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 di usion 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

We combine the data from Safegraph with data from the 2018 American Community

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 in uence 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 re ects that ACS de nes 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 speci cations with cases per capita and also the log of cases, and found similar results.

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 de ned 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,

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

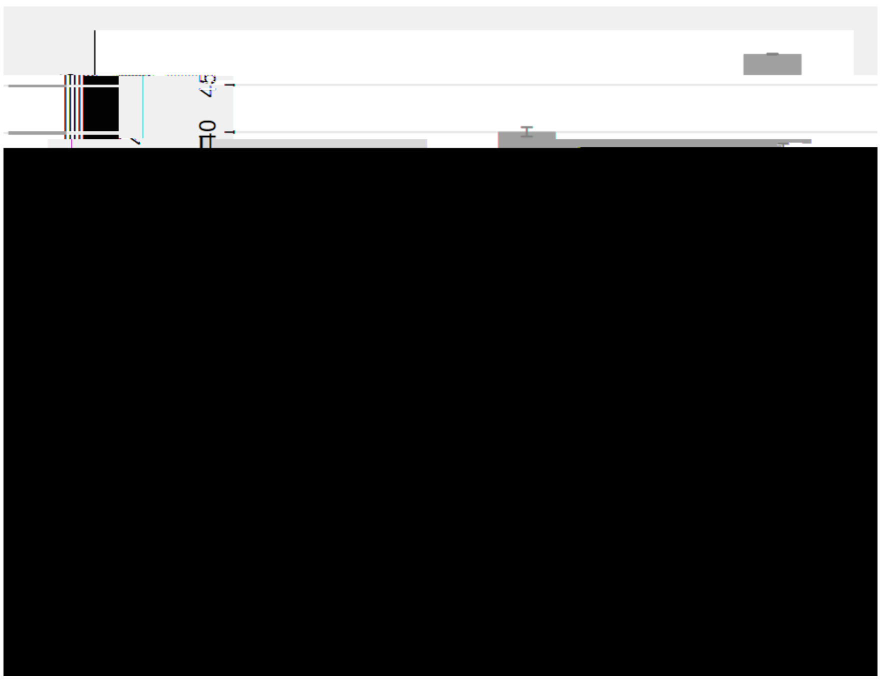
Notes: February 1-April 7 2020 data. \High Internet" is de ned by whether that PUMA region has broadband penetration that is above the median.

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 de ned by whether that PUMA region has household income that is above the median. \High Internet" is de ned by whether that PUMA region has broadband penetration that is above the average.

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



Notes: February 1-April 7 2020 data. This gure 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 E ects	Yes	Yes	Yes	Yes	Yes	Yes
State Fixed E ects	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 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 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

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 $_t$ 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

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

	Weekends (1) % Stay Home	Weekdays (2) % Stay Home
State Directive	0.685 (0.0529)	-0.223 (0.0499)
State Directive High Internet	3.478 (0.0710)	4.447 (0.0726)
State Directive High Income	3.660 (0.0707)	3.849 (0.0725)
Reported Cases	0.000802 (0.0000147)	0.000955 (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 rst run a regression that splits our sample by

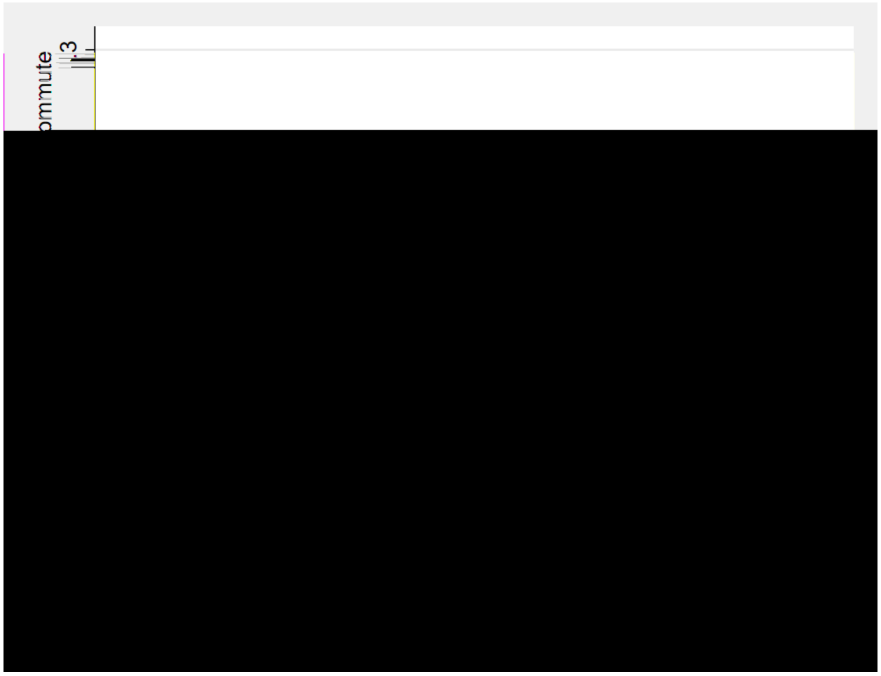


Figure 5: Proportion of Jobs Which are Relatively Easy to Conduct Remotely By Internet and Income

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 di usion 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.

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 xed e ects combined with the

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 a ord. *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.

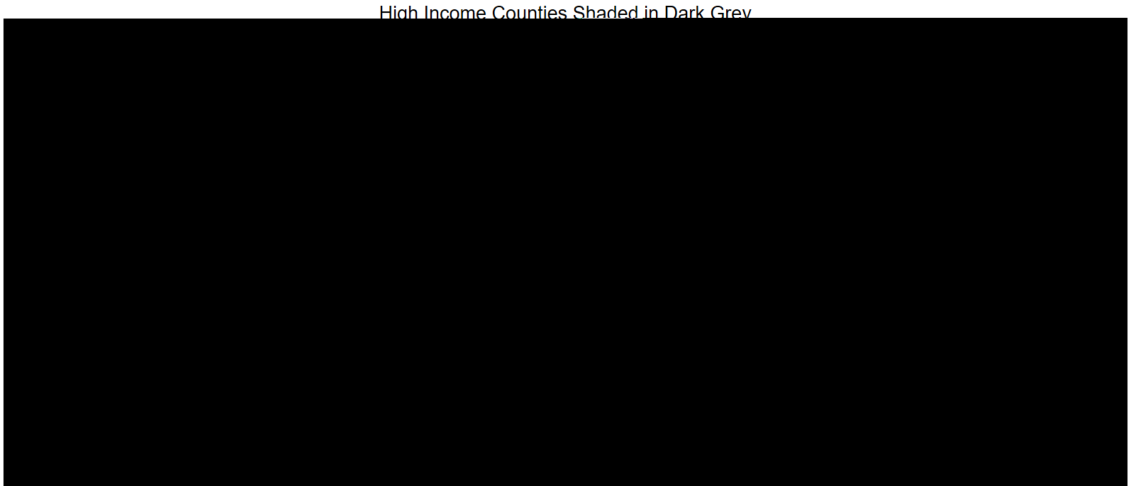


Figure A1: Distribution of High-Income Counties

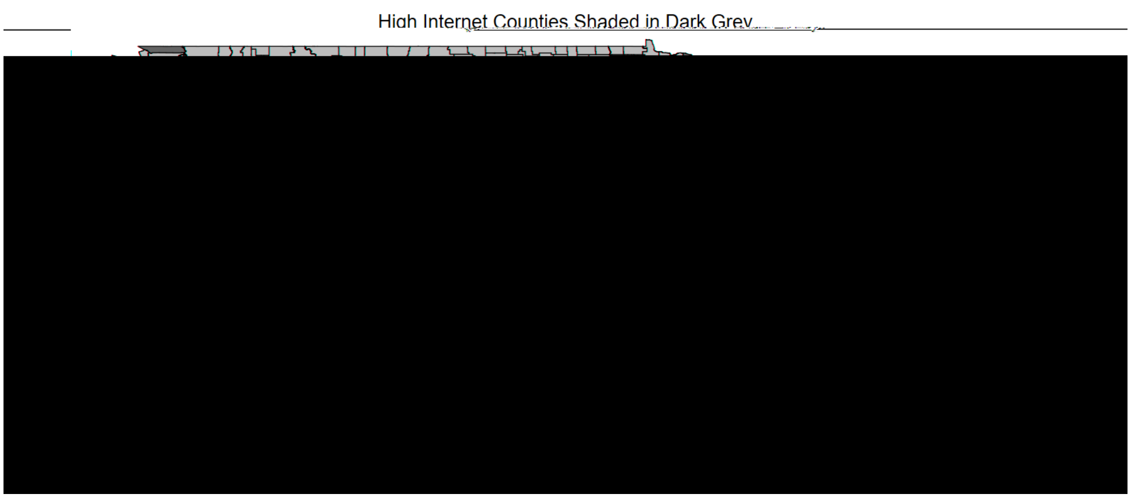


Figure A2: Distribution of High-Internet Counties

[redacted]

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

[redacted]

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

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

	(1)	(2)	(3)	(4)
	% Stay Home	% Stay Home	% Stay Home	% 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

Table A2: Staying at Home: The E ect 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 E ects	Yes	Yes	Yes	Yes
Census Tract Fixed E ects	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. * $p < 0{:}05$, ** $p < 0{:}05$, *** $p < 0{:}001$