Investigating the Connection Between Mobility and Covid19

Aaron Jacobson

Introduction

As the Covid19 pandemic has lasted almost two years now, we have begun to collect data that helps us better understand just how it works. Non-pharmaceutical interventions like mask mandates and social distancing guidelines have been a key way we have tried to slow the spread of Covid19. Many of these interventions are based on the idea that the more time people spend being close to others, the more the virus will spread. The key question I hoped to address with this project was "How do levels of human activity correlate with the spread of Covid19?" I used publicly available data from John Hopkins, Apple, and OpenTable to gain some insight. Given the data I have, I also attempt to investigate whether different kinds of transportation are associated with increases in Covid19.

Background

Much has been written about Covid19 in both academic and popular sources. One article, "An Evidence Review of Face Masks against COVID-19", finds "the preponderance of evidence indicates that mask-wearing reduces transmissibility" (Howard, et al. 2021). When it comes to public transit, Scientific American's article "There Is Little Evidence That Mass Transit Poses a Risk of Coronavirus Outbreaks" discusses evidence from France and Japan that shows that those country's contact tracing systems found very few Covid19 clusters connected to their transit systems (Joselow 2020). BBC's article, "Coronavirus: What's the Risk of Covid on Public Transport?" discusses the various risks posed by the close quarters found in public transit (BBC 2021). NBC's article "Public Transit in Cities Not Huge Contributor in Spread of Respiratory Diseases: Study" finds that the actual spread associated with public transit is minimal (NBC New York, 2021).

Methodology

To answer this question I used data from John Hopkins on new Covid19 cases in various counties in New York State, Apple's Mobility Trends Report data, and OpenTable's State of Industry data. I created a lagged variable that represents the % change from the previous days in the increase in new daily reported cases. To deal with data issues relating to day-of-the-week effects I ignored data from Mondays, Saturdays, and Sundays. I then ran OLS regression with heteroskedastic robust standard errors using the % change in the number of new daily cases as the response variable with the level of driving, transit, walking, and restaurant reservation activity as the explanatory variables.

Findings

To assess the robustness of statistical significance I first ran simple OLS regressions investigating the relationship between each of my explanatory variables and the dependent variable.

Below is the regression table for the regression model of the form:
Percent Change in New Daily Covid Cases ~ constant + coefficient * Transit_Level

Dep. Variable:	Cases_Change %	**R-squared:**	0			
Model:	OLS	**Adj. R-squared:**	0			
Method:	Least Squares	**F-statistic:**	1.906			
Date:	Tue, 14 Dec 2021	**Prob (F-statistic):**	0.167			
Time:	19:22:14	**Log-Likelihood:**	-4416.7			
No. Observations:	4038	**AIC:**	8837			
Df Residuals:	4036	**BIC:**	8850			
Df Model:	1					
Covariance Type:	HC3					
	coef	**std err**	**z**	**P>|z|**	**[0.025**	**0.975]**
const	0.2359	0.027	8.769	0	0.183	0.289
Transit_Level	-0.0004	0	-1.381	0.167	-0.001	0
Omnibus:	3279.97	**Durbin-Watson:**	1.845			
Prob(Omnibus):	0	**Jarque-Bera (JB):**	124588.692			
Skew:	3.612	**Prob(JB):**	0			
Kurtosis:	29.236	**Cond. No.**	198			

Below is the regression table for the model of the form:
Percent Change in New Daily Covid Dases ~ constant + coefficient * Driving_Level

Dep. Variable:	Cases_Change	**R-squared:**	0.001			

	%					
Model:	OLS	**Adj. R-squared:**	0			
Method:	Least Squares	**F-statistic:**	2.239			
Date:	Tue, 14 Dec 2021	**Prob (F-statistic):**	0.135			
Time:	19:22:24	**Log-Likelihood:**	-4416.3			
No. Observations:	4038	**AIC:**	8837			
Df Residuals:	4036	**BIC:**	8849			
Df Model:	1					
Covariance Type:	HC3					
	coef	**std err**	**z**	**P>|z|**	**[0.025**	**0.975]**
const	0.2636	0.044	6.059	0	0.178	0.349
Driving_Level	-0.0005	0	-1.496	0.135	-0.001	0
Omnibus:	3273.944	**Durbin-Watson:**	1.845			
Prob(Omnibus):	0	**Jarque-Bera (JB):**	123662.875			
Skew:	3.604	**Prob(JB):**	0			
Kurtosis:	29.135	**Cond. No.**	441			

Below is the regression table for the regression model of the form:
Percent Change in New Daily Covid Cases ~ constant + coefficient * Walking_Level

Dep. Variable:	Cases_Change %	**R-squared:**	0			
Model:	OLS	**Adj. R-squared:**	0			
Method:	Least Squares	**F-statistic:**	1.265			
Date:	Tue, 14 Dec 2021	**Prob (F-statistic):**	0.261			
Time:	19:22:32	**Log-Likelihood:**	-4417			
No. Observations:	4038	**AIC:**	8838			
Df Residuals:	4036	**BIC:**	8851			
Df Model:	1					
Covariance Type:	HC3					

	coef	std err	z	P>|z|	[0.025	0.975]
const	0.2359	0.032	7.267	0	0.172	0.3
Walking_Level	-0.0002	0	-1.125	0.261	-0.001	0
Omnibus:	3277.438	**Durbin-Watson:**	1.845			
Prob(Omnibus):	0	**Jarque-Bera (JB):**	124302.556			
Skew:	3.608	**Prob(JB):**	0			
Kurtosis:	29.205	**Cond. No.**	383			

Below is the regression table for the regression model of the form:
Percent Change in New Daily Covid Cases ~ constant + coefficient * Table_Level

Dep. Variable:	Cases_Change %	**R-squared:**	0.001			
Model:	OLS	**Adj. R-squared:**	0.001			
Method:	Least Squares	**F-statistic:**	3.546			
Date:	Tue, 14 Dec 2021	**Prob (F-statistic):**	0.0598			
Time:	19:22:39	**Log-Likelihood:**	-4415.1			
No. Observations:	4038	**AIC:**	8834			
Df Residuals:	4036	**BIC:**	8847			
Df Model:	1					
Covariance Type:	HC3					
	coef	**std err**	**z**	**P>|z|**	**[0.025**	**0.975]**
const	0.2435	0.025	9.57	0	0.194	0.293
Table_Level	-0.0011	0.001	-1.883	0.06	-0.002	4.55E-05
Omnibus:	3272.852	**Durbin-Watson:**	1.843			
Prob(Omnibus):	0	**Jarque-Bera (JB):**	124084.647			
Skew:	3.6	**Prob(JB):**	0			
Kurtosis:	29.185	**Cond. No.**	75.7			

Below is the regression table produced by the python library StatsModels using the model described in the methods section.

Dep. Variable:	Cases_Change %	**R-squared:**	0.001			
Model:	OLS	**Adj. R-squared:**	0			
Method:	Least Squares	**F-statistic:**	0.9666			
Date:	Tue, 14 Dec 2021	**Prob (F-statistic):**	0.425			
Time:	19:14:56	**Log-Likelihood:**	-4414.7			
No. Observations:	4038	**AIC:**	8839			
Df Residuals:	4033	**BIC:**	8871			
Df Model:	4					
Covariance Type:	HC3					
	coef	**std err**	**z**	**P>|z|**	**[0.025**	**0.975]**
const	0.241	0.051	4.758	0	0.142	0.34
Transit_Level	0.0004	0.001	0.793	0.428	-0.001	0.002
Driving_Level	-0.0002	0.001	-0.299	0.765	-0.002	0.001
Walking_Level	9.37E-05	0	0.22	0.826	-0.001	0.001
Table_Level	-0.0015	0.001	-1.243	0.214	-0.004	0.001
Omnibus:	3271.424	**Durbin-Watson:**	1.841			
Prob(Omnibus):	0	**Jarque-Bera (JB):**	123908.319			
Skew:	3.598	**Prob(JB):**	0			
Kurtosis:	29.166	**Cond. No.**	806			

Previous work I did found statistically significant relationships between the level of human activity and the percent change in new covid cases but after fixing some bugs in my code the statistically significant relationships went away.

Discussion

While the model with just the restaurant table reservations had a statistically significant coefficient for the explanatory variable, the full model has no statistically significant coefficients for the explanatory variables. Were this the only evidence on the matter, it would suggest there is likely no relationship between human activity and the spread of Covid19. Given the evidence from other sources, both popular and peer-reviewed, these results may be due to the limitations of the data used.

Limitations

The limitations of this study can be grouped into two categories, limitations of the data used and limitations of the model used. The first limitation of the Apple mobility data is the fact that it only represents people who use Apple Maps. Additionally, the Apple mobility data reflects the level of activity relative to January 13, 2020, but not the amount of activity in absolute terms. The OpenTable data is limited in that it represents only a small portion of restaurants and in that it is a statewide, as opposed to countywide, level of data. The main limitation in the John Hopkins Covid19 case data is the lack of use of the test positivity rate in the models. Some big limitations of the model I used include bad lagged autocorrelation choices, the model not reflecting the incubation period of Covid19, and a lack of other controls such as some representation of changes in mask use over time.

Conclusion

My models suggest there is likely no relationship between the level of human activity in New York State and the spread of Covid19 or a relationship between any specific form of mobility and the spread of Covid19. However, other public research using other data sets has found contradictory results. Given the contradiction between my results and the results of others, people should still take precautions to avoid catching and spreading Covid19.

References

Joselow, Maxine. “There Is Little Evidence That Mass Transit Poses a Risk of Coronavirus Outbreaks.” Scientific American, Scientific American, 28 July 2020, https://www.scientificamerican.com/article/there-is-little-evidence-that-mass-transit-poses-a-risk-of-coronavirus-outbreaks/.

“Coronavirus: What's the Risk of Covid on Public Transport?” BBC News, BBC, 21 May 2021, https://www.bbc.com/news/health-51736185.

NBC New York. “Public Transit in Cities Not Huge Contributor in Spread of Respiratory Diseases: Study.” NBC New York, NBC New York, 3 Dec. 2020, https://www.nbcnewyork.com/news/coronavirus/public-transit-in-cities-not-huge-contributor-in-spread-of-respiratory-diseases-study-finds/2759012/.

Hudson, Alex. “As Seattle Slowly Reopens, Focus on the Facts, Not Fears, of Bus Ridership.” The Seattle Times, The Seattle Times Company, 10 May 2021, https://www.seattletimes.com/opinion/as-seattle-slowly-reopens-focus-on-the-facts-not-fears-of-bus-ridership/.

Stabley, Justin, and Vignesh Ramachandran. “Covid-19 Changed Public Transportation. Here's How.” PBS, Public Broadcasting Service, 10 June 2021, https://www.pbs.org/newshour/nation/public-transit-post-pandemic.

Howard, Jeremy, et al. “An Evidence Review of Face Masks against COVID-19.” Proceedings of the National Academy of Sciences - PNAS, vol. 118, no. 4, 2021, p. 1.

Data Sources

John Hopkins University Covid-19 Data:
https://www.kaggle.com/antgoldbloom/covid19-data-from-john-hopkins-university
Apple Mobility Trends Reports:
https://covid19.apple.com/mobility
OpenTable State of Industry:
https://www.opentable.com/state-of-industry