

THE EFFECT OF PIRACY WEBSITE BLOCKING ON CONSUMER BEHAVIOR¹

Brett Danaher

Argyros School of Business and Economics, Chapman University,
Orange, CA 90041 U.S.A. {danaher@chapman.edu}

Jonathan Hersh

Argyros School of Business and Economics, Chapman University,
Orange, CA 90041 U.S.A. {hersh@chapman.edu}

Michael D. Smith

School of Information Systems and Management, Heinz College, Carnegie Mellon University,
Pittsburgh, PA U.S.A. {mds@cmu.edu}

Rahul Telang

School of Information Systems and Management, Heinz College, Carnegie Mellon University,
Pittsburgh, PA U.S.A. {rtelang@andrew.cmu.edu}

In this study, we ask what drives the success or failure of various supply-side anti-piracy enforcement actions such as piracy website blocking. We do this in the context of three court-ordered events affecting consumers in the United Kingdom: We first study Internet Service Providers' blocking of 53 video piracy sites in 2014 and of 19 piracy sites in 2013, and we then study the blocking of a single dominant site, "The Pirate Bay," in 2012.

We show that blocking 53 sites in 2014 caused treated users to decrease piracy and to increase their usage of legal subscription sites between 7% and 12%. It also caused an increase in new paid subscriptions. We find similar results for the blocking of 19 piracy sites in 2013. However, blocking a single site in 2012 caused no increase in usage of legal sites but instead caused users to increase visits to other unblocked piracy sites and VPN sites. We find evidence that increased search and learning costs associated with piracy drive the effectiveness of blocking multiple sites rather than just one primary site.

This suggests that to increase legal IP use when faced with a dominant piracy channel, the optimal policy response must block multiple channels of access to pirated content, a distinction that the current literature has not made clear.

Keywords: Piracy, regulation, digital distribution, motion picture industry, natural experiment

¹Gal Oestreicher-Singer was the accepting senior editor for this paper. Jui Ramaprasad served as the associate editor.

Introduction

One of the most important challenges facing media industries today is whether and how copyright policy should be adapted to the realities of the digital age. The invention and subsequent adoption of file sharing technologies² have eroded the strength of copyright law across many countries. In the 10 years following the introduction of Napster in 1999 worldwide revenues from recorded music fell by 50% (Goldman 2010), and in the four years after the introduction of BitTorrent, home video sales declined in the film industry by 27% (Zentner 2012). The vast majority of the academic literature find that digital piracy causes a significant reduction in sales of music and motion picture content (for a review of this literature, see Danaher, Smith, and Telang 2014).

Given the well-established economic harm from piracy, it is important to understand how to design and enforce copyright policy in an age of filesharing technologies. Government regulators, copyright holders, and Internet platforms are pursuing a variety of efforts to respond to piracy through both direct enforcement against consumers (i.e., demand-side enforcement) and direct enforcement against websites and distribution technologies that facilitate piracy (i.e., supply-side enforcement).

The literature on demand-side enforcement has been relatively uniform in finding that efforts targeting consumers can be effective at increasing legal sales in the context of threats of lawsuits against music sharers (e.g., Bhattacharjee et al. 2006), laws making it easier for rights holders to take pirates to court (e.g., Adermon and Liang 2014), and notice/penalty sending actions against pirates (Danaher, Smith, Telang, and Chen 2014). However, taking direct action against users is costly, and frequently generates negative publicity or political sentiment. For these reasons, recent anti-piracy efforts have focused on the supply-side of piracy: targeting the sites and networks that make pirated content available to consumers. The literature on the effectiveness of these interventions is divided. While some studies show that supply-side anti-piracy actions can increase legal sales of blockbuster entertainment products (Danaher and Smith 2014; Peukert et al. 2017), others show that supply-side interventions have no impact on total piracy levels (Poort et al. 2014) or on legitimate consumption (Aguiar et al. 2018). Our study seeks to explain this empirical contrast in the literature by testing a

specific hypothesis regarding why some supply-side efforts are effective at reducing the harm from piracy while others are not.

Specifically, we argue that the Megaupload shutdown studied by Danaher and Smith (2014) and Peukert et al. (2017) involved the complete shutdown of a major piracy cyberlocker,³ a shutdown that caused the removal of vast amounts of copyright infringing content from the Internet. This also affected piracy “link sites,” pirate sites that do not host infringing content but serve as convenient, trusted aggregators of links to find pirated content hosted on cyberlockers. The removal of all of the content on Megaupload rendered many of these sites and links defunct and thus may have meaningfully increased the inconvenience of piracy for a number of illegal downloaders. Moreover, with such a large amount of content removed, the cost to pirates to source all of this content and replace a site like Megaupload would be high.

In contrast, Poort et al. (2014) and Aguiar et al. (2018) study two different interventions that cut off access to pirated content through particular websites, without removing the actual source content from the Internet. Neither study found a decrease in total piracy levels as pirates simply found other paths to the same content and, in particular, Aguiar et al. found that shutting down a piracy link site caused the emergence of a number of new piracy linking sites to replace it. The ineffectiveness of these two supply-side efforts is perhaps explained by the fact that, because the source piracy content still exists on the Internet, the costs to users to discover new sites that provided links to their desired content was sufficiently low that the users did not need to switch their consumption to legal channels. As well, the cost to pirates to replace such sites was also likely low, given that no source content had to be replaced.

The removal of source piracy content from the Internet may be effective at changing consumer behavior, but it is not always convenient or politically viable to shut down entire sites. As a result, supply-side anti-piracy interventions like the Megaupload shutdown are relatively uncommon. It is increasingly more common for governments to attempt to disrupt access to pirated content through website blocking strategies, whereby Internet Service Providers (ISPs) within a country are ordered to not resolve domain names pertaining to a website that has been shown to facilitate illegal copyright infringement. As with the interventions in Aguiar et al. and

²As is customary in the economics and information systems literatures, we use the terms *filesharing* and *piracy* interchangeably to refer collectively to all of the major forms of Internet media piracy, including BitTorrent and other peer-to-peer protocols, direct cyberlocker downloads, and illegal linking/streaming sites.

³A cyberlocker is a cloud site or server that provides file storing and sharing. In the context of piracy, cyberlockers are repositories of illegal content that users can download, whereas other types of sites merely provide links or tracker files that link to pirated content stored elsewhere.

Poort et al., such actions do not remove any source pirate content from the Internet, but instead aim to make said content more difficult to access.

Based on the two studies cited above, one would conclude that such actions are unlikely to decrease piracy or increase legal consumption. We suggest, however, that this may not tell the full story. While disrupting access to pirated content through *a single channel* may not be effective at changing consumer behavior, we hypothesize that simultaneously disrupting access to pirated content through *multiple channels* may decrease piracy and increase legal consumption, and we find evidence that supports this hypothesis.

There are two primary reasons why disrupting access to pirated content through multiple channels/sites might have a different impact than disrupting just one primary site. First, Internet users incur search costs and learning costs to switch between Internet sites/portals, and prior research suggests that these costs may be lower when consumers are already aware of the other sites (Chen and Hitt 2002; Goldfarb 2006). Such fixed costs do appear to apply to piracy websites as piracy sites with quality content may be difficult to find, difficult to learn to use, or present the risk of content loaded with malware (Danaher et al. 2010). When multiple piracy sites are blocked, the probability that an individual is already aware of and proficient with other (trusted) unblocked sites will be lower, increasing the expected fixed cost to the user of future piracy. When this is true, adopting a new, well-known, easy-to-use, and safe legal site may be seen as a viable alternative to continued piracy.

Second, it may be that disrupting access to pirated content sends a signal to consumers about overall anti-piracy enforcement severity, and the perceived strength of the signal may depend on the number of channels/paths/sites that are disrupted. This is supported by the “broken windows” theory of criminology, which suggests that visible signs of unpoliced crime or antisocial behavior (in this case, freely available piracy sites) sends a signal of low enforcement that encourages further crime and disorder (Kelling and Wilson 1982) and that the salience of law enforcement activity increases its effectiveness (Dur and Vollard 2019). The implication is that actions that disrupt access to pirated content through multiple channels may have an increased chilling effect on users’ overall propensity to pirate.

However, it remains entirely possible that any disrupted channels to source content will simply be replaced with new ones linking to the same content (as in Aguiar et al.), or that pirates will easily find other channels to the same source content when it has not been removed from the Internet, and so our study seeks to test the hypothesis that the number of channels

disrupted (and thus the strength of the intervention) impacts the effectiveness of this type of supply-side anti-piracy enforcement. We do this using a series of interventions that occurred in the same country and over a relatively short time period, allowing us to better test under what circumstances supply-side interventions are most likely to increase consumption through legal channels, and why.

Specifically, we examine three waves of piracy website blocking in the United Kingdom: the blocking of “The Pirate Bay” in May 2012 (the most popular piracy site at the time), the blocking of 19 major video piracy sites in November 2013, and the further blocking of 53 video piracy sites in November 2014. In each case, no pirated content was removed from the Internet, but the enforcement actions attempted to block UK users from reaching pirated content through particular domains. As such, our study is uniquely able to study multiple instances of the same type of supply-side intervention in the same country and at varying degrees of strength, and we are uniquely able to test the hypothesis that the number of piracy channels disrupted moderates the effectiveness of this type of supply-side anti-piracy intervention.

We employ panel datasets on the behaviors of a large number of UK Internet users and implement a generalized version of the difference-in-differences design. Although all individuals in the UK were blocked from accessing the sites, the intensity—or “bite”—of this treatment varied from user to user. An individual who had previously used the blocked sites frequently was more heavily shocked than an individual who was only an infrequent user of the blocked sites, and both of these individuals were more heavily shocked than non-users of the blocked sites. Thus, our generalized version of the difference-in-difference approach asks whether an individual’s pre-block usage of the blocked sites was correlated with her pre-post change in visits to legal media sites and/or alternate unblocked piracy sites, a methodology we further describe and justify in the empirical section below.

Our results show that, consistent with Poort et al. and Aguiar et al., the 2012 blocking of one major piracy site caused no increase in legal consumption, as users of the blocked site increased their visits to unblocked piracy sites. However, we find that the 2013 blocking of 19 major video piracy sites and the 2014 blocking of 53 major video piracy sites caused meaningful decreases in total piracy as well as a 7% to 12% increase in usage of paid legal streaming sites among users affected by the blocks. Thus our results confirm earlier findings that disrupting access to content through a single dominant channel is unlikely to be effective at changing consumer behavior. However, our results paint a more complete picture of the likely effectiveness of supply-side interventions

by showing that interventions that disrupt multiple paths to pirated content, even those that do not remove source pirated content from the Internet, can affect a consumer's choice between legal and illegal media channels. We also find evidence pointing to increased search and learning costs as the mechanism driving this result.

Background on Video Filesharing and Anti-Piracy Enforcement

The advent of the BitTorrent filesharing protocol in 2003 led to a rapid spread of Internet piracy for motion pictures, and a number of studies (described in more detail below) have causally linked this spread of motion picture piracy with significant losses in major film and television sales channels. In our study, we analyze under what conditions supply-side efforts to combat pirate are most likely to be effective.

Methods of Supply-Side Anti-Piracy Enforcement

The main categories of supply-side enforcement actions include piracy takedown notices, piracy site shutdowns, and piracy website blocking, and each of these involves a different action and different legal and technical requirements. Because different categories of piracy sites operate differently, these actions may have different consequences depending on the site targeted.

Website shutdowns occur when government enforcement organizations target a site known to be primarily dedicated to copyright infringement and shut down that site by seizing the servers as well as the domain associated with that site. Examples of this include the Megaupload shutdown, the Kino.to shutdown, and the MegafilesHD shutdown. If the site in question hosted pirated content (e.g., a piracy cyberlocker), then the shutdown removes pirated content from the Internet, and reproducing a new copy of that site may be difficult and costly. But if the site was a piracy linking site and merely provided convenient links to download or stream pirated content hosted on other sites, then a shutdown is more like an attempt to disrupt access to pirated content through a convenient channel. Notably, when a site that hosted content (such as Megaupload.com) is shut down, many of these linking sites are negatively affected as well if they pointed to content on the host site.

Takedown notices involve using legal channels to send notices to websites compelling them to take down specific pirated content or links from their pages. Again, whether this

involves the removal of content or the disruption of access depends on whether the site targeted was hosting the content or simply providing links to pirated content.

Finally, an increasingly common anti-piracy method is piracy website blocking, a strategy that has been attempted in over 25 countries to date (Corey 2016). This involves requiring ISPs to block access to websites that facilitate illegal consumption of content (by not resolving domain requests to those sites), and thus by nature does not target the removal of any pirated content. Website blocking of this sort may be an attractive alternative strategy to site shutdowns because it does not involve cross-country cooperation for non-domestic websites. However, while blocking access to websites may be easier than shutting them down, website blocks are easily circumventable for three reasons. First, users of blocked sites may access the same content or links to the same content on other unblocked sites. Second, because all of the blocked piracy site content remains on the Internet, a new pirate site with a new (but often similar) domain can be created to mirror the content on the blocked site or provide a proxy through which it can be accessed. In either case, if the new domain is not blocked, then individuals can access the same content or links through the new site. Finally, users may subscribe (typically for a fee) to a Virtual Private Network (VPN) and access the blocked sites through this connection.⁴

Piracy Website Blocking in the UK

The UK has used website blocking to fight piracy since October 2011 when British Telecom and five other UK ISPs were ordered by the High Court to block their customers from accessing Newzbin2, an indexing site for pirated content posted to the Usenet.⁵ Following the Newzbin2 precedent, as of April 2015, over 125 copyright infringing sites were subject to court-ordered blocks in the UK.

Our present analysis concerns three waves of UK blocks that occurred in 2012, 2013, and 2014. Specifically, in April 2012 six major UK ISPs were ordered by the courts to block access to The Pirate Bay, a major website that indexes the tracker files necessary to gain access to pirated media files through BitTorrent.⁶ These ISPs made up 98% of the market, and so

⁴By using a VPN, a user can appear to be attempting to access a blocked site from another country, and thus the request to the site will resolve.

⁵<https://www.bbc.com/news/technology-14322957>

⁶Specifically the ISPs Everything Everywhere, Sky, TalkTalk, Telefónica, and Virgin Media were ordered to block access in April resulting in the block occurring in May. The sixth ISP, British Telecom, implemented the blocks in June.

the vast majority of Internet users were subject to the blocks.⁷ The Pirate Bay was the most popular piracy site in the UK at the time, with reportedly 3.7 million users.⁸ In November 2013, these six ISPs were ordered to block access to 19 additional piracy websites that provided access to copyrighted video content. Finally, in November 2014, they were again ordered to block access to a total of 53 additional piracy sites.

Existing Literature

There is a significant body of work on the relationship between piracy and sales of video content, including Bai and Waldfogel (2012), Bounie et al. (2006), Danaher et al. (2010), DeVany and Walls (2007), Hennig-Thurau et al. (2007), Herz and Kiljanski (2018), Ma et al. (2014), McKenzie and Walls (2016), Rob and Waldfogel (2007), Smith and Telang (2009), and Zentner (2012). The majority of this literature finds evidence of sales displacement caused by video piracy, although Bounie et al. finds that this displacement does not extend to the French box office and Smith and Telang find a lack of displacement late in the lifecycle of a film (once it is being broadcast on cable).

There have also been a number of studies on whether anti-piracy enforcement actions by governments can influence pirates to turn to legal channels of consumption (see Table 1 for a summary of the empirical literature evaluating government anti-piracy enforcement actions). Studies on the effects of demand-side anti-piracy interventions targeting consumers of pirated content show that these actions tend to be effective in reducing pirated content and/or increasing legitimate consumption (Adermon and Liang 2014; Bhattacharjee et al. 2006; Danaher, Smith, Telang, and Chen 2014), although McKenzie (2017) shows no impact of such interventions on box office revenues. However, demand-side interventions are frequently seen as draconian, and political taste for such enforcement activity has diminished (Danaher et al. 2017).

With a decline in demand-side enforcement, there has been an increase in supply-side enforcement efforts. As noted above, the literature is divided as to the effectiveness of such efforts with Danaher and Smith (2014) and Peukert et al. (2017) finding that supply-side actions are effective at increasing legal consumption of blockbuster films, and Poort et al. (2014) and Aguiar et al. (2018) finding no impact from supply-side interventions. We suggest that this divergence in

findings is related to whether supply-side enforcement sufficiently increases the search and learning costs—or signals the strength of anti-piracy legal enforcement—to consumers of pirated content. The three supply-side studies that considered the removal of pirated source content from the Internet (and thus rendered all other links to that content defunct) found an increase in legal sales. Specifically, both studies on the shutdown of Megaupload.com, which hosted over 25 petabytes of mostly pirated content, found that it increase sales of large blockbuster films.⁹ And while Reimers (2016) studied another type of supply-side enforcement—the use of a private company to seek out pirated e-books and legally compel websites to take them down and delist them in search results—that also led to the removal of source content from the web and caused an increase in sales.

In contrast, the two studies that focused on disrupting access to pirated content through a dominant channel found such efforts to be ineffective. Aguiar et al. studied the shutdown of Kino.to, but this was a popular German piracy linking site. This shutdown did not lead to the removal of pirated content from the Internet, but merely disrupted conveniently aggregated access to pirated content hosted on other sites. Poort et al. studied the ISP blocking of The Pirate Bay in Germany; as discussed, the fact that it was merely blocked meant that there were still a number of ways consumers might gain access to the source content on the site. Thus, both studies considered anti-piracy enforcement actions designed to disrupt access to content through a single dominant channel, and both studies found no decrease in total piracy or increase in legitimate consumption.

However, it is impossible to conclude whether the differences in the results of these studies are due to differences in the types of interventions studied (and especially the relative strengths of each intervention) or other factors unrelated to the strength of the intervention (e.g., differences in the countries affected or the timeframes analyzed). Our data provide a unique opportunity to bridge this gap in the literature by analyzing a series of website blocks of different strengths affecting users in the same country implemented over a relatively short timeframe.

By analyzing the UK blocking of The Pirate Bay in 2012 and two successive multi-site waves of blocks, we ask whether simultaneously blocking multiple avenues of access to pirated

⁷See <https://www.ispreview.co.uk/review/top10.php>

⁸<http://www.theguardian.com/technology/2012/apr/30/british-isps-block-pirate-bay>

⁹Megaupload also had a sister site, Megavideo.com, which provided convenient links to stream pirated content hosted on Megaupload.com, and it is likely that many other link sites pointed to content at Megaupload. Megavideo and all of these other links became inactive once Megaupload was shut down. Further, Peukert et al. found that shutting down Megaupload also harmed sales of smaller, independent films.

Table 1. Summary of Empirical Literature on Government Anti-Piracy Enforcement*

<i>Authors</i>	<i>Topic</i>	<i>Demand or Supply Side?</i>	<i>Source Content Removed?</i>	<i>Result</i>
Danaher, Smith, Telang and Chen (2014)	HADOPI "three strikes law" in France	Demand		Approximately 25% increase in digital music sales
Adermon and Liang (2014)	IPRED law in Sweden	Demand		36% increase in music sales for six months, then return to normal levels after lax enforcement of law
Bhattacharjee et al. (2006)**	Highly publicized legal threats by industry against individual filesharers	Demand		Decreased tendency to share copyright infringing files, but majority of content remained available
McKenzie (2017)	Graduated response anti-piracy laws in 6 countries	Demand		No increase in box office revenues of films
Aguiar et al. (2018)	Shutdown of Kino.to (popular German piracy streaming/linking site)	Supply	No	No increase in legal consumption, increase in piracy at other sites, emergency of new piracy link sites to replace Kino.to
Poort et al. (2014)	Dutch ISP domain blocking of The Pirate Bay	Supply	No	No lasting decrease in total Dutch piracy
Danaher and Smith (2014)	Shutdown of Megaupload.com	Supply	Yes	6.5-8.5% increase in digital revenues from Hollywood films
Peukert et al. (2017)	Shutdown of Megaupload.com	Supply	Yes	Increase in box office for large films, decrease in box office for smaller, indie films.
Reimers (2016)**	Piracy "takedown notices" and search de-listing	Supply	Yes	15% increase in sales for book titles whose pirated counterparts were removed from websites and delisted from search engines.

*In addition to papers analyzing government-sponsored anti-piracy enforcement, Sivan et al.(2019) show experimentally that deprioritizing pirate links for movies from search engine results causes a significant increase in legal consumption for those movies.

**Technically these actions were taken by private parties, but they relied on the legal regime to enforce.

content sufficiently increases the (search and learning) costs of continued piracy or sends a strong enough signal about anti-piracy enforcement activity to increase legal consumption even while blocking one dominant site does not. We note that our study may also be seen as an empirical test of Dey et al. (2018), who provide analytical results suggesting that the effectiveness of supply-side anti-piracy enforcement will depend on the strength of the enforcement action.

Data

We obtained data from an anonymous Internet consumer panel tracking company, which we refer to as PanelTrack in

this paper.¹⁰ PanelTrack offers individuals compensation to participate in their panel, and then subsequently installs software that monitors a user's PC Internet activity unnoticeably in the background for as long as the user remains in the panel.¹¹

¹⁰Because our study is about piracy, PanelTrack required that the company remain anonymous. However, this tracking company is one of several leaders in the field and their data has been used in other peer reviewed papers to study the behavior of consumers on the Internet.

¹¹ Although this observation occurs in the background, we cannot rule out that the sample is biased due to Hawthorne effects. However, these data are the standard in the entertainment industry as well as others for learning about Internet consumer behavior over time. Because we study changes within users before and after the blocks, our methodology helps to difference out any degree to which users behave differently as a result of participating in the panel.

We use each wave of blocks as a natural experiment affecting piracy at the blocked sites to determine how consumers respond. They may change usage of remaining unblocked piracy sites, increase usage of legal sites, circumvent the blocks by using VPNs, or simply stop consuming the media that they had been pirating.

2014 Blocks

We use a panel of 24,620 UK Internet users, covering the time period from August 2014 to February 2015, to study the impact of the 53 site blocks in November 2014. This panel includes each individual's monthly visits to (1) blocked piracy sites, (2) unblocked piracy sites, (3) VPN sites, and (4) paid legal streaming sites like Netflix or LoveFilm.¹² The panel is unbalanced; a number of these users joined the sample after our study began or left before it ended. As such, we observe 67,098 user-months. Our difference-in-difference approach can be effectively applied to unbalanced panels, but we also show in Appendix D that all of our 2014 results hold when estimated using only the balanced panel of individuals observed in all seven months.

It is important to ask what constitutes a site "visit." PanelTrack defines a visit as a "session," which can include a number of page views at a site. In other words, should an individual visit www.netflix.com, and from there navigate to www.netflix.com/browse, and then watch a film, this would count as one visit in our data as it is a single session. Should the individual close her browser or navigate to another site and then visit Netflix again, that would become a second visit.¹³ We believe that in the context of our empirical methodology, changes in the relative number of visits to legal sites most closely proxies for changes in films or television programs viewed. Of course, it is not a perfect measure, as an individual may visit a site and then choose not to view anything. For piracy sites, a visit to a site likely proxies for the intent to download or watch something, but it is possible that a visit is associated with multiple pirated downloads since it is possible to download a number of files within one session/visit. A visit may also lead to no pirated consumption if the user does not find the content for which she is looking.

It is reasonable to ask what type of content (TV shows, movies, music, etc.) is being accessed on piracy sites. Using our data, it is not feasible to track which files are being downloaded or watched, only that a piracy site was visited. As we describe in Appendix B, the piracy sites we track hold a variety of content, but we eliminated piracy sites that were wholly dedicated to music, anime, adult, games, or eBook content. Still, the remaining sites likely contain a mix of content, and measuring pirated content by type is difficult given the illegal nature of such activity. One study (Watters et al. 2011) scraped popular BitTorrent tracking sites and categorized each tracker by type of content. They found that almost 43.3% of BitTorrent downloads were movies, 16.5% were music, 29.1% were TV shows, 3.7% were pornography, 4.4% were games, and the remainder was a variety of other content. These results plus the fact that we eliminated sites dedicated to non-television/film content mean that the most likely purpose of a visit to a piracy site in our data is to consume movies or television shows.

Table 2 shows some descriptive statistics for the panel of UK Internet users before and after the blocks in November 2014. We present average monthly site visits, prior to and after the blocks, to the categories of sites considered as our outcome variables of interest. We exclude observations from November from this table, as the blocks were in the process of being implemented, and thus November is considered "partially" treated. We define treatment intensity as the average monthly pre-block visits to sites that were subsequently blocked in November 2014, under the logic that an individual who was using the blocked sites more heavily in the pre-period was more intensively treated by the blocks than an individual making no or light use of the blocked sites. We see that the November 2014 blocks were effective at reducing visits to blocked sites. Visits to blocked sites dropped by 88% from the 3 months before the blocks to the 3 months after.¹⁴ It also appears that visits to unblocked sites decreased and visits to paid sites increased. However, we will investigate this more rigorously using a generalized difference-in-difference analysis to control for the time trends underlying the blocks that may be driving the changes.

¹²See Appendix A for a list of the blocked sites in each wave of blocks and Appendix B for an explanation of how we determined the sets of legal subscription websites, unblocked piracy sites, and VPN sites.

¹³PanelTrack would also count a second visit if an individual spent a sufficiently long period of time inactive at a site and then began re-engaging with that site again.

¹⁴There are several reasons why the drop may not be 100%. First, only the top six ISPs were compelled to implement the blocks, and 2% of users connected to Internet through nonparticipating ISPs. Second, users on VPNs could still access the blocked sites and PanelTrack's machine-side software would still detect such visits. Finally, it may be that some ISPs did not fully operationalize blocks to all of the sites by the beginning of our post-period.

Table 2. 2014 Summary Statistics

	<i>Blocked Sites</i>		<i>Unblocked Sites</i>		<i>VPN Sites</i>		<i>Paid Streaming Sites</i>		<i>N</i>
Treatment Intensity (average monthly pre-block blocked site visits)	Pre	Post	Pre	Post	Pre	Post	Pre	Post	
0-1	0.0	0.1	3.7	4.5	0.1	0.1	1.5	2.2	19525
1-5	1.9	0.5	9.9	7.9	0.1	0.1	1.4	2.2	3323
5-10	6.8	1.3	20.6	11.8	0.0	0.1	2.4	2.7	798
10-50	20.2	3.6	47.0	24.9	0.1	0.2	2.6	3.1	852
50+	84.7	15.6	179.2	62.1	0.1	0.2	1.0	3.2	122

Notes: This table shows average monthly site visits by site category and treatment intensity. Treatment intensity is measured by average monthly visits, prior to the block, to sites that would eventually be blocked. N shows number of users within each bucket of treatment intensity over which averages are calculated.

2012 and 2013 Blocks

Due to a change in policy surrounding privacy concerns, PanelTrack would not release individualized data for the 2012 and 2013 waves of blocks. We instead received monthly data with individuals aggregated by groups stratified by users' pre-block usage of websites that were subsequently blocked. Thus, instead of measuring the treatment intensity (the bite of the blocks) on individual users by how many visits each user made to blocked sites prior to the blocks, for these waves we observe aggregate group behavior. The treatment intensity of the blocks for each group of users is defined as that group's overall average monthly visits to the blocked sites in the three months before the blocks. For example, during the 2012 Pirate Bay block, one of the groups averaged only 1 visit per user per month to The Pirate Bay before it was blocked, while the group with the heaviest Pirate Bay use averaged 230 visits per user per month to that site in the months before it was blocked.¹⁵ Thus for the 2012 wave and for the 2013 wave we have separate datasets, each of which observes aggregate visits for 10 different consumer groups for the 7 months surrounding each wave of blocks. Importantly, because we only observe behavior at the group level and cannot observe individuals entering and exiting the panel, we required that the groups be formed only of individuals who were observed during all 7 months, and thus the groups were created from a balanced panel of users.

Again because of privacy concerns, PanelTrack would not reveal the number of individuals comprising each group in our aggregate group data. They provided visit counts scaled to the UK population of Internet households by using sample weights to expand individual data to the population sample. We were assured these sample weights were designed such that the sample is representative of the population of UK Internet users.¹⁶ Thus, what we observe for each group during each month is the aggregate population-scaled visits for that group during that month to each of the site types in question (legal, illegal, VPNs). Because we also know the number of projected users in each group, we can divide scaled visits by projected users to determine the average number of visits per user per month. Importantly, we confirmed with PanelTrack that each group of users in both datasets is comprised of at least 200 raw users, and thus the data points are generated by a relatively large sample within each group.

A positive feature of our 2012 and 2013 data is that we were able to separate the unblocked piracy sites into two categories of piracy sites—unblocked torrent sites and unblocked cyberlocker sites—and obtain each groups' aggregate monthly visits to each of these subtypes of piracy sites.¹⁷ This allows us to measure whether any increase or decrease in unblocked piracy was disproportionately driven by a particular piracy protocol.

¹⁵Exact details of how PanelTrack sorted their users into these ten groups are in Appendix C. This explains why there is no "control group" per se, although such a concern does not prohibit drawing inferences from the generalized version of the difference-in-difference model with a continuous treatment intensity variable.

¹⁶Although we would have preferred raw data for these two waves of blocks, PanelTrack's practice of scaling the data is consistent with industry practices such as scaling television ratings data to determine population audience sizes.

¹⁷Piracy linking sites were less common during this time, but those that were present we categorized with cyberlockers as they mostly linked to content on cyberlockers.

Table 3. Average Monthly Visits Per User Before and After 2012 Pirate Bay Block

	<i>The Pirate Bay</i>		<i>Unblocked Torrent Sites</i>		<i>Unblocked Cyberlockers</i>		<i>VPN Sites</i>		<i>Paid Streaming Sites</i>	
Group	Treatment Intensity /Pre	Post	Pre	Post	Pre	Post	Pre	Post	Pre	Post
1	0.8	0.1	4.9	3.7	4.0	1.6	0.1	0.0	0.3	0.3
2	2.0	0.2	11.2	9.9	7.2	4.2	0.2	0.1	0.6	0.3
3	2.1	0.4	3.5	3.1	3.0	3.9	0.2	0.2	0.4	0.5
4	4.2	0.4	13.9	10.8	6.4	3.5	0.2	0.1	0.5	0.5
5	6.8	0.5	16.4	11.9	7.0	6.5	0.1	0.2	2.1	1.0
6	12.8	1.5	40.4	29.1	13.1	4.9	0.3	0.2	3.6	0.3
7	17.1	2.6	25.6	21.2	11.0	14.9	0.4	0.2	2.5	0.9
8	38.5	2.1	32.2	28.9	13.1	7.6	0.4	0.3	2.0	1.3
9	55.0	4.5	42.9	50.8	12.3	11.0	0.9	0.8	1.6	2.1
10	231.2	11.7	80.2	102.6	15.4	12.0	0.6	1.7	2.7	2.0

Notes: This table shows average monthly visits per user by site category for each treatment intensity group. Users were aggregated into these groups by PanelTrack based on their treatment intensity, or pre-period visits to The Pirate Bay, in accordance with the guidelines in Appendix C.

Table 3 provides average monthly visits per user by type of site during the pre-period (February, March, and April 2012) and post period (June, July, and August 2012). We exclude May 2012 from this table as the block was in the process of being implemented this month and we consider it “partially” treated.

The second column in Table 3 indicates each group’s average monthly pre-block visits to The Pirate Bay, and thus it is our measure of treatment intensity for that group. Clearly there is dispersion across groups in usage of The Pirate Bay, indicating that the bite of the treatment was different for each group. Visits to blocked sites drop by 80% to 95% across the various groups, indicating an effective block. Heavy users of The Pirate Bay were also heavier users of other torrent and cyberlocker sites. Visits to pirate sites appear more common than visits to paid streaming sites. This may be because during this period, a number of paid streaming sites were in their infancy (Netflix, for example, launched in January 2012) and thus were not yet widely adopted.

Table 4 reports the summary statistics for the data surrounding the 19 site blocks in November 2013. Again, the second column is average monthly pre-block visits to the blocked sites, and thus indicates the bite of the treatment on each group. All groups decrease their usage of blocked sites by 80% to 90%. Visits to paid streaming sites appear higher in 2013 than they were in 2012, likely due to increased adoption levels of these services. Visits to unblocked piracy sites appear lower, but this is likely because 19 major piracy

sites (as well as a number of their mirror sites) are included in the blocked sites rather than just one, leaving less piracy remaining at unblocked sites.

Because visits to blocked sites drop by 80% to 90% in each wave of blocks, we have clear discrete shocks to piracy at the blocked sites. In the next section, we present our empirical model to analyze these experiments and determine their causal effect on consumer behavior.

Empirical Model and Results

November 2014 Blocking of 53 Major Piracy Sites

We first turn our attention to how blocking 53 major piracy sites in November 2014 affected consumer use of legal and illegal media channels. Because changes in outcome variables, such as use of paid streaming channels, might change over time for reasons other than the block, we employ a generalized version of the difference-in-difference model using a continuous treatment variable. This is a common method when the treatment being studied is not binary but rather varies in intensity, which is the case with our data (see, for example, the seminal case of the Card (1992) study on minimum wage and unemployment rates). Here the treatment variable is a measure of the bite of the treatment on each affected user. We define each user’s treatment intensity as proportional to their average monthly visits to the 53 blocked

Table 4. Average Monthly Visits Per User Before and After November 2013 Blocks

	Blocked Sites		Unblocked Torrent Sites		Unblocked Cyberlockers		VPN Sites		Paid Streaming Sites	
Group	Treatment Intensity /Pre	Post	Pre	Post	Pre	Post	Pre	Post	Pre	Post
1	1.4	0.3	1.2	1.1	0.7	0.5	0.0	0.0	2.0	2.1
2	2.6	0.4	1.7	1.0	1.1	0.7	0.0	0.0	2.2	2.4
3	3.4	0.4	2.3	1.6	1.5	0.5	0.0	0.0	3.2	2.0
4	3.9	0.5	1.6	1.7	1.3	0.6	0.3	0.0	1.9	2.1
5	5.1	0.9	2.0	1.9	1.5	1.1	0.3	0.1	3.1	4.1
6	5.6	0.7	1.5	1.4	2.2	0.9	0.1	0.1	1.9	1.8
7	7.5	0.8	2.4	2.9	1.9	1.4	0.0	0.1	2.4	2.3
8	12.5	1.3	3.2	3.1	2.8	1.2	0.0	0.3	2.4	3.2
9	20.0	2.8	4.0	4.6	3.6	1.7	0.1	0.4	2.0	3.2
10	51.5	5.0	6.1	6.8	8.8	2.6	0.1	0.7	3.9	6.8

Notes: This table shows average monthly visits per user by site category for each treatment intensity group. Users were aggregated into these groups by PanelTrack based on their treatment intensity, or pre-period visits to blocked sites, in accordance with the guidelines in Appendix C.

sites before the blocks were enacted. Our logic is that users who visited these blocked sites more before they were blocked were more impacted by the treatment than users who visited them less.

In line with prior use of this generalized difference-in-difference model, we identify the causal effect of the blocks by comparing individuals' pre-post changes in the outcome variables of interest with those individuals' treatment intensity. We acknowledge that individuals are self-selecting into different measures of treatment intensity (based on their tendency to visit the blocked sites). We control for time-invariant differences across users by including individual fixed effects in our model. Our approach relies on several assumptions. First, the allocation of treatment cannot be based on expectations of the outcome variable. In this case, the same intervention was applied to all individuals (the website blocks), and the intensity of treatment only varies because individuals had varying (static) preexisting levels of use of the blocked sites. Thus treatment was not allocated based on expectation of our outcome variables, such as unblocked piracy site visits or legal subscription site visits. Second, our approach relies on the assumption that a user's month to month changes in the outcome variable would be uncorrelated with treatment intensity in the absence of the treatment, which is the parallel trends assumption. Because we observe the individual for three months before the blocks, we can partly test this assumption by testing whether there is a correlation during the pre-period.

Specifically, we estimate a model of the form

$$Visits_{it} = \beta_0 + \beta_1 \cdot month_t + \beta_2 TreatmentIntensity_i + \beta_3 TreatmentIntensity_i \cdot month_t + \mu_i + \varepsilon_{it} \quad (1)$$

Where $Visits_{it}$ indicates the number of website visits by individual i in month t to the set of sites in question (paid legal subscription sites, unblocked piracy sites, or VPN sites), is a vector of month fixed effects, $TreatmentIntensity_i$ is the average number of monthly visits prior to the block made by individual i , μ_i is an individual fixed effect and ε_{it} is the error term. Our outcome variables are visits to sites and should be modeled as count data as they can only take on non-negative integer values.¹⁸ We estimate equation (1) using a negative binomial fixed effects regression because of this feature, and prefer it over Poisson because of the former's better handling of over-dispersion as is present in our data. We use the negative binomial fixed effects model as developed by Hausman et al. (1984). We note that negative binomial fixed effects models do not estimate a true fixed effect due to the incidental parameters problem, and thus can still estimate a coefficient for $TreatmentIntensity_i$ in this model. However, we also demonstrate in Appendix D that our results are robust in sign and significance to estimation using Poisson as well as using OLS with log of visits plus one as the outcome variable (both of which estimate a true fixed effect and are not subject to the incidental parameters problem).

¹⁸Cameron and Trivedi (2005) suggest using generalized linear models for count data when the arrival parameter or expected value is less than 10, as is the case here.

The chief coefficient of interest is β_3 , which indicates the degree to which treatment intensity is correlated with individuals' month-to-month changes in site visits of interest. Under the parallel trends assumption, we expect β_3 to be statistically indistinguishable from 0 during the months before the treatment. Then, after the treatment, β_3 gives the causal effect of the blocks on site visits. Figure 1 shows the estimation of equation (1) plotting and its standard errors over time. We observe that the 2014 website blocks caused a decrease in unblocked piracy site visits, an inconclusive impact on VPN site visits, and a persistent increase in legal subscription site visits. The parallel trends assumption appears to hold for VPN sites visits and legal subscription site visits. Though it does not appear to hold for unblocked piracy site visits, the discrete drop in the post period is much larger than the slight preexisting downward trend.

The natural next step is to estimate the size of these effects. To do so, we estimate the following model:

$$\begin{aligned} Visits_{it} = & \beta_0 + \beta_1 Post_t + \beta_2 TreatmentIntensity_i + \\ & \beta_3 Post_t \cdot PartialIntensity_i + \beta_4 \cdot PartialTreatment_t \\ & + \beta_4 Partial_t \cdot TreatmentIntensity_i + \mu_i + \varepsilon_{it} \end{aligned} \quad (2)$$

Here we have replaced the month dummies with an indicator variable for the "partial treatment" period (November 2014) as it is a partial treatment month and then an indicator for the post period, equal to 1 for the months of December, January, and February. Under the identifying assumption, the interaction of treatment intensity and the partial treatment indicator represents the impact of the blocks on the outcome variable in the month they were being implemented and the interaction of treatment intensity with the post dummy represents the effect of the blocks during the following three months.

In Table 5, we see from the coefficient on the post treatment dummies that piracy at unblocked sites was generally decreasing (for users with 0 treatment intensity) during this time while traffic to legal subscription sites was increasing. The coefficient on treatment intensity is estimated due to the aforementioned caveats associated with negative binomial fixed effects models; it implies that heavier users of the blocked sites made lighter use of paid streaming sites in the pre-period (consistent with descriptive statistics in Table 1). While the interactions between the partial treatment dummy and treatment intensity may be interesting, we focus on analysis of the effect of the blocks once they were fully implemented, which is represented by the interaction between the post dummy and treatment intensity. We see that it is negative and significant for unblocked piracy sites, positive but insignificant for VPN sites, and positive and significant

for paid legal streaming sites. In column 3, the coefficient on the interaction term is 0.0104; this implies that an individual who visited the blocked sites one more time in the pre-period increased his usage of paid legal streaming sites by 1.04% more than he otherwise would have had the blocks not affected him (i.e., were his treatment intensity zero).

It is clear from Table 2 that the panel is unbalanced as some individuals were not observed during some months. Although fixed effects models are robust to unbalanced panels, one might worry whether the months in which individuals are not observed are somehow selected with bias. For example, if individuals periodically choose to be unobserved due to their behaviors. As a check on this, in Table D1 in Appendix D we re-estimate our results using only a strictly balanced panel of individuals who are observed in all seven months of the dataset. Our results remain similar in sign and significance. However, in the balanced panel, the coefficient on the interaction between post and treatment intensity for legal subscription sites is .0169. This indicates that a user with one additional pre-block visit to blocked sites increases usage of paid legal streaming by 1.69% more after the blocks then she otherwise would have. While the balanced panel may be preferred for the reason stated above, the unbalanced panel has a much larger number of observations. As well, the balanced panel may select for users who are more consistent consumers of media overall if users in the panel sometimes drop out of observation specifically when they are not visiting any sites. Because the unbalanced panel and the balanced panel each have advantages, we consider these estimates as indicating the range of possible effects of the blocks.

Due to the fact that we estimate the effect of the blocks on multiple outcome variables that may substitute for each other, the error terms across these equations could be correlated. Testing for the robustness of inference to a possibly correlated error structure naturally suggests the implementation of the seemingly unrelated regression (SUR) model (Zellner 1962). This is most feasibly implemented as a linear model. Thus we estimate a SUR model of equation (2) in Table D2 in Appendix D using $\log(visits + 1)$ as the outcome variables. The results are qualitatively and quantitatively similar to the same model estimated using OLS in Table D3.

Although we have checked for the existence of preexisting differential trends, we also consider two placebo/falsification tests. In Table D8 in Appendix D, we drop months 4 through 7 and estimate a placebo model where we presume the blocks happened just after the first month and a placebo model where we presume the blocks happened just after the second month. In all cases we fail to reject the null when the outcome variable is visits to legal subscription sites. When the outcome is

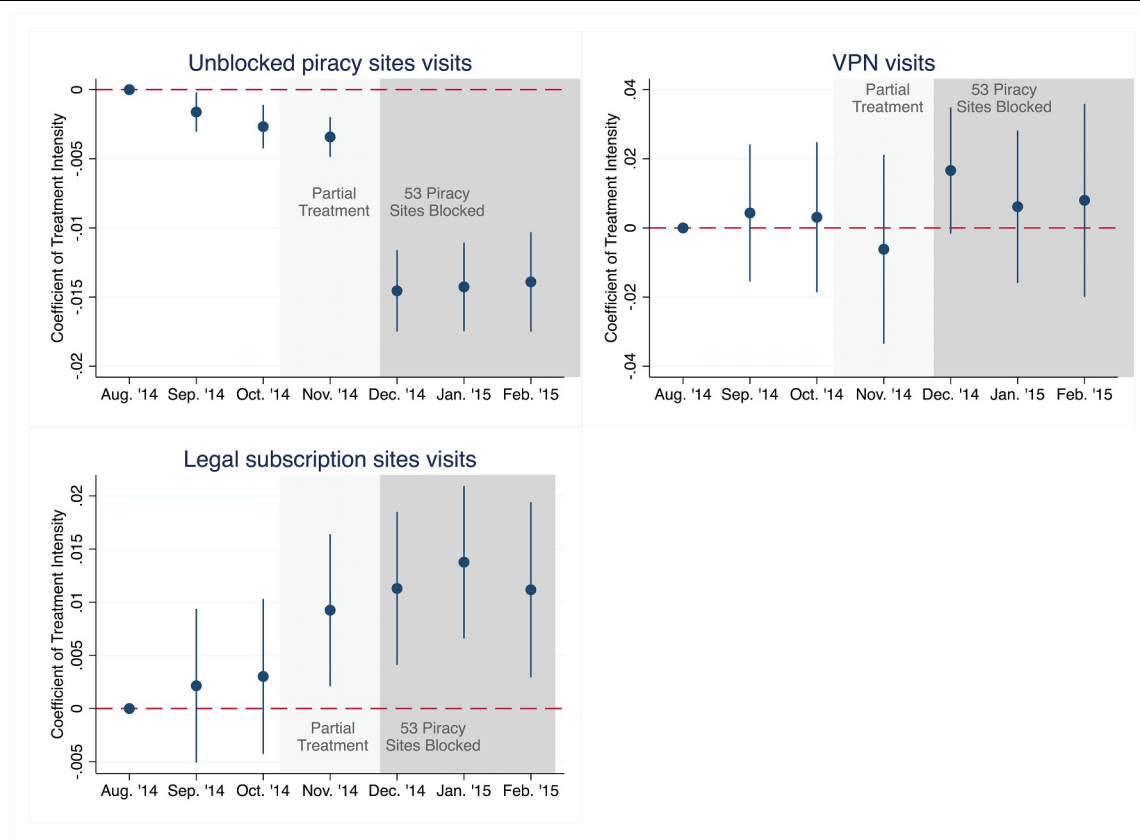


Figure 1. Effect of 2014 Blocks on Outcomes

Table 5. Estimated Impact of 53 Site Block in November 2014 on User Site Visits

	(1)	(2)	(3)
Dependent Variable:	Unblocked Piracy Sites	VPN Sites	Legal Subscription Sites
Post Treatment	-0.250*** (0.0113)	-0.0995 (0.0828)	0.130*** (0.0178)
Treatment Intensity	0.00701*** (0.000608)	-0.0266** (0.00815)	-0.0180*** (0.00230)
Post × Treatment Intensity	-0.0125*** (0.000969)	0.00999 (0.00610)	0.0104*** (0.00229)
Partial Treatment	-0.196*** (0.0132)	-0.152 (0.111)	-0.0194 (0.0234)
Partial × Treatment Intensity	-0.00173*** (0.000175)	-0.00851 (0.0122)	0.00743** (0.00288)
Constant	0.369*** (0.0115)	-0.0637 (0.113)	-0.271*** (0.0188)
Individual FE?	Y	Y	Y
N	46733	2546	29685
Individuals	11847	556	7095
Log-Likelihood	-88399.33	1706.17	-36720.16

Notes: Standard errors are shown in parentheses and clustered by user. + $p < 0.10$ * $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

visits to unblocked piracy sites, we do reject the null hypothesis in our pre-period placebo tests, but the estimated coefficient of interest is much smaller than it is for our real estimates in Table 5. This is consistent with the small negative preexisting trend that we observed in Figure 1, and our interpretation of these results is made accordingly.¹⁹

One concern with our identification strategy is that high treatment intensity users, though most affected by the blocks, might be the least likely to turn to legal channels due to their affinity for piracy. If this were true, it would bias our result towards zero, making it harder to find an effect. However, this is not what we observe, as more heavily treated user decrease their use of unblocked piracy sites, and disproportionately turn towards legal sites.

While our results demonstrate that the 2014 website blocks caused an increase in visits to legal subscription sites, one might ask whether all of these additional visits came from users who were already subscribed or if the blocks caused some non-subscribers to begin paying for legal subscriptions. We do not have e-commerce data on actual signups or subscriptions, but we can assume that an individual who made no visits to legal subscription sites in the 3 months before the blocks was not a paying subscriber. We first limit our sample to only individuals who made no visits to subscription sites in the pre-period. We then define a binary variable equal to 1 if the user made a number of visits to legal subscription sites above some threshold in the post period, and equal to zero otherwise. Using this variable we estimate the following cross-sectional model on the post period:

$$NewSubscriber_i = \beta_0 + \beta_1 TreatmentIntensity_i + \varepsilon_i \quad (3)$$

Equation (3) measures whether treatment intensity—pre-blocked usage of subsequently blocked sites—is associated with a higher likelihood of becoming a new subscriber in the post period. We estimate this model via a logistic regression, the results of which are shown in Table 6. Making a single visit to subscription sites in the post period may indicate exploration of the site without actually signing up, which is why we vary the threshold number of visits necessary in the post period required to indicate becoming a new subscriber (columns 1 through 3).

The coefficient on treatment intensity in each column indicates a positive and statistically significant relationship between pre-period usage of blocked sites and post-period likelihood of becoming a new subscriber. For each additional

visit to blocked sites in the pre-period, an individual's probability of becoming a new subscriber in the post period increases by about 1% over the baseline probability. The similarity of these coefficient across all three columns indicates that the threshold number of visits required to indicate an actual subscription does not materially impact our estimates. We acknowledge that without a fixed effects model, one might suggest that individuals with higher treatment intensity are more likely to use paid subscription sites as well. We have partly controlled for this problem by only looking at individuals who, in spite of their varying levels of treatment intensity, were non-subscribers during the pre-period. Thus we are only looking at individuals with similarly low propensity to subscribe. We have shown that users more affected by the blocks were more likely to become new paid subscribers in the post period than users less affected by the blocks, which suggests a causal interpretation. In Table D5 in Appendix D, we also estimate this model for the balanced panel and find coefficients as high as 0.018. We discuss the economic significance of this in section 6.2.

We next ask whether our findings change when fewer sites are blocked. To do so, we examine the blocking of 19 sites in November 2013 and also a single site in May 2012.

November 2013 Blocking of 19 Major Piracy Sites

Recall that because of privacy concerns PanelTrack would only release monthly data aggregated into consumer groups for 2012 and 2013. While using aggregate grouped data is clearly inferior to using individual data, our analysis is able to recover the impact of website blocks on legal and illegal media usage with the appropriate inferential statistics. We rely on the estimator and inference suggested by Donald and Lang (2007), who discuss methods for correcting for common group errors when the treatment is assigned at the group level such as in our data. The estimator they recommend as most efficient in many circumstances is the *between-group estimator*, which is in fact a regression of grouped means against group average outcomes. The approach relies on the fact that each data point is based on the aggregated behavior of a sufficiently large underlying group and thus is measured with greater precision than if each observation were generated by one individual, which is precisely the data for the website blocks in 2012 and 2013 we have at our disposal. Inference for our estimators is given by a t_{G-2} distribution, where G indicates the number of groups.²⁰

¹⁹Specifically, Figure 1 and our placebo tests suggest that the blocks did cause an increase in visits to unblocked piracy sites, but that our difference-in-difference approach likely overestimates the magnitude of this effect. We thus make no claims about the overall size of the effect.

²⁰Note this distribution is more conservative than using a t-distribution with $t_{M_1+M_2+\dots+G-2}$ degrees of freedom that would be used were we to estimate pooled OLS. This point is made in lecture notes by Wooldridge (2007).

Table 6. Impact of 53 Site Block in November 2014 on New Legal Subscriptions Estimated Via Logistic Regression

	(1)	(2)	(3)
<i>Dependent variable:</i>	> 1 Post-Period Legal Subscription Visits	> 2 Post-Period Legal Subscription Visits	> 3 Post-Period Legal Subscription Visits
Treatment Intensity	0.00909** (0.00341)	0.0111** (0.00385)	0.0102* (0.00444)
Constant	-1.096*** (0.0313)	-1.918*** (0.0404)	-2.378*** (0.0484)
N	5759	5759	5759
Log-Likelihood	-3262.513	-2233.499	-1697.341

Notes: Standard errors are shown in parentheses. + $p < 0.10$ * $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$. Model is estimated on a subset of users who made zero pre-period visits to paid subscription sites.

Specifically, we estimate the following model:

$$\ln \text{Visits}_{jt} = \gamma_0 + \gamma_1 \text{month}_t + \gamma_2 \text{TreatmentIntensity}_j \cdot \text{month}_t + \mu_j + \mu_{jt} \quad (4)$$

where all terms are the same as in (1) except that the j subscript now denotes a group of consumers (as opposed to i indexing the individual). Because our data are aggregated across groups, the resulting visits are large enough that we can estimate OLS. However, we log these data because visits are right skewed and because we expect trends across the groups to be comparable on a relative (percent) basis. γ_2 is the coefficient of interest, and as a test of the parallel trends assumption, we ask if γ_2 is 0 for all months before the blocks. We plot all γ_2 coefficients below for the various outcome variables.

In the post period, it appears as if visits to unblocked cyberlocker piracy sites decreased as a result of the blocks while visits to legal subscription sites increased; both of these results are consistent with our results from 2014. It also appears as if VPN visits and visits to unblocked torrent sites increased as a result of the November 2013 blocks.

While the parallel trends assumption holds for cyberlocker visits, for legal subscription, and (almost) for VPN site visits, it fails for unblocked torrent site visits. Heavier users of the blocked sites appeared to increase their usage of unblocked torrent sites more in the pre-period than lighter users. We are not certain why this is the case. It is possible that because the court case which ordered the blocks occurred during October 2013, some users of pirate sites may have had advance knowledge of the blocks and started to rely more other sites. Alternately, some of the unblocked torrent sites may in fact be proxy or mirror sites for the blocked sites. Either way, any results for the effect of the 2013 wave of blocks for visits to unblocked torrent sites must be taken with caution as they

may not be causal, but we can say that we find no evidence of a decrease in usage of unblocked torrent sites caused by the 2013 blocks.

To measure the overall effect of the November 2013 blocks on the outcome variables and to determine statistical significance, we estimate the following model:

$$\ln \text{Visits}_{jt} = \gamma_0 + \gamma_1 \text{Post}_t + \gamma_2 \text{TreatmentIntensity}_j \cdot \text{Post}_t + \gamma_3 \text{PartialTreatment}_t + \gamma_4 \text{Partial}_t \cdot \text{TreatmentIntensity}_j + \mu_j + \epsilon_{jt} \quad (5)$$

Model (5) is similar to (2) except that indexes each group and the outcome variable is logged visits due to our ability to estimate using OLS.

In Table 7, we see from the Post Treatment dummy that all of the outcome variables were decreasing over time, though the decreases are relatively small for VPN sites and legal subscription sites. The coefficients of interest are those on the post \times treatment intensity interaction term. Here, we see an increase usage of unblocked torrent sites (significant at alpha = 0.1). Because we know from the time plot in Figure 2 that this may be an extension of preexisting trends we cannot make a strong claim as to whether these blocks increased usage of unblocked torrent sites, but there is no evidence of a decrease. We do observe a causal decrease in visits to unblocked cyberlocker sites, a causal increase in visits to VPN sites, and a causal increase in visits to legal subscription sites. Thus, our results indicate that, like the blocking of 53 sites in November 2014, the blocking of 19 sites did drive some users to paid legal streaming sites and reduced at least some forms of piracy (cyberlockers). An individual who made one more visit per month to blocked sites during the pre-period increased her monthly visits to legal subscription sites 1.34% more than she would have if not for the blocks.

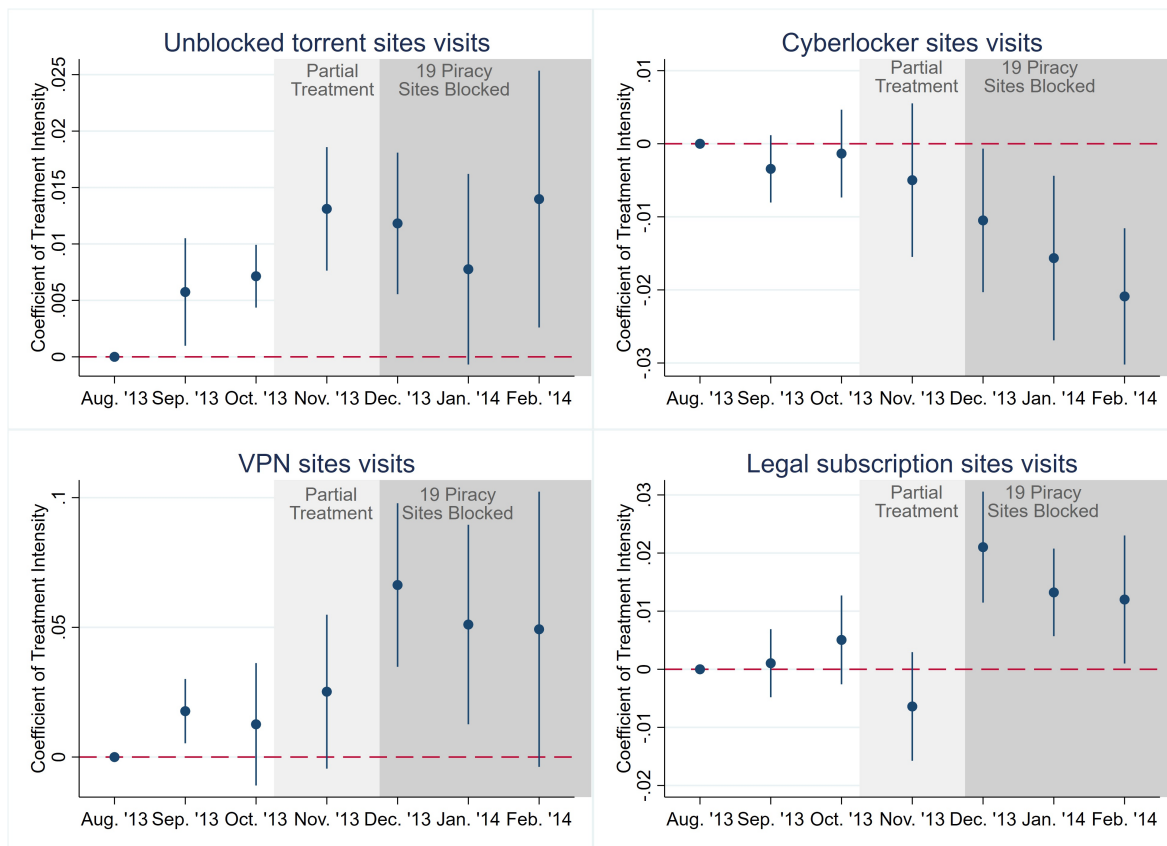


Figure 2. Effect of November 2013 Blocks on Outcomes

Table 7. Estimated Impact of 19 Site Block in November 2013 on User Site Visits

	(1)	(2)	(3)	(4)
<i>Dependent Variable:</i>	Unblocked Torrent Sites	Cyberlockers	VPN Sites	Legal Subscription Sites
Post Treatment	-0.149 (0.0934)	-0.527** (0.121)	-0.0540 (0.418)	-0.0339 (0.0974)
Post × Treatment Intensity	0.00689+ (0.00324)	-0.0141** (0.00305)	0.0454** (0.0140)	0.0134** (0.00359)
Partial Treatment	0.0295 (0.0697)	-0.214 (0.122)	0.456+ (0.225)	-0.00699 (0.0691)
Partial × Treatment Intensity	0.00881** (0.00186)	-0.00339 (0.00343)	0.0151 (0.0109)	-0.00843* (0.00364)
Constant	13.74*** (0.0318)	13.55*** (0.0472)	9.927*** (0.147)	13.78*** (0.0380)
User Group FE?	Y	Y	Y	Y
N	70	70	70	70
User Groups	10	10	10	10
Adjusted R ²	.125	.633	.221	.255

Notes: Standard errors are shown in parentheses and clustered by user group. + $p < 0.10$ * $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

We followed Donald and Lang (2007) in computing p-values when outcome variables are aggregate data from large groups, and believe this to be sufficiently correct for any downward bias in standard errors. However, because our number of clusters is small, we also impute even more conservative p-values using the wild cluster bootstrap approach (Cameron et al. 2008). These p-values on the coefficients of interest can be found in Table D6 in Appendix D (there the coefficient for visits to paid legal subscription has a p-value of 0.08).

Both the 2013 and 2014 waves of blocks involved the blocking of a number of major piracy sites. Next, we ask whether the blocking of one major piracy site—an experiment more akin to those in Poort et al. (2014) and Aguiar et al. (2018)—demonstrates similar outcomes or produces a different set of results.

May 2012 Blocking of The Pirate Bay

The data we obtained from PanelTrack to study the blocking of The Pirate Bay in 2012 are similar to the data from 2013: we observe outcomes by consumer group by month, generated from a balanced panel of consumers observed in all months. Again, PanelTrack sorted consumers into groups based on pre-block usage of the blocked site, in this case, The Pirate Bay.

We estimate model (4) for each of the outcome variables and plot the coefficients of interest for the models in Figure 3.

The 2012 blocking of the Pirate Bay appears to have caused an increase in visits to unblocked torrent sites as well as visits to VPN sites. We observe no clear effect on visits to cyberlockers and while we may see an increase in usage of legal subscription sites in the month after the blocks, it disappears by the second and third months after the blocks. The parallel trends assumption appears to hold for visits to unblocked torrent sites, VPN sites, and (nearly) for unblocked cyberlockers. However, the parallel trends assumption fails for visits to legal subscription sites as treatment intensity appears positively correlated with changes in visits to subscription sites during the pre-period. There is a compelling explanation for this fact: one of the paid subscription sites, Netflix, was introduced to the UK in January 2012, and it quickly became popular due to the fame of the brand. During the initial adoption period, we argue that people who were pirating a lot of content (relative to people who were pirating little) are more likely to have an initial interest in Netflix and therefore subscribe. This would explain the elevated coefficients in March and April. The direction of the preexisting trend would actually suggest that the correlation should have increased in the post period, and instead we see it remain flat

or decrease other than in June 2012. Thus we conclude that blocking The Pirate Bay caused no lasting increase in paid legal consumption and at most a temporary one month increase.

Next, we estimate (5) for each of the outcome variables and present the results below.²¹

In Table 8 we observe a statistically significant increase in usage of other unblocked torrent sites such that a person visiting The Pirate Bay one more time during the pre-period increased her usage of other torrent sites 0.22% more than she would have after the blocks if she had not been using The Pirate Bay. We observe a statistically significant increase in usage of VPN sites, indicating that, after the block, some heavy users of the Pirate Bay turned to using a VPN to circumvent the blocking of the site. However, the economic significance of this may be small as the constant term here is small: visits to VPN sites in the data are relatively low. Finally, the coefficient for paid legal streaming sites is positive but small and statistically insignificant. Typically, this might lead to an inconclusive interpretation: Is the increase positive or 0? From Figure 3 we know that any increase in the post period is driven entirely by the first month, after which it disappears. And we also know that even this first month effect may be the result of a preexisting trend. Thus, like Aguiar et al. (2018) we find no lasting causal effect of the May 2012 blocking of The Pirate Bay on legitimate consumption.

Summary of Empirical Results

In summary, we found that the 2014 blocking of 53 major piracy sites not only decreased visits to the blocked sites but also caused a decrease in usage of other unblocked piracy sites. We observe that it causally increased usage of paid legal streaming sites and may have been associated with an increase in new paid subscriptions. Together, these results imply that supply-side anti-piracy enforcement can be effective in turning users of illegal piracy channels toward paid legal consumption. In November 2013 when 19 major piracy sites were blocked, we do not observe a causal decrease on visits to unblocked torrent sites but we do observe a causal decrease in visits to unblocked cyberlocker sites. We also observe a statistically significant increase in usage of paid legal streaming sites. Finally, consistent with the literature, we found that the May 2012 blocking of The Pirate Bay

²¹ As with the 2013 blocks, we present in Table D7 in Appendix D the same estimates but with standard errors estimated using the wild cluster bootstrap approach. The increase in visits to unblocked torrent sites remains significant with a p-value of 0.036.

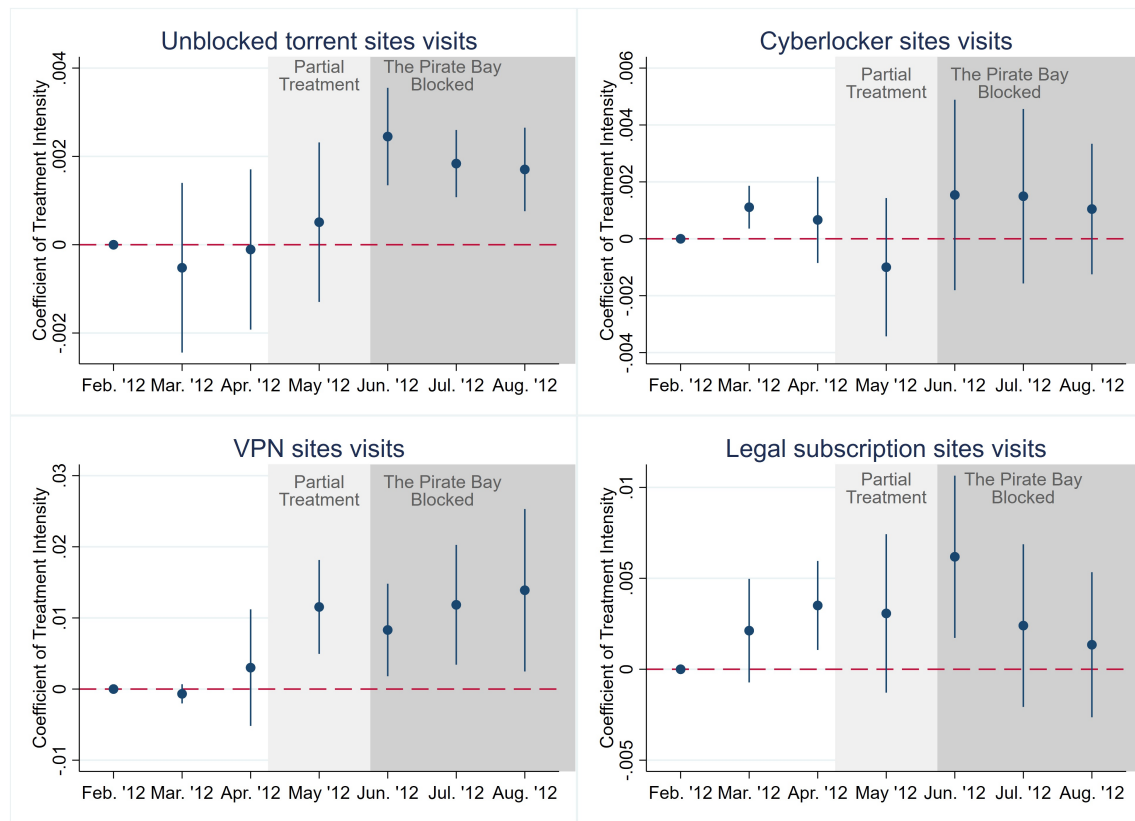


Figure 3. Effect of May 2012 Pirate Bay Block on Outcomes

Table 8. Estimated Impact of The Pirate Bay Block in May 2012 on User Site Visits

	(1)	(2)	(3)	(4)
<i>Dependent Variable:</i>	Unblocked Torrent Sites	Cyberlockers	VPN Sites	Legal Subscription Sites
Post	-0.207** (0.0434)	-0.388+ (0.173)	-0.962+ (0.508)	-0.576+ (0.292)
Post × Treatment Intensity	0.00221*** (0.000363)	0.000769 (0.000935)	0.0106** (0.00261)	0.00143 (0.00148)
Partial Treatment	-0.312** (0.0811)	-0.340+ (0.160)	-1.085* (0.392)	-0.707+ (0.322)
Partial × Treatment Intensity	0.000720 (0.000436)	-0.00159+ (0.000835)	0.0108** (0.00233)	0.00119 (0.00166)
Constant	14.67*** (0.0216)	13.85*** (0.0785)	9.897*** (0.217)	11.90*** (0.141)
User Group FE?	Y	Y	Y	Y
N	70	70	70	70
User Groups	10	10	10	10
Adjusted R ²	0.274	0.288	0.050	0.203

Notes: Standard errors are shown in parentheses and clustered by user group. + $p < 0.10$ * $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

caused users to increase their visits to other unblocked piracy sites and to circumvent the blocks through use of VPNs. We found no increase in usage of paid subscription sites as a result of this block.

Possible Mechanisms

We motivated our research by describing two primary reasons why blocking access to multiple piracy sites might have a different effect on consumer behavior than blocking access to only one site. First, we suggested that the fixed cost involved with switching to a new piracy sites (which involves both search and learning costs) could be higher and affect more individuals when more sites are blocked, thus some individuals might choose to substitute legal consumption for piracy. Second, we considered the possibility of a chilling effect, whereby blocking a large number of piracy sites sends a stronger signal to pirates about the severity of the anti-piracy enforcement regime or increases its salience. The question remains whether our results can distinguish between these two mechanisms.

Recall that in both 2013 and 2014, we found that the waves of blocks not only decreased visits to blocked sites but also caused decreases in visits to at least some other unblocked piracy sites. In 2013 when we had data on the type of piracy sites, we observed that this causal decrease was driven by visits to piracy cyberlockers and did not extend to unblocked torrent sites. If the mechanism driving the effectiveness of blocking multiple sites in 2013 and 2014 were a chilling effect (based on the broken windows theory of crime and the increased perception of enforcement activity), we would expect this to affect illegal behavior at torrent sites and cyberlockers. Because we do not observe a causal decrease in visits to unblocked torrent sites, we infer that a signaling effect about anti-piracy enforcement is less likely to be driving our results.

If that is true, then by elimination we are left with the increased search and learning costs of piracy associated with blocking multiple sites as the mechanism driving our results. This explanation is highly consistent with our results. Some of the piracy sites blocked in 2013 and 2014 were popular piracy link sites, which by definition direct users to content hosted on piracy cyberlockers. If the value of these piracy links sites is that they conveniently reduce the search and learning costs to users for finding content spread across many cyberlockers, then when these link sites are blocked it could cause a decrease in visits to the unblocked cyberlockers at which their links pointed. Because link sites operate by pointing to full video files hosted on cyberlockers and not by pointing to torrent sites (which simply index torrent tracker

files for P2P downloads), we would not expect blocking access to piracy link sites to decrease visits to unblocked torrent sites. This is exactly the pattern that we observe in 2013, and so we believe the most likely explanation for why blocking multiple sites has a greater effect than blocking just one large site is the increase in search and learning costs. In short, we suggest that the drop in piracy cyberlocker usage in 2013 (and 2014) is the result of making the content on those sites harder to find by blocking access to a number of convenient link piracy sites, and this story is consistent with the lack of a drop in usage of torrent sites following the blocks.

Economic Impact of Website Blocking

While the effect of the 2013 and 2014 waves of blocks on legal channels were statistically significant, it is important to ask whether they were economically significant. In 2014, we start with each individual's observed post-treatment visits to paid legal subscription sites. We estimate their counterfactual post-treatment visits to these subscription sites by predicting what they would have been if treatment intensity were zero (our estimate of the counterfactual, or what they would have been had the individual not been affected by the blocks). We aggregate the difference across all individuals between observed visits to legal sites and counterfactual visits to determine the total causal uplift in visits to legal subscription sites and divide this by the total counterfactual visits to get the overall percent increase. If we use the coefficient estimate from the unbalanced panel estimates in Table 5 (0.0104), we find that users of the blocked sites in 2014 increased their usage of legal subscription sites by 7% relative to what they would have done in the absence of the blocks. If we use the coefficient estimate from the balanced panel estimates in Table D1 (0.0169), we find that this increase was 12%. As both the balanced and unbalanced panels have strengths and weaknesses (discussed earlier), we suggest that the effect of the 2014 blocks on the usage of legal sites by treated users was somewhere between 7% and 12%.

Performing the same analysis for the 2013 blocks (but at the group level rather than the individual), we find that on average the blocks caused treated users to increase their visits to paid subscription sites by 8% relative to what they would have done if not for the blocks. Thus both the 2014 wave and 2013 wave appear to have had similar impacts on legal consumption.

It is worth asking if such increases are economically significant, particularly given the low average visits to subscription streaming sites in our data. We consider our 2014 data and recall that a "visit" in our data measures one continuous session at a legal subscription streaming site, and thus might

roughly be equivalent to watching a film or watching one or more episodes of a television show. Because treated individuals averaged 2.37 legal subscription visits per month in the post period, our estimated 7% to 12% causal increase in visits implies 0.155 to 0.253 more legal subscription visits per person per month. In our sample, 26.3% of users were treated (used blocked sites at least once before the blocks), and in 2014 the Office for National Statistics reported that there were 22 million Internet connected households²² in Great Britain. If our sample is representative, then around 5.78 million households were directly affected by the 2014 blocks. Assuming these households responded similarly to our panel, we estimate that the blocks caused an increase of 896 thousand to 1.46 million legal subscription streaming sessions per month in Great Britain (and more than that across the United Kingdom as a whole). More sessions would lead to a higher perceived value by users, which would increase their willingness to pay for the service, although the precise measurement in terms of demand elasticity is beyond the scope of this paper.

Of course, it is natural to enquire as to the actual number of new subscriptions caused by the 2014 blocks. The lowest coefficient on treatment intensity from our logit model on new subscriptions was 0.009 in the unbalanced panel and the highest was .018 in the balanced panel. The average treatment intensity for treated individuals in this sample was 6.47 monthly blocked piracy site visits in the pre-period. Holding all other parameters fixed at their mean, we calculate the logs odds ratio (of becoming a new subscriber) given a treatment intensity of 6.47 versus the log odds ratio at a treatment intensity of zero. We compute the difference between the two and convert it to a difference in likelihood of subscribing to legal streaming sites. This corresponds to a 1.1 (1.5) percentage point average increase in the probability of subscribing for treated individuals in the unbalanced (balanced) panel. In our data, 18.3% of individuals used blocked sites and did not use legal subscription sites in the pre period. If we assume that this is representative of the 22 million Internet households in Great Britain, then roughly 4.03 million households in Great Britain were affected by the blocks but did not have a paid legal subscription prior to the blocks. A 1.1 to 1.5 percentage point increase in probability of subscribing to a service each month implies an expected 44,000 to 60,000 additional subscribers per month. UK Netflix subscriptions alone grew by 1.9 million between 2014

and 2015,²³ and so our implied increase in total monthly subscribers to all legal streaming services is roughly 2.3% to 3.1% of Netflix's growth that year. We interpret these findings as economically meaningful, given that the monthly price of a subscription to, say, Netflix in the United Kingdom is £6.99 to £9.99 per month. Of course, this result does not account for existing subscription customers who would have otherwise left the service but were retained as a result of the blocks, or the beneficial price elasticity effects of causing existing users to view 7% to 12% more content on these sites.

Discussion

While the use of supply-side anti-piracy actions has increased greatly in recent years as a tool in the fight against intellectual property theft, there are relatively few studies that have empirically analyzed their effectiveness in changing user behavior. While the studies that considered the takedown of pirated content found uplifts in legal sales, studies that focused on cutting off or blocking access to content through a dominant channel found no effect on legitimate consumption. By analyzing the blocking of a single major piracy site in the United Kingdom in 2012, we confirm these prior findings: pirates continued to access illegal content by increasing piracy through other sites or by finding ways to circumvent the blocks. But we demonstrate that the effect of supply-side anti-piracy policies are more nuanced when more sites are involved: Our results show that disrupting access to content through a number of the most popular sites causes decreases in overall piracy and increases in usage of paid legal channels. And we find evidence suggesting that the mechanism driving this is that enough sites have to be blocked to sufficiently increase the search and learning costs associated with additional piracy.

One objection to the causal interpretation of our results might be that legal subscription sites could have started advertising their services more heavily around the time of the blocks in 2014 and 2013. However, we believe this interpretation is unlikely to explain our results for two main reasons. First, our difference-in-difference model is able to capture common time trends through the time fixed effects, so this would only be a concern if legal services could somehow target high piracy individuals more than low piracy individuals with such advertisements. Second, we observe a lack of differential preexisting time trends, and so this counter explanation is only relevant if legal services started targeting heavier users of the blocked sites (and not lighter users) with increased

²²Although our PanelTrack data are at the Internet user level, we extrapolate to households for two reasons. First, PanelTrack generally attempts to capture users from unique households in their sample. Second, decisions on whether to subscribe to a legal subscription service and how much to pay generally occur at a household level rather than an individual level, since multiple individuals in a household may use one account.

²³<https://www.statista.com/statistics/324092/number-of-netflix-subscribers-uk/>

advertising, and they did so exactly at the timing of both the 2013 and the 2014 blocks. The discrete jumps in legal consumption following each of these blocks were not followed by continuing upward trends, and so the timing of the correlation between treatment intensity and discrete changes in legal consumption is telling. In short, while no quasi-experimental claims of causality are ever 100% perfect, we suggest that alternative explanations for our results are unlikely.

There are of course several limitations to this study. First, we were only able to study legal consumption of media through paid legal subscription sites. Users may consume media legally in other ways, such as by digital purchase/rental, physical purchase/rental, or legal free ad-supported viewing channels. Because PanelTrack observes clickstream data but not actual e-commerce, we cannot infer *a la carte* purchases or rentals (e.g., people visit a site like Amazon.com for many reasons other than purchasing movies or television).²⁴ Second, because 2% of ISPs (weighted by market share) did not implement the blocks and the ones that did may have only fully implemented the blocks by some time in the post period, our results may underestimate the true effect of website blocking on legal consumption. Third, we only observe three months after each wave of blocks, and thus we do not know how long our measured impacts lasted. Although the effects on legal subscription visits appeared persistent in our data, it remains possible that increases in legal consumption caused by the blocks fade over time as consumers eventually identify and grow to trust alternate piracy sites. Finally, we are not able to fully estimate the social welfare implications of these blocks because we do not know the costs of these blocks and because we have no data on the long-run impact of increased firm profitability on industry output. Future work should focus on these issues to obtain a better understanding of the broader impacts of site blocking and other anti-piracy measures.

Given the accumulated evidence, how should policymakers view supply-side interventions to curb illegal piracy? We consider by analogy the Greek myth of the Hydra, the mythical, multi-headed beast. The Hydra is one of most difficult animals to kill in Greek mythology. Decapitating any single one of its heads only results in several more growing back to replace it, an excellent analogy for our results and those of prior researchers. It is only when a sword is plunged into its heart that it dies. Removing the source of the pirated content stored in cyberlockers and linked to by many other sites is akin to stabbing the Hydra in the heart (and akin to shutting down Megaupload.com); this is effective but may not

always be feasible. Blocking a single site is akin to decapitating only one of the Hydra's heads. The result will only be a more diffuse network of piracy sites, with no curb on pirating activity. Blocking multiple sites at once is akin to decapitating several of the Hydra's heads. With the network of sites significantly disrupted, this could possibly be a mortal wounding. We have shown that users' behavior is sufficiently disrupted and that some increase the use of legal channels, and reduce illegal ones.

Acknowledgments

This research was conducted as part of Carnegie Mellon University's Initiative for Digital Entertainment Analytics (IDEA), which receives unrestricted (gift) funding from the Motion Picture Association of America. Danaher acknowledges support from an NBER Economics of Digitization research grant. This research was conducted independently without any oversight or editorial control. The authors presented an earlier version of this paper at the December 2014 Workshop on Information Systems and Economics and to seminar participants at the University of Arizona's Eller School of Management and thank participants for their helpful feedback. The authors thank Jesse Newby for excellent research assistance. All findings and errors are entirely our own.

References

- Adermon, A., and Liang, C. Y. 2014. "Piracy and Music Sales: The Effect of an Anti-Piracy Law," *Journal of Economics Behavior and Organization* (105), pp. 90-106.
- Aguiar, L., Peukert, C., and Claussen, J. 2018. "Catch Me If You Can: Effectiveness and Consequences of Online Copyright Enforcement," *Information Systems Research* (29:3), pp. 656-678.
- Bai, J., and Waldfogel, J. 2012. "Movie Piracy and Sales Displacement in Two Samples of Chinese Consumers," *Information Economics and Policy* (24:3), pp. 187-96.
- Bhattacharjee, S., Gopal, R., Lertwachara, K., and Marsden, J. 2006. "Impact of Legal Threats on Online Music Sharing Activity: An Analysis of Music Industry Legal Actions," *Journal of Law and Economics* (49:1), pp. 91-114.
- Bounie, D., Bourreau, M., and Waelbroeck, P. 2006. "Piracy and the Demand for Films: Analysis of Piracy Behavior in French Universities," *Review of Economic Research on Copyright Issues* (3:2), pp. 15-27.
- Card, D. 1992. "Do Minimum Wages Reduce Unemployment? A Case Study of California," *Industrial and Labor Relations Review* (46:1), pp. 38-54.
- Cameron, A. C., Gelbach, J. B., and Miller, D. L. 2008. "Bootstrap-Based Improvements for Inference with Clustered Standard Errors," *Review of Economics and Statistics* (90:3), pp. 414-427.
- Cameron, A. C., and Trivedi, P. K. 2005. *Microeconometrics: Methods and Applications*, New York: Cambridge University Press.

²⁴PanelTrack does also track when a user opens the iTunes application on their computer, a common channel for purchasing digital media. However, many things (including plugging in one's devices) cause an app like this to open and so we chose not to purchase these data from PanelTrack.

- Chen, P., and Hitt, L. M. 2002. "Measuring Switching Costs and the Determinants of Customer Retention in Internet-Enabled Businesses: A Study of the Online Brokerage Industry," *Information Systems Research* (13:3), pp. 255-274.
- Corey, N. 2016. "How Website Blocking Is Curbing Digital Piracy Without 'Breaking the Internet,'" Information Technology & Innovation Foundation (<http://www2.itif.org/2016-website-blocking.pdf>).
- Danaher, B., Dhanasobhon, S., Smith, M. D., and Telang, R. 2010. "Converting Pirates Without Cannibalizing Purchasers: The Impact of Digital Distribution on Physical Sales and Internet Piracy," *Marketing Science* (29:6), pp. 1138-1151.
- Danaher, B., and Smith, M. D. 2014. "Gone in 60 Seconds: The Impact of the Megaupload Shutdown on Movie Sales," *International Journal of Industrial Organization* (33), pp. 1-8.
- Danaher, B., Smith, M. D., and Telang, R. 2014. "Piracy and Copyright Enforcement Mechanisms," *Innovation Policy and the Economy* (14), pp. 31-67.
- Danaher, B., Smith, M. D., and Telang, R. 2017. "Copyright Enforcement in the Digital Age: Empirical Evidence and Policy Implications," *Communications of the ACM* (60:2), pp. 68-75.
- Danaher, B., Smith, M. D., Telang, R., and Chen, S. 2014. "The Effect of Graduated Response Anti-Piracy Laws on Music Sales: Evidence from an Event Study in France," *Journal of Industrial Economics* (62:3), pp. 541-553.
- De Vany, A. S., and Walls, W. D. 2007. "Estimating the Effects of Movie Piracy on Box-Office Revenue," *Review of Industrial Organization* (30), pp. 291-301.
- Dey, D., Kim, A., and Lahiri, A. 2018. "Online Piracy and the 'Longer Arm' of Enforcement," *Management Science* (65:3), pp. 955-1453.
- Donald, S. G., and Lang, K. 2007. "Inference with Difference-in-Differences and Other Panel Data," *The Review of Economics and Statistics* (89:2), pp. 221-233.
- Dur, R. and Volland, B. 2019. "Salience of Law Enforcement: A Field Experiment," *Journal of Environmental Economics and Management* (93), pp. 208-20.
- Goldfarb, A. 2006. "State Dependence at Internet Portals," *Journal of Economics & Management Strategy* (15:2), pp. 317-352.
- Goldman, D. 2010. "Music's Lost Decade: Sales Cut in Half," *CNN Money*, February 3 (https://money.cnn.com/2010/02/02/news/companies/napster_music_industry).
- Hausman, J., Hall, B. H., and Griliches, Z. 1984. "Econometric Models for Count Data with an Application to the Patents-R&D Relationship," *Econometrica* (52), pp. 909-938.
- Hennig-Thurau, T., Henning, V., and Sattler, H. 2007. "Consumer File Sharing of Motion Pictures," *Journal of Marketing* (71), pp. 1-18.
- Herz, B., and Kiljanski, K. 2018. "Movie Piracy and Displaced Sales in Europe: Evidence from Six Countries," *Information Economics and Policy* (43), pp. 12-22.
- Kelling, G. L., and Wilson, J. Q. 1982. "Broken Windows," *Atlantic Monthly* (249:3), pp. 29-38.
- Ma, L., Montgomery, A., Smith, M. D., and Singh, P. 2014. "The Effect of Pre-Release Movie Piracy on Box Office Revenue," *Information Systems Research* (25:3), pp. 590-603.
- McKenzie, J. 2017. "Graduated Response Policies to Digital Piracy: Do They Increase Box Office Revenues of Movies?," *Information Economics and Policy* (38), pp. 1-36.
- McKenzie, J., and Walls, W. D. 2016. "File Sharing and Film Revenues: Estimates of Sales Displacement at the Box Office," *B. E. Journal of Economic Analysis and Policy* (16:1), pp. 25-57.
- Peukert C., Claussen J, and Kretschmer, T. 2017. "Piracy and Movie Revenues: Evidence from Megaupload," *International Journal of Industrial Organization* (52), pp. 188-215.
- Poort, J., Leenheer, J, van der Ham, J., and Dumitru, C. 2014. "Baywatch: Two Approaches to Measure the Effects of Blocking Access to The Pirate Bay," *Telecommunications Policy* (38:1), pp. 383-392.
- Reimers, I. 2016. "Can Private Copyright Protection Be Effective? Evidence from Book Publishing," *Journal of Law and Economics* (59), pp. 411-440.
- Rob, R., and Waldfogel, J. 2007. "Piracy on the Silver Screen," *Journal of Industrial Economics* (55:3), pp. 379-393.
- Sivan, L., Smith, M. D., and Telang, R. 2019. "Do Search Engines Influence Media Piracy? Evidence from a Randomized Field Study," *MIS Quarterly* (43:4), pp. 1143-1154.
- Smith, M., and Telang, R. 2009. "Competing with Free: The Impact of Movie Broadcasts on DVD Sales and Internet Piracy," *MIS Quarterly* (33:2), pp. 312-338.
- Watters, P. A., Layton, R., and Dazeley, R. 2011. "How Much Material on Bittorrent Is Infringing Content? A Case Study," *Information Security Technical Report* (16:2), pp.79-87.
- Wooldridge, J. 2007. "What's New in Econometrics? Lecture 10 Difference-in-Differences Estimation," NBER Summer Institute (http://www.nber.org/WNE/Slides7-31-07/slides_10_diffindiffs.pdf).
- Zellner, A. 1962. "An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias," *Journal of the American Statistical Association* (57:298), pp. 348-368.
- Zentner, A. 2012. "Measuring the Impact of File Sharing on the Movie Industry: An Empirical Analysis Using a Panel of Countries," Working Paper, University of Texas at Dallas, Dallas, TX.

About the Authors

Brett Danaher is an assistant professor of Economics and Management Science at Chapman University. His research interests center around the application of data analytics to the entertainment industries as well as questions of copyright policy in the digital era. He regularly consults for firms in the entertainment industry, helping them to apply new methods in data analytics to optimize their businesses and drive consumer demand for their content. His classes include quantitative methods, industrial organization, and a course on digital change in the entertainment industries. He received his Ph.D. in Applied Economics from the University of Pennsylvania and his B.S. in Economics from Haverford College.

Jonathan Hersh is an assistant professor of Economics and Management Science at the Argyros School of Business at Chapman University. His research interest include the economics of infor-

mation systems and digitization, management of information systems, and the application of predictive models from machine learning and artificial intelligence on business and management processes. He frequently consults with businesses and international organizations on how to build predictive models to augment data-driven decision making. His work often focuses on using machine learning to fill “data gaps” that inform management processes, some of which include using AI applied to satellite imagery. Jonathan teaches courses in machine learning and data science for managers to undergraduates and MBAs. His classes focus on creative applications of AI and machine learning, in addition to hard coding and technical skills. Previously he was a lecturer at MIT and Wellesley College. He received his Ph.D. from Boston University.

Michael D. Smith is the J. Erik Jonsson Chaired Professor of Information Technology and Marketing and co-director of the Initiative for Digital Entertainment Analytics at Carnegie Mellon University’s Heinz College. His research specializes in entertain-

ment analytics, marketing, and management and he is a co-author of the book *Streaming, Sharing, Stealing: Big Data and the Future of Entertainment* (MIT Press, 2016). He received a Bachelor of Science in Electrical Engineering (summa cum laude) and a Master of Science in Telecommunications Science from the University of Maryland, and a Ph.D. in Management Science from the Sloan School of Management at MIT.

Rahul Telang is a professor of Information Systems and co-director of the Initiative for Digital Entertainment Analytics at the Heinz College, Carnegie Mellon University. His major research focus is on Economics of Digitization where he studies how content digitization is affecting firms, users and our policy. His work has appeared in top scientific journals and won numerous awards. He is co-author of the book, *Streaming, Sharing, Stealing: Big Data and the Future of Entertainment* (MIT Press, 2016). He received his Ph.D. in Information Systems from Carnegie Mellon University’s Graduate School of Industrial Administration.

Appendix A

List of Blocked Sites in Each Wave

2012 Block Wave—The Pirate Bay Site Only

thepiratebay.se

2013 Block Wave—19 Unique Video Piracy Sites

In 2013, the following 19 sites were blocked. A number of known mirror sites with similar domains (different suffixes) were blocked along with these. For example, while torrentz.eu was ordered blocked, ISP’s were also ordered to block sites such as torrentz.net when it was verified that they contained the same content.

1337x.org	rapidlibrary.com	torrentz.eu
bitsnoop.com	solarmovie.us	torrentz.eu
Extratorrent.cc	torrentcrazy.com	tubeplus.me
filecrop.com	torrentdownloads.me	vodly.com
filestube.com	torrenthound.com	watchfreemovies.com
monova.org	torrentreactor.net	yify-torrents.com
primewire.net		

2014 Block Wave—53 Unique Video Piracy Sites Plus Known Mirrors

nowtorrents.com	btloft.com	iwannawatch.to
torrentdb.li	picktorrent.com	warez-bb.org
watchseries.to	seedpeer.me	icefilms.info
heroturko.me	torlock.com	Tehparadox.com
torrentbytes.net	torrentbit.net	scnsr.me
seventorrents.org	torrentdownload.ws	rapidmoviez.com

tormovies.org	torrentexpress.net	isohunt.to
bts.to	torrentfunk.com	torrentz.pro
limetorrents.com	torrentproject.com	torrentbutler.eu
torrentus.eu	torrentroom.com	iptorrents.com
vitorent.org	torrents.net	sumotorrent.sx
movie25.cm	torrentz.cd	torrentday.com
iwatchonline.to	torrentzap.com	torrenting.com
losmovies.com	watchseries.lt	bitsoup.me
torrents.fm	stream-tv.me	yourbittorrent.com
bittorrent.am	watchserieshd.eu	demonoid.ph
btdigg.org	cucirca.eu	torrent.cd
vertor.eu	rarbg.com	

Appendix B

Creation of Illegal and Legal Site Lists

While the list of blocked sites in each wave were publicly available based on court orders and summarized (with citations) on Wikipedia,²⁵ an important question is how we created the lists of legal sites and other unblocked piracy sites that we provided to PanelTrack in order for them to provide clickstream visits.

The list of legal sites for each wave of blocks was relatively easy to put together based on a combination of Internet research for available legal services as well as conversations with film and television industry contacts.

However, the list of piracy websites was more difficult since such sites do not necessarily advertise themselves. It is important that we capture the majority of piracy activity in this list or else it could be that pirates thwarted from the blocked sites turn to other unblocked sites and we would not observe it. As such, we went through a multi-step process to determine the set of unblocked piracy sites available in the UK during each wave of blocks.

1. We collected lists of potential piracy sites from various sources, including
 - The City of London Police's "infringing website" list, available at <https://www.cityoflondon.police.uk/advice-and-support/fraud-and-economic-crime/pipcu/Pages/Operation-creative.aspx>
 - The Google Transparency Report provides lists of websites with removal requests due to copyright infringement, available at <https://www.cityoflondon.police.uk/advice-and-support/fraud-and-economic-crime/pipcu/Pages/Operation-creative.aspx>
 - The MPAA's "Online Notorious Market Report", provided to us by the MPAA.
2. We then also asked PanelTrack for their list of piracy sites. PanelTrack categorizes many of the sites that show up in their clickstream data and one of the categories is piracy.
3. After merging #1 and #2, we connected to a UK VPN (in order to appear to be accessing websites from the UK) and attempted to connect to each site to determine if it truly was a site mostly dedicated to piracy, and for 2012 and 2013 to determine whether it was a torrent/P2P site or a cyberlocker/link site. We removed any sites that were not piracy sites, or that were clearly dedicated only to music, anime, adult, games, or eBook content since the blocked sites and legal sites that we studied were largely dedicated to standard television and film content.
4. Finally, we provided the resulting lists to PanelTrack (for each wave, this list was hundreds of sites) who then confirmed based on their data that most of the piracy visits were concentrated within a small number of these sites. We therefore believe it is unlikely that we missed major relevant piracy sites.

²⁵https://en.wikipedia.org/wiki/List_of_websites_blocked_in_the_United_Kingdom#Court_ordered_implementations_targeting_copyright_and_trademark_infringement

Appendix C

Explanation of Consumer Group Formations in 2012 and 2013

As described in the data section, our datasets for 2012 and 2013 include aggregated observations for groups of consumers (each month) rather than individual level data. While this is not ideal, this was mandated by PanelTrack due to issues of privacy and the topic of our research (piracy).

If PanelTrack had randomly assigned consumers in their panel into groups, we would expect very little variation in treatment intensity (pre-blocked visits to blocked sites) across group due to the central limit theorem. Thus, we asked that PanelTrack ensure variation in treatment intensity by bucketing consumers into bins based on their pre-block usage of blocked sites in the three months before the blocks. Instead, PanelTrack bucketed consumers based on their visits to blocked sites in the first month of the study (for example, for the 2012 data, consumers were placed into groups based on February 2012 visits to thepiratebay.se). We consider average visits to blocked sites in the three months before the blocks to be a more accurate representation of a consumer's pre-block behavior than just one month's worth of visits, and so we use the former as our measure of treatment intensity for each group.

One consequence of this is that there is no group with 0 treatment intensity, because some consumers who didn't make visits to the blocked sites in the first month of the panel still made some in the second and third months, leading to a treatment intensity greater than 0. We note that even were there a group with 0 treatment intensity, this group would not truly be a pure "control" group as it remains possible that some individuals who did not visit blocked sites before the blocks would have attempted to afterward, and thus would have been treated by the blocks. Pre-block usage of blocked sites is merely a measure of the "bite" of the treatment on each group, and a zero-value group is unnecessary for identification in this generalized form of the difference-in-differences model.

Appendix D

Supporting Tables and Figures

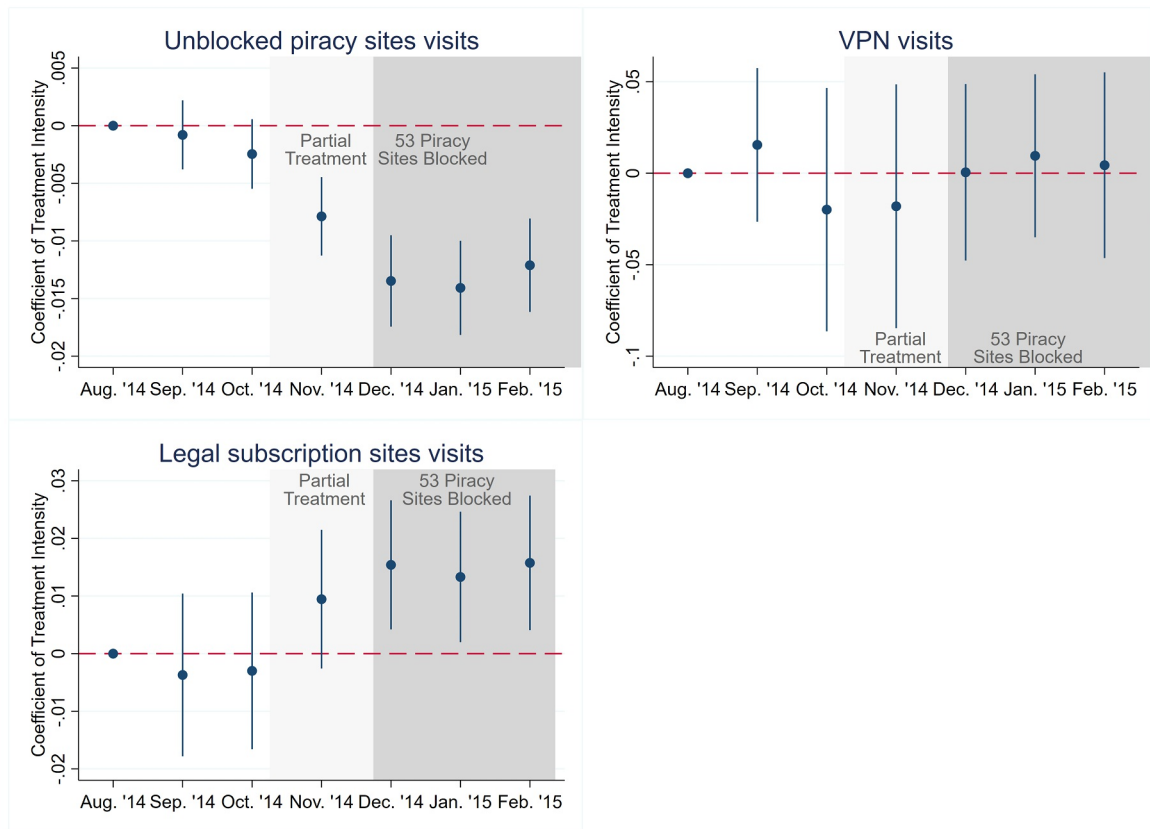


Figure A1. 2014 Blocks—Estimates of Model (1) Using Balanced Panel

Table D1. 2014 Blocks—Negative Binomial Estimates of Model (2) Using Balanced Panel

	(1)	(2)	(3)
<i>Dependent Variable:</i>	Unblocked Piracy Sites	VPN Sites	Legal Subscription Sites
Post Treatment	-0.267*** (0.0201)	-0.0452 (0.140)	0.155*** (0.0338)
Partial Treatment	0.131*** (0.0288)	-0.0533 (0.220)	-0.0902+ (0.0470)
Partial × Treatment Intensity	0.00533** (0.00171)	-0.0226 (0.0307)	-0.00536 (0.00423)
Treatment Intensity	0.0137*** (0.00134)	-0.0477+ (0.0253)	-0.0247*** (0.00363)
Post × Treatment Intensity	-0.0121*** (0.00121)	0.00197 (0.0137)	0.0169*** (0.00350)
Constant	0.487*** (0.0239)	-0.0678 (0.193)	-0.310*** (0.0378)
Individual FE?	Y	Y	Y
N	9478	826	7133
Individuals	1354	118	1019
Log-Likelihood	-25326.04	-571.30	-11136.01

Notes: Standard errors are shown in parentheses and clustered by user. + $p < 0.10$ * $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

Table D2. 2014 Blocks—Seemingly Unrelated Regression Estimates of Model (2) With Log(Visits + 1) as Outcome, Using Balanced Panel

	(1)	(2)	(3)
<i>Dependent Variable:</i>	Unblocked Piracy Sites	VPN Sites	Legal Subscription Sites
Post Treatment	-0.22137*** (0.01594)	-0.0012 (0.00274)	0.05771*** (0.01310)
Partial Treatment	0.1056*** (0.02255)	0.00055 (0.00387)	-0.03524+ (0.01852)
Partial Treatment × Treatment Intensity	0.00894*** (0.00178)	-0.0003 (.000307)	-0.00231 (0.00147)
Post × Treatment Intensity	-0.01881*** (0.00126)	0.00023 (0.00022)	0.00527*** (0.00104)
Individual FEs?	Y	Y	Y
N	10,661	10,661	10,661
Individuals	1523	1523	1523
R ²	0.7719	0.8048	0.74

Notes: Standard errors are shown in parentheses and clustered by user. + $p < 0.10$ * $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

Table D3. 2014 Blocks—OLS Estimates of Model (2) with Log(Visits + 1) as Outcome, Using Balanced Panel

	(1)	(2)	(3)
<i>Dependent Variable:</i>	Unblocked Piracy Sites	VPN Sites	Legal Subscription Sites
Post Treatment	-0.222*** (0.0172)	-0.00112 (0.00297)	0.0580*** (0.0141)
Partial Treatment	0.108*** (0.0243)	0.000281 (0.00420)	-0.0336+ (0.0200)
Partial Treatment × Treatment Intensity	0.00897*** (0.00192)	-0.000296 (0.000332)	-0.00227 (0.00158)
Post × Treatment Intensity	-0.0189*** (0.00136)	0.000175 (0.000234)	0.00524*** (0.00112)
Individual FEs?	Y	Y	Y
N	10,661	10,661	10,661
Individuals	1523	1523	1523
R ²	0.7719	0.8048	0.74

Notes: Standard errors are shown in parentheses and clustered by user. + $p < 0.10$ * $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

Table D4. 2014 Blocks—Poisson Estimates of Model (2)

	(1)	(2)	(3)
<i>Dependent Variable:</i>	Unblocked Piracy Sites	VPN Sites	Legal Subscription Sites
Post Treatment	-0.262*** (0.00390)	-0.284*** (0.0314)	-0.0213** (0.00678)
Partial Treatment	-0.153*** (0.00425)	-0.0995* (0.0414)	-0.0509*** (0.00868)
Partial Treatment × Treatment Intensity	-0.00208*** (0.0000563)	-0.0296** (0.0111)	0.00297** (0.00107)
Post × Treatment Intensity	-0.00766*** (0.000146)	0.0187*** (0.00432)	0.00817*** (0.000795)
Individual FEs?	Y	Y	Y
N	46733	2546	29685
Individuals	11847	556	7095
Log-likelihood	-178247.95	-2846.39	-68904.83

Notes: Standard errors are shown in parentheses and clustered by user. + $p < 0.10$ * $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

Table D5. Impact of 53 Site Block in November 2014 on New Legal Subscriptions Estimated Via Logistic Regression, Using Balanced Panel

	(1)	(2)	(3)
<i>Dependent Variable:</i>	>1 Post-Period Legal Subscription Visits	>2 Post-Period Legal Subscription Visits	>3 Post-Period Legal Subscription Visits
Treatment intensity	0.0114* (0.00559)	0.0170** (0.00582)	0.0186** (0.00611)
Constant	-0.542*** (0.0783)	-1.374*** (0.0930)	-1.779*** (0.106)
N	826	826	826
Log-Likelihood	-547.46	-430.27	-358.71

Notes: Standard errors are shown in parentheses. + $p < 0.10$ * $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

Table D6. 2013 Estimates of Model (5) Using Wild Cluster Bootstrap Standard Errors

	(1)	(2)	(3)	(4)
<i>Dependent Variable:</i>	Unblocked Torrent Sites	Cyberlockers	VPN Sites	Legal Subscription Sites
Post × Treatment Intensity	0.00689 [0.205]	-0.0141 [0.188]	0.0454 [0.132]	0.0134+ [0.082]

Notes: All other variables from model (4) were estimated but suppressed. Table contains only the estimates for the coefficient of interest. P-values computed using wild cluster bootstrapping of standard errors are displayed in brackets below the estimates.

Table D7. 2012 Estimates of Model (5) Using Wild Cluster Bootstrap Standard Errors

	(1)	(2)	(3)	(4)
<i>Dependent Variable:</i>	Unblocked Torrent Sites	Cyberlockers	VPN Sites	Legal Subscription Sites
Post × Treatment Intensity	0.00221* [0.036]	0.000769 [0.469]	0.0106 [0.193]	0.00143 [0.355]

Notes: All other variables from model (4) were estimated but suppressed. Table contains only the estimates for the coefficient of interest. P-values computed using wild cluster bootstrapping of standard errors are displayed in brackets below the estimates.

Table D8. 2014 Negative Binomial Estimates of Model (2) Placebo Tests

	(1)	(2)	(3)	(4)	(5)	(6)
<i>Dependent Variable:</i>	Unblocked Piracy Sites, Placebo Treatment Month 2	Unblocked Piracy Sites, Placebo Treatment Month 3	VPN Sites, Placebo Treatment Month 2	VPN Sites, Placebo Treatment Month 3	Legal Subscription Sites, Placebo Treatment Month 2	Legal Subscription Sites, Placebo Treatment Month 3
Post Placebo	-0.120*** (0.0129)	0.0408** (0.0132)	-0.156+ (0.0924)	-0.309** (0.106)	-0.172*** (0.0223)	-0.0756** (0.0234)
Placebo Treatment Intensity	0.00597*** (0.000522)	0.00514*** (0.000740)	-0.0175 (0.0202)	-0.0308 (0.0262)	-0.00426 (0.00338)	-0.00780* (0.00363)
Post Placebo × Treatment Intensity	-0.00656*** (0.000833)	-0.00824*** (0.00104)	-0.00587 (0.00820)	-0.0126 (0.0130)	-0.00224 (0.00245)	-0.00237 (0.00284)
Constant	0.541*** (0.0191)	0.440*** (0.0168)	0.580** (0.192)	0.546** (0.180)	0.157*** (0.0334)	0.0805** (0.0295)
Individual FEs?	Y	Y	Y	Y	Y	Y
N	19396	23078	790	904	9790	11334
Individuals	7473	9314	292	349	3757	4529
Likelihood	-31796.89	-36094.77	-581.11	-627.1	-11316.28	-12665.10

In column (1), (2), and (3) we drop months 4 through 7 and we falsely assume that the blocks happened just before month 2. In columns (4), (5), and (6) we again drop months 4 through 7 but we falsely assume that the blocked happened just before month 3.

Notes: Standard errors are shown in parentheses and clustered by user. + $p < 0.10$ * $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

Copyright of MIS Quarterly is the property of MIS Quarterly and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.