Social Distancing, Internet Access and Inequality

Lesley Chiou and Catherine Tucker*

April 13, 2020

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 track 19 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 drives much of this observed income effect. We examine compliance with state-level directives to remain at 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 an individual's ability to self-isolate.

*Lesley Chiou is Professor of Economics at Occidental College. Catherine Tucker is the Sloan Distinguished Professor of Marketing at MIT Sloan School of Management, Cambridge, MA and Research Associate at the NBER. See https://mitmgmtfaculty.mit.edu/cetucker/disclosure/ for disclosures. All errors are our own. We thank Safegraph for sharing the data used in the paper.

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 incomes. For example, *The New York Times* published an article entitled, "White-Collar Quarantine Over Virus Spotlights Class Divide."[1] Our research highlights that while the popular 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 how access to the Internet and income affects an individual's ability to stay at home, we use data from a panel of 19 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. These data allow us to observe when people leave their homes, and when they stay at home for the entire day. We supplement this with data on income levels and Internet use by region from the 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 the 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

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

speed Internet—and the fact that Internet access at home is correlated with income—explains 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 stay at home. In other words, the combination of high-income and high-Internet diffusion appears to be a large driver in observed inequality. We document two mechanisms for this result. First, that people who live in high-income and high-Internet areas also tend to have jobs that are amenable to telecommuting. Second, that people in high-income and high-Internet areas show a relative decline in physical visits to convenience stores after the directive.

This paper contributes to what we believe will be a large literature that tries to understand the economic consequences of the COVID-19 pandemic. Multiple papers are trying 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 their 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). 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 isolate itself in the wake of a pandemic. We present evidence that 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 the particular combination of high-speed Internet access and high-income can exacerbate inequality. This is an unexpected spillover from the diffusion of the Internet.

Our paper also helps to inform policymaking and the 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 with greater penetration of Internet and high income levels, individuals have an increased ability to stay at home after a state directive. This suggests that from a policy perspective, the digital divide may be a more relevant policy issue than ever.

2 Data

We use data provided by Safegraph for the purposes of studying the spread of coronavirus during February, March 2020 and the first week of April 2020. This data is derived from a panel of around 19 million devices that collect anonymous location data.[2] Each of the users of these devices gave 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 that 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

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

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.' 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 from the census block level up 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 optin 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 align with Census data.[3]

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

We combine the data from Safegraph with data from the 2018 American Community Survey (ACS).[4] We use the ACS data to construct estimates of household income, local demographic 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 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. If the order was effective after midday, we used the next date as the effective date for the order. 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 from the *New York Times* data repository.[6] In our specifications, we use the number of reported cases as our 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 at the census tract level.

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

[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

However, it is the most granular data we can obtain. We also recognize that this is an imperfect measure of actual spread of coronavirus due to lack of available testing. However, it seems reasonable that reported cases are correlated with the number of actual cases, and also that the reported number of cases may influence people's behavior.

Table 1 provides summary statistics of the key variables. We have data on 72,374 census tracts each day for Feb 1-April 7 2020.

Table 1: Summary Statistics

	Mean	Std Dev	Min	Max
% Stay Home	28.8	11.3	0.19	92.2
Device Count in Census Tract	255.7	230.5	0	62608
Reported Cases	143.3	774.8	0	16610
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
Local Urban Population Share	0.84	0.34	0	1
Proportion College Degrees	0.23	0.11	0.028	0.71
Highspeed Internet	0.61	0.14	0.18	0.91
State Directive	0.18	0.39	0	1

Our key dependent variable, which is the fraction of devices that stay at home, is on average 26.9%. In other words, on average throughout our time period, nearly 27% of devices did not leave their home on a given day. On average, a region had 61% of households reporting access to high-speed Internet. The proportion of unemployed people appears low at 0.5%, but this reflects that ACS defines unemployed 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 later period.[7]

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

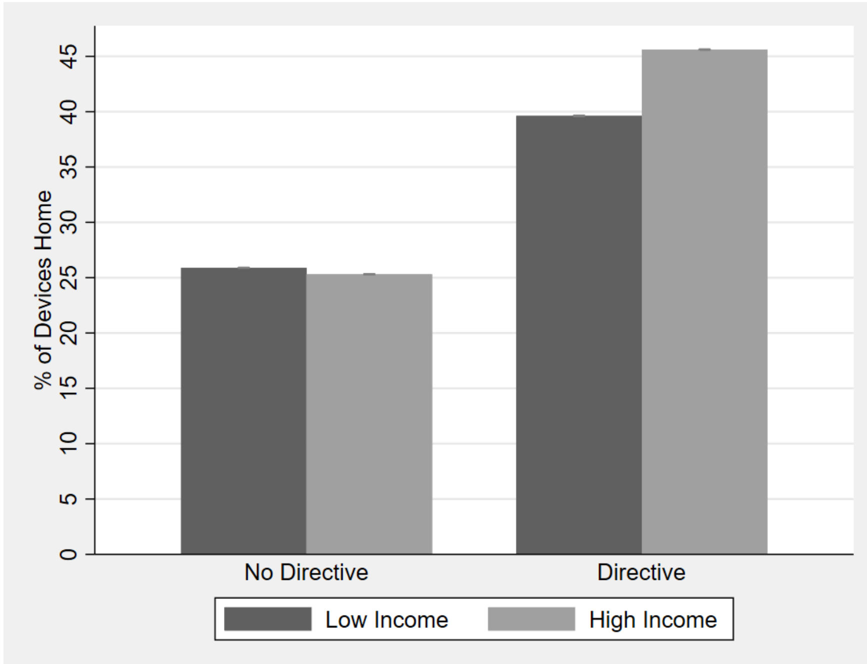
3 Empirical Analysis

We present some preliminary evidence of the two main results in the paper. Figure 1 illustrates 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 with people living in high-income areas far less likely to leave their homes than people living in low-income areas. The figure also shows that after a state directive was issued, all individuals were indeed more likely to comply and 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. The figure suggests that while a concentration of high-income counties exists on the coasts, a reasonable geographic spread of high-income counties exists across the US.

Figure 2 shows a plot of the fraction of people staying at home by high versus low Internet penetration. As seen in the figure, in areas with high Internet, people were more likely to comply with a state directive. What is striking is how similar Figures 1 and 2 are in magnitude and pattern. This reflects the fact that high income and high broadband penetration are highly correlated within a region. Indeed, the correlation is 0.65 for the indicator variables of high income and high Internet, and 0.74 for the raw variables of income level and Internet penetration. This paper tries to tease apart the extent to which the inequality in self-isolation that is attributed to disparities in income can actually be explained by disparities in Internet access. One immediate question is whether 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 illustrates that regions with high Internet penetration have reasonably broad coverage across the country and not are concentrated within any one region.

It is also worth discussing in the context of Figure 2 why broadband Internet might

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



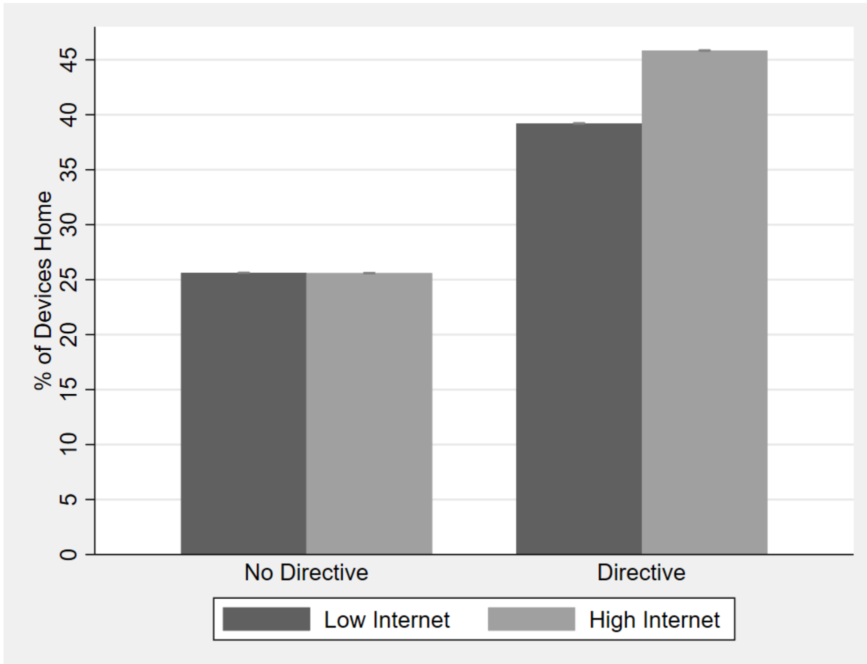
Notes: February 1-April 7 2020 data. High income is defined by whether that PUMA region has above-median household income.

play a large role in behavior even if people have access to 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, cellular plans typically have data limits that make it prohibitively expensive to use mobile phones for data-intensive uses, such as watching movies or conducting video calls. Second, most cellular plans have limits on whether their users can "tether" their cellphones—that is, use them as a general Internet modem for a home.[8] Compared to an Internet-powered desktop or laptop, mobile phones may also be a less than ideal substitute for many work or recreational purposes.

Figure 3 expands the analysis by allowing for whether or not a region had above- or below- median Internet penetration. The figure shows that after the directive, 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/`

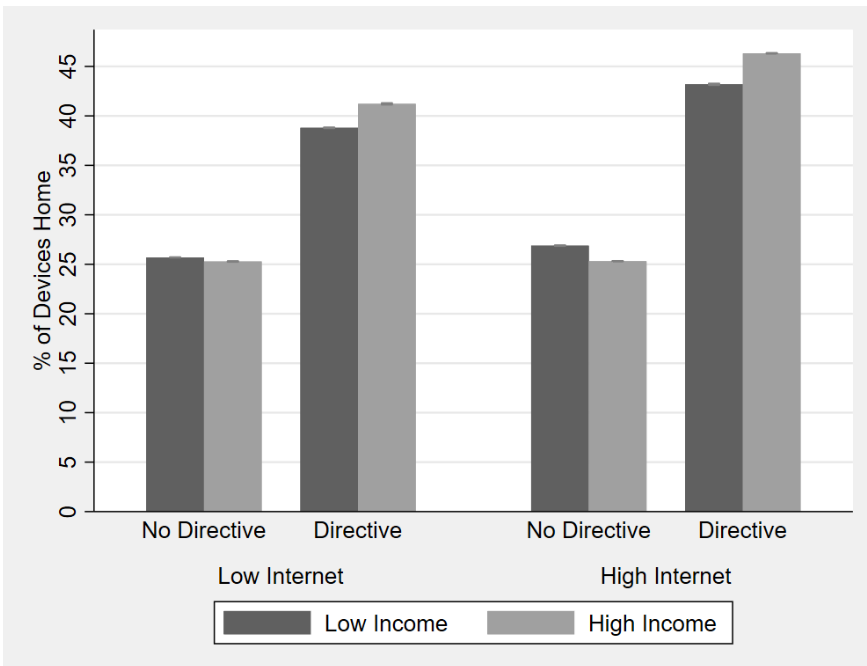
Figure 2: High-Internet Areas are More Likely to Stay at Home Following a State Directive



Notes: February 1-April 7 2020 data. "High Internet" is defined by whether that PUMA region has broadband penetration that is above the median.

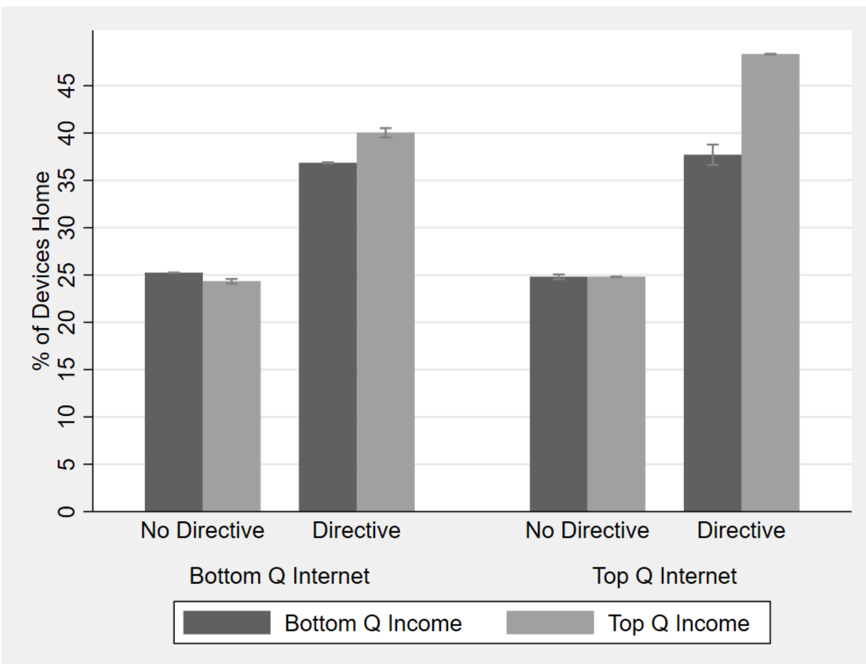
Figure 4 shows that the pattern is even more exaggerated for the top and bottom quartiles of Internet and income, compared to looking at the simple measure of above- and below- the median.

Figure 3: Internet Access Improves Everyone's Ability to Stay at Home



Notes: February 1-April 7 2020 data. "High Income" is defined by whether that PUMA region has household income that is above the median. "High Internet" is defined by whether that PUMA region has broadband penetration that is above the average.

Figure 4: Differences in Behavior among Income Groups for Bottom and Top Quartiles of Internet Penetration



Notes: February 1-April 7 2020 data. This figure repeats the analysis of Figure 3 but examines the top and bottom quartiles of high-speed broadband penetration and household income instead of using above and below median measures.

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***	-0.435***	-0.675***	0.437***	-0.127***	-0.00409
	(0.00765)	(0.0111)	(0.0136)	(0.0102)	(0.0147)	(0.0181)
Reported Cases	0.108***	0.00124***	0.00918	0.000937***	0.000917***	0.000766***
	(0.0166)	(0.0000170)	(0.0161)	(0.0000171)	(0.0000165)	(0.0000153)
Highspeed Internet		7.336***	-0.923***		12.93***	6.638***
		(0.166)	(0.218)		(0.212)	(0.264)
Proportion Black			4.918***			5.076***
			(0.122)			(0.153)
Proportion Asian			1.809***			15.37***
			(0.266)			(0.344)
Proportion Unemployed			23.13***			41.08***
			(4.145)			(4.986)
% 60+			3.944***			0.861
			(0.391)			(0.467)
Local Urban Population Share			0.548***			2.144***
			(0.0646)			(0.0703)
Proportion College Degrees			-0.389			0.998**
			(0.234)			(0.320)
Date Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
State Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2095224	4840472	2095224	2745248	2745248	2745248
R-Squared	0.369	0.609	0.379	0.618	0.626	0.640

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. * $p < 0.05$, ** $p < 0.05$, *** $p < 0.001$

While the results of Figures 1 and 2 are useful, they do not control for other shifts that occurred during the time period, and nor do they control for underlying differences in regions. To address this, we employ an econometric specification. Our initial regressions focus on whether demographic and economic characteristics are correlated with whether devices stay at home.

Table 2 shows how the decision to stay home varies with regional social, economic and demographic characteristics. In this specification, we use state fixed effects, so we only examine variation in income and Internet within a state rather than across states. In other words, the results do not reflect differences across states due to different timings 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. Column (1) shows that during February people were more likely to leave their house if they lived in a higher income region. Column (4) shows that this pattern reverses in March, and people in 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, people in high-income regions appear to be more likely to stay at home because 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. Overall, comparing February with March suggests that broadband penetration has a large positive and statistically significant effect on whether people leave their homes during the pandemic.

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

Other demographic variables included in Table 2 also offer interesting results. It appears that fewer devices leave the home in areas with a higher fraction of individuals who are African-American. 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. In general, areas with urban populations are more likely to stay at home. The average education level of the region, which we measure as the proportion of people who have college degrees, does not appear to be correlated with behavior.

One potentially troubling effect that we measure is the behavior of regions which have a larger proportion of people over the age of 60. This group is more likely to be vulnerable to coronavirus complications. Fortunately, we do not observe in March any negative and significant relationship with this measure of the proportion of elderly in a local region, which might cause alarm from a public health standpoint. In addition, across all specifications, the number of reported coronavirus cases in the local region has a significant effect at encouraging people to stay at home.

Though the results of Table 2 are useful from a policy perspective, they primarily provide correlations. 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 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 coefficients β_1 and β_2 reflect the focal policy effects studied in this paper because they capture the relationship between the interaction of a state directive with above median household income and Internet. The coefficient β_3 captures the level effect of the state directive, which 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 often people leave their homes. Note that Table 2 uses fixed

[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	4.448***	0.850***	0.713***	0.0404
	(0.0423)	(0.0468)	(0.0470)	(0.0476)
State Directive × High Income		6.546***		3.793***
		(0.0538)		(0.0689)
State Directive × High Internet			6.687***	4.159***
			(0.0537)	(0.0690)
Reported Cases	0.00122***	0.000974***	0.000947***	0.000907***
	(0.0000155)	(0.0000147)	(0.0000145)	(0.0000144)
Date Fixed Effects	Yes	Yes	Yes	Yes
Census Tract Fixed Effects	Yes	Yes	Yes	Yes
Observations	4840472	4840472	4840472	4840472
R-Squared	0.731	0.743	0.743	0.745

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 from Feb 1-April 7. * $p < 0.05$, ** $p < 0.05$, *** $p < 0.001$

effects at the state level in order to examine 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 current specification. For example, the main effect of household income drops out of our specification because of the presence of census-tract fixed effects. The coefficient δ_t is a vector of fixed effects for each date in the sample period.

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

Table 3 presents the results, exploring the effect of state directives for the combined sample from Feb 1- April 7. We report similar results for the non-February subsample in appendix Table A1. Column (1) shows the raw effect of the state directive whereby only essential businesses could stay open. It suggests an increase of 4.4 percentage points in devices remaining at home after the order. This is a small effect relative to an average of 26.9% of devices not leaving the house in Table 1. Perhaps this is unsurprising though, given the nationwide trend in people distancing themselves regardless of state directives.

Column (2) examines how the effect of the state ban was moderated by income. The results are striking. State directives appear to be far more effective at ensuring that people stayed at home in high-income regions. Columns (3) suggests a very similar pattern with a positive effect for the region having a high degree of broadband penetration. 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 examine the effect of including interactions for both income and Internet. In both cases, the effects of each variable are attenuated, but still positive and significant. The negligible coefficient on the baseline coefficient for the state directive, suggests that without the presence of either high-income or high Internet penetration, state directives did not exert a large effect on an individual's decision to stay at home.

Table A2 in the appendix extends these results by investigating whether other potential moderators may explain the effects in Table 3. In particular, it allows for the effect of the directive to vary by the average age in the district, whether that area is rural, and whether the area has a high fraction of individuals with a college education. All of these potential moderators have some effect, but the main effect of living in a high Internet and a high income area remains robust. Of these potential drivers, the largest effect can be seen for rural areas where devices were less likely to stay at home after a directive.

Given the strong correlation between high-income areas and the diffusion of high-speed Internet, a natural question is whether it is income or the diffusion of the Internet that drives

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	0.0353	2.268***	-0.0451	2.174***
	(0.0963)	(0.0515)	(0.102)	(0.0515)
State Directive × High Income	4.476***	3.002***		
	(0.0961)	(0.0980)		
State Directive × High Internet			4.875***	3.450***
			(0.101)	(0.0937)
Reported Cases	0.000790***	0.00105***	0.000801***	0.00107***
	(0.0000173)	(0.0000264)	(0.0000172)	(0.0000258)
Date Fixed Effects	Yes	Yes	Yes	Yes
Census Tract Fixed Effects	Yes	Yes	Yes	Yes
Observations	2418460	2422012	2404116	2436356
R-Squared	0.789	0.707	0.785	0.712

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 from Feb 1-April 7. * $p < 0.05$, ** $p < 0.05$, *** $p < 0.001$

our results. The correlations in Table 2 suggest collinearity with Internet was driving the apparent effect of high-income people staying at home.

We examine this more closely in Table 4. We exploit variation across regions and examine regions which have high Internet diffusion and low income, or low internet diffusion and high income. These regions are of course potentially non-representative. For example, areas of high Internet diffusion and low income regions disproportionately reside in states such as North Carolina. It appears that there are spillovers of provision of high-speed Internet from the "Triangle Area" to poorer areas nearby. Some high income and low Internet diffusion areas are found in states such as Wyoming where presumably topography and density interfere with the provision of Internet. However, we also observe other states such as Florida where lack of high-speed Internet does not appear to be driven by topography or density. Therefore, though it seems likely that unobserved characteristics of regions drive whether or not they deviate from the usual correlation between high income and high Internet diffusion, they do not seem entirely systematic. We also report a series of charts that maps out counties by whether they are above or below the median level of income or 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.

Columns (1) and (2) of Table 4 report a specification that splits our sample by whether a region is high- or low-income (above- or below-median income), or high- or low- internet (above or below median broadband penetration). These specifications provide similar insights. Columns (1) suggests that in high-Internet regions, a state-level directive only induces people to stay home if they live in the subset of high-Internet high-income regions. Column (2) suggests that in regions with low Internet, a state directive induces people with high and low incomes to stay at home. Column (3) suggests that in high-income regions, a directive encourages people to stay at home only when the region is high-internet. Column (4) suggests that in low-income regions, then the presence of high internet diffusion

Table 5: Staying at Home: Differences Between Weekend and Weekdays

	Weekends	Weekdays
	(1)	(2)
	% Stay Home	% Stay Home
State Directive	0.685***	-0.223***
	(0.0529)	(0.0499)
State Directive × High Internet	3.478***	4.447***
	(0.0710)	(0.0726)
State Directive × High Income	3.660***	3.849***
	(0.0707)	(0.0725)
Reported Cases	0.000802***	0.000955***
	(0.0000147)	(0.0000151)
Date Fixed Effects	Yes	Yes
Census Tract Fixed Effects	Yes	Yes
Observations	1444850	3395622
R-Squared	0.744	0.739

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 Feb 1-April 7 2020. * $p < 0.05$, ** $p < 0.05$, *** $p < 0.001$

encourages people to stay at home, and there is still a positive effect of the state directive.

Table A3 in the appendix, presents results for a version of Table 4 with just data for March onwards. Though the magnitudes of the coefficients are generally smaller, the relative size and direction are similar.

3.1 Suggestive Evidence about the Mechanism

Our results so far suggest that the combination of high income and living in a high Internet access area drives a household's ability to successfully self-isolate. Two potential reasons may explain why. One may be that the type of work pursued by people who live in high-income and high-Internet areas can be done at home. Another explanation is that people who live in high-income and high-Internet areas are able to avoid visiting retailers by using online delivery platforms, or potentially being able to stockpile supplies more successfully (Orhun and Palazzolo, 2019).

To investigate the two hypotheses, we first run a regression that splits our sample by weekends versus weekdays. We hypothesize that if our results are driven by the ability of individuals with higher incomes to work at home (facilitated by the Internet), then the effect of the Internet would be higher on weekdays—when people are conventionally working. Table 5 suggest that is indeed the case, the coefficient for the effect of the Internet on people staying home is greater on weekdays. By contrast, the effect of income on the ability to stay home is larger on weekends. This suggests that potentially both mechanisms are present—the combination of high Internet and high income enabled people to self-isolate on weekdays and not have to commute to work, while the combination of high Internet and high income also allowed people avoid leaving their homes to go shopping at weekends.

To further explore the first hypothesis, we use data from the Census's County Business Patterns to examine the types of employment in the local county. We then build on work by Dingel and Neiman (2020) that examines which categories of jobs can successfully be done remotely. We identify a variety of sectors of employment, where Dingel and Neiman (2020) calculated that over 70% of jobs could be done remotely. These sectors spanned professional management jobs, finance jobs, information-sector jobs, and IT jobs. We then calculate the proportion of these jobs within the local county. Figure 5 shows that indeed it is high-income and high-Internet areas that tend to have the highest proportion of jobs where it is possible to telecommute.

To further explore the second hypothesis of whether non-work related movement outside the home (in order to purchase goods) accounts for our results, we turn to a different dataset that Safegraph shares with researchers, which looks at the number of visits to stores. We calculate the number of visits to supermarkets and convenience stores divided by the number of devices that Safegraph tracks within that county. We focus on supermarkets and convenience stores because they represent essential businesses that are allowed to remain open

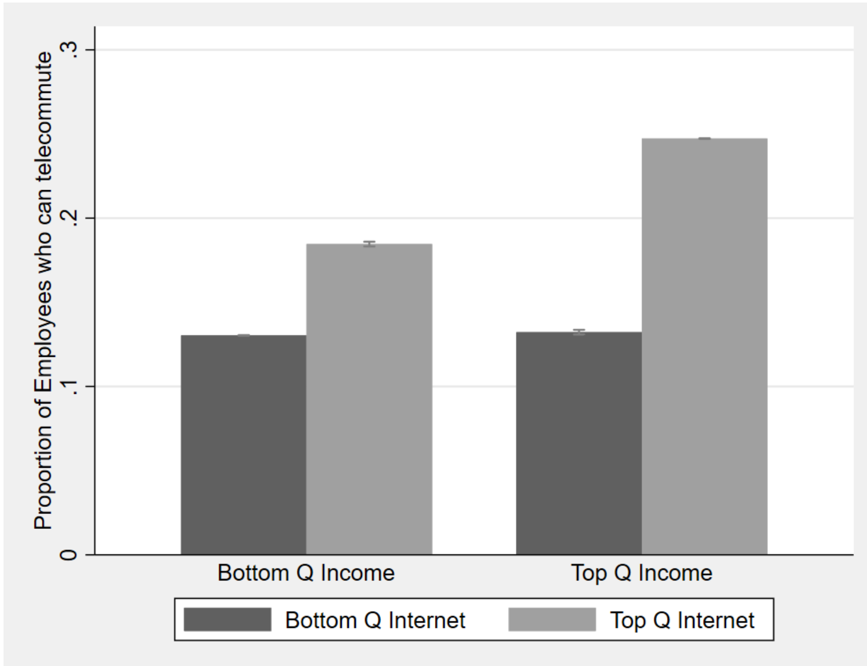


Figure 5: Proportion of Jobs Which are Relatively Easy to Conduct Remotely By Internet and Income

amid the directives that encourage people to stay at home.

Figure 6 illustrates the results for visits to convenience stores, and Figure 7 shows the results for supermarket visits. Figure 6 suggests that the presence of high Internet diffusion did lead to a relative reduction in visits to convenience stores, in particular for those with high incomes after the directive. That is, within each income bracket, those with high Internet had a larger proportional reduction in visits to convenience stores compared to those with low Internet.

By contrast Figure 7 shows a different pattern is present for supermarkets. Though average visits to supermarkets fell after the directive, there is little evidence that a larger relative change existed either for those living in a high income region or people living in a high Internet region. For instance, within each income bracket, it seems that those with high Internet did not relatively reduce visits to supermarkets compared to those with low Internet. Similarly, within each Internet bracket, there does not appear to be evidence of a

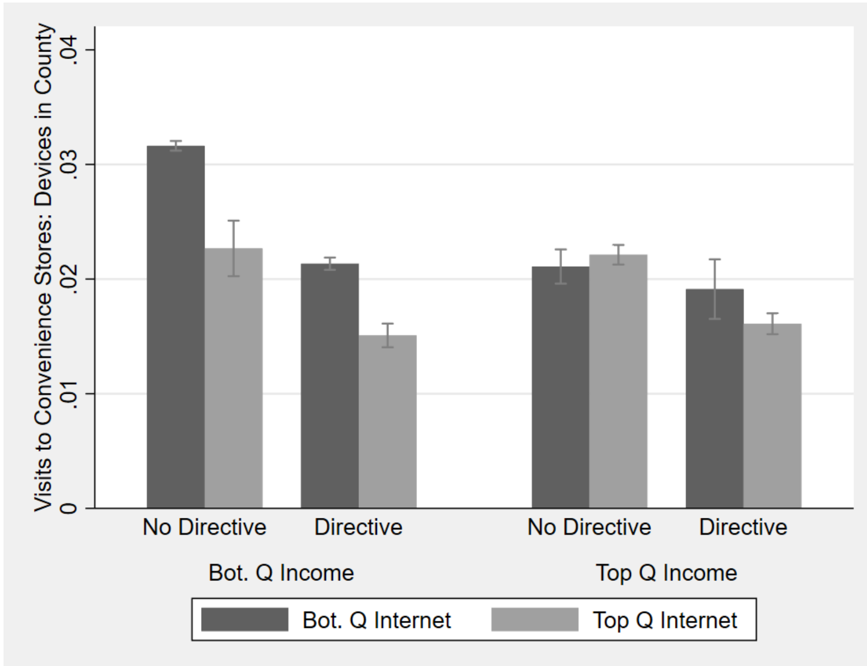


Figure 6: Trips to Convenience Stores After the Directive

dramatically larger proportional drop in visits by those with income compared to those with low income.

Overall, we observe that people who live in regions with high Internet, regardless of income, exhibit a reduction in trips to stores after the directive is in place, and that this reduction is greatest for trips to convenience stores among people who live in high-income regions. By contrast, within each income bracket, we do not see evidence of a strong differential effect in the reduction of trips to the supermarkets for those with high-speed Internet compared to those without.

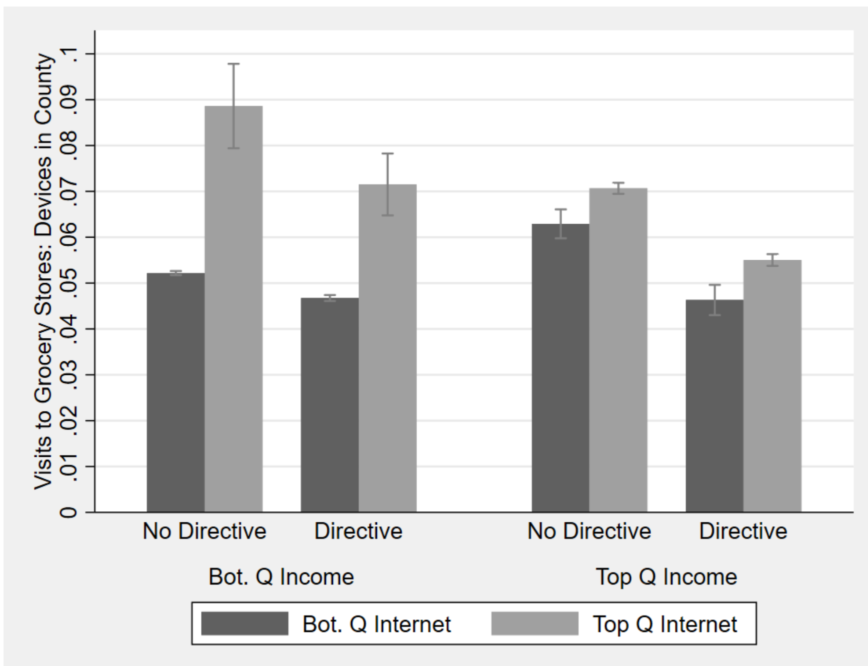


Figure 7: Trips to Supermarkets After the Directive

4 Conclusions

This paper is a first attempt at understanding the role of income inequality in moderating the effectiveness of social distancing measures in wake of the spread of coronavirus. Our results suggest that people who live in high-income areas are 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 self-isolated more and not left their home. 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.

We present some suggestive evidence about why this occurs. There appear to be two mechanisms. On weekdays, access to the Internet matters for high income areas because these areas have a high number of jobs that can be performed remotely with high-speed Internet. On weekends, income seems slightly more important than Internet access in its effect on staying home after a directive. We document that there does appear a reduction in trips to smaller retailers by those who have high incomes and live in a high-internet region after a directive.

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 be concerned about the effectivness of self-isolation policies in regions with low Internet penetration and low household 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).

Dingel, J. and B. Neiman (2020). How many jobs can be done at home? Technical report, National Bureau of Economic Research.

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.

Orhun, A. Y. and M. Palazzolo (2019). Frugality is hard to afford. *Journal of Marketing Research 56*(1), 1–17.

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.

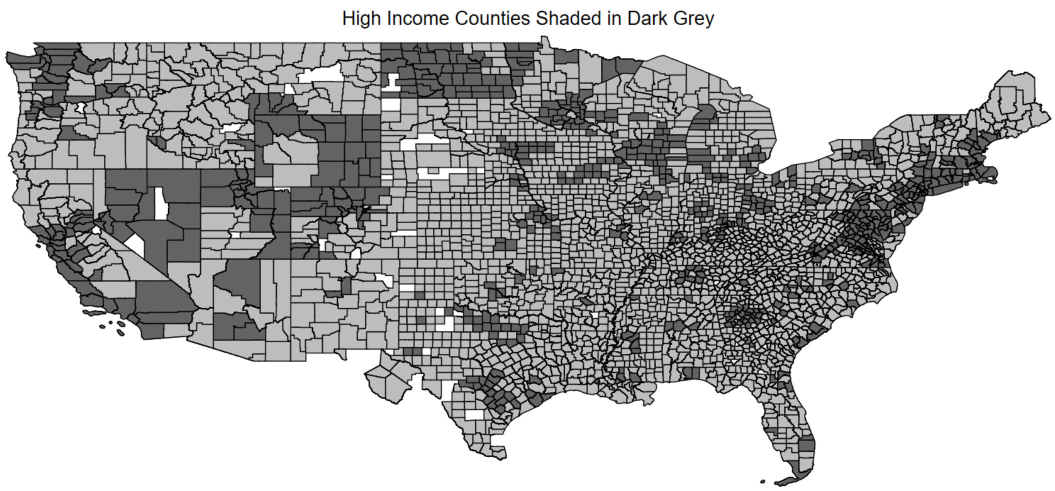


Figure A1: Distribution of High-Income Counties

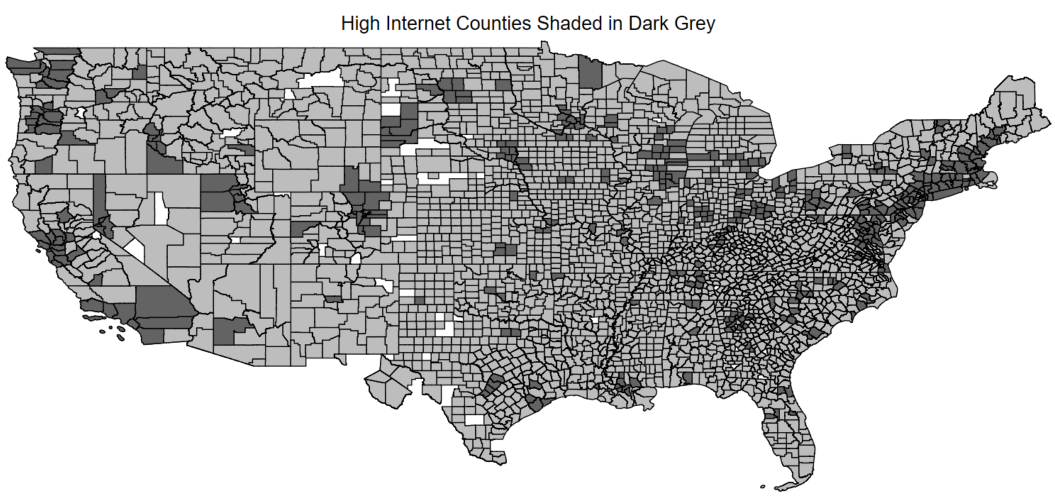


Figure A2: Distribution of High-Internet Counties

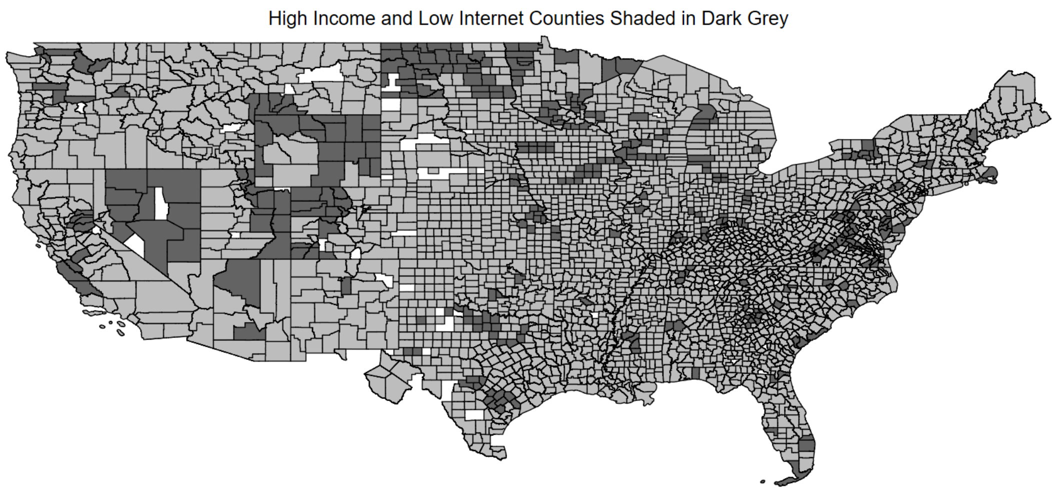


Figure A3: Distribution of High-Income and Low-Internet Counties

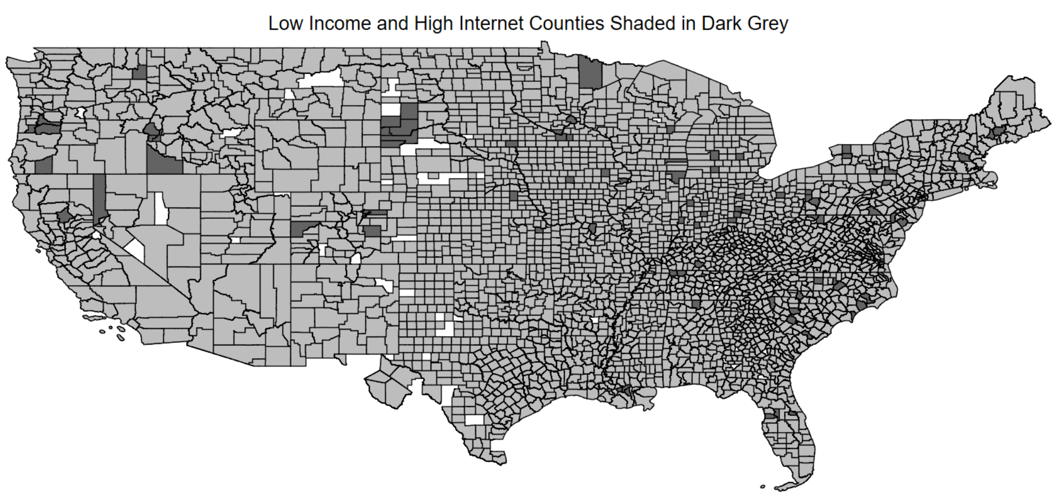


Figure A4: Distribution of Low-Income and High-Internet Counties

Table A1: Staying at Home: The Effect of State Directives (Excluding February)

	(1) % Stay Home	(2) % Stay Home	(3) % Stay Home	(4) % Stay Home
State Directive	3.543***	1.035***	0.986***	0.529***
	(0.0307)	(0.0355)	(0.0356)	(0.0458)
State Directive × High Income		4.705***		2.867***
		(0.0431)		(0.0719)
State Directive × High Internet			4.736***	3.002***
			(0.0431)	(0.0720)
Reported Cases	0.000833***	0.000667***	0.000651***	0.00146***
	(0.0000118)	(0.0000113)	(0.0000113)	(0.0000321)
Date Fixed Effects	Yes	Yes	Yes	Yes
Census Tract Fixed Effects	Yes	Yes	Yes	Yes
Observations	2745248	2745248	2745248	2239635
R-Squared	0.791	0.798	0.798	0.781

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 1-April 7 2020 only. * $p < 0.05$, ** $p < 0.05$, *** $p < 0.001$

Table A2: Staying at Home: The Effect of State Directives

	(1) % Stay Home	(2) % Stay Home	(3) % Stay Home	(4) % Stay Home
State Directive	0.926***	-0.121*	-0.417***	0.301***
	(0.0552)	(0.0577)	(0.0489)	(0.0607)
State Directive × High Income	3.709***	3.959***	3.075***	2.839***
	(0.0707)	(0.0713)	(0.0759)	(0.0758)
State Directive × High Internet	3.773***	4.493***	3.688***	3.193***
	(0.0729)	(0.0720)	(0.0729)	(0.0753)
Directive × Rural=1	-3.107***			-3.018***
	(0.0648)			(0.0683)
Directive × Above Median Age=1		0.193***		0.504***
		(0.0554)		(0.0567)
Directive × Above Median College=1			2.554***	2.230***
			(0.0662)	(0.0661)
Date Fixed Effects	Yes	Yes	Yes	Yes
Census Tract Fixed Effects	Yes	Yes	Yes	Yes
Observations	4840472	4840472	4840472	4840472
R-Squared	0.744	0.743	0.744	0.745

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. * $p < 0.05$, ** $p < 0.05$, *** $p < 0.001$

Table A3: Staying at Home: The Interaction Between Internet and Income (Excluding February)

	High Internet (1) % Stay Home	Low Internet (2) % Stay Home	High Income (3) % Stay Home	Low Income (4) % Stay Home
State Directive	0.595***	2.295***	0.649***	2.222***
	(0.0959)	(0.0486)	(0.103)	(0.0487)
State Directive × High Income	3.185***	2.508***		
	(0.0995)	(0.104)		
State Directive × High Internet			3.311***	2.714***
			(0.105)	(0.0984)
Reported Cases	0.00119***	0.00172***	0.00125***	0.00170***
	(0.0000373)	(0.0000562)	(0.0000371)	(0.0000561)
Date Fixed Effects	Yes	Yes	Yes	Yes
Census Tract Fixed Effects	Yes	Yes	Yes	Yes
Observations	1118999	1120636	1112353	1127282
R-Squared	0.810	0.755	0.810	0.759

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 1-April 7 2020 only. * $p < 0.05$, ** $p < 0.05$, *** $p < 0.001$