Does Warmer Weather Provide Relief from COVID Spread?

Philip Rundall

May 22, 2020

Abstract

The COVID-19 Pandemic has shaken up the world as we know it, mainly due to the fact that there is so much uncertainty surrounding the virus's future, cures, and treatments. One of the main hopes of researchers and officials is that as the temperature increases, the spread of COVID-19 will slow down, just as other diseases such as the seasonal flu. Investigating this using the double-lasso technique we determine that temperature has a statistically significant positive effect on the spread of COVID-19 during April 2020. There are several economic speculations to explain this contrarian result.

1 Introduction

Many news articles cite 77 Degrees as the temperature which kills COVID-19 providing a false sense of security in warm climates. These studies are performed in laboratories which do not accurately simulate how disease spread is encouraged/discouraged with temperature changes. Using empirical data from five sources, we look at how the average temperature for April in each country affected the growth of COVID-19 from April 1st, 2020 - May 1st, 2020. We also include a multitude of covariates that possibly affect the spread of COVID - 19. With this data, the double - lasso technique can determine the true effect that April's temperature had on the increase in cases of COVID-19 during April. Similarly double - lasso allows us to calculate asymptotic variances for confidence intervals of our estimates.

2 Data

2.1 John Hopkins University[1]

Growth rate of total cases between 4/1/20-5/1/20 (DEPENDENT) - is calculated as a simple growth rate on cumulative cases of COVID. A growth rate of 1 would imply no new cases appearing in this month.

Total cases per POP. on 4/1/20 - accounts for the current position on the logistic growth curve when our natural temperature experiment begins. A country initially exposed to COVID at the beginning of April will have a higher expected growth rate compared to a country already having a large proportion of its population infected.

2.2 Climate.gov[2]

April Temp - is the variable of interest. It is the average temperature for the month of April for each country given in degrees celsius.

April Precip - is the average inches of rainfall in each country. This is important as heavy rain could possibly cause forced social distance or weakened immune systems.

2.3 Global Health Data Exchange[3]

Gov_Health_Spending_US_Thousands - Annual amount a government spends on healthcare and health services a year, in US thousands of dollars. This shows how a country's government prioritizes health care as well as its ability to afford health care.

Prepaid_Private_Health_Spending_US_Thousands - Annual amount spent by health insurance companies on healthcare in a country. This is important because privatized health insurance can indicate availability of high quality health care but also limited access to those who can afford it.

Out_Pocket_Health_Spending_US_Thousands - Annual amount individuals in a country spends on healthcare and health services a year, in US thousands of dollars. This shows how much healthcare individuals in a country can afford or prioritize.

2.4 Our World in Data - Global Health[4]

2019_Life_Expectancy - Average life expectancy per country in 2019. Higher life expectancy can indicate a more developed country or healthcare system.

Child_Mortality_Rate_Percent - Child mortality rate, in percentage. A lower child mortality rate indicates a more developed healthcare system

Maternal_Mortality_Per_100000 - Number of mother deaths per 100,000 live births. A lower number of deaths indicates better healthcare.

DALY_Per_100000 - Disability adjusted life years (DALY) per 100,000 people. DALY indicates how many years are lost due to disease or disability. This shows the burden of disease in a country.

Deaths_HIV_AIDS - Number of deaths due to HIV or AIDS. This can measure the level of healthcare in a country by its ability to treat a chronic disease that is not new.

New_Cases_HIV_AIDS - Number of new HIV or AIDS cases. This measures a country's ability to prevent the spread of disease already reached its peak

People_Living_With_HIV_AIDS_x10 - Number of people living with HIV or AIDS in 10's. This is important because deaths due to and number of new cases of HIV or AIDS can be relative to the number of people currently living with it.

2.5 Kaggle - Countries of the World[5]

Region - is transformed into 7-factor variables which is vital as there may be fixed effects for each region.

Population - Population of a country, has a positive correlation with how fast a virus can spread.

Area (sq. mi.) - Area of a country in square miles, a larger country can mean more people or it can mean that the population is less dense.

Pop Density - Population density, shows if people live more densely or spread out in a country which can affect the spread of a disease

Coastline (coast/area ratio) - Ratio of coastile to country's area, countries with more coastline could have increased travel due to beaches or shipping due to ports which can both increase spread.

Net migration - Net migration of people immigrating and emigrating, larger immigration can lead to increased spread of disease.

GDP ($ per capita) - GDP per capita in US dollars, indicates the wealth or development of a country which affects its healthcare.

Literacy (%) - Percent of population that is literate, usually indicates how developed and educated a country is which will affect the level of health care.

Phones (per 1000) - Number of people with phones per 1000 people, another indicator of how developed a country is.

Arable (%) - Percent of land that is suitable for growing crops, shows how naturally suited a country is to grow food. This can show how country's ability to feed its population

Table 1: Asymptotic Confidence Interval for Alpha

	BRT	BHHC	OLS
Without Interactions	[0.433, 9.82]	[0.533 , 9.846]	[-3.11, 16.476]
With 2nd Order Interations	[-3.519, 10.01]	[0.836, 9.935]	[-13.63, 13.282]

Crops (%) - Percent of land used for growing crops, also shows a country's ability to feed its population

Other (%) - Percent of land used for other purposes, shows how much of a country's land is used for other purposes.

Climate - Indicates how much rain or humidity a country generally gets. This can show how April's temperature coincides with or goes against the general climate.

Birthrate - Number of new births, shows how quickly a country's population is increasing.

Death rate - Number of deaths due to all causes, shows how quickly a country's population is declining.

Agriculture - Portion of GDP due to agriculture, indicates what a country's economy is based on

Industry - Portion of GDP due to industry, indicates what a country's economy is based on

Service - Portion of GDP due to services, indicates what a country's economy is based on

After concatenation and dropping observations missing features our data consists of 109 countries and 49 independent variables.

3 Results

Summarized in table 1:

3.1 Fitting with Bickel-Rito-Tsybakov

We first attempted to use the Bickel-Ritov-Tsybakov Rule to determine our lambdas. We fit two double-lasso models, with and without interaction terms.

3.1.1 Double Lasso with no interactions

First, we used double-lasso on our original dataset without including interactions. When using the Bickel-Ritov-Tsybakov Rule to determine lambdas, the result was 1.26 for first stage lasso regression and 107.81 for second stage lasso regression. This set 39 of 41 variables to 0. After running the two regressions, the double-lasso estimate for the temperature coefficient was 5.12, with the 95% confidence interval being [0.433, 9.82]. This shows that as temperature increases by, the growth rate of COVID-19 also increases.

3.1.2 Double Lasso with Interactions

To make the model more inclusive, we added polynomial effects to the second degree, then ran double lasso as before. This increased the number of variables to 902. When using the Bickel-Ritov-Tsybakov Rule to determine lambdas, the result was 0.619 for first stage lasso regression and 18.797 for second stage lasso regression. This set 899 of 902 variables to 0. After running the two regressions, the double-lasso estimate for the temperature coefficient was 3.25, with the 95% confidence interval being [-3.519, 10.01]. By including interaction terms, the effect of temperature on the spread of COVID-19 is still positive but has decreased. The confidence interval also includes a negative effect, which may mean the results are not as robust as previously thought.

3.2 Fitting with Belloni-Chen-Chernozhukov-Hansen

We next attempted to use the Belloni-Chen-Chernozhukov-Hansen Rule to determine our lambdas. To determine the effect of temperature on COVID-19, we again build two double-lasso models, with and without interaction terms.

3.2.1 Double Lasso with no Interactions

First, we used double-lasso on our original dataset without including interactions. When using the Belloni-Chen-Chernozhukov-Hansen Rule to determine lambdas, the result was 1.14for first stage lasso regression and 61.64 for second stage lasso regression. This set 40 of 41 variables to 0. After running the two regressions, the double-lasso estimate for the temperature coefficient was 5.39, with the 95% confidence interval being [0.533 , 9.846]. This shows that as temperature increases by, the number of COVID-19 cases also increases.

3.2.2 Double Lasoo with Interactions

Adding polynomial terms, the Belloni-Chen-Chernozhukov-Hansen Rule to determine lambdas, the result was 1.734 for first stage lasso regression and 107.81 for second stage lasso regression. This set 901 of 902 variables to 0. After running the two regressions, the double-lasso estimate for the temperature coefficient was 3.25, with the 95% confidence interval being [0.836, 9.935]. By including interaction terms, the effect of temperature on the spread of COVID-19 is still positive and is very similar to not including the interaction terms which proves the robustness of the coefficient and Double - Lasso.

3.3 OLS Comparison

3.3.1 OLS with no Interactions

We first ran a normal OLS regression, regressing growth rate on our original covariates. This gave us a beta estimate of 6.68 but a 95% confidence interval of [-3.11, 16.476]. As we can see, the estimate is still positive, but the confidence interval is much larger and also includes a negative effect, showing the variance issue with OLS.

3.3.2 OLS with Interactions

We ran a normal OLS regression again, but this time used the dataset with second degree polynomial effects. This gave us an alpha estimate of -0.17618334359603427 and a 95% confidence interval of [-13.63, 13.282]. This validates the robustness issues OLS has when dealing with a large amount of covariates.

3.4 Quadratic Temperature

Most diseases' spread tends to slow down as temperature increases, however our results show that the spread actually increases. We hypothesized that this was due to people disobeying quarantine orders as the temperature increases due to weather being nicer. To test this we also estimated the effect of temperature squared. If this is negative, it will confirm our hypothesis that at first as temperature increases, people will go out more, but as it increases further, it will slow down the spread of the virus.

3.4.1 Estimated Effect of Temperature Squared

We estimated the effect of temperature squared using the dataset that included second order polynomial effects, calculating our lambdas with the Bickel-Ritov-Tsybakov Rule. Our first lambda was 0.619, and second was 18.797. The estimated effect was 4.03, with a 95% confidence interval of [-3.778, 11.83]. This neither invalidates or validates our hypothesis of the spread of COVID-19 will start to decrease as temperature further increases and more data would be needed to tell.

4 Conclusions and Future Works

Accounting for over 50 covariates Double Lasso generated a statistically significant estimate of 5.4: which is interpreted as each degree celsius warmer we expect 540% increase in a month of cumulative COVID cases. Interpreted daily: each degree Celsius warmer should see 5.8% more cases.

While our results may seem to contradict the popular opinion on COVID's reaction to temperature this is only a Local Average Treatment Effect. Anecdotally this result makes complete sense. Southern countries of Europe were hit much harder, Southern California residents flooded parks when the temperature got tolerable and Florida not being unable to keep people off the beaches.

To explain this, we would expect a negative quadratic relationship between growth and temperature. Low temperature would curb the spread by keeping people indoors. As temperatures increase people would interact more spurring growth rates until an inflection temperature is reached. Any hotter would then kill the virus and the effect witnessed in labs would occur. However we found no significance to the quadratic term on temperature implying the temperatures observed in April may not have been high enough to observe a negative quadratic effect on COVID growth.

To further improve this work, we would like to include data that accounts for public activity or people leaving their homes. For example, the number of guests at a public park or beach would be helpful. This will allow us to better account for people leaving their homes due to nicer weather and we could further test our hypothesis. Another way to boost significance would be to go more granular and perform Double - Lasso on city level data. Doing this would create more observations with metrics more descriptive.

References

[1] https://coronavirus.jhu.edu/map.html

[2] https://www.climate.gov/maps-data/dataset/monthly-climatic-data-world-data-tables

[3] http://ghdx.healthdata.org/record/ihme-data/global-health-spending-1995-2017

[4] https://ourworldindata.org/health-meta

[5] https://www.kaggle.com/fernandol/countries-of-the-world