Dean Parker, dap2180 5/13/2020
STAT4232/5232 Final Project

Effect of Temperature on the Spread of COVID-19 in New York City

Background

Since the COVID-19 became widely known around the world, there has been conjecture as to whether warmer temperatures resulting from the transition of winter to spring would slow or help contain the spread of the disease. This was based on experiences from past outbreaks, including SARS in 2003 and yearly Influenza seasons (Stiepan, 2020). It is known that UV light (which is more plentiful as the days lengthen) kills viruses, so this conjecture makes sense in those terms (Decontamination and Reuse of Filtering Facepiece Respirators, 2020). However, the World Health Organization claims that the virus spreads just as easily in tropical, warm climates. Therefore, it is worth examining the difference in spread in the temperate New York climate.

Data

The data for COVID-19 related information (Cases, Hospitalizations, and Deaths) was obtained from the New York City Department of Health. Temperature data (Daily High, Daily Low, Daily Average) for Central Park was obtained from the National Weather Service Forecast Office. Additional indicator variables representing control measures put in place by municipal and state governments (Schools Closed, Non-Essential Businesses Closed, Masks Mandatory) were sourced from various news sites. Data was truncated at May 9th because this is the last reliable day of data. Additional derived data was created to aid in analysis:

- Day First Death: A number representing days since the first death recorded in the NYC Health Data. It begins at -11 on February 29th and ends at 59 on May 11th.
- Daily Change: Change in daily average temperature since the day prior.

- Daily Dev: Difference of Daily Average against the whole average of Daily Average. Centralizes around 0, but preserves the variance.

Analysis

The model of choice for the analysis was a Negative Binomial model with a Log-link. This model was chosen for the following reasons:

- The goal was to model count data in the form of new cases, hospitalizations, and deaths.
- The data was clearly overdispersed as the disease is very dynamic. Therefore, the Poisson model was not very applicable.
- The parameters should represent a percent increase in the data, as opposed to a direct increase as with the identity link.

To fit each Negative Binomial, it is assumed that the variance of each μ_i is scaled by a constant ϕ. This value is determined through a Poisson regression to give an estimate, $\hat{\phi}$. The formula for the estimation (as given in class) is $\hat{\phi} = \frac{1}{n-p} \frac{\sum_{i=1}^{n}(y_i - \widehat{\mu_i})^2}{\widehat{\mu_i}}$, where n is the number of data points, p is the number of regressors, and $\hat{\mu}_i$ is the estimate for the i-th point from the Poisson regression.

First, analysis was conducted to determine which result (cases, hospitalizations, or deaths) had the greatest relationship with daily temperatures. This was done through preliminary regressions of each result against "Daily_Dev" with an intercept. The results are below:

	Cases	**Hospitalizations**	**Deaths**
Coefficient	0.0106	0.0102	0.0366
P-Value	0.986	0.977	0.869

These are very poor effects, both in terms of magnitude and significance. This is a bad sign for the rest of the analysis. This could be because the spring of 2020 had been oddly cold, so there is not much variability of the daily temperature and the effects do not have a chance to pronounce themselves as much as they might have. However, the effect on deaths is the most significant (not significant in strict terms), so the rest of the analysis will continue to be based on isolating some relationship with regard to that result.

There are other variables which may impact the spread of COVID-19 and the resulting deaths. The most obvious was a variable representing the temporal aspect of the disease which was codified in the previously mentioned “Day First Death”. Including this alongside the original regressor gives the following results:

```
==============================================================================
                 Generalized Linear Model Regression Results
==============================================================================
Dep. Variable:                 Deaths   No. Observations:                   71
Model:                            GLM   Df Residuals:                       68
Model Family:        NegativeBinomial   Df Model:                            2
Link Function:                    log   Scale:                          1.0000
Method:                          IRLS   Log-Likelihood:                -594.10
Date:                Wed, 13 May 2020   Deviance:                       1.7227
Time:                        16:42:34   Pearson chi2:                    0.565
No. Iterations:                    36
Covariance Type:            nonrobust
===================================================================================
                      coef    std err          z      P>|z|      [0.025      0.975]
-----------------------------------------------------------------------------------
Intercept           3.1835      2.405      1.323      0.186      -1.531       7.898
Daily_Dev           0.0161      0.228      0.071      0.944      -0.431       0.463
Day_First_Death     0.0776      0.078      0.998      0.318      -0.075       0.230
===================================================================================
```

As reported, the already low significance of the “Daily_Dev” regressor has been reduced by the inclusion of the new regressor. This shows that the majority of the impact had been a result of the coincidental timing of the virus following its natural course at a time when the weather was heating up. The coefficients imply that on a given day, if the temperature were to deviate up a single degree, it would imply an average increase in deaths of 1.61%. This is a very miniscule impact, which cannot be accepted due to the high p-value.

The dynamics of the disease and its progression were further controlled for by introducing the previously mentioned variables for local response, including making masks mandatory in public, closing public schools, and closing non-essential businesses. These are introduced as mixed effects with days since the death, as the policies take time to affect the population. The results are as follows:

```
==================================================================================
                 Generalized Linear Model Regression Results
==============================================================================
Dep. Variable:                 Deaths   No. Observations:                   71
Model:                            GLM   Df Residuals:                       62
Model Family:        NegativeBinomial   Df Model:                            8
Link Function:                    log   Scale:                          1.0000
Method:                          IRLS   Log-Likelihood:                -466.50
Date:                Wed, 13 May 2020   Deviance:                       1.1786
Time:                        17:08:49   Pearson chi2:                    0.890
No. Iterations:                    21
Covariance Type:            nonrobust
=========================================================================================================
                                          coef    std err          z      P>|z|      [0.025      0.975]
---------------------------------------------------------------------------------------------------------
Intercept                              -1.3595      1.863     -0.730      0.466      -5.012       2.293
Daily_Dev                              -0.0009      0.087     -0.010      0.992      -0.170       0.169
Day_First_Death                         0.9078      0.893      1.017      0.309      -0.842       2.658
Schools_Closed                          1.6683      8.087      0.206      0.837     -14.182      17.519
Schools_Closed:Day_First_Death         -0.5490      1.551     -0.354      0.723      -3.589       2.491
Non_Essential_Closed                    3.4206      8.152      0.420      0.675     -12.556      19.398
Non_Essential_Closed:Day_First_Death   -0.2685      1.271     -0.211      0.833      -2.760       2.223
Mask_Mandatory                          5.3551      6.205      0.863      0.388      -6.807      17.517
Mask_Mandatory:Day_First_Death         -0.1712      0.154     -1.110      0.267      -0.474       0.131
=========================================================================================================
```

As reported, the significance of daily temperature has fallen even further. The newly introduced regressors are relatively significant (still not acceptably significant) and have the expected signs (they are negative and reduce the spread of the virus). Introducing these control variables has also altered the sign of the daily temperature to be slightly negative, a change from previous regressions. However, the impact remains insignificant.

Conclusion

After examining the baseline model, as well as models controlling for the time element and government responses, the above results can be safely interpreted as meaning that temperature has no discernable effect on the transmission or health effects of COVID-19. In fact, the models show that there are discernable results from edicts of local and state government, and this is positive news. Additional investigation could be conducted into determining whether there is a time delay element to the data, based on the incubation time of the virus, and what the most significant lag would be. Overall, this is an effective study in determining that a relationship is not present.

References

Chaffin, J. (2020, April 15). *New York orders people to wear masks in public places*. Retrieved from Financial Times: https://www.ft.com/content/8be8f9d0-55b5-4bc0-994b-a31242922eab

COVID-19: Data. (2020, May 13). Retrieved from NYC Department of Health: https://www1.nyc.gov/site/doh/covid/covid-19-data.page

Decontamination and Reuse of Filtering Facepiece Respirators. (2020, April 30). Retrieved from Centers for Disease Control: https://www.cdc.gov/coronavirus/2019-ncov/hcp/ppe-strategy/decontamination-reuse-respirators.html

Morosky, K. (2020, March 20). *New York is shutting down*. Retrieved from Riverhead Local: https://riverheadlocal.com/2020/03/20/new-york-is-shutting-down-most-businesses-ordered-closed-all-non-essential-workers-must-stay-home-ban-on-social-gatherings/

Online Weather Data. (2020, May 13). Retrieved from weather.gov: https://w2.weather.gov/climate/index.php?wfo=okx

Shapiro, E. (2020, March 15). *NYC Schools Closed*. Retrieved from New York Times: https://www.nytimes.com/2020/03/15/nyregion/nyc-schools-closed.html

Stiepan, D. (2020, March 10). *What to expect from COVID-19 as the weather warms up*. Retrieved from Mayo Clinic: https://newsnetwork.mayoclinic.org/discussion/what-to-expect-from-covid-19-as-the-weather-warms-up/

Dean Parker, dap2180 5/13/2020
STAT4232/5232 Final Project

Python Script

```
import pandas as pd
import numpy as np
import statsmodels.api as sm
from patsy import dmatrices

data = pd.read_csv("./data.csv", parse_dates=True, index_col=[0])

def run_regression(s, p):
    y, X = dmatrices(s, data, return_type='dataframe')
    n = len(y)
    poisson = sm.GLM(y, X, family=sm.families.Poisson()).fit()
    mus = poisson.mu
    phi_hat = np.sum((y.values[:, 0] - mus)**2 / mus) / (n - p)
    nb2 = sm.GLM(y, X, family=sm.families.NegativeBinomial(alpha=phi_hat,
link=sm.families.links.log)).fit()
    return nb2

print(run_regression("Cases ~ Daily_Dev", 1).summary())
print(run_regression("Hospitalizations ~ Daily_Dev", 1).summary())
print(run_regression("Deaths ~ Daily_Dev", 1).summary())

print(run_regression("Deaths ~ Daily_Dev + Day_First_Death", 2).summary())

print(run_regression("Deaths ~ Daily_Dev + Day_First_Death +
Schools_Closed*Day_First_Death + Non_Essential_Closed*Day_First_Death +
Mask_Mandatory*Day_First_Death", 8).summary())
```