Staying at Home: Mobility Effects of Covid-19*

Sam Engle[†], John Stromme[‡], Anson Zhou[§]

April 2020

Abstract

We combine GPS data on changes in average distance traveled by individuals at the county level with COVID-19 case data and other demographic information to estimate how individual mobility is affected by local disease prevalence and restriction orders to stay-at-home. We find that a rise of local infection rate from 0% to 0.003%[1] is associated with a reduction in mobility by 7.44%. An official stay-at-home restriction order corresponds to reducing mobility by 6.71%. Counties with larger shares of population over age 65, lower share of votes for Trump in 2016, and higher population density are modestly more responsive to disease prevalence and restriction orders.

Keywords: COVID-19, Location Data, Restriction Order

---

*Working paper: Circulated for discussion and comment purposes. This paper has not been peer-reviewed.

[†]University of Wisconsin–Madison — `sengle2@wisc.edu`

[‡]University of Wisconsin–Madison — `jstromme@wisc.edu`

[§]University of Wisconsin–Madison — `anson.zhou@wisc.edu`

[1]This is the median infection rate (number of confirmed cases divided by county population) across counties with positive number of confirmed cases as of 3/20/2020.

1 Introduction

In the face of the rapidly growing threat posed by the COVID-19 pandemic, public health experts and economists alike are relying on epidemic models to make predictions and evaluate policies. In the standard SEIR model (e.g. Wang et al. (2020)), the effective reproduction rate R_t measures the actual average number of secondary cases per infected case at time t. It is widely acknowledged that R_t reflects both the nature of the virus (including the basic reproduction rate R_0), as well as the effectiveness of various protective measures taken by individuals and governments in response to available information. In the case of COVID-19, the key policy measure to reduce R_t is a restriction order to stay-at-home. To date, this policy has been promoted by governments across the globe. It is an open question, however, to what extent individuals alter their mobility in response to government orders and the increasing prevalence of coronavirus.

Mobility statistics provide invaluable information as to whether people are actively reducing their exposure to COVID-19 by reducing distances traveled and avoiding social contact. In this paper, we use a novel dataset from Unacast, a location data firm, to estimate how changes in distance traveled (Δ_{it}) is related to perceived local disease prevalence (Ω_{it}) and restriction orders I_{it}. We also investigate how these relationships depend on demographic characteristics (X_i). The methodology here is similar to that in Auld (2006), where the author estimates elasticities of risky behavior to local prevalence of AIDS, and explored heterogeneity across observable characteristics.

The estimates obtained here contribute to the current discussion in three ways. First, our results provide an estimate of how much human behavior, in our case average distance traveled, responds to perceived local disease prevalence. Second, the results give us a sense of how important government announcements are in affecting people's behaviour. Lastly, by considering demographics, political attitude and population density, we evaluate whether characteristics of the underlying population play a role in determining the effectiveness of restriction orders and responses to disease prevalence. In particular, since older individuals are at higher risk[2], we would be interested in whether counties with relatively high elderly populations have altered their behavior more than younger counties. There has also been some discussion that political partisanship is an indicator of skepticism in the legitimacy of the COVID-19 outbreak [3].

In the rest of the paper, we outline a simple model in section 2. In 3 we discuss our novel data source on daily travel patterns and how we have augmented it with COVID-19 data. In 4 we present our preliminary results and argue that even the simple model provides a solid baseline. We finish with 5 where we summarize our current progress and outline our plan for current and future work.

2 Simple Model

In this section, we present a simple model that relates an individual's travel decision to perceived disease prevalence. This model provides theoretical motivation for the estimation strategy in Section 4.

Consider an individual that derives utility $U(d) = d^\sigma/\sigma$ from distance traveled (d) with $0 < \sigma < 1$. The cost of traveling each unit of distance is composed of one component that is independent of the epidemic Π, and one component that is the product of perceived disease prevalence level Ω and the utility cost of contracting the disease Z. An individual's utility is given by:

$$\begin{aligned} U(d) &= d^\sigma/\sigma - \Pi d - \Omega Z d \\ &= d^\sigma/\sigma - \Pi\left(1 + \Omega\frac{Z}{\Pi}\right)d \\ &\approx d^\sigma/\sigma - \Pi e^{\frac{Z}{\Pi}\Omega} d \end{aligned} \tag{1}$$

[2]See, e.g. the CDC guidelines: https://www.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/people-at-higher-risk.html

[3]e.g. https://www.nytimes.com/interactive/2020/03/21/upshot/coronavirus-public-opinion.html

The solution of the utility maximization problem is therefore:

$$d^* = \left(\Pi e^{\frac{Z}{\Pi}\Omega}\right)^{\frac{1}{\sigma-1}} \tag{2}$$

As can be seen, when an individual perceives a higher risk level, corresponding to higher Ω, individuals decide to travel less, i.e. by reducing d. We will carefully define Ω later.

If we compare an individual's decision to travel at time t under unit cost Ω_t versus some benchmark date t_0 with Ω_0, we get a measure of change in distance traveled:

$$\Delta_t = \frac{d_t^*}{d_0^*} - 1 = e^{\frac{Z}{(\sigma-1)\Pi}(\Omega_t - \Omega_0)} - 1 \equiv e^{\beta_0 + \beta_1 \Omega_t} - 1 \tag{3}$$

Equation (3) is suggestive of a strategy to estimate how perceived disease prevalence, Ω, affects the percentage change in distance traveled from date t relative to date 0. We propose to estimate β_1 via nonlinear least squares, after we consider an appropriate definition of Ω_t below in Section 4.[4]

3 Data

We construct a county-level panel data for the contiguous United States. with dates covering 2/24/2020 to 3/25/2020. Our data includes the following information:

1. Daily confirmed coronavirus cases compiled by The New York Times.[5]

2. Daily changes in average distance traveled relative to the same weekday pre-COVID-19, provided by Unacast. Unacast use GPS signals from mobile devices to calculate average distance traveled by device-holders in each county at a daily frequency.[6]

[4]Since change in distance traveled (Δ_t) is large in the data, we do not approximate it by $\log(d_t^*/d_0^*)$.

[5]We also compared this data to case data compiled by the Johns Hopkins University Center for Systems Science and Engineering and found the data to be essentially identical. Our results are robust to both sources.

[6]For more information on methods of data collection and aggregation, visit `unacast.com` and Unacast COVID-19 Social Distancing Scoreboard.

3. Enacted social-distancing policies (stay-at-home restriction orders) as of 3/28/2020 as compiled by the New York Times.

4. Demographic data is sourced from the MIT Election Data and Science Lab (MEDSL). MEDSL data conveniently matches demographic information from the 2012-2016 5-year ACS, to county-level 2016 Presidential Election Results.

In total, our data covers 3142 U.S. counties with 94,116 observations. The summary statistics are given in Table 1.

Figure 1 plots the changes in average distance traveled relative to the same weekday pre-COVID-19 on 2/24/2020. The overall light color in the figure indicates that at the beginning of the epidemic when there were very few cases confirmed (see Figure 3), there was not much change in population mobility.

When we turn to a more recent date, 3/23/2020, Figure 2 shows that average distance traveled decreases significantly in most counties across the U.S., with particularly large drop in New York, California, Colorado and Florida. Figure 4 shows that these are also places with relatively large number of reported COVID-19 cases.

Figure 5 and 6 show the share of counties and population that is under a stay-at-home order respectively. Both measures start to grow on 3/19 as national cases surpass 10,000. As of 3/25/2020, more than 30% of the counties and 55% of the national population is under government orders to stay-at-home unless for essential activities.

Figure 7 shows the 10th quantile, median, and 90th quantile of the changes in average distance traveled across counties in our sample. We can see that mobility starts to decrease for median counties at around 3/10, well before the announcements of restriction orders as shown in Figure 5 and 6.

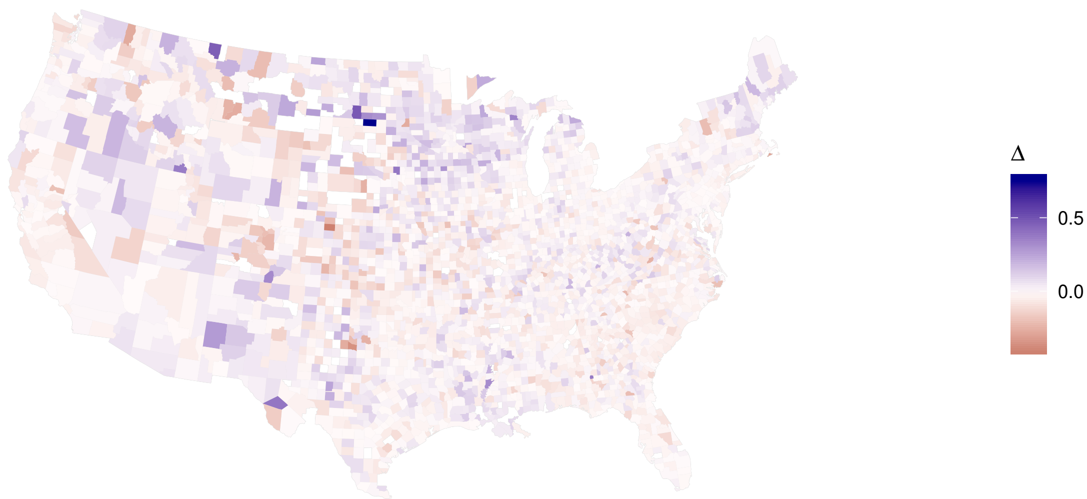

Figure 1: Change in distance traveled relative to the same weekday pre-COVID-19, 2/24/2020

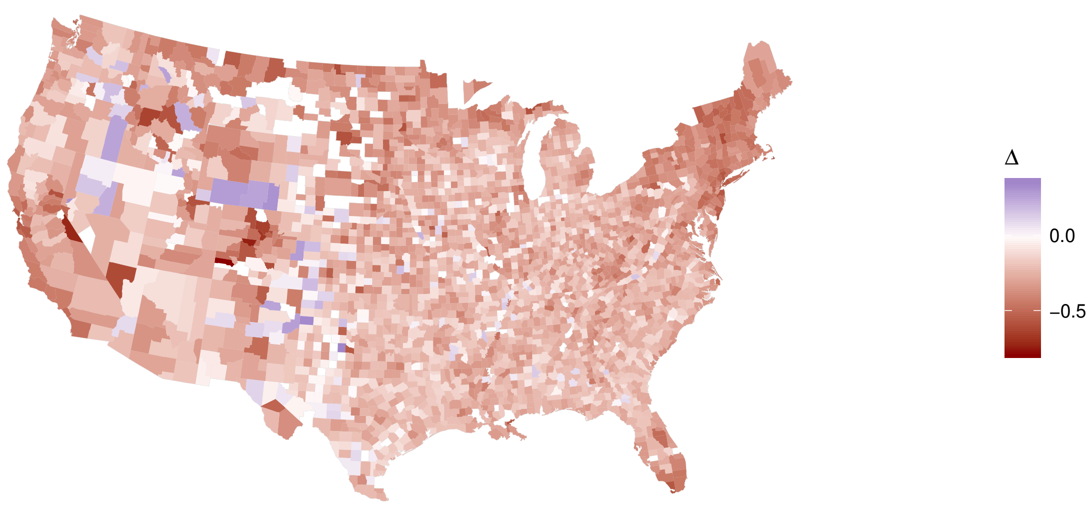

Figure 2: Change in distance traveled relative to the same weekday pre-COVID-19, 3/23/2020

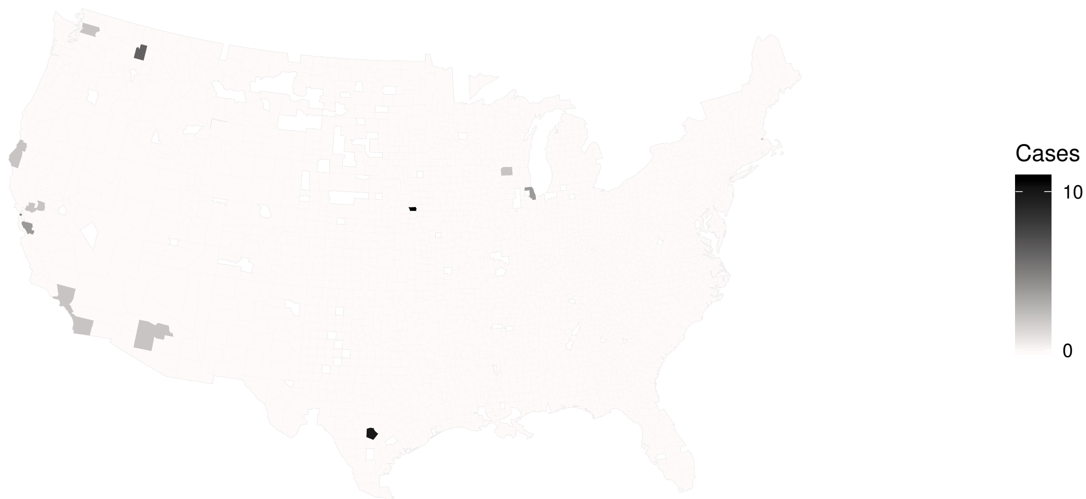

Figure 3: Number of confirmed cases, 2/24/2020

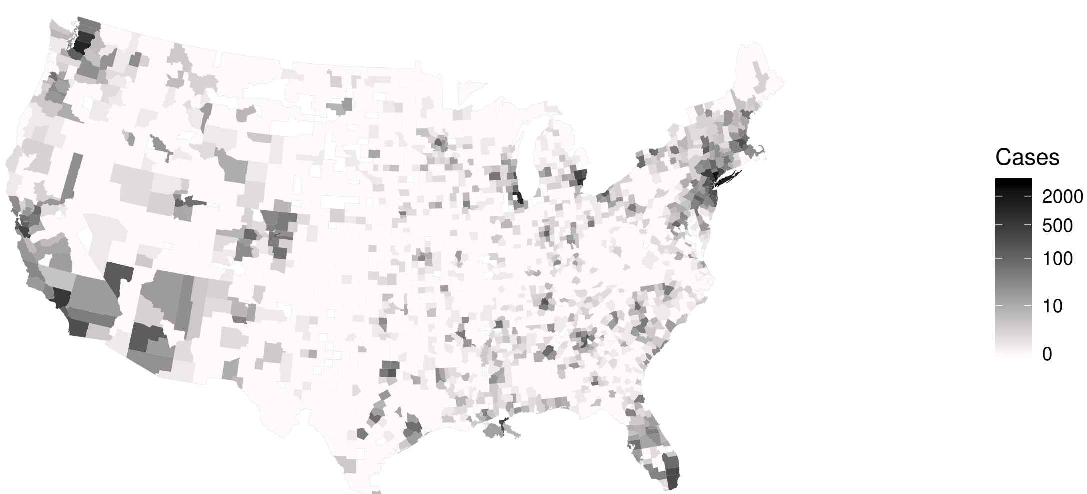

Figure 4: Number of confirmed cases, 3/23/2020

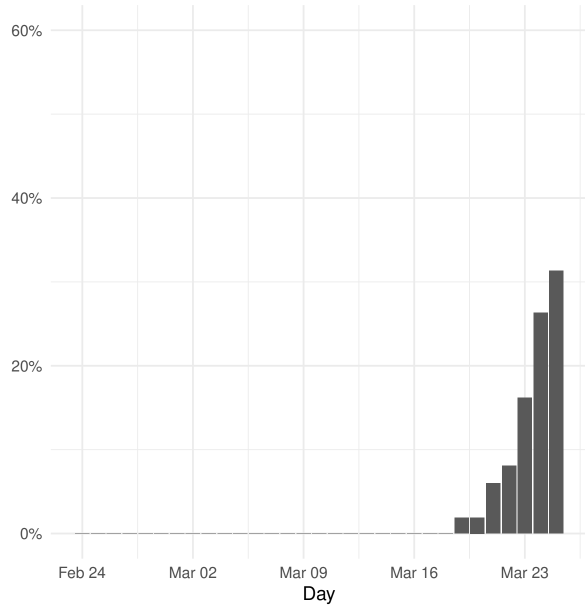

Figure 5: Share of Counties under Stay-at-Home

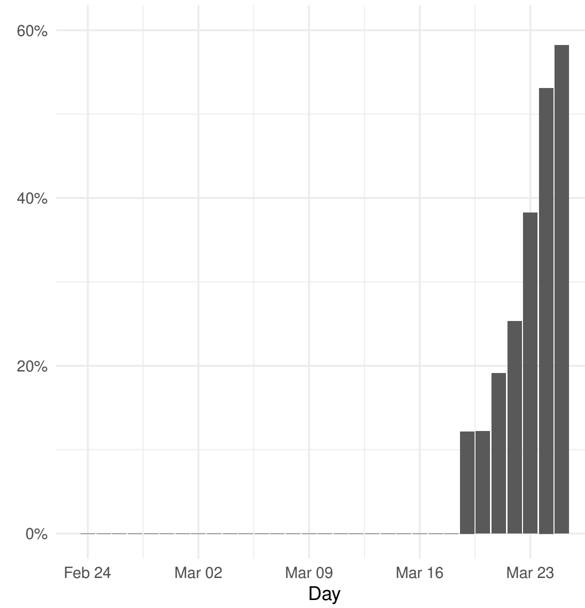

Figure 6: Share of Pop under Stay-at-Home

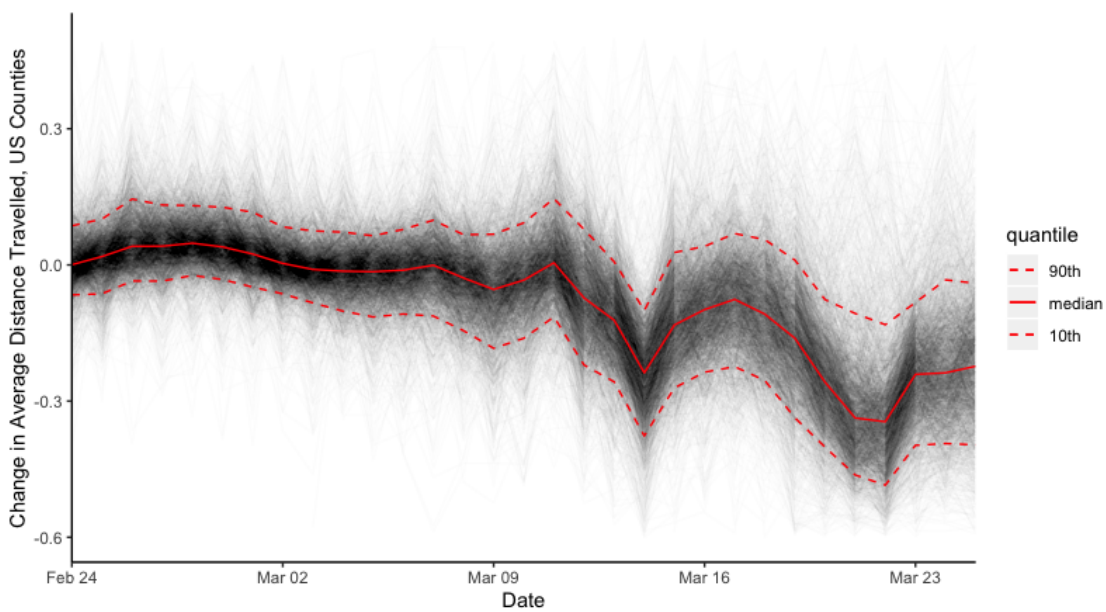

Figure 7: Quantiles of Changes in Average Distance Traveled Across Counties

Table 1: Summary Statistics of Pooled Dataset

Statistic	Mean	St. Dev.	Min	Max
Total Cases	3.004	61.546	0.000	6,154.000
Cases, Share of Pop	0.001%	0.01%	0.00%	0.5%
Neighbor Cases, Weighted	0.005%	0.01%	0.00%	0.2%
Pct Chg in Distance Travelled	−0.079	0.158	−0.879	1.388
Share of Pop Over Age 65	17.502	4.319	3.855	53.106
Share of Trump Votes, 2016	0.629	0.156	0.041	0.916
Population Size, Thousands	104.648	330.248	1.233	10,057.160
Density: Thousands per Sq. Mile	0.268	1.753	0.001	69.468

Note: See Section 4 for formal definition of the *Neighbor Cases* variable.

4 Empirical Results

Motivated by the model outlined in (3), our baseline regression model is given by

$$\Delta_{it} = \beta_0 + \beta_1 e^{\beta_2 \Omega_{it}} + \beta_3 I_{it} + \beta_{I\Omega} e^{\beta_2 \Omega_{it}} I_{it} + \beta_X [e^{\beta_2 \Omega_{it}}, I_{it}] \times X_i + \rho \Delta_{i,t-1} + \epsilon_{it} \tag{4}$$

where the variables are defined as follows:

Dependent variable $\boldsymbol{\Delta_{it}}$ Mobility change is measured by the percentage difference in average daily distance in county i at time t, compared to the average of that in the same weekday before COVID-19 outbreak.

Perceived disease prevalence $\boldsymbol{\Omega_{it}}$ We assume that individuals approximate Ω_{it} as a linear combination of COVID-19 prevalence in both local and neighboring counties:

$$\Omega_{it} = \gamma_1 C_{i,t-1} + \gamma_2 \sum_{j \neq i} w_{ij} C_{j,t-1} \tag{5}$$

where each term in (5) denotes

1. $C_{i,t-1}$, is total confirmed cases divided by population at county i at time $t-1$. To ease interpretation of coefficients, we then normalize the median prevalence

level for counties with positive COVID-19 cases on 3/20/2020 to be one (0.003% - Mecklenburg, North Carolina).

2. $\sum_{j \neq i} w_{ij} C_{j,t-1}$, a weighted average of disease prevalence at neighboring counties measured at time $t-1$. For simplicity, we let weight be proportional to the inverse of distance between county centroids with $\sum_{j \neq i} w_{ij} = 1$.We adopted the same normalization as in $C_{i,t-1}$

We normalize $\gamma_1 = 1$ since the overall scale of $\{\gamma_1, \gamma_2\}$ cannot be separately identified from β_2. Using the fraction of total population infected as a proxy for contact rate follows the SEIR model's assumption of random meeting between individuals.

Restriction orders $\boldsymbol{I_{it}}$ Restriction order I_{it} is a dummy variable that takes the value of one if an order to stay-at-home is in effect in county i at date t; zero otherwise.

County demographics $\boldsymbol{X_i}$ County-level demographics include age structure (share of population over 65), political attitude (share of the population that voted for Trump in the 2016 Presidential election), and population density (thousand people per square mile).

Our main results are shown in Table 2. Combining estimates of β_1 and β_2 provides the direct effect of COVID-19 prevalence on changes in mobility. An increase of local infection rate from 0% to 0.003% of the population (i.e. $C_{i,t-1}$ increases from 0 to 1 due to scaling) leads to a reduction of mobility by 7.44%. These estimates suggests that decreases in mobility could take place well before the official announcement of restriction orders, which is line with the findings in Figure 7 and evidence from OpenTable reservations in Kaplan et al. (2020). On the other hand, the estimate of β_3 indicates that enactment of restriction order reduces mobility by 6.71% on average.

The estimate of γ (relative weight on disease prevalence in neighboring counties) indicates larger spatial spillovers of information across regions. This is driven by individuals adjusting their travel behavior in response to COVID-19 epidemics in Seattle or New York

Table 2: Main results

	Baseline Specification
β_0	-0.1204^{***}
	(0.0011)
β_1	0.1339^{***}
	(0.0027)
β_2	-0.5881^{***}
	(0.0847)
β_3	-0.0671^{***}
	(0.0105)
γ	2.1104^{***}
	(0.3096)
$\beta_{T,\Omega}$	0.0119^{***}
	(0.0032)
$\beta_{T,I}$	0.1792^{***}
	(0.0143)
$\beta_{O,\Omega}$	-0.0011^{***}
	(0.0001)
$\beta_{O,I}$	-0.0034^{***}
	(0.0005)
$\beta_{D,\Omega}$	-0.0025^{***}
	(0.0006)
$\beta_{D,I}$	-0.0004
	(0.0003)
$\beta_{I,\Omega}$	-0.1142^{*}
	(0.0460)
ρ	0.5177^{***}
	(0.0029)
$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$	
Num. obs.	91080

City even though there have not been confirmed cases locally. It highlights the importance of incorporating state or national information into the analysis of behaviors at a local level.

Coefficients $\beta_{X,\Omega}$ and $\beta_{X,I}$ shows the interaction effect of demographic characteristics $X \in \{T, O, D\}$ with perceived prevalence Ω and restriction order I. Characteristics T, O and D measure the fraction of population that voted for Trump in the 2016 Presidential election, share of population with age over 65, and population density (in thousands of people per square mile) respectively. Our results indicate that counties with lower votes for Trump, with older population, and with higher population density reacts more strongly to disease prevalence Ω and announcement I. Except for $\beta_{D,I}$, all coefficients are significant at 0.05 level.

The coefficient of the interaction between Ω and I, $\beta_{I,\Omega}$, is negative with an absolute value smaller than β_1. This implies that an increase in prevalence level Ω still decreases mobility when restriction order is announced, but with a smaller impact. Lastly, the estimate of ρ indicates that the Δ_{it} process is serially correlated.

5 Discussion and Conclusion

In this paper, we combine a novel GPS location dataset with COVID-19 cases and population characteristics at the county level to estimate the effects of disease prevalence and restriction orders on individual mobility. We find that population mobility reacts strongly to disease prevalence and a government stay-at-home announcement: a rise of local infection rate from 0% to 0.003% reduces mobility by 7.44%, and a government restriction order to stay-at-home reduces mobility by 6.71%. Additionally, we find that these effects of information on individual behaviour depends on characteristics of the underlying population. In particular, counties with larger shares of population over age 65, lower share of votes supporting Trump, and higher population density are more responsive to disease prevalence and restriction orders.

We plan to do a number extensions in the future. First, we would like to include more demographic controls such as industry composition and share of workers in essential jobs. These could affect the substitutability between on-site work and work-from-home, hence affecting changes in mobility. Furthermore, the guidelines for stay-at-home orders vary by profession. Second, we would like to explore county-level heterogeneity. Counties have varying unobservable characteristics that affect both the baseline Δ process, as well as how the disease spreads in the community. For example, social distancing practices and guidelines may differ from county to county. Third, since the speed of disease spread is going to be affected by individual decisions, we would like to extend the model to include these dynamics. In the current setting, we cannot make causal claims, as Δ_{it} and C_{it} are jointly determined by the true disease prevalence. The social distancing guidelines indicate the following feedback loop: disease prevalence causes confirmed cases to increase. Social distancing measures are designed to slow the spread of disease, and act on the true prevalence, which in turn affects the number of cases. Our current proposed strategy is to break this endogeneity by estimating a disease spread equation based on the exponential growth model of the spread of COVID-19, incorporating the delayed effects of mobility reduction on confirmed cases two weeks later; we can (approximately) think of two week lags of Δ_{it} as instruments for Ω_{it}. This exercise will benefit greatly from extending our dataset forward, through the predicted "flattening of the curve." Lastly, we will keep updating our dataset to get better estimates with more observations, and begin to identify the feedback loop previously outlined.

References

Auld, M. C. (2006). Estimating behavioral response to the aids epidemic. The BE Journal of Economic Analysis & Policy 5(1).

Hauck, K. (2018). The economics of infectious diseases. Oxford University Press.

Kaplan, G., B. Moll, and G. Violante (2020). Pandemics according to hank.

MIT Election Data + Science Lab (2020, Accessed: March 28th). 2018 election analysis dataset. `https://github.com/MEDSL/2018-elections-unoffical`.

The New York Times (2020a, Accessed: March 30th). Coronavirus (covid-19) data in the united states. `https://github.com/nytimes/covid-19-data`.

The New York Times (2020b, Accessed: March 28th). See which states and cities have told residents to stay at home. `https://https://www.nytimes.com/interactive/2020/us/coronavirus-stay-at-home-order.html`.

Unacast (2020, Accessed: March 28th). Covid-19 location data. `https://https://www.unacast.com/covid19`.

Wang, H., Z. Wang, Y. Dong, R. Chang, C. Xu, X. Yu, S. Zhang, L. Tsamlag, M. Shang, J. Huang, et al. (2020). Phase-adjusted estimation of the number of coronavirus disease 2019 cases in wuhan, china. Cell Discovery 6(1), 1–8.