.1 percentage points less support for Sanders compared to the 2016 vote share he enjoyed in the average county. This result is reinforced at the intensive margin, as illustrated by the negative and significant coefficient in column 2, suggesting that a standard deviation increase in the number of confirmed cases (roughly 99 new cases in the DMA) corresponds to a 0.12 standard deviation decline in support for the Sanders' campaign, or roughly 1.5 percentage points, relative to 2016.

However, the results in columns 1 and 2 rely on the assumption that insulated and exposed counties are valid counterfactuals for each other after controlling for a variety of demographic and economic county-level factors. In columns 3 and 4, we reduce our reliance on this assumption by employing a nearest-neighbors matching strategy in which we identify the most similar control county for each treated county in our dataset based on the same county-level covariates. We use Mahalanobis distance measures to summarize the difference across our twelve county-level covariates and choose the county that is most similar to each treated county in terms of this distance measure.[8]

Substantively, this approach strengthens our conditional independence claim that we are comparing otherwise similar counties that differ only in the timing of their exposure and the number of cases experienced, which are both exogenous events. As indicated in columns 3 and 4 of Table 3, this matching strategy strengthens our conclusions, suggesting that exposure to the pandemic reduces support for the Sanders' campaign by almost 60% of a standard deviation for the binary measure (column 3), and a third of a standard deviation for the continuous measure (column 4). These suggest that the substantive impact on COVID-19 exposure is non-trivial, accounting for approximately 7 percentage points slippage for Sanders vote share between 2016 and 2020.

However, matching strategies such as the method we implement require us to jettison a substantial number of observations. As indicated at the bottom of Table 3, we rely on less than 15% of our total observations to draw these conclusions, choosing only those control observations that are most similar to the treated according to the Mahalanobis distance measure across the 12-dimensional covariate space. As a final test, we instead employ a weighting strategy that re-weights the control observations to best approximate the treated observations, without throwing any information away. Specifically, we implement the optimal weighting method of Zubizarreta (2015), achieve good balance across all observables, as summarized in Table 4 in our Supporting Information.

Columns 5 and 6 summarize the weighted estimates, suggesting that exposure to the novel coronavirus predicts a decline in support quite similar to the unmatched regression in columns 1 and 2 - with Sanders' support declining slightly less than half a standard deviation for the binary treatment measure, and 0.12 standard deviations in response to a standard deviation increase in cases. Notably, our predictive power increases meaningfully from an R^2 of just under 0.40 to over 0.45 with the weighting method employed in columns 5 and 6.

[8]We achieve good balance, as demonstrated in the balance tests reported in the Supporting Information, specifically Figure 8.

Covid Economics 3, 10 April 2020: 54-84

Table 2: Main Results: Change in Sanders Support ∼ Exposure

	Dependent variable: Δ Sanders Vote Share					
	Basic		Matching		Weighting	
	Bin.	Cont.	Bin.	Cont.	Bin.	Cont.
	(1)	(2)	(3)	(4)	(5)	(6)
Treatbin	−0.361		−0.600**		−0.472**	
	(0.231)		(0.251)		(0.226)	
Treatcont		−0.102***		−0.300***		−0.119***
		(0.012)		(0.036)		(0.006)
Tot pop	0.036	0.032	0.131**	0.109*	−0.015	−0.016
	(0.030)	(0.028)	(0.065)	(0.062)	(0.011)	(0.011)
Old age dep ratio	0.024	0.017	0.047	0.055	0.098**	0.096**
	(0.046)	(0.045)	(0.087)	(0.091)	(0.045)	(0.044)
Bachelor's degree	−0.071	−0.066	0.156	0.194*	−0.030	−0.024
	(0.054)	(0.055)	(0.115)	(0.110)	(0.060)	(0.060)
Female HH no husband	0.380***	0.383***	0.346***	0.329***	0.363***	0.357***
	(0.056)	(0.056)	(0.115)	(0.099)	(0.068)	(0.068)
Md inc HH	0.084	0.087	−0.136	−0.121	0.027	0.027
	(0.064)	(0.064)	(0.184)	(0.177)	(0.056)	(0.055)
Manufacturing	0.125***	0.129***	0.080	0.085	0.059	0.062
	(0.048)	(0.048)	(0.078)	(0.067)	(0.047)	(0.047)
Speak only english	−0.309***	−0.306***	−0.195*	−0.190**	−0.360***	−0.360***
	(0.051)	(0.051)	(0.101)	(0.094)	(0.041)	(0.042)
Below poverty level	0.021	0.020	−0.106	−0.110	−0.015	−0.017
	(0.039)	(0.039)	(0.094)	(0.089)	(0.050)	(0.050)
White	−0.069	−0.067	0.012	−0.006	−0.132**	−0.134**
	(0.066)	(0.066)	(0.098)	(0.093)	(0.062)	(0.063)
LFPR	−0.073*	−0.072*	−0.163	−0.122	−0.059	−0.061
	(0.041)	(0.040)	(0.145)	(0.160)	(0.055)	(0.055)
Unem rate	0.040	0.043	0.147**	0.150**	0.034	0.037
	(0.041)	(0.041)	(0.074)	(0.070)	(0.036)	(0.036)
Rural	0.017	0.027	−0.136	−0.085	−0.066	−0.064
	(0.037)	(0.037)	(0.096)	(0.102)	(0.059)	(0.058)
Constant			0.300***	0.000	−0.083	−0.116
			(0.108)	(0.110)	(0.071)	(0.072)
Observations	1,657	1,657	234	234	1,657	1,657
R^2	0.420	0.424	0.389	0.388	0.455	0.455
Election FE	Y	Y	N	N	N	N

Note: *p<0.1; **p<0.05; ***p<0.01
Robust standard errors clustered at the DMA-election indicated in parentheses.

Covid Economics 3, 10 April 2020: 54-84

Differences-in-Differences

The preceding results exploit temporal variation in exposure, but operationalize this variation in cross-sectional statistical analyses. In the following section, we instead turn to a difference-in-differences specification in which we compare the difference between treated and control counties prior to the outbreak with the difference in Sanders support among these groups of counties following the outbreak.

The left panel of Figure 2 plots the simple averages of treated (blue) and control (red) groups prior to (left) and following (right) the outbreak of the virus. Based solely on this simple difference-in-differences, one might draw several conclusions. First, there appears to be a decline in support for Sanders among both treated and control counties following the outbreak of the novel coronavirus. Second, there is some evidence suggesting that the counties that were exposed to the virus and voted after the outbreak shifted more strongly against Sanders than those counties that were not exposed. The right panel of Figure 2 presents bivariate regressions across groups, suggesting that there is a weak positive association between the number of cases (logged, x-axis) and the change in Sanders vote share between 2016 and 2020 in the pre-outbreak period (March 3rd and earlier, indicated in red). Conversely, there is a clear negative correlation following the outbreak, as indicated by the blue line.

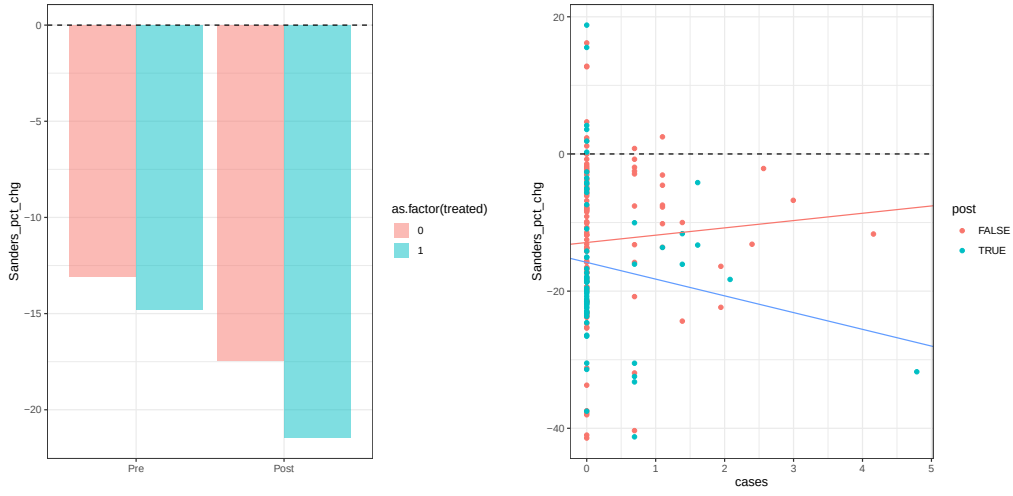


Figure 2: Descriptive differences between treated and control voting behavior before and after the outbreak, defined as starting on March 4th. Left panel groups counties by whether they were exposed as of March 9th, right panel plots the logged cases as of March 9th by whether the county voted prior to, or following, the outbreak.

These plots are descriptive, and are not meant to support well-identified inferential conclusions. As such, we turn to our conditional difference-in-difference regression specifica-

Covid Economics 3, 10 April 2020: 54-84

tions. We examine both the basic conditional results as well as the matched and weighted results using different dates for the beginning of "treatment" in Figure 3. When we set the treatment period to March 1st and include the exposed counties voting in Super Tuesday among our treated group, we find significant evidence that exposure leads to declining support for Bernie Sanders. However, this effect declines over time, with the result attenuating to a null when we define the outbreak starting after Super Tuesday and even some suggestive evidence that the virus actually benefited Sanders among the counties voting on March 17th. We suspect that these patterns reflect a broadening of the national coverage of the outbreak, prompting SUTVA violations when we define exposure at the DMA. We test this suspicion in our Supporting Information, and find that redefining the unit of exposure at the state level recovers our main results. This suggests that the "flight to safety" occurs throughout, but that as time passes information about diagnoses induces panic not just within the DMA but state-wide.

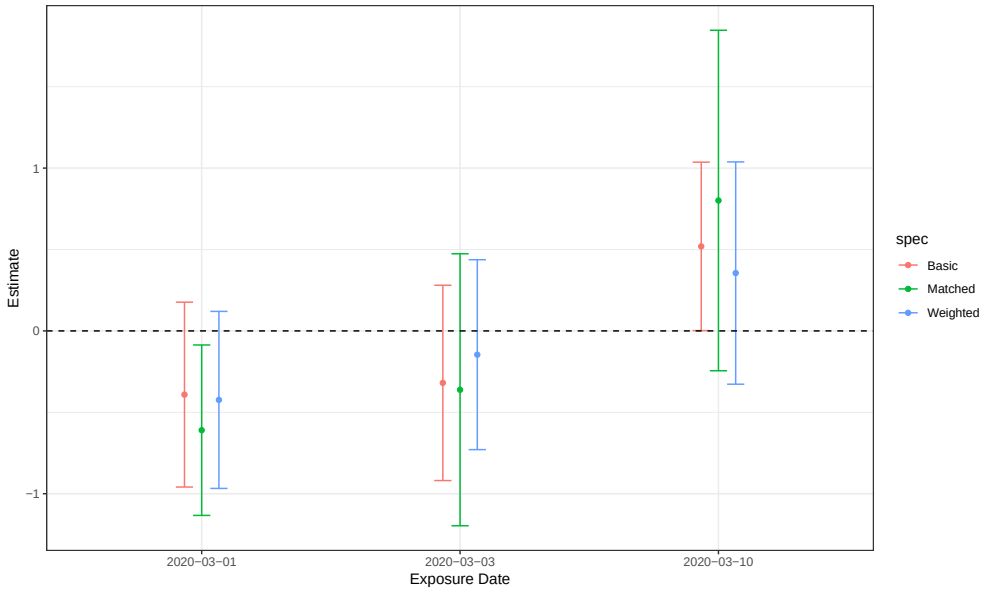


Figure 3: Diff-in-diff estimates for different start-dates of the outbreak (x-axis).

The Supplementary Information includes additional analyses specified in the pre-analysis plan, including robustness checks, exploring sensitivity to shifting definitions of treatment, choices of matching strategy, and balancing weights. These tests confirm the main findings described above, with stronger results if we define treatment using deaths due to COVID-19 instead of confirmed cases.

One of the primary inquiries we pursue in the Supplementary Information is to explore whether alternative mechanisms explain the pattern of results. We find that differential turnout is not a likely explanation of these findings, with turnout not being suppressed

until the March 17th election and no evidence that age was particularly determinative. We were also concerned that our results might be driven by a "party consolidation" effect. To explore this we run a placebo test by permuting treatment, breaking the observed relationship between exposure and voting behavior. We find consistent null results, suggesting that our findings are not being driven by any secular Democratic party elite consolidation behind Biden over time. Finally, we examine whether our results are driven by a selection effect in which areas predisposed to vote against Sanders were disproportionately exposed to the outbreak. If anything, our results indicate the opposite – counties that were *more* supportive of Sanders in 2016 were more likely to be exposed in 2020, indicating that our results obtain despite a conservative selection effect.

While not the only possible interpretation, we believe these results provide suggestive evidence that voter anxiety and a flight to safety is the mechanism underlying the decline in Sanders' vote share induced by COVID-19's appearance across space and time.

5 Discussion

These findings explore the substantive political effects of the novel coronavirus. We conclude that exposure to the outbreak lead to a relative decline in Sanders support, showing in our primary specification (Table 3) that COVID-19 exposure in a local media market depresses vote shares for Sanders by up to 7 percentage points. The size of the COVID-19 effect is not large enough to have made Bernie Sanders the front-runner in the absence of the novel coronavirus' appearance. It is, however, a far from trivial effect.

We posit that these results are consistent with a political "flight to safety", complementing work by Campante, Depetris-Chauvin and Durante 2020, who find Ebola-induced fear had substantial electoral consequences in the 2014 US midterm elections. We recognize that there may be other alternative pathways that explain the effect we document. We disconfirm several alternative mechanisms, including the coincident rallying of Democratic party support behind Biden, the possibility that the outbreak disproportionately suppressed Sanders supporters, and the selection effect of more anti-Sanders areas being more exposed. We leave an explicit test of the proposed mechanism for future research.

Covid Economics 3, 10 April 2020: 54-84

Covid Economics 3, 10 April 2020: 54-84

References

Abramson, Paul R, John H Aldrich, Jill Rickershauser and David W Rohde. 2007. "Fear in the voting booth: The 2004 presidential election." *Political Behavior* 29(2):197–220.

Adrian, Tobias, Richard K Crump and Erik Vogt. 2019. "Nonlinearity and Flight-to-Safety in the Risk-Return Trade-Off for Stocks and Bonds." *The Journal of Finance* 74(4):1931–1973.

Blackwell, Matthew, Stefano Iacus, Gary King and Giuseppe Porro. 2009. "cem: Coarsened exact matching in Stata." *The Stata Journal* 9(4):524–546.

Campante, Filipe, Emilio Depetris-Chauvin and Ruben Durante. 2020. "The Virus of Fear: The Political Impact of Ebola in the US.".

Canes-Wrone, Brandice, Michael C Herron and Kenneth W Shotts. 2001. "Leadership and pandering: A theory of executive policymaking." *American Journal of Political Science* pp. 532–550.

Eilperin, Juliet. 2020. "Bernie Sanders's climate record in Congress: Lots of advocacy, no compromise." *Washington Post* .
URL: *https://www.washingtonpost.com/climate-environment/bernie-sanderss-climate-record-in-congress-lots-of-advocacy-no-compromise/2020/03/12/7aa2772a-5fd7-11ea-b29b-9db42f7803a7_story.html*

Fegenheimer, Matt and Katie Glueck. 2020. "Joe Biden Had a Big Night. He Needs Another in 72 Hours." *New York Times* .
URL: *https://www.nytimes.com/2020/02/29/us/politics/joe-biden-super-tuesday.html*

Getmansky, Anna and Thomas Zeitzoff. 2014. "Terrorism and voting: The effect of rocket threat on voting in Israeli elections." *American Political Science Review* 108(3):588–604.

Green, Donald Philip and Bradley Palmquist. 1994. "How stable is party identification?" *Political behavior* 16(4):437–466.

Gutiérrez, Sandra Jessica Ley. 2014. Citizens in fear: political participation and voting behavior in the midst of violence PhD thesis tesis de doctorado, Durham, Duke University, 2014), en especial 1-10 ....

Imai, Kosuke and Marc Ratkovic. 2014. "Covariate balancing propensity score." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 76(1):243–263.

Inghelbrecht, Koen, Geert Bekaert, Lieven Baele and Min Wei. 2013. "Flights to safety." *NBER WORKING PAPER SERIES* 19095:1–51.

Leonhardt, David. 2020. "A Complete List of Trump's Attempts to Play Down Coronavirus." *New York Times* .
URL: *https://www.nytimes.com/2020/03/15/opinion/trump-coronavirus.html*

Maskin, Eric and Jean Tirole. 2004. "The politician and the judge: Accountability in government." *American Economic Review* 94(4):1034–1054.

Montalvo, Jose G. 2011. "Voting after the bombings: A natural experiment on the effect of terrorist attacks on democratic elections." *Review of Economics and Statistics* 93(4):1146–1154.

Nezi, Roula. 2012. "Economic voting under the economic crisis: Evidence from Greece." *Electoral studies* 31(3):498–505.

Remmer, Karen L. 1991. "The political impact of economic crisis in Latin America in the 1980s." *American Political Science Review* 85(3):777–800.

Roose, Kevin and Gabriel J.X. Dance. 2020. "The Coronavirus Revives Facebook as a News Powerhouse." *New York Times* .
URL: *https://www.nytimes.com/2020/03/23/technology/coronavirus-facebook-news.html*

Smith, Alastair. 1998. "International crises and domestic politics." *American Political Science Review* 92(3):623–638.

Staff, FiveThirtyEight. 2020. "Who Will Win the 2020 Democratic Primary?" *FiveThirtyEight.com* .
URL: *https://projects.fivethirtyeight.com/2020-primary-forecast/?cid=rrpromo*

Stewart, Cameron. 2020. "Now it's Bernie v Biden." *The Australian* .
URL: *https://www.theaustralian.com.au/inquirer/super-tuesday-now-its-bernie-v-biden/news-story/11aff0ef309b7ce6be787d4457ab9510*

Yglesias, Matthew and Zack Beauchamp. 2020. "4 winners and 3 losers on Super Tuesday." *Vox.com* .
URL: *https://www.vox.com/2020/3/3/21163826/super-tuesday-winners-losers-biden-sanders*

Zubizarreta, José R. 2015. "Stable weights that balance covariates for estimation with incomplete outcome data." *Journal of the American Statistical Association* 110(511):910–922.

Covid Economics 3, 10 April 2020: 54-84

Covid Economics 3, 10 April 2020: 54-84

Supplementary Information

"Party Decides" Placebo Tests

The main results suggest that exposure to the novel coronavirus results in a greater decline in support for a Sanders presidency than what we observe in relatively insulated counties or those that voted prior to the outbreak. However, even with our matching and weighting strategies to argue that the outbreak is as-good-as-randomly assigned conditional on observables, there remains a concern with regards to timing. Specifically, our definition of "exposure" is defined as any county residing within a DMA that had confirmed cases of the virus as of March 9th, 2020. Effectively, this definition risks conflating other contemporaneous changes in the political landscape that occurred between Super Tuesday (March 3rd), and the 10 states that voted afterwards (7 on March 10th, 3 on March 17th). Specifically, this period saw the Democratic party rally around the establishment candidacy of Joseph Biden as several candidates dropped out of the race and endorsed Biden.

To confirm our results are not simply picking temporal variation and the momentum shift that occurred on Super Tuesday, we run a placebo test in which we permute our explanatory variable while keeping our definitions of pre and post exposure at March 3rd. If our main results are driven by the "party decides" phenomenon, we should still find a significant negative relationship between Sanders' declining vote share and our permuted treatment. We bootstrap sample our data, each time drawing a permuted explanatory variable, and re-estimate our main specifications. As illustrated in Figure 4, our results are noisily estimated nulls.

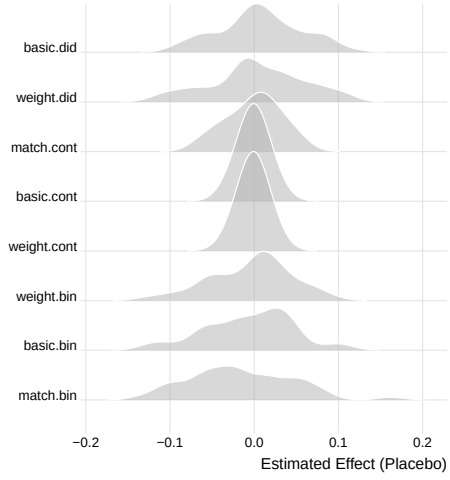


Figure 4: Placebo test bootstrapped estimates generated by permuting the COVID-19 cases while keeping the exposure start date starting after Super Tuesday.

Covid Economics 3, 10 April 2020: 54-84

The results summarized above use outbreak dates to separate treated and control elections as per Table 1, meaning that all elections prior and including a given cutoff are defined as control, and all elections following the cutoff are treated. We also re-run our analyses by conducting a series of pairwise comparisons in which one election is defined as control and the other is defined as treated. Doing so allows us to identify where (and more precisely, when) our effects obtain. We treat all primary elections prior to Super Tuesday as one group in order to include multiple states in each treatment and control condition. Figure 5 summarizes these results for every specification at our disposal. The Democratic party consolidated support behind Biden ahead of Super Tuesday. As Figure 5 demonstrates, the results do not depend on comparing the period before Super Tuesday to the period after, and thus are not collinear with a "party consolidation" effect, though we cannot rule out that such an effect may also contribute to the findings in the panel comparing Super Tuesday to pre-Super Tuesday voting states.

We can also reverse the temporal sequencing of these results, creating placebo tests for our conclusions. We treat the later election as the control, and the earlier as the treated, and estimate the effect of future COVID-19 cases on vote choice. Our results are summarized in Figure 6

Turnout and Age

Our main results suggest that Bernie Sanders was hurt by the outbreak of the virus, although the effect attenuated over time. We argue that this is consistent with our theorized mechanism of an electoral "flight to safety". However, an alternative mechanism might be that the outbreak differentially reduced turnout among different voting groups. One plausible scenario might be that those most threatened by exposure might be less likely to turn out. If this group is also more likely to support Sanders, there is an alternative explanation for the effects we document. Of course, Sanders' popularity among young voters is well-documented, while the elderly are most threatened by the virus. As such, if this mechanism is operating, it should be the case that older voters are *less* likely to turn out, and that therefore we should see an *increase* in support for Sanders from younger voters.

We examine this alternative mechanism by replacing the change in Sanders' vote share with the change in county-level turnout. As illustrated in Figure 7, there is little evidence to suggest that such an age-based dynamic is at play. As illustrated, it appears that turnout wasn't suppressed by the virus until after the March 10th elections, when it reduced turnout for Illinois, Arizona, and Florida. (Ohio chose to delay its primaries due to concerns about the virus.) Furthermore, the marginal effects of exposure across different populations of older voters suggests that COVID-19 was *less* depressing to turnout in counties with more elderly constituents, although the interaction effects are not statistically significant. Taken together, these findings suggest that turnout was not appreciably influenced by COVID-19 exposure as of March 3rd (95% confidence intervals for the marginal effects always include zero) and that, although there is some evidence of heterogeneity across counties by share of the population

Covid Economics 3, 10 April 2020: 54-84

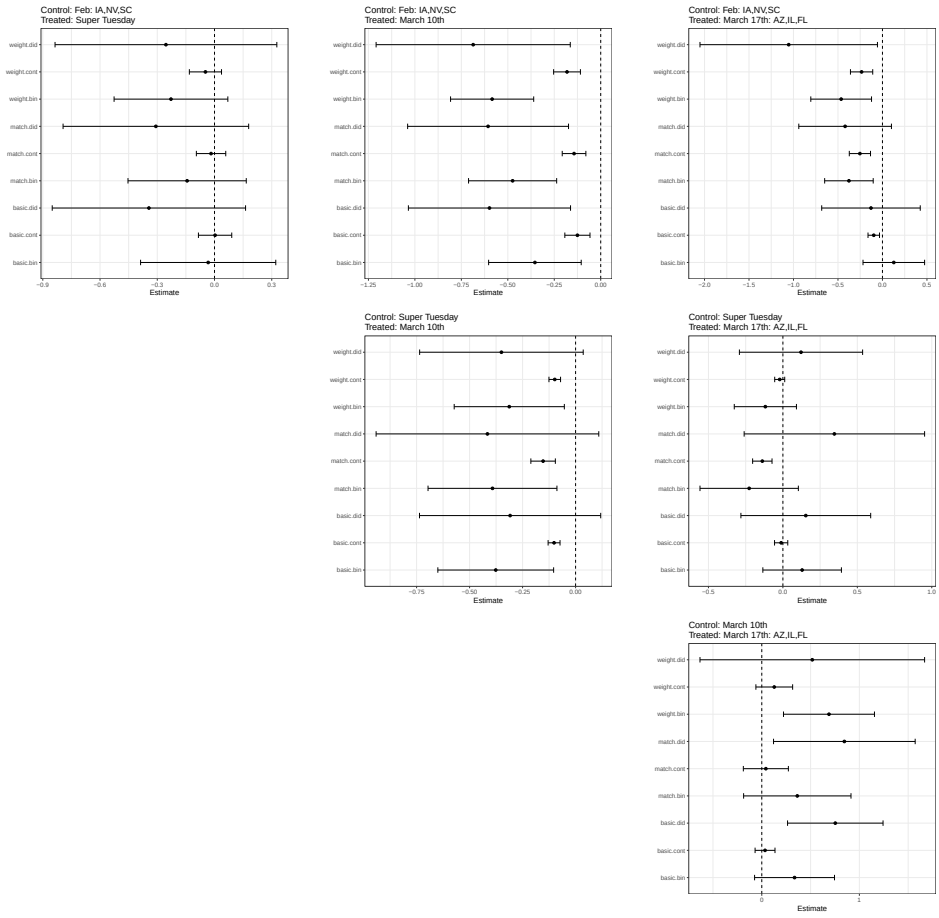


Figure 5: Pairwise election comparisons by control (rows) and treatment (columns) given in plot titles.

older than 64 years of age, these interaction coefficients are also insignificant. As such, we are confident in our conclusion that the reduction in Sanders' support is attributable to more than simply shifting turnout dynamics across the period of analysis.

Selection Effects

Our results would be spurious if the outbreak disproportionately affected parts of the country that were already anti-Bernie to begin with. Although our matching and weighting strategies are one solution to minimizing this risk, we can also evaluate the identification

Covid Economics 3, 10 April 2020: 54-84

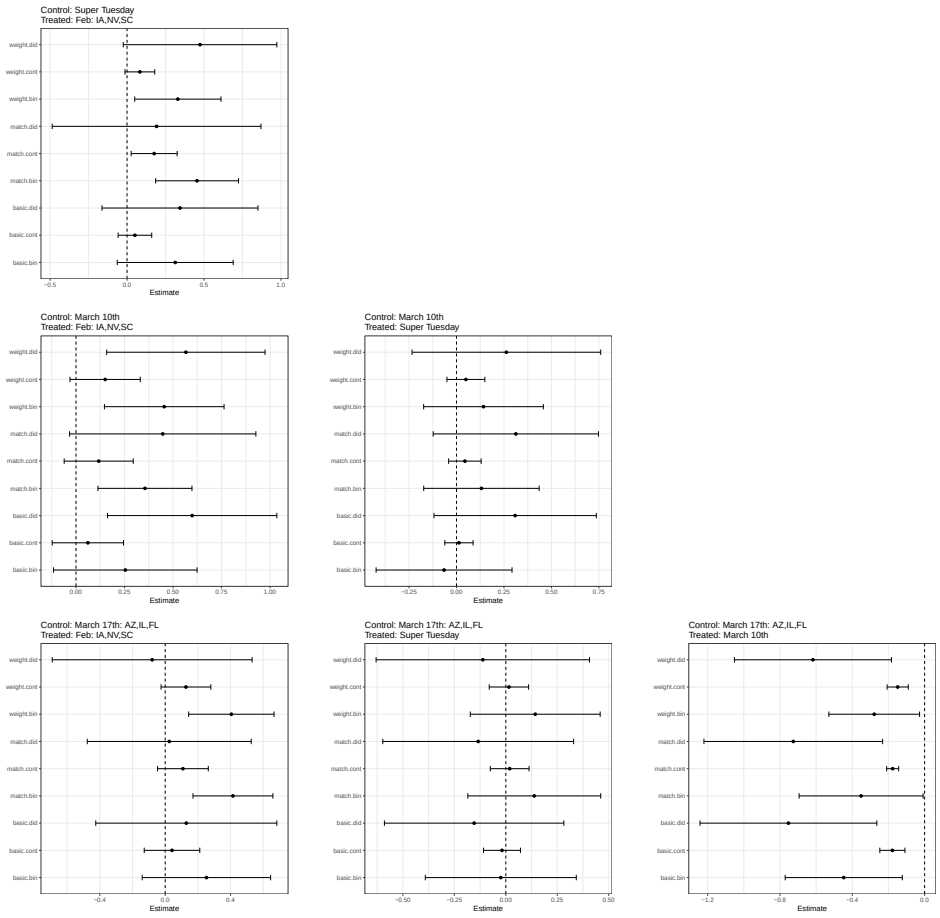


Figure 6: Pairwise election comparisons by control (rows) and treatment (columns) given in plot titles.

challenge directly. We replace our main outcome variable with Sanders' 2016 voteshare, testing whether 2020 exposure rates also predict 2016 Sanders support. We find, if anything, a source conservative bias as shown in Table **??**. Specifically, the counties that were more exposed to the outbreak in 2020 were *more* supportive of Sanders in 2016, revealing a selection effect that works against our main results.

Covid Economics 3, 10 April 2020: 54-84

Table 3: Selection Effects: 2016 Sanders Voteshare ∼ Exposure

	Dependent variable: 2016 Sanders Vote Share					
	Basic		Matching		Weighting	
	Bin.	Cont.	Bin.	Cont.	Bin.	Cont.
	(1)	(2)	(3)	(4)	(5)	(6)
Treatbin	0.738***		0.779***		0.657***	
	(0.268)		(0.258)		(0.254)	
Treatcont		0.162***		0.356***		0.159***
		(0.007)		(0.040)		(0.007)
Tot pop	0.013	0.027	−0.100*	−0.074	0.039***	0.039***
	(0.025)	(0.018)	(0.061)	(0.059)	(0.009)	(0.009)
Old age dep ratio	−0.146***	−0.123***	−0.322***	−0.333***	−0.218***	−0.216***
	(0.045)	(0.044)	(0.109)	(0.117)	(0.053)	(0.053)
Bachelor's degree	0.111**	0.094**	0.051	0.001	0.056	0.048
	(0.045)	(0.045)	(0.104)	(0.097)	(0.071)	(0.071)
Female HH no husband	−0.442***	−0.451***	−0.352***	−0.337***	−0.436***	−0.428***
	(0.048)	(0.051)	(0.109)	(0.097)	(0.088)	(0.087)
Md inc HH	−0.109*	−0.102	0.185	0.176	−0.055	−0.056
	(0.064)	(0.066)	(0.177)	(0.174)	(0.075)	(0.075)
Manufacturing	−0.130***	−0.142***	0.016	0.009	−0.037	−0.041
	(0.040)	(0.039)	(0.076)	(0.058)	(0.054)	(0.054)
Speak only english	0.096**	0.099***	0.040	0.033	0.150***	0.150***
	(0.038)	(0.039)	(0.094)	(0.093)	(0.044)	(0.045)
Below poverty level	0.097**	0.103**	0.220**	0.226**	0.116**	0.119**
	(0.042)	(0.043)	(0.110)	(0.094)	(0.059)	(0.060)
White	0.319***	0.326***	0.068	0.087	0.256***	0.259***
	(0.052)	(0.052)	(0.080)	(0.085)	(0.063)	(0.065)
LFPR	0.108***	0.101***	0.038	−0.018	0.150**	0.152**
	(0.037)	(0.039)	(0.159)	(0.178)	(0.076)	(0.075)
Unem rate	0.047	0.052	−0.003	−0.003	0.116***	0.112***
	(0.035)	(0.034)	(0.094)	(0.085)	(0.037)	(0.038)
Rural	−0.071*	−0.111***	0.201**	0.136	0.098	0.095
	(0.037)	(0.034)	(0.092)	(0.096)	(0.067)	(0.070)
Constant	−0.052	0.000	−0.390***	−0.000	0.228***	0.275***
	(0.056)	(0.056)	(0.099)	(0.133)	(0.076)	(0.085)
Observations	1,657	1,657	234	234	1,657	1,657
R^2	0.504	0.498	0.436	0.410	0.374	0.368
Election FE	Y	Y	N	N	N	N

Note: *p<0.1; **p<0.05; ***p<0.01
Robust standard errors clustered at the DMA-election indicated in parentheses.

Covid Economics 3, 10 April 2020: 54-84

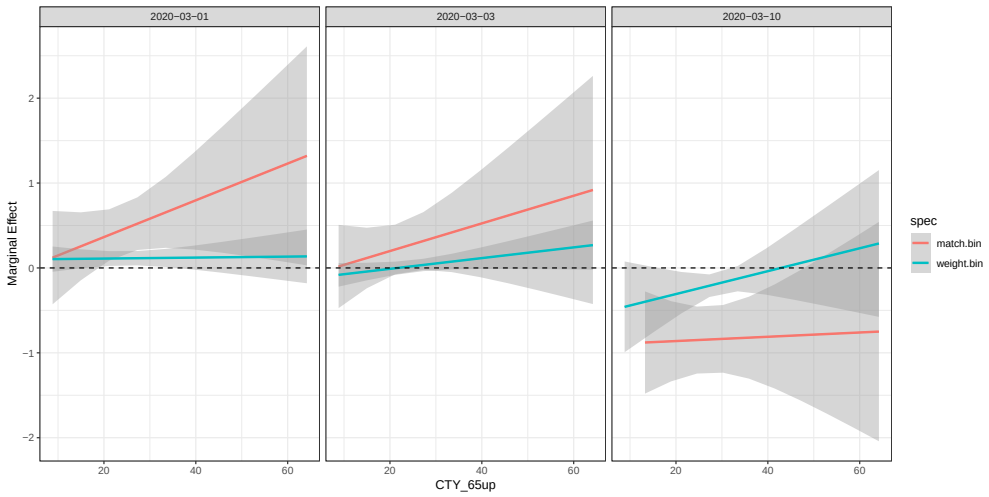


Figure 7: Marginal effects of exposure on turnout (y-axes) across counties with smaller and larger proportions of their population older than 64 years of age (x-axes) by outbreak onset date (panels). None of the marginal effects are themselves significant at conventional levels.

Balance and Weighting Robustness

We achieve good balance on both the matching and weighting strategies employed in the body of our paper. Figure 8 plots the improvements to balance on observables between treated and control units generated by our choice of nearest-neighbor matching using minimized Mahalanobis distance. And Table 4 summarizes the differences in treated and control covariates prior to, and following the `optmatch` weights. In both cases, we successfully adjust our data to better reflect the distribution of observables in an experimental context in which treatment is randomly assigned.

We also confirm the robustness of our main findings to different choices about the matching strategy and the balancing weights. Specifically, we re-estimate our main findings replacing the `optweight` method of Zubizarreta (2015) with covariate balancing propensity scores (CBPS, Imai and Ratkovic 2014), and replacing the nearest neighbor matching strategy with coarsened exact matching (CEM, Blackwell et al. 2009). The former robustness check yields substantively and statistically similar findings to our main results, as illustrated in Tables 5 and 6.

Moving from nearest neighbor matching based on Mahalanobis distance to the CEM method requires us to reduce the number of county-level covariates we use for matching. This is due to the default parameter settings yielding only two matched observations, precluding our ability to estimate treatment effects. We reduce our set of covariates to select the following six across which we can obtain reasonably good performance on our balance tests

Covid Economics 3, 10 April 2020: 54-84

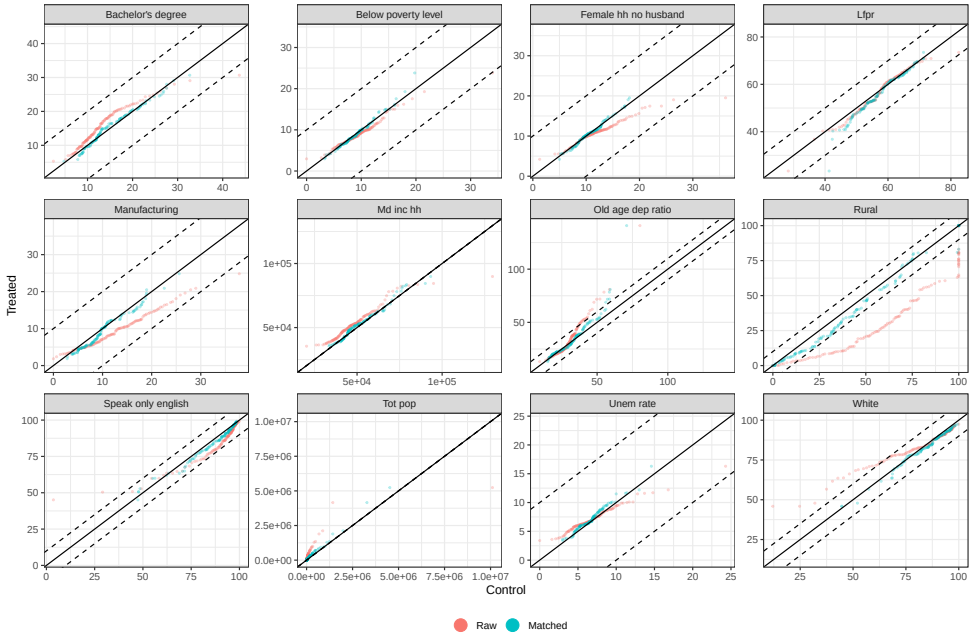


Figure 8: Balance of treated and control covariates before (red) and after (blue) matching. 45 degree line indicates perfect match.

Table 4: Weighting Balance Checks

	Covs	Diff_Unm	Bal_Test_Unm	Diff_Match	Bal_Test_Match
1	County Pop	0.320	Not Balanced, >0.05	0	Balanced, <0.05
2	Old-Age Dep Ratio	0.240	Not Balanced, >0.05	0	Balanced, <0.05
3	% Bachelor's	0.370	Not Balanced, >0.05	0	Balanced, <0.05
4	Female HH, No Hub	-0.410	Not Balanced, >0.05	0	Balanced, <0.05
5	Median HH Inc	0.370	Not Balanced, >0.05	0	Balanced, <0.05
6	% Manuf	-0.690	Not Balanced, >0.05	0	Balanced, <0.05
7	% Speak English	-0.300	Not Balanced, >0.05	0	Balanced, <0.05
8	% Below Pov	-0.160	Not Balanced, >0.05	0	Balanced, <0.05
9	% White	0.250	Not Balanced, >0.05	0	Balanced, <0.05
10	LFPR	-0.100	Not Balanced, >0.05	0	Balanced, <0.05
11	Unemp Rate	0.130	Not Balanced, >0.05	0	Balanced, <0.05
12	%Rural	-0.770	Not Balanced, >0.05	0	Balanced, <0.05

while also obtaining enough observations for statistical inference:

- % 65 and older
- % with bachelor's degree

Covid Economics 3, 10 April 2020: 54-84

Table 5: Main results using CBPS instead of optweights

	Dependent variable:					
	bsc.bin	bsc.cont	mtc.bin	mtc.cont	wgt.bin	wgt.cont
	(1)	(2)	(3)	(4)	(5)	(6)
Treatbin	−0.361		−0.600**		−0.496**	
	(0.238)		(0.258)		(0.225)	
Treatcont		−0.102***		−0.300***		−0.106***
		(0.013)		(0.038)		(0.011)
Tot pop	0.036	0.032	0.131**	0.109*	0.017	0.013
	(0.030)	(0.028)	(0.065)	(0.063)	(0.026)	(0.021)
Old age dep ratio	0.024	0.017	0.047	0.055	0.065*	0.047
	(0.046)	(0.046)	(0.087)	(0.088)	(0.035)	(0.030)
Bachelor's degree	−0.071	−0.066	0.156	0.194*	0.012	0.049
	(0.054)	(0.055)	(0.114)	(0.106)	(0.078)	(0.078)
Female HH no husband	0.380***	0.383***	0.346***	0.329***	0.354***	0.317***
	(0.053)	(0.054)	(0.115)	(0.096)	(0.081)	(0.075)
Md inc HH	0.084	0.087	−0.136	−0.121	−0.096	−0.081
	(0.061)	(0.061)	(0.183)	(0.173)	(0.087)	(0.082)
Manufacturing	0.125***	0.129***	0.080	0.085	−0.011	0.002
	(0.048)	(0.048)	(0.077)	(0.064)	(0.068)	(0.055)
Speak only english	−0.309***	−0.306***	−0.195**	−0.190**	−0.211***	−0.215***
	(0.050)	(0.050)	(0.097)	(0.088)	(0.066)	(0.065)
Below poverty level	0.021	0.020	−0.106	−0.110	−0.070	−0.084*
	(0.040)	(0.040)	(0.093)	(0.083)	(0.052)	(0.049)
White	−0.069	−0.067	0.012	−0.006	−0.111*	−0.130**
	(0.064)	(0.064)	(0.091)	(0.084)	(0.059)	(0.063)
LFPR	−0.073**	−0.072**	−0.163	−0.122	−0.008	−0.023
	(0.036)	(0.036)	(0.143)	(0.158)	(0.067)	(0.070)
Unem rate	0.040	0.043	0.147*	0.150**	0.138**	0.165***
	(0.040)	(0.040)	(0.077)	(0.073)	(0.057)	(0.060)
Rural	0.017	0.027	−0.136	−0.085	−0.153***	−0.131*
	(0.038)	(0.039)	(0.091)	(0.090)	(0.057)	(0.069)
Constant			0.300***	0.000	−0.107	−0.285*
			(0.107)	(0.124)	(0.092)	(0.163)
Observations	1,657	1,657	234	234	1,657	1,657
R^2	0.420	0.424	0.389	0.388	0.406	0.418

Note: *p<0.1; **p<0.05; ***p<0.01

Covid Economics 3, 10 April 2020: 54-84

Table 6: Balance results for CBPS

	Covs	Diff_Unm	Bal_Test_Unm	Diff_Match	Bal_Test_Match
1	CTY_tot_pop	0.320	Not Balanced, >0.05	0	Balanced, <0.05
2	CTY_Old_age_dep_ratio	0.240	Not Balanced, >0.05	0.110	Not Balanced, >0.05
3	CTY_Bachelor_s_degree	0.370	Not Balanced, >0.05	0.010	Balanced, <0.05
4	CTY_Female_hher_no_husbandhh	-0.410	Not Balanced, >0.05	-0.050	Balanced, <0.05
5	CTY_Md_inc_hhs	0.370	Not Balanced, >0.05	0.010	Balanced, <0.05
6	CTY_Manufactur	-0.690	Not Balanced, >0.05	-0.020	Balanced, <0.05
7	CTY_Speak_only_English	-0.300	Not Balanced, >0.05	0.010	Balanced, <0.05
8	CTY_Below_poverty_level_AGE_18_64	-0.160	Not Balanced, >0.05	-0.030	Balanced, <0.05
9	CTY_White	0.250	Not Balanced, >0.05	0.010	Balanced, <0.05
10	CTY_Labor_Force_Part_Rate_pop_16_over	-0.100	Not Balanced, >0.05	-0.070	Not Balanced, >0.05
11	CTY_Unem_rate_pop_16_over	0.140	Not Balanced, >0.05	0	Balanced, <0.05
12	CTY_POPPCT_RURAL	-0.770	Not Balanced, >0.05	0	Balanced, <0.05

- Median household income
- % speak only English
- County unemployment rate
- % White

These choices reduce the number of total observations to 152 but yield substantively and statistically similar results to our main findings, as illustrated in Table 7. The balance test results are visualized in Figure 9.

With counties nested within DMAs and states, our data facilitate multilevel models as an alternative to standard linear regression analyses, as well as allowing for more rigorous fixed effects at the DMA or state level. Per our PAP we implement both in examining the results of March 17th in Figure 10 which summarizes the impact of different fixed effects / mixed effects, producing noisier but still negative estimates for most checks.

Diff-in-Diff Over Time

We summarize the descriptive difference-in-differences for March 1st, March 3rd, and March 10th in the plots below. As illustrated, prior to Super Tuesday, counties that would become exposed were more supportive of Sanders than counties that would remain in control over the period of analysis (although the general shift was still away from Sanders relative to 2016). In addition, the March 17th primary voters in exposed counties also shifted less away from Sanders compared to the control counties.

Covid Economics 3, 10 April 2020: 54-84

Table 7: Main Results Estimated with CEM instead of nearest neighbor matching

	Dependent variable:					
	bsc.bin	bsc.cont	mtc.bin	mtc.cont	wgt.bin	wgt.cont
	(1)	(2)	(3)	(4)	(5)	(6)
Treatbin	−0.392		−0.924**		−0.531**	
	(0.262)		(0.408)		(0.252)	
Treatcont		−0.107***		−0.257***		−0.114***
		(0.014)		(0.029)		(0.011)
Bachelor's degree	−0.178***	−0.181***	−0.080	−0.118	−0.031	−0.011
	(0.042)	(0.041)	(0.100)	(0.114)	(0.065)	(0.061)
Md inc HH	−0.018	−0.015	−0.123	−0.064	−0.069	−0.044
	(0.049)	(0.049)	(0.093)	(0.118)	(0.066)	(0.070)
Speak only english	−0.301***	−0.293***	−0.071	−0.090	−0.329***	−0.320***
	(0.051)	(0.051)	(0.091)	(0.097)	(0.063)	(0.059)
White	−0.292***	−0.290***	−0.230**	−0.243**	−0.313***	−0.309***
	(0.059)	(0.059)	(0.104)	(0.105)	(0.055)	(0.055)
Unem rate	0.117**	0.118**	0.181**	0.171**	0.164***	0.180***
	(0.049)	(0.048)	(0.072)	(0.075)	(0.060)	(0.061)
Old age dep ratio	−0.042	−0.048	−0.092	−0.075	0.009	0.013
	(0.042)	(0.041)	(0.078)	(0.076)	(0.044)	(0.031)
Constant			0.175	−0.009	−0.011	−0.213
			(0.118)	(0.128)	(0.077)	(0.156)
Observations	1,657	1,657	219	219	1,657	1,657
R^2	0.363	0.366	0.315	0.258	0.373	0.386
Note:						*p<0.1; **p<0.05; ***p<0.01

Covid Economics 3, 10 April 2020: 54-84

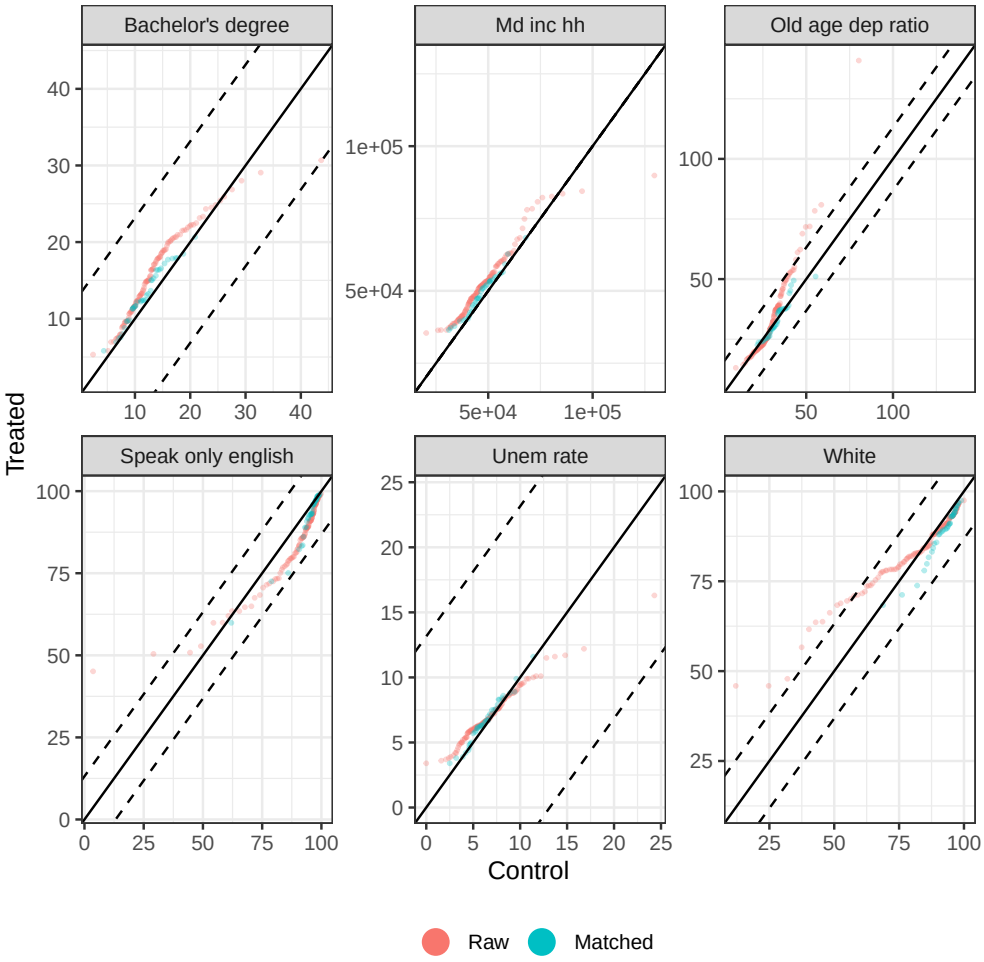


Figure 9: Balance performance across 6 covariates using CEM (Blackwell et al., 2009).

Geographic Units of Treatment

Our main findings define treatment as a function of the local media market in which a county resides. The intuition is that, as the virus was initially spreading, local media was more likely to report on the virus when cases appeared in their market. However, by March 17th news about the virus was a constant fixture on national stations, suggesting that the DMA would no longer be an appropriate border by which to define exposure to the fear and uncertainty generated by the outbreak. To the extent that larger units grew more salient as

Covid Economics 3, 10 April 2020: 54-84

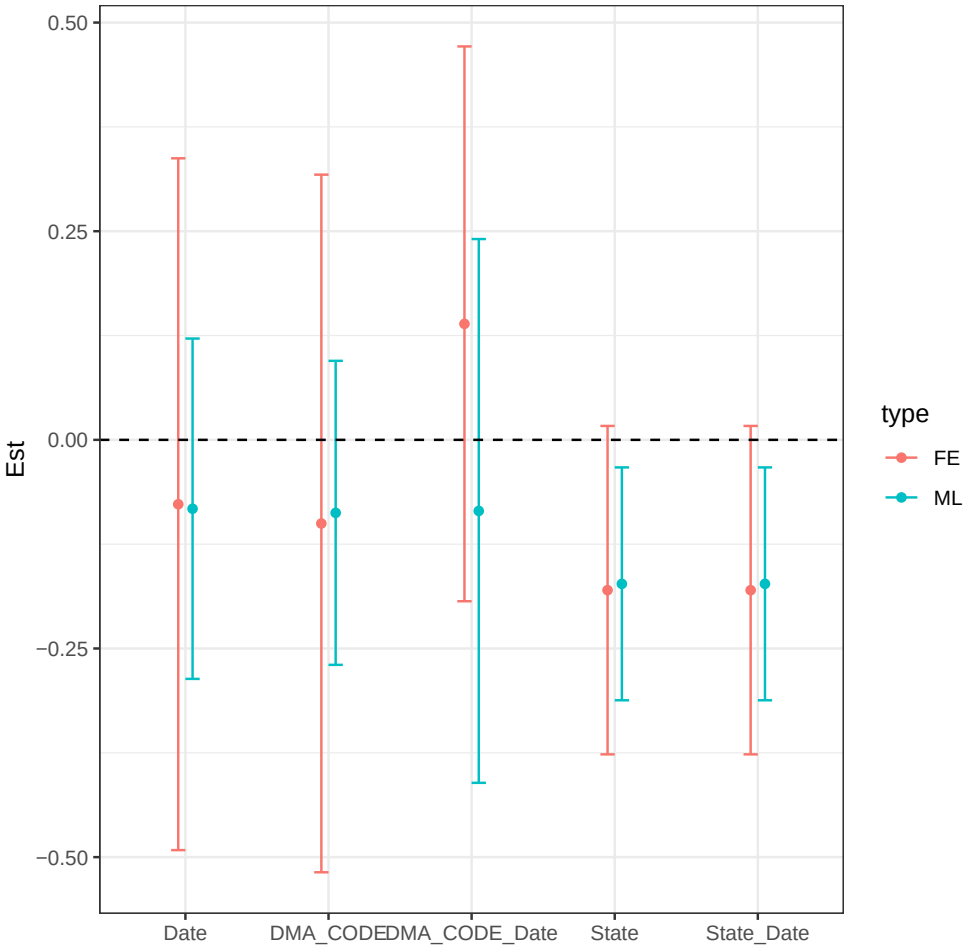


Figure 10: Estimates subject to different choices of fixed (red) and mixed (blue) effects.

overall coverage of the outbreak increased, we compare the estimates generated by defining the virus at the county, the DMA, and the state in Figure 12. As illustrated, aggregating at smaller geographic units produces null to positive results when defining the outbreak as starting after March 10th. Conversely, the negative findings persist when defining treatment assignment at the state level, although this unit appears too large for the earlier days of the outbreak when local news sources would be more appropriate for transmitting information and uncertainty.

Covid Economics 3, 10 April 2020: 54-84

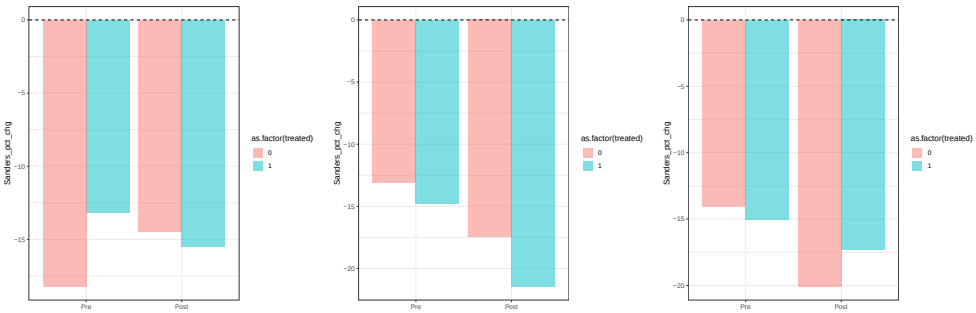


Figure 11: Descriptive DID data for outbreak dates of March 1st, March 3rd, and March 10th

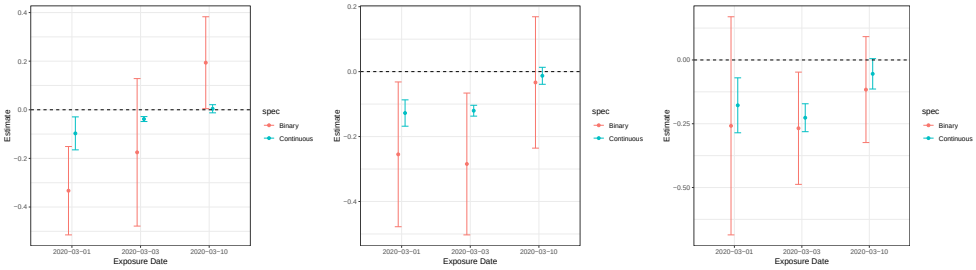


Figure 12: Matching estimates for impact of exposure on the change in Sanders' vote share when treatment is defined at the level of the county (left plot), the DMA (center plot), or at the state (right plot).

Deaths

The main results use confirmed cases of the virus to define treatment. However, we also have data on deaths due to COVID-19. We reproduce our main table, replacing confirmed cases with deaths, and find even stronger results, as illustrated in Table 8.

The associated descriptive statistics are presented in Figure 13.

Covid Economics 3, 10 April 2020: 54-84

Table 8: Relationship between Sanders vote share and COVID-19 Deaths

	Dependent variable: Δ Sanders Vote Share 2020-2016								
	Basic		Matched		Weighted		Diff-in-Diff		
	Bin	Cont	Bin	Cont	Bin	Cont	Basic	Match	Weight
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
treatBin	−0.922***		−0.995***		−0.922***				
	(0.191)		(0.250)		(0.185)				
treatCont		−0.124***		−0.424***		−0.124***			
		(0.010)		(0.078)		(0.011)			
treatGroup							−0.113	0.424	−0.030
							(0.092)	(0.344)	(0.094)
post							−0.203*	−0.116	−0.161
							(0.116)	(0.305)	(0.136)
Tot Pop	0.021	0.023	0.097	0.089	0.001	0.001	0.028	0.106	0.008
	(0.025)	(0.025)	(0.145)	(0.143)	(0.024)	(0.025)	(0.028)	(0.148)	(0.028)
Old Age Dep	0.024	0.018	0.111	0.228	0.058	0.057	0.023	0.109	0.063*
	(0.049)	(0.049)	(0.264)	(0.200)	(0.037)	(0.037)	(0.050)	(0.264)	(0.037)
Bach Deg	−0.065	−0.063	0.125	0.066	−0.042	−0.040	−0.070	0.147	−0.053
	(0.051)	(0.051)	(0.227)	(0.217)	(0.061)	(0.061)	(0.053)	(0.240)	(0.065)
Female HH	0.369***	0.383***	0.176	0.216*	0.321***	0.319***	0.366***	0.173	0.303***
	(0.054)	(0.053)	(0.157)	(0.126)	(0.083)	(0.083)	(0.054)	(0.161)	(0.089)
Med HH Inc	0.079	0.084	−0.172	0.060	0.053	0.058	0.082	−0.223	0.054
	(0.065)	(0.065)	(0.281)	(0.239)	(0.062)	(0.061)	(0.064)	(0.297)	(0.062)
% Manuf	0.120**	0.127**	0.009	0.085	0.089*	0.092*	0.121**	0.018	0.094*
	(0.050)	(0.050)	(0.113)	(0.097)	(0.054)	(0.053)	(0.050)	(0.114)	(0.052)
Speak English	−0.321***	−0.320***	−0.273*	−0.223	−0.416***	−0.415***	−0.301***	−0.272*	−0.398***
	(0.043)	(0.043)	(0.139)	(0.152)	(0.055)	(0.055)	(0.048)	(0.143)	(0.064)
Below Pov	0.005	−0.002	−0.010	0.054	0.041	0.039	0.011	−0.032	0.042
	(0.040)	(0.041)	(0.222)	(0.187)	(0.045)	(0.046)	(0.039)	(0.227)	(0.047)
% White	−0.116*	−0.113*	0.078	0.097	−0.083	−0.084	−0.096	0.073	−0.071
	(0.066)	(0.066)	(0.172)	(0.175)	(0.101)	(0.101)	(0.064)	(0.175)	(0.100)
LFPR	−0.079*	−0.074*	−0.113	−0.022	−0.063	−0.064	−0.082**	−0.099	−0.066
	(0.043)	(0.043)	(0.221)	(0.200)	(0.049)	(0.049)	(0.041)	(0.217)	(0.049)
Unemp Rate	0.036	0.038	0.049	0.077	0.048	0.051	0.054	0.057	0.066
	(0.043)	(0.042)	(0.065)	(0.058)	(0.047)	(0.047)	(0.041)	(0.066)	(0.047)
% Rural	0.022	0.034	−0.147	−0.110	−0.006	−0.007	0.011	−0.159	−0.030
	(0.035)	(0.035)	(0.121)	(0.124)	(0.049)	(0.049)	(0.037)	(0.124)	(0.056)
trtGrp:post							−0.677***	−1.326***	−0.791***
							(0.210)	(0.415)	(0.219)
Constant	0.027	0.000	0.498***	0.000	−0.059	−0.086	0.091	0.521**	−0.015
	(0.063)	(0.062)	(0.155)	(0.170)	(0.085)	(0.086)	(0.087)	(0.240)	(0.100)
Observations	1,657	1,657	96	96	1,657	1,657	1,657	96	1,657
R^2	0.406	0.398	0.394	0.306	0.349	0.338	0.414	0.400	0.355

Note: *p<0.1; **p<0.05; ***p<0.01

Covid Economics 3, 10 April 2020: 54-84

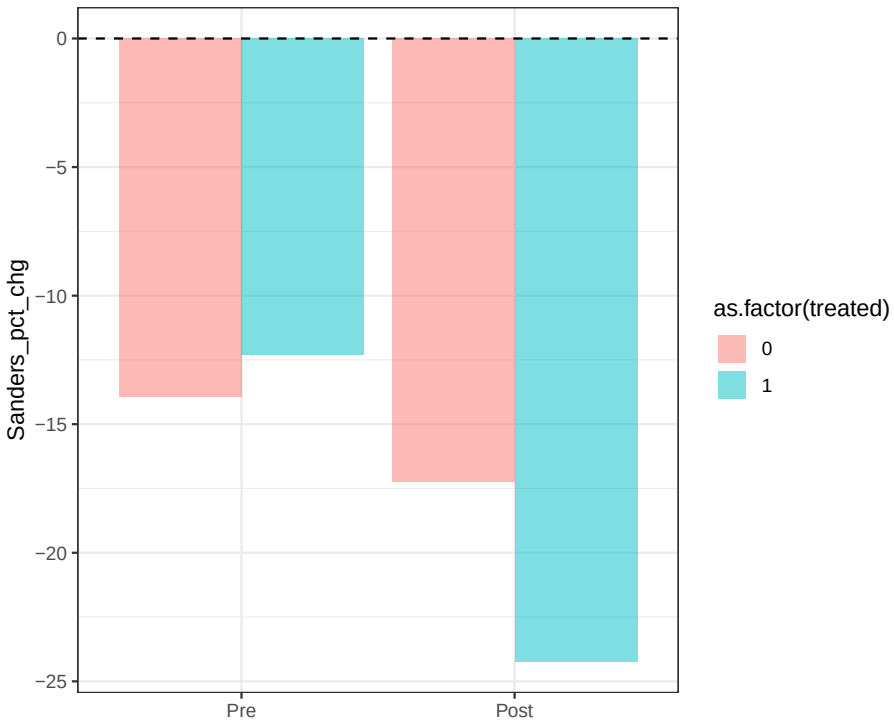


Figure 13: Change in support for Sanders (y-axis) between 2016 and 2020 by exposure to COVID-19 deaths in the DMA prior to and following the outbreak (dated to after March 3rd, x-axis).

CEPR PRESS

Covid Economics **Issue 3, 10 April 2020**

Sectoral effects of social distancing

Jean-Noël Barrot,[1] Basile Grassi,[2] and Julien Sauvagnat[3]

Date submitted: 4 April 2020; **Date accepted:** 5 April 2020

The health crisis caused by the outbreak of the Covid-19 virus has led many countries to implement drastic measures of social distancing. By reducing the quantity of labour, social distancing in turn leads to a drop in output which is difficult to quantify without taking into account relationships between sectors. Starting from a standard model of production networks, we analyse the sectoral effects of the shock in the case of France. We estimate that six weeks of social distancing brings GDP down by 5.6%. Apart from sectors directly impacted by social distancing measures, those whose value-added decreases the most are upstream sectors, i.e. sectors most distant from final demand. The same exercise is carried out for other European countries, taking into account national differences in sectoral composition and propensity to telework. Finally, we analyse the economic impact of phasing out social distancing by sector, region or age group.

Covid Economics 3, 10 April 2020: 85-102

1 Associate Professor, HEC Paris.
2 Assistant Professor, Bocconi University.
3 Associate Professor, Bocconi University.

CEPR PRESS

Sectoral effects of social distancing

Jean-Noël Barrot, Basile Grassi, and Julien Sauvagnat [1]

Work in progress

The health crisis caused by the outbreak of the Covid-19 virus leads many countries to implement drastic measures of social distancing. By reducing the quantity of labor, social distancing in turn leads to a drop in output which is difficult to quantify without taking into account relationships between sectors. Starting from a standard model of production networks, we analyze the sectoral effects of the shock in the case of France. We estimate that six weeks of social distancing brings GDP down by 5.6%. Apart from sectors directly concerned by social distancing measures, those whose value added decreases the most are upstream sectors, i.e. sectors most distant from final demand. The same exercise is carried out for other European countries, taking into account national differences in sectoral composition and propensity to telework. Finally, we analyze the economic impact of phasing out social distancing by sector, region or age group.

Covid Economics 3, 10 April 2020: 85-102

Modern economies are characterized by the many interdependencies formed by companies in their production processes. These interdependencies are well identified in the literature as facilitating the propagation of non-systemic shocks (Barrot and Sauvagnat, 2016) and their aggregation (Acemoglu et al., 2012; Baqaee and Farhi, 2019), with applications for public policies (Grassi and Sauvagnat, 2019). For a recent review, see Carvalho and Tahbaz-Salehi (2018). Analyzing the effect of the social distancing rules implemented to curb the spread of the Covid-19 virus requires to estimate its effects on the active workforce, and to measure its impact across the production network [2].

Effect of social distancing on the workforce

Administrative closings The decree of March 14, 2020 prohibits certain categories of establishments from opening to the public [3]. Exceptions are granted by the decree of March 15, 2020, and relate in particular to the food and basic necessities trade. To estimate the reduction in active workforce due to administrative closings in each sector, we proceed as follows. Starting from the finest sector classification, the NAF rev. 2 in 732 sector classification, we identify the sectors corresponding to the decree of March 14, 2020, for which

1. Jean-Noël Barrot is affiliated with HEC Paris (barrot@hec.fr). Basile Grassi (basile.grassi@unibocconi.it) and Julien Sauvagnat (julien.sauvagnat@unibocconi.it) are affiliated with Bocconi.

2. In what follows, we abstract from economic policy initiatives implemented in response to the crisis.

3. Hearing rooms, conferences, meetings, shows or for multiple use ; Sales stores and Shopping centers, except for their delivery and order picking activities ; Restaurants and drinking places, except for their takeaway delivery and sales activities, room service in hotel restaurants and bars and contract catering ; Dance halls and play rooms ; Libraries, documentation centers ; Exhibition halls ; Covered sports establishments ; Museums ; Marquees, tents and structures ; Outdoor establishments ; Educational, educational, training establishments, holiday centers, leisure centers without accommodation.

Covid Economics 3, 10 April 2020: 85-102

we consider that the active workforce is zero. By aggregation, using the number of workers by sector in the 2016 census data available on the INSEE website, we obtain the share of the inactive workforce for each of the 38 sectors of the aggregated NAF rev. 2 classification. The share of the total workforce affected by administrative closings stands at 10.9% and is concentrated in directly affected sectors : hotel and restaurants, arts and leisure, wholesale and retail, social work.

Closures of nurseries, schools, secondary schools and high schools In addition, all nurseries, schools, colleges and high schools were closed from March 16, 2020, in accordance with the decree of March 14, 2020. To estimate the effects on the workforce in each sector, we use data from the census to identify, in each of the 38 sectors of NAF rev. 2, the share of working people with dependent children under 16 and therefore forced into inactivity [4]. The share of the total workforce affected by childcare caused by the closings of nurseries, secondary schools and high schools stands at 13.2% [5], and varies, if we leave aside the sectors concerned by administrative closings and the health sector, from 11.7% (Agriculture) to 19.4% (Pharmaceuticals).

Confinement On the other hand, to prevent the spread of the Covid-19 virus, traffic restrictions are imposed, as well as the strict compliance with a safety distance of one meter between each individual. So that these rules do not lead to the shut down of business, the Minister of Labor asked firms to facilitate remote work as much as possible (telework), and urged companies to bring together their Social and Economic Committee (CSE) to adapt working conditions to health guidelines. In the absence of better data, the share of the active population in each sector likely to continue working at home is estimated using data from the European Community survey on the use of ICT and electronic commerce in businesses carried out by INSEE for Eurostat on a sample of 12,500 companies. This provides, for each sector, the share of employees of companies with more than 10 employees using a portable device provided by the company, connected to the Internet via the network of mobile phones (laptop, smartphone, tablet, etc.) in 2019. This stands at 32%, which is consistent with some recent telework surveys [6]. However, confinement should lead companies to increase their use of telework. The ICT survey also provides the share of employees of companies with more than 10 employees using a computer (including a portable device) with internet access for professional use (fixed or mobile connection), which averages 62%. We note that this indicator is significantly correlated (correlation = 0.5) with the share of employees in telework in 2017 estimated by DARES [7]. Some sectors being excluded from the survey (agriculture, financial services, public administration), the missing variable is imputed by applying the average ratio between the survey variable and the share of employees in telework estimated by DARES. In

4. More specifically, we consider that an active person has dependent children if there is not another inactive person in the household, who could take care of them. If there are several active adults in the household, we consider the drop in activity to be evenly distributed among these adults. We exclude from this calculation those who are forced to inactivity because of administrative closings, and health workers whose children are taken care of in the school system.

5. Sadique et al (2008) obtain a similar proportion based on English data.

6. See for example the 2020 Telework study by Malakoff Humanis, March 2020

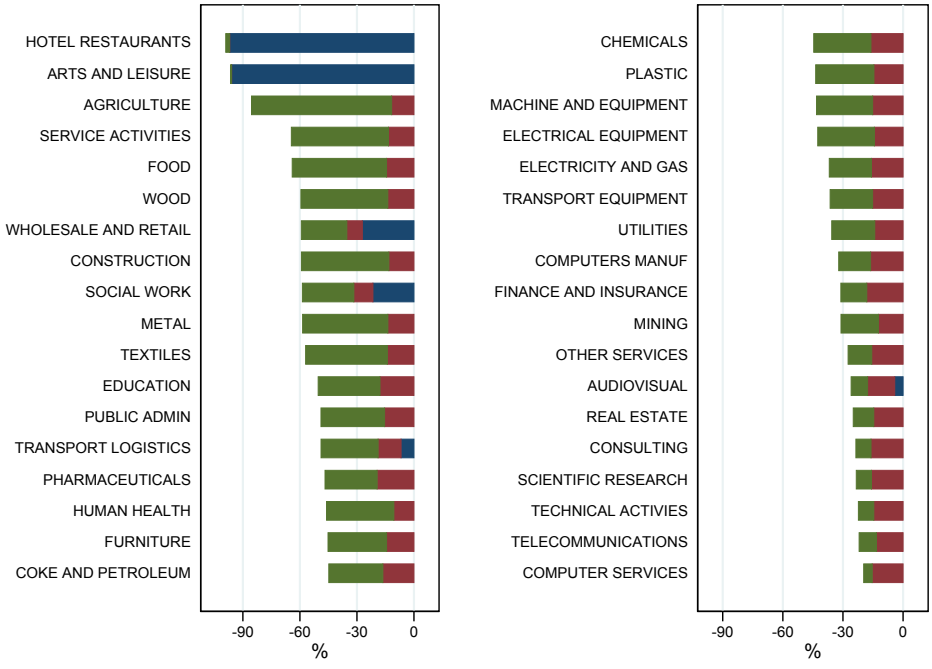
7. DARES Analyses, November 2019, Number 051

Covid Economics 3, 10 April 2020: 85-102

the limit case of absolute confinement, only these employees could continue to work, either because they can work at home, or because it is probably easier to reorganize their work environment in accordance with social distancing rules.

Figure 1
Decrease in active workforce caused by social distancing measures

This figure shows the effects of social distancing measures on the workforce by sector (in %). Blue bars represent the decline in the active workforce due to administrative closings. Red bars, the additional effect linked to school closings. Green bars, the residual effect related to confinement.



Cumulative effect By combining the effect of administrative closings, that of childcare imposed by the closings of nurseries, schools, colleges and high schools, and that of strict confinement allowing only people usually working with a computer to continue to do so, we obtain an overall drop in the active workforce of 52% [8]. The detail by sector is presented in Figure 1. The effect is broken down according to the origin of the shock. In blue, the effect of administrative closings ; in red, the additional effect of school closings ; and in green, the residual effect of confinement. Unsurprisingly, "arts and leisure" and "hotel restaurants" are

8. This number is remarkably close to estimates by Google of the drop in workplace mobility, -56% relative to baseline in France, see COVID-19 Community Mobility Report as of March 29, 2020

Covid Economics 3, 10 April 2020: 85-102

the hardest hit, due to administrative closings. Next comes "agriculture" or "business services", where the share of the workforce who does not work on a computer is high. Conversely, "technical activities ", "telecommunications " or "computer services " are relatively spared.

Description of the production network

Companies, and consequently the sectors of the French economy, are linked to each other through the network of customer-supplier relationships. There is no data to trace these business-to-business relationships. We therefore rely on the input-output table produced by INSEE, which describes and synthesizes transactions in goods and services in product and branch of activity. Figure 2 shows the structure of the French production network according to the 38 branches of activity of the NAF rev. 2 for 2015. [9]

In the first panel, each column represents the production of a sector. Each line represents the intermediate consumption of a sector, i.e. the inputs of its production process. In short, the column sectors are the suppliers, the row sectors are the customers. Each box in the table gives the intensity of use by a sector (on the column) of the input (on the row) in its production process. The darker the blue, the more quantitatively important the input. We thus verify that for the row (client sector) "hotel restaurants", the column (supplier sector) "food" is quantitatively important. We note that certain supplier sectors, in columns, contribute significant inputs from a large number of sectors. These are "business services", " consulting", and " wholesale and retail " activities.

In the second panel of Figure 2 are represented the links between sectors. Each point represents a sector, and its size is proportional to the total volume of its inputs. Each line represents a relation between a supplier sector and a client sector, and its width is proportional to the share of the input in the total of the inputs of the client. This graph highlights the chains of links : thus, the "agriculture" sector is an important input of the "food" sector, itself an important input of the "hotel restaurants" sector.

Finally, we report in Table 3 two key network statistics, Bonacich-Katz centrality and upstreamness, for each sector of the economy. The centrality of a sector measures the importance as a supplier to the economy, whereas the upstreamness measures the number of nodes between a given sector and the final demand.

Description of the model

To analyze the effect of social distancing on GDP and on the value added of each sector, we construct a standard model of production networks. Each sector produces a good by using labor and intermediate consumption produced by the other sectors, and by choosing the quantities so as to maximize its profit. Households consume goods produced by each sector [10], and provide a fixed amount of labor to each sector so as to maximize their utility.

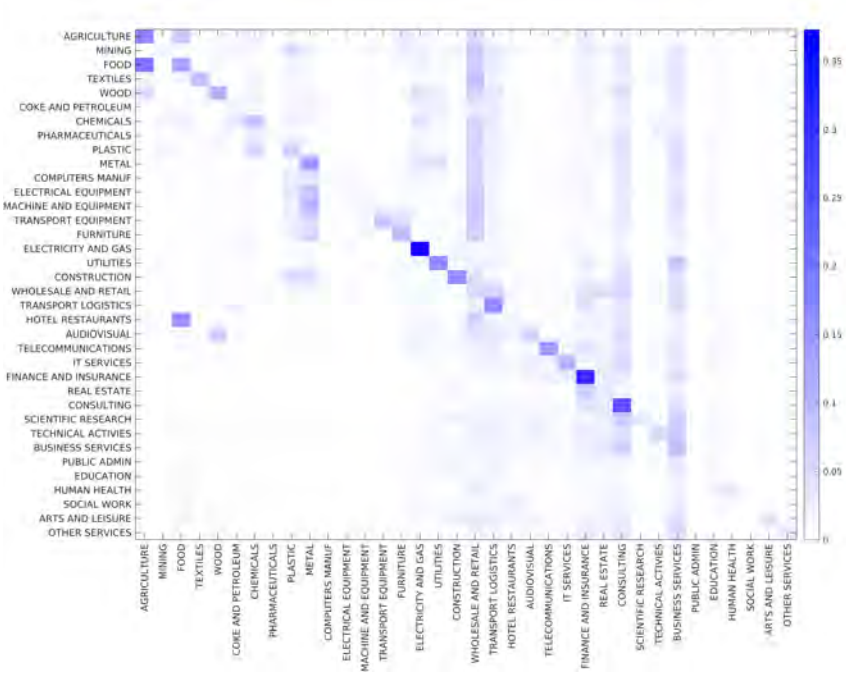
The economy's response to the shock depends on two parameters. The elasticity of substitution between goods drives the responses of household consumption to changes in relative prices. If the elasticity is greater than 1, an increase in the price of a given good leads to

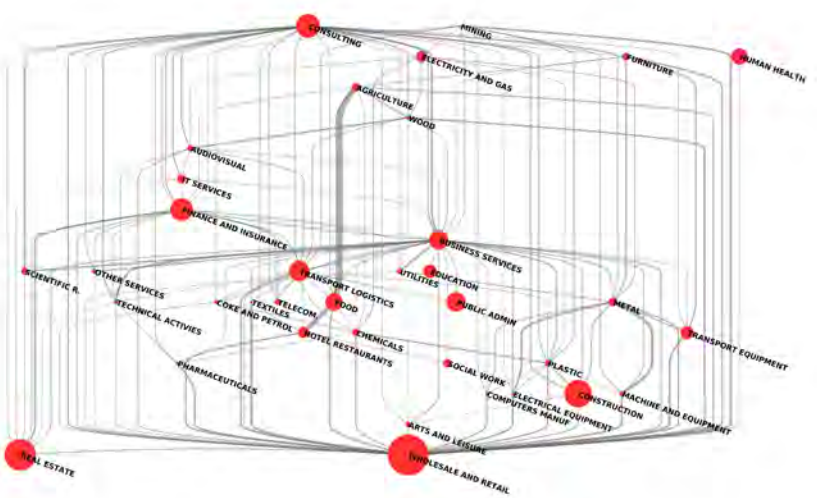
9. We report the full description of each branch in Table 2.

10. In such a closed economy model, household consumption includes public spending, investment and exports

Figure 2
The French production network

This figure shows the structure of the French production network according to the 38 branches of activity of the NAF rev. 2. In the first panel, each column represents the production of a sector. Each line represents the intermediate consumption of a sector. In the second panel, each point represents a sector, and its size is proportional to the total volume of its inputs. Each line represents a relation between a supplier sector and a client sector, and its width is proportional to the share of the input in the total of the inputs of the customer.





Covid Economics 3, 10 April 2020: 85-102

Covid Economics 3, 10 April 2020: 85-102

a decrease in its share in the household consumption basket, and vice versa. Similarly, the elasticity of substitution between intermediates describes the response of sectors to a change in the relative prices of the inputs they use. The higher it is, the more a sector can substitute inputs between them. This elasticity is lower when the horizon is short - it can be difficult to quickly substitute inputs between them -, and the level of aggregation is high - it is easier for a company to change supplier, than for a sector to do without an upstream sector. We rely on the literature to calibrate the elasticity of substitution between final goods at 3, and the elasticity of substitution between intermediate inputs at 0.5. We check that the results are robust to alternative values.

The model is useful for estimating the effect of the supply shock linked to social distancing, and its propagation throughout the production network. It does not take into account international trade [11]. Economic shocks affecting foreign countries with domestic repercussions are not quantified here. Furthermore, the model does not integrate the effect of automatic stabilizers and economic support policies announced in response to the crisis, such as the extension of partial unemployment, the suspension of contributions and tax charges, or the solidarity fund for the self-employed and very small businesses in France. The model also does not integrate the effects of possible demand shocks caused by the health crisis : increased demand for medical and surgical equipment, or the consumption of digital services. The model also ignores possible changes in the structure of consumption (or preference parameters in the utility function) of households linked to the consequences of the outbreak of the virus. Finally, it ignores the amplification effects linked to potential business bankruptcies, the destruction of customer-supplier relationships, and more generally the destruction of companies' relational or organizational capital.

Effect of social distancing on GDP

Table 1
Effect of 6 weeks of social distancing on GDP

	Administrative closings	Administrative closings + School closings	Administrative closings + School closings + Confinement
GDP growth	-0.9%	-2.5%	-5.6%

The model makes it possible to estimate the effect of social distancing on the value added

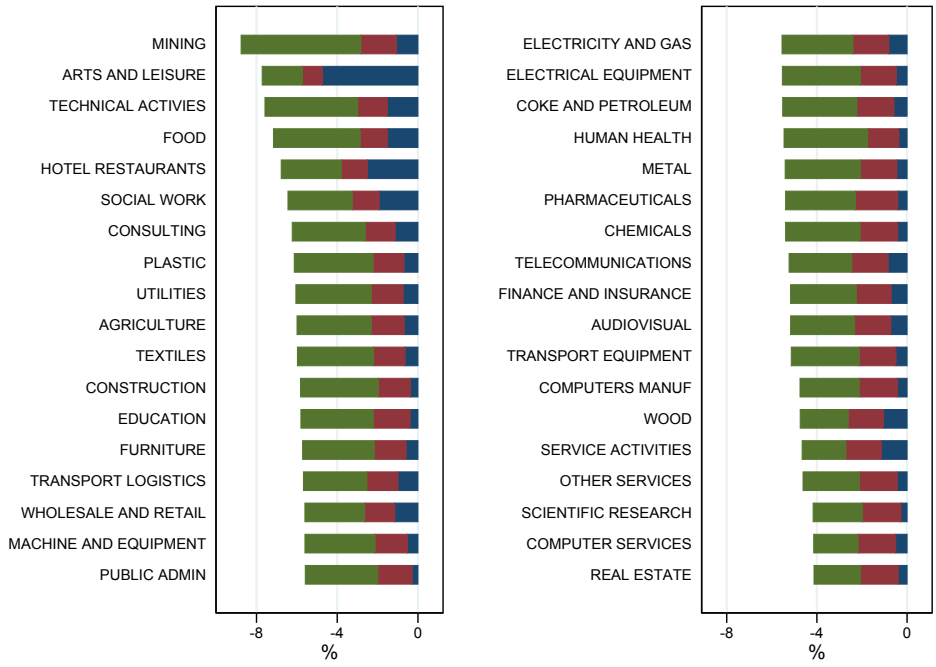
11. The effects of the shock in China on the French economy is studied, for example, by Gerschel et al. (2020).

Covid Economics 3, 10 April 2020: 85-102

of each sector, the weighted sum of which forms GDP. The results are presented for a period of 6 weeks. The fall in annual GDP is -5.6% [12] . We decompose this figure according to the origin of the shock, and present the result in Table 1. Administrative closings cause a decrease of 0.9%. When we add the closings of nurseries, colleges and high schools, the drop is 2.5%. The residual difference, ie 3.1 percentage points, is explained by confinement.

Figure 3
Value added growth for 6 weeks of social distancing (%)

This figure shows the effects of social distancing on value added growth for each sector (in %). Blue bars represent the decline in the active workforce due to administrative closings. Red bars, the additional effect linked to school closings. Green bars, the residual effect linked to confinement.



The model makes it possible to estimate the impact of the shock separately for each sector. Figure 3 shows the effect of six weeks of social distancing on annual value added growth in each sector. The effect is broken down according to the origin of the shock. In blue, the effect of administrative closings ; in red, the additional effect of school closings ; and in green, the residual effect of confinement. The drop varies from -8.8% to -4.1% depending on

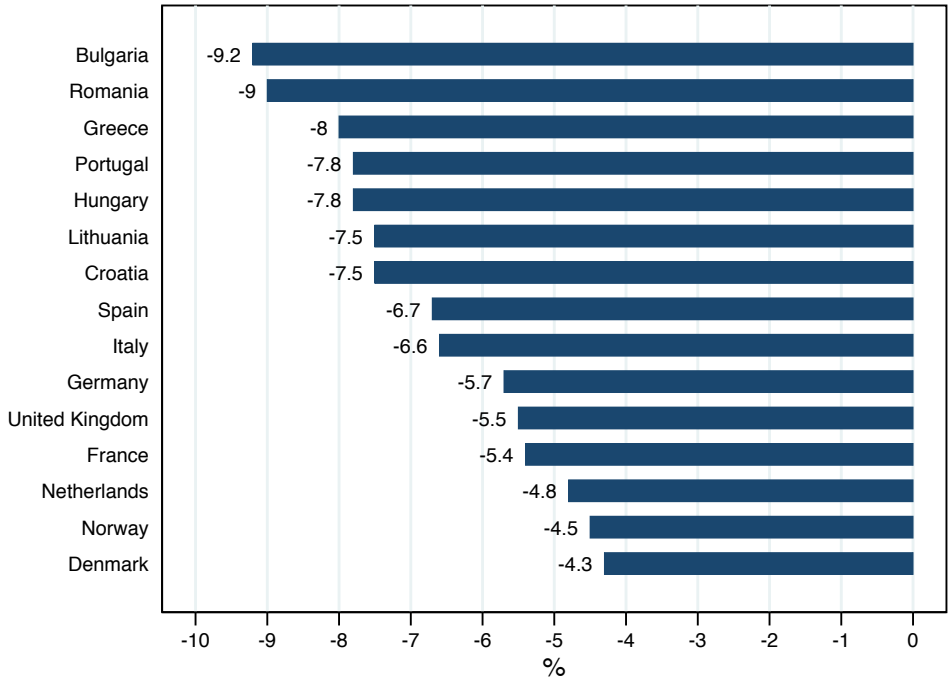
12. This estimate is higher than that presented by INSEE in its Conjoncture Point of March 26, 2020 from contemporary shock data, which finds 3% for a month of social distancing, and to that presented by the OFCE in its Policy Brief of March 30, 2020. This difference can be explained by the fact that the model does not take into account automatic stabilizers and support policies.

the sector. Among the sectors most affected, are some of those directly impacted by social distancing measures, such as "arts and leisure" (-7.7%) and "hotel restaurants" (-6.8%). However, among the most affected sectors are also upstream ones, i.e. those most distant from final demand, such as "mining" (-8.8%), and "technical activities" (-7.6%), "consulting" (-6.2%) or "utilities" (-6.0%). Thus, if the downstream sectors seem more directly disturbed by administrative closings in terms of active workforce, upstream sectors suffer most significantly in terms of value added.

Extension to other European countries

Figure 4
GDP drop by country for 6 weeks of social distancing

This figure shows the effects of social distancing on the GDP growth of European countries (in %). We assume that all countries apply the same restrictions, and that social distancing is in place for 6 weeks in each country. Only the sectoral composition and the propensity to telework vary.



We next analyze the effect that social distancing would have on other European countries, taking into account only national differences in sectoral composition, and in the telework propensity. It is therefore assumed that all countries take the same decisions on administrative

Covid Economics 3, 10 April 2020: 85-102

closings, closings of schools and confinement [13]. The structure of the production network in 54 branches is obtained from the "World Input-Output" database (WIOD, version 2016), which provides for each country the 2014 input-output table. The propensity of each sector in each country to telework comes from the community survey on the use of ICT and electronic commerce in businesses described above.

The results are presented in Figure 4. The GDP drops on average by 6.6% in the sample, for six weeks of social distancing [14]. The fall in GDP ranges from 4.3% (Denmark) to 9.2% (Bulgaria). These differences are partly explained by the sectoral composition, and partly by the propensity to telework, as shown in Figure 8 which shows the correlation [15] between propensity to telework and decline in GDP.

Progressive phasing out of social distancing

Phasing out of social distancing is anticipated to be implemented progressively. The model allows us to predict the marginal effect on GDP that phasing out would have on each sector, region or age group, taken in isolation. We consider each sector (or region, or age group) one after the other assuming that social distancing is lifted after 4 weeks instead of 6, and we measure the effect on marginal on GDP. We then normalize the implied GDP in euros by the number of released workers, which gives an approximation of the marginal benefit per worker of phasing out social distancing. The results are presented in Figure 5-7. The effect on GDP (Panel A) varies by a factor of 4 across sectors and across regions, but is relatively stable by age group. The marginal effect per worker (Panel B) is stable by region and age group, but varies very strongly by sector.

These results must be interpreted with caution, and within the limits of the model's assumptions. They correspond to the effect of the decline in the workforce linked to social distancing and do not take into account international trade or public policies undertaken to support the economy. They describe the supply side response at the sectoral level and make it possible to identify the most affected sectors. They do not in any way challenge the importance of social distancing, which is well identified by the medical literature as an effective means of slowing down the epidemic propagation, whose human, social and economic costs are considerable.

13. All countries in the sample closed their schools, but not all of them imposed administrative closings and confinement.

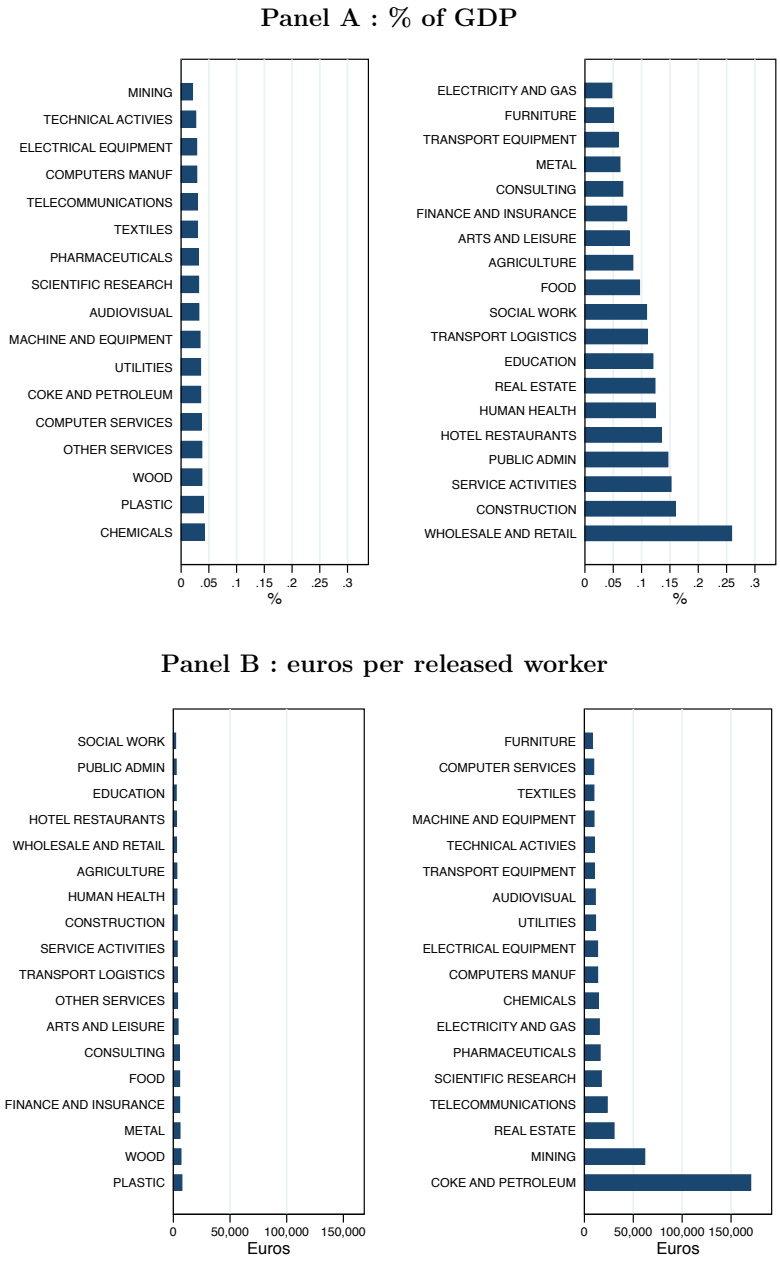
14. We note that France undergoes a drop in GDP of 5.4%, very close to the 5.6% estimated above from 38 sectors, instead of 54.

15. A 10 percentage points increase in telework propensity increases GDP by 1%.

Covid Economics 3, 10 April 2020: 85-102

Figure 5
Effect on GDP of differentiated phasing out by sector

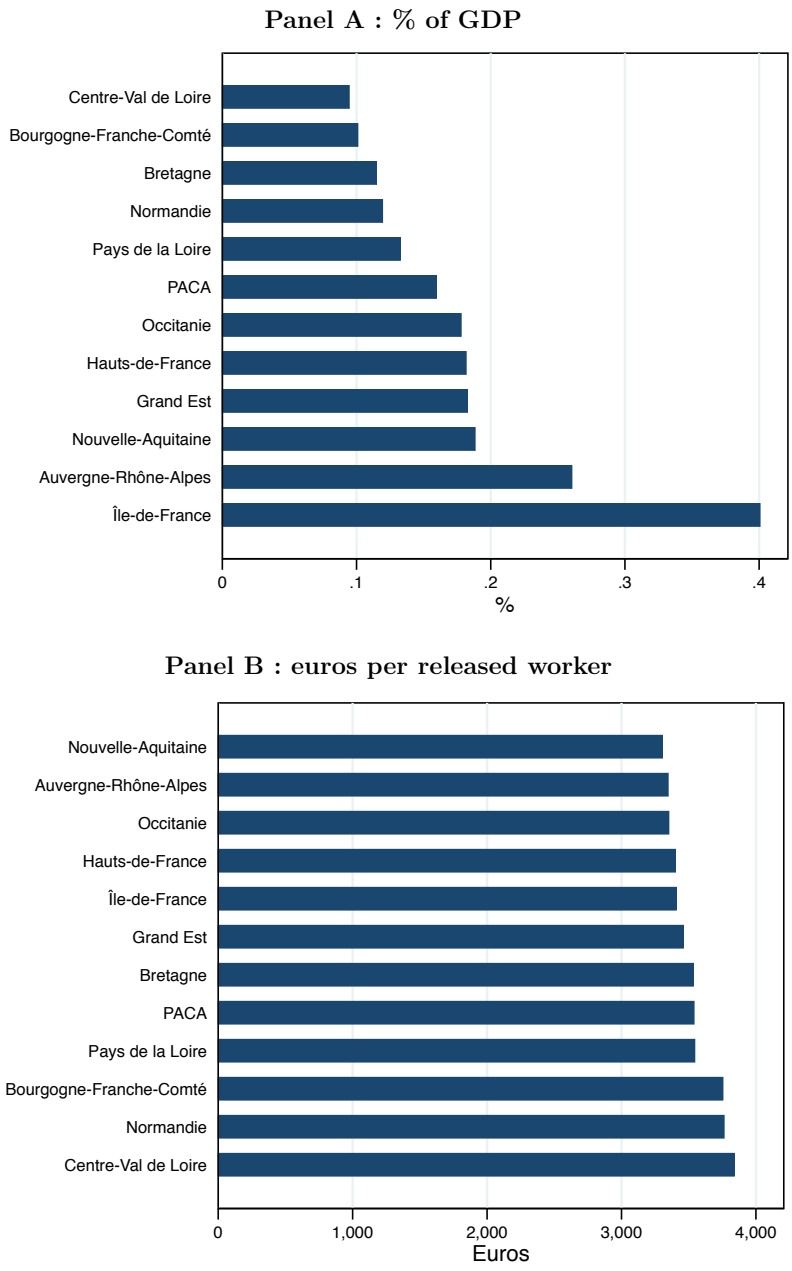
This figure shows the effects on GDP of phasing out social distancing in each sector individually, after 4 weeks instead of 6.



Covid Economics 3, 10 April 2020: 85-102

Figure 6
Effect on GDP of differentiated phasing out by region

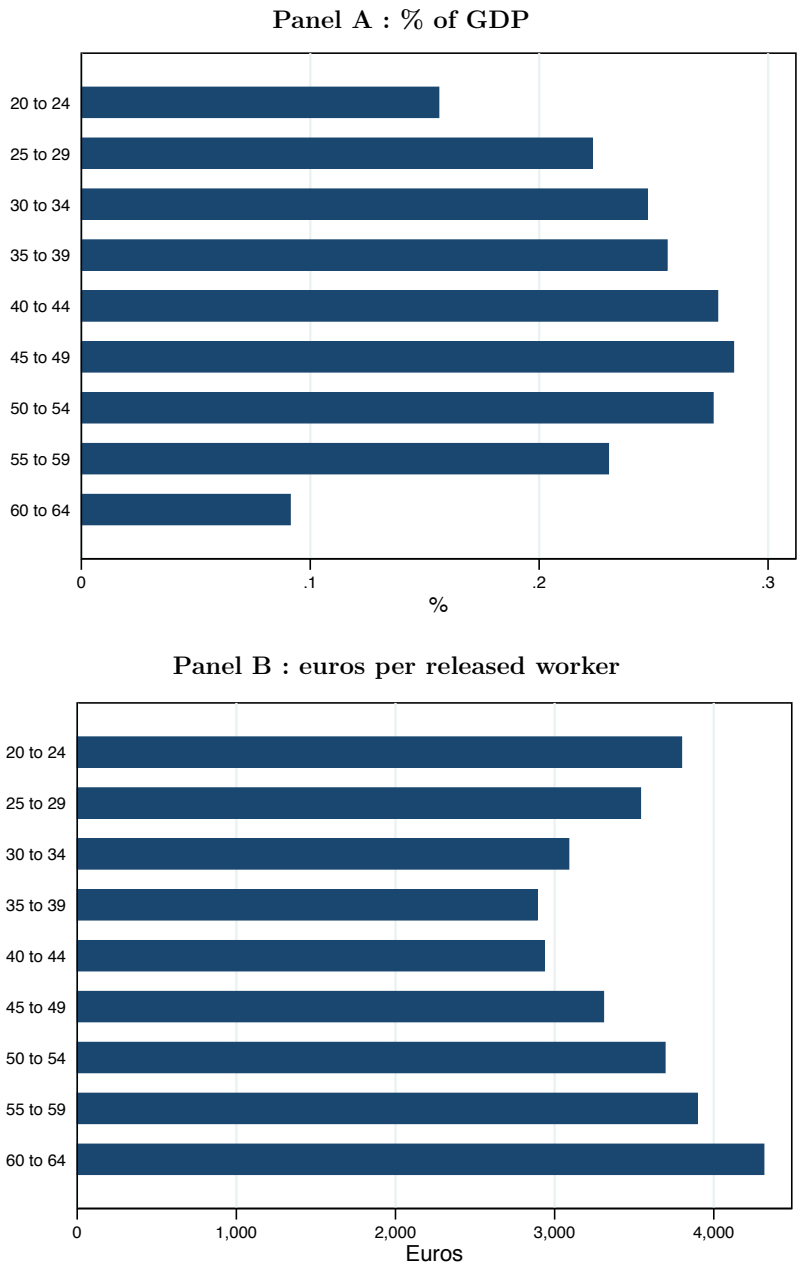
This figure shows the effects on GDP of phasing out social distancing in each region individually, after 4 weeks instead of 6.



Covid Economics 3, 10 April 2020: 85-102

Figure 7
Effect on GDP of differentiated phasing out by age group

This figure shows the effects on GDP of of phasing out social distancing in age group individually, after 4 weeks instead of 6.



Covid Economics 3, 10 April 2020: 85-102

Références

Acemoglu, D., V. M. Carvalho, A. Ozdaglar, and A. Tahbaz-Salehi (2012, 09). The network origins of aggregate fluctuations. *Econometrica 80*(5), 1977–2016.

Baqaee, D. R. and E. Farhi (2019). Networks, barriers, and trade. *Working Paper*.

Barrot, J.-N. and J. Sauvagnat (2016). Input Specificity and the Propagation of Idiosyncratic Shocks in Production Networks. *The Quarterly Journal of Economics 131*(3), 1543–1592.

Carvalho, V. M. and A. Tahbaz-Salehi (2018). Production networks : A primer. *Annual Review of Economics*.

Gerschel, E., A. Martine, and I. Mejen (2020, 3). Propagation of shocks in global value chains : the coronavirus case. *IPP Policy brief* (53).

Grassi, B. and J. Sauvagnat (2019, 12). Production networks and economic policy. *Oxford Review of Economic Policy 35*(4), 638–677.

Covid Economics 3, 10 April 2020: 85-102

Appendix

Covid Economics 3, 10 April 2020: 85-102

Covid Economics 3, 10 April 2020: 85-102

Table 2
List of Sectors

CODE A38	ACRONYM A38	DESCRIPTION A38
AZ	AGRICULTURE	agriculture, forestry and fishing
BZ	MINING	mining and quarrying
CA	FOOD	manufacture of food products, beverages and tobacco products
CB	TEXTILES	manufacture of textiles, wearing apparel and leather products
CC	WOOD	manufacture of wood and paper products, and printing
CD	COKE AND PETROLEUM	manufacture of coke and refined petroleum products
CE	CHEMICALS	manufacture of chemicals and chemical products
CF	PHARMACEUTICALS	manufacture of basic pharmaceutical products and pharmaceutical preparations
CG	PLASTIC	manufacture of rubber and plastics products, and other non-metallic mineral products
CH	METAL	manufacture of basic metals and fabricated metal products, except machinery and equipment
CI	COMPUTERS MANUF	manufacture of computer, electronic and optical products
CJ	ELECTRICAL EQUIPMENT	manufacture of electrical equipment
CK	MACHINE AND EQUIPMENT	manufacture of machinery and equipment n.e.c.
CL	TRANSPORT EQUIPMENT	manufacture of transport equipment
CM	FURNITURE	manufacture of furniture ; other manufacturing ; repair and installation of machinery and equipment
DZ	ELECTRICITY AND GAS	electricity, gas, steam and air conditioning supply
EZ	UTILITIES	water supply ; sewerage, waste management and remediation activities
FZ	CONSTRUCTION	construction
GZ	WHOLESALE AND RETAIL	wholesale and retail trade, repair of motor vehicles and motorcycles
HZ	TRANSPORT LOGISTICS	transportation and storage
IZ	HOTEL RESTAURANTS	accommodation and food service activities
JA	AUDIOVISUAL	publishing, audiovisual and broadcasting activities
JB	TELECOMMUNICATIONS	telecommunications
JC	IT SERVICES	computer programming, consultancy and related activities ; information service activities
KZ	FINANCE AND INSURANCE	financial and insurance activities
LZ	REAL ESTATE	real estate activities
MA	CONSULTING	legal and accounting activities ; activities of head offices ; management consultancy activities ; architecture and engineering activities ; technical testing and analysis
MB	SCIENTIFIC RESEARCH	scientific research and development
MC	TECHNICAL ACTIVITIES	advertising and market research ; other professional, scientific and technical activities ; veterinary activities
NZ	BUSINESS SERVICES	administrative and support service activities
OZ	PUBLIC ADMIN	public administration and defence ; compulsory social security
PZ	EDUCATION	education
QA	HUMAN HEALTH	human health activities
QB	SOCIAL WORK	social work activities
RZ	ARTS AND LEISURE	arts, entertainment and recreation
SZ	OTHER SERVICES	other service activities

Covid Economics 3, 10 April 2020: 85-102

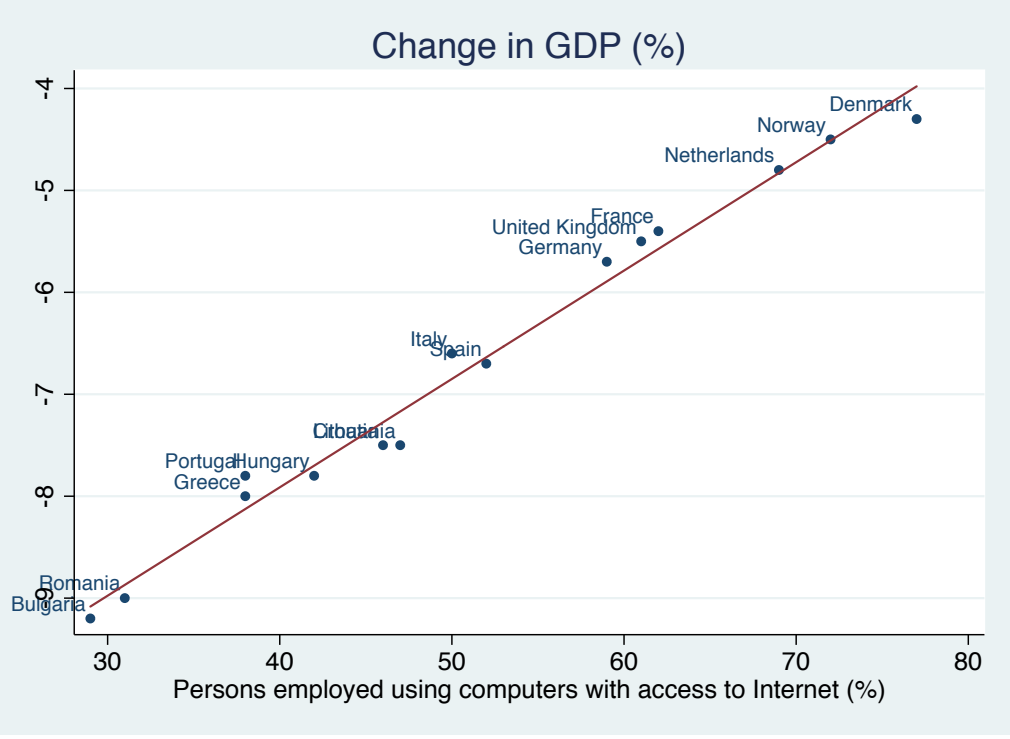
Table 3
Sector characteristics and network statistics

(1)	(2)	(3)	(4)	(5)
Secteur	Sector Characteristics		Network Statistics	
	Final Demand	Employment	Upstreamness	Network Centrality
AGRICULTURE	2.7%	1.1%	2.09	0.032
MINING	0.1%	0.0%	2.43	0.002
FOOD	2.3%	4.6%	1.52	0.070
TEXTILES	0.4%	0.5%	1.43	0.007
WOOD	0.8%	0.4%	2.27	0.015
COKE AND PETROLEUM	0.0%	0.9%	1.69	0.015
CHEMICALS	0.5%	1.9%	1.47	0.027
PHARMACEUTICALS	0.3%	1.0%	1.04	0.011
PLASTIC	1.0%	0.7%	2.04	0.021
METAL	1.5%	1.3%	1.91	0.033
COMPUTERS MANUF	0.5%	0.9%	1.09	0.010
ELECTRICAL EQUIPMENT	0.4%	0.7%	1.28	0.008
MACHINE AND EQUIPMENT	0.7%	1.3%	1.25	0.015
TRANSPORT EQUIPMENT	1.3%	4.6%	1.14	0.052
FURNITURE	1.1%	1.7%	1.59	0.029
ELECTRICITY AND GAS	0.7%	1.4%	2.36	0.043
UTILITIES	0.7%	0.6%	2.17	0.020
CONSTRUCTION	6.5%	8.4%	1.35	0.112
WHOLESALE AND RETAIL	12.8%	11.8%	1.46	0.167
TRANSPORT AND LOGISTICS	5.1%	3.6%	1.92	0.080
HOTEL RESTAURANTS	4.0%	3.0%	1.48	0.042
AUDIOVISUAL	0.9%	1.3%	1.75	0.022
TELECOMMUNICATIONS	0.5%	1.0%	2.05	0.022
IT SERVICES	1.6%	2.2%	1.61	0.034
FINANCE AND INSURANCE	3.5%	2.7%	2.36	0.091
REAL ESTATE	1.4%	10.1%	1.37	0.128
CONSULTING	4.4%	2.5%	2.40	0.097
SCIENTIFIC RESEARCH	0.7%	2.4%	1.03	0.025
TECHNICAL ACTIVITIES	1.0%	0.3%	2.33	0.014
BUSINESS SERVICES	5.7%	2.0%	2.28	0.077
PUBLIC ADMIN	9.8%	7.8%	1.00	0.078
EDUCATION	7.7%	4.4%	1.26	0.053
HUMAN HEALTH	7.1%	6.2%	1.06	0.065
SOCIAL WORK	7.6%	3.3%	1.00	0.033
ARTS AND LEISURE	1.6%	1.7%	1.20	0.020
OTHER SERVICES	3.0%	1.3%	1.40	0.017
Moyenne	2.8%	2.8%	1.64	0.044
Min	0.0%	0.0%	1.00	0.002
Max	12.8%	11.8%	2.43	0.167

Covid Economics 3, 10 April 2020: 85-102

Figure 8
Cross-country correlation between telework and GDP change

This graph shows the correlation between the share of persons employed using computers with access to Internet and the change in GDP (in %) across 15 European countries.



CEPR PRESS

Covid Economics **Issue 3, 10 April 2020**

Measuring the economic risk of Covid-19

Ilan Noy,[1] Nguyen Doan,[2] Benno Ferrarini[3] and Donghyun Park[4]

Date submitted: 4 April 2020; **Date accepted:** 5 April 2020

We measure the economic risk of Covid-19 at a geo-spatially detailed resolution. In addition to data about the current prevalence of confirmed cases, we use data from 2014-2018 to compute measures for exposure, vulnerability, and resilience of the local economy to the shock of the epidemic. Using a battery of proxies for these four concepts, we calculate the hazard and the principal components of exposure and vulnerability to it, and of the economy's resilience (i.e., its ability of the recover rapidly from the shock). We find that the economic risk of this pandemic is particularly high in most of Africa, the Indian subcontinent, the Persian Gulf, and Southeast Asia. These results are consistent when comparing an ad hoc equal weighting algorithm for the four components of the index, an algorithm that assumes equal hazard for all countries, and one based on an estimated weights using previous aggregated disability-adjusted life years losses associated with communicable diseases.

1 Chair in the Economics of Disasters and Climate Change, Victoria University of Wellington.
2 PhD student, Victoria University of Wellington.
3 Principal Economist, Asian Development Bank.
4 Principal Economist, Asian Development Bank.

Covid Economics 3, 10 April 2020: 103-118

Covid Economics 3, 10 April 2020: 103-118

Introduction

The economic risk of an epidemic, any epidemic, is very distinct from the health (morbidity and mortality) risk. The basic framework that assesses disaster risk is typically constructed around four concepts – hazard, exposure, vulnerability and resilience – and it is the interaction of these that leads to the economic consequences. The hazard in these frameworks is the natural trigger – in the present circumstances, it is the SARS-Cov-2 virus which causes Covid-19. Since the economic risk is determined not only by the hazard but also by the exposure, vulnerability and resilience, this risk has plausibly very different spatial variability than the spread of the virus.

Even with no significant case load or mortality associated with it, the epidemic can lead to very adverse changes within and outside an affected economy that can lead to dramatic economic effects. Given the paucity of data on epidemic cases in the recent past (the period for which comprehensive economic and demographic records are available) and the unprecedented nature of this event, our aim here is not to precisely measure the likely consequence of this pandemic, but rather to evaluate comparatively where the economic risk of Covid-19 is currently concentrated using several alternative algorithms.

Our risk measure is premised on the observation that a disaster, including an epidemic, occurs when a hazard (in this case, the disease) interacts with an exposed population that is vulnerable to this hazard, thus causing harm to people. Epidemics always arise out of a natural pathogen (very often zoonotic) but the pathogen by itself does not create the epidemic, and definitely not its economic consequences. For that, the pathogen must encounter a society, people and an economy that are both exposed and vulnerable to it. Resilience, in this framework, is conceptualised and measured as the ability of the economy to bounce back given the magnitude of the shock (generated by the intersection of the hazard, exposure, and vulnerability).[1] The degree of resilience in a system (in this case, the economy) is thus determined by the speed in which the recovery process occurs and when the system reverts back to its pre-shock level (i.e., when full recovery is achieved).

As defined by the UN Office for Disaster Risk Reduction, a disaster is "a serious disruption of the functioning of a community or a society at any scale due to hazardous events interacting with conditions of exposure, vulnerability and capacity, leading to one or more of the following: human, material, economic and environmental losses and impacts. The effect of the disaster can be immediate and localised, but is often widespread and could last for a long period of time" (UNDRR, 2017).

Exposure in the UNDRR definition refers to the population and the economic activity that is located in areas that are being exposed to the pathogen or that is indirectly exposed to the changing behaviour that is induced by the presence of this pathogen (e.g., Epstein, 2009). Vulnerability, in this case, refers to the ability of the pathogen to adversely affect the exposed

[1] An alternative term that is sometimes used similarly to 'resilience' is 'capacity' [for recovery]. We prefer the term 'resilience' (or 'socio-economic resilience') as it is defined by Hallegatte (2014) and the World Bank (2018).

economy. A higher degree of vulnerability will lead to a more adverse outcome for the economy, given the same exposure to the SARS-Cov-2 virus.[2]

Over time, the economic losses will depend on the depth of the shock and on the economy's resilience (its ability to bounce back). A more resilient economy, in this framework, is one that manages to minimise the post-shock cumulative loss of income during the recovery process for a given size of the shock (Hallegatte, 2014).[3] As Prager et al. (2017) note, resilience policies are often not really plausible to pursue during the rapid phase of the spread of the epidemic. What is more plausible is to make up for lost production once the epidemic has abated, and potentially prepare the economy for the recovery period while the epidemic is still ongoing (as many governments are trying to do now for Covid-19).[4] The ability to implement such policies, as determined by both financial and institutional capacity, is therefore an important determinant of economic resilience.

In a previous paper (Noy et al., 2019), we analysed the economic risk of a generic epidemic. Here, instead of focusing on a generic emerging infectious disease event, we focus on Covid-19 (Figure 1). SARS-Cov-2 fits perfectly the pattern of a zoonotic pathogen emerging from the interaction of a wild animal population with a food market that epidemiologists have been warning about (e.g., Allen et al., 2017). However, the economic characteristics of this unprecedented event are somewhat different (for example, the total collapse of the international tourism is unique) and we therefore modified our risk analysis to fit the new experience with Covid-19.

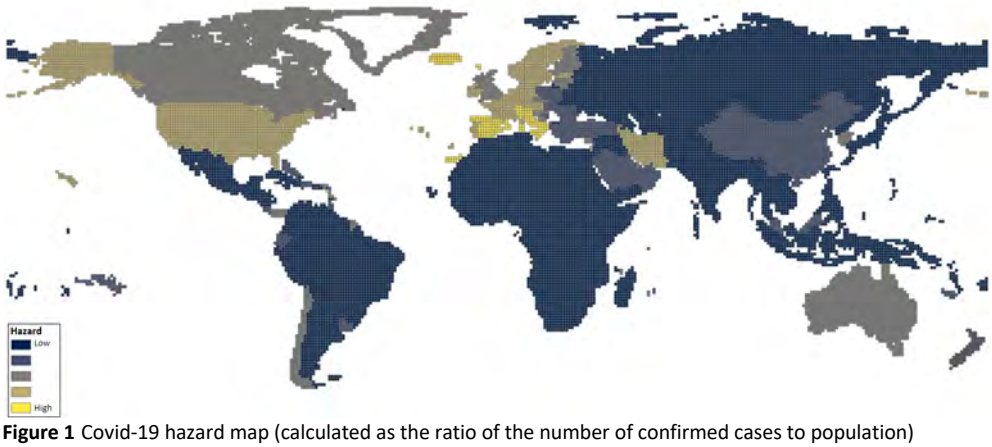


Figure 1 Covid-19 hazard map (calculated as the ratio of the number of confirmed cases to population)
Note: Data updated on 30 March 2020.

[2] These distinctions are always imperfect, and that is also the case for epidemics. Even the basic epidemiological parameter, R_0, may be a function of the socio-economic environment, as the contact rate, the probability of transmission upon contact, and the duration of infection are all also determined by social factors (e.g., poverty or nutrition); see Janes et al. (2012).
[3] Prager et al. (2017), in their modelling of influenza in the United States use a somewhat broader definition of resilience. They define 'economic resilience' as the capacity "to maintain functionality and dampen business interruption losses in the aftermath of a disaster" (p. 6).
[4] See: https://www.imf.org/en/Topics/imf-and-covid19/Policy-Responses-to-Covid-19.

Covid Economics 3, 10 April 2020: 103-118

Measured at the level of grid cells, *g*, we model the risk associated with the economic impact of epidemics as a linear combination of hazard plus a local economy's exposure and vulnerability to it, minus its resilience or ability to bounce back:

$$\widehat{Risk}_g = \alpha_1 Hazard_g + \alpha_2 Exposure_g + \alpha_3 Vulnerability_g - \alpha_4 Resilience_g \quad (1)$$

We collect a large group of sub-national and national measures from recent years (2014-2018) to proxy for exposure, vulnerability, and economic resilience. The selection of variables is based on the literature measuring disaster risk, as reviewed in Yonson and Noy (2018), and on the current experience of Covid-19. We then use principal component analysis (PCA) to compute a standardised index for each exposure, vulnerability, and resilience. Using the first component of exposure, vulnerability, and resilience index, in addition to the number of confirmed cases of Covid-19, we compute a risk index in relation to the economic risk of epidemics. In our simplest specifications, we assume $\alpha_i = \alpha_j$ for all *i* and *j*; in alternative algorithms, we assume $\alpha_1 = 0$ (the hazard is equal for all countries), or we estimate the α_i based on a disability-adjusted life years (DALY)-based least-squares regression algorithm.

Results

Figure 2 shows descriptive information and principle component analysis (PCA) results of all variables we use to measure exposure, vulnerability and resilience. The principal component index is the output of linear combination of the original variables. We use the first principal component for each exposure, vulnerability and resilience index. As the first component accounts for most variation in the data and contribute the most explanation in the combining procedure. The proportion of eigenvalues indicates the explanatory importance of the factor, which are 4.0, 3.4, and 2.8 for exposure, vulnerability and resilience, respectively. Economic activities, demographic measures and infrastructure density all positively explain exposure. High-income areas with better healthcare quality (as measured by lower infant mortality, health spending, hospital infrastructure) are related to less vulnerable areas. Tourism areas and high numbers of the elder are associated with higher vulnerability. For resilience, areas with higher social and cultural disparity have a lower index. Countries with a lower ratio of government debt and higher expenditure are more resilient.

Covid Economics 3, 10 April 2020: 103-118

Covid Economics 3, 10 April 2020: 103-118

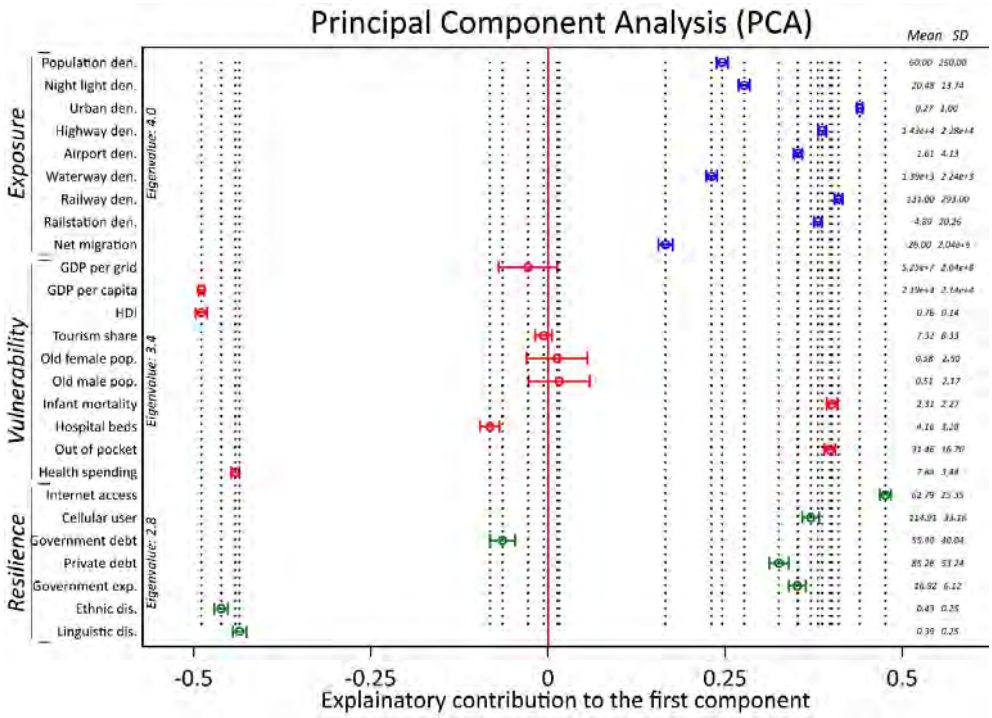


Figure 2 Descriptive data and principal component analysis (PCA) results

Note: The lower and upper caps represent standard errors of each variable in the first component.

We normalise all exposure, vulnerability, and resilience indices. Figure 3 presents the cumulative distribution of main results for hazard, exposure, vulnerability, resilience, and economic risk. For hazard, we use the number of confirmed cases of Covid-19 per 1 million people from the Worldometer website, which has the most frequently updated data. We calculate the economic risk by an equal-weight linear combination of the four indices.

Covid Economics 3, 10 April 2020: 103-118

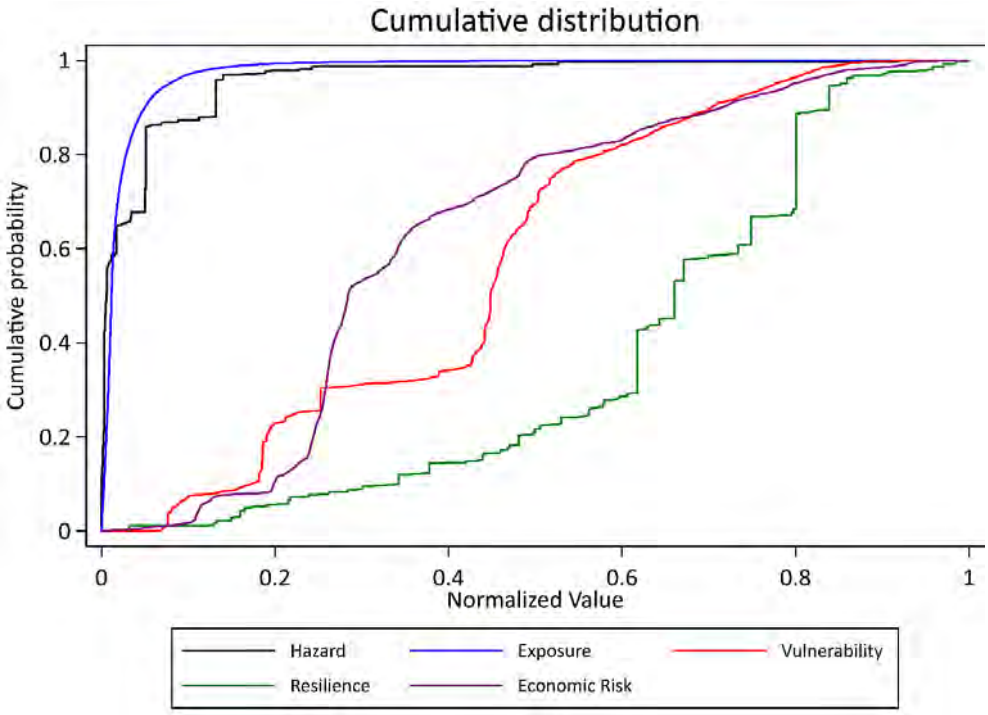


Figure 3 The cumulative distribution of indices

We find that the economic risk of epidemics is especially high in most of Latin America (except for the Southern Cone countries), most of Africa, South Asia, Iran, and much of Southeast Asia (Figure 4). Fundamentally, areas with the greatest exposure to the prevalence of Covid-19 align with high economic risk. The economic risk is high in Africa and Southeast Asia, as these are the most vulnerable areas with low income and healthcare quality. Resilience, intentionally or otherwise, plays a role in reducing the economic risk from epidemics. For example, in Southern Cone countries (Argentina and Chile) resilience is higher than in neighbouring countries due to less fractionalised socio-cultural characteristics (lower ethnic and linguistic disparity) and higher incomes. Saudi Arabia and Russia have lower economic risks because their domestic economies are focused on huge amounts of (oil) exports, and hardly rely on tourism at all.

In Figure 5, we assume that the hazard (the presence of the virus) is identical in all countries. This can be motivated either by the expectation that eventually, the spread of the virus will reach epidemic levels in all countries, or because of the widely held view that differences in the testing regimes account for a lot of the differences in the number of confirmed cases (probably especially relevant for low-income countries). As such, the assessment of the economic risk that is caused by this virus should not be based on the present known spread of the virus, but on its global potential. Besides some expected differences, however, the results presented in Figure 5 (uniform spread of the virus) and Figure 4 (hazard based on the current spread of the virus) are very similar. The only distinctive difference is that the countries of Southern Europe that have very high official infection rates (e.g., Italy and Spain)

are assessed to be a comparably lesser risk in Figure 5 (when we assume a uniform level of the hazard).

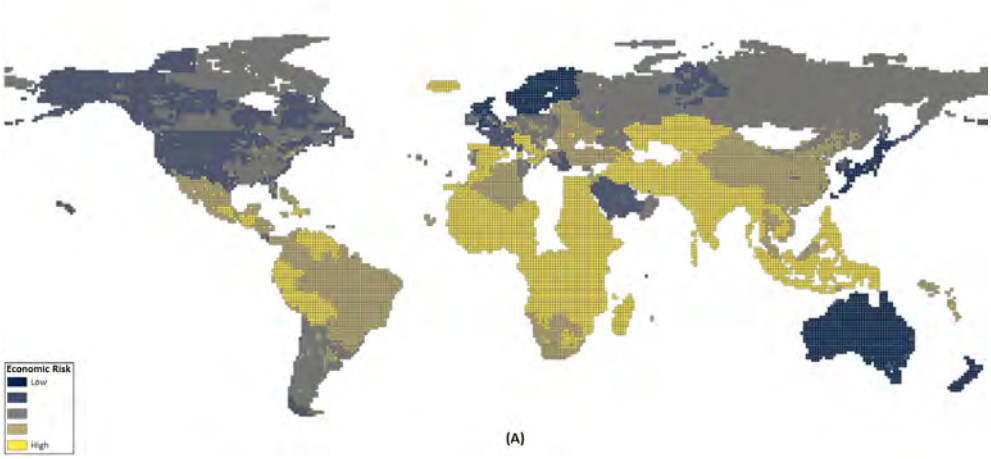


Figure 4 Economic risk of Covid-19 using equation (1) with equal weights

Notes: Observations are divided into five classes by Jenks natural breaks classification method, a data clustering method. The method optimally minimizes the average deviation from each class and maximises the deviation across classes.

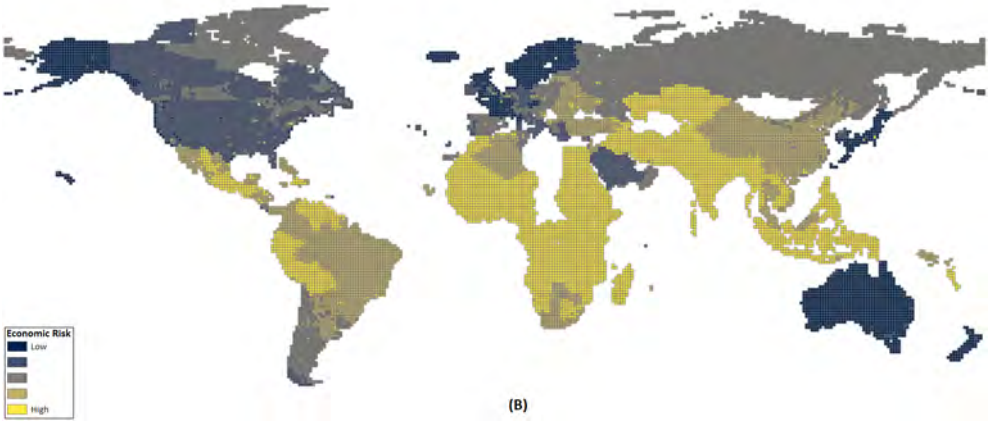


Figure 5 Economic risk of Covid-19 using a modified equation (1) with hazard calibrated so all countries have an equal hazard (all are susceptible to Covid-19)

A less ad hoc weighting scheme, instead of equal weights as in Figures 4 and 5, relies on the DALY measure of overall disease burden. Since previous DALYs associated with communicable disease is the outcome of previous events, it could be a good source for understanding the interactions between the (mostly zoonotic) hazard and exposure, vulnerability, and resilience to it. DALYs are the sum of years lost due to ill-health, disability or premature death from communicable diseases. Weights for each of the four components are derived by ordinary

Covid Economics 3, 10 April 2020: 103-118

least squares regression with the country-level DALYs as the dependent variable, as in Equation 2 (we assign the same DALY value for all grid cells within each country):

$$DALY_g = \beta_0 + \beta_1 Hazard_g + \beta_2 Exposure_g + \beta_3 Vulnerability_g + \beta_4 Resilience_i + \varepsilon_g \quad (2)$$

The estimated weights and the constant are then plugged into the risk function (i.e., $\alpha_g = \beta_g$) which now places considerably more weight on exposure than on hazard, resilience and vulnerability:

$$\widehat{WRisk}_g = 0.03 + 0.02 Hazard_g + 0.62 Exposure_g + 0.20 Vulnerability_g - 0.13 Resilience_g \quad (3)$$

The spatial patterns of the DALY-weighted risk map in Figure 6 are somewhat similar to those observed in the unweighted maps (Figures 4 and 5). As before, the areas at highest risk of economic losses from epidemics remain sub-Saharan Africa and South Asia. But much of Central Asia, and South East Asia is considered less risky with this approach, as some areas, though relatively poor, are not too densely populated (unlike Central America, for example).

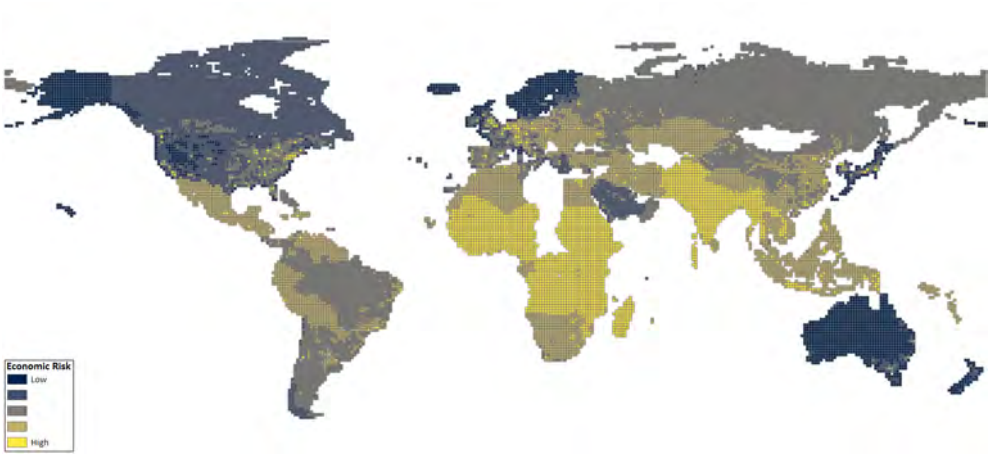


Figure 6 Economic risk of Covid-19 using the DALY-weighted index

Discussion and conclusions

We developed a measurement tool to estimate the economic risk of epidemics. With growing globalisation and inter-connectedness among far-flung populations comes increased exposure to the risk of epidemics, with potentially dire implications for the world economy; as is quite apparent with the current Covid-19 crisis.

The economic consequences of an epidemic, like any other natural hazard shock, can be delineated into damages, direct losses and indirect losses (Noy, 2016). Direct losses included lost income and output due to death and symptomatic illness as well as increased healthcare costs. If measured through the standard statistical tools used by governments to evaluate the cost of life (the value of statistical life, or VSL), the direct costs of pandemics like Covid-19 due to illness and mortality are probably much smaller than the indirect losses. This is, of course, especially true now for countries in which the epidemic has not yet spread indiscriminately,

Covid Economics 3, 10 April 2020: 103-118

but that are very exposed to the global shock it created (for example, tourism-dependent economies in the Global South).

When we account for the ways an epidemic creates economic losses, we need to measure not only the direct reductions in economic activity that are attributable to changes in government policy (e.g., mandatory lockdowns), but also behavioural changes that are caused by changing subjective judgements about the risk of contraction among the still healthy population. These behavioural changes may be influenced not only by the characteristics of the epidemic contagion process and the disease virulence, but also by its media coverage and the fear it might generate.

As public health systems have improved over the past century, this pandemic's health impacts are unlikely to be of the magnitude of the 1918-19 Influenza pandemic, though they may still be on a catastrophic scale (at the time of writing, the number of confirmed cases is reaching a million). However, what remains maybe even more salient is the pandemic's economic consequences. The exposure, vulnerability and resilience to these economic consequences were not ameliorated as much when public health systems developed throughout the last century. In contrast, globalisation of trade, increased tourism and labour flows, and the more recent advent of social media are all likely to have amplified behavioural responses and created additional vulnerabilities, thus potentially exacerbating the economic losses that will be experienced before this pandemic is over.

If the extensive behavioural reaction to the SARS crisis in 2003 could be typified as a high-prevalence-elasticity response to a disease outbreak; when the public response to an epidemic results in significant behavioural changes that increase in severity with the number of infected persons. The SARS case fits with the argument of Philipson (2000) that when private behaviour is strongly prevalence-elastic, the main economic cost of a disease outbreak is likely to arise out of preventative actions rather than directly from infections. This even more true with Covid-19, and it is true whatever the merits are of these preventative actions.

A study by Perrings et al. (2014) highlights the importance of government intervention that targets the private costs and benefits of disease avoidance so that they induce individual behavioural responses, such as social distancing, which align with the overall interests of the wider society. This concerns the trade-off that individuals make regarding their respective costs and benefits from, for example, social distancing in an epidemic situation. If the benefits of social interactions for an individual are high (e.g., they are necessary to earn the income required to meet daily subsistence costs) then this could result in continued interaction during an epidemic and, while reducing the economic impact, can potentially increase the disease reproduction rate. It can also work in the opposite direction: if the individual costs of public avoidance are very low and benefits very high, then mass public avoidance in an epidemic, where the mortality and contagiousness are not significant enough to warrant such a response, will lead to unnecessarily large economic and welfare losses. Improved understanding of the dynamics of individual trade-offs could help to prioritise public health interventions beyond what is suggested from our measure of economic risk.

To summarise, what is most apparent from our analysis is that the greatest economic risk from Covid-19 is not found in Italy or New York City, where most of the media and public

Covid Economics 3, 10 April 2020: 103-118

attention is concentrated. Rather, the highest economic risks are in countries and regions that do not get much global attention in normal times (such as sub-Saharan Africa) and get even less in the midst of the frantic reporting from the immediate frontlines of the pandemic's spread. This is unfortunate as, ultimately, the economic costs will be borne there, away from the public eye.

Appendix: Data and methods

Principal components analysis

Before going through the dimensionality reduction procedure, we standardise all variables and analyse the correlations to choose a proper set of variables. To compute a coherent index for exposure, vulnerability and resilience separately, we use principal components analysis (PCA), an algorithm to compress a large set of variables while retaining most of the information in the initial data (Ringnér, 2008). The eigenvalues, latent roots, capture the variations in the set of variables for that component. PCA is based on the eigen decomposition of positive semi-definite matrices and the singular value decomposition of rectangular matrices (Abdi and Williams, 2010).[5] Mathematically, PCA is executed on a square symmetric matrix: (i) pure sums of squares and cross products (SSCP matrix); (ii) scaled sums of squares and cross products (covariance matrix); (iii) sums of squares and cross products from standardized data (correlation matrix). The correlation matrix performs well when there are significant differences in the variances and the units of measurement of original variables.

DALY weighting method

Based on DALY data from the Institute for Health Metrics and Evaluation, we calculate the average of DALY in the period 2012-2017 from three communicable causes: (i) diaorrhea and common infectious diseases; (ii) Malaria and neglected tropical diseases; (iii) Other communicable diseases. We use this aggregate measure of DALYs lost to these infectious diseases as an alternative proxy for the risk of epidemics. One DALY equals one lost year of healthy life, either from year of life lost or year lived with a disability. Since the DALY aggregates are calculated for each country, we merge the country-level data into grid cell data. The assumption is the current health situation and an ideal health status are identical within country. We then estimate the following model by ordinary least squares (OLS):

$$DALY_g = \beta_0 + \beta_1 Hazard_g + \beta_2 Exposure_g + \beta_3 Vulnerability_g + \beta_4 Resilience_g + \varepsilon_g$$

where $Hazard_g$ is the prevalence of Covid-19 in grid g. $Exposure_g$, $Vulnerability_g$, and $Resilience_g$ is the first component of principal component analysis for exposure, vulnerability, and resilience in grid g.

Table 1 Estimation results for national DALY

Hazard	13.619**	(6.346)
Exposure	502.462***	(46.564)
Vulnerability	158.968***	(4.809)

[5] Abdi and Williams (2010) provide proof and statistical inference of PCA.

COVID ECONOMICS
VETTED AND REAL-TIME PAPERS

Covid Economics 3, 10 April 2020: 103-118

Resilience	-101.399***	(4.477)
_cons	27.727***	(3.025)
Obs.	16654	
R-squared	0.166	

Robust standard errors are in parenthesis. *** p<0.01, ** p<0.05, * p<0.1.

In Table 1, the result of three causes shows that 1 percentage point increase in the probability of exposure is positively related to 502 DALYs, while 1 percentage point increase in vulnerability is positively associated with an additional 159 DALYs points. The relation between hazard and DALY is smallest, the slope is almost 14. Whereas a 1-percentage-point increase of resilience relates to 101-DALY-point decrease. Building on these results, an alternative functional form to measuring economic risk uses the weights implied in the coefficients described in Table 1. The weights are calculated by $\hat{\beta}_j(\sum_{j=0}^{4}|\hat{\beta}_j|)^{-1}$, then:

$$\widehat{WRisk}_g = 0.03 + 0.02Hazard_g + 0.62Exposure_g + 0.20Vulnerability_g - 0.13Resilience_g$$

Hazard indicators

We use the number of Covid -19 confirmed cases per 1 million people. Data is updated on 30 March 2020.

Exposure indicators

In terms of economic exposure, we use population and nighttime light density to measure human presence and economic activity. Nightlight data is used as a proxy for economic wealth; the data is described in Román et al. (2018). Transport density provides another relevant indicator for population density. An urban metropolitan area likely has a denser network of highways and air links. To get a coherent layer of transportation density, we use all types of transport as described in Lloyd et al. (2017). Transport databases from Open Street Map (OSM) include: Highway, waterway, railway network, railway station and airport. Last, we use the number of net incoming migrants to proxy for external economic exposure. Data for each variable to proxy for exposure are collected as raster format with higher resolution than data for hazard. Hence, we can plausibly merge with data about epidemic into grid 1 degree by 1 degree by WGS84 projection.

Vulnerability indicators

Likewise, we use a set of data on economic outcomes, human development, tourism, and health quality to measure vulnerability. Drake et al. (2012) argue that the vulnerability to infectious disease outbreak is much higher in low- and middle-income countries, especially the vulnerability to mortality and morbidity risk. The United Nations' Human Development Index (HDI) and total GDP in each grid cell, are collected from the data described in Kummu et al. (2018). Kummu et al. (2018) estimate Gross Grid-Cell Product by multiplying country-

Covid Economics 3, 10 April 2020: 103-118

level GDP per capita (PPP) with 30 arc-sec population data.[6] To get sub-national data on HDI, Kummu et al. (2018) develop scaling factors to combine sub-national and national data.

Tatem et al. (2012) survey the need and availability of sub-national detailed demographic data that might be useful in understanding disease exposure and vulnerability. They argue that for improvement in our understanding of disease transmission and control, we require detailed spatially-referenced demographic data (for example, distinguished by cohorts and gender). This data is only available in low frequency in countries that conduct a comprehensive census. We lack data on health quality at the sub-national level; except for spatially-detailed data on the old population density and infant mortality rate, we use country-level measures of healthcare spending and number of hospital beds per 1000 population. These data are from the World Bank Development Indicators (WDI) and World Health Organization (WHO). We merge the country-level data into the grid cell data by assigning the same value for all grid-cells within the same country.

Resilience indicators

Hallegatte et al. (2016) argue that early warning systems possibly reduce asset losses. We assume information about epidemics is accessed via the internet and mobile phones, so we associate higher penetration rates of these with higher resilience. We use data from the WDI and the International Telecommunication Union. Next, we assume that the capacity of government to implement economic relief policy, and household to access loans are positively associated with resilience. Last, we use data about ethnic and linguistic diversity to measure socio-cultural disparity (Alesina et al., 2003). We assume that the diversity plausibly affects the behaviour of individuals and communities in a hazard event.

[6] The strategy to estimate Gross Cell Product is very similar to Nordhaus and Chen (2016), but the Kummu et al. (2018) data were updated more recently, and are available at a higher resolution.

Covid Economics 3, 10 April 2020: 103-118

Covid Economics 3, 10 April 2020: 103-118

Table 3: Details of variables

	Variable name	Description	Unit of measurement	Kind of indicators	Spatial availability	Year released/ updated	Data coverage by grid	Source
1	Covid-19	Number of confirmed cases per 1 million people	Number of people	Hazard	Country-level	30 March 2020	100%	Worldometer
2	Population density	Number of persons per square kilometre in 2015	Number of people per km^2	Exposure	Resolution: 0.5' (1 km)	2017	100%	(CIESIN, 2018)
3	Night-time lights	Night-time light intensity in 2016	Index	Exposure	Resolution: 1.5' (3 km)	2017	100%	Román et al. (2018)
4	Urban built-up	Human impact on land by urbanization activity	Index	Exposure	Resolution: 0.5' (1 km)	2014	100%	Tuanmu and Jetz (2014)
5	Transport networks in 2016	Highway density	Index	Exposure	Resolution: <1 km	2016	100%	Lloyd et al. (2017)
		Airport density						
		Waterway density						
		Railway network						
		Rail station density						
6	Net migration	Number of in-migrants minus out-migrants	Number of people	Exposure	Resolution: 0.5' (1 km)	2015	100%	de Sherbinin et al. (2015)
7	GDP	Gross Domestic Product (PPP) per grid in 2015 (constant 2011 USD).	USD	Vulnerability	Resolution: 0.5' (1 km)	2018	100%	Kummu et al. (2018)
8	GDP per capita	Gross Domestic Product per capita (PPP) per grid in 2015 (constant 2011 USD).	USD	Vulnerability	Resolution: 5' (10 km)	2018	98%	World Bank (WDI)
9	HDI	Human Development Index [0-1]	Index	Vulnerability	Resolution: 0.5' (1 km)	2018	100%	Kummu et al. (2018)
10	Tourism	Share of travel and tourism to GDP	Percent	Vulnerability	Country level	2018	94%	World Bank (WDI)
11	Old population density	Number of female/male aged 70 or more per square kilometre in 2020	Number of people per km^2	Vulnerability	Resolution: 0.5' (1 km)	2017	100%	WorldPop and CIESIN (2018)

Covid Economics 3, 10 April 2020: 103-118

12	Infant mortality rate	The number of children who die before their first birthday per 1,000 births in 2017	Proportion	Vulnerability	Resolution: 0.5' (1 km)	2018	100%	(CIESIN, 2019)
13	Hospital beds	The number of hospital beds per 1,000 population	Number of beds	Vulnerability	Country level	2015	95%	World Health Organization (WHO)
14	Out-of-pocket	Share of Out-of-Pocket Expenditure on Healthcare	Percent	Vulnerability	Country level	2014	96%	World Bank (WDI)
15	Health spending	Total health care expenditure as GDP	Percent	Vulnerability	Country level	2014	96%	World Bank (WDI)
17	Internet access	Share of population using the Internet	Percent	Resilience	Country level	2017	99%	World Bank (WDI)
18	Cellular user	Mobile cellular subscriptions per 100 people	Numeric	Resilience	Country level	2017	99%	International Telecommunication Union (ITU)
19	Public and private debt	Ratio of central government debt to GDP	Percent	Resilience	Country level	2018	98%	IMF and WDI
		Ratio of domestic credit to private sectors to GDP						
20	Government expenditure	Ratio of government expenditure to GDP	Percent	Resilience	Country level	2018	98%	World Bank (WDI)
21	Socio - Cultural disparity	Ethnic disparity [0-1]	Index	Resilience	Country level	2016	99%	Alesina et al. (2003)
		Linguistic disparity [0-1]						

Covid Economics 3, 10 April 2020: 103-118

References

Abdi, H., and Williams, L. J. (2010). Principal component analysis. *Wiley interdisciplinary reviews: computational statistics*, 2(4), 433-459.

Alesina, A., Devleeschauwer, A., Easterly, W., Kurlat, S., and Wacziarg, R. (2003). Fractionalization. *Journal of Economic growth*, 8(2), 155-194.

Allen, T., Murray, K. A., Zambrana-Torrelio, C., Morse, S. S., Rondinini, C., Di Marco, M., Breit, N., Olival, K. J., and Daszak, P. (2017). Global hotspots and correlates of emerging zoonotic diseases. *Nature communications*, 8(1), 1124.

CIESIN. (2018). Center for International Earth Science Information Network, Columbia University. Gridded Population of the World, Version 4 (GPWv4): Population Density Adjusted to Match 2015 Revision UN WPP Country Totals, Revision 11. Retrieved from: https://doi.org/10.7927/H4F47M65

CIESIN. (2019). Center for International Earth Science Information Network, Columbia University: Global Subnational Infant Mortality Rates, Version 2. Retrieved from: https://doi.org/10.7927/H4PN93JJ

de Sherbinin, A., Levy, M., Adamo, S., MacManus, K., Yetman, G., Mara, V., Razafindrazay, L., Goodrich, B., Srebotnjak, T., Aichele, C., and Pistolesi, L. (2015). Global Estimated Net Migration Grids by Decade: 1970-2000. Retrieved from: https://doi.org/10.7927/H4319SVC

Drake, T. L., Chalabi, Z., and Coker, R. (2012). Cost-effectiveness analysis of pandemic influenza preparedness: what's missing? *Bulletin of the World Health Organization*, 90, 940-941.

Epstein, J. (2009). Modelling to contain pandemics. *Nature* 460, 687.

Hallegatte, S. (2014). Economic Resilience: Definition and Measurement. *World Bank Policy Research Working Paper* 6852.

Hallegatte, S., Bangalore, M., and Vogt-Schilb, A. (2016). Assessing socioeconomic resilience to floods in 90 countries: The World Bank.

Janes, C. R., Corbett, K. K., Jones, J. H., and Trostle, J. (2012). Emerging infectious diseases: the role of social sciences. *The Lancet*, 380(9857), 1884-1886.

Kummu, M., Taka, M., and Guillaume, J. H. (2018). Gridded global datasets for gross domestic product and Human Development Index over 1990–2015. *Scientific data*, 5, 180004.

Lloyd, C. T., Sorichetta, A., and Tatem, A. J. (2017). High resolution global gridded data for use in population studies. *Scientific data*, 4, 170001.

Nordhaus, W., and Chen, X. (2016). Global gridded geographically based economic data (G-econ), version 4. In NASA Socioeconomic Data and Applications Center (SEDAC).

Noy, I. (2016). Tropical storms: The socio-economics of cyclones. Nature Climate Change, 6(4), 343.

Noy, I., Doan N., Ferrarini B., and Park, D. (2019). Measuring the Economic Risk of Epidemics. *CESifo Working Paper 8016.*

Noy, I., and Shields, S. (2019). The 2003 Severe Acute Respiratory Syndrome Epidemic: A Retroactive Examination of Economic Costs. *Asian Development Bank Working Paper*.

Perrings, C., Castillo-Chavez, C., Chowell, G., Daszak, P., Fenichel, E. P., Finnoff, D., Horan, R. D., Kilpatrick, A. M., Kinzig, A. P., and Kuminoff, N. V. (2014). Merging economics and epidemiology to improve the prediction and management of infectious disease. *EcoHealth*, 11(4), 464-475.

Philipson, T. (2000). Economic epidemiology and infectious diseases. *Handbook of health economics*, 1, 1761-1799.

Prager, F., Wei, D., and Rose, A. (2017). Total economic consequences of an influenza outbreak in the United States. *Risk Analysis*, 37(1), 4-19.

Ringnér, M. (2008). What is principal component analysis? *Nature biotechnology*, 26(3), 303.

Román, M. O., Wang, Z., Sun, Q., Kalb, V., Miller, S. D., Molthan, A., Schultz, L., Bell, J., Stokes, E. C., and Pandey, B. (2018). NASA's Black Marble nighttime lights product suite. *Remote Sensing of Environment*, 210, 113-143.

Tatem, A. J., Adamo, S., Bharti, N., Burgert, C. R., Castro, M., Dorelien, A., Fink, G., Linard, C., John, M., and Montana, L. (2012). Mapping populations at risk: improving spatial demographic data for infectious disease modeling and metric derivation. *Population health metrics*, 10(1), 8.

Tuanmu, M. N., and Jetz, W. (2014). A global 1-km consensus land-cover product for biodiversity and ecosystem modelling. *Global Ecology and Biogeography*, 23(9), 1031-1045.

UNDRR. (2017). Terminology on Disaster Risk Reduction, UNDRR. https://www.unisdr.org/we/inform/terminology.

WHO. (2018). WHO methods and data sources for global burden of disease estimates 2000-2016. World Health Organisation, Geneva.

WorldPop (www.worldpop.org - School of Geography and Environmental Science, University of Southampton; Department of Geography and Geosciences, University of Louisville; Departement de Geographie, Universite de Namur) and Center for International Earth Science Information Network (CIESIN), Columbia University (2018). Global High Resolution Population Denominators Project - Funded by The Bill and Melinda Gates Foundation (OPP1134076). https://dx.doi.org/10.5258/SOTON/WP00646

Yonson, R., and I. Noy (2018). Economic Vulnerability and Resilience to Natural Hazards: Concepts and Measurements. *Sustainability*, 10, 2850.

Covid Economics 3, 10 April 2020: 103-118