digital society project

November 2019

Terrorism and Internet Censorship

Stephen A. Meserve
Daniel Pemstein

Terrorism and internet censorship*

Stephen A. Meserve† and Daniel Pemstein‡

Digital Society Project
Working Paper #3

Abstract

The internet provides a powerful tool to terror organizations, enhancing their public messaging, recruitment ability, and internal communication. In turn, governments have increasingly moved to disrupt terror organizations' internet communications, and even democracies now routinely work to censor terrorist propaganda, and related political messaging, in the name of national security. We argue that *democratic* states respond to terror attacks by increasing internet censorship and broadening their capacity to limit the digital dissemination of information. This article builds on previous work suggesting this relationship, substantially improving measurement and estimation strategy. We use latent variable modeling techniques to create a new measure of internet censorship, cross nationally and over time, from internet firm transparency reports, and compare this measure to an expert-survey based indicator. Leveraging both measures, we use a variety of panel specifications to establish that, in democracies, increases in terror predict surges in digital censorship. Finally, we examine the posited relationship using synthetic control models in a liberal democracy that experienced a large shock in terror deaths, France, showing that digital censorship ramped up after several large terrorist attacks.

Keywords: Internet censorship, terrorism, latent variable models

*This research was supported, in part, by a Google Faculty Research Award.

†Assistant Professor, Department of Politics and International Affairs, Northern Arizona University. Email: stephen.meserve@nau.edu

‡Associate Professor, Department of Criminal Justice and Political Science, North Dakota State University. Email: daniel.pemstein@ndsu.edu (corresponding author)

1 Introduction

Since a core interest of the state is to protect itself, its agents, and its rule (Tilly 1990), theories focused narrowly on the states' survival interests suggest that governments might meet activities that threaten them with repression. Over time, a large literature has identified the ways in which digital communications technology facilitates groups that oppose the state, by enabling them to recruit, organize, and coordinate action (Tufekci 2017).[1] In response, governments developed and refined tools to restrict communications in the digital realm (Tufekci 2017, Roberts 2018), in order to—among other things—protect against terrorism and insurgency.

But how do findings about tools for digital control and repression explain behavior by democratic regimes? Most definitions of democracy rest on a bedrock of freedom of expression, and all measures of democracy demand that governments at least allow opposition to freely organize, compete and dissent politically. Democracies may, therefore, abhor digital censorship. The literature provides suggestive but limited answers about democratic digital control (Deibert, Palfrey, Rohozinski & Zittrain 2008, Deibert, Palfrey, Rohozinski & Zittrain 2010, MacKinnon 2012), but most of what we systematically know about digital content restriction rests on studies of autocracies that are relatively unconstrained with respect to freedom of expression. Do democratic states control digital spaces, or do they maintain free and open communication, in response to threats? To the extent that they do restrict online speech, how do they censor?

We provide evidence that democracies do respond to internal threats with repressive and controlling behavior. But they use a different suite of tools than authoritarian regimes. We focus in particular on their use of intermediaries to remove content: treating online content providers as "points of control" (Zittrain 2003). We hypothesize that the imperatives of state survival apply to democracies under threat, despite countervailing pressures to protect

[1] These expectations in the digital realm parallel recent theory, measurement, and empirical research on traditional press and media freedom and civil conflict (Whitten-Woodring & Van Belle 2017).

broad citizen freedoms. Democracies tend to work within legal frameworks to remove digital content. But our empirical picture of what this process looks like is far from complete, and understanding democratic digital censorship is fundamental to timely policy debates about who—governments or firms—should be responsible for policing internet content.

While previous work has explored the association between terrorism and digital content controls, the measurement strategies and findings were relatively limited, exploring a brief time frame and using a single measure of digital restriction, based on the products of a single company, Google LLC (Meserve & Pemstein 2018). Here we delve into the topic of digital censorship and security threats in greater empirical depth and specificity, refining the measurement of internet restrictions, and demonstrating the robustness of the relationship between digital censorship and terrorism to new data and methods, with a greater attention to causation. We apply panel data and synthetic control techniques to newly developed measures of internet freedom and state digital censorship. Specifically, we create a new measure of internet censorship, using a latent variable model to scale internet transparency reports from major multi-national firms. We also perform tests using new, expert-rated, measures of internet freedom, in order to confirm that our firm-based measures produce results that are consistent with a broader, but more subjective, assessment of content regulation.

We find consistent, compelling evidence that violent opposition induces the state to censor digital content and reduce internet freedoms. Of substantive note, our data show that restricting digital content and restricting internet freedom in response to terrorism is not simply a behavior performed by illiberal regimes like Turkey (Meserve & Pemstein 2018, Gohdes 2018). We find that even liberal democracies tend to respond to terrorism and insurgency by tightening content restrictions through the use of legal mechanisms that force a variety of online content providers (OCPs) to censor content on their behalf, rather than relying on the direct infrastructure control and filtering techniques pioneered by countries like China. To this point, after presenting our panel data evidence, we examine the case of France, using synthetic control methods to show how sensitive digital freedoms are to security threats to the state. After several deadly terrorist events, France greatly tightened its legal rules on digital content, going so far as to authorize a state of emergency that resulted in a tremendous number of content restrictions and overall reduction in internet freedom. We show that internet freedom in liberal democracies is sensitive to internal threats, and that democratic governments, like their autocratic counterparts, restrict digital freedom when faced with terrorism and insurgency.

Controlling communications on the internet

Opposition groups, broadly, are thought to have seen their capacity to organize and act collectively strengthened by the spread of digital technology. Digitally networked movements can use a wide variety of tools unavailable to previous generations of social movements and opposition groups. Tufekci (2017) outlines the ways in which networked protest and opposition groups use digital tools to help found, organize, and coordinate protest movements ranging from the Arab Spring to Occupy Wall Street. Participation by peripheral, less committed individuals, is critical to the success of collective action, and digital tools provide networked movements with the ability to reach and persuade critical, fringe, individuals to join and act (Barberá, Wang, Bonneau, Jost, Nagler, Tucker & González-Bailón 2015, Steinert-Threlkeld 2017). While an oversimplification, digital technology, initially, made it generally easier to reach and network diverse people with similar grievances. Moreover, digital networks also make it easier to cross boundaries and evade governments by acting trans-nationally, pushing issues into and out of different national jurisdictions and facilitating the provision of material support to those in other political environments or regime types (Keck & Sikkink 1998).

While some of the literature focuses on democratic opposition groups and protests like the Arab Spring, or the suppression of Turkish opposition, the same tools that facilitate networked protest are also used by violent opposition groups to assist their activities. Digitally empowered domestic and international terrorists, for example, recruit and plan across borders, away from prying government eyes. At the subnational level, access to cell phone networks and social media are associated with more insurgent violence (Warren 2015), but digital technologies also may facilitate loyal groups collaborating with government forces (Shapiro & Siegel 2015). The interaction between access to digital tools, and government restrictions of internet content even modulates how governments target repression against regime opponents, and how much violence insurgents perpetrate (Gohdes 2015, Bak, Sriyai & Meserve 2018, Gohdes forthcoming).

In practice, digital content scholars suggest that states facing internal pressures and direct violent threats reasserted their power over citizen digital communications using what DeNardis (2014, 199) calls 'the dark arts of internet governance.' Broadly, states played technological and infrastructural catch-up to master systems of digital control in order to combat opposition movements in the early years of the 21st century. While 'the costs to governments of fear-based censorship are more severe in the information age' (Roberts 2018, 54), after initial missteps, state authorities have pioneered new digital content control tools and have used those tools to minimize the damage of networked violent opposition.

Crucially, the bulk of the aforementioned literature focuses on authoritarian regimes, where we would expect an unconstrained state response. Authors have directly explored how authoritarian regimes leverage digital tools like social media to stabilize the regime (Gunitsky 2015), using it as a form of repression technology (Rød & Weidmann 2015). The most well-studied example, China, constructed a vast censorship apparatus, which features all manner of coercive control—from human censors, relatively porous blocking of the ability to see outside country content, and the production of a flood of misinformation that makes finding the truth difficult online for Chinese citizens (King, Pan & Roberts 2013, Roberts 2018). We know, fairly comprehensively, that authoritarian regimes do their best to control digital spaces in response to regime threats.

But democracies are, arguably, substantially more constrained in their ability to control digital spaces. Do the implications of the large literature on digital censorship in autocracies carry over to the democratic space? We provide systematic evidence that democracies do their own kind of filtering, often pressuring the multinational firms that control the majority of potentially dangerous content to censor for them (Deibert et al. 2008, Deibert et al. 2010, MacKinnon 2012). Indeed, "while billions of people use the internet, a small number of services capture or shape most of their activities"—including protest, mobilization, and organization (Tufekci 2017, 135). This leads democracies to engage in 'delegation of censorship' to OCPs to control content that endangers the state (Seltzer 2008). Additionally, firms have limited resources to contest state pressure, often have little financial incentive to fight individual requests to take down content, and, because the process of censorship is off-loaded onto firms, censorship through private points of control exhibits less oversight than 'old-fashioned' censorship (Adler 2011). OCP-based restrictions are therefore potentially attractive to democracies, as censorship can be codified in legal systems, can be offloaded financially to firms, and, especially in less liberal democracies, can be manipulated to effect political censorship that would not stand up to strict legal scrutiny (Adler 2011, Marsden 2011, Meserve & Pemstein 2018). We provide robust, systematic evidence that democracies respond to violent opposition by censoring digital content, and do so specifically through private points of control.

2 Data and Methods

We test the above argument using worldwide bi-annual data from the 2009-2017 period. We chose this period for two reasons. First, takedown data become available in the second half of 2009, making analysis of takedowns impossible before 2009. Second, while other (e.g., V-Dem's) measures of internet censorship stretch further back in the past, widespread internet penetration, and especially social media use, is spotty during the first decade of the 2000s. Thus, while we might arguably have pushed back our analysis of this measure to 2008, or, optimistically, 2006, we decided to use the availability of transparency data as our starting point. Because we are focused on democracies, we conduct our core analysis on countries classified as democracies by V-Dem's Regimes of the World (RoW) measure (Lührmann, Tannenberg & Lindberg 2018), although we include non-democracies in some descriptive analyses of the data, presented in this section. In this section we describe our measures of internet censorship effort, terrorism, and internal instability, and a number of control variables. We also describe our estimation strategy.

Measuring internet censorship effort

Takedown requests

Increasingly, OCPs have sought to increase perceptions of transparency, by releasing (semi)-annual takedown request reports, detailing the extent to which firms fielded requests to remove content from their platforms that governments allege contravenes local law. All requests in this analysis emanate from government executives and judiciaries, including local, regional, and national authorities. Requests are generated from a variety of executive and judicial processes such as legal rulings, military and police requests, or bureaucratic actions. We rely on data from four large multinational content-providing firms: Facebook, Google, Microsoft, and Twitter.[2] Firms vary in their tendency to comply with requests, although all four claim to evaluate requests with respect to local law. Because previous work relying on Google transparency reports alone may reflect the peculiarities of Google's products and global reach, we use latent variable modeling techniques to combine data from all four firms, extending coverage and focusing attention on patterns that are common to all four firms.[3]

Request data Facebook has published biannual transparency reports since the first half of 2013, and data on takedown requests since the second half of that year (Facebook

[2]We choose these firms for a number of reasons. First, all these firms distribute products used widely internationally, critical for the purposes of cross-national analysis. Second, on a practical level, the firms chosen provide the most consistent data in their transparency reports. Third, these specific firms, with the exception of perhaps Twitter, allow us to incorporate a broad bundle of digital products across the areas of search, social media, business, etc. For more details, see our discussion in section A.

[3]We discuss takedown requests in more conceptual detail in appendix section B.

Incorporated 2018). These requests cover material that governments flag as violating local law. Facebook reports each request, rather than each piece of content, as a single data point. Google provides biannual transparency reports, starting in the second half of 2009 (Google LLC 2018). These reports cover formal requests from governments to remove content, based on local law. Each request may reference one or more pieces of content, but repeated requests to remove the same piece of content count as multiple requests. The data do not include removals that Google performs without prompting, such as the removal of child pornography, and excludes requests to remove intellectual property that are not the result of a court case—that is, Google has another system designed to field and arbitrate such requests directly from firms. Microsoft has released biannual content removal request reports since 2015 (Microsoft Corporation 2018). Like Facebook and Google, data reflect requests, rather that content items flagged. Microsoft's reports focus specifically on requests initiated by governments. Finally, Twitter provides biannual data on government-initiated requests, starting in 2012 (Twitter Incorporated 2018). As a consequence, the takedown data are biannual country-level observations of the number of takedown requests for each firm. In the next section, we scale together firm observations to arrive at a latent measure of biannual country content removal effort.

Latent content removal effort Given the difference in product portfolios and global market share across firms, simply summing up content removal requests across reports is likely to lead to erroneous conclusions. In other words, 5 Google takedown requests do not equate to 5 Facebook requests. Reliable cross-national and temporal market share data are also difficult to obtain, limiting our ability to weight contributions by market share. Nonetheless, to average over firm idiosyncrasies, and to leverage all the information available to us, it makes sense to create a composite measure from all four reports. To achieve this goal we treat requests as observable manifestations of underlying effort expended by governments to remove content from the internet, and use Bayesian factor analysis to estimate this latent

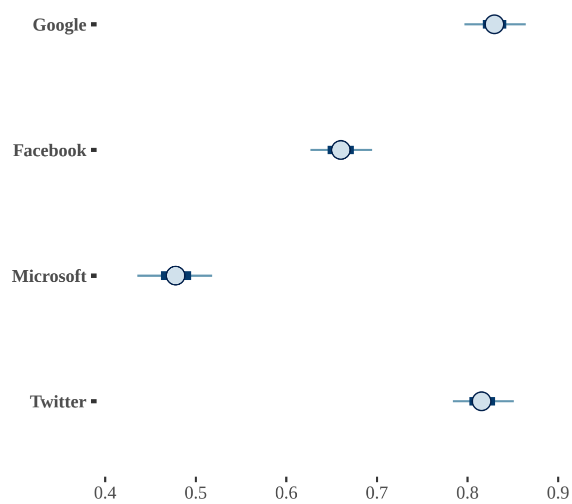


Figure 1: Takedown request factor loadings

variable from the firms' transparency reports.

We use a simple one-dimensional model.[4] Section D, in the appendix, provides information about model specification, estimation, and diagnostics. Figure 1 displays factor loadings, with 95 (thin line) and 50 (thick line) per cent credible intervals and around posterior means. As expected, all four series of takedown requests load positively on the latent trait. Notably, Google and Twitter takedown requests load most highly on the trait, with factor loadings above 0.8, Microsoft requests are more moderately associated with the latent trait with a loading just below 0.5, while Facebook requests fall somewhere in between, approaching 0.7. With the possible exception of the Microsoft loading, factor analysis folk wisdom would classify all of these loadings as 'strong,' providing evidence that Takedown requests reflect a consistent latent process across firms. Microsoft's somewhat weaker loading may reflect its relatively smaller market share in the social media and content provision spaces, and its limited time converage.

Figure 2 provides time-series plots for six countries, displaying logged takedown request counts and the latent score across the observation period. We can see that the latent scores

[4]Exploratory 2-dimensional analyses provided little evidence for two latent dimensions.

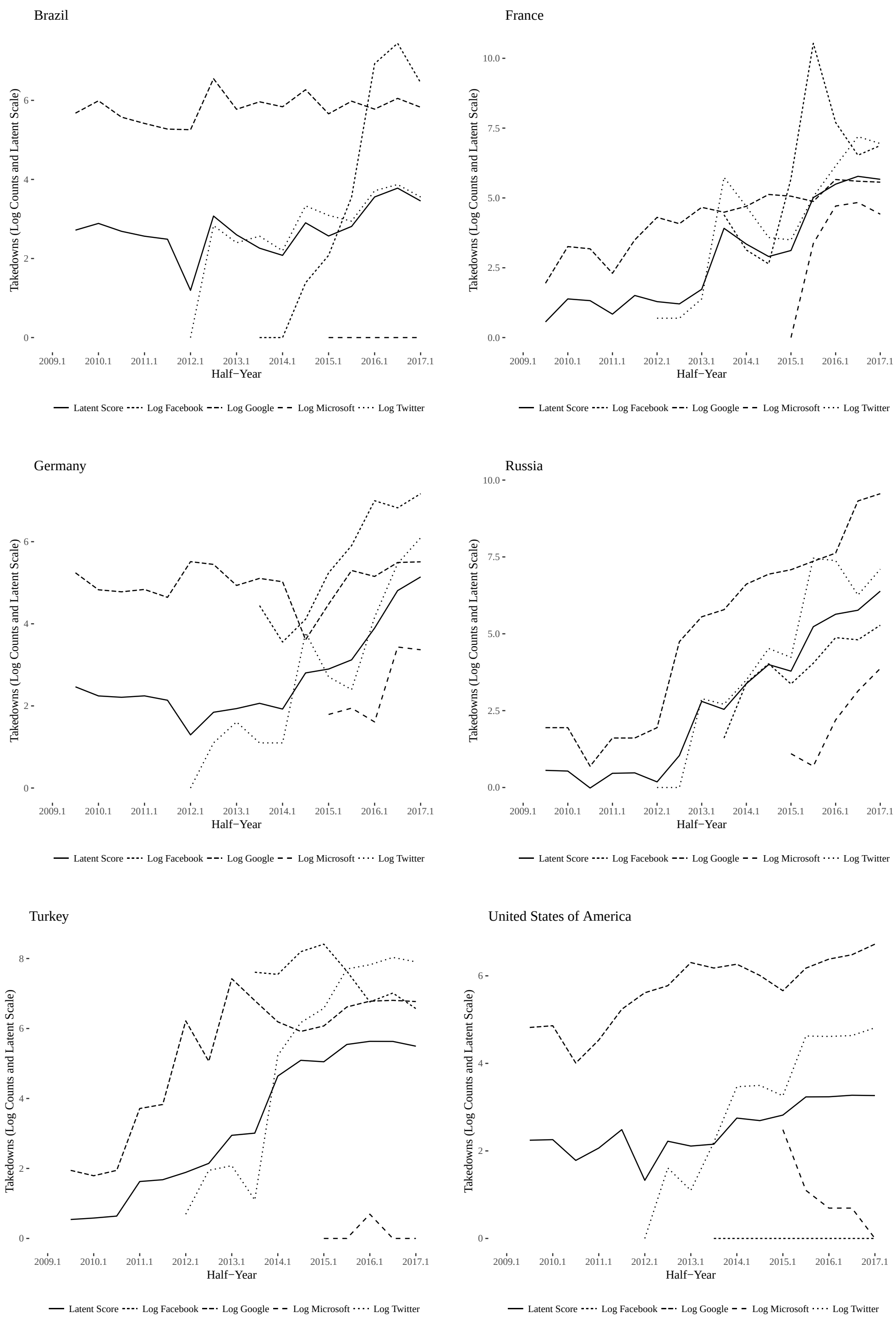


Figure 2: Time-series plots of logged takedown requests, latent takedown scores, and V-Dem expert scores for six countries

track broad trends in the request counts, while smoothing over volatility in the individual scores. They also allow one to take advantage of the long-standing Google data, while incorporating information from the other providers as they become available. The figure also highlights the dearth of Microsoft data, which only becomes available in 2015. This short time series may also help to explain the lower loading shown in figure 1.

Internet censorship effort

While takedown requests provide observable information about government efforts at internet censorship, they are an imperfect measure of political censorship, both because they capture significant non-political censorship, since only a handful of firms provide takedown request reports, and because takedowns capture only one mechanism through which governments censor digital content. We therefore make use of an alternative measure of this latent concept, based on an expert survey fielded by the Varieties of Democracy project (V-Dem). We use a question about government censorship effort (Coppedge, Gerring, Knutsen, Lindberg, Skaaning, Teorell, Altman, Bernhard, Cornell, Fish, Gjerløw, Glynn, Hicken, Krusell, Luhrmann, Marquardt, McMann, Mechkova, Moa, Paxton, Pemstein, Seim, Sigman, Staton, Sundstrom, Tzelgov, Uberti, Wig & Ziblatt 2018), that asked roughly five experts (per observation) to rate country-years on the effort and success government officials have in blocking internet content. For the exact survey question wording, clarification information, and measurement decisions, see appendix section C.

Comparing latent internet censorship measures

These two approaches to measuring internet censorship—constructing a latent measure from reported takedown requests, and leveraging the subjective ratings—are potentially complementary. They tap different strategies for measuring the censoring of digital content. One is based on objective counts of reported events, while the other leverages the subjective evaluations of topic and country experts. The question of which of these measures is more valid

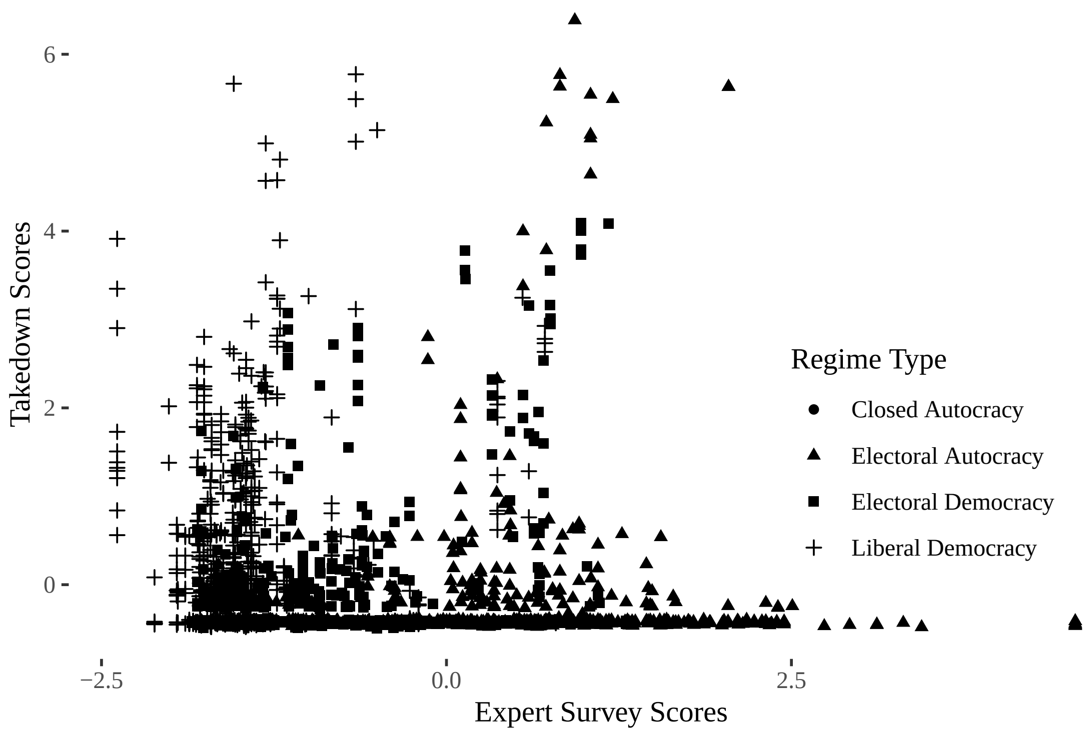


Figure 3: Expert scores and takedown requests

is, however, debatable, so we include both in our analysis. The takedown approach measures cross-nationally comparable behavior, reported by a third party without any ulterior interest in obscuring censorship practices. From this perspective, we believe that the takedown based measure represents a valid, sensitive, behavioral measure of the intensity of censorship. It also focuses on censorship through private points of control, which is the core quantity of interest for our analysis. A potential worry, however, is that some countries do not use private points of control, or use other censorship strategies more intensively. The V-Dem measure captures the subjective perception of experts about censorship within countries. This has the distinct drawback of not being based on actual country behavior. On the other hand, these measures allow us to test whether the conclusions drawn from our behavioral analyses have wide generality, since the subjective assessments by experts will, in principle capture both the behavior that we focus on here and other forms of censorship.

Interestingly, in practice, the two measures are largely uncorrelated ($r = -0.08$). Figure 3 is a scatter-plot of the two scores, broken down by regime type. Clearly, not all countries take advantage of takedown requests. In particular, closed autocracies do not bother with takedowns, likely relying on more forceful measures, and are absent from our data. Electoral autocracies exhibit a non-linear relationship. The raw correlation between measures in this subset is 0.05, but takedown effort can be quite intense among electoral autocracies with expert scores in the zero to one range, while it is rare in electoral autocracies with especially high or low expert censorship scores. If we lump democracies together, we find a correlation of 0.06, but correlations of 0.27 and 0.15 emerge in electoral, and liberal, democracies, respectively.

Figure 4 plots coefficients from three models in which we regress expert scores on takedown scores, V-Dem's polyarchy measure (Teorell, Coppedge, Lindberg & Skaaning 2019)][5] and its square, and country fixed effects, for the entire dataset, autocracies, and democracies, as classified by RoW. In each case we find small, but positive, and highly statistically signif-

[5]This is a continuous measure of electoral democracy, ranging between 0 and 1. The ordinal RoW measure that we use to divide the dataset is based on this measure, under the hood.

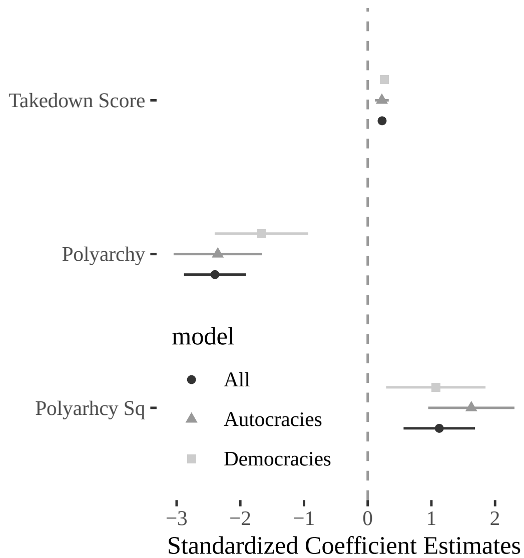


Figure 4: Predicting expert scores with takedown scores

icant ($t > 10$) relationships between our two measures of internet censorship effort. While our two measures capture distinct aspects of the internet censorship, we find that, across the democracy range, takedown request effort predicts expert assessments of censorship effort. The V-Dem measure likely captures a broader range of censorship activities, but takedowns are a good predictor of *digital* censorship scores once we control for the broad package of civil liberties baked into the polyarchy measure.

Independent Variables

Key predictors

We measure our key independent variable, terrorism, as the logged total number of terrorist events—and alternatively, as a robustness check, the logged total of deaths caused by terrorist events—in each half year, reported in the Global Terrorism Database (National Consortium for the Study of Terrorism and Reponses to Terrorism (START) 2018). We also include an alternative indicator, the World Governance Indicator's (WGI) 'political stability and absence of violence' index (Kaufmann, Kraay & Mastruzzi 2013). While our primary focus is on the relationship between terrorism and internet censorship, we include this predictor in alternative specifications to test the more general relationship between internal unrest and digital repression.

Covariates

We include a number of covariates to adjust for potential omitted variable bias. First, countries that produce substantial intellectual property (IP) are likely to police content more aggressively than their counterparts and may also face more terrorism. We use population, drawn from the World Development Indicators (WDI), via the Quality of Governance (QoG) dataset (The World Bank 2018, Teorell, Dahlberg, Holmberg, Rothstein, Pachon & Svensson 2018), to create a patents per capita measure. Similarly, we use the World Intellectual Property Organization's IP filing database (World Intellectual Property Organization 2018) to measure the number of patent applications originating from each state in the dataset across the time period. Economic development predicts internet use, and therefore, likely digital censorship, and potentially terrorism. We therefore include GDP, again from the WDI, via QoG. Because more efficient states are more likely to be able to effectively leverage private points of control to censor, we include a measure of bureaucratic efficiency—the number of days it takes to start a business, again from the WDI—in our specifications. We use the WDI's measure of the percentage of citizens who regularly use the internet to capture the importance of digital platforms and use V-Dem's polyarchy variable to control for level of democracy, both of which may causally relate to both terrorism and censorship. Section E, in the appendix, contains summary statistics of our data, for both the full sample, and for the sample of democracies.

Estimation

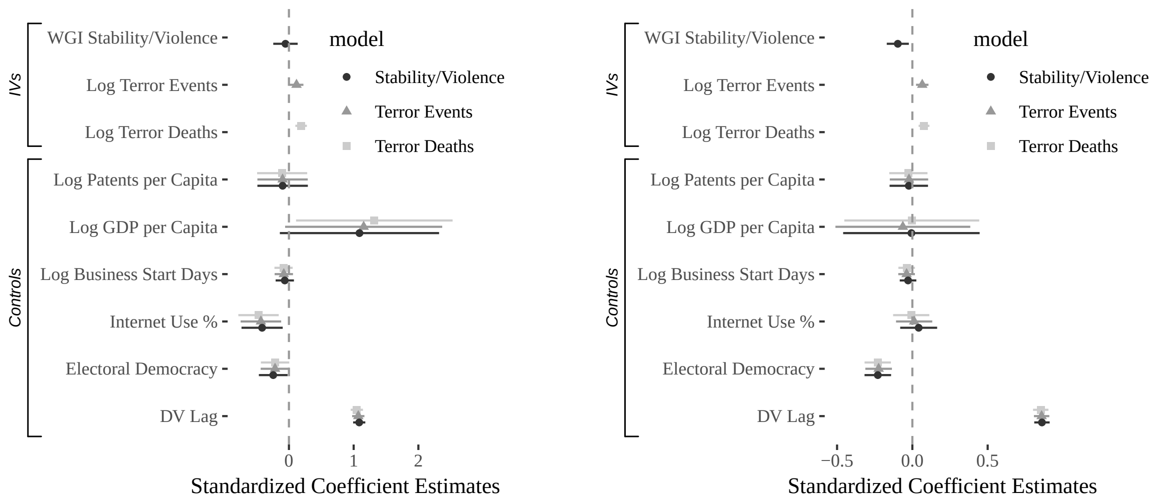
We examine the relationship between terrorism—and internal unrest—and internet censorship in two ways. First, we use panel data techniques to examine the extent to which terrorism and internal unrest predict changes in both takedown requests and expert assessments of governments' internet censorship efforts. Specifically, we use two-way fixed effects models, controlling for country and year. We include the above-described battery of covariates in these models. In particular, trends in internet penetration might plausibly covary with both terrorism and takedown requests, violating the parallel trends assumption inherent in fixed effects regression. Finally, we include lagged dependent variables to help account for endogeneity.[6]

Second, sacrificing generality, but addressing the parallel trends assumption in the fixed effects models, we present a short case study of terrorism and internet censorship in France, and use synthetic control techniques to demonstrate that France significantly increased its internet censorship efforts after experiencing a wave of large terrorist incidents. We chose France because it experienced the largest terrorist incident—measured by deaths—of any liberal democracy—measured using V-Dem's RoW indicator—in our sample period. We apply synthetic control techniques (Abadie, Diamond & Hainmueller 2010) to both of our dependent variables. We include all of the above-mentioned independent variables, and lags of our dependent variables, as predictors when generating synthetic matches.

3 Results

Figure 5 displays coefficient estimates from a two sets of two-way fixed effects regression models of internet censorship effort—measured with (a) takedown scores and (b) expert scores, respectively—on three measures of terrorism or internal unrest, a battery of controls, and a lagged dependent variable. We replicate Meserve & Pemstein's (2018) initial finding

[6] Putting a lagged DV in a fixed effects regression can induce bias. The appendix reports results of separate fixed effects and lagged dependent variables models, in tables 6 and 7, plausibly placing bounds on effect sizes (Angrist & Pischke 2009). The direction and statistical significance of our key coefficients are largely robust to specification. The one exception is that the relationship between WGI S&V and takedown scores is statistically insignificant in the combined and fixed effects models, but statistically significant in the ldv-only model, while the relationship between WGI S&V and the latent score is statistically significant in the combined and fixed effects models, but statistically insignificant in the ldv-only model. The WGI S&V coefficient is negative in all models.



(a) Takedown request scores

(b) Expert survey scores

Figure 5: Two-way fixed effect regression coefficients

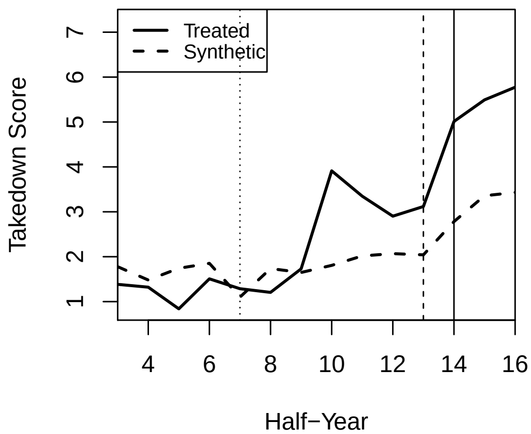
that democracies that experience terror censor the internet more aggressively. While the previous research relied on a short panel of Google takedown requests, we find this effect across both multi-firm takedown scores and expert-based measures. While the standardized coefficients are reasonably small, the effects are, nonetheless, substantively significant. For example, a 30-terror-death half year is associated with a quarter of a standard deviation increase in takedown effort, and about a fifth of a standard deviation increase in expert-rated censorship effort.

The effect holds whether we measure terrorism in terms of event counts or deaths. We also use more robust estimation techniques. The original finding was based on a random-effects regression with no lagged dependent variable. Here, leveraging our longer panel, we include fixed effects for both country and year, and a lagged dependent variable (and, in the appendix, bounding specifications using only fixed effects or lagged DV). The WGI stability and violence index also predicts expert and takedown scores in the expected directions but statistical significance is sensitive to specification (see footnote 6). In sum, we find robust evidence that democracies increase digital repression in response to terrorism, and some evidence that they so in response to instability, more generally. This suggests that the liberality of digital freedom is not a given in even the most democratic countries, but instead is, in part, conditional on the existence of internal and external threats to the government.

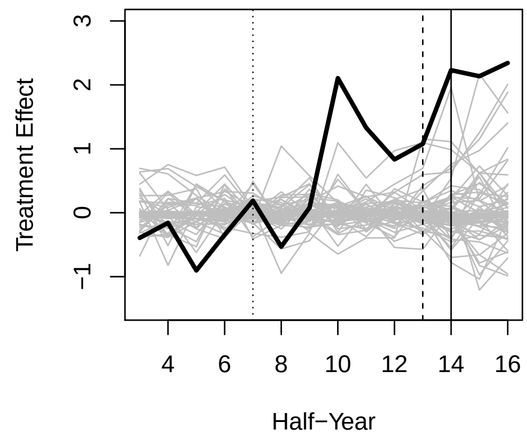
France

While the panel results establish a general relationship between terrorism and digital censorship, studies of cases that experienced terrorism shocks can help to better establish the plausibility of a causal relationship between terrorism and digital censorship. Meserve & Pemstein (2018) use synthetic control techniques to show that Turkey greatly increased its use of Google takedown requests after a spike in attacks by the Kurdistan Workers' Party (PKK), and the Gezi Park protests. Turkey was, arguably, an electoral democracy at the time, and these events presaged a rapid period of autocratization. Here, we examine a second, perhaps more worrying case. In particular, we focus on *liberal* democracies, again as indicated by V-Dem's RoW measure, and examine the case within this set, France, that experienced the largest terrorist attack, as measured by deaths, during our observation period. France experienced three half-years—half-years 13 and 14 in 2015 and half-year 16 in 2016—with deaths from terrorist attacks exceeding two standard deviations above the liberal democracy average, in our observation period. These include the Charlie Hebdo attack in January 2015, the Paris attacks in November 2015 (the largest attack on a liberal democracy in the dataset), and the Nice attack in July 2016. The latter two attacks represent the highest casualty incidents among liberal democracies in our dataset.

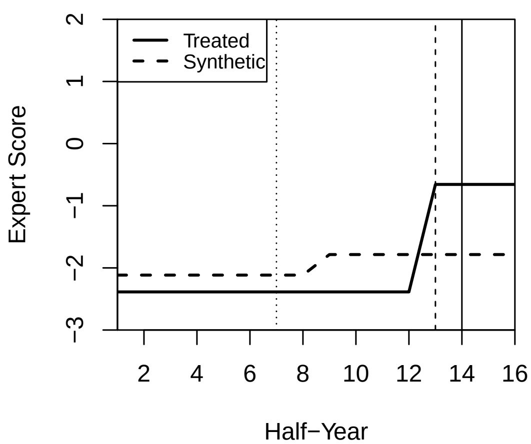
Figure 6 provides the results of a synthetic control analysis of the French case. Panels (a) and (b) present patterns in takedown scores, while panels (c) and (d) examine expert scores. The left-hand panels (a and c) compare France to a synthetic control case, while the right-hand panels (b and d) plot treatment effects for France (thick black line) and placebos constructed from every other democracy in the dataset. Each graph include three vertical lines. The first line indicates the 7th half-year (early 2012), when a single gunman killed



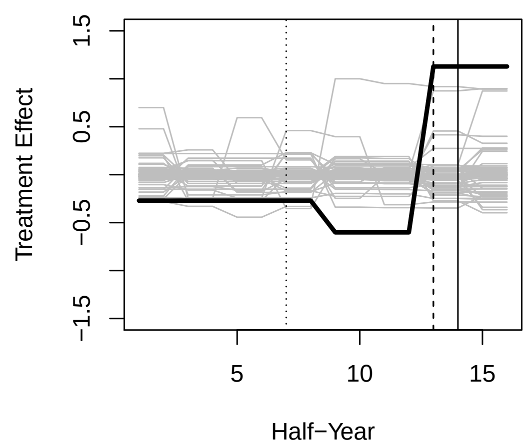
(a) Takedown score treated vs. synthetic



(b) Takedown score treatment effects



(c) Expert score treated vs. synthetic



(d) Expert score treatment effects

Figure 6: Synthetic control method, France. Vertical lines denote attacks: Montauban/-Toulouse [dotted], Charlie Hebdo [dashed], and Nice [solid]. We treat the pre-Hebdo period (half-years 1–12) as pre-treatment.

seven people, over two days, in Montauban and Toulouse. While this time period did not exhibit a particularly high overall count of terrorism deaths, the protracted nature of the event produced substantial news coverage, and this event arguably kicked off a period of heightened awareness of terrorism in France. The second two lines, in half-years 13 and 14, demarcate the Charlie Hebdo and Paris attacks. The Nice attack occurred in half-year 16. We take half-year 13, the Charlie Hebdo attack, as our treatment initiation period, when generating synthetic controls.

Looking first at panel (a), we see that takedown rates in France look similar to the control case until the tenth period, where they spike briefly. They then taper back towards the control before jumping dramatically in the period following the Charlie Hebdo attack, and growing after half-year 14. In panel (b) we can see that none of the placebo cases exhibits as large an estimated treatment effect as France. Turning to the bottom panels, panel (c) shows a close correspondence between France and the control case, until half-year 13, when experts report a substantial increase in France's censorship effort.[7] France jumps three full standard deviations—among democracies—on V-Dem's internet censorship measure. Mirroring the results for takedowns, no other case exhibits as large a treatment effect, if we consider Charlie Hebdo, or the Paris attacks, as the treatment period.

Taken together, our synthetic control analysis highlights the extent to which the Charlie Hebdo and Paris attacks triggered a period of heightened censorship in France. The placebo tests, depicted in panels (b) and (d), show that this correspondence between terror and censorship is unlikely to be a matter of chance. At the same time, for both dependent variables, the quality of our synthetic controls leaves a bit to be desired. France exhibits a more volatile takedown trajectory in the pre-Hebdo period than the control, and diverges substantially from it in half-year ten. The pre-treatment period match for the expert measure is substantially more clean, although France exhibits a reasonably large negative gap in the

[7]V-Dem data are yearly, so this jump probably reflects the second half of the year, when France instituted a state of emergency, in the wake of the Paris attacks. The half-yearly takedown data spikes in half-year 14, consistent with this interpretation.

half-years 9-12. Nonetheless, the magnitude of the treatment effect dwarfs this pre-treatment gap, and the analysis, as a whole, is largely consistent with a substantial, and unusual, increase in censorship in the wake of the Hebdo and Paris attacks.

4 Conclusion

Our results indicate that even the most liberal, consolidated, democracies respond to terrorism and internal threat by clamping down on the freedom of digital spaces. In contrast to existing literature showing this behavior in autocracies, our evidence comes from states which have fundamental protections for civil liberties. In practice, terrorists and insurgent groups may even accomplish some of their goals of shaking the liberal norms of democracies in the digital sphere by forcing regimes to tighten their control over internet speech, due to its potential use for recruiting, organizing, and coordinating dangerous activities. We show that, measured in a number of ways, and using panel techniques that demand a lot of the data, we nevertheless find a robust relationship between violence and digital censorship and control. While earlier work has highlighted this mechanism of censorship within democracies (Deibert et al. 2008, Deibert et al. 2010), and provided some initial tests (Meserve & Pemstein 2018), our findings provide substantial evidence of the generality and robustness of this argument. Further, our synthetic control study of France demonstrates our proposed mechanism in action, and shows how major terror shocks can cause liberal democracies to inhibit digital freedoms

Whether our paper represents a normatively disappointing finding depends on one's perspective about the inherent normative 'bad' of digital censorship itself. To an early web crusader, who imagined digital spaces as a neutral new frontier where individuals were reasonably free of state control, it is anathema that even the most liberal democracies will monitor and control internet speech between citizens. Yet, it is important to note that when regimes tighten their control over digital spaces, it is not necessarily the case that this control is done unlawfully or illegitimately. Indeed, in the liberal democracies our models describe, such as France, tightened digital control may be performed at the behest of elected officials, using institutionally provided powers, with relatively broad public approval. What our results suggest is that even democracies' digital spaces are subject to the state and its security interests when governments are threatened. Finding content restriction regulatory regimes that balance security interests of the state and the freedom to communicate online is already becoming a significant political flash point in democratic (and authoritarian) politics, just as the balance between security and media freedom was before it, in the 19th and 20th centuries.

References

Abadie, Alberto, Alexis Diamond & Jens Hainmueller. 2010. "Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program." *Journal of the American Statistical Association* 105(490):493–505.

Adler, Julie. 2011. "The Public's Burden in the Digital Age: Pressures on Intermediaries and the Privatization of Internet Censorship." *Journal of Law and Policy* 20(1):231–266.

Angrist, Joshua David & Jörn-Steffen Pischke. 2009. *Mostly harmless econometrics: an empiricist's companion*.

Bak, Daehee, Surachanee Sriyai & Stephen Meserve. 2018. "The internet and state repression: A cross-national analysis of the limits of digital constraint." *Journal of Human Rights* 17(5):642–659.

Barberá, Pablo, Ning Wang, Richard Bonneau, John T. Jost, Jonathan Nagler, Joshua Tucker & Sandra González-Bailón. 2015. "The Critical Periphery in the Growth of Social Protests." *PLOS One* 10(11). `https://doi.org/10.1371/journal.pone.0143611`.

Coppedge, Michael, John Gerring, Carl Henrik Knutsen, Staffan I. Lindberg, Svend-Erik Skaaning, Jan Teorell, David Altman, Michael Bernhard, Agnes Cornell, M. Steven Fish, Haakon Gjerløw, Adam Glynn, Allen Hicken, Joshua Krusell, Anna Luhrmann, Kyle Marquardt, Kelly McMann, Valeriya Mechkova, Olin Moa, Pamela Paxton, Daniel Pemstein, Brigitte Seim, Rachel Sigman, Jeffrey Staton, Aksel Sundstrom, Eitan Tzelgov, Yi-ting Uberti, Luca Wang, Tore Wig & Daniel Ziblatt. 2018. Varieties of Democracy Codebook v8. Technical report Varieties of Democracy Project: Project Documentation Paper Series.

Deibert, R.J., J.G. Palfrey, R. Rohozinski & J. Zittrain. 2008. *Access Denied: The Practice and Policy of Global Internet Filtering*. Cambridge, MA: MIT Press.

Deibert, Ronald, John Palfrey, Rafal Rohozinski & Jonathan Zittrain, eds. 2010. *Access Controlled: The Shaping of Power, Rights, and Rule in Cyberspace*. Cambridge, MA: MIT Press.

DeNardis, Laura. 2014. *The Global War for Internet Governance*. New Haven: Yale University Press.

Facebook Incorporated. 2018. "Transparency Report: Content Restrictions Based on Local Law." `https://transparency.facebook.com/content-restrictions` (Last accessed April 10, 2018).

Gelman, Andrew & Donald B Rubin. 1992. "Inference from Iterative Simulation using Multiple Sequences." *Statistical Science* 7:457–511.

Gohdes, Anita R. 2015. "Pulling the Plug: Network Disruptions and Violence in Civil Conflict." *Journal of Peace Research* 52(3):352–367.

Gohdes, Anita R. 2018. "The Relationship between State and Corporate Censorship." *APSA Comparative Politics Newsletter* 18(2):31–38.

Gohdes, Anita R. forthcoming. "Repression technology: Internet Accessibility and State Violence." *American Journal of Political Science* . Accessed at `http://www.anitagohdes.net/uploads/2/7/2/3/27235401/gohdes_repressiontech_ajps.pdf`.

Google LLC. 2018. "Transparency Report: Government Requests to Remove Content." `https://transparencyreport.google.com/government-removals/overview` (Last accessed April 10, 2018).

Gunitsky, Seva. 2015. "Corrupting the Cyber-Commons: Social Media as a Tool of Authoritarian Stability." *Perspectives on Politics* 13(1):42–54.

Kaufmann, Daniel, Aart Kraay & Massimo Mastruzzi. 2013. "World Governance Indicators." `http://info.worldbank.org/governance/wgi/index.aspx#home` (Last accessed January 27th, 2013).

Keck, Margaret E. & Kathryn Sikkink. 1998. *Activists Beyond Borders*. Ithaca: Cornell University Press.

King, Gary, Jennifer Pan & Margaret E. Roberts. 2013. "How Censorship in China Allows Government Criticism but Silences Collective Expression." *American Political Science Review* 107(02):326–343.

King, Gary & Jonathan Wand. 2007. "Comparing Incomparable Survey Responses: Evaluating and Selecting Anchoring Vignettes." *Political Analysis* 15(1):46–66.

Lührmann, Anna, Marcus Tannenberg & Staffan I. Lindberg. 2018. "Regimes of the World (RoW): Opening New Avenues for the Comparative Study of Political Regimes.".

MacKinnon, R. 2012. *Consent of the Networked: The Worldwide Struggle for Internet Freedom*. New York: Basic Books.

Marsden, C.T. 2011. *Internet Co-Regulation: European Law, Regulatory Governance, and Legitimacy in Cyberspace*. Cambridge: Cambridge University Press.

Meserve, Stephen. 2018. "Why Do Governments Censor? Expanding from State Survival to Content Regulation Theories in Political Science." *APSA Comparative Politics Newsletter* 18(2):55–59.

Meserve, Stephen A. & Daniel Pemstein. 2018. "Google Politics: The Political Determinants of Internet Censorhsip in Democracies." *Political Science Research & Methods* 6:245–263.

Microsoft Corporation. 2018. "Content Removal Request Report." `https://www.microsoft.com/en-us/corporate-responsibility/crrr/` (Last accessed April 10, 2018).

National Consortium for the Study of Terrorism and Reponses to Terrorism (START). 2018. "Global Terrorism Database [Data file].". Retrieved from http://www.start.umd.edu/gtd (Last accessed 11/30/2018).

Pemstein, Daniel, Brigitte Seim & Staffan I. Lindberg. 2016. "Anchoring Vignettes and Item Response Theory in Cross-National Expert Surveys.".

Pemstein, Daniel, Kyle L. Marquardt, Eitan Tzelgov, Yi-ting Wang, Joshua Krusell, & Farhad Miri. 2018. "The V-Dem Measurement Model: Latent Variable Analysis for Cross-National and Cross-Temporal Expert-Coded Data." *Varieties of Democracy Institute Working Paper* 21(3rd Ed).

Roberts, Margaret E. 2018. *Censored: Distraction and Diversion Inside China's Great Firewall*. Princeton: Princeton University Press.

Rød, Espen Geelmuyden & Nils B Weidmann. 2015. "Empowering Activists or Autocrats? The Internet in Authoritarian Regimes." *Journal of Peace Research* 52(3):338–351.

Seltzer, Wendy. 2008. "The Politics of Internet Control and Delegated Censorship." *American Society of International Law*. Available at SSRN: https://ssrn.com/abstract=1496056.

Shapiro, Jacob N & David A Siegel. 2015. "Coordination and Security: How Mobile Communications Affect Security." *Journal of Peace Research* 52(3):312–322.

Stan Development Team. 2015. "Stan: A C++ Library for Probability and Sampling, Version 2.9.0.".
URL: `http://mc-stan.org/`

Steinert-Threlkeld, Zachary. 2017. "Spontaneous Collective Action: Peripheral Mobilization During the Arab Spring." *American Political Science Review* 111(2):379–403.

Teorell, Jan, Michael Coppedge, Staffan Lindberg & Svend-Erik Skaaning. 2019. "Measuring Polyarchy Across the Globe, 1900–2017." *Studies in Comparative International Development* 54(1):71–95.
URL: `https://doi.org/10.1007/s12116-018-9268-z`

Teorell, Jan, Stefan Dahlberg, Sören Holmberg, Bo Rothstein, Natalia Alvarado Pachon & Richard Svensson. 2018. "The Quality of Government Standard Dataset, version Jan18." University of Gothenburg: The Quality of Government Institute, `http://www.qog.pol.gu.sedoi:10.18157/QoGStdJan18`.

The World Bank. 2018. "World Development Indicators." `http://data.worldbank.org/data-catalog/world-development-indicators`.

Tilly, Charles. 1990. *Coercion, Capital, and European States 990-1990*. Cambridge: Blackwell.

Tufekci, Zeynep. 2017. *Tiwtter and Tear Gas: The Power and Fragility of Networked Protest*. New Haven: Yale University Press.

Twitter Incorporated. 2018. "Transparency Report: Removal Requests." `https://transparency.twitter.com/en/removal-requests.html` (Last accessed April 10, 2018).

Warren, T Camber. 2015. "Explosive Connections? Mass Media, Social Media, and the Geography of Collective Violence in African States." *Journal of Peace Research* 52(3):297–311.

Whitten-Woodring, Jenifer & Douglas A. Van Belle. 2017. "The Correlates of Media Freedom: An Introduction of the Global Media Freedom Dataset." *Political Science Research and Methods* 5(1):179–188.

World Intellectual Property Organization. 2018. "IP Statistics Data Center.".
URL: `http://https://www3.wipo.int/ipstats/`

Zittrain, J. 2003. "Internet Points of Control." *Boston College Law Review* 44:653–688.

A Online Content Providers (OCP)

To expand on our firm selection choices (Facebook, Google, Microsoft, Twitter), we provide additional explanation in this section, as well as a menu of available firm choices in table 1. The OCP table is taken from Google's transparency report (Google LLC 2018), which represents to our knowledge the most complete list of the firms that issue transparency reports. Note, however, that this does not mean that every firm on the table produces systematic data or even regular biannual transparency reports, as we address shortly. Our overall philosophy was to choose the few firms that provide international products with lengthy time and country data coverage that are large enough to produce a diverse set of digital products.

First, we choose firms that do international business. Even otherwise large and important firms that issue transparency reports may be of extremely limited use, due to their scope. Consider firms like Comcast, a U.S. domestic internet provider and content producer, or Rogers, a Canadian internet service provider. These firms provide no significant utility for a cross-national analysis due to their largely single country focused activities. By contrast, the four firms we choose conduct operations in a large variety of country markets and are subject to government pressure from nearly all of them.

Second, some of the producers of transparency reports in the list do not consistently produce transparency data statistics at all, and their transparency reports are focused on organically describing transparency steps the firms take or aggregate numbers. Smaller firms sometimes produced a single or several reports but not systematic data; Cheezburger, for example, produced a single report in 2014 and never reported again. As another example, Naver, a Korean firm, produces requests data but only in the aggregate—overall number of government requests received with any country of origin information omitted. Furthermore, time period is a significant issue. Often, firms only started their transparency reports recently, in the mid-2010s, which limits their usefulness for panel data. Compare that to Google which starts in 2009 or Twitter who starts in 2012. We want to note that in five to ten years, gathering a large swathe of international firms with significant data will be much

Table 1: Firms Producing Transparency Reports

AOL	Apple	AT&T	Cheezburger
Cloudflare	Comcast	Credo Mobile	CyberGhost
Daum Kakao	Deutsche Telekom	Dropbox	Facebook
GitHub	Google	Hong Kong Transparency Report	Kickstarter
Korea Internet Transparency Report	LeaseWeb	LinkedIn	Lookout
Microsoft	Naver	Nest	Oath
Pinterest	reddit	Rogers	SaskTel
Snap	TekSavvy	TeliaSonera	Telstra
TradeMe	Tumblr	Twilio	Twitter
Uber	University of California, Berkeley	Verizon	Vodafone
Wikimedia Foundation	WordPress	Yahoo!	

List taken from Google LLC (2018).

more useful; we hope to perform more analysis when additional time-series data is available.

Finally, we skewed towards firms who produce a large number of products with different characteristics. Because multinational OCPs are now extremely consolidated, these data incorporate many of the most used digital products by average citizens around the world. Requests to takedown Google data, for example, represents the following cornucopia of vastly different products: Adsense, Android Market, Blog Search, Blogger, Buzz, Chrome Store, Feedback, Gmail, Google Adwords, Google App Engine, Google Books, Google Cloud Storage, Google Code, Google Docs, Google Earth, Google Fiber, Google Groups, Google Images, Google Maps, Google News, Google Notebook, Google Photos, Google Places, Google Play Apps, Google Play Music, Google Product Search, Google Profiles, Google Scholar, Google Sites, Google SMS Channels, Google Translate, Google URL Shortener, Google Videos, Google Video Search, Google Voice, Google+, iGoogle, Knowledge Graph, Orkut, Panoramio, Street View, Textcube, The Internet, Web Search, and Youtube. Similarly, Microsoft and Facebook own a tremendous number of widely used products—Facebook owns Instagram, WhatsApp, and Oculus VR, for example. Choosing large firms involved in a diverse set of content markets is an attempt to make the conclusions drawn from our analyses relatively broad and representative as well.

B Takedown Requests

Takedown requests provide a window into the otherwise difficult to observe process of internet censorship. They capture both 'good' and 'bad' censorship, and should be regarded largely as a value-free measure of when governments attempt to remove content from the internet (Meserve 2018). On the one hand, many requests reflect legally valid—and socially desirable—reasons for censorship. For instance, court orders may remove material that threatens national security, violates individuals' privacy, or unjustly defames a person. On the other hand, governments may seek to leverage content takedown requests to pursue political censorship. Indeed, both Google and Facebook provide vignettes of takedown requests that exhibit the latter flavor (Facebook Incorporated 2018, Google LLC 2018). For instance, the Italian police demanded that Google take down a video embarrassing to Silvio Berlusconi, and Google provides multiple examples of politicians attempting to leverage defamation, and even copyright, law to remove critical content. Indeed, even legal takedown demands can provide cover for political censorship. For example, an autocratizing Turkey made extensive use of legal takedowns to restrict political speech during the period of the Gezi Park protests.

Meserve & Pemstein (2018) provides an extended discussion of takedown requests, and, relying on Google's reports, describe the types and frequency of requests—national security, defamation, privacy, and so on—that OCPs regularly field. They also briefly describe compliance rates. We do not seek to extend this descriptive analysis. Rather, our focus is on using data from multiple reports to create a better general measure of the effort that governments put into internet content removal.

Takedown request reports provide an observable, and cross-nationally comparable, behavioral indicator of governments' content removal effort. But takedown request data are relatively sparse. The longest series, Google's, started only in 2009, and different firms vary in their cross-national market penetration, and the types of services that they provide.

C V-Dem Expert Censorship Question Details

The question used in the construction of the dependent variable for the expert analysis is worded as follows:

> **Question:** Does the government attempt to censor information (text, audio, or visuals) on the Internet?
>
> **Clarification:** Censorship attempts include Internet filtering (blocking access to certain websites or browsers), denial-of-service attacks, and partial or total Internet shut-

downs. We are not concerned with censorship of topics such as child pornography, highly classified information such as military or intelligence secrets, statements offensive to a particular religion, or defamatory speech unless this sort of censorship is used as a pretext for censoring political information or opinions. We are also not concerned with the extent of Internet access, unless there is absolutely no access at all (in which case the coding should be 0).[8]

Responses:

1. The government successfully blocks Internet access except to sites that are pro-government or devoid of political content.
2. The government attempts to block Internet access except to sites that are pro-government or devoid of political content, but many users are able to circumvent such controls.
3. The government allows Internet access, including to some sites that are critical of the government, but blocks selected sites that deal with especially politically sensitive issues.
4. The government allows Internet access that is unrestricted, with the exceptions mentioned above.

Note that we reverse the polarity of this question throughout our analysis, so that higher scores mean more censorship, and lower scores mean less censorship. V-Dem uses rater responses to this question, and an accompanying set of anchoring vignettes (King & Wand 2007, Pemstein, Seim & Lindberg 2016), to estimate the latent variable of internet censorship effort, using item response theory techniques (Pemstein, Marquardt, Tzelgov, Wang, Krusell, & Miri 2018). This approach to measuring internet censorship has the advantage of generality. As the clarification makes clear, it attempts to capture a wide variety of censorship approaches, from shutting down the internet, to active content filtering, and denial of service attacks. It is also closely focused on political or 'bad' censorship in a way

[8]Zero codings were moved to another variable and do not occur in the time period that we study.

that is impossible to capture with takedown request data. Nonetheless, it asks a lot of experts who must make inferences about a set of practices that can be difficult to observe, and its broad approach to defining censorship makes it an (at least) triple-barreled question, tapping multiple dimensions of censorship simultaneously. It may therefore lack consistency across cases.

D Factor analysis specification & diagnostics

We use one-dimensional Baysian factor analysis to estimate latent—or at least, difficult to directly observe—internet censorship effort from reported takedown requests in firm transparency reports. We assume a diagonal variance-covariance structure; that is, we assume that the errors are independent across manifest variables. The likelihood function for the model is

$$y_{i,j} \sim \mathcal{N}(\lambda_j \cdot \phi_i, \psi_{jj}), \tag{1}$$

where i indexes observations (country-years), j indexes manifest variables (firms), each ϕ_i is the latent censorship effort for observation i, each λ_j is the factor loading for the manifest variable j, and each ψ_{jj} is a diagonal element of the variance-covariance matrix, or error variance for manifest variable j, given our assumption of independent errors. We use vaguely informative conjugate priors:

$$\phi_i \sim \mathcal{N}(0, 1), \tag{2}$$

$$\psi_{jj} \sim \mathcal{IG}(0.01, 0.01), \tag{3}$$

$$\lambda_1 \sim \mathcal{TN}_{(0,\infty)}(0, 10), \text{and}, \tag{4}$$

$$\lambda_{j>1} \sim \mathcal{N}(0, 10). \tag{5}$$

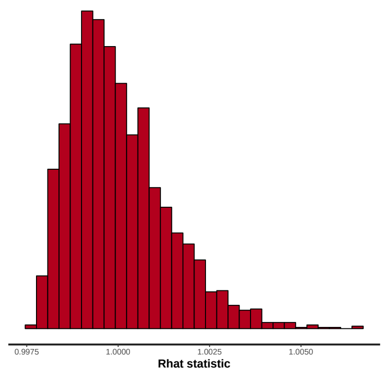
The standard normal prior on the latent traits ($\boldsymbol{\phi}$) establishes scale, while the priors on the remaining parameters are commonly used vague conjugates. Note that we use a truncated

```
data {
  int<lower=1> N;                    // Country-years
  int<lower=2> J;                    // # Manifest variables
  real<lower=-999> manifest[N,J];  // Data
  real a[J];                         // IG prior on Psi diagonal elements
  real b[J];                         // IG prior on Psi diagonal elements
}

parameters {
  real<lower=0> Psi_diag[J];       // Var-cov diagonal
  real phi[N];                       // Latent trait
  real<lower=0.01> Lambda1;         // First constrained loading
  real Lambda2p[J-1];              // Remaining parameters
}

transformed parameters {            // Roundabout way to constrain L1>0
  real Lambda[J];
  Lambda[1] = Lambda1;
  for (i in 2:J)
  Lambda[i] = Lambda2p[i-1];
}

model {
  for (i in 1:N)
  phi[i] ~ normal(0, 1);
  Lambda[1] ~ normal(0,10)T[0,];
  for (j in 2:J)
  Lambda[j] ~ normal(0, 10);
  for (j in 1:J) {
    Psi_diag[j] ~ inv_gamma(a[j], b[j]);
    for (i in 1:N) if (manifest[i,j] != -999) {
      manifest[i,j] ~ normal(Lambda[j] * phi[i], Psi_diag[j]);
    }
  }
}
```

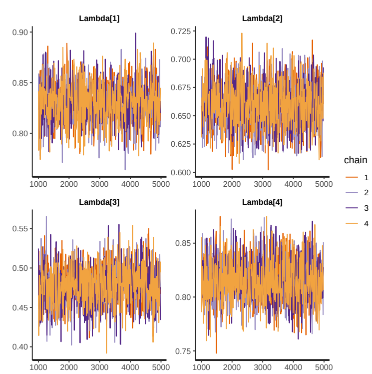
Figure 7: Stan code

normal prior on the first element of the loading vector ($\boldsymbol{\lambda}$) to identify the model with respect to rotation.

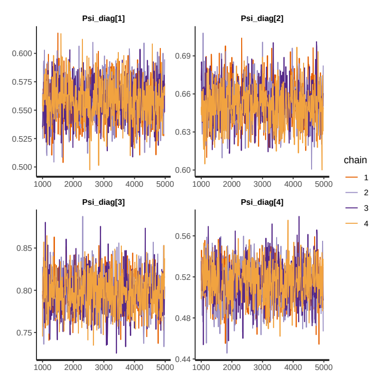
We implemented the model using Stan (Stan Development Team 2015). Figure 7 provides the source code. We logged all manifest variables and, to facilitate interpretation of factor loadings, standardized them before fitting. We ran four chains for 5000 iterations each, discarding the first 1000 iterations and retaining every 10th iteration from each chain, simulating a sample of 1600 posterior draws. Figure 8 provides convergence diagnostics, including the distribution of Gelman & Rubin (1992) convergence diagnostics, $\hat{r}$ for all estimated parameters—all of which fall below 1.1, and traceplots for the factor loadings and error variance parameters, which display good mixing.



(a) $\hat{r}$ Distribution



(b) Factor Loading Traceplot



(c) Error Variance Traceplot

Figure 8: Convergence checks

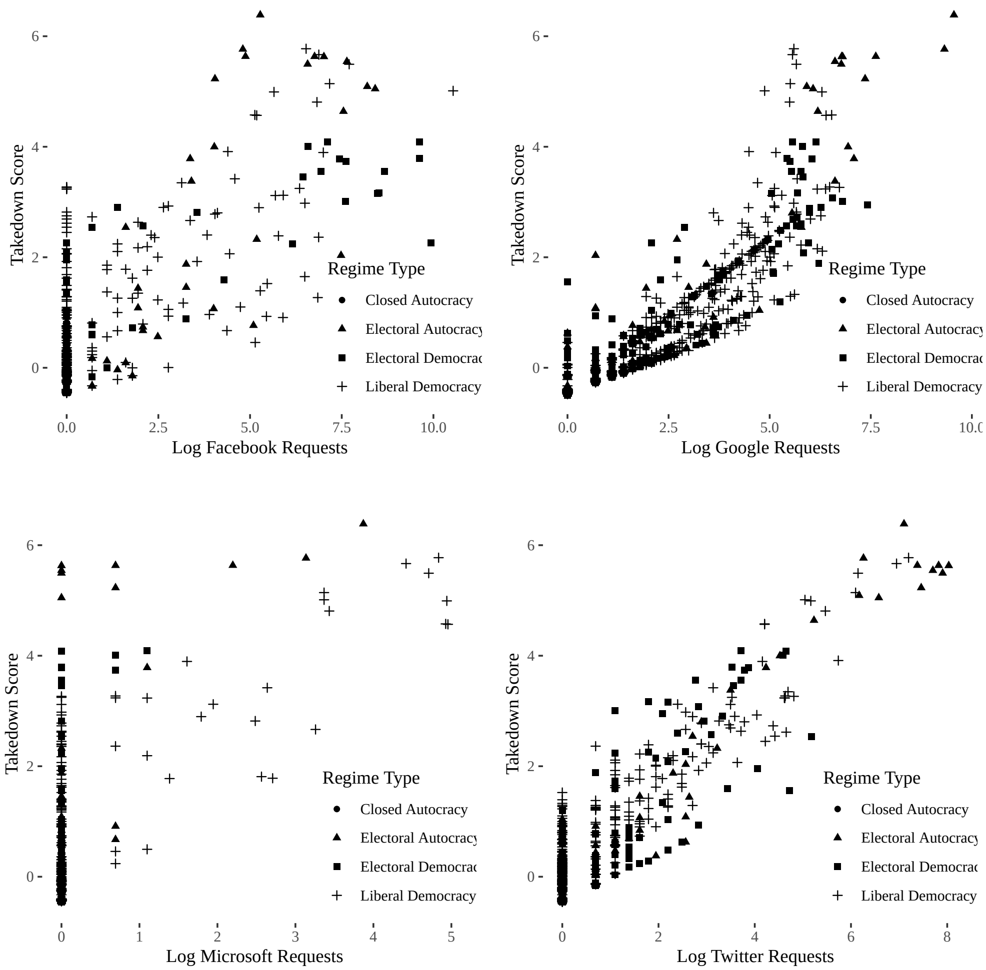


Figure 9: Latent-manifest variable scatterplots

Table 2 provides the correlation matrix between the latent variable and a the logged manifest variables. As we would expect, each manifest variable is more highly correlated with the latent variable than with any of the other manifest variables. Each manifest variable correlates quite highly with the latent variable, with the possible exception of logged Microsoft takedown request counts ($r = 0.64$). Figure 9 plots latent against manifest variables, to provide context.

Latent	**Facebook**	**Google**	**Microsoft**	**Twitter**
1.00	0.81	0.93	0.64	0.92
0.81	1.00	0.66	0.51	0.67
0.93	0.66	1.00	0.51	0.73
0.64	0.51	0.51	1.00	0.56
0.92	0.67	0.73	0.56	1.00

Table 2: Pairwise correlations between latent and manifest variables

E Descriptive Statistics & Regression Table

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
Takedown score	2,392	−0.03	0.89	−0.50	−0.43	−0.07	6.39
Expert score	2,918	−0.31	1.37	−2.39	−1.45	0.61	4.63
WGI S&V	2,722	−0.19	0.96	−3.31	−0.78	0.57	1.53
GTD events	2,920	29.33	132.78	0	0	3	2,431
GTD deaths	2,920	66.89	392.83	0	0	2	7,399
Patents pc	2,616	0.0003	0.001	0.00	0.00	0.0001	0.01
GDP pc 2010	2,612	12,891.49	18,648.83	218.28	1,296.01	14,473.47	111,001.00
Business start days	2,293	25.09	42.06	0.50	8.00	30.00	694.00
Internet use %	2,686	38.38	29.09	0.00	11.12	63.66	98.24
Polyarchy	2,920	0.54	0.26	0.02	0.32	0.78	0.94

Table 3: Descriptive statistics, full sample

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
Takedown score	1,543	0.09	0.93	−0.50	−0.43	0.18	5.77
Expert score	1,641	−1.13	0.70	−2.39	−1.60	−0.83	1.30
WGI S&V	1,530	0.19	0.80	−2.81	−0.28	0.86	1.53
GTD events	1,641	13.96	60.37	0	0	2	625
GTD deaths	1,641	23.18	192.42	0	0	0	4,702
Patents pc	1,494	0.0005	0.001	0.00	0.0000	0.0004	0.01
GDP pc 2010	1,510	17,515.77	21,214.80	319.66	2,811.16	24,461.02	111,001.00
Business start days	1,452	23.37	47.91	0.50	7.00	27.00	694.00
Internet use %	1,514	47.70	29.10	0.26	21.98	73.43	98.24
Polyarchy	1,641	0.74	0.13	0.50	0.63	0.87	0.94

Table 4: Descriptive statistics, democracies only

	Dependent variable:					
	Takedown score			Expert score		
	(1)	(2)	(3)	(4)	(5)	(6)
WGI S&V	−0.03 (0.06)			−0.06*** (0.02)		
Log GTD Events		0.04** (0.02)			0.02*** (0.01)	
Log GTD deaths			0.07*** (0.02)			0.03*** (0.01)
Log GDP pc	0.39* (0.23)	0.42* (0.22)	0.48** (0.22)	−0.002 (0.08)	−0.02 (0.08)	−0.001 (0.08)
Log Bus. start days	−0.03 (0.04)	−0.04 (0.04)	−0.04 (0.04)	−0.02 (0.01)	−0.02 (0.01)	−0.02 (0.01)
Internet use %	−0.01** (0.003)	−0.01*** (0.003)	−0.01*** (0.003)	0.001 (0.001)	0.0002 (0.001)	−0.0001 (0.001)
Polyarchy	−0.93** (0.43)	−0.82* (0.43)	−0.82* (0.43)	−0.87*** (0.17)	−0.85*** (0.17)	−0.88*** (0.17)
Log patents pc	−44.68 (91.06)	−45.03 (90.88)	−48.12 (90.36)	−11.05 (30.27)	−10.42 (30.23)	−12.37 (30.14)
Lag takedown score	0.61*** (0.03)	0.60*** (0.03)	0.59*** (0.03)			
Lag expert score				0.65*** (0.02)	0.65*** (0.02)	0.65*** (0.02)
Constant	−2.61 (2.09)	−2.88 (2.04)	−3.39* (2.03)	0.31 (0.77)	0.53 (0.75)	0.37 (0.75)
Observations	1,184	1,184	1,184	1,342	1,342	1,342
R^2	0.90	0.90	0.90	0.96	0.96	0.96
Adjusted R^2	0.89	0.89	0.89	0.96	0.96	0.96
Note:						*p<0.1; **p<0.05; ***p<0.01

Table 5: Fixed effects with lagged DV (figure 5 in main text)

F Robustness Checks

	Dependent variable:					
	Takedown score			Expert score		
	(1)	(2)	(3)	(4)	(5)	(6)
WGI S&V	-0.11			-0.14^{***}		
	(0.07)			(0.03)		
Log GTD Events		0.10^{***}			0.05^{***}	
		(0.02)			(0.01)	
Log GTD Deaths			0.14^{***}			0.06^{***}
			(0.02)			(0.01)
Log GDP pc	0.71^{***}	0.72^{***}	0.83^{***}	0.09	0.04	0.08
	(0.25)	(0.25)	(0.24)	(0.11)	(0.11)	(0.11)
Log Bus. start days	-0.005	-0.02	-0.03	-0.03	-0.04^{*}	-0.04^{*}
	(0.04)	(0.04)	(0.04)	(0.02)	(0.02)	(0.02)
Internet use %	-0.01^{*}	-0.01^{**}	-0.01^{***}	-0.001	-0.002	-0.002
	(0.003)	(0.003)	(0.003)	(0.002)	(0.001)	(0.001)
Polyarchy	-1.78^{***}	-1.50^{***}	-1.52^{***}	-2.17^{***}	-2.13^{***}	-2.17^{***}
	(0.49)	(0.49)	(0.48)	(0.23)	(0.23)	(0.23)
Log patents pc	-7.96	-8.61	-15.71	-32.51	-30.92	-34.61
	(95.16)	(94.38)	(93.12)	(41.54)	(41.44)	(41.18)
Constant	-4.80^{**}	-4.97^{**}	-5.87^{***}	-0.07	0.46	0.13
	(2.32)	(2.25)	(2.23)	(1.05)	(1.03)	(1.03)
Observations	1,263	1,263	1,263	1,342	1,342	1,342
R^2	0.84	0.84	0.85	0.92	0.92	0.93
Adjusted R^2	0.83	0.83	0.83	0.92	0.92	0.92
Note:						*p<0.1; **p<0.05; ***p<0.01

Table 6: Fixed effects only

	Dependent variable:					
	Takedown score			Latent score		
	(1)	(2)	(3)	(4)	(5)	(6)
WGI S&V	-0.04^{**}			-0.01		
	(0.02)			(0.01)		
Log GTD Events		0.03^{***}			0.01^{***}	
		(0.01)			(0.003)	
Log GTD Deaths			0.02^{***}			0.01^{***}
			(0.01)			(0.003)
Log GDP pc	0.02	0.01	0.02	-0.01	-0.01	-0.01
	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)
Log Bus. start days	-0.01	-0.01	-0.01	0.01	0.01	0.01
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Internet use %	0.0004	0.0002	0.0002	0.0004	0.0004	0.0003
	(0.001)	(0.001)	(0.001)	(0.0004)	(0.0004)	(0.0004)
Polyarchy	0.01	0.01	-0.05	-0.07	-0.07	-0.07
	(0.14)	(0.13)	(0.13)	(0.06)	(0.06)	(0.06)
Log patents pc	9.75	10.30	6.58	4.97	5.91	5.10
	(11.79)	(11.62)	(11.59)	(5.22)	(5.17)	(5.15)
Lag takedown score	0.97^{***}	0.96^{***}	0.97^{***}			
	(0.01)	(0.01)	(0.01)			
Lag expert score				0.96^{***}	0.96^{***}	0.96^{***}
				(0.01)	(0.01)	(0.01)
Constant	-0.17	-0.13	-0.12	0.06	0.07	0.05
	(0.14)	(0.13)	(0.13)	(0.06)	(0.05)	(0.06)
Observations	1,184	1,184	1,184	1,342	1,342	1,342
R^2	0.87	0.87	0.87	0.94	0.94	0.94
Adjusted R^2	0.87	0.87	0.87	0.94	0.94	0.94
Note:					*p<0.1; **p<0.05; ***p<0.01	

Table 7: Lagged DV only

	Dependent variable:											
	Facebook	Google	Microsoft	Twitter	Facebook	Google	Microsoft	Twitter	Facebook	Google	Microsoft	Twitter
WGI S&V	−0.64**	−0.60**	0.14	−0.63**								
	(0.32)	(0.26)	(0.58)	(0.25)								
Log GTD Events					0.10	0.08	−0.04	0.08				
					(0.06)	(0.06)	(0.10)	(0.06)				
Log GTD Deaths									0.14***	0.15***	−0.04	0.15***
									(0.05)	(0.06)	(0.07)	(0.06)
Log GDP pc	−3.58***	−4.38***	−9.51***	−4.27***	−3.41***	−4.46***	−9.63***	−4.36***	−3.48***	−4.45***	−9.64***	−4.35***
	(1.19)	(1.07)	(3.48)	(1.07)	(1.19)	(1.08)	(3.44)	(1.07)	(1.19)	(1.07)	(3.44)	(1.07)
Log Bus. start	−0.17	−0.33**	0.02	−0.29**	−0.19	−0.36**	0.04	−0.32**	−0.17	−0.35**	0.02	−0.31**
	(0.16)	(0.14)	(0.25)	(0.14)	(0.16)	(0.14)	(0.25)	(0.14)	(0.16)	(0.14)	(0.25)	(0.14)
Internet use %	0.0001	0.005	−0.002	0.005	−0.005	0.001	0.001	0.001	−0.005	0.0005	−0.001	0.0001
	(0.02)	(0.01)	(0.04)	(0.01)	(0.02)	(0.01)	(0.04)	(0.01)	(0.02)	(0.01)	(0.04)	(0.01)
Polyarchy	−2.37	−3.98**	−5.01	−3.98**	−2.68	−4.39**	−5.11	−4.45**	−2.48	−4.19**	−5.08	−4.23**
	(1.91)	(1.80)	(3.22)	(1.80)	(1.90)	(1.80)	(3.23)	(1.79)	(1.89)	(1.79)	(3.21)	(1.78)
Log patents pc	533.25	740.28*	−1.70	743.67*	440.94	674.02	8.26	673.76	436.72	663.02	9.90	662.41
	(442.85)	(419.59)	(582.89)	(418.35)	(441.95)	(420.40)	(583.20)	(419.44)	(439.36)	(418.21)	(582.87)	(417.21)
Lag Facebook	0.54***				0.55***				0.55***			
	(0.05)				(0.05)				(0.05)			
Lag Google		0.15**				0.15**				0.14**		
		(0.06)				(0.06)				(0.06)		
Lag Microsoft			−0.31***				−0.31***				−0.32***	
			(0.11)				(0.11)				(0.11)	
Lag Twitter				0.17***				0.16***				0.16***
				(0.06)				(0.06)				(0.06)
Constant	34.48***	42.62***	91.67***	41.85***	33.76***	44.25***	92.65***	43.60***	34.30***	43.99***	92.83***	43.39***
	(11.02)	(9.86)	(31.33)	(9.82)	(11.06)	(9.86)	(31.12)	(9.84)	(10.98)	(9.80)	(31.11)	(9.78)
Observations	502	590	249	590	502	590	249	590	502	590	249	590
R^2	0.89	0.86	0.95	0.86	0.89	0.86	0.95	0.86	0.89	0.86	0.95	0.86
Adjusted R^2	0.87	0.83	0.92	0.83	0.87	0.83	0.92	0.83	0.87	0.83	0.92	0.83

Note: *p<0.1; **p<0.05; ***p<0.01

Table 8: Manifest DVs, fixed effects OLS with lagged DV