NBER WORKING PAPER SERIES

SOCIAL DISTANCING, INTERNET ACCESS AND INEQUALITY

Lesley Chiou
Catherine Tucker

Working Paper 26982
http://www.nber.org/papers/w26982

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
April 2020

At least one co-author has disclosed a financial relationship of potential relevance for this research. Further information is available online at http://www.nber.org/papers/w26982.ack

Social Distancing, Internet Access and Inequality
Lesley Chiou and Catherine Tucker
NBER Working Paper No. 26982
April 2020
JEL No. L96,L98,M15

ABSTRACT

This paper measures the role of the diffusion of high-speed Internet on an individual's ability to self-isolate during a global pandemic. We use data that tracks 20 million mobile devices and their movements across physical locations, and whether the mobile devices leave their homes that day. We show that while income is correlated with differences in the ability to stay at home, the unequal diffusion of high-speed Internet in homes across regions drives much of this observed income effect. We examine compliance with state-level directives to avoid leaving your home. Devices in regions with either high-income or high-speed Internet are less likely to leave their homes after such a directive. However, the combination of having both high income and high-speed Internet appears to be the biggest driver of propensity to stay at home. Our results suggest that the digital divide---or the fact that income and home Internet access are correlated---appears to explain much inequality we observe in people's ability to self-isolate.

Lesley Chiou
Occidental College
1600 Campus Road
Los Angeles, CA 90041
lchiou@oxy.edu

Catherine Tucker
MIT Sloan School of Management
100 Main Street, E62-533
Cambridge, MA 02142
and NBER
cetucker@mit.edu

1 Introduction

Countries across the world are experiencing unparalleled disruption due to the coronavirus (COVID-19) pandemic. In order to avoid overburdening the health system, many countries and regions decided to announce directives that encourage individuals to remain in their homes. The idea is that policies of social distancing will help stem the spread of a viral pandemic (Glass et al., 2006). However, as with many forms of governmental interventions, questions about equity arise. Various commentators in the press speculated that in practice, remaining at home as a policy is only accessible to those with high income. For example The New York Times published an article entitled, "White-Collar Quarantine Over Virus Spotlights Class Divide."[1] This paper highlights that while the debate is focused on income, much of the observed inequality is explained by differences in the diffusion of high-speed Internet to homes.

To investigate this question, we use data from a panel of 20 million mobile devices provided by a company named Safegraph. Safegraph tracks the location of these devices after people provide consent for mobile apps to track their precise location over time. This data allow us to observe when people leave their homes, and when they stay at home for the entire day. We combine this with data on income levels and Internet use by region from the Census American Community Survey.

We show that in February 2020, when the effects of the coronavirus pandemic were not clear to most people in the US, devices located in high-income regions were more likely to leave the home. However, in March 2020 this pattern reversed, and remaining at home became strongly and positively correlated with household income by region. This correlation disappears when we control for access to high-speed internet. It appears that access to high speed Internet—and the fact that Internet access at home is correlated with income—explains

[1] See `https://www.nytimes.com/2020/03/27/business/economy/coronavirus-inequality.html`

much of the observed disparity between high-income and low-income regions.

We then show that when states enacted directives encouraging people to stay at home, people living in high-income or high-Internet areas were more likely to increase their propensity to stay at home. We find also that the particular combination of a region having high-income and having more access to high-speed Internet, leads people to be far more likely to stay at home. In other words, the combination of high-income and high-Internet diffusion appears to be a large driver in observed inequality.

This paper contributes to what we believe will be a large literature that tries to understand the economic consequences of the COVID-19 pandemic. There are already multiple papers that try to calibrate the likely effect of social distancing measures on the spread of coronavirus within the US (Greenstone and Nigam, 2020; Stock, 2020; Berger et al., 2020). Other papers examine recent data from China to try to measure the effect of self-isolation on the spread of the virus (Fang et al., 2020). By contrast, we investigate the underlying economic factors that drive an individual's ability to self-isolate and protect themselves and the community from the spread of coronavirus.

Our paper also builds on a literature in digital economics that tries to measure the relationship between access to the Internet and inequality. Since the early days of the Internet, concerns existed that access to the Internet might echo or even reinforce existing sources of inequality (Keller, 1995; Servon, 2008). Early research documented the digital divide in electronic commerce (Hoffman et al., 2000) and Internet usage (Goldfarb and Prince, 2008). Since then, there have been some efforts to try to quantify the effects of certain digital technologies on the rich relative to the poor (Miller and Tucker, 2011; Tucker and Yu, 2019). We contribute to this literature by being the first paper to our knowledge that examines whether the relationship between access to the Internet and income affects a community's ability to enforce self-distancing. We present evidence that in general, high-speed Internet penetration helps regions comply with social distancing. However, regions with both high-speed Internet and high-income are far more likely to stay at home after a state directive, suggesting that it is the particular combination of high-speed Internet access and high-income that can exacerbate inequality. This is an unexpected spillover from the diffusion of the Internet.

Our paper also helps to inform policy and public debate about the likely consequences of self-distancing measures. Community spread may be most severe in regions with low broadband Internet penetration. In regions where there is a greater degree of broadband Internet penetration and high-income level, it appears that there is an increased ability to stay at home after a state directive. This suggests that from a policy perspective, the digital divide may a be more relevant policy issue than ever.

2 Data

We use data provided by Safegraph for the purposes of studying the spread of coronavirus for February and March 2020. This data is built on a panel of 20 million devices that collect anonymous location data.[2] Each of the users of these devices has given permission for their location to be tracked by a variety of mobile apps. Safegraph matches the location of these devices to a variety of locations of branded physical retail locations within the US, and its main business is focused on providing data on retail traffic to firms and analysts. However, Safegraph also shared data with researchers who are working on measuring the effects of the spread of coronavirus. For this purpose, it released data which tracks whether a device appears to leave its home or whether the device stayed at home the entire day.

Safegraph describes its data collection process as follows: "The data was generated using a panel of GPS pings from anonymous mobile devices. We determine the common nighttime location of each mobile device over a 6 week period to a Geohash-7 granularity ($\sim$153m $\times$ $\sim$ 153m). For ease of reference, we call this common nighttime location, the device's 'home.'

[2] https://www.safegraph.com/blog/what-about-bias-in-the-safegraph-dataset

We then aggregate the devices by home census block group and provide the metrics set out below for each census block group." While this is a reasonable procedure for determining a device's natural home, it may lead to some misleading results for a variety of circumstances (e.g., if someone works at night, or regularly sleeps at a romantic partner's house.) However, given the need to ensure that the data is anonymized and not related to any one individual, there is no easy way of correcting such issues. Another issue is that someone may leave the house without taking their device, though that would be unusual. Given that we focus on changes over time, we believe that any measurement error introduced by these types of errors will not affect the direction of our results. We aggregate these measures up from the census block level to the census tract level.

This dataset is focused on the US. Two avenues exist whereby the data is potentially not representative. First, it does not represent behavior of people who do not have smartphones. The 2018 American Community Survey (ACS) suggested that 66% of the US population have a smartphone. Another source of bias is that the people who opt-in to allow their location to be tracked may not be representative of the population. Goldfarb and Tucker (2012) show that willingness to divulge personal information decreases with age. Athey et al. (2017) suggests that to some extent the decision to divulge information is highly contextual and can be easily shifted through small incentives and changes in interfaces, implying that at least potentially some randomness in the decision to divulge may exist. Though some selection bias appears likely, Safegraph performs a variety of analyses that suggest their data does line-up with Census data.[3]

We combine the Safegraph data with data from the 2018 American Community Survey (ACS).[4] We use the ACS data to construct estimates of household income, local demographic

[3] https://colab.research.google.com/drive/1u15afRytJMsizySFqA2EPlXSh3KTmNTQ#offline=true&sandboxMode=true

[4] This was released in November 2019 https://www2.census.gov/programs-surveys/acs/data/pums/2018/?

composition, and access to the Internet. The American Community Survey "Public Use Microdata Sample" (PUMS) is released at the geographic level of the "Public Use Microdata Areas" (PUMA). PUMAs are the smallest geography available. They are designed to have a population of roughly 100,000 or more people. We were able to match 98.70% of the census tracts in our Safegraph data to the appropriate PUMA. To measure the diffusion of the Internet, we use the measure of whether "Broadband (high-speed) Internet service such as cable, fiber optic, or DSL service" is available in the household. For income, we use household income—which is total household income reported for the last 12 months.

In addition, we collected data from the National Governors' Association about when each state had (if at all) issued an order which required only "essential businesses" to stay open.[5] In each case, we verified what date the order went into place and used that date. These directives differed in the extent to which they were framed as a "shelter in place" order, or a "safety at home" order. However, the policies did share a similar intention of minimizing social interactions by closing down the physical manifestations of most businesses. For the purposes of our paper, we measure the average effect of these orders and do not try and distinguish between their different features or what they defined as an essential business.

Last, we also collected data on the spread of coronavirus itself at the county level. We collected this data from The New York Times data repository.[6] In our specifications, we focus on using the spread of cases as our focal measure of the spread of coronavirus in the local county. This data is less granular than our data on whether devices stay at home, which is on the census tract level. However, it is the most granular data we can obtain. We also recognize that this is an imperfect measure of actual coronavirus spread due to unavailability of testing. However, it seems reasonable to think that reported cases are correlated with the real number of cases, and also that the reported number of cases may in itself have been a

[5] https://www.nga.org/wp-content/uploads/2020/03/Appendix-I-Essential-Business_3.31.20.pdf

[6] https://raw.githubusercontent.com/nytimes/covid-19-data/master/us-counties.csv

meaningful influence on people's behavior.

Table 1 provides summary statistics of the key variables. We have data on 72,444 census tracts each day for the months of February 2020 and March 2020.

Table 1: Summary Statistics

	Mean	Std Dev	Min	Max
% Stay Home	26.9	10.1	0.19	89.5
Device Count in Census Tract	256.8	235.4	0	62608
Reported Cases	43.6	280.4	0	9326
HH Income (0000)	6.33	2.28	2.10	18.6
Proportion Black	0.14	0.17	0	0.94
Proportion Asian	0.054	0.079	0	0.68
Proportion Unemployed	0.0050	0.0044	0	0.044
% 60+	0.21	0.054	0.058	0.59
Highspeed Internet	0.61	0.14	0.18	0.91
State Directive	0.11	0.32	0	1

Our key dependent variable is 26.9%. In other words, on average throughout our time period, nearly 27% of devices did not leave their house on a given day. The average region had 61% of households reporting access to high-speed Internet. The proportion of unemployed people appears low at 0.5%, but that is because the ACS defines this category as people who have been without a job for 5 years. Average penetration of high-speed Internet is 61% of households. The average household income is around $63,000. On average, there were 43 reported cases of coronavirus at the county level in our dataset, but this is skewed in particular by Seattle in the earlier period and New York in the latter period.[7]

3 Empirical Analysis

We begin by presenting some raw evidence of the two main results in the paper. Figure 1 shows that prior to the state directives, areas with above median income typically had more devices leave their homes than areas with below median income. However, after state directives were issued, this pattern reversed, and people living in high-income areas were less

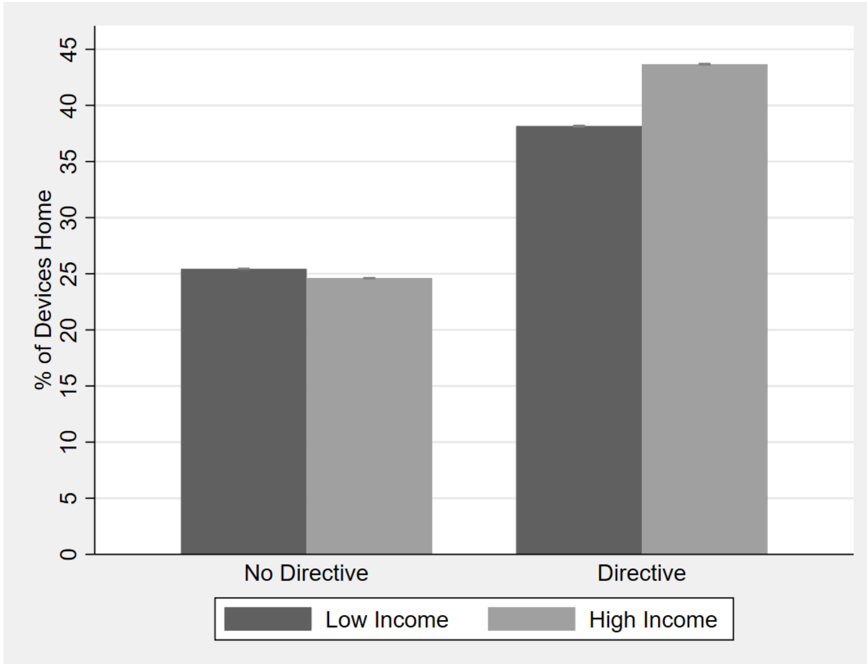
[7] We ran specifications with cases per capita and also the log of cases with similar results.

likely to leave their homes than people living in low-income areas. It also shows that after a state directive was issued, people were indeed more likely to stay in their homes. It is also useful to understand the geographical spread of high-income regions in the US. We map this in Figure A1. It suggests that though there is a concentration of high-income counties on the coasts, there is a reasonable spread of high-income counties geographically across the US.

Figure 2 shows the same plot for whether the region had high or low Internet penetration. It shows that in areas with high Internet, people were more likely to comply with a state directive. What is striking though is how similar Figures 1 and 2 are in magnitude and pattern. This reflects the fact that having a high income in a region and high broadband penetration is highly correlated. Indeed, the correlation is 0.65 for the indicator variables, and 0.74 for the raw variables. This paper tries to tease apart the extent to which much of the inequality in self-isolation that has been attributed to disparities in income can actually be explained by disparities in Internet access. One immediate question is whether or not the spread of high speed Internet is focused in certain regions and whether that might be correlated in a systematic way with the spread of the coronavirus. Figure A2 in the Appendix indicates that regions with high Internet penetration have reasonably broad coverage across the country and not are concentrated in one region.

It is also worth discussing in the context of Figure 2 why it is that broadband Internet might still play a large role in behavior, even if people have access to their Internet through their mobile devices. Evidently all individuals in the sample have access to the Internet, as they are being tracked through their mobile devices. We highlight three potential avenues. First, in general cellular plans have data limits which mean that using mobile phones for data-intensive uses, such as watching movies or conducting video calls, may be prohibitively expensive. Second, most cellular plans have limits on whether their users can "tether"

Figure 1: High-Income Areas are More Likely to Stay at Home Following a State Directive



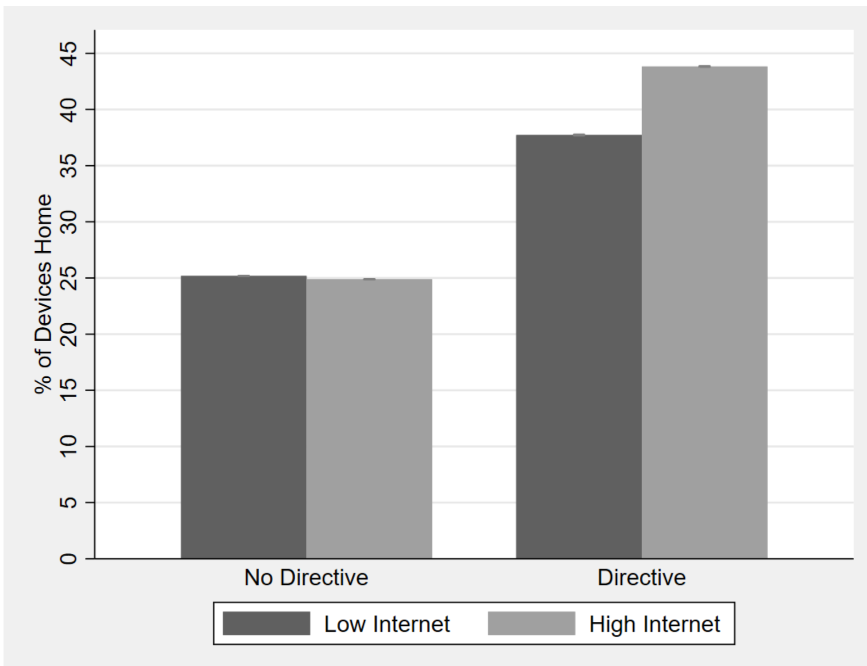
Notes: February and March 2020 data. High income is defined by whether that PUMA region has above-median household income.

them—that is, use them as a general Internet modem for a house.[8] Mobile phones may also be a less than ideal substitute for an Internet-powered desktop or laptop for many work or recreational purposes.

Figure 3 expands this result by allowing for whether or not the region had above- or below- median Internet. It shows that in general after the directive, having access to high speed Internet led more people to stay at home in both high and low income regions.

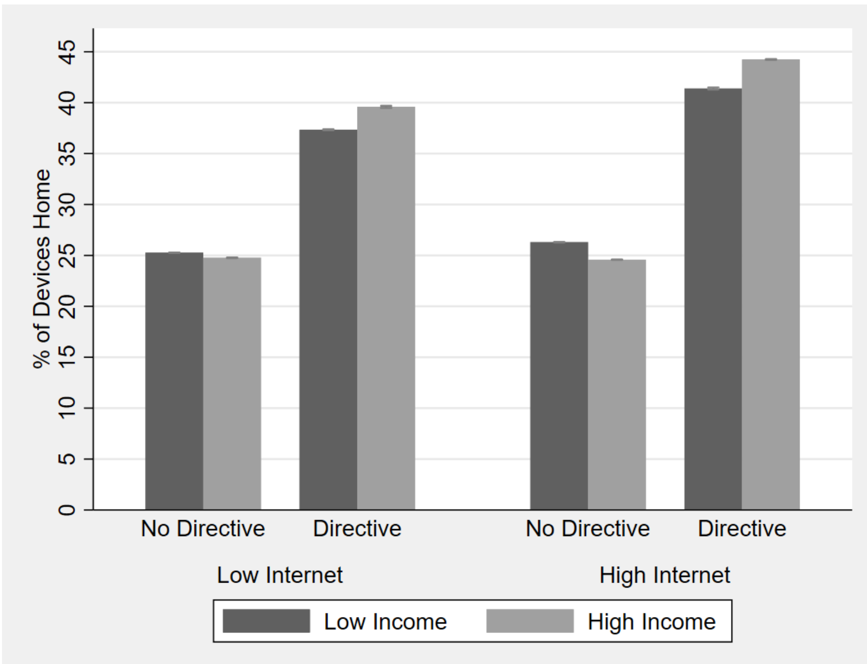
[8] `https://fortune.com/2015/09/17/cellphone-unlimited-data/`

Figure 2: High-Internet Areas are More Likely to Stay at Home Following a State Directive



Notes: February and March 2020 data. High-internet is defined by whether that PUMA region has above-median broadband penetration.

Figure 3: Internet Access Improves Everyone's Ability To Stay At Home



Notes: February and March 2020 data. High income is defined by whether that PUMA region has above median household income. High-internet is defined by whether that PUMA region has above average broadband penetration.

Table 2: Correlations Between Regional Characteristics And Staying at Home

	Feburary			March		
	(1)	(2)	(3)	(4)	(5)	(6)
	% Stay Home	% Stay Home	% Stay Home	% Stay Home	% Stay Home	% Stay Home
HH Income (0000)	-0.825***	-0.538***	-0.698***	0.248***	-0.248***	-0.128***
	(0.00765)	(0.0108)	(0.0119)	(0.00995)	(0.0144)	(0.0150)
Reported Cases	0.107***	0.00326***	0.0118	0.00253***	0.00250***	0.00219***
	(0.0166)	(0.0000569)	(0.0161)	(0.0000533)	(0.0000523)	(0.0000485)
Highspeed Internet		5.797***	-0.0962		11.37***	9.054***
		(0.163)	(0.192)		(0.209)	(0.225)
Proportion Black			5.130***			6.132***
			(0.119)			(0.148)
Proportion Asian			1.710***			13.28***
			(0.255)			(0.331)
Proportion Unemployed			24.82***			42.86***
			(4.139)			(5.020)
% 60+			3.470***			-0.428
			(0.382)			(0.454)
Date Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
State Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2097446	4267162	2097446	2169716	2169716	2169716
R-Squared	0.368	0.544	0.378	0.579	0.586	0.598

Notes: Dependent variable is the percentage of devices which did not leave the designated home in that census tract. Robust standard errors clustered at census tract level are in parentheses. * $p < 0.05$, ** $p < 0.05$, *** $p < 0.001$

Though the results of Figures 1 and 2 are useful, they also do not control for many other shifts that were taking place during the time period. Neither do they control for underlying differences in regions. To address this, we turn to an econometric specification. Initially, we focus on a correlational analysis between regional demographic and economic characteristics and whether devices stay at home.

Table 2 presents some simple correlations about how the decision to stay home varies with various regional social, economic and demographic characteristics. In this specification, we use state fixed effects. This means that we are only looking at variation in income and Internet within a state rather than state-to-state variation. This also means that we are not picking up differences across states due to different timing of the outbreak.

The first three columns of Table 2 present results for February 2020, and the second three columns present results for March 2020. In general, during February people were more likely to leave their house if they lived in an area with a higher income. Column (4) shows that this pattern reverses in March, and high-income regions are more likely to stay at home. However, in Column (5) we show that this reversal was driven by the presence of high-speed Internet. In other words, the reason it appears as though high-income regions were more likely to stay at home, is that they also had access to high-speed Internet.

A comparison of Columns (2) and (5) illustrates that broadband penetration enables individuals to not leave their homes. Comparing February with March suggests that broadband penetration has a large and positive significant effect on whether people leave their homes during the pandemic.

Looking at the results for incremental demographics in Columns (3) and (6) also offer insights. In general, high levels of prolonged unemployment (more than 5 years) affect whether people leave their homes, but this is more pronounced in March, reflecting potentially the fact that those without jobs are more able to avoid leaving their house.

The additional demographic variables included in Table 2 also offer interesting results. It appears that areas where there are a higher fraction of individuals who are African-American, fewer devices leave the home. This trend only shifts slightly in March after the likely effects of the pandemic were more transparent. The largest shift we see in behavior is in areas with a high proportion of individuals who are Asian. These regions show a far higher increase of instances where the device did not leave the home in March.

One potentially troubling effect that we measure is the behavior of regions which have a larger proportion of people over the age of 60. In general, this group is more likely to be vulnerable to coronavirus complications. We do not observe in March any negative and significant relationship with this measure of the elderly in a local region, as might be ideal from a public health standpoint. However, we do observe that this compares favorably to the positive and significant coefficient we see in February. In addition, across all specifications, it is clear that the number of reported coronavirus cases in the local region have a significant effect at encouraging people to stay at home.

Though the results of Table 2 are useful from a policy perspective, they are correlational. We next turn to a more rigorous specification which exploits the precise timing of the state directives in our data.

Equation (1) presents our base econometric specification, where the proportion of devices located on services located in census tract i in county c in state s on date t are a function of:

$$\begin{aligned}\%StayHome_{it} = {} & \beta_1 HighIncome_i \times StateDirective_{st} \\ & + \beta_1 HighInternet_i \times StateDirective_{st} \\ & + \beta_3 StateDirective_{st} + \beta_4 ReportedCases_{ct} + \gamma_i + \delta_t + \epsilon_i \end{aligned} \tag{1}$$

We estimate this specification using ordinary least squares (OLS).[9] The focal policy effect studied in this paper is captured by β_1 and β_2, which capture the relationship between the enactment of a state directive and above median household income and Internet. The coefficient β_2, which captures the level effect of the state directive, is also of policy interest.

The coefficient γ_i is a vector of census-level fixed effects intended to capture baseline regional differences in how much people leave their house. In Table 2, a specification where we have fixed effects at the state level, allows us to see the direct effects of demographics on people's likelihood of staying home. However, since these demographics are of course collinear with the census-tract-level fixed effects, they are not present in this specification.

[9]Since our dependent variable is naturally bounded between 0 and 100, we also estimated an aggregate logit model that directly accounted for this. The results were similar, so we report OLS for ease of interpretation.

Table 3: Staying at Home: The Effect of State Directives

	(1) % Stay Home	(2) % Stay Home	(3) % Stay Home	(4) % Stay Home
State Directive	4.227***	0.320***	0.162**	-0.540***
	(0.0486)	(0.0565)	(0.0571)	(0.0582)
State Directive × High Income		6.853***		4.017***
		(0.0683)		(0.0886)
State Directive × High Internet			6.954***	4.236***
			(0.0684)	(0.0888)
Reported Cases	0.00302***	0.00247***	0.00245***	0.00235***
	(0.0000525)	(0.0000505)	(0.0000501)	(0.0000498)
Date Fixed Effects	Yes	Yes	Yes	Yes
Census Tract Fixed Effects	Yes	Yes	Yes	Yes
Observations	4267162	4267162	4267162	4267162
R-Squared	0.690	0.701	0.701	0.703

Notes: Dependent variable is the percentage of devices which did not leave the designated home in that census tract. Robust standard errors clustered at census tract level are in parentheses. * $p < 0.05$, ** $p < 0.05$, *** $p < 0.001$

For example, due to the census-tract fixed effects, the main effect of household income drops out of our specification. The coefficient δ_t is a vector of date fixed effects for each date in the sample period.

In general, these estimates can be interpreted causally—in the same manner as a regression discontinuity design, due to the combination of date and census-block fixed effects and the sharp timing differences across states of when these directives were imposed. However, we caution that it is appropriate to think of these estimates as reflecting everything that happened on that day in the state which led to the directive being issued, rather than necessarily the causal effect of the directive alone.

Table 3 presents results exploring the effect of state directives for the combined sample of February and March. We report similar results for the March-only subsample in Table A1 of the appendix. Column (1) shows the raw effect of the state directive that only essential businesses could stay open. It suggests that there is a 4.2 percentage point increase in devices remaining at home after the order. This is a relatively small effect relative to an average of 26.9% of devices not leaving the house suggested by Table 1. Perhaps this is unsurprising though, given the nationwide trend in people distancing themselves regardless of state directives.

Column (2) looks at how the effect of the state ban was moderated by income. The results are striking. In general, they suggest that state directives were far more effective at ensuring people stayed at home in high-income regions. Columns (3) suggests that there is a very similar pattern positive effect for the state having a high degree of broadband penetration. In general, these results suggest that in both high-speed Internet and high-income regions, state directives are more effective at encouraging people and their associated devices to remain at home. In Column (4), we look at the effect of including interactions for both income and Internet. In both cases, the individual effect is attenuated, but still positive and significant. The apparent negative effect of the state directive in column (4) may be an artifact of the combination of census tract and date fixed effects and the implied baseline. However, it does suggest that without either high-income or high-Internet being present, state directives did not have a large effect on people' staying at home.

A natural question given the strong correlation between high-income areas and the diffusion of high-speed Internet, is whether it is income or the diffusion of the Internet which is driving these results. This is especially the case given the correlations in Table 2 which suggested collinearity was driving the apparent effect of high-income on getting people to stay home.

We turn to examine this in Table 4. In this table we try and exploit variation across regions and examine regions which have high-Internet, but low-income, or low-Internet but high-income. These are of course potentially non-representative. For example, high-Internet but low-income regions are disproportionately found in states such as North Carolina. There, it appears that there have been spillovers of provision of high-speed Internet for the "triangle"

Table 4: Staying at Home: The Interaction Between Internet and Income

	High Internet (1) % Stay Home	Low Internet (2) % Stay Home	High Income (3) % Stay Home	Low Income (4) % Stay Home
State Directive	-0.665***	1.880***	-0.615***	1.740***
	(0.118)	(0.0615)	(0.130)	(0.0612)
State Directive × High Income	4.737***	3.215***		
	(0.121)	(0.130)		
State Directive × High Internet			5.037***	3.540***
			(0.132)	(0.120)
Reported Cases	0.00188***	0.00311***	0.00195***	0.00321***
	(0.0000558)	(0.0000835)	(0.0000558)	(0.0000818)
Date Fixed Effects	Yes	Yes	Yes	Yes
Census Tract Fixed Effects	Yes	Yes	Yes	Yes
Observations	2133014	2134148	2117902	2149260
R-Squared	0.750	0.667	0.745	0.673

Notes: Dependent variable is the percentage of devices which did not leave the designated home in that census tract. Robust standard errors clustered at census tract level are in parentheses. Data for March 2020 and February 2020. * $p < 0.05$, ** $p < 0.05$, *** $p < 0.001$

to poorer areas nearby. Some high-income but low Internet areas are driven by states such as Wyoming where presumably topography and density interfere with the provision of Internet. However, we also see other states such as Florida where lack of spread of high speed Internet does not seem to be driven by topography or density. Therefore, though it seems likely that there are unobserved characteristics of regions which drive whether or not they deviate from the usual correlation between high-income and high-Internet, they do not seem entirely systematic. In the Appendix we report a series of charts which maps out counties by whether they are above or below the median income or level of broadband diffusion. Figures A3 and A4 in the Appendix, suggest that these unusual regions were not concentrated in one area, but instead were spread out geographically.

We report in Columns (1) and (2) of Table 4 a specification which divides up the sample by whether it is high or low income (above or below median income), or high or low Internet (above or below median broadband penetration). These specifications give us similar insights. Columns (1) suggests that in general in regions where there is high-Internet, that a state-level directive only induces people to stay home if they live in the subset of high-Internet regions with high-incomes. Column (2) suggests that in low Internet regions the state directive positively affects both high-income and low income regions within this subset. Column (3) suggests that in regions with high-income, it is only when there is a high-Internet presence that the directive encourages people to stay at home. Column (4) suggests that in regions with low income, then the presence of high-Internet encourages people to stay at home in that subset of regions, but in general there is still a positive effect of the state directive.

One slightly jarring result presented in Table 4 is that in Columns (1) and (3) there is a negative coefficient on the state directive. At face value, this suggests that the state directive led regions that were not displaying the natural correlation between income and Internet penetration, to leave the house more after a state directive. However, it should be remembered that this measured effect is from a baseline which is determined by the combination of census tract fixed effects and date effects. Instead, it should be placed in the context of Columns (3) and (4)) which suggests that either the presence of high-income in a low Internet region, or the presence of high-Internet in a low-income region, does have a small positive effect.

In the Appendix in Table A2 we also present results for a version of Table 4 where we just examine data for March. Though the size of the coefficients are generally smaller, the relative size and direction are similar.

4 Conclusions

This paper is a first attempt at understanding the role of income inequality in moderating the effectiveness of social distancing measures taken in wake of the spread of coronavirus. Our results suggest that people who live in high-income areas are in general more likely to engage in activities outside the home. However, since March 2020, and in particular since the enactment of state directives defining what essential businesses were allowed to stay open, people living in high-income areas have in general have self-isolated more and not left their home. But this seems to be driven by the fact that high-income areas are also likely to have higher broadband diffusion. We present evidence that the presence of above average high-speed Internet in a region increases the ability of all residents to self-distance. However, it also exacerbates the difference between high-income and low-income regions, further cementing the digital divide.

This paper aims to guide policy. The sheer scale of state executive orders encouraging people to stay at home, is unparalleled in recent US history. Therefore, it is useful to measure whether they are effective and in what contexts they are likely to be less effective. Our results suggest that policymakers should worry about the likely lack of success of self-isolation policies in regions where there is low Internet penetration and where people have low incomes. The results also highlight unforeseen consequences of the scale (or lack of scale) of deployment of high speed Internet across the US in potentially exacerbating the effects of income inequality in the ability to self-isolate.

There are of course limitations to this research. First, this paper is written using data for February and March 2020, and it is uncertain how the pandemic will evolve, how policies to tackle it will evolve, and how the trends documented in this paper will evolve. Second, the paper is descriptive. Though our combination of date fixed effects combined with the exact timing of a state-wide stay at home order provides a sharp discontinuity to measure the effect of the state-wide order, we do not have exogenous variation that allows us to attribute causality to income or high speed Internet as moderating variables. Third, this is a paper focused on the US, and it is not clear how it applies to other countries during a global pandemic. Notwithstanding these limitations, we believe this paper represents a useful first contribution to the policy debate about the role of income inequality in moderating outcomes in the wake of national emergencies.

References

Athey, S., C. Catalini, and C. Tucker (2017). The digital privacy paradox: Small money, small costs, small talk. *National Bureau of Economic Research Working Paper*.

Berger, D., K. Herkenhoff, and S. Mongey (2020). An seir infectious disease model with testing and conditional quarantine. *University of Chicago, Becker Friedman Institute for Economics Working Paper* (2020-25).

Fang, H., L. Wang, and Y. Yang (2020, March). Human mobility restrictions and the spread of the novel coronavirus (2019-ncov) in china. Working Paper 26906, National Bureau of Economic Research.

Glass, R. J., L. M. Glass, W. E. Beyeler, and H. J. Min (2006). Targeted social distancing designs for pandemic influenza. *Emerging infectious diseases 12*(11), 1671.

Goldfarb, A. and J. Prince (2008). Internet adoption and usage patterns are different: Implications for the digital divide. *Information Economics and Policy 20*(1), 2–15.

Goldfarb, A. and C. Tucker (2012). Shifts in privacy concerns. *American Economic Review: Papers and Proceedings 102*(3), 349–53.

Greenstone, M. and V. Nigam (2020). Does social distancing matter? *University of Chicago, Becker Friedman Institute for Economics Working Paper* (2020-26).

Hoffman, D. L., T. P. Novak, and A. Schlosser (2000). The evolution of the digital divide: How gaps in internet access may impact electronic commerce. *Journal of computer-mediated communication 5*(3), JCMC534.

Keller, J. (1995). Public access issues: An introduction. In *Public access to the Internet*, pp. 34–45. MIT Press.

Miller, A. R. and C. E. Tucker (2011). Can health care information technology save babies? *Journal of Political Economy 119*(2), 289–324.

Servon, L. J. (2008). *Bridging the digital divide: Technology, community and public policy.* John Wiley & Sons.

Stock, J. H. (2020). Data gaps and the policy response to the novel coronavirus. Technical report, National Bureau of Economic Research.

Tucker, C. E. and S. Yu (2019). Does IT Lead to More Equal Treatment? An Empirical Study of the Effect of Smartphone Use on Customer Complaint Resolution. *Mimeo, MIT).*

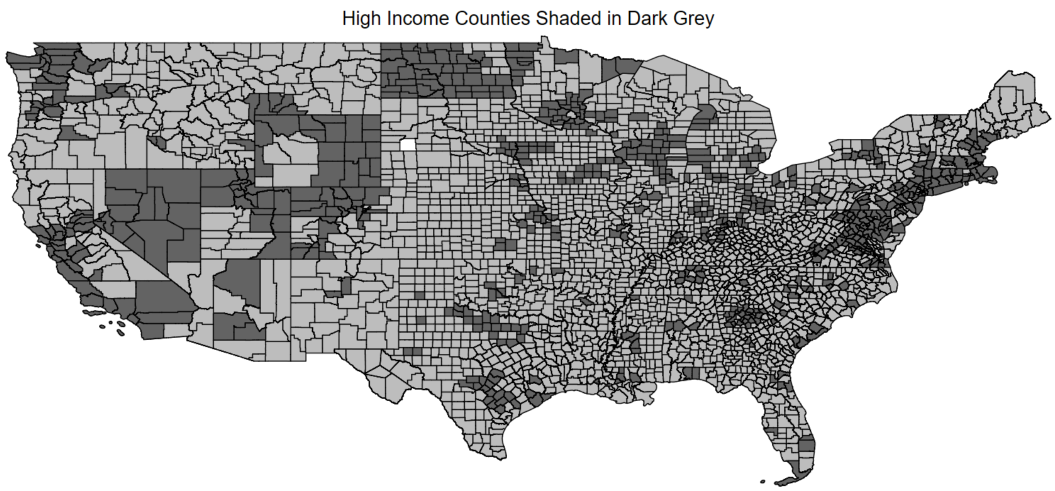


Figure A1: Distribution of high-income Counties

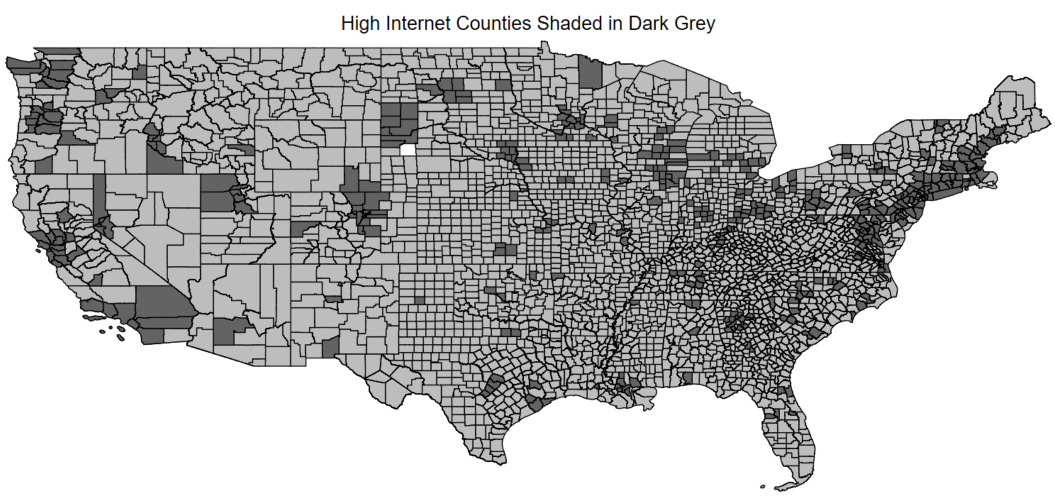


Figure A2: Distribution of High-Internet Counties

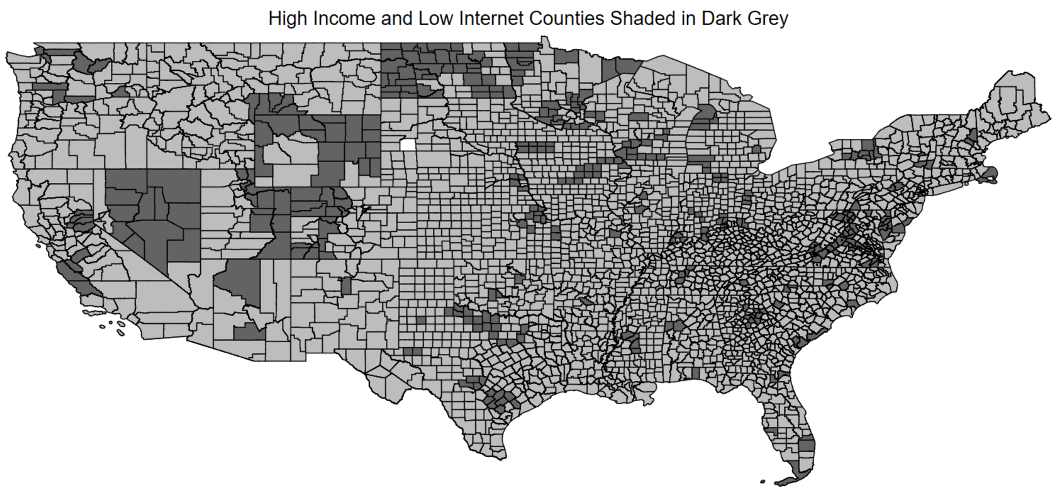


Figure A3: Distribution of high-income and Low-Internet Counties

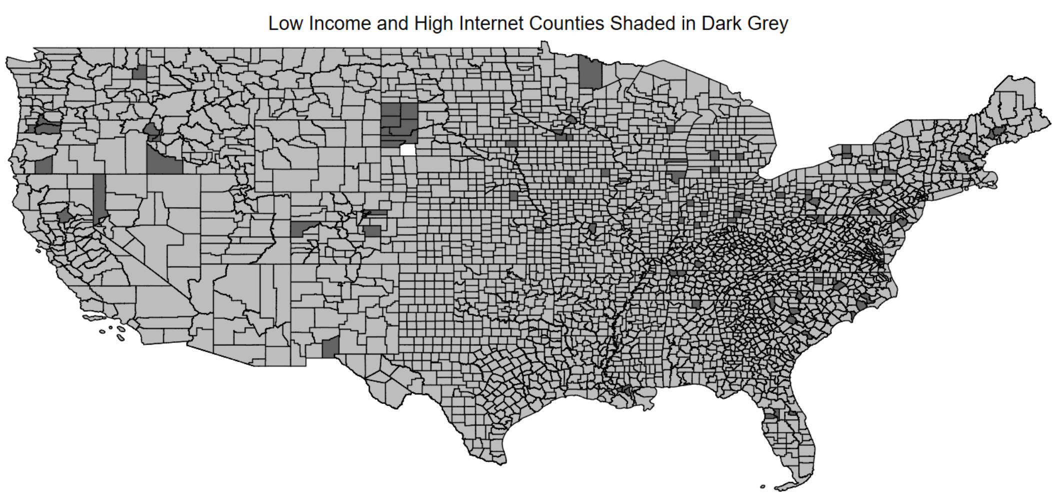


Figure A4: Distribution of Low-Income and High-Internet Counties

Table A1: Staying at Home: The Effect of State Directives (March Only)

	(1) % Stay Home	(2) % Stay Home	(3) % Stay Home	(4) % Stay Home
State Directive	3.376***	0.410***	0.338***	-0.216***
	(0.0374)	(0.0445)	(0.0449)	(0.0459)
State Directive × High Income		5.279***		3.183***
		(0.0556)		(0.0723)
State Directive × High Internet			5.284***	3.136***
			(0.0557)	(0.0723)
Reported Cases	0.00226***	0.00186***	0.00186***	0.00178***
	(0.0000403)	(0.0000389)	(0.0000388)	(0.0000385)
Date Fixed Effects	Yes	Yes	Yes	Yes
Census Tract Fixed Effects	Yes	Yes	Yes	Yes
Observations	2169716	2169716	2169716	2169716
R-Squared	0.771	0.779	0.779	0.780

Notes: Dependent variable is the percentage of devices which did not leave the designated home in that census tract. Robust standard errors clustered at census tract level are in parentheses. Data for March 2020 only. * $p < 0.05$, ** $p < 0.05$, *** $p < 0.001$

Table A2: Staying at Home: The Interaction Between Internet and Income (March Only)

	High Internet (1) % Stay Home	Low Internet (2) % Stay Home	High Income (3) % Stay Home	Low Income (4) % Stay Home
State Directive	-0.384***	1.691***	-0.152	1.623***
	(0.0966)	(0.0481)	(0.102)	(0.0482)
State Directive × High Income	3.622***	2.669***		
	(0.100)	(0.104)		
State Directive × High Internet			3.592***	2.708***
			(0.105)	(0.0992)
Reported Cases	0.00144***	0.00225***	0.00151***	0.00225***
	(0.0000440)	(0.0000671)	(0.0000438)	(0.0000672)
Date Fixed Effects	Yes	Yes	Yes	Yes
Census Tract Fixed Effects	Yes	Yes	Yes	Yes
Observations	1084567	1085149	1076869	1092847
R-Squared	0.808	0.753	0.808	0.757

Notes: Dependent variable is the percentage of devices which did not leave the designated home in that census tract. Robust standard errors clustered at census tract level are in parentheses. Data for March 2020 only. * $p < 0.05$, ** $p < 0.05$, *** $p < 0.001$