Federal Reserve Bank of Philadelphia

Working Papers
Research Department

WP 20-38
September 2020
https://doi.org/10.21799/frbp.wp.2020.38

The Geography of Travel Behavior in the Early Phase of the COVID-19 Pandemic

Jeffrey C. Brinkman
Federal Reserve Bank of Philadelphia Research Department

Kyle Mangum
Federal Reserve Bank of Philadelphia Research Department

ISSN: 1962-5361

The Geography of Travel Behavior in the Early Phase of the COVID-19 Pandemic

Jeffrey C. Brinkman* and Kyle Mangum*

September 2020

Abstract

We use a panel of county-level location data derived from cellular devices in the U.S. to track travel behavior and its relationship with COVID-19 cases in the early stages of the outbreak. We find that travel activity dropped significantly as case counts rose locally. People traveled less overall, and they specifically avoided areas with relatively larger outbreaks, independent of government restrictions on mobility. The drop in activity limited exposure to out-of-county virus cases, which we show was important because such case exposure generated new cases inside a county. This suggests the outbreak would have spread faster and to a greater degree had travel activity not dropped accordingly. Our findings imply that the scale and geographic network of travel activity and the travel response of individuals are important for understanding the spread of COVID-19 and for policies that seek to control it.

JEL codes: R11, I18, H11

Keywords: travel behavior, mobility, COVID-19 pandemic, spatial dynamics, spatial networks, cellular device location

---

*Federal Reserve Bank of Philadelphia. We thank Heidi Artigue and Solomon Tarlin for research assistance. We gratefully acknowledge Victor Couture, Jonathan Dingel, Allison Green, Jessie Handbury, and Kevin Williams (cited below) for making publicly available the data on which this paper is based and for commenting on earlier versions of the manuscript.

1 Introduction

In the early stages of the COVID-19 outbreak, people drastically reduced their travel. Governments enacted numerous policies including stay-at-home orders, business closures, and limits on mass gatherings to reduce exposure and slow the spread of the virus. The change in travel behavior may reflect the implementation of these policies but also may be attributed to people responding to available information about the number of new coronavirus cases in their proximity. How did people reduce their travel behavior during the onset of the outbreak? Did they avoid places with larger outbreaks? And how did this response affect exposure and slow the spread of the disease?

In this paper, we use data on the movement of cell phones between U.S. counties to study the change in travel behavior and exposure in the early stages of the outbreak. The data provide daily measures of the network of bilateral travel flows between counties.[1]

We use these data to construct a mobility index for each county and examine how this index changed in response to government policies and the emergence of local cases. While government policies reduced mobility, travel also declined in response to the number of local cases in the county. This provides initial evidence that people adjusted travel behavior based on available information about the geography of the outbreak. Importantly, failure to account for this behavioral response would lead to overestimates of the effectiveness of shutdown orders on travel.

People not only traveled less, but also avoided locations that had higher numbers of cases. Using gravity regressions, we show that flows between locations declined in response to increased cases in both origin and destination.[2] This result holds even when controlling for government orders. Cases alone predict much of the decline in mobility, although they fail to capture the rise in travel that started in mid April.

[1] The measures were constructed by Couture et al. (2020) using cell phone data provided by PlaceIQ and generously made publicly available at https://github.com/COVIDExposureIndices.

[2] As we will explain, the notions of "origin" and "destination" are somewhat blurred in the cellular device data, although we use the terminology for expositional convenience.

Changes in mobility had large effects on overall virus exposure. We construct a measure of nonlocal (out-of-county) exposure as a sum of flows between counties that is weighted by the number of confirmed cases in the counties visited. We then compare the observed overall exposure to a counterfactual experiment that assumes mobility never changed from the average in the pre-pandemic period. In the counterfactual, exposure would have been twice as high at the end of April had people not changed their travel behavior. Furthermore, when we decompose this reduction in exposure, we find that roughly one third of the difference came from changes in the travel network, as opposed to overall declines in travel. The important policy implication of this result is that providing accurate timely information about the geography of an outbreak should be a policy priority.

The reduction in nonlocal exposure matters because such exposure led to increases in new COVID-19 cases. Using both least squares and instrumental variables methods, we find that a 1 percent increase in the exposure measure led to a 0.12 to 0.20 percent increase in new cases. This is a direct effect estimated in the early stages and does not reflect the long-run dynamics of the pandemic. Changes in travel patterns likely had significant benefits in reducing the spread of the disease by decreasing exposure.

Finally, we provide a simple model of the spatial dynamics of an outbreak. The model is used to illustrate the importance of the connectedness of locations and the mobility response of individuals to the geographic spread of new cases early in a pandemic. The important takeaway from the illustration is that connectedness between counties can both *speed the spread* in the short run and *perpetuate the outbreak* over the longer run, while a mobility response mitigates both of these effects. The model does not include important features of an epidemiological model such as recovery rates, deaths, or immunity, that would determine the arc of a disease outbreak. However, it is suggestive that reductions in mobility will reduce aggregate infections.

Our findings complement other recent research that has looked at declines in mobility during the outbreak using daily location data. Gupta et al. (2020) find that government policies led to significant declines in mobility, while Engle et al. (2020) find that policy as well as local case levels reduced mobility. There is also recent evidence that reductions in mobility and government policies mitigated the outbreak including work by Courtemanche et al. (2020), Fang et al. (2020), Fenichel et al. (2020), and Kraemer et al. (2020). In contrast to these studies, our research explicitly considers changes to the travel network in addition to declines in mobility levels. In addition, we are able to disentangle the role of policy interventions from behavioral responses to information.

Glaeser et al. (2020) study the effect of mobility on case growth using measures of within-metropolitan area travel activity for several large cities in the U.S., focusing on New York City. They find that mobility drops in response to cases, but when instrumenting for changes in mobility, higher within-city mobility drives faster case growth at the neighborhood level. Our study focuses on regional trip activity instead of neighborhoods, and we construct a measure of case exposure in addition to generic trip rates, which we find is an important determinant of case growth.

Other researchers have looked at the role of networks during the pandemic following work by Christakis and Fowler (2010) and Bailey et al. (2018). Kuchler et al. (2020) show that social networks in New York and Lodi, Italy predict the spread of COVID-19, while Coven and Gupta (2020) perform a similar analysis for New York, but also consider differences in mobility among demographic groups. In contrast to these papers, we consider how the observed travel network changed in response to the outbreak, and how this affected the spread of the disease. Monte (2020) also shows how the connectedness of counties shrank during the pandemic, but does not explicitly study the effects on exposure or case growth.

Lastly, our research connects other work that seeks to inform policies that restrict mobility. For example, Atalay et al. (2020) and Dingel and Neiman (2020) study the ability of workers to work from home in different occupations and industries. Our work, along with papers like these, can also help inform current theoretical research that seeks to understand the tradeoff between controlling the outbreak and economic welfare, including work by Farboodi et al.

(2020), Guerrieri et al. (2020), and Kaplan et al. (2020). These papers do not directly address the spatial dynamics of the outbreak and the importance of travel behavior.

2 Data Description

There are two main datasets used in our analysis. The first is the record of COVID-19 daily case diagnoses by county as reported by Johns Hopkins University.[3] We combine this with a listing of state-level activity restrictions including stay-at-home orders, closure of "nonessential" businesses and services, and restrictions on mass gatherings.[4] The second is a unique record of spatial travel activity as registered by cellular device locations, which we will describe in detail.

2.1 Case Prevalence

First we present some basic statistics from the daily COVID-19 case data.[5] In the spring of 2020, the early phase of the pandemic, COVID-19 cases were relatively concentrated in the Northeast U.S., and especially the New York City metro area, although there was some presence of cases throughout the country.

Table 1 reports summary statistics on case prevalence in terms of per capita cumulative diagnoses and rates of new diagnoses in each month of our period of interest. The distribution of cases is skewed with a long right tail, with many counties having low rates but some having major outbreaks. The ratio of the 99th percentile to the median per capita infection rate is at or above 19 for each month in our sample. The mass of the distribution shifted to the right

[3] Johns Hopkins University Coronavirus Resource Center. Data were retrieved from https://coronavirus.jhu.edu/.

[4] These data were collected by the Institute for Health Metrics and Evaluation at the University of Washington. They were downloaded from https://covid19.healthdata.org/.

[5] We have taken care to adjust the data for changes in reporting format or geography (e.g., counties that report together in some periods and separately in others) and to exclude outliers and values outside the domain of possible outcomes (e.g., negative case growth). Nevertheless, the data are reported subject to some discretion by health care providers and state and local health departments that introduces unavoidable measurement error.

Table 1: Summary of Case Prevalence

Time	Mean	SD	25th	50th	75th	90th	99th
Cases per 1k Residents							
Last Week of March	0.150	0.459	0.022	0.059	0.135	0.278	1.413
Last Week of April	1.582	2.991	0.351	0.716	1.548	3.280	16.499
Last Week of May	3.000	5.087	0.685	1.414	3.252	6.757	26.790
New Cases in Preceding 2 Weeks							
Last Week of March	65.99	578.89	0.71	3.00	13.00	60.00	843.00
Last Week of April	192.08	893.21	4.29	15.93	70.14	267.43	4,318.71
Last Week of May	150.99	644.21	4.71	17.79	78.43	278.57	2,350.00

NOTES: The table reports summary statistics of COVID-19 cases per capita and new case growth as reported by Johns Hopkins University.

as the virus percolated throughout the country. The peak of new infections was in early to mid April (although these rates have been surpassed by additional outbreaks in the summer of 2020).

Our study will examine the implications of the skewed geographic pattern of virus prevalence on exposure and the spread of the virus across regions.

2.2 Measuring Travel Behavior From Cellular Devices

Our measure of mobility is an anonymized summary of movement between counties derived from a microdata record of cellular device locations. The measure was constructed and generously made publicly available by Couture et al. (2020) (hereafter, CDGHW) using data provided by analytics firm PlaceIQ. The individual device location "pings" are aggregated at the county level.[6] We often refer to the cross-county travel behavior depicted by this dataset interchangeably as "mobility" or "spatial activity."

Specifically, we observe: (i) the number of devices registering in a county each day, and (ii) the fraction of those devices that registered in each county elsewhere in the U.S. sometime in the preceding 14 days. For convenience we will refer to the current county as "home" and the previous county as the "visit(ed)" county because that is how the public data were

[6] Using the same underlying cellular device data, CDGHW also maintain a measure of device exposure defined by points of interest (shops, parks, restaurants, etc.). Some of the work cited above studies activity in the pandemic by point of interest. In this article, we have chosen to focus on regional travel activity.

constructed by CDGHW, although as we will explain, the notions of origin and destination are not completely straightforward. In our analysis, we use data running from January 20, 2020 until May 25, 2020.[7]

To protect confidentiality and to limit the size of matrix, CDGHW limited the reported counties to those that had at least 1,000 devices registering over a one week period from November 2019 to early January 2020. The resulting dataset consists of a time-consistent list of 2,018 counties in the U.S., comprising 4,072,324 possible pairs, some of which may be zero if no device from the home county has visited the destination county in the previous two weeks. The average county exhibits 1,732 unique nonzero pairs per week in the base period of January 20 to February 23, the first five weeks of data and before the onset of widespread COVID-19 diagnoses. We do not observe activity outside the 2,018 county network, but the observed counties encompass 97 percent of the U.S. population.

Under CDGHW's selection method of a 1,000 device threshold, larger counties are more likely to register in the data, but some rural areas and small towns are represented as well. Among metro areas, 90 percent of counties (comprising 99 percent of metro population) appear in the data. Counties in metro areas making up 87 percent of total U.S. population comprise 86 percent of the devices registering in the dataset. Rural and micropolitan counties appear at a rate of 48 percent, although 80 percent of the population in these areas is represented. These smaller counties, making up 11 percent of U.S. population, account for the remaining 14 percent of devices we observe.

Table 2 reports summary statistics for one of the main objects of interest, the fraction of devices in a county that were also present in other counties in the previous 14 days. The typical county has a same-county ping rate of 90 percent, meaning 10 percent of devices present today are "new" and were not present in the preceding two weeks. When limiting to the non-reflexive counties, the average ping rate is dramatically lower, but with a clear geographic pattern. The average county pair has a ping rate of 0.3 percent. This rises to 0.5

[7]CDGHW continue to update the data on a regular basis, but we stop on May 25 in order to focus our analysis on the initial phase of the outbreak.

Table 2: Summary Statistics of Device Ping Rates by Geography

	Pairs	NT	Mean	SD	P10	P50	p90
Same County	na	70,630	89.08	4.78	82.69	90.01	94.19
Other Counties:							
Neighbor	11,384	398,283	34.83	29.76	6.13	23.16	89.33
Within Metro Area	5,420	189,700	17.90	20.79	0.95	8.60	54.19
Within State	128,351	4,429,513	3.08	7.82	0.09	0.62	6.97
Within Division	573,239	17,131,215	1.00	4.29	0.02	0.12	1.43
Within Region	1,388,653	35,309,674	0.55	3.06	0.03	0.06	0.65
Any	3,905,332	84,612,867	0.27	2.01	0.01	0.04	0.30

NOTES: The table reports summary statistics of device ping activity occurring over a 14 day window in the pre-pandemic period (January 20 to February 23, 2020). Statistics count only observations with nonzero ping rates. There are 2,018 counties included in the dataset, and 1,004 in metro areas. Source: Couture et al (2020).

percent within region, 2 percent within metro area (among urban counties), 3 percent within state, and up to 34 percent among neighboring counties, with some neighbors much higher.

The cellular device data roughly resemble other spatial connectedness measures normally used by economists, such as commuting, migration, or trade flows, but with a few specific features that are worth bearing in mind.

First, the public use data as produced by CDGHW embed a particular lag structure to mobility events. The indicator measure is whether a given device has visited another county in the prior two weeks.[8] This mechanically builds in a moving-average nature to the data, in that a single visit on one day will register for 14 days to come.

The second mechanical feature of the data is that the visit propensity measure for each county is the probability of a binary event–whether it was in the county in the prior two weeks–and not a transition matrix over an exhaustive set of alternatives. There are multiple (2,018) potential binary events, and a device can trigger the affirmative in any and all of these, so the probabilities are separate outcomes, not a vector of transition probabilities summing to one. Moreover, because each county-event measure is a binary probability, we cannot distinguish a small amount of mobility from a large number of devices from a large amount of mobility for a small number of devices. For example, it is observationally equivalent for a single device to ping in multiple counties, registering in each place, or many devices

[8]The data were constructed in this manner to capture potential exposure and are based on the estimates of the COVID-19 incubation period that have been widely reported.

disparately visiting single, separate counties. The visit probability measures are thus better interpreted as indices of trip activity.

Third, by definition, a device visiting multiple counties is affirmatively "present" in each one, and there is no natural definition of "origin" and "destination," or even of "home" and "away" for a given device. The cross-county visit rate is best viewed as a measure of *connectivity* rather than a true directional trip. For example, a metro area commuter crossing from a suburban county to a central county would register as present in both his county of residence and his county of work on a given day, though we cannot distinguish which county occupies either status.[9] Bearing this caveat in mind, for ease of exposition, we will refer to the focal county, i.e., the one in which devices register today, as "home," and the visit counties, i.e., the set of possibilities for which the indicator measures whether the focal devices were present, as a "destination."

Lastly, we note that visits are simply frequency counts, and do not measure any sort of visit intensity, such as time spent or persons encountered in the visited county. At the extreme, a device could register a visit by simply passing through in a car, having never actually contacted anyone or anything in the county.

For all of these reasons, we avoid cardinal statements and cross-sectional comparisons in our analysis, opting instead to form indices of mobility, normalizing each county or county-pair by its pre-period regular activity. This allows us to control for idiosyncrasies in the data and focus our attention on behavior in "pandemic times" compared with "normal times."

3 Mobility During the Pandemic

In this section we describe the dynamics of mobility over the early phase of the COVID-19 pandemic and examine how travel activity changed with case counts. We first construct a

[9]The further implication is that prior visits would register for each county. If he visited a third county a week ago, that visit would register for both the residence and the work counties–each of which we will read as "home" for the visit event and for which we will separately count the events. In this way, the measure of mobility is nonlinear in that one county crossing leads to observation of multiple other-county visits.

measure of mobility for each county that captures changes in both overall activity and travel to outside counties. We use this measure to show that while mobility declined in response to government orders, there was also an independent response to proximate virus cases.

3.1 Dynamics of Mobility

Figure 1, panel A plots the two main pieces of information provided in the cellular device dataset. To emphasize the changes over time, each of these is indexed within county so that the period of January 20 to February 23 (the first five weeks of data) averages to one, and the graph plots the median index value.

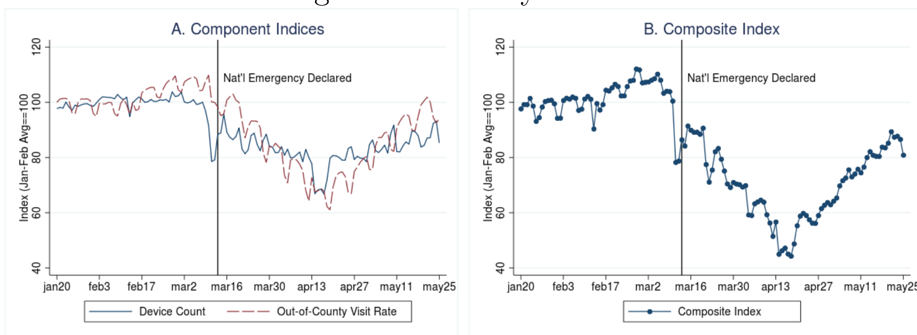
The blue solid line depicts the number of devices registering in each county j on each day t, an object we will denote below as d_{jt}. The more spatially active a device is within or between counties, the more likely it is to ping a cellular tower and register its location, and therefore the count of devices in the data forms a sort of mobility metric. This has a clear downward cycle, falling about 30 percent from normal times to the trough in mid-April and recovering somewhat by the end of May to about a 15 percent reduction from normal times.

The red dashed line depicts the sum of out-of-county pings, the sum of the binary event probabilities. Denoting the binary event of a visit from a device in j to a location i occurring in the preceding 14 days t as σ_{ijt}, the total activity for a county j is $\sum_i \sigma_{ijt}$. This sum forms another form of mobility metric, as higher rates of out-of-county location indicate higher prior mobility of the devices registering in the county. The out-of-county ping rate also shows a clear downward cycle, falling 40 percent to the trough in mid April and recovering to only about 5-10 percent down by the end of May.

Panel B depicts a composite of the two mobility indices. The solid blue line is the device ping count, the device-weighted sum of out-county pings, which essentially combines the two highly variable series from panel A as

$$m_{jt} = \sum_i \sigma_{ijt} d_{jt}. \tag{1}$$

Figure 1: Mobility Indices



NOTES: The figures present the median index value taken across counties in the 2,018 national sample. Panel A displays the median of the two subcomponents of device mobility, the count of active devices and the sum of out-of-county pings ("visit rate"). Panel B reports the composite index, the active device-weighted visit rate. Source: Authors' calculations using Couture et al (2020).

We then index this composite series as $\tilde{m_{jt}} = \frac{m_{jt}}{\bar{m_{j0}}}$, where $\bar{m_{j0}}$ is the mean of the county's index in the pre-pandemic period (January 20 to February 23). Under this metric, mobility fell by half from February to mid-April, recovering by late May to about 20 percent below normal.[10]

Figure 2 plots the 90th and 10th percentiles in addition to the median, showing the entire distribution of mobility indices moved to the left, though there is some heterogeneity in the magnitude of declines. Some counties showed mobility declines greater than 60 percent while others declined less than 40 percent.

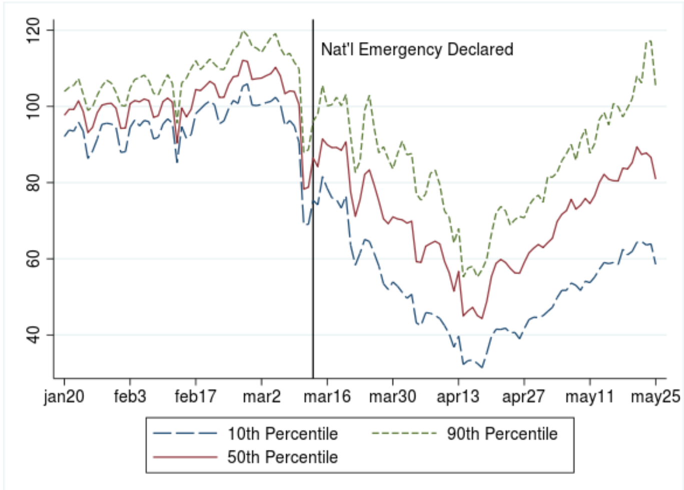
3.2 Mobility and the Onset of the Pandemic

Next we use our composite mobility index to examine the how spatial activity changed in the early stages of the outbreak. Figure 3 shows the dynamics of mobility in comparison to common state-level mobility restrictions of bans on mass gatherings, closures of "nonessential" businesses, and stay-at-home orders (Panel A), as well as cumulative cases (Panel B) and case growth (Panel C). The orders are plotted as the share of counties under each type of restriction. A vertical line marks the national emergency declaration on March 13.

One notable feature is that the first drop in mobility occurred immediately following the

[10]We also considered a metric where we weighted the sum by distance between counties. The dynamics of a distance-weighted mobility index are very similar to the unweighted measure we present here.

Figure 2: Mobility Index Distribution Over Time



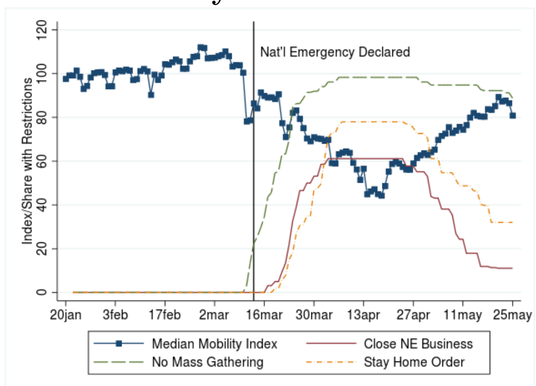
NOTES: The figure displays the time series of the 10th, 50th, and 90th percentile of the composite mobility index. Statistics are unweighted, taken across all 2,018 counties in the national sample. Source: Author's calculations using Couture et al (2020).

initial run-up in cases–and before mobility restrictions were enacted–as it became clear that the U.S. was experiencing community spread and not just isolated cases due to foreign travel. From March 1 to March 14, though no county was yet under stay-at-home order, mobility dropped by 20 percent as cases rose 500%.

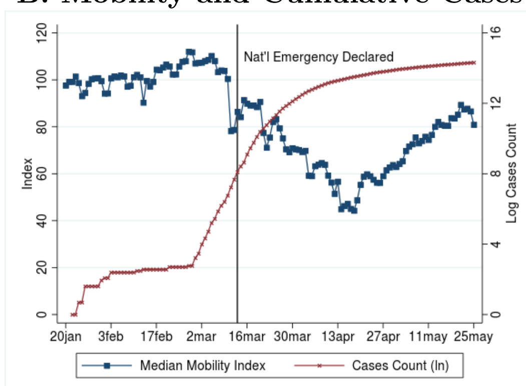
Spatial activity continued its downward trend from that point into April as case counts continued an exponential rise and stay-at-home orders and other mobility restrictions were more widely enacted. Mobility reached a bottom in mid-April (coinciding with the Easter and Passover holiday observances), with a trough on April 18 of 44; that is, 56 percentage points below its pre-virus average. At that point, however, new case rates had started to flatten and some mobility restrictions began to lift, and mobility crept back upward. Together, these patterns suggest that households may have been responding to information about (or at least perceptions of) the virus prevalence as well as formal emergency declarations and restrictions, a hypothesis we will test.

Figure 3: National Mobility Index, Mobility Restriction, and Case Growth

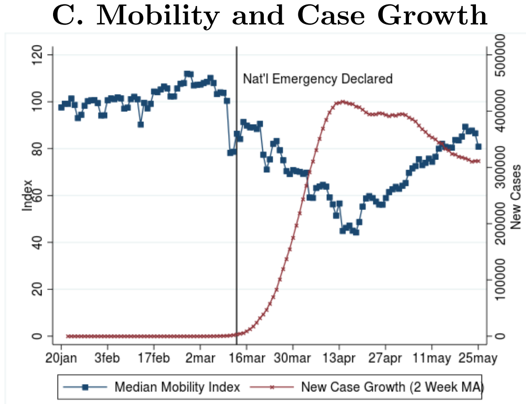
A. Mobility and Restrictions



B. Mobility and Cumulative Cases



C. Mobility and Case Growth



NOTES: The figures plot the median composite mobility index against: the fraction of counties under government restrictions (A), the log of the national case count (B), and the number of new cases reported nationally in the preceding two weeks (C). Sources: Couture et al (2020), all panels; healthdata.org (2020), panel A; Johns Hopkins University (2020), panels B and C.

3.3 Mobility and Local Case Counts

The national series mask a fair amount of regional heterogeneity. For example, the New York metro area alone accounted for 41 percent of cases on April 18. Moreover, states varied in the timing and intensity of their spatial activity restrictions. Did households in more affected areas respond more strongly?

To answer this question, we leverage the spatial variation in the mobility metric and case counts by county in addition to state-level restrictions on mobility in the following model:

$$\tilde{m_{jt}} = \beta_c c_{j,t-13:t} + \sum_q \beta_q I(R^q_{jt}) + \epsilon_t, \tag{2}$$

where the left-hand side is the indexed mobility rate, $c_{j,t-13:t}$ denotes new case diagnoses in the home county in the preceding two weeks,[11] and the R_q terms denote type q government restrictions on activity. These take the form of indicator variables $I()$ for whether the restriction is in place at time t.

Table 3 reports coefficients from this regression of county-level mobility rates on local cases and restrictions.[12] In the first column, we use (log) local new case counts from the previous two weeks as the lone explanatory variable. The coefficient of -7.06 means that a 100 log point rise in new cases resulted in a 7 percentage point drop in the mobility index. For some context, the average county saw about a 300 log point rise in case growth in April (and some as large as 1,000 log points).

Mobility dropped across the board, however, and not only in places with outbreaks (see Figure 2). Column 2 adds time dummies to pick up the common trend, meaning the estimated coefficient is therefore the marginal effect, relative to national trends in mobility, of additional cases in one's own county.[13] Column 2 shows that a 100 log point increase in local cases

[11]Case growth in a two week window corresponds with the time lag of visits embedded in the cellular device data. We have checked that these results hold for cumulative cases, cases per capita, new cases per capita, and deaths.

[12]Recall that each county has been indexed to a pre-period average of 100 so that level differences in mobility levels between counties will not affect the covariances we estimate.

[13]Because the "common trend" was due to the outbreak of COVID-19 cases, however, it is debatable which

results in a 3.7 percent drop in mobility relative to national trend. This coefficient is a more conservative estimate of the direct effect of local cases instead of the more ethereal fear of the virus.[14] We continue to use time effects hereafter.

Around the same time cases were growing, state and local governments enacted restrictions designed to limit mobility of residents and suppress the spread of the virus. In column 3, we include indicator variables for the presence of the three most common of these measures, closure of nonessential services, bans on mass gatherings, and direct stay-at-home orders, in order to measure their effects on cross-county activity.[15] Each independently depressed activity to a significant degree, although their inclusion scarcely affects the estimate of local cases, suggesting people were still reacting to public information about the virus. In column 4, we remove the local case growth variable. Failure to include public information about local cases causes a larger estimate of the marginal effect of government activity restrictions.

Of course, to exit a county, there must be a county to go, and these destinations may themselves be under closures and restrictions. In columns 5 and 6, we measure destination restrictions as the pre-pandemic mobility weighted average of other counties' prevalence of restrictions. The destination closure measure is $R(dest)^q_{jt} = \sum_{i \neq j} \bar{\sigma}_{ij0} I(R^q_{it})$.[16]

Column 5 includes the destination measures jointly with home county restrictions, finding a significant effect of nonessential business closures, perhaps the result of there being fewer businesses to visit. The stay-at-home contribution is negligible, and the destination mass gatherings ban is effectively collinear with the time dummies and not well estimated. In column 6, we remove the home county restrictions and find larger partial effects of destinations'

---

of these models estimates the *causal* effect of the virus more generally. The model with time effects serves to isolate the contribution of local cases.

[14]In unreported results, we found a decaying spatial pattern in the marginal effect of cases on mobility. County cases had the highest effect, while state and regional case counts contributed independently but to a smaller degree.

[15]Many of these restrictions were designed to limit within-county activity as much as between-county activity, which we do not study in this paper. These effects may show up in our measures to the extent that they depressed the number of active devices in a county, d_{jt}.

[16]The timing for other counties' restrictions is effectively the same for all places (but for the omission of one's own county from the destination network), but the measures will vary according to how frequently that county interacted with the others according to baseline visit rates, $\bar{\sigma}_{ij0}$. As it turns out, the destination mass gatherings index is effectively collinear with the time trend.

Table 3: Mobility Index: The Effect of State Restrictions and Observed Cases

	1	2	3	4	5	6
Log Cases (2 wk lag)	-7.066	-3.674	-3.519		-2.560	-2.889
	(0.071)	(0.130)	(0.127)		(0.126)	(0.125)
Home County:						
Nonessential Services			-2.143	-2.988	-0.872	
			(0.392)	(0.468)	(0.757)	
Mass Gatherings			-0.748	-0.523	-13.327	
			(0.540)	(0.545)	(0.747)	
Stay Home			-3.152	-4.458	-3.604	
			(0.437)	(0.498)	(0.797)	
Destination County:						
Nonessential Services					-0.575	-0.614
					(0.322)	(0.122)
Mass Gatherings					6.521	5.538
					(0.304)	(0.210)
Stay Home					-0.019	-0.765
					(0.331)	(0.134)
Constant	90.858	86.923	88.737	99.319	99.319	99.319
	(0.126)	(0.214)	(0.412)	(0.116)	(0.116)	(0.116)
Time Effects		y	y	y	y	y
R^2	0.329	0.767	0.771	0.736	0.815	0.801
NT	252,250	252,250	252,250	252,250	252,250	252,250

NOTES: The outcome variable is the county-level index of mobility as defined in equation 1, and indexed by the pre-pandemic average for each county. Units are percentage points. Standard errors are clustered by county. Home state restrictions are indicator variables, and destination state restrictions are a visit-weighted index of exposure to restrictions in potential destinations; the units on the home and destination side are not directly comparable. Source: Authors' calculations using data retrieved as described in Section 2.

closures and stay-home orders. In neither case do these significantly affect the coefficient on local cases. The destination restriction measures are additionally useful in that they predict mobility for a county using outcomes outside of the county, a feature we exploit below in estimating the effect of virus exposure from travel on new case growth.

4 The Pattern of Changes in Mobility

Did mobility drop in a uniform way, or were some locations avoided more than others? That is, what does the network of origins and destinations tell us about travel behavior during the early phase of the pandemic? To examine these questions, we return to the raw, uncollapsed structure of the data to see which sets of visits were declining.

4.1 A Gravity Model of Visits

We use a gravity model to examine how the network of travel flows between counties changed in response to cases in both home and destination counties. Specifically, we regress county pair ping rates on local case counts and mobility restrictions.[17]

$$\sigma_{ijt} = \omega_i c_{i,t-13:t} + \omega_j c_{j,t-13:t} + \beta_{i,q} I(R_{it}) + \beta_{j,q} I(R_{jt}) + \rho_{ij}. \tag{3}$$

In using a gravity regression on bilateral flows, we can separately enter the case counts on the focal (origin) and visit (destination) county sides of a trip. This is more informative regarding exposure than the summarized mobility regression in Table 3. Because the visit frequency is already reported as a moving average (and to economize on computational requirements), we limit to weekly observations. All specifications include county pair and week fixed effects, and the visit rates and case counts are log transformed.

Table 4 reports the results of the gravity regressions. Column 1 shows the coefficients on the two-week new case count for the home and visit counties. Cases in the home county reduce visits outside the county, but cases in the visit destination also limit the visit frequency. That is, conditional on making a trip outside a county, devices are less likely to visit counties with relatively higher infection rates.

The distribution of visit frequency to particular counties is highly skewed, as Table 2

[17]Note we are using the ping rate–the probability of a binary event of a visit between i and j counties–without the device count. This is because the device count only affects the origin side. Results using a device-weighted ping rate look similar, but the coefficients on the origin side variables are mechanically larger.

Table 4: Changes in Mobility - Gravity Regressions

	1	2	3	4
Cases in Home County	-0.0305	-0.0307	-0.0304	-0.0305
	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Cases in Visited County	-0.0493	-0.0480	-0.0486	-0.0473
	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Cases in Visited X Baseline Visit Rate		-0.2746		-0.2746
		(0.0049)		(0.002)
Stay at Home, Home County			-0.0013	-0.0012
			(0.0003)	(0.0003)
Stay at Home, Visit County			-0.0306	-0.0306
			(0.0003)	(0.0003)
Constant	-7.5035	-7.5038	-7.4952	-7.4955
	(0.0003)	(0.0003)	(0.0002)	(0.0002)
Time Effects	y	y	y	y
County Pair Effects	y	y	y	y
R^2	0.895	0.896	0.896	0.896
NT	41,253,269	41,253,269	41,253,269	41,253,269
Pairs	3,564,207	3,564,207	3,564,207	3,564,207
Weeks	18	18	18	18

NOTES: The table reports results from a gravity regression of log visit rate in the two weeks preceding observation date on case growth and stay-at-home orders; see equation 3. The observation level is a weekly observation of a directed county pair (i.e., $A \to B \neq B \to A$). All specifications include directed county pair and week of year fixed effects. Standard errors are clustered by directed county pair. Source: Authors' calculations using data retrieved as described in Section 2.

showed, and it is a possibility that the elasticity of visit frequency is a function of the base rate, which might in turn affect the parameter estimates. Therefore, in column 2 we add an interaction of the baseline rate and the cases in the destination. More frequently visited counties are substantially more affected by the count of cases registering there.[18]

The remaining columns of Table 4 add controls for mobility restrictions such as stay home orders. These are highly correlated with each other, so are we deliberately parsimonious in

[18]We have examined the elasticity of the visit frequency with respect to its baseline rate, which we found to be U-shaped. Rarely visited places, though very numerous in numbers of pairs, account for little of the total outside visit frequency and even less of the change in mobility, with some actually increasing in visit frequency in the depths of the pandemic. Then, the proportional change in visit rate is decreasing in the baseline rate up until the highest centiles of the frequency distribution (which, as seen in Table 2, are mostly the neighboring counties) at which point the rate of change is smaller. Some of this pattern is mechanical in that visit frequency is bounded between zero and one, and the largest and smallest base rates move by lower percentages. Some of it might indicate the nature of travel however, in that trips in the tails of the distribution might be the most essential and least likely to change. We have run the regressions of Table 4 with controls for a time-varying slope of the elasticity of the visit frequency with respect to the baseline rate. This does not materially affect our estimates.

their inclusion. Stay home orders on both sides of the visit are negatively associated with visit frequency, but the inclusion of these controls does not affect the coefficients on case counts in the home or visit county.

In summary, it appears that not only are households traveling less, they are avoiding more affected places. The data do not reveal precisely *why* people avoid places with high case counts. It could be a fear of being exposed to the virus in high caseload areas. It could be that reduced activity in the destination produces less of a reason to travel there; for example, if more workplaces and stores are closed (beyond the government's prescription). The outcome is plausibly a combination of these mechanisms.[19]

Whatever the cause, the pattern is meaningful in two ways. First, it explains most of the mobility decline in the pandemic (although less so its recovery in May), as we show next. Second, it means that households were less exposed to virus cases than if they had continued spatial activity as they had in the days before the pandemic, a topic we treat in more detail in the subsequent section.

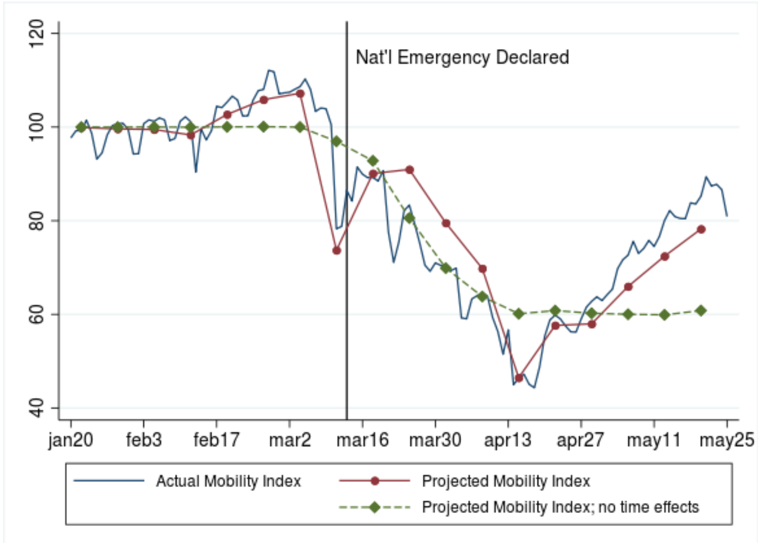
4.2 Case Avoidance and the Drop in Mobility

The avoidance of cases explains most of the mobility decline early in the U.S. outbreak. This is illustrated in Figure 4, which plots the median mobility index from predicted values of regression 3 (column 2 from Table 4) alongside the actual mobility index from Figures 1, 2, and 3. There are two versions of the projection: one with the time dummies factored in and one with them excluded so that the projection relies only on case counts.[20] In either version, the projection does a remarkably good job of predicting the fall in mobility, indicating that case avoidance was critically important in explaining the drop in spatial activity. The version without time dummies fails to predict two blips in activity–just before the emergency

[19] Here, we note again that our designation of "origin" and "destination" is for convenience, as these are difficult to distinguish given the way the cellular device locations are reported. What is important is that activity between counties depends on case prevalence on either side of the connection.

[20] The time dummies are used in the regression in both cases. The difference is whether the time dummy coefficients are included in the projection.

Figure 4: Actual and Projected Median Mobility Index Over Time



NOTES: The figure displays the median composite mobility index against the median mobility index as projected by the coefficients from the Table 4, column 2 estimate of (3). One projection uses the time dummies in its forecasted values, and the other uses the coefficient estimates from the same model but omits the time dummies. Source: Authors' calculations using data retrieved as described in Section 2.

declaration and in the depths of the trough in mid-April–but more notably, it fails to predict the rise in mobility in mid to late May, as cases were still fairly prevalent. A version of the projection using the stay-at-home orders produces similar results and only slightly better predicts the recovery in activity in May.

5 Case Exposure

Devices crossing county lines may indicate people coming in contact with cases from outside their local areas. Are these encounters consequential for case growth? In order to study this question, we develop a measure of case exposure that summarizes the number of case contacts a county is incurring by encountering other places with active cases. Then we consider how the pattern of case avoidance affected exposure and subsequently altered the trajectory of virus spread.

Table 5: Summary of Case Exposure

Time	Mean	SD	25th	50th	75th	90th	99th
Last Week of March	296.2	553.1	122.8	191.5	301.7	494.1	2,375.5
Last Week of April	798.3	937.2	321.1	506.8	897.2	1,554.2	5,316.3
Last Week of May	1,035.1	942.1	483.9	751.2	1,258.7	2,030.4	4,811.1

NOTES: The table reports summary statistics of exposure to nonlocal cases as defined in equation (4) for each listed point in time. Source: Authors' calculations using data retrieved as described in Section 2.

5.1 A Metric for Nonlocal Case Exposure

We define nonlocal *case exposure* as

$$x_{jt} = \sum_{i \neq j} (\sigma_{ijt} d_{jt})^{\alpha_1} n_{it}^{\alpha_2}, \tag{4}$$

where n_{it} represents cases in the destination county and $\sigma_{ijt} d_{jt}$ is a pairwise mobility measure (as in equation (1), but with the summation outside the exponent). The case exposure measure is thus a sum of travel flows from an origin weighted by the number of cases in the destination county. For the majority of what follows, we set the exponents as $\alpha_1 = 1, \alpha_2 = 1$, but we examine robustness to alternative exponents. We focus on exposure to new cases in a two week window, although the general patterns we find are robust to cumulative cases and cases per capita.

Table 5 presents summary statistics of this exposure measure for the counties in our mobility sample at each checkpoint as the case summary in Table 1. This measure averages values in the hundreds, and like the cases themselves, is highly skewed to the right tail. Exposure rose over time even as mobility fell because cases became more widespread.

Exposure measured in this way could be high for a given county because of some combination of (i) high frequency of travel and (ii) travel to high caseload areas. In Table 6, we present some decompositions to illustrate the source of higher exposures. The general pattern is that more exposed counties tend to have greater contact with extremely high caseload areas and not necessarily higher levels of overall mobility.

The upper panel of Table 6 shows statistics for the whole sample and for a split between

Table 6: Sources of Case Exposure

Panel A: Highly and Lesser Exposed

	1	2	3	4	5	6	7	8	9	10	11
				Destinations in Top 2 Pct. Of Cases				Top 50 Destinations			
Counties, N	Average Exposure	Total Visits	Cases Per 1k Residents	Visits To	Visit Share	Exposure In	Exposure Share	Visits To	Visit Share	Exposure In	Exposure Share
All, 2,018	464.0	3.16	0.58	0.14	0.05	216.6	0.53	2.30	0.73	293.6	0.67
Lesser Exposed, 1,968	451.6	3.17	0.56	0.14	0.04	209.7	0.52	2.31	0.73	285.9	0.67
Top 50 Most Exposed, 50	4,448.5	2.87	7.54	1.60	0.58	4,289.2	0.96	2.26	0.81	4,357.7	0.98

Panel B: Selected Cities

	1	2	3	4	5	6	7	8	9	10	11
				Destinations in Top 2 Pct. Of Cases				New York Metro			
Location, N Counties	Average Exposure	Total Visits	Cases Per 1k Residents	Visits To	Visit Share	Exposure In	Exposure Share	Visits To	Visit Share	Exposure In	Exposure Share
Philadelphia, 11	2,456.2	3.04	3.48	1.21	0.41	1,901.1	0.77	0.26	0.09	518.6	0.20
Pittsburgh, 7	453.8	2.61	0.88	0.12	0.05	218.0	0.48	0.03	0.01	92.9	0.20
Chicago, 13	4,052.6	3.19	1.85	0.63	0.22	3,512.2	0.82	0.02	0.01	43.2	0.01
Miami, 3	959.4	1.42	2.50	0.70	0.49	917.3	0.95	0.05	0.04	25.5	0.03
Houston, 10	1,761.5	2.84	0.76	0.72	0.27	1,562.2	0.83	0.01	0.00	13.6	0.01
Los Angeles, 2	1,406.3	1.10	1.10	0.49	0.42	1,263.4	0.69	0.01	0.01	8.1	0.02
San Francisco, 5	410.0	2.19	0.97	0.15	0.07	104.6	0.26	0.02	0.01	13.5	0.03

NOTES: The table reports summary statistics of the visit rates, cases per capita, and exposure to nonlocal cases as defined in equation (4) for each listed point in time.
Source: Authors' calculations using data retrieved as described in Section 2.

the 50 highest exposure areas and the lesser exposed areas. What caused the great difference in exposures? The highest exposure areas had relatively less mobility on average, making outside county trips at a rate of 287 percent compared to 317 percent for lesser exposed counties. (Recall that the mobility measure is a sum of binary event probabilities and can therefore sum to more than 100; other caveats from Section 2 apply.)

The differences appear when splitting by destination. Columns 4, 5, 6 and 7 show statistics for contact with destinations among the highest two percent of cases per capita in the U.S. The highest exposure areas visited these drastically more often the typical county, encountering a high case area at a trip rate of 160 percent (or 58 percent share of all out-of-county visits) versus 14 percent (4 percent share) for lesser exposed counties. These high case areas comprise nearly all of a highly exposed county's exposures (and still a significant fraction of the lesser exposed places too).

As a benchmark, columns 8 to 11 show the travel to the counties' top 50 most visited destinations. The more exposed counties did not travel to their usual partners at any higher frequency than the lesser exposed, but they accumulated much more exposure because these partners had larger outbreaks. Thus, exposure is largely a function of outbreak size in a county's usual network.

To illustrate these patterns, the bottom panel of the table shows statistics for selected metro areas from various regions of the country and with differing levels of outbreaks. In this panel, we replace the "top 50 destinations" with a particularly hard-hit region, the New York metro area.

The cities of Philadelphia and Pittsburgh provide a telling example because, among several other similarities, they were both under the guidance of the same state government, Pennsylvania. Philadelphia, however, had much greater exposure to high caseload areas in the Northeast corridor, including a relatively high contact rate with the New York metro area. Hence, its case exposure level was far greater than Pittsburgh's, and its size of outbreak far greater as well.

5.2 Case Avoidance and the Effect on Exposure

Next we examine the importance of changes in travel behavior on exposure. To do so we compare actual exposure to counterfactual exposure measures that assume travel behavior did not change despite the increase in cases. We then decompose the differences in exposure to understand the importance of declines in overall travel activity versus the avoidance of highly affected locations.

Figure 5 plots the median actual exposure, measured in counts of cases, and the median exposure that would have been obtained had each county continued with business as usual. (To obtain these series, each county's exposure was scaled by its pre-pandemic mobility in order to allow cross-sectional comparisons.) Clearly, the pullback in mobility significantly altered the degree of outside exposure to virus cases: The median county would have been exposed to twice as many cases if travel behavior had not adjusted.

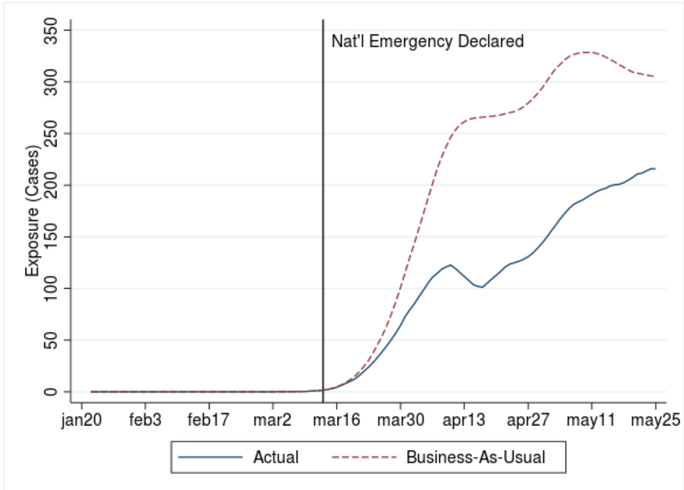
From equations (1) and (4), there are three ways the contact intensity could change. First, the number of devices registering as active could fall. Second, the total frequency of out-of-county pings could change. Both of these did trend lower as activity dropped, as Figure 1 showed. The third way is for the network of visit destinations to change, as Table 4 suggests did occur and Table 6 indicates could be important for explaining exposure.

Table 7 shows the ratio of counterfactual exposures to actual exposures for three points in time (end of March, April, and May). The table lists the combined effect of all forms of mobility and the univariate decompositions.[21] To quantify the pattern from Figure 5, the table shows that had spatial activity continued as usual, the median county would have had exposure to 52 percent more cases at the end of March, 115 percent more cases in April, and 44 percent more cases in May.

Each component of spatial activity contributed to this drop. For example, in April, had the active devices changed but out-of-county mobility continued as usual, case exposure would have been 24 percent higher (Column 2). Had total out-of-county ping activity continued

[21] The decompositions do not add to the total because each is a median of a univariate calculation.

Figure 5: Actual Exposure Compared With "Business As Usual" Exposure



NOTES: The figure plots the median exposure index across the sample of counties in the cellular device data. The unit of measure is number of cases per unit of mobility in the pre-pandemic period; its scale is comparable across counties. Source: Authors' calculations using data retrieved as described in Section 2.

Table 7: Decomposition of Actual Exposure Relative to "Business As Usual," By Mobility Component

		Partial Effect Of:		
Time	Combined	Device Count	Visit Rate	Visit Geo. Network
	1	2	3	4
Last Week of March	1.54	1.17	1.16	1.13
Last Week of April	2.09	1.24	1.34	1.22
Last Week of May	1.40	1.14	1.02	1.19

NOTES: The table reports the median ratio of counterfactual exposure, projected using pre-pandemic period mobility rates, relative to actual exposure for each listed point in time. Nonlocal case exposure is defined in equation (4). Column 1 is the combined exposure index, and columns 2 through 4 are its components. Column 2 holds fixed total active devices, column 3 holds fixed out-of-county pings per device, and column 4 holds fixed the destination share each origin's travel network. Source: Authors' calculations using data retrieved as described in Section 2.

as usual, case exposure would have been 35 percent higher (Column 3). Had the network of destinations remained the same, case exposure would have been 25 percent higher (Column 4). This last term is especially interesting because it shows a substantial amount of the change in exposure resulted not just from staying home, but from avoiding places with higher levels of outbreak when traveling. Notably, even as the level of mobility edged higher in May, a reduction in exposure resulted from people avoiding counties with high caseloads.

6 The Effect of Exposure on Case Growth

The remaining major question is whether out-of-county exposure to virus cases creates new cases in the home county. To test this, we regress case growth in a county on our index of exposure to out-of-county cases, controlling for lagged cases and other county attributes. The baseline model is

$$n_{j,t} = \theta_1 n_{j,t-1} + \theta_2 x_{jt} + \theta_3 m_{jt} + Z_j' \gamma + \varepsilon_{jt}, \tag{5}$$

where j, t are subscripts for county and time, respectively, n denotes new cases, x is the out-of-county exposure from equation (4), and m denotes the mobility index from (1).

The θ's are parameters of interest, and principally, the exposure parameter θ_2. The regression also independently controls for the county's mobility index as a form of placebo test; the mobility index term will reveal if higher mobility results in more cases for reasons other than exposure. Z is a set of controls for county characteristics, such as population size and density, or fixed effects to capture attributes nonparametrically. The ε is the error term.

The outcome variable is the natural logarithm of one plus the number of new cases reported in the last week.[22] The observation level is county by week beginning the first week of March when community-spread cases began to emerge in the U.S. The lag cases variable is the log number of new cases in the county in the preceding two weeks (i.e., one to three weeks prior to the observation date).

Column 1 of table 8 presents results from the baseline model. The regression shows two features of the contagion. First, and unsurprisingly, lagged cases in the county create new cases. A one-percent rise in past cases is associated with a 0.74 percent rise in new case growth. Second, and more novel, exposure to out-of-county cases also increases case growth. A one percent rise in outside exposure is associated with a 0.12 percent increase in case growth locally. However, mobility alone, absent an accounting of cases in the visited

[22] We use weekly case growth at the county level to limit the noise present in the daily series.

Table 8: Case Exposure and New Case Growth

	1	2	3	4	5	6	7
Model	OLS	OLS	OLS	OLS	IV 2SLS	IV GMM	IV GMM
Network Case Expo.	0.123	0.192	0.133	0.115	0.151	0.120	0.217
	(0.006)	(0.010)	(0.007)	(0.008)	(0.007)	(0.012)	(0.016)
Mobility Index	-0.002	-0.005	-0.003	-0.003	-0.003	0.006	-0.022
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.002)	(0.003)
Local Cases	0.742	0.619	0.739	0.712	0.737	0.755	0.696
	(0.004)	(0.005)	(0.005)	(0.005)	(0.005)	(0.008)	(0.010)
Population	0.096		0.117	0.156	0.071	0.128	-0.048
	(0.009)		(0.010)	(0.011)	(0.010)	(0.021)	(0.027)
Pop. Density	0.052		0.030	0.027	0.049	0.051	0.048
	(0.006)		(0.007)	(0.008)	(0.006)	(0.006)	(0.009)
Constant	-1.297	0.114	-1.474	-1.378	-0.253	-1.402	2.149
	(0.071)	(0.081)	(0.077)	(0.084)	(0.077)	(0.379)	(0.487)
Fixed Effects							
Level(s)	Week	County; Week	Region X Week	State X Week	Week	Week	Week
Number	12	2018; 12	108	612	12	12	12
Instruments:							
Home County Orders						y	
Destination Orders							y
Projected Exposure					y	y	y
R^2	0.861	0.868	0.865	0.875	0.861	0.858	0.848
NT	24,038	24,038	24,038	24,038	24,038	24,038	24,038

NOTES: The table reports regression results of the model represented by equation (5). The outcome variable is the log number of new cases in the county. The observation level is county by week. Standard errors are clustered by county. Source: Authors' calculations using data retrieved as described in Section 2.

network, does not explain the case growth. Its coefficient is slightly negative, owing to the pullback in mobility during the period of peak case growth. Thus it appears the actual network exposure to active cases is meaningful, but travel activity, generally, is not.[23]

As noted above, there was regional heterogeneity in the severity of the outbreak and a predictable geographic component to the observed travel network, and thus an alternative explanation to the effect of exposure is spatial correlation in travel and case outcomes. In other words, specification 1 could be simply picking up overlap in the regional components of each variable. To address the potential for spatial correlation, column 2 includes county

[23]One specific implication is that counties with less reduction in mobility fared no worse in terms of case growth, except to the extent captured by the exposure measure.

fixed effects to absorb any county-level attributes that may contribute to case growth, relying on within-county variation in exposure over time. This results in an even larger exposure estimate than column 1. Columns 3 and 4 include time interacted with census division and state, respectively. (The observation level is county by week.) Each of these arrives at similar estimates as column 1. We conclude that despite the spatial correlation in the outcome variable, the travel network is introducing independently meaningful variation through case exposure.

We have already shown that mobility decreased in response to rising case counts, which suggests some degree of reverse causality is present and may bias the effects we estimate. We therefore instrument for actual exposure with the predicted exposure that would have occurred had the county's mobility not changed from the pre-pandemic period, reported in column 5. The coefficient on exposure rises somewhat to 0.15, suggesting some attenuation from reverse causality. For the mobility index, the stay-at-home orders and related closures form natural instrumental variables (IV). The most straightforward IVs to use are the home county's social distancing orders. Column 6 expands to a GMM estimator, instrumenting for exposure and mobility using predicted exposure and the three forms of governmental activity restrictions from Table 3 within the home county. This results in no discernible effect on the exposure coefficient, but the mobility index does flip to the intuitive sign. However, the home county orders may not satisfy the exclusion restriction if they suppress new case diagnoses through a channel other than out-of-county mobility (as it appears was their intended use), such as fewer within-county contacts. Therefore our preferred instrument for the mobility index is the pre-period mobility-weighted sum of closures in potential destination counties from column 4 of Table 3. These indices affect out-of-county mobility but are mechanically detached from the home county. Column 7 reports this specification, showing no impact on the mobility coefficient but an increase in the exposure coefficient.

In summary, we find robust evidence that out-of-county exposure via the travel network affects new case diagnoses. In the appendix, we run a gamut of robustness checks on the functional form and spatial scale of the exposure measure. Specifically, we build permutations of alternative exposure indices of $x_{j,t} = \sum_i (d_{j,t}\sigma_{ijt})^{\alpha_1} n_{it}^{\alpha_2}$, where the α terms are in the list $\{0.5, 1, 1.5, 2\}$. These allow extreme observations to be more (when $\alpha_{1(2)} > 1$) or less (when $\alpha_{1(2)} < 1$) important to the index value. As Table 6 would suggest, the extreme values of cases tend to be more predictive, but overall, none of the conclusions are materially changed. The baseline linear model provides the best fit.

The robustness checks for spatial scale split the exposure along geographic lines: in or out of metro area, state, neighbor/adjacency status, or frequently-visited pair (top 10 destinations for a county). Interestingly, the more distant exposures are more predictive than the proximate/frequently visited destinations, but again, qualitative conclusions are unaffected.

Also in the appendix we present a brief look at heterogeneity in effects. We find that own-county case spread is smaller in locations with stay-at-home orders, but also smaller in larger counties. The effects of out-of-county exposure, however, are larger in more populous and denser counties, and also in counties with stay-at-home orders enacted. These patterns suggest that outside exposure is multiplied in bigger, more urban areas, and that outside exposure can undo some of the suppression effects of stay-at-home orders.

7 Geographic Exposure and Virus Spread

We have marshaled evidence for three important facts: (i) Spatial activity dropped significantly as case counts rose, with a particular avoidance of areas with relatively larger outbreaks; (ii) Such a drop in activity limited exposure to out-of-county virus cases; (iii) Out-of-county exposure matters for predicting new case growth, suggesting cases would have been higher had travel activity not dropped in response to cases. We are now prepared to summarize the combined effect of these components and, in particular, evaluate conjectures about spread of the virus in alternative travel activity scenarios.

Our last exercise is to simulate a simple model of virus spread over space. We construct the following spatial autoregressive (SAR) model of mobility, case exposure, and case growth. The primary outcome of interest is new case growth, in equation (6a), which is affected by own-county and out-of-county exposure (the latter being the "spatial" element of the model). We take point estimates from Table 8, column 3. To derive exposure, we model the effect of mobility on exposure to out-of-county cases (6b), from our definition in equation (4). Exposure is a function of mobility, which is itself affected by cases (equation 6c). We take point estimates from Table 4, column 2. To illustrate the importance of endogenous spatial activity in affecting the rate of disease spread, we simulate the model without mobility as a default, and then with mobility but with and without the feedback effect of cases on mobility.

$$n_{jt} = \theta_1 n_{j,t-1} + \theta_2 x_{j,t-1} \tag{6a}$$

$$x_{j,t} = \sum_i \sigma_{ij,t} d_{j,t} n_{i,t} \tag{6b}$$

$$\sigma_{ij,t} d_{j,t} = \beta n_{j,t}(\delta_1 n_{j,t} + \delta_2 n_{i,t} + \delta_3 \bar{\sigma}_{ij,0} n_{i,t}) \tag{6c}$$

We must emphasize that this is only an SAR process and not an epidemiological model. In particular, there are no notions of recovery, death, or immunity among the population. (Indeed, our unit of analysis is a spatial area, not a person or household.) Our focus is the rate of growth of new cases with and without endogenous mobility, not the arc of a contagious disease in a population. We will note the values the model produces for the sake of exposition, but we intend this exercise to be more illustrative than empirical.

Accordingly, to keep the model simple, we illustrate a three location system. Two counties are calibrated with symmetric mobility rates to represent two closely connected counties, like that of typical neighbors from Table 2. A third location has connectedness of a typical distant within-region county.

The model is used for the following thought experiment: if an outbreak of new cases exogenously appears in one of the two connected locations, what happens to the spread of the disease locally and throughout the system? To illustrate the importance of endogenous spatial activity in affecting the rate of disease spread, we simulate the model in three scenarios: (1) without mobility as a default ((6a) alone), (2) with mobility but without the feedback effect of cases on travel ((6a) and (6b)), and (3) with mobility and endogenous feedback ((6a), (6b), and (6c)). Figure 6 plots the impulse responses for an experiment of 10 new cases dropped into the "treated" county.

The rate of own-county spread is below one, so that if there were no mobility (and consequently no out-of-county spread), the virus would eventually (but very slowly) die out in the treated county, as illustrated by the "isolated/no mobility" lines. But with out-of-county contact and infection, the disease jumps to the other locations, which then themselves perpetuate through own-county spread, leading to the subsequent re-infection of other places in the system, keeping the disease alive in perpetuity. In this calibration, the case growth rates reach a steady state.[24]

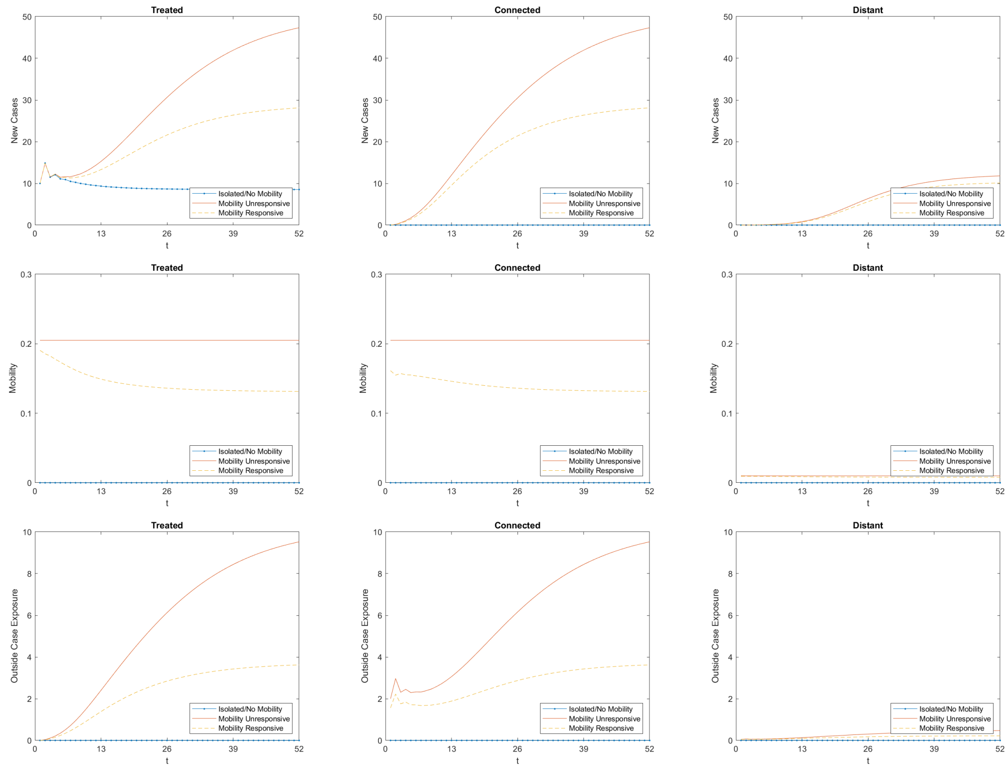
The second row of subfigures shows the mobility rate. The "unresponsive" scenario is fixed to have no endogenous change in mobility and mechanically results in the flat lines. In the endogenous mobility scenario, we see spatial activity fall as the outbreak occurs. In the two focal counties, mobility falls by almost 40 percent, though the untreated county drops first because the effect of avoiding the destination is larger. The consequence of the mobility drop is a reduction in exposure, shown in the third row of subfigures, which falls by over 60 percent by the end of the simulation period.

The drop in exposure then lowers the rate of disease transmission, creating the gap in new case growth among scenarios shown in the first row of subfigures. In the treated county with the exogenous outbreak, there is an initially oscillating pattern, as own-county case growth slows but exposure to outside cases rises. Eventually, the new case growth reaches a constant level, but about 70 percent higher in the scenario without mobility response.

In the connected county, spared the initial outbreak, it was out-of-county exposure that

[24] A steady state is not a general feature of the model. The parameters we find at the lags we estimate them (2 weeks) happen to have a steady state growth rate, but in general, explosive paths are possible.

Figure 6: Simulated Viral Spread Across Locations



NOTES: The figures report the time path of the variables in the dynamic system represented by equations 6a, 6b, and 6c. Each line refers to a separate scenario using different assumptions about the reaction of mobility to local and nonlocal cases. Source: Authors' calculations using estimates from Tables 4 and 8.

seeded the local outbreak. When mobility does not decline in response to the outbreak, the rate of new case growth is faster, and it reaches the same steady state level as the treated location by the end of the simulation.

The distant county experiences its own outbreak, although its lower connectivity translates into a lower long run average rate of exposure, so its steady state is lower than the two closely integrated counties.

In summary, the model shows why spatial connectedness matters for both the spread and the perpetuation of the virus. Most directly, nonlocal exposure allows the virus to jump from one area to another. Moreover, nonlocal exposure affects not only the rate of growth of cases, but also the steady state level. That is, connectedness generates higher caseloads as travel compounds baseline levels of transmission through reinfection across areas.

8 Conclusion

This paper has used county level location data from cellular devices to document the change in spatial activity during the early phase of the COVID-19 pandemic in the U.S. We find that mobility across counties dropped substantially as case counts rose. Relatively larger case counts decreased spatial activity on both sides of a trip: Mobility decreased more in counties with more cases, and moreover, the activity that did occur tended to avoid areas with higher case counts. Empirically, case avoidance is able to explain most of the observed drop in activity. Restrictions on movement, such as closure of nonessential services and stay-home orders, also contributed to the drop in activity but to a lesser extent than case avoidance.

Understanding the nature of the change in spatial activity is important because mobility across county lines produces contact with nonlocal cases. Such case exposure contributes to local case growth which in turn has a feedback effect on nonlocal case growth, in that it creates exposure for other localities, and so on in an endogenous loop. Such connectedness means the presence of the virus anywhere in the system is a threat everywhere else in the system.

Our findings have several implications for policy and practice. First, public information about the spread of the virus is important. We find people responding to such information by restricting their activity in rational ways–both in level and in direction. In a sense, a "healthy fear" of the virus appears an important applicator for social distancing and similar behavioral interventions, perhaps even more so than government mandates.

Second, because spatial activity never entirely disappears, localities should coordinate responses and share information. Connectedness will cause the virus to percolate through the system, or in other words, there are spatial externalities. A policy that suits one area may inadvertently produce a threat to a connected area. The adage that "we are all in it together" seems to apply.

This is especially true for more connected regions and localities, as we found that most exposure is produced by contact with areas of high outbreak. Our third implication is guidance for ways that health authorities can monitor the outbreak across areas and evaluate their own level of exposure. When two cities have a high amount of activity between them, an outbreak in one poses a direct and immediate threat to the other. This too suggests that fragmented policy across regions could inhibit society's ability to control the spread of COVID-19.

References

Atalay, Enghin, Shigeru Fujita, Sreyas Mahadevan, and Ryan Michaels, "Essential Industries, Contact Intensity, and Working from Home" Manuscript 2020.

Bailey, Michael, Rachel Cao, Theresa Kuchler, Johannes Stroebel, and Arlene Wong. "Social connectedness: Measurement, determinants, and effects." *Journal of Economic Perspectives* 32, no. 3 (2018): 259-80.

Christakis, Nicholas A., and James H. Fowler. "Social network sensors for early detection of contagious outbreaks." *PLOS ONE* 5, no. 9 (2010).

Courtemanche, Charles, Joseph Garuccio, Anh Le, Joshua Pinkston, and Aaron Yelowitz. "Strong Social Distancing Measures In The United States Reduced The COVID-19 Growth Rate." *Health Affairs* 39, no. 7 (2020).

Couture, Victor, Jonathan I. Dingel, Allison E. Green, Jessie Handbury, and Kevin R. Williams. "Measuring movement and social contact with smartphone data: a real-time application to COVID-19." No. w27560. National Bureau of Economic Research, 2020.

Coven, Joshua, and Arpit Gupta. "Disparities in mobility responses to COVID-19." NYU Stern Working Paper, 2020.

Dingel, Jonathan I., and Brent Neiman. "How many jobs can be done at home?." No. w26948. National Bureau of Economic Research, 2020.

Engle, Samuel, John Stromme, and Anson Zhou. "Staying at home: Mobility effects of COVID-19." Working Paper (2020).

Fang, Hanming, Long Wang, and Yang Yang. "Human mobility restrictions and the spread of the novel coronavirus (2019-ncov) in China." No. w26906. National Bureau of Economic Research, 2020.

Farboodi, Maryam, Gregor Jarosch, and Robert Shimer. "Internal and external effects of social distancing in a pandemic." No. w27059. National Bureau of Economic Research, 2020.

Fenichel, Eli P., Kevin Berry, Jude Bayham, and Gregg Gonsalves. "A cell phone data driven time use analysis of the COVID-19 epidemic." *medRxiv* (2020).

Glaeser, Edward L., Caitlin S. Gorback, and Stephen J. Redding. "How Much does COVID-19 Increase with Mobility? Evidence from New York and Four Other U.S. Cities." NBER Working Paper No. 27519. (2020).

Guerrieri, Veronica, Guido Lorenzoni, Ludwig Straub, and Ivan Werning. "Macroeconomic Implications of COVID-19: Can Negative Supply Shocks Cause Demand Shortages?." No. w26918. National Bureau of Economic Research, 2020.

Gupta, Sumedha, Thuy D. Nguyen, Felipe Lozano Rojas, Shyam Raman, Byungkyu Lee, Ana Bento, Kosali I. Simon, and Coady Wing. "Tracking public and private response to the COVID-19 epidemic: Evidence from state and local government actions." No. w27027. National Bureau of Economic Research, 2020.

Johns Hopkins University Coronavirus Resource Center. https://coronavirus.jhu.edu/ (2020).

Kaplan, Greg, Benjamin Moll, and Gianluca Violante. "Pandemics according to HANK." Powerpoint presentation, LSE 31 (2020).

Kraemer, Moritz UG, Chia-Hung Yang, Bernardo Gutierrez, Chieh-Hsi Wu, Brennan Klein, David M. Pigott, Louis Du Plessis et al. "The effect of human mobility and control measures on the COVID-19 epidemic in China." *Science* 368, no. 6490 (2020): 493-497.

Kuchler, Theresa, Dominic Russel, and Johannes Stroebel. "The geographic spread of COVID-19 correlates with structure of social networks as measured by Facebook." No. w26990. National Bureau of Economic Research, 2020.

Monte, Ferdinando. "Mobility Zones." No. w27236. National Bureau of Economic Research, 2020.

Appendix: Additional Exhibits

Table 9 reports robustness checks on the baseline exposure-to-case growth model in Table 8. The upper panel displays results from permutations of the exponents used to weight the exposure metric in (4). In the second set of rows (in which $\alpha_1 = 1$), columns 2 and 6 correspond to the estimates in the baseline (Table 8, columns 1 and 2). Coefficients decline mechanically as the mean of the index increases, but otherwise results are similar across specifications. Over weighting the mobility component relative to the case count component depresses coefficients, likely because what matters for variance in exposure is contact with high caseload areas (see Table 6).

The lower panel reports results when exposure is measured separately for mutually exclusive geographic areas. The table shows larger marginal affects for contact with farther-away areas than more proximate areas.

Table 10 reports results from a county heterogeneity analysis of the baseline exposure model. The results indicate the effect of local cases on new case growth is depressed by stay-at-home orders. The effect of outside exposure is magnified in larger and denser counties, and also in locations under stay-at-home orders.

Table 9: Robustness Checks: Case Exposure and New Case Growth

A. Functional Form

	1	2	3	4	5	6	7	8
	OLS				OLS FE			
$\alpha_2 =$	$\frac{1}{2}$	1	$\frac{3}{2}$	2	$\frac{1}{2}$	1	$\frac{3}{2}$	2
$\alpha_1 = \frac{1}{2}$								
[mean]	[8.331]	[10.86]	[14.07]	[17.63]	[8.331]	[10.86]	[14.07]	[17.63]
Coef.	0.169	0.165	0.130	0.102	0.297	0.317	0.226	0.160
(se)	(0.011)	(0.009)	(0.007)	(0.006)	(0.017)	(0.014)	(0.010)	(0.008)
R^2	0.859	0.860	0.860	0.860	0.866	0.867	0.867	0.867
$\alpha_1 = 1$								
[mean]	[10.11]	[12.48]	[15.50]	[18.93]	[10.11]	[12.48]	[15.50]	[18.93]
Coef.	0.126	0.123	0.101	0.085	0.153	0.192	0.161	0.127
(se)	(0.006)	(0.005)	(0.004)	(0.004)	(0.009)	(0.008)	(0.006)	(0.005)
R^2	0.860	0.861	0.861	0.860	0.866	0.868	0.868	0.867
$\alpha_1 = \frac{3}{2}$								
[mean]	[12.76]	[14.87]	[17.54]	[20.68]	[12.76]	[14.87]	[17.54]	[20.68]
Coef.	0.067	0.074	0.069	0.062	0.075	0.109	0.111	0.096
(se)	(0.004)	(0.003)	(0.003)	(0.003)	(0.005)	(0.005)	(0.005)	(0.004)
R^2	0.860	0.860	0.861	0.860	0.866	0.867	0.867	0.867
$\alpha_1 = 2$								
[mean]	[15.84]	[17.75]	[20.12]	[22.91]	[15.84]	[17.75]	[20.12]	[22.91]
Coef.	0.040	0.048	0.048	0.046	0.044	0.066	0.077	0.075
(se)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)	(0.003)	(0.003)
R^2	0.859	0.860	0.860	0.860	0.865	0.866	0.867	0.867

B. Destination

	1	2	3	4	5	6	7	8
	OLS				OLS FE			
α_2	State	CBSA	Neighbor	Top 10 Visit	State	CBSA	Neighbor	Top 10 Visit
Inside/Same								
[mean]	[10.60]	[4.126]	[8.679]	[10.61]	[10.60]	[4.126]	[8.679]	[10.61]
Coef.	0.044	0.009	0.036	0.047	0.061	0.029	0.060	0.058
(se)	(0.002)	(0.000)	(0.001)	(0.002)	(0.003)	(0.001)	(0.002)	(0.003)
Outside/Other								
[mean]	[11.92]	[12.30]	[12.29]	[11.97]	[11.92]	[12.30]	[12.29]	[11.97]
Coef.	0.089	0.091	0.082	0.075	0.158	0.151	0.126	0.135
(se)	(0.005)	(0.006)	(0.005)	(0.006)	(0.009)	(0.009)	(0.009)	(0.010)
R^2	0.861	0.860	0.862	0.861	0.868	0.868	0.871	0.868

NOTES: The table reports robustness results for the model represented by equation (5). The outcome variable is the log number of new cases in the county. The observation level is county by week. The upper panel uses different calibration for the exponents in the exposure measure, equation 4. The lower panel uses exposure measured at different geographies. Source: Authors' calculations using data retrieved as described in Section 2.

Table 10: County Heterogeneity: Case Exposure and New Case Growth

	1	2
Model	OLS	OLS FE
Population	-0.213	
	(0.038)	
Pop. Density	-0.011	
	(0.026)	
Local Cases	0.750	0.575
	(0.006)	(0.007)
X Population	-0.007	-0.007
	(0.004)	(0.005)
X Pop. Density	0.006	-0.007
	(0.003)	(0.005)
X Stay-at-home Order	-0.049	-0.065
	(0.006)	(0.007)
Mobility Index	0.002	0.003
	(0.000)	(0.000)
Network Case Expo.	0.299	0.377
	(0.032)	(0.035)
X Population	0.199	0.356
	(0.018)	(0.019)
X Pop. Density	0.045	0.105
	(0.016)	(0.016)
X Stay-at-home Order	0.027	0.042
	(0.006)	(0.008)
Constant	0.343	-0.883
	(0.182)	(0.080)
Time Effects	y	y
County FEs		y

NOTES: The table reports regression results of the model represented by equation (5) but with county attributes interacted with the main variables of interest. The outcome variable is the log number of new cases in the county. The observation level is county by week. Source: Authors' calculations using data retrieved as described in Section 2.