Effectiveness and Compliance to Social Distancing During COVID-19

Kristi Bushman
Dpt. of Computer Science
University of Pittsburgh
k.bushman@pitt.edu

Konstantinos Pelechrinis
Dpt. of Informatics & Networked
Systems
University of Pittsburgh
kpele@pitt.edu

Alexandros Labrinidis
Dpt. of Computer Science
University of Pittsburgh
labrinid@cs.pitt.edu

ABSTRACT

In the absence of pharmaceutical interventions to curb the spread of COVID-19, countries relied on a number of nonpharmaceutical interventions to fight the first wave of the pandemic. The most prevalent one has been stay-at-home orders, whose the goal is to limit the physical contact between people, which consequently will reduce the number of secondary infections generated. In this work, we use a detailed set of mobility data to evaluate the impact that these interventions had on alleviating the spread of the virus in the US as measured through the COVID-19-related deaths. To establish this impact, we use the notion of Granger causality between two time-series. We show that there is a unidirectional Granger causality, from the median percentage of time spent daily at home to the daily number of COVID-19-related deaths with a lag of 2 weeks. We further analyze the mobility patterns at the census block level to identify which parts of the population might encounter difficulties in adhering and complying with social distancing measures. This information is important, since it can consequently drive interventions that aim at helping these parts of the population.

1 INTRODUCTION

Since the first reported case of COVID-19 in early December 2019 in Wuhan, China, the world has experienced dramatic changes in an effort for societies to deal with the pandemic. Given the absence of pharmaceutical interventions (i.e., a medical treatment or a vaccine), governments and health officials have relied on non-pharmaceutical interventions, including *shelter-at-home* orders, contact tracing and volume testing. The reasoning behind shelter-at-home interventions is to limit the physical contacts between people, which furthermore limits the transmission of the virus. Given the absence of a vaccine, this does not mean that the virus will be eradicated but

rather, limiting people's mobility will allow the health systems to operate under capacity and be as effective as possible, consequently limiting the number of fatalities.

Of course, these measures have not been without controversy. Hence, it is important to examine whether they are effective in achieving their goal. Using a granular mobility dataset for the US obtained from SafeGraphs (details provided in the following sections) and COVID-19-related fatalities we show that average fraction of people staying home weekly *Granger-causes* the number of COVID-19-related fatalities with a 3-weeks lag. We also examined for and did not find any evidence of bidirectional Granger causality, i.e., feedback effects of people altering their mobility as a response to the change of the number of fatalities (e.g., as a reaction to the news). We also provide a short-term prediction model for the number of COVID-19 related fatalities in US one-week out, using only information about population-level mobility behavior and fatalities over the past three weeks.

Given the effectiveness of these measures it is critical to understand who in the population complies and to what extent. Differences in compliance levels are not necessarily by choice. For example, many people are essential workers and hence, need to spend time outside of their home. Others might not have to physically be at work, but they have to take care of family members living in other households etc. Identifying these parts of the population can provide critical information on possible policies/interventions that could further increase compliance, without compromising the needs of people. Therefore, in this work we build a framework using a beta regression model to predict the percentage of time spent daily at home within a census block based on demographic characteristics. Using these models we can then examine various hypotheses on whether specific demographics of interest are associated with a change in mobility above and beyond of what was expected from the mobility patterns prior to stay-at-home orders. We focus on two particular demographics, age and race, and show that show that minorities and older people, while significantly increasing their stay at home, this increase is smaller compared to that white and younger people. We further provide some possible mechanisms that lead to this observation and show that income disparities can explain a sizable part of this difference. The main contributions of our work can be summarized as follows:

- Provide a Granger-causality analysis on the impact of stay-at-home orders on COVID-19-related fatalities
- Design a framework for quantifying adherence to social distancing according to various demographics
- Design a dynamic dashboard to visualize both the *raw* mobility data as well as, the results from our analysis.

We believe that our work can provide critical information to local officials and policy makers. The rest of the paper is organized as follows. Section 2 provides a description of the data we used for our analysis, as well as, a brief review on related to our study literature. Section 3 provides our Granger-causality analysis, while Section 4 introduces our framework for identifying the relationship between social distancing compliance and demographics. We conclude our work and discuss its limitations and directions for future work in Section 5.

2 DATA AND RELATED WORK

In this section we describe the dataset we use for our analysis, as well as, relevant to our study literature. The code for the analysis presented in the paper can be found on our github repository: https://github.com/kpelechrinis/epiDAMIK20-COVID.

SafeGraph data: SafeGraph has released a detailed mobility dataset based on the locations of about 18 million mobile phones across the US. This information is obtained through various mobile applications that partner with SafeGraph. This provides diverse population coverage, while the data are provided in an aggregated manner, with steps taken towards satisfying differential privacy requirements. While a detailed description can be found on Safe-Graph's COVID-19 data consortium page [25], the main information that we will use is the daily mobility patterns for census block groups (CBG). In particular, for each day and each census block group since 01/01/2020 we obtain - among other - the following daily information:

- `completely_home_device_count`: This is the number of devices within the CBG of interest that did not leave their home.
- `distance_traveled_from_home`: This is the median distance traveled during the day from all the devices whose home is within the CBG of interest
- `median_percentage_time_home`: This is the median percentage of time spent at home during the day from devices whose home is within the CBG of interest
- `destination_cbgs`: This is the CBGs that were visited during the day from devices whose home is within the CBG of interest. Each destination block is also associated with the number of devices in the SafeGraph dataset that performed this transition.

COVID-19 data: In order to evaluate any (Granger causal) impact between mobility and COVID-19-related fatalities we need to utilize data related to the number of confirmed cases and deaths. While an accurate number for the daily number of infections would be the most appropriate variable for this analysis, it is widely known that the reported numbers are a severe undercount of the actual number of infections. On the other hand the number of fatalities is also inaccurate but it is considered a more robust signal for the prevalence of the disease. Albeit it is a lagged signal, with an average of 15-20 days delay [15]. We obtain our data from the NY Times github repository [28].

COVID-19 and mobility: Excluding clinical interventions (potential treatments, vaccine, etc.), limiting mobility and inter-personal contacts has been the most central intervention in an effort to contain the pandemic. As such, several studies have analysed the changes in human mobility across various regions using granular mobility data (e.g., [4, 9, 13, 23] with the list being non-exhaustive). Aleta et al. [2] further utilize these mobility information to drive agent-based simulators in order to understand the impact of contact tracing and testing on a possible second wave of the disease. Zhang et al. [31] have further analyzed contact surveys from the early epidemic stage in China and built transmission models to quantify the impact of social distancing and school closures. This line of research is of course still developing as restrictions are lifted, new measures potentially coming in the possibility of a second wave etc.

Public health non-pharmaceutical evaluation: Of course, similar non-pharmaceutical interventions have been applied in the past as well and there is a volume of research that evaluates their impact. For example, Ahmed et al. [1] provide a review study on social distancing measures in workplace. Their review includes both epidemiological as well as, modeling studies and they concluded that overall workplace social distancing reduced the influenza attack rate approximately 23%. Similarly, Rashid et al. [24] reviewed studies that evaluated various measures (school closings, work-from-home etc.) for dealing with the 2009 influenza pandemic. They identified that workplace interventions provide moderate reduction in transmissions (20-30%). Other non-pharmaceutical interventions include the banning of mass events. While intuitively this seems to be a particularly effective strategy, prior literature has shown that this is true only in combination with other interventions [12, 16]. One of the reasons for this is the contact time at such events is relatively small compared to the time spent in schools, workplaces, or other community locations [7]. The literature aforementioned is not exhaustive. However, to the best of our knowledge, there is no study that uses the notion of Granger causality for non-pharmaceutical interventions. Contrary to the majority of existing studies that rely on large-scale simulation models, or, analyzing a small case (e.g., a restaurant, a specific workplace etc.), we take a macroscopic approach, looking at the aggregate adherence to these interventions and the macroscopic results (e.g., total fatalities).

3 EVALUATING SOCIAL DISTANCING

In this section we will begin by introducing the notion of Granger causality between two time series and then we will see how it applies to our case.

3.1 Granger Causality

Granger causality is a statistical test that aims at identifying whether a time-series $x(t)$ provides useful information in predicting time-series $y(t)$ [10]. It is eminent to understand that Granger causality is what Granger himself described, "temporally related" or "predictive causality", rather than the traditional notion of causality. Simply put, $x(t)$ is said to Granger-cause $y(t)$ if it precedes in time and is able to improve the predictions of $y(t)$ beyond auto-regressive models. While this might not be a useful notion for what is needed in areas like clinical treatments, it is particularly useful and has been extensively used in econometrics, public policy etc. (e.g., [3, 5, 6, 11, 18] with the list being non-exhaustive).

Formally, the examination of whether $x(t)$ Granger-causes (G-causes for short) $y(t)$ one needs to build the following two models:

$$M_0: \quad y(t) = a_{00} + \sum_{i=1}^{m} a_{0i} y(t-i) + \epsilon_0 \qquad (1)$$

$$M_1: \quad y(t) = a_{10} + \sum_{i=1}^{m} a_{1i} y(t-i) + \sum_{i=1}^{p} b_i x(t-i) + \epsilon_1 \qquad (2)$$

The first model (Eq. 1) is essentially a pure auto-regressive model on y up to lag m (called the restricted model), while the second one includes lagged terms from the time-series $x(t)$ to be explored as a potential Granger cause (called the unrestricted model). Given this setting the following null hypothesis is examined: via conducting an F-test:

$$H_0: \quad b_1 = b_2 = \cdots = b_p = 0 \qquad (3)$$

The null here is the hypothesis that no explanatory power is jointly added from the lags of x. So eventually, we retain all the lagged values of x that are individually statistically significant (t-statistic), but in order to reject H_0 that x does not G-cause y, all these lags need to add explanatory power (as compared to the restricted model). We would like to note here that the time series need to be stationary before performing the Granger test. Hence, if the original data are not stationary they should be transformed to eliminate the possibility of autocorrelation (e.g., through differentiation).

3.2 Shelter-at-home and COVID-19 fatalities

We are interested in examining whether the mobility of people in the US G-causes the number of fatalities from COVID-19. Here, we would like to emphasize that for the latter, we are using the number of COVID-19 deaths ϕ reported from health authorities as discussed in Section 2. We do not make use of any information related to excess fatalities, or any attempt to estimate the under-reporting factor in fatalities. For the G-cause variable, we first obtain the fraction of devices in each census block group b that stayed exclusively at home daily[1] h_b. We then obtain a weighted average value over all the CBGs, $h_{US}(t)$, for each day t, where the weights are the sample size in each block. We further aggregate the data weekly, since there are known inconsistencies and delays in reporting cases and deaths. Weekly aggregation should remove some of the associated noise with COVID-19 daily reports.

Figure 1 shows the two weekly time-series of interest for the period between 01/21/2020 (when the COVID-19 cases started being recorded) and 07/03/2020. We apply the Kwiatkowski–Phillips–Schmidt–Shin test [14] and we identify that these time-series are not stationary. However, differentiating both time-series will lead to stationarity. Running the Granger causality test for lags up to 6 weeks (given the length of our time-series longer lags cannot be tested), we obtain the results in Table 1. As we can see, there is evidence that mobility G-causes COVID-19-related fatalities at a lag of about 2 weeks. We also examined for bidirectional G-causality, i.e., people listening to the news and number of fatalities, and reacting with changes in their mobility. However, we did not find any supporting evidence.

Given the results from our Granger causality analysis we can build a time-series prediction model for estimating the weekly number of fatalities in the near-future (e.g., one week ahead). We

[1] We also examined the median percentage of time spent at home, with similar results. In fact, the median percentage of time spent at home and the fraction of people staying completely at home daily are highly correlated, with a correlation coefficient $\rho = 0.93$.

	Lag					
	1	**2**	**3**	**4**	**5**	**6**
b_1	263.4*	156.0	236.9*	230.9*	344.4**	289.5*
b_2	-	401.4**	432.7**	539.6**	447.8**	574.9*
b_3	-	-	348.1*	516.2*	675.9**	760.1*
b_4	-	-	-	186.7	−18.1	65.1
b_5	-	-	-	-	145.6	−64.3
b_6	-	-	-	-	-	−10.3
Adj R^2	0.46	0.69	0.77	0.79	0.91	0.9
F-test	5.08	10.3	11.1	18.9	8.5	12.6
(p-val)	(0.03)	(<0.01)	(<0.01)	(<0.01)	(<0.01)	(<0.01)

$^{**}p < 0.01, \; ^{*}p < 0.05, \; \cdot p < 0.1$

Table 1: Individual coefficients' significance and F-test result for various lags.

experiment with two different models, namely, a Vector AutoRegression (VAR) and a Long-Short Term Memory neural network. The VAR model is essentially the unrestricted model in the Granger-causality test (Equation 2), where $m = p = 3$ Table 2 shows the corresponding model. As we can see, increased fraction of people staying home will result in a reduction in the predicted number of fatalities 3 weeks ahead. We also examined a stacked LSTM architecture, with 2 layers with 50 hidden units each, followed by a dense layer with ReLU activation. We use again a sequence of size of 3 and train the model over multiple epochs using early stopping. The results from our two models are presented in Figure 2. In particular, we provide predictions for the last 5 weeks (as of this writing) and we train each model using all the data up to the week of interest. Consequently we make our out-of-sample predictions with each model which are overlaid with the actual fatalities. Overall, both models perform relatively well, especially given the short span of the time-series, as well as, the simplicity of the models in terms of input features. We would like to note here that these models are not appropriate for longer term predictions (e.g., fatality count in 4 months), which is the focus of most of the fatality-related prediction models developed (https://www.cdc.gov/coronavirus/2019-ncov/covid-data/forecasting-us.html).

4 QUANTIFYING SOCIAL DISTANCING BEHAVIOR PER DEMOGRAPHICS

In the previous section, we saw that there is strong evidence that limiting mobility Granger-causes fewer fatalities from COVID-19. Therefore, it is important to understand if and which parts of the population are not able to *adhere* to the guidelines. This information is critical to be communicated to health officials and policy makers, since it can drive interventions that will help everyone follow the recommendations to the extent possible. In this section, we present a framework based on a beta regression model from the daily percentage of time spent home and the difference-in-differences method that can identify the relationship between demographics of interest and the way they relate to social distancing behavior.
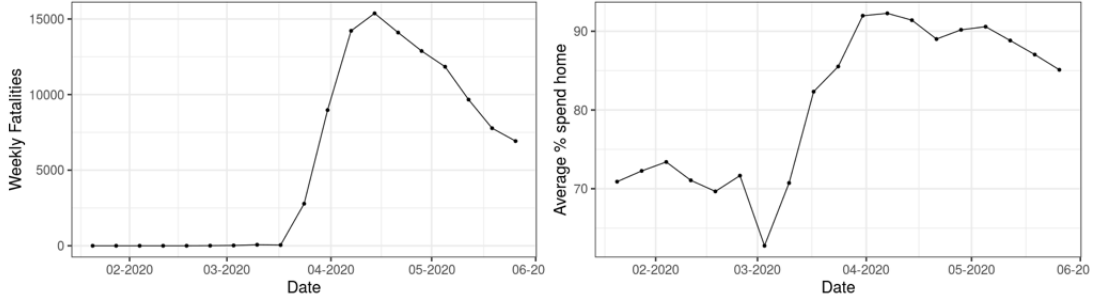
Figure 1: Weekly time series for COVID-19-related fatalities (left) and percentage of people staying at home (right).
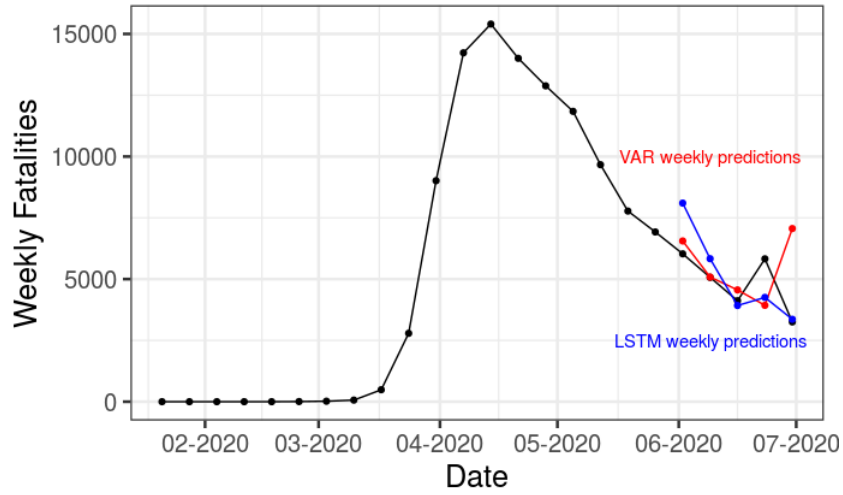

Figure 2: Both models perform reasonably out-of-sample, relative to the amount of data available for learning and the simplicity of the input features.

Variable	Coefficient	p-val
$h_{US}(t-1)$	137.9	0.21
$h_{US}(t-2)$	252.4	0.15
$\mathbf{h_{US}(t-3)}$	−384.7	< 0.01
y(t-1)	1.35	< 0.01
y(t-2)	−0.59	0.16
y(t-3)	0.19	0.40
R^2	0.86	
SE_{res}	1250	

$^{***}p < 0.01,\ ^{**}p < 0.05,\ ^{*}p < 0.1$

Table 2: VAR model for predicting weekly fatalities one-week-out.

4.1 Beta regression model

Our goal is to model the percentage of time $h_{\mathcal{P}}$ that a specific population $\mathcal{P}$ spends home daily. Given that our dependent variable $h_{\mathcal{P}}$ is real-valued, bounded in the unit interval a linear regression is not an appropriate model. Hence, we choose to use a beta regression model [8], where essentially the data are assumed to follow a beta distribution. A useful parametrization of the beta distribution for this type of models is given by:

$$f(y|\mu, \phi) = \frac{\Gamma(\phi)}{\Gamma(\mu\phi) \cdot \Gamma((1-\mu)(\phi))} y^{\mu\phi-1}(1-y)^{(1-\mu)\phi-1} \quad (4)$$

where μ is the mean of the beta distribution and ϕ is a parameter called precision. ϕ *controls* the variance of the distribution; for a fixed μ, higher precision leads to smaller variance. With this setting the beta regression model for $h_{\mathcal{P}}$ is:

$$g(\overline{h}_{\mathcal{P}}) = \mathbf{x}_{\mathcal{P}}^T \cdot \mathbf{b} + \epsilon \quad (5)$$

where $\overline{h}_{\mathcal{P}}$ is the average daily fraction of time spent home for $\mathcal{P}$, $\mathbf{x}_{\mathcal{P}}$ is the vector of the model's covariates, $\mathbf{b}$ is the vector of the regression's coefficients and $g(\cdot)$ is a link function (strictly increasing and twice differentiable). This model is very similar to a generalized linear model (e.g., logistic, Poisson or negative

binomial regression) and it is solved through a Maximum Likelihood Estimation (MLE). The MLE identifies the coefficients **b**, but also the precision parameter ϕ, which is a constant and not a function of $\mathbf{x}_{\mathcal{P}}$ [2].

4.2 Demographics Analysis

In this section we will begin by modeling the fraction of time spent at home daily in each census block as a function of specific demographics of the population. We start with race, where census data provide information on the percentage of people within each census block that belong to the following categories: White, Black, Hispanic, Asian, American Indian or Native Alaskan, and Other races [3]. Since we want to estimate the relationship between these demographics and the changes observed after the social distancing recommendations, we build two separate models; one that captures the mobility prior to stay-at-home orders (M_{pre}) and one that captures mobility after these orders were put in place (M_{post}). One of the problems is that different parts of the country put these measures in place in different times through the course of the pandemic. Given that the majority of the orders were put in place sometime within March 2020, we build M_{pre} using data from February 2020, and M_{post} using data from April 2020. Table 3 presents the results of these regressions. Using these results we can start examining the average percentage of time spent daily at home by the population of a hypothetic census block group (HCBG) with a specific racial demographic composition. For example, Figure 3 presents the beta distribution for racially homogeneous (hypothetical) census block groups. As we can see, there are differences across these hypothetical census block groups, both for the same time period, as well as, their shift as the stay-at-home orders were put in place. More specifically, Table 4 presents the average stay home percentage for each of the hypothetical blocks.

Variable	M_{pre}	M_{post}
White%	−0.45***	−0.48***
Black%	−0.27***	−0.56***
Hispanic%	0.29***	0.39***
Asian%	−0.40***	1.87***
Natives+Others%	−0.51***	−0.93***
constant	1.39***	2.5***
ϕ	14.5	5.8
N	201,917	201,917

***$p < 0.01$, **$p < 0.05$, *$p < 0.1$

Table 3: Beta regression model for the average daily percentage of time of stay home at a census block group before (02/20) and after (04/20) stay-at-home orders.

Table 4, while providing us with a quick view of how specific parts of the population might comply with the social distancing recommendations (in terms of staying home), it does not provide the

[2]There are extensions of this model [26] that models the precision as a function of a set of regressions **z**, i.e., $g'(\phi) = \mathbf{z}^T \cdot \mathbf{c} + \epsilon$.
[3]For the purposes of our analysis we merge the American Indian and Native Alaskan category with the Other races category.

Hypothetical Block	Pre	Post
White	71.8%	89.6%
Black	75.6%	88.6%
Hispanic	84.4%	94.9%
Asian	72.9%	98.9%
Natives+Others	70.7%	84.8%

Table 4: Percentage of time spent home daily for hypothetical racially homogeneous census block groups based on the beta regression models from Table 3.

whole picture. In particular, we can see that different demographics are *associated* with different levels of mobility outside of their home even before the stay-at-home order. So any change observed after the orders were put in place, they need to be compared with the original difference. This process is visualized in Figure 5, where we see two populations $\mathcal{P}_1$ and $\mathcal{P}_2$, with their pre-lockdown daily percentage of staying home, as well as, their post-lockdown daily percentage of staying home. While Δ_2 provides us with information about what is happening in the two populations after the stay-at-home orders were put in place, it does not adjust for the *behavior* of the two populations prior to the intervention, and the difference $\delta(\mathcal{P}_1, \mathcal{P}_2) = \Delta_2 - \Delta_1$ is more informative. Hence, in order to identify demographic discrepancies between two populations, $\mathcal{P}_1$ and $\mathcal{P}_2$, in complying with stay-at-home orders, we performed the following hypothesis test:

$$H_0: \quad \delta(\mathcal{P}_1, \mathcal{P}_2) = 0 \quad (6)$$
$$H_1: \quad \delta(\mathcal{P}_1, \mathcal{P}_2) \neq 0 \quad (7)$$

In order to perform this test, we use the full beta distribution for each population-time combination and repeatedly sample them to build the distribution of $\delta(\mathcal{P}_1, \mathcal{P}_2)$. Then we can perform the above hypothesis test. Table 5 presents the results for the various comparisons between the minority HCBG and the white one. As we can see all minority HCBG - except the Asian one - exhibit a smaller increase as what was expected based on their pre-intervention patterns. Particularly interesting is the case of the Hispanic HCBG, which even though exhibits the second highest daily percentage of staying home after the stay-at-home orders, the observed increase is smaller as compared to the white HCBG. Furthermore, it is interesting that the Asian HCBG exhibits a 7.5% higher *compliance* as compared to the white HCBG. While the reasons for this are not clear - and we cannot identify them through the data we have - there are a few reasons that are plausible, including the increase of racist attacks targeting Asians in the US at the wake of the pandemic [19–22, 29, 30].

While for the Asian population, staying at home more might also be a way of avoiding racist attacks, the question remains, why are there discrepancies for the rest of the minorities as compared to the white HCBG ? One plausible explanation is that a large fraction of these minorities are essential workers and while overall they increase their stay at home, they really need to go to their work. Another possible reason is that minorities live in inner cities and as such they are close to their families. Furthermore, these minorities
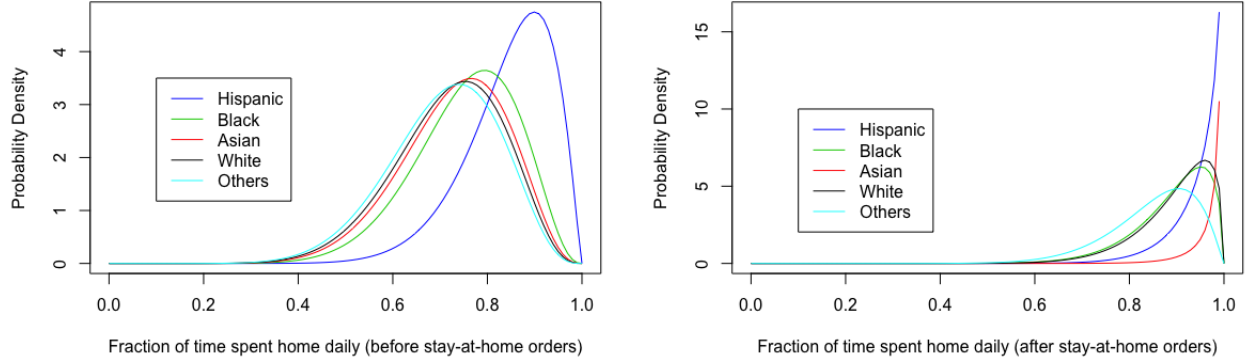
Figure 3: Beta distribution for hypothetical - racially homogeneous - census block group before (left) and after (right) stay-at-home orders across the US.

$\mathcal{P}_1$	$\mathcal{P}_2$	$\delta(\mathcal{P}_1, \mathcal{P}_2)$
Black	White	-4.8%***
Hispanic	White	-6.2%***
Asian	White	7.5%***
Natives+Others	White	-3.6%***

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 5: Minority HCBGs exhibit lower percentage of stay-at-home, as compared to white HCBGs.

Variable	M_{pre}	M_{post}
White%	-0.43***	-0.61***
Black%	-0.29***	-0.3***
Hispanic%	0.27***	0.7***
Asian%	-0.29***	0.87***
Natives+Others%	-0.52***	-0.79***
Median Income	$-9.9 \cdot 10^{-7}$ ***	$9.9 \cdot 10^{-6}$ ***
constant	1.43***	2.13***
ϕ	14.6	6.34
N	201,917	201,917

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 6: Beta regression model for the average daily percentage of time of stay home at a census block group before (02/20) and after (04/20) stay-at-home orders adjusting for median income (expressed in thousands of $s) in the CBGs.

$\mathcal{P}_1$	$\mathcal{P}_2$	$\delta(\mathcal{P}_1, \mathcal{P}_2)$
Black	White	$6 \cdot 10^{-4}\%$
Hispanic	White	-4.3%***
Asian	White	5.7%***
Natives+Others	White	$-5 \cdot 10^{-3}\%$

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 7: When adjusting for income a large percentage of the mobility differences between HCBGs during stay-at-home orders is explained.

have come to rely and support their extended families [27] and hence, they might be providing them with help (e.g., childcare support for essential workers etc.) during this time, leading to higher mobility outside the home. Other plausible reasons include the relationship between these groups and technology. In particular, ethnic minorities have traditionally been slower in adopting new technology for a variety of reasons [17] and this could mean in a situation like the current pandemic, their inability or unwillingness to use online platforms for essential errands such as grocery shopping. While we cannot show with our current data whether any of these plausible reasons are in play, we can examine one additional factor that is relevant to all of the above possibilities; their median income. Low income families typically live in inner-city and are of ethnic minorities, they have issues with accessing and adopting technology, while many of the essential workers are low-paid employees (e.g., grocery store workers, delivery, etc.). Tables 6 and 7 present the same results when we adjust for the median income of an HCBG. As we can see, the mobility differences between black and white HCBGs, as well as native and other minorities and white HCBGs, disappears, while for Hispanic and Asian HCBGs the differences are reduced.

We also examined another demographic attribute, namely, age. While census provides a breakdown of the age of a census block group in several age brackets, we aggregated them into two bins; younger or older than 50 year old[4]. Again, we build a beta regression model with the same dependent variable as before but the

---

[4]Obviously one can repeat the analysis with more bins, but we wanted to keep things simpler mainly for presentation purposes.

independent variable is the percentage of the population in the CBG that is older than 50 years old. The results are presented in Table 8, where as we can see the older population is associated with a reduced stay-at-home daily time as compared to younger population (less than 50). Figure 4 further visualizes the beta distributions for hypothetical CBGs with only population older or younger than 50 years old. Furthermore, by performing a similar hypothesis test as in Eq. (6)-(7), we find that the HCBG with population older than 50 years old stays at home 2.6% (p-val < 0.01) **less** time at home on average as compare to younger population and based on their pre-intervention patterns. In contrast to the race case, when adjusting for the median income, the difference remains (-2.5%, p-val < 0.01). A potential reason for this difference between population in the opposite side of the 50 years old mark, can be their *technology fluency*. Younger people that are avid users of (mobile) technology can take advantage of various services that can help people complete their errands (e.g., grocery shopping), while staying at home. This might not be the case for older people (at least to the same extent). Again, while this is a plausible mechanism that can drive the observed difference, the data in our disposal does not allow us to further examine this.

Variable	M_{pre}	M_{post}
Older50%	0.09***	−0.28***
constant	0.98***	2.39***
ϕ	14.2	5.5
N	201,917	201,917

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 8: Beta regression model for the average daily percentage of time of stay home at a census block group before (02/20) and after (04/20) stay-at-home orders based on the percentage of the population that are older than 50 years old.

4.3 Dashboard

We have also created a dashboard to visualize this mobility information in an interactive manner[5]. Figure 6 presents a screenshot from the dashboard that depicts the census block tracts of Allegheny County on the left half. The user can choose a tract (the selected tract will be colored red as in the figure) and information about the outgoing mobility (i.e., movements of people whose home CBG is the selected one) and incoming mobility (i.e., movements from people whose home CBG is not the selected one but they visited it) associated with it is presented. The choice between outgoing and incoming mobility can be made through the control buttons above the map. For example, in Figure 6 outgoing mobility information for people whose home CBG is the selected origin CBG (420035231001) is presented on the map. The color for each census block group tract i represents the fraction of the total foot traffic from the residents of the origin CBG, over the period selected from the user[6], that visited CBG i. On the right half of the figure, there
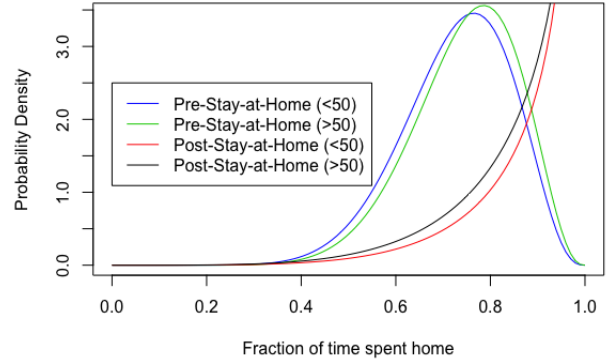

Figure 4: Beta distributions for the daily fraction of time spent at home for population older and younger than 50.
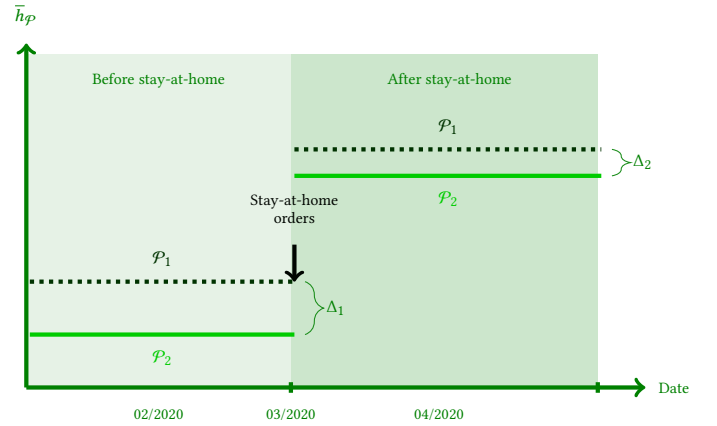

Figure 5: When comparing the mobility post-lockdown for different populations, we need to consider the pre-lockdown mobility as well.

are two time-series depicted that provide temporal information for the CBG that the user is currently hovering over (say CBG_h). In the specific situation depicted here, this is CBG 420035231001. The top time series provides the daily number of visits in CBG_h from the origin CBG, while the bottom time series represents the fraction of time residents of CBG_h spent at home. It is interesting here to note that if we hover over the origin CGB, i.e., CBG_h is the selected CBG, then the top time-series represents *self-loops*. That is, traffic from residents of the CBG that was destined to other venues/points of interest within the CBG. Finally, we also present a table with some basic demographic information about the origin CBG related to our analysis, such as racial and age composition of the population, median income and total population.

We would also like to note here that this dashboard is still *work-in-progress* in the sense that new features are being added prior

---

[5]http://mobility.pittsmartliving.com/.
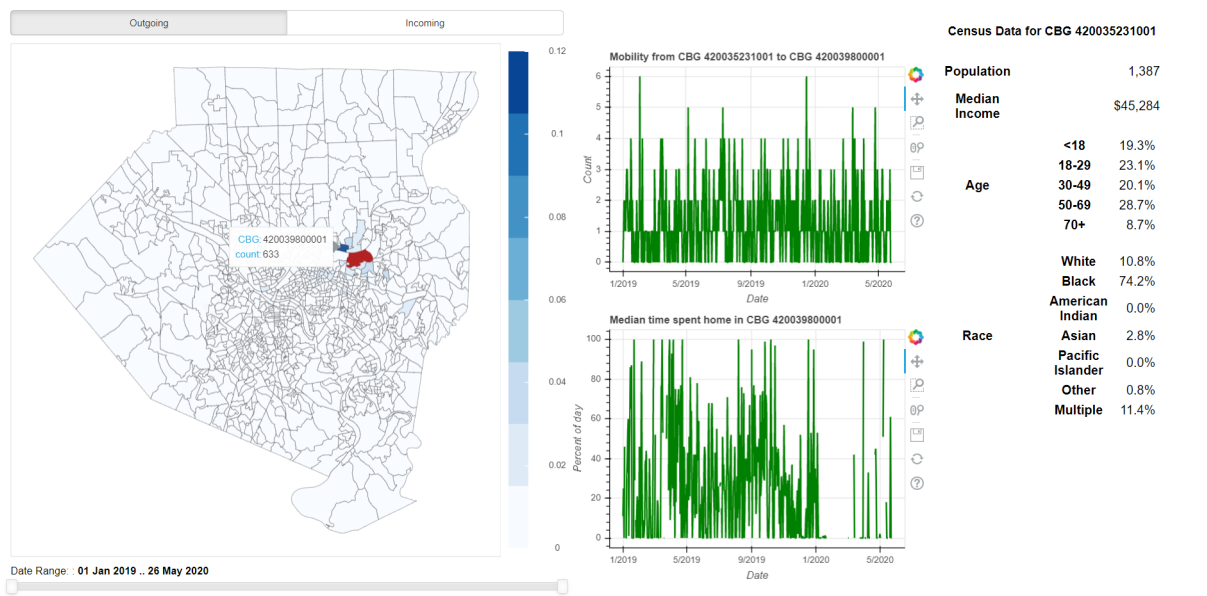[6]The user can select the time period through the slider under the map.

Figure 6: Our dashboard for Allegheny County showing the outgoing mobility from the selected CBG (red) alongside with demographic information. The two time-series plots further provide information related to the interaction between the selected CBG and another CBG that the user hovers over.

to going publicly live. For example, our immediate future plan is to visualize information about specific businesses and their geographical reach (i.e., where do customers of different establishments come from?). This information can be very helpful for local health authorities when identifying a plan for interventions and the corresponding protocols.

5 CONCLUSIONS AND DISCUSSION

In this study we perform a macroscopic analysis of the effectiveness of social distancing measures in the US during the COVID-19 pandemic using the notion of Granger causality. Our analysis indicate that the average daily fraction of population staying completely at home Granger-causes the number of COVID-19 fatalities in a 3-week period. We further examine the presence of bidirectional Granger causality and we do not find any supporting evidence. Using this observation, we also build two simple prediction models for weekly COVID-19-related fatalities, using auto-regressive and mobility features. We further provide a framework to identify the relationship between demographics and social distancing behavior. While this analysis does not provide causal relationships, it can certainly provide important information for policy makers while thinking of ways to increase *compliance*. Finally, we provide a visualization dashboard with the raw data as well as, the results from our analysis. This dashboard is constantly being updated with new results and data.

We would like to emphasize here that even though we have included a prediction model in our analysis, this is only to showcase in practise the conclusions from the Granger causality analysis[7].

---

[7]Furthermore, there are several well-performing prediction models in the public sphere - tracked by CDC as well - and our goal is certainly not to add yet another model.

Furthermore, while the model performs well out-of-sample, several improvements can be achieved by including even more informative features. For instance, just an aggregate number of how many hours a person spends out of their home does not capture factors important for the prediction of infections. Was this movement to a high-risk location (e.g., a grocery store) or was it for a stroll around the neighborhood? Disentangling this is certainly not trivial and we are working in methods for identifying the number of potential contacts a person from a specific CBG is expected to have based on their mobility and the POI foot traffic data. Furthermore, it will be particularly useful to extend our analysis to a more (spatially) fine granularity, focusing on a microscopic analysis (e.g., at the county, or city, level). This will allow us to identify the exact time points of interventions and possibly attempt to extract causal relationships using quasi-experimental methods, such as instrumental variables and difference-in-differences.

REFERENCES

[1] Faruque Ahmed, Nicole Zviedrite, and Amra Uzicanin. 2018. Effectiveness of workplace social distancing measures in reducing influenza transmission: a systematic review. *BMC public health* 18, 1 (2018), 518.

[2] Alberto Aleta, David Martin-Corral, Ana Pastore y Piontti, Marco Ajelli, Maria Litvinova, Matteo Chinazzi, Natalie E Dean, M Elizabeth Halloran, Ira M Longini, Stefano Merler, et al. 2020. Modeling the impact of social distancing, testing, contact tracing and household quarantine on second-wave scenarios of the COVID-19 epidemic. *medRxiv* (2020).

[3] Mak B Arvin, Rudra P Pradhan, and Neville R Norman. 2015. Transportation intensity, urbanization, economic growth, and CO2 emissions in the G-20 countries. *Utilities Policy* 35 (2015), 50–66.

[4] J. Bayham, J. Adams, D. Ghosh, and P. Jackson. 2020. Colorado Mobility Patterns During the COVID-19 Response. *Technical Report* (2020).

[5] Johan Bollen, Huina Mao, and Xiaojun Zeng. 2011. Twitter mood predicts the stock market. *Journal of computational science* 2, 1 (2011), 1–8.

[6] Willem H Buiter. 1984. Granger-causality and policy effectiveness. *Economica* 51, 202 (1984), 151–162.

[7] Neil Ferguson, Daniel Laydon, Gemma Nedjati Gilani, Natsuko Imai, Kylie Ainslie, Marc Baguelin, Sangeeta Bhatia, Adhiratha Boonyasiri, ZULMA Cucunuba Perez, Gina Cuomo-Dannenburg, et al. 2020. Report 9: Impact of non-pharmaceutical interventions (NPIs) to reduce COVID19 mortality and healthcare demand. (2020).

[8] Silvia Ferrari and Francisco Cribari-Neto. 2004. Beta regression for modelling rates and proportions. *Journal of applied statistics* 31, 7 (2004), 799–815.

[9] Song Gao, Jinmeng Rao, Yuhao Kang, Yunlei Liang, and Jake Kruse. 2020. Mapping county-level mobility pattern changes in the United States in response to COVID-19. *Available at SSRN 3570145* (2020).

[10] Clive WJ Granger. 1969. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: journal of the Econometric Society* (1969), 424–438.

[11] MV Hood, Quentin Kidd, and Irwin L Morris. 2008. Two sides of the same coin? Employing Granger causality tests in a time series cross-section framework. *Political Analysis* 16, 3 (2008), 324–344.

[12] David A Ishola and Nick Phin. 2011. Could influenza transmission be reduced by restricting mass gatherings? Towards an evidence-based policy framework. *Journal of epidemiology and global health* 1, 1 (2011), 33–60.

[13] Brennan Klein, T LaRocky, S McCabey, L Torresy, Filippo Privitera, Brennan Lake, Moritz UG Kraemer, John S Brownstein, David Lazer, Tina Eliassi-Rad, et al. 2020. Assessing changes in commuting and individual mobility in major metropolitan areas in the United States during the COVID-19 outbreak.

[14] Denis Kwiatkowski, Peter CB Phillips, Peter Schmidt, Yongcheol Shin, et al. 1992. Testing the null hypothesis of stationarity against the alternative of a unit root. *Journal of econometrics* 54, 1-3 (1992), 159–178.

[15] Stephen A Lauer, Kyra H Grantz, Qifang Bi, Forrest K Jones, Qulu Zheng, Hannah R Meredith, Andrew S Azman, Nicholas G Reich, and Justin Lessler. 2020. The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: estimation and application. *Annals of internal medicine* 172, 9 (2020), 577–582.

[16] Howard Markel, Harvey B Lipman, J Alexander Navarro, Alexandra Sloan, Joseph R Michalsen, Alexandra Minna Stern, and Martin S Cetron. 2007. Non-pharmaceutical interventions implemented by US cities during the 1918-1919 influenza pandemic. *Jama* 298, 6 (2007), 644–654.

[17] Karen Mossberger, Caroline J Tolbert, Daniel Bowen, and Benedict Jimenez. 2012. Unraveling different barriers to Internet use: Urban residents and neighborhood effects. *Urban Affairs Review* 48, 6 (2012), 771–810.

[18] Paresh Kumar Narayan and Stephan Popp. 2012. The energy consumption-real GDP nexus revisited: Empirical evidence from 93 countries. *Economic Modelling* 29, 2 (2012), 303–308.

[19] NBC News. 2020. Smashed windows and racist graffiti: Vandals target Asian Americans amid coronavirus. https://www.nbcnews.com/news/asian-america/smashed-windows-racist-graffiti-vandals-target-asian-americans-amid-coronavirus-n1180556

[20] NPR. 2020. New Site Collects Reports Of Racism Against Asian Americans Amid Coronavirus Pandemic. https://www.npr.org/sections/coronavirus-live-updates/2020/03/27/822187627/new-site-collects-reports-of-anti-asian-american-sentiment-amid-coronavirus-pand

[21] PBS. 2020. Asian Americans describe ́gut punch ́of racist attacks during coronavirus pandemic. https://www.pbs.org/newshour/nation/asian-americans-describe-gut-punch-of-racist-attacks-during-coronavirus-pandemic

[22] The Washington Post. 2020. As the coronavirus spreads, so does online racism targeting Asians, new research shows. https://www.washingtonpost.com/technology/2020/04/08/coronavirus-spreads-so-does-online-racism-targeting-asians-new-research-shows/

[23] Giulia Pullano, Eugenio Valdano, Nicola Scarpa, Stefania Rubrichi, and Vittoria Colizza. 2020. Population mobility reductions during COVID-19 epidemic in France under lockdown. *Epicx-lab Technical Report* (2020).

[24] Harunor Rashid, Iman Ridda, Catherine King, Matthew Begun, Hatice Tekin, James G Wood, and Robert Booy. 2015. Evidence compendium and advice on social distancing and other related measures for response to an influenza pandemic. *Paediatric respiratory reviews* 16, 2 (2015), 119–126.

[25] SafeGraph. 2020. SafeGraph COVID-19 Data Consortium. https://www.safegraph.com/covid-19-data-consortium

[26] Alexandre B Simas, Wagner Barreto-Souza, and Andréa V Rocha. 2010. Improved estimators for a general class of beta regression models. *Computational Statistics & Data Analysis* 54, 2 (2010), 348–366.

[27] Robert Joseph Taylor, Linda M Chatters, Amanda Toler Woodward, and Edna Brown. 2013. Racial and ethnic differences in extended family, friendship, fictive kin, and congregational informal support networks. *Family relations* 62, 4 (2013), 609–624.

[28] The New York Times. 2020. COVID-19 Data. https://github.com/nytimes/covid-19-data

[29] The New York Times. 2020. How Asian-American Leaders Are Grappling With Xenophobia Amid Coronavirus. https://www.nytimes.com/2020/03/29/us/politics/coronavirus-asian-americans.html

[30] USA Today. 2020. We just want to be safe: Hate crimes, harassment of Asian Americans rise amid coronavirus pandemic. https://www.usatoday.com/story/news/politics/2020/05/20/coronavirus-hate-crimes-against-asian-americans-continue-rise/5212123002/

[31] Juanjuan Zhang, Maria Litvinova, Yuxia Liang, Yan Wang, Wei Wang, Shanlu Zhao, Qianhui Wu, Stefano Merler, Cécile Viboud, Alessandro Vespignani, et al. 2020. Changes in contact patterns shape the dynamics of the COVID-19 outbreak in China. *Science* (2020).