Reasons behind Mobility Decline in the US during Covid-19- Evidence from Twitter and Local Policy Data

Mengshan Zhao

1 Research Object

Covid-19 became the global event in 2020. It influenced many aspects of individual lives. The government policy also played a dramatic role during the process, especially in March, when the local governments issued Stay-at-home order while the central government introduced CARES action. The situation requires researchers to understand how human behaviors change and their effect on the economy.

This study focused on a currently focused question-- Which is the major force of citizen’s mobility decrease during the Covid-19? Is that due to the lockdown policy or social awareness? If it was led by social awareness, what kind of emotional reaction will influence mobility mostly?

There are generally two directions for Covid-19 studies: The first one looks at how the disease or some policies (lockdown or CARES, for example) influence people's behaviors, while the second orientation suggests their direct economic impact. My research is closer to the first type.

There have been some papers discussing the impact of social awareness and lockdown policy: Goolsbee and Syverson (2020) argued that the lockdown policy can only explain a 7 percent decline in consumption flow, while the majority decrease should be due to the increasing social awareness. Goldfarb and Tucker (2020) applied retail activity data to evaluate citizen’s mobility. What’s more, studies on the social network also support that social awareness is playing a big role in human behaviors. For example, Bailey, Cao, Kuchler & Stroebel (2018a) found that the social network will influence people’s decisions on housing. That suggests researchers find other reasons and to measure their impacts on the mobility decrease. Because if the mobility decline was not only due to the lockdown policy, the economy flow may not increase only by removing the lockdown policy.

Besides studies in the economic area, researchers already notice the social media data have been a great tool to analysis people’s response towards a particular issue. Because it can capture the effect directly and promptly. With the help of machine learning methods and other tools, researchers are able to build models to help identify emotions of large set literal content. Many researchers have summarized the commonly used tools in sentiment study for social media studies, like Kharde & Chapman (2018). Giachanou & Crestani (2016), Sahayak, Shete, Pathan (2015). They found that machine learning and Lexicon-Based models are most employed. There are also many studies that applied the sentiment study on Twitter. For instance, Roberts, Sadler & Moore (2019) used Twitter data to determine people’s emotional response to urban green spaces. Allen et al. (2016) trained machine learning model to monitor influenza outbreaks with Twitter data.

In this research, I want to measure the real impact on citizen’s mobility by the sudden lockdown policy and the gradual awareness employing time-series datasets. Although the lockdown policy also generated a social awareness impact, the impact will show after it announced, which is normally earlier than its delivered.

2 Data Description

In this research, I combined three main panel data: The mobility dataset, the Twitter data, and the Lockdown dataset. The first one is the response variable and the latter two are the explanatory variables.

2.1 Mobility Dataset

The mobility data collected by Descartes Labs.[[1]](#footnote-1) They collected anonymous locations data from the mobile device using cloud computing strategies, then computed the daily mobility of citizens living in each county in the US. It is worth noticing that the data at the county level is important as the detailed lockdown dates are varied between counties. Some counties began lockdown before the state government’s decision. The time range of the mobility dataset is from March first to the current, and they keep updating.

The citizen’s mobility is calculated in the following way: According to their explanation, the mobile device will automatically report a new position by its latitude, longitude, and an estimate of accuracy with other identified information like time and its id. The researchers will collect the data once a day, which generally around 100GB, and use an algorithm to normalize the data. After the data clearing, they obtained the max-distance mobility per individual within the given county and select the median of the max-distance mobility-- m50 as the county’s mobility. What’s more, they use a normalization function to normalize m50. The normalization function is as follows:

Where m50 stands for the median value of the max-distance mobility in a given region and m50\_norm is the standard level of mobility in that region, defined as the median m50 from an earlier period of the region.

Table 1 shows an example of the mobility data, and Table 2 is the distribution of the m50 and m50\_index. The two distributions have significant differences as the function denotes, m50 is in kilometers while the index is in percentage.

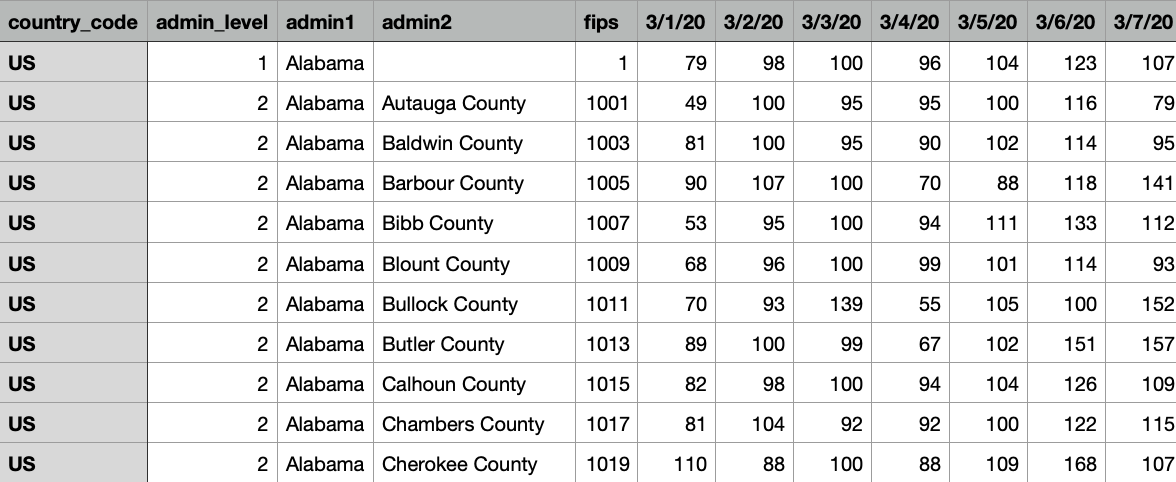


Table 1 Sample of Mobility Data

Fig. 1 and 2 are two sample graphs of the mobility dataset. Fig.1 is the change of mobility of six different states, and the Fig.2 shows the trend of mobility of different counties in Wisconsin state during March. We can see on the state level, the change pattern is very similar. Moreover, there is a decline at the beginning of April, when most of states begin the stay at home policy. However, we can still find the gap in the county level.

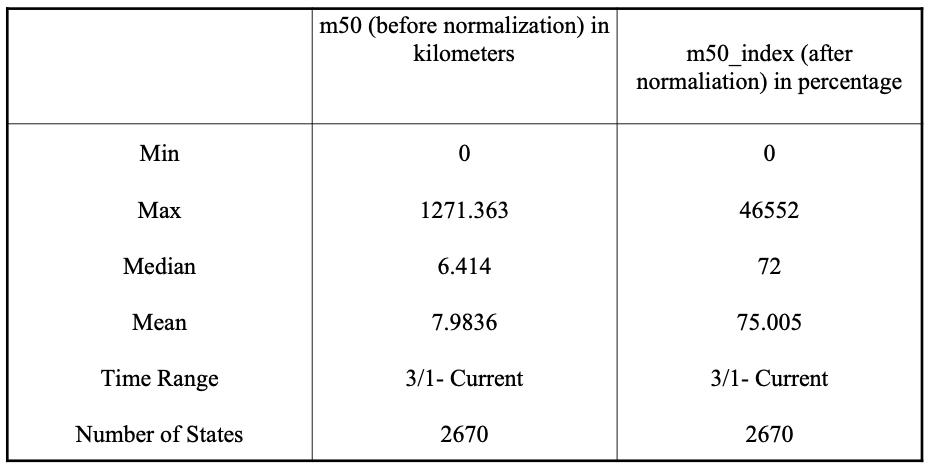
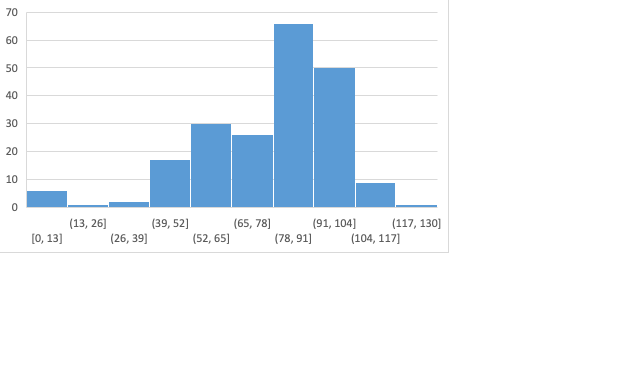


Table 2 Distributions of m50 and m50\_index

Fig 3 is a histogram of one state, Alabama, we can see that after the normalization process, the most frequent value of m50\_index is around [80,90].

2.2 Twitter Chatter Dataset

The second dataset is the Covid-19 Twitter chatter dataset maintained by Georgia State University’s Panacea Lab.[[2]](#footnote-2) This dataset tracked all the tweets and re-tweets that are relevant to Covid-19 from January, including all languages and nations. It included two versions of data, the first type is the whole dataset, and the second one is the “clean set” without all the retweeting. Table 3 shows the summary statistics for the clean versions. The researchers also collected the top 1000 words used each day from all twitters, which is more helpful with the natural language processing (NLP) task. Due to the privacy restrictions of Twitter, they can only share the Twitter ID instead of the real content of Twitters. However, researchers can trace back the original content through the hydrating process with some open-source packages and use it for non-profit study. [[3]](#footnote-3)

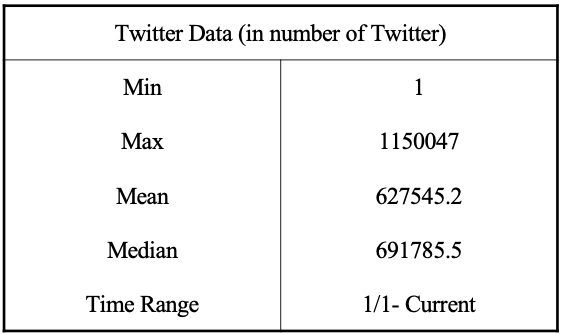


Table 3. Distribution of Twitter Data

Fig. 4 shows the trend of the number of tweets, and Fig.5 is the trend of the clean set.

2.3 The Local Policy Dataset

The third dataset is the Lockdown Policy Dataset collected by Goolsbee, Austan(2020). It included the Stay-at-home order beginning and end date of all the counties in the United States. As the authors of the dataset mentioned, the county-level data is useful as some counties take action before the states, especially in the states with a larger number of cases.

Fig 6 shows a combination of the three datasets. The orange line is the mobility change of Wisconsin during March, and the blue line is the number of Twitter during that period. We can see there is a pattern that increasing in the number of Twitter leads to a decline in mobility. Moreover, the mobility already decreased a lot when the State began lockdown on March 25th.[[4]](#footnote-4)

2.4 Other Datasets

There are other county-level datasets added as control variables. Currently, I use Local Area Unemployment data from the United States Bureau of Labor Statistics, and Number of Cases Data summarized by the New York Times.

Other time-invariant variables like poverty rate, age distribution, gender ratio are not included.

3 Methodology

3.1 Logic of the model

The reasons behind mobility decrease can be generated by two major parts. The first one is the obligation due to policy restriction, as it states what an individual cannot do. The second reason is people’s self-awareness. The self-awareness states what an individual should or should not do. In this case, the self-awareness is majorly shaped by social education. However, individuals may have different understandings of the same information. In that sense, to measure the impact of social awareness, the researchers could not only rely on the media reports but also need to analyze individual reactions to the reports. Different from traditional newspapers, social network platforms like Twitter can provide a relatively big picture of how the information is created, processed, and learned. Researchers can not only learn how much people focused on a particular issue, but also know their emotions towards the issue. The method to classify a paragraph is Positive/Negative/ Neutral emotional is called Sentiment Analysis. In this way, the Twitter data may serve as a great data source to evaluate the impact of social awareness on mobility decrease. Fig 7 displays the flow of how the two approaches contribute to mobility decline.

With this setting, the model will use Regression Discontinuity Design in time series (RDiT) to analysis the effect of Lockdown Policy, and the number of Tweets as well as their emotions on mobility decline. RDiT is employed as the Lockdown order has a specific date, and there are not other significant changes before and after the few days of its implemented, then by design it can measure the change around a threshold.[[5]](#footnote-5)

Diagram

Description automatically generated

Fig 7. Reasons behind Mobility Decrease

Moreover, the model is aimed to test the following four hypotheses:

Hypothesis 1: The Lockdown Policy has a negative effect on mobility.

Hypothesis 2: The Number of Tweets related has a negative effect on mobility.

Hypothesis 3: The effect of the percentage of Negative Sentiment Tweets are larger thant the Positve ones on mobility.

Hypothesis 4: The Number and Sentiments of Tweets play a more important role in mobility decline than Lockdown policy.

3.2 The RDiT Model

RD in time series can be employed for the Lockdown policy analysis as the policy stated a specific time on the beginning and ending time in each county. According to the discussion provided by Hausman and Rapson (2018), unlike the traditional Regression Discontinuity Design (RDD), Researchers may not able to be argued RDiT is close to a natural experiment. As in our case, the date of Lockdown is determined and announced before its implement, so citizen may change their behaviors according to it rather than random. However, if the analysis used a relatively narrow time window, The RDiT is able to show how people behaviors change around the threshold with a flexible time trend (Lucas, 2018).

The lockdown data in my analysis is the date at which the lockdown was implemented, rather than the date it introduced. In the later robustness test part, the introduced day will be used instead, which is expected to give a similar result.

The formula below shows how RDiT works there. In this formula, , is the object variable, mobility of county i at the time t. The date variable is , which denotes how many days from the day t to the implementation day for county i. The value is negative when the day t is prior to the implementation date, and positive if it was after that date. is a dummy variable, it takes one after the policy implementation, and zero before it.

The most concerned control variable is the sentiment variables: is a percentage of Tweets have a positive emotion at time t, while is a percentage of Tweets are negative at time t. is the number of Tweets at time t. These three categories are defined by the existing Sentiment Analysis literature on Twitter data: Negative, Positive, and Neutral. With this method, we can track the basic emotions change in the social media. Our analysis used the vaderSentiment package by Hutto, C.J. & Gilbert, E.E (2014). They developed pre-trained machine learning and Lexicon based model target to classify different types of sentiments on social media data and had examined the model with multiple datasets. In the result part, I will discuss the accuracy of the model. [[6]](#footnote-6)After that, I calculated the proportion of Tweets that are negative, positive or neutral. The sentiment is time-variant but the same for each county, as people from all counties share the same Twitter platform. However, it will not absorb by the time fixed effect in the model, because the time fixed effects are calculated at month level.

Besides the sentiment variables, county fixed effect and month-level time fixed effect , the model also takes into account control variables that are both time-variant and county-variant. includes local unemployment rates, whether the day is weekend[[7]](#footnote-7), infection and death numbers at county i, day t.

Where  is the time fixed effect of month m, is the county fixed effect of a county i. is a polynomial function of which allows for non-linear relationships or different functions before or after Lockdown. The time fixed effect including other big events in national level, the weather, or whether this day is workday or not, etc. The county fixed effect covers all generally statistics at a county level, like average education, gender ratio, age distribution, income distribution, family size.

4 Results

The main results can be divided by two parts, the first one is the result of sentiment analysis, and the second part is the results for RDiT model.

4.1 Sentiment Analysis Result

The results of sentiment analysis using vaderSentiment are displayed by Fig 8 and Table 4, where the sentiments of Tweets are divided into three categorical: positive, negative and neutral

. This result is calculated based on a randomly selected sample of each day during that time period. The overall sample size is 826,310 in terms of Tweets.[[8]](#footnote-8) All Tweets chosen are in English. The time range of the sample is from March 16 to April 7, according to the range of first lockdown issued.

Table

Description automatically generated

Table 4. Sentiment Percentage Change

From the graph, we can see the sentiment of the Tweets are relatively steady during the lockdown period, Negative sentiment takes the majority at first, but the positive sentiment increased with the time being and takes the lead at the beginning of April. This result is consistent to what we have learned from stock market, as the price increases with lockdown policy implemented.

For the accuracy of the model, as our sample is unlabeled, I employed human-labeled Twitter dataset Sentiment140 as test set. The results are still under possessing. Moreover, alternative models using Lexicon method or deep learning are included in the robustness tests part to justify the result produced by our model.

4.2 RDiT Model Results

Table 5 shows the results of the RDiT model. “Before” states the day before the policy implemented, “Equal” is the day of implemented, and “After” denotes the result 1 day after implemented. Because there only includes three-days data, the “Before” dummy is omitted due to collinearity, therefore, it serves as base for the model. The results are calculated by a linear assumption on the time parameters. The first column is the result without all the controls, the second column is the result with Tweets and Weekend controls, and the third column it the result with Tweets, Weekend, Number of Cases, and Death controls. All the three models use county and month fixed effect controls.

The results are consistent to our assumptions: Firstly, comparing to the after Lockdown implemented based group, people have higher mobility score before the lockdown, or at the day of lockdown. According to the third model, the mobility score is decreased by 4.863 on average at the date of implemented, and 12.38 after one day of the implemented comparing to one day before the Lockdown implemented.

Moreover, when compare to the neutral sentiment, increasing in the percentage of negative feelings lead to a significant decline in mobility. According to the model 3, one percentage increases in the negative sentiment leads to 0.42 decline in mobility score on average. The significance levels are varied across model, it is 0.001 level of significance for the model 2, and 0.1 level of significance at model 3. In the meantime, the positive sentiment has a positive effect on citizen’s mobility.

However, the time parameters may have a non-linear relationship with mobility change, that requires us to look into longer time period before and after the Lockdown and included functions with higher degree or different shape before or after the Lockdown.[[9]](#footnote-9)

Table

Description automatically generated

Note: ‘\*\*\*’means a significant level of 0.001; ‘\*\*’ means a significant level of 0.01; ‘\*’ means a significant level of 0.05, and ‘.’ means a significant level of 0.1. [[10]](#footnote-10)

Table 5. Linear RDiT results

Table 5. Linear Outputs of RDiT Model

Reference:

Bailey, Michael, et al. "The economic effects of social networks: Evidence from the housing market." *Journal of Political Economy* 126.6 (2018): 2224-2276.

Baker, Scott R., et al. *Income, liquidity, and the consumption response to the 2020 economic stimulus payments*. No. w27097. National Bureau of Economic Research, 2020.

Goolsbee, Austan, and Chad Syverson. *Fear, lockdown, and diversion: Comparing drivers of pandemic economic decline 2020*. No. w27432. National Bureau of Economic Research, 2020.

Bartik, Alexander W., et al. *How are small businesses adjusting to covid-19? early evidence from a survey*. No. w26989. National Bureau of Economic Research, 2020.

Jordà, Òscar, Sanjay R. Singh, and Alan M. Taylor. *Longer-run economic consequences of pandemics*. No. w26934. National Bureau of Economic Research, 2020.

Kaplan, Greg, Benjamin Moll, and Giovanni L. Violante. *The Great Lockdown and the Big Stimulus: Tracing the Pandemic Possibility Frontier for the US*. No. w27794. National Bureau of Economic Research, 2020.

Baker, Scott R., et al. *How does household spending respond to an epidemic? consumption during the 2020 covid-19 pandemic*. No. w26949. National Bureau of Economic Research, 2020.

Hausman, Catherine, and David S. Rapson. "Regression discontinuity in time: Considerations for empirical applications." *Annual Review of Resource Economics* 10 (2018): 533-552.

Goldfarb, Avi, and Catherine Tucker. *Which Retail Outlets Generate the Most Physical Interactions?*. No. w27042. National Bureau of Economic Research, 2020.

Goolsbee, Austan, Nicole Bei Luo, Roxanne Nesbitt, and Chad Syverson. 2020. "COVID-19 Lockdown Policies at the State and Local Level." BFI Working Paper 2020-116, https://bfi.uchicago.edu/working-paper/2020-116/

Allen, Chris, et al. "Applying GIS and machine learning methods to Twitter data for multiscale surveillance of influenza." *PloS one* 11.7 (2016): e0157734.

Agarwal, Apoorv, et al. "Sentiment analysis of twitter data." *Proceedings of the workshop on language in social media (LSM 2011)*. 2011.

Giachanou, Anastasia, and Fabio Crestani. "Like it or not: A survey of twitter sentiment analysis methods." *ACM Computing Surveys (CSUR)* 49.2 (2016): 1-41.

Kharde, Vishal, and Prof Sonawane. "Sentiment analysis of twitter data: a survey of techniques." *arXiv preprint arXiv:1601.06971* (2016).

Kouloumpis, Efthymios, Theresa Wilson, and Johanna Moore. "Twitter sentiment analysis: The good the bad and the omg!." *Fifth International AAAI conference on weblogs and social media*. 2011.

Gautam, Geetika, and Divakar Yadav. "Sentiment analysis of twitter data using machine learning approaches and semantic analysis." *2014 Seventh International Conference on Contemporary Computing (IC3)*. IEEE, 2014.

Kumar, Shamanth, Fred Morstatter, and Huan Liu. *Twitter data analytics*. New York, NY: Springer New York, 2014.

Colnerič, N., & Demšar, J. (2018). Emotion Recognition on Twitter: Comparative Study and Training a Unison Model. IEEE Transactions on Affective Computing, PP (99), 1. [*https://doi.org/10.1109/TAFFC.2018.2807817*](https://doi.org/10.1109/TAFFC.2018.2807817)

Roberts, Helen, Jon Sadler, and Lee Chapman. "The value of Twitter data for determining the emotional responses of people to urban green spaces: A case study and critical evaluation." *Urban studies* 56.4 (2019): 818-835.

Documenting the Now. (2020). Hydrator [Computer Software]. Retrieved from https://github.com/docnow/hydrator

Hausman, Catherine, and David S. Rapson. "Regression discontinuity in time: Considerations for empirical applications." *Annual Review of Resource Economics* 10 (2018): 533-552.

Lee, David S., and Thomas Lemieux. "Regression discontinuity designs in economics." *Journal of economic literature* 48.2 (2010): 281-355.

Brodeur, Abel, et al. "Covid-19, lockdowns and well-being: Evidence from google trends." (2020).

Aiken, Abigail RA, et al. "Demand for Self-Managed Online Telemedicine Abortion in Eight European Countries During the COVID-19 Pandemic: A Regression Discontinuity Analysis." *medRxiv* (2020).

Bækgaard, Martin, et al. "Rallying around the flag in times of Covid-19: Societal lockdown and trust in democratic institutions." *Journal of Behavioral Public Administration* 3.2 (2020).

Davis, Lucas W. "The effect of driving restrictions on air quality in Mexico City." *Journal of Political Economy* 116.1 (2008): 38-81.

Hutto, C.J. & Gilbert, E.E. (2014). VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. Eighth International Conference on Weblogs and Social Media (ICWSM-14). Ann Arbor, MI, June 2014.

1. Link of their project: <https://github.com/descarteslabs/DL-COVID-19> [↑](#footnote-ref-1)
2. Link to their project: <http://www.panacealab.org/covid19/> [↑](#footnote-ref-2)
3. Link of the hydrator software: <https://github.com/docnow/hydrator> [↑](#footnote-ref-3)
4. Different from some states, all the counties in Wisconsin use the same day of lockdown. [↑](#footnote-ref-4)
5. As the data already been normalized using the mobility at the beginning of the event, we did not need lagged control variables here. [↑](#footnote-ref-5)
6. vaderSentiment is not the only commonly used package in sentiment analysis. My current plan is to label a test set myself using the existing Tweets at a size around 1000, and test which packages/models have the best accuracy rate. [↑](#footnote-ref-6)
7. Similar to the sentiments, whether it is weekend should be county-invariant, however, due to the threshold day is variant between counties, so it is also variant with different counties. [↑](#footnote-ref-7)
8. The original sample size is 1,038,305, however, because about 20% of Tweets are deleted by the users or Twitter company, we can only recall this sample size. Moreover, sampling is employed due to time limit, I will work on the whole [↑](#footnote-ref-8)
9. Will include later. [↑](#footnote-ref-9)
10. Due to difference in county names documenting, the observations for model 3 is less than model 1 and 2, which leads to NA in the data. I will keep working on that to find where the difference lies. [↑](#footnote-ref-10)