TECHNOLOGY POLICY INSTITUTE

Do Pirated Video Streams Crowd Out Non-Pirated Video Streams? Evidence from Online Activity

Sarah Oh, Scott Wallsten and Nathaniel Lovin

January 2020

409 12th Street SW ● Suite 700 ● Washington, DC 20024
Tel: (202) 828-4405 ● Email: info@techpolicyinstitute.org ● www.techpolicyinstitute.org

Do Pirated Video Streams Crowd Out Non-Pirated Video Streams? Evidence from Online Activity

January 22, 2020

Sarah Oh,* Scott Wallsten,‡ & Nathaniel Lovin§

Does watching more pirated streaming video mean spending less time watching non-pirated streaming video? This study measures whether, and how much, time spent watching pirated video crowds out time spent on streaming video apps. While prior studies have estimated the impact of piracy on sales revenues, our study measures the impact of piracy on time spent on free and paid streaming apps. We combine big data tools with standard econometric techniques, including a two-stage least squares model, to analyze 5.25 terabytes of online activity data from 19,764 American households and their 468,612 devices from 2016 to 2017. The analysis suggests that every minute spent engaged with pirated video sites crowded out about 3.5 minutes of time spent streaming video. Because pirated video files are generally more compressed than non-pirated video files and because they are frequently downloaded as entire files rather than streamed, as with non-pirate sites like Netflix and Amazon, we conclude that our results exhibit closer to a 1-to-1 crowding out effect of piracy on over-the-top streaming video services.

Keywords: Piracy, digital content distribution, intellectual property

JEL Classification: D69, L1, L11, L8, L82, M31, O30

* Senior Fellow, Technology Policy Institute. The authors thank Michael Smith and participants of the 2018 TPI Aspen Forum for comments. Any mistakes are our own.
‡ President and Senior Fellow, Technology Policy Institute.
§ Research Associate, Technology Policy Institute.

Introduction

Americans watch more streaming video each year, with 64 million U.S. households watching streaming content in 2019.[1] Households that watch streaming video spend, on average, nearly three hours per day watching over-the-top (OTT) video streaming services.[2]

While the big four streaming apps–Netflix, YouTube, Hulu, and Amazon–account for nearly 80 percent of the streaming video market, dozens of other subscription services–such as Disney+, Peacock, HBO Max, Apple TV+ and smaller niche streaming services–offer direct-to-consumer access to video content. Nearly 46 percent of American households today subscribe to streaming video services, up from 20 percent in 2014.[3]

Yet, even with an abundance of video content with high-definition quality and low monthly subscription cost, suppliers of pirated video streams continue to distribute pirated copies of premium movies, television shows, and music to American households.[4]

A key policy question is whether pirated video streams complement or substitute for non-pirated video.[5] Our research, based on a very large dataset of household Internet use, concludes that pirated streaming crowds out non-pirated streaming. We find that every minute that is spent downloading or streaming pirated video crowds out about 3.5 minutes of non-pirated streaming video. Taking into account different compression rates and delivery via file download or streaming, we conclude that time spent watching pirated video displaces nearly the same amount of time that would have been spent watching non-pirated video.

Previous Research on the Effects of Pirated Video

As many as 7.3 percent of American households accessed pirated video streams in 2018.[6] What is the effect of this piracy on consumer behavior? Does time spent watching pirated content displace (crowd out) time otherwise spent watching non-pirated content?

In principle, consumption of pirated content could either increase or decrease consumption of non-pirated content. Piracy could decrease consumption of non-pirated content if consumers

---

[1] https://www.comscore.com/Insights/Presentations-and-Whitepapers/2019/State-of-OTT, https://www.adweek.com/tv-video/the-number-of-ott-only-u-s-homes-has-tripled-over-the-last-5-years/, http://www.thevab.com/wp-content/uploads/2018/03/OTT-Ecosystem-Overview-Final.pdf.

[2] *Id.* (ComScore State of OTT Report, June 2019).

[3] https://deadline.com/2019/10/half-of-broadband-homes-have-multiple-streaming-subscriptions-parks-associates-1202750361/ (citing a study tracking the growth of over 235 subscription streaming services).

[4] https://www.cnbc.com/2019/10/20/netflix-and-hbo-shows-are-getting-pirated-on-teatv-and-other-sites.html.

[5] A second key policy question is what effect piracy has on content creation and innovation. Our data cannot address this question.

[6] https://www.sandvine.com/blog/global-internet-phenomena-spotlight-video-piracy-in-north-america, https://www.sandvine.com/inthenews/pirate-tv-services-are-taking-a-bite-out-of-cable-company-revenue-ars-technica, https://www.sandvine.com/hubfs/downloads/archive/whitepaper-video-and-television-piracy-ecosystem-and-impact.pdf.

watch any given piece of content via a pirate source rather than a source offering the same content via licensed means. This displacement is a crowding out effect. On the other hand, piracy could increase consumption of non-pirated content by serving as a way for consumers to "sample" content before buying, or via "indirect appropriability" or "network effects" that increase the value of the pirated content by increasing its popularity and visibility.

Capturing the true effect of piracy is difficult. Even assuming one accurately measures consumption of pirated and non-pirated content, a key problem for empirical analysis is endogeneity.[7]

For example, a negative correlation between time spent watching pirated video and time spent watching non-pirated video is consistent with the crowding out hypothesis. But if, for example, people with less disposable income are more likely to view pirated content and would not otherwise purchase non-pirated content, then, without controlling for income, such a negative correlation would not necessarily support the crowding out hypothesis. Instead, a negative correlation would be explained by distributional and pricing effects, rather than crowding out.

Similarly, a positive correlation between the consumption of pirated and non-pirated content could imply crowding in, consistent with the "sampling" or "network effects" hypotheses. But if people who consume a lot of pirated content consume a lot of all types of content, the positive correlation might be identifying content-hungry people, not crowding in. Additionally, if popular movies are more likely to be pirated, a positive correlation might identify how popular a particular piece of content is, not the causal effect of piracy.

Thus, simple correlations between consumption of pirated and non-pirated content may yield spurious results. Scholars have attempted to deal with the endogeneity problem in empirical studies of piracy. Causal effects have been measured through natural experiments with treatment and control groups, product-level analysis, city or country-level data, individual-level survey data, and instrumental variables.[8] Instrumental variables and novel datasets have included German school vacations, broadband penetration rates, and more.

In general, economists have found evidence of displacement of non-pirated content by pirated content in software, movies, music, and television. Some studies, however, have found conflicting evidence on the effect of piracy on sales. Oberholzer-Gee and Strumpf (2007) used data on broadband access before and after a German secondary school vacation in order to estimate piracy's effects on music album sales.[9] Based on data collected in 2002, they found little impact of file sharing on music sales.

In their results, German school vacations did not affect U.S. sales, thus leading to their conclusion that file sharing had little impact on music sales. The model relied on the assumption that during German school vacations students presumably had more time to share files online,

---

[7] Brett Danaher, Michael D. Smith, and Rahul Telang, "Piracy and Copyright Enforcement Mechanisms" (NBER Working Paper 19150, June 2013), 4, https://www.nber.org/papers/w19150.
[8] Danaher, Smith, and Telang, 6–7.
[9] Felix Oberholzer-Gee and Koleman Strumpf, "The Effect of File Sharing on Record Sales: An Empirical Analysis," *Journal of Political Economy* 115, no. 1 (2007): 1–42.

which should have shown more piracy from Germany and thus lower U.S. sales. Liebowitz (2017), however, explained that their use of German school vacations as an instrument for piracy was problematic, generating an ongoing debate between the authors.[10] The main critique of the Oberholzer-Gee and Strumpf paper was that German school vacations were not an effective instrument for a variety of reasons.

In the last 20 years, scholars have sought to establish a more robust empirical literature to investigate the crowding out hypothesis. These studies found crowding out of sales of DVDs, music, and motion picture ticket sales. Zentner (2009) used country-level data on broadband penetration as an instrumental variable to find 58 to 92 percent decline in sales from piracy across 49 countries from 1997 to 2008. Zentner (2012) used country-level data on broadband penetration to measure effects of piracy on motion picture sales from 2001 to 2008, with a before and after comparison around 2003, the year that BitTorrent was introduced. He found a strong negative relationship between increased broadband penetration and DVD sales, but no statistical relationship of broadband on movie sales.[11] Ma, Montgomery, Singh, and Smith (2014) exploited a time difference in pre-release pirated copies of movies in order to measure the effects of piracy on non-pirated movie box office releases. Using data from 2006 to 2008, they found a 19.1 percent decrease in revenue from pre-release piracy compared to post-release piracy on box office revenues. Piracy's effects in software markets have been studied as well. Athey and Stern (2013) exploited differences between countries to identify determinants of software piracy of Windows 7. They found a negative relationship between piracy and GDP per capita, with controls for institutional quality, broadband access, and business environment in poor and wealthy countries.[12]

Studies of Internet time use behavior also inform the piracy literature. Because the amount of time in a day is fixed, the opportunity cost of doing one activity is not doing another activity, multi-tasking notwithstanding. Time spent watching pirated streams must mean less time doing something else. Likely, it means less time spent watching other sources of video, but it could also mean less time spent doing other things like online web browsing or offline activities.

Crowding out of offline activity by online activity has been found in survey data. Wallsten (2015) found that online activity crowds out offline activities in American households, where online leisure time displaced time formerly spent working, sleeping, or engaging in educational activity.[13] Using person-level data from the American Time Use Survey from 2003 to 2011, Wallsten measured net benefits and marginal gains from displacement of online and offline activity in the context of measuring the economic surplus generated from the Internet. The study incorporated fixed effects for American household demographics, serving as a critical control in

[10] Stan J. Liebowitz, "Responding to Oberholzer-Gee & Strumpf's Attempted Defense of Their Piracy Paper," *SSRN Electronic Journal*, 2017, https://doi.org/10.2139/ssrn.2887122.
[11] Alejandro Zentner, "Measuring the Impact of File Sharing on the Movie Industry: An Empirical Analysis Using a Panel of Countries," University of Texas at Dallas, Working Paper, http://dx.doi.org/10.2139/ssrn.1792615.
[12] Susan Athey and Scott Stern, "The Nature and Incidence of Software Piracy: Evidence from Windows," National Bureau of Economic Research, Working Paper No. 19755 (December 2013), http://www.nber.org/papers/w19755.
[13] Scott Wallsten, "What Are We Not Doing When We Are Online?" NBER Economics of Digitization Group, in *Economic Analysis of the Digital Economy*, eds. Avi Goldfarb, Shane M. Greenstein, and Catherine E. Tucker (University of Chicago Press, April 2015): 55–82, http://www.nber.org/chapters/c13001.

empirical studies of online activity. An earlier study also documented the importance of fixed effects for household demographics for studies of household Internet use. Goldfarb and Prince (2008) found from a survey of 18,439 Americans that high-income, educated people were more likely to adopt Internet use. But, after conditioning on adoption rates, they observed that low-income, less-educated people were likely to spend more time online than others.[14]

Other studies have measured crowding out effects of online activity. Liebowitz and Zentner (2012) used Nielsen television data and broadband penetration rates to find that Internet use reduces television viewing by 11 percent with extensive analysis of different demographic categories.[15] Chen, Hu, and Smith (2018) considered the effects of eBook sales on print book sales. They exploited an exogenous shock in a release delay of Kindle eBooks in 2010 to measure effects on print sales, finding no effect of cannibalization of eBook sales on print book sales.

Our study builds on this literature on piracy and the digital economy. As far as we know, this is a first study of its kind that exploits device-level data using a weighted panel of American homes to investigate the effects of pirated video on non-pirated streaming video services. With over one trillion observations of raw Internet traffic data from 19,764 American households and their 468,612 devices, we use various econometric techniques to measure whether pirated streams displaced time spent watching Netflix, Hulu, YouTube, and Amazon Video.

Data

Data collection and analysis are possible today at a scale previously unavailable to economists who studied piracy in years past. We cleaned and processed 5.25 terabytes of raw data of online activity from ComScore's Total Home Panel, which is a population-weighted database containing enormous detail on Internet traffic flowing into and out of American homes. This section describes the data and how we used it to measure non-pirated and pirated video streams.

The Sample

The ComScore Total Home Panel consists of households who choose to participate in the company's data program.[16] These households provide demographic information, along with other information such as their Internet service provider and any cable or satellite television

---

[14] Avi Goldfarb and Jeff Prince, "Internet Adoption and Usage Patterns Are Different: Implications for the Digital Divide," *Information Economics and Policy* 20 (2008): 2–15.

[15] Stan J. Liebowitz and Alejandro Zentner, "Clash of the Titans: Does Internet Use Reduce Television Viewing?" *The Review of Economics and Statistics* 94, no. 1 (2012): 234–45.

[16] ComScore Media Metrix Methodology (2016). Over half of the households who were asked to participate in the program consented to install meters, and approximately a quarter of those households were qualified to be included in the panel. (*Id*.). Each household is assigned a weight using an iterative sequential stratification technique. (Metrix Methodology, at 27). The household weight represents the demographic of that household out of approximately 90 million Internet households. This weight is adjusted in a monthly enumeration survey with methods to reduce data volatility each reporting period (*Id*.).

subscriptions. Via an electronic meter and proprietary software, ComScore then collects raw data from each device in the household that connects to the Internet via the home Internet connection, including each device's brand name, family name, model name, manufacturer, and operating system.

Raw data includes details on each data packet sent to and from the web and each device in participating households. This data allows us to observe the full activity of each device with timestamps to the hundredth of a millisecond. Our time frame spans eight alternating months between September 2016 and November 2017. The panel includes 19,764 unique households and 468,612 unique devices. Over one trillion observations of online activity are logged, amounting to 5.25 terabytes of raw data.[17]

This data offers advantages for an empirical study of crowd out effects compared to previous studies that relied on survey data or estimates of broadband penetration rates. Survey data, such as the American Time Use Survey (ATUS), rely on panelists to accurately remember and truthfully reveal their online activity. By contrast, the ComScore Total Home Panel collects a precise, instantaneous record of online behavior for every second of everyday for every device for each person in a participating household.

For all its advantages, raw data collection introduces other challenges. For this project, a key technical issue is differentiating between types of video streams. To overcome this problem, we took care to identify sources of non-pirated and pirated video streams while matching the timestamped data flows to them, as described below. Since our raw data includes every packet of data transmitted—including banner advertisements, auto-refresh pages, and parent-child framed pages—we needed to take care to identify the correct data records for active Internet usage.[18] For example, a good deal of advertising video plays automatically on many web properties, and should not be treated as time spent watching streaming video if the user did not make an active choice to view it.

Non-Pirated OTT Video Streaming Sites

We identified 67 non-pirated over-the-top (OTT) video streaming services in the raw data. These include Netflix, Hulu, YouTube, Slingbox, Amazon Video, and dozens of others (Table 1). The list of non-pirated video streaming services is based on a data analytics report we obtained from ComScore on video streaming apps appearing in the Total Home Panel over 244 days from September 2016 to November 2017.[19]

[17] We used Google BigQuery tools to analyze 5.25 TB of data (the equivalent of 30,337 files of 110 MB each in cold storage).
[18] ComScore Media Metrix Methodology, at 49, 54.
[19] ComScore Streaming Apps Report.

Table 1. Non-Pirated OTT Video Streaming Sites

A&E	FXNow	Spectrum TV
ABC News	FYI TV	Starz
ACORN.TV	HBO Go	Syfy Now
Amazon Video	HBO Now	TBS
AMC Mobile	HGTV Watch	Tubi TV
Apple TV iTunes	History	TWCable TV
Bravo Now	Hulu	Twitch
CBS All Access	Lifetime	USTVNOW.COM
CBS News	Mixer - Streaming	VEVO
CNN Go	NBC	Viewster
Crackle - Movies & TV	Netflix	VUDU Movies and TV
CWTV	OVGuide	Watch ABC
Dailymotion	PBS	Watch Food Network
Directv Now	PBS KIDS Video	Watch TNT
Disney Entertainment	Playstation Vue	Watch Travel Channel
DIY Watch	Pluto.TV	WatchESPN
DramaFever	QVC	WWE
FandangoNOW	Showtime	XBox Movies & TV
FOX News	Showtime Anytime	Xumo
Fox Now	Sling	YouTube

Identifying the sites required manually identifying web domains in the raw data that corresponded to each streaming service.[20] Netflix, for example, distributes streaming video from several domains, including "nflxvideo.net" and a few static numeric Internet Protocol (IP) addresses. Amazon Video proved the most challenging to identify with video streams flowing through domains such as "aiv-cdn.net" and multiple "akamaihd.com" subdomains.[21]

After identifying the non-pirated streaming services in the raw data, we were able to calculate descriptive statistics about American streaming. We see daily and weekly patterns of online activity in the Total Home Panel that conforms with what we already know about American leisure time use (Wallsten, 2015).

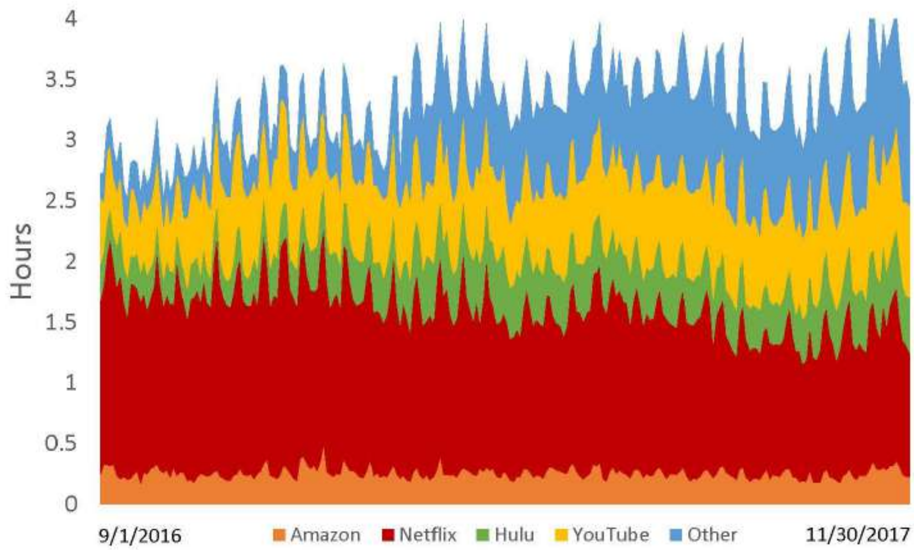
Figure 1 displays an aggregate time series of online video streaming in the Total Home Panel from September 2016 to November 2017, where regular peaks for weekends are apparent.

[20] For each web domain, we compared our manually identified web domains with ComScore's report on streaming applications by timestamp, household, and device. ComScore offered a list of streaming applications in its report without particularly specifying the web domain as it would appear in the raw data. As a robustness check, we confirmed that our efforts to back-engineer ComScore's reports from the same raw data was successful. At each timestamp, we confirmed that we used the same web domains that ComScore used to measure non-pirated video streams in their commercial product. YouTube posed a particular challenge due to its mix of pirated and non-pirated content. For the purposes of this study, we treat all YouTube traffic as non-pirated video content. Since a non-zero percentage of those streams are pirated video streams, our estimates of crowding out effects will be understated. *See generally* https://www.cnbc.com/2017/12/06/how-to-watch-nfl-games-on-facebook-youtube.html.

[21] For YouTube, we identified domain "googlevideo.com" and mimetype = "video/webm" or "video/mp4." For Netflix, we identified video streams with mimetype = "application/octet-stream" and domain = "45.57.28.132," among others. We observed that these streaming services have consistent methods of delivering video within the time span of 2016 to 2017 but delivery methods differ across firms.

The figure shows time spent watching the top four non-pirated streaming sites and another 63 sites combined. This figure shows average daily OTT viewing by households that view non-pirated video streams.[22] Including households that watch no OTT channels reduces the national average to less than one hour of viewing per day. Total hours of OTT viewing by households that watch non-pirated streaming sites increased from three hours per day to nearly four hours per day between 2016 and 2017.

Figure 1. Average Daily OTT Viewing by Households that View OTT



The figure shows changes in the market share of the big four streaming sites. Netflix held the lead in 2017 with the largest market share among non-pirated streaming services, with YouTube second, Hulu third, Amazon fourth, and a rising share of other niche channels combined.

Pirated Video Streaming Sites

Identifying the pirated video streaming sites in the raw data proved more challenging than identifying non-pirated streaming domains. We used several methods to identify web domains that distribute pirated content. We first compiled a list of 2,632 pirated video streaming domains from several sources. We combined a list of top well-known pirate domains, with a list of domains in the Google Transparency Report, a list of well-known Kodi repositories, and a list of domains found on popular subreddits (Table 2).

[22] *See also* ComScore's State of OTT Report, June 2019 (trends on households that view OTT streaming services).

Table 2. Sample of Pirated Video Streaming Sites (2600+)

123movies.to	watchfree.to	torrentkim3.net	mp4upload.com
hdmovieswatch.net	megashare.sc	300mbfilms.co	hdmovie14.net
180upload.com	movie2k.tl	watchmovies-online.ch	watchepisodes.tv
nowvideo.li	indavideo.hu	movpod.in	megashara.com
vid.ag	h33t.to	nowvideo.ch	oneclickwatch.ws
watchseriesus.tv	movie25.com	vidplay.net	watchseries.ag
watchepisodes1.com	watchepisodes1.tv	vidlockers.ag	alltube.tv
lostfilm.tv	mediafire.vc	megarapid.net	moevideo.net
movietube.cc	my-hit.ru	nowfilms.ru	oteupload.com
piratbit.net	series.ly	seriesfree.biz	seriestvix.net
toorgle.com	ultramegabit.com	videopw.com	watch-tv-series.to
watchseriestv.to	watchtvseries.ch	clicknupload.link	fanstash.eu
pirateproxy.tv	watchserieshd.eu	wawa-film.net	akstream.net
cinetube.es	clicktoview.org	cuevana.tv	donevideo.com
filejungle.com	filmifullizle.com	pirateproxy.net	watchseries-online.li

These lists have some drawbacks for identifying pirated video streams in raw Internet traffic data. First, many of these sites serve both non-pirated and pirated content. Labeling some sites as fully pirated content, when they also serve non-pirated content, might cause us to overstate piracy in our dataset. To the extent that we counted some file-sharing sites in our lists as used primarily for piracy, our results will overstate the effects of crowding out.

Second, even if the sites on our list are used primarily for pirated content, they may not directly stream video from their domains, but rather present links to Google Drive files and other file locations for downloads on BitTorrent or elsewhere. We did not include some of the largest file-sharing domains in our list, like Google Drive, because we cannot distinguish between pirated and non-pirated download behavior on them. To the extent that we are missing instances of pirated video downloads in our dataset, our results will understate the effects of crowding out.

Third, pirate domain names change frequently. Many are active only for a limited time and are then taken down or replaced with other names. Pirates have created their own domain naming conventions, such as "123movies.com" or "234movies.com," with a string of numbers followed by keywords such as "movies." These domain names are bulk-generated and changed frequently to evade enforcement authorities.[23] Increasing the costs of tracking piracy is part of the business model for pirates. The harder it is for copyright enforcers to identify new pirate domains, the longer these pirated video streaming sites can operate under the cover of non-pirate activity. Our lists may be missing a large number of domains that were created in the 2016 to 2017 timeframe but were not included in the lists of known pirate sites compiled in later years. To the extent that we are missing pirate sites, our results will understate the crowding out effects of piracy.

[23] Some DNS registrars support bulk or algorithmic domain name generation which are used to evade enforcement authorities. *See generally* https://blog.malwarebytes.com/security-world/2016/12/explained-domain-generating-algorithm/.

With this understanding of the ecosystem, we finalized our list of pirate sites to include 2,632 web domains.[24] Starting with a list of top 1,000 pirate domains,[25] we added approximately 1,000 additional domains from the Google Transparency Report with an Alexa ranking over 20,000,[26] and manually collected domain names from the /r/piracy subreddit for top movies and television programs.[27]

Despite our efforts, additional pirate domains likely exist that of which we are unaware. We also recognize that some of the sites in our list offer non-pirated video content as well as pirated video. These countervailing factors offset each other in the under- and overcounting of pirate sites. Still, based on our knowledge of how piracy works and the data we observe, we are confident that we have reasonably captured the lion's share of pirating behavior in the Total Home Panel.

Time Spent on Pirated Video Streaming

Having identified sources of pirated video, we faced a challenge of measuring the time that American households spent watching video from those sites. Identifying video streams in the raw data is challenging because streaming technology varies by site.

Non-pirate sites tend to use easily recognizable data-delivery technology and typically stream the data rather than deliver an entire file. Pirate sites, understandably, do not always use file types easily identifiable in traffic data, at least partly with the intention of making pirated data flows harder to police.[28] In the raw data, we identified piracy by first looking for recognizable file names for motion pictures and television series. Then, we observed the types of data files used to store video. Finally, we searched for file types in the raw data, and manually spot-checked to identify that sites were indeed hosting pirated programming. To the extent that our methodology

---

[24] We spent considerable time browsing through our list of pirate domains to observe the advertising model and streaming technology by MIME-type and packet delivery methods.

[25] Incorpro Ltd. is a company that tracks infringement of intellectual property and sells a database with piracy intelligence on domains that make available copyrighted content. Sites are closely investigated and given an "Infringement Index" score between 0 and 1 based on a technical and manual assessment of each website. The infringing nature of the website depends on the standard as defined in *Twentieth Century Fox Film Corporation & Ors v. Newzbin Limited [2010] EWHC 608 (Ch)*.

[26] https://transparencyreport.google.com/copyright/overview?hl=en; https://www.alexa.com/siteinfo

[27] https://www.reddit.com/r/Piracy/

[28] Certain streaming devices created measurement challenges, especially in distinguishing between non-pirated and pirated video streams. As far as we can tell, no or almost no pirated content flows over, for example, Roku devices. Other devices, however, are more complicated. Some devices use the "Kodi" platform. Kodi software is available on devices sold as "Kodi boxes," but can also be installed on devices like Amazon Fire sticks. Installing and using Kodi software is legal, but is often used to access pirated content, with 68.6 percent of Kodi users having add-ins installed to make such access easier. *See* https://www.sandvine.com/hubfs/downloads/archive/2017-global-internet-phenomena-spotlight-kodi.pdf. To identify sources of pirated streams through Kodi software, we included a list of 286 Kodi repositories known to host pirated content and pirate add-ons. We include these domains in our list of pirated video streaming sites with the caveat that not all of the traffic from each site is pirated, but that the bulk of engagement with these domains is indeed pirate activity. Lastly, we included a manually-collected list of links from the /r/fullmoviesongoogle subreddit which include specific Google Drive file locations with full-length high-definition motion picture files. *See* https://www.reddit.com/r/fullmoviesongoogle/.

was not comprehensive enough to screen for all the pirated files in the raw data, our results will understate the amount of piracy and crowding out.

The file types that we screened for included ".mkv" which contain direct copies of BluRay or DVD discs, ".mp4" for video in a more compressed format, ".avi" an older version of the .mp4 format, and the ".x265" format.[29] Other common pirated video file types also include the ".vtt" and ".x264" format.[30]

Figure 2 and Figure 3 show examples of pirated video files in our dataset, such as "Moana.2016.BluRay.vtt," "Zootopia.2016.1080p.3D.HEVC.BluRay.x265.mk," "Homeland.S03E05.HDTV.XviD-AFG.avi.mp4," and "Sherlock.S04E02.WEBRip.XviD-FUM.avi."

Figure 2. Video Piracy in Motion Pictures

Moana.2016.BluRay..vtt
Zootopia.2016.1080p.3D.HEVC.BluRay.x265.mk
The Huntsman Winters War 2016 Extended BluRay 720p DTS AC3 x264-ETRG.mkv

Figure 3. Video Piracy in Original Series

Homeland.S03E05.HDTV.XviD-AFG.avi.mp4
Sherlock.S04E02.WEBRip.XviD-FUM.avi
House.of.Cards.2013.S04E11.720p.WEBRip.HEVC.x265

[29] *See generally*, https://handbrake.fr/; https://www.makemkv.com/; https://www.macworld.com/article/3179350/how-to-rip-dvds-and-blu-ray-discs-with-makemkv-and-handbrake.html.
[30] https://www.reddit.com/r/Piracy/comments/a4pgsq/x264_or_x265_question/.

After identifying these pirated video files, we needed a method for estimating the time spent watching these video files. Our method takes into account file size, compression, and data-delivery technology.

Pirated video files tend to come in one of three sizes. In standard definition, pirated video files are often delivered in file sizes of 200 MB to 300 MB for a 20-minute television episode.[31] In high-definition, pirated video files are delivered in file sizes of 1.4 GB or more for 40-minute episodes. Pirated video files often are compressed up to a fourth of the size of non-pirated video.[32] For example, full-length pirated movie downloads are frequently in the range of file sizes of 1.9 GB and 2.7 GB for 90-minute feature films,[33] while the same videos sent by Netflix are delivered in file sizes of 5 GB to 6 GB for high-definition 4K video streams.[34]

To estimate the time spent watching these pirated video files, we counted visits to pirate domains in ten-minute increments. Ten-minute time blocks were selected over 15, 30, or 60-minute time blocks because we concluded that ten-minute increments best assemble into the length of typical pirated video files. Television episodes are typically 20- or 40-minutes long, while movies are closer to 90-minutes long.

Our methodology likely understates the time spent watching pirated video. First, pirated video tends to be more compressed than non-pirated video. Additionally, pirated video is often delivered in a single file request rather than continuous cached streaming connections over time (such as those delivered by Netflix or Amazon). For example, a 40-minute interaction with a pirate site may enable a user with just a few clicks to download four 90-minute movie files for 360 minutes of viewing time, while 40 minutes on Netflix would deliver one television episode of 40 minutes of viewing time. In other words, the time interacting with, and downloading from, the pirate site is probably less than the time spent actually watching the video, whereas almost all the time on Netflix or another non-pirate site involves watching the video.

Recognizing these limitations, we apply this method to estimate time spent watching pirated video from our list of 2,632 pirate sites.

Non-Pirated and Pirated Video Streaming in American Households

Figure 4 shows an indexed comparison of non-pirated and pirated video streaming activity by time of day according to our methodology. The figure shows that video streaming behavior follows consistent time use patterns of leisure activity. Households watch video streams in the evening primetime hours. Pirated video viewing patterns generally follow non-pirated video

---

[31] How-to instructions exist online for pirating content from BluRay discs, with tips for optimal compression for transmission, exchange, and viewing quality. *See, e.g.*, https://www.reddit.com/r/Piracy/, https://www.reddit.com/r/Rainierland/.

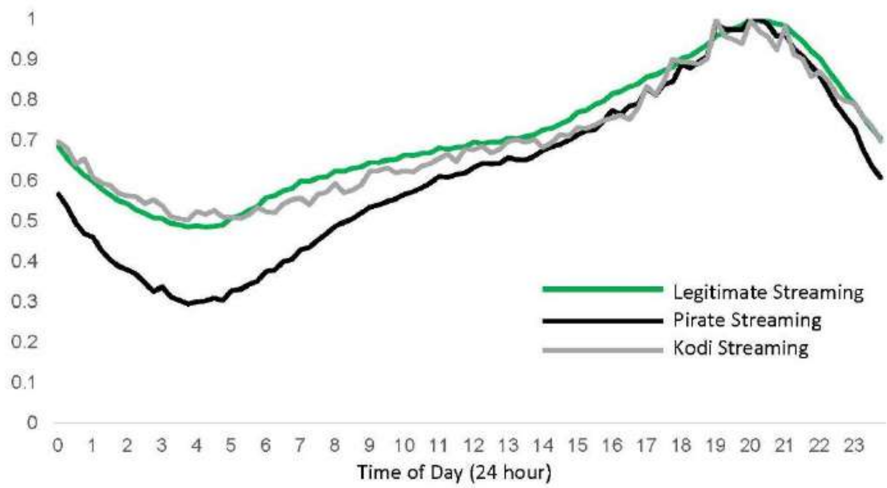
[32] https://news.ycombinator.com/item?id=17537665; https://video.stackexchange.com/questions/7338/how-to-create-a-high-quality-small-file-size-mp4-from-mov.

[33] https://www.reddit.com/r/Piracy/comments/bp6feg/yts_and_yify_are_considered_very_low_quality_on/, https://www.reddit.com/r/Piracy/comments/bpm3p0/what_tools_do_yall_use_for_redbox_dvds/.

[34] https://www.howtogeek.com/338983/how-much-data-does-netflix-use/.

viewing patterns, providing evidence that we are identifying reasonable metrics for time spent watching pirated video streams. Time use activity shows identifiable hours of sleep and work over a 24-hour cycle.

Figure 4. Non-Pirated and Pirated Video Streaming by Time of Day (Indexed 1 = Largest Number of Households Watching)



After checking the reasonableness of our method of identifying pirated video streams in the raw data, we proceeded to categorize over one trillion millisecond-level observations by ten-minute increments. While the raw data includes passive data flows from advertisements and browser-page reloads,[35] our methodology is not affected by the volume of activity within each ten-minute increment. As long as the user is engaged with a domain within a ten-minute increment, the amount of advertising or other delivered data does not affect our estimates of time spent watching pirated video.

We collapsed the raw data into 120 million observations by device-household-hour from online activity measured in ten-minute increments. From this dataset, we took a random sample to generate five million observations on which to run two-stage least squares analysis (2SLS). Descriptive statistics of the sample are available in the Appendix (Table A1).

Empirical Analysis

We begin by looking at a simple correlation. Non-pirated streaming and pirated streaming are positively correlated in our sample. Results from a simple regression show that for every

---

[35] Passive browser activity likely occurs only on desktops, laptops, and tablets. Other devices such as mobile phones, connected televisions, and TV device sticks, do not appear to have as many video advertisements such as banner ads and pop-up video streams.

additional minute spent watching pirated streams, the average American Internet household also watches an additional 0.37 minutes more of non-pirated video streams (Table 3).

Table 3. Simple Correlation

	Non-Pirated Video Streaming
Pirated Video Streaming	0.37***
	(0.01)
Observations	4,682,880
R-squared	0.003
Number of Households	19,412

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

This positive correlation is consistent with the "sampling" or "indirect appropriation" theory of piracy, where consumption of pirated content stimulates viewership of non-pirated content. However, as discussed above, it may also simply be identifying something about the type of people who watch pirated video content. In the following sections we estimate more robust models.

Method

Our empirical strategy is to instrument for pirating behavior and control for other factors that affect viewing behavior, including fixed effects for the month of each panel and household weights for all regressions. Our instrumental variable for pirating is whether a household has a Windows device that streams any video, whether non-pirated or pirated. We consider the Windows operating system a reasonable instrument for several reasons.

First, software piracy is more common on Windows devices than devices running other operating systems.[36] If the software to distribute and watch pirated streams is more readily available on Windows, and household members have a predilection for piracy, then the household is more likely to operate Windows devices. Additionally, since most non-pirated streaming occurs on non-Windows devices, households that have Windows devices are not likely to have more non-pirated streaming than other households, except as influenced by the predilection for piracy (which is correlated with the Windows operating system).

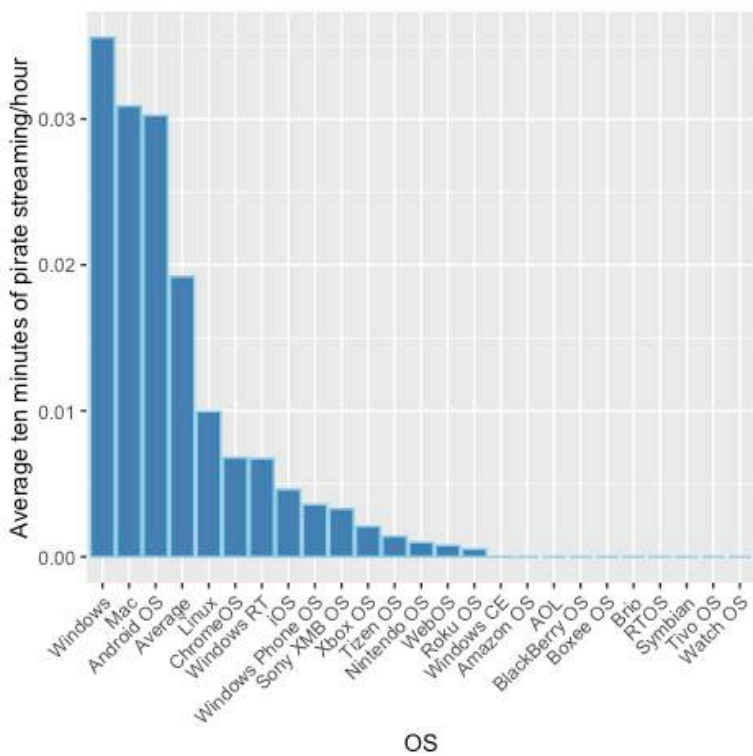
To confirm this intuition, we checked the prevalence of pirated and non-pirated streaming by device operating system in the raw data. We find that Windows devices have the highest average proportion of time spent on pirated streaming, compared to other uses such as non-pirated streaming or web browsing (Figure 5).

[36] https://torrentfreak.com/why-mac-users-are-better-pirates-090206/; https://en.wikipedia.org/wiki/Usage_share_of_operating_systems.

Figure 5 shows average minutes spent on pirated streaming per device per hour.[37] Windows PC devices (that is, not including Windows RT, Windows Phone OS, or Windows CE) show higher proportions of time spent on pirate sites than devices with other types of operating systems.[38]

Other operating systems with high levels of piracy include Mac, Android OS, Linux, and Chrome OS. The frequency distribution of piracy per operating system reveals that most devices used for piracy are represented by a few top operating systems. The remaining operating systems have a smaller proportion of time spent on pirated content but are less prevalent among devices in the panel. These additional operating systems include Windows RT, iOS, Windows Phone OS, Sony XMB OS, Xbox OS, Tizen OS, Nintendo OS, WebOS, Roku OS, Windows CE, Amazon OS, AOL, BlackBerry OS, Boxee OS, Brio, RTOS, Symbian, Tivo OS, and Watch OS.

Figure 5. Pirated Video Streaming by Operating System

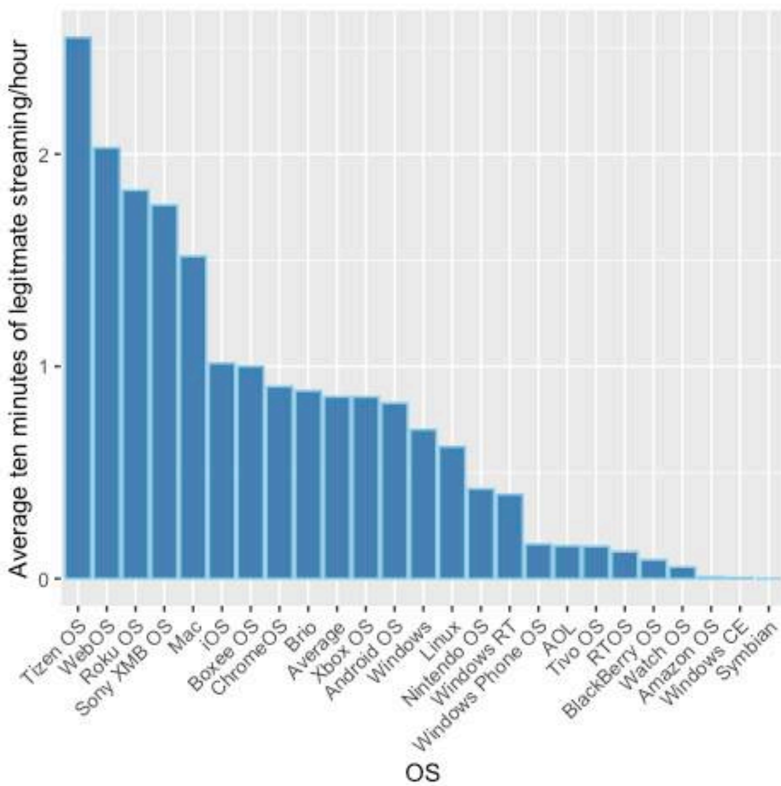


Non-pirated streaming per device per hour by operating system is another relevant consideration in our decision to select the Windows operating system as an instrumental variable. We find that that average time of non-pirated streaming on Windows devices is near the overall average proportion of usage behavior for non-pirated streaming by all operating systems (Figure 6).

[37] This figure does not show total minutes or number of devices per household.
[38] The number of devices with Windows RT, Windows Phone OS, or Windows CE is small enough that we did not include them as additional instrumental variables.

Figure 6 shows that devices with Tizen OS,[39] WebOS, Roku OS, Sony XMB OS, and Mac are used heavily to visit non-pirated streaming sites, with an average of over ten minutes per hour of device usage. Windows devices have below average levels of non-pirated streaming. This descriptive statistic tells us that these devices are used for other content, such as web browsing or pirated streaming.

Figure 6. Non-Pirated Video Streaming by Operating System



Empirical Model

After selecting Windows devices to instrument for piracy, we ran a two-stage model to estimate the crowding out effects of pirated video streams on non-pirated video streams.

We exclude from the model times of day that do not include any video consumption because other activities such as sleep and work take up most of the hours of each day for the average household. As a result, we limit our study to the effects of time spent on one source of video on another source of video. While we note that online activities have been seen to crowd out offline activities (Wallsten, 2015), we focus this study on effects of crowding out from the source of video, whether pirated or non-pirated streaming sites. If we sought to study the crowding out effects of pirated video on a broader range of offline activities, such as sleep or work, then we would apply the model to all hours of the day.

---

[39] Samsung TVs stream via Tizen OS, which is why it is so prevalent in American homes. Samsung Z-Series phones used Tizen, as well, although its more recent phones do not.

Equations 1 and 2 show the two-stage least squares (2SLS) model that includes controls for household demographics, income effects, education effects, and fixed effects for household, day, and hour.

(1) $Pirated\ Streaming_{it} = f(Z)$

(2) $Non - Pirated\ Streaming_{it} = f\left((Z), \widehat{Pirated\ Streaming_{it}}\right)$

Where $\boldsymbol{Z}$ is a vector of the following variables:

$Web\ Browsing_{it}, Head\ of\ Household\ Age_i, Number\ of\ Teenage\ Girls_i, Number\ of\ Teenage\ Boys_i,$

$Number\ of\ TVs_i, Income_i, Education_i,$ *Month Fixed Effects*

We control for differences in household demographics to better estimate the crowding out effect in isolation. Table 4 describes the mean and median values for the control variables in the dataset. Web Browsing is hours of web use for household i on day t for all other web activity besides video streaming, whether non-pirated or pirated. Income is household income for household i. Education is level of highest education for household i. Number of Teenage Girls is the number of teenage girls in household i. Number of Teenage Boys is the number of teenage boys in household i. Number of TVs is the number of TVs in household i.

Table 4. Control Variables

Variable	**Description**	**Mean**	**Median**
$Web\ Browsing_{it}$	Number of ten-minute blocks of non-streaming web activity for household *i* in hour *t*	4.30	5
$Head\ of\ Household\ Age_i$	Age of head of household *i*	50.12	50
$Household\ Size_i$	Number of people in household *i*	3.00	3
$Number\ of\ TVs_i$	Number of TVs in household *i*	2.85	3
$Presence\ of\ Children_i$	Presence of Children in household *i* (1: Yes, 0: No)	0.41	0
$Number\ of\ Teenage\ Boys_i$	Number of teenage boys in household *i*	0.15	0
$Number\ of\ Teenage\ Girls_i$	Number of teenage girls in household *i*	0.16	0
$Income_i$	Level of income in household *i* (1: <$25,000, 2: $25,000-$50,000, 3: $50,000-$75,000, 4: $75,000-$125,000, 5: $125,000+)	3.39	4
$Education_i$	Highest level of education in household *i* (1: 8th grade or less, 2: Some high school, 3: High school, 4: Post-secondary technical or vocational, 5: Associate, 6: Some college, 7: College, 8: Graduate)	4.81	5

Household weights ensure that results are representative of the American Internet household population. Month fixed effects control for monthly and seasonal variation.

We estimated each equation seven times, using seven different measures of the dependent variable. First, we combined all non-pirated streaming from the list of 67 streaming services into a single variable (labeled, "All Non-Pirated"). Then, we estimate the effects of pirated video on each of the top five non-pirated streaming services separately. The results show the effects of pirated video streaming on all non-pirated streaming services, on Netflix, Hulu, Amazon, YouTube, and Sling individually, and the remaining 62 other non-pirated streaming services combined.

Results

Our results show that the time that American households spent on pirate video sites crowded out time spent on non-pirated streaming apps (Table 5). Column 1 shows that every ten-minute time period spent on a pirated streaming site is associated with about 35 fewer minutes on a non-pirated streaming site. The results suggest overall a strong crowding out effect of pirated on non-pirated viewing.

Table 5. Estimated Crowding-Out Effects of Pirated Video on Non-Pirated Streaming Services

Variables	(1) All Non-Pirated Streaming	(2) Netflix	(3) Hulu	(4) Amazon	(5) YouTube	(6) Sling	(7) Other Non-Pirated Streaming
Pirated Streaming	-3.54 (2.45)	-4.96*** (1.87)	-0.14 (0.33)	-4.23*** (1.61)	6.73*** (2.04)	0.19 (0.26)	-3.43** (1.64)
Non-Streaming Activity	0.19*** (0.01)	0.04*** (0.01)	0.01*** (0.00)	0.06*** (0.01)	0.04*** (0.01)	0.00* (0.00)	0.07*** (0.01)
Head of Household Age	0.01*** (0.00)	0.00*** (0.01)	0.00*** (0.00)	0.00*** (0.00)	-0.00 (0.00)	-0.00 (0.00)	0.00 (0.00)
Household Size	0.05*** (0.02)	0.01 (0.01)	-0.00 (0.00)	0.05*** (0.01)	-0.01 (0.01)	-0.00 (0.00)	0.03*** (0.01)
Number of Televisions	0.01 (0.01)	0.02** (0.01)	0.00 (0.00)	0.02*** (0.01)	-0.02** (0.01)	0.01** (0.00)	0.00 (0.01)
< $25,000	0.00 (1.00)	0.00 (1.00)	0.00 (1.00)	0.00 (1.00)	0.00 (1.00)	0.00 (1.00)	0.00 (1.00)
$25,000 - $49,000	-0.10*** (0.02)	-0.05*** (0.02)	-0.01*** (0.01)	-0.05*** (0.01)	-0.02 (0.02)	0.00*** (0.00)	0.00 (0.01)
$50,000 - $74,999	-0.14*** (0.04)	-0.10*** (0.32)	-0.02*** (0.01)	-0.11*** (0.03)	0.06* (0.04)	0.00 (0.00)	-0.02 (0.03)
$75,000 - $124,999	-0.15*** (0.05)	-0.08** (0.04)	-0.02*** (0.01)	-0.11*** (0.03)	0.03 (0.04)	0.01 (0.00)	-0.05* (0.03)
> $125,000	-0.19*** (0.05)	-0.09*** (0.04)	-0.03*** (0.01)	-0.09*** (0.03)	0.03 (0.04)	0.01** (0.01)	-0.08*** (0.03)
Presence of Children	0.03 (0.02)	0.01 (0.02)	0.00 (0.00)	0.02 (0.02)	-0.01 (0.03)	0.00* (0.00)	0.01 (0.02)
Number of Teenage Boys	0.04** (0.02)	0.02 (0.02)	-0.001 (0.00)	-0.03** (0.01)	0.04* (0.02)	0.00 (0.00)	-0.00 (0.01)
Number of Teenage Girls	0.03 (0.03)	0.05** (0.02)	-0.004 (0.01)	0.06*** (0.02)	-0.09*** (0.03)	-0.00 (0.00)	0.05** (0.02)
Observations	4,682,880	4,682,880	4,682,880	4,682,880	4,682,880	4,682,880	4,682,880
F Statistic	3502.95	99.75	244.53	654.74	638.49	115.29	755.32

Includes month fixed effects. Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

Not all streaming services are affected by piracy equally, however. Columns 2 through 7 show results for each of the top five non-pirated streaming sites separately and 62 other non-pirated streaming sites combined. For Netflix (Column 2), Amazon (Column 4), and "Other" (Column 7), the coefficients on pirated streaming show crowding out effects similar in magnitude and statistical significance, consistent with the displacement theory. We see no statistically significant effect on Hulu (Column 3) or Sling (Column 6).

YouTube streaming yields the one different result in Column 5, with a positive, statistically significant coefficient, implying a crowding in effect of piracy, rather than a crowding out. Additionally, the magnitude is large, suggesting that every ten minutes of pirated streaming is associated with an additional 67 minutes of YouTube viewing.

This result likely reflects idiosyncratic viewing patterns on YouTube. While Netflix, Hulu, and Amazon stream only licensed, legal content that is typically at least as long as a half-hour television show (which is around 20 minutes without commercials), YouTube content is more eclectic and broader in scope. The most-viewed videos on YouTube are short music videos,[40] with an enormous range of global content including how-to videos, lectures, television shows, and much more. As discussed earlier, some YouTube content also includes pirated content, although the vast majority of content is user-generated video.[41]

At least two possibilities may explain the positive relationship between YouTube and pirated video. One is that people looking for videos through web searches may end up finding pirated content wherever it is, including on YouTube. In this case, our model is identifying a type of person who likes to pirate, despite our attempts to control for that phenomenon. A related possibility is that the results reflect a pricing effect, as both YouTube and pirated content are free, unlike non-pirated content on other platforms, which are generally either subscription-based or pay-per-view.[42]

Overall, online activity data from 19,764 American households shows that watching pirated video had a crowding out effect on non-pirated over-the-top streaming apps, with some variation across the top five streaming services.

---

[40] https://en.wikipedia.org/wiki/List_of_most-viewed_YouTube_videos.

[41] https://9to5google.com/2017/09/05/google-drive-youtube-copyright-pirates-dmca/.

[42] Hulu transitioned to subscription only shortly before the time period of our sample. Household preferences for free content is an area of research that deserves for more attention than the scope of this study. For example, in our data, we observe that richer households watch more YouTube, despite their ability to pay for subscription services. Our results also show that households with older heads of household stream more, perhaps capturing older people who are retired and spend more time at home. It also shows that lower-income families stream more video than do wealthier families. Finally, and not surprisingly, larger households and households engaged in more non-streaming activities, such as web browsing, also stream more non-pirated content. *See generally* https://www.theregister.co.uk/2011/12/21/ofcom_piracy_research/; https://www.theregister.co.uk/2013/03/08/hug_a_pirate/; https://www.theregister.co.uk/2017/07/10/file_sharing_survey/.

Discussion

The literal interpretation of the analysis is that each minute engaged with a pirate video site crowds out about 3.5 minutes of non-pirated streaming time on sites like Netflix and Amazon. At first blush this result seems improbably large. What could explain such a large crowding out effect?

We suspect that the answer derives from the differing compression techniques combined with the way pirated and non-pirated video files are delivered. Specifically, as discussed above, pirated content tends to be more compressed than non-pirated content and is often delivered as a file download rather than streamed over time. Both of those reasons could cause us to understate time spent on pirated content. For example, if compression and delivery differences mean that a pirated video can be downloaded in one-fourth the amount of time as a non-pirated stream takes to watch, then our model would show a four-minute crowd out of non-pirated viewing for every one minute of pirated viewing.

To the extent that pirated video files are highly compressed and hard to find in raw Internet traffic data, our results understate the amount of piracy and crowding out in the Total Home Panel. On balance, we believe our results fall within a reasonable range of empirical results for a nearly 1-to-1 crowding out effect of pirated video on non-pirated streaming apps.

Conclusion

Pirate sites compete with non-pirated streaming services for a growing share of time that American households spend each day watching online video. Using big data with standard econometric tools, we estimate a crowding out effect of about 3.5—every minute engaged with a pirate site crowds out about 3.5 minutes of time engaged with non-pirated streaming apps like Netflix and Amazon. Because pirated video files are more compressed than non-pirated video files, often by a factor of four, and because pirated video is frequently downloaded in full and non-pirated video is streamed, we conclude that time spent watching pirated video displaces nearly the same amount of time spent watching over-the-top streaming apps.

References

Athey, S., Stern, S. 2013. "The Nature and Incidence of Software Piracy: Evidence from Windows." *NBER Working Paper* No. 19755, http://www.nber.org/papers/w19755.

Chen, H., Yu, J.H., Smith, M.D. 2018. "The Impact of Ebook Distribution on Print Sales: Analysis of a Natural Experiment." *Management Science*, 38(3): 345-359.

Danaher, B., Smith, M.D., Telang, R. 2013. "Piracy and Copyright Enforcement Mechanisms." *NBER Working Paper* No. 19150, http://www.nber.org/papers/w19150.

Hong, S. 2007. "The Recent Growth of the Internet and Changes in Household-Level Demand for Entertainment." *Information Economics and Policy*, 19 (2007): 304-318.

Liebowitz, S.J. 2008. "File-Sharing: Creating Destruction or Just Plain Destruction?" *Journal of Law and Economics*, 49: 1-28.

Liebowitz, S.J. 2008. "Testing File-Sharing's Impact by Examining Record Sales in Cities." *Management Science*, 54(4): 852-859.

Liebowitz, S.J. 2011. "The Metric is in the Message: How Much of the Decline of Sound Recording Sales is Due to File-Sharing?" *Journal of Cultural Economics*. http://dx.doi.org/10.2139/ssrn.1932518.

Liebowitz, S.J. 2017. "Responding to Oberholzer-Gee & Strumpf's Attempted Defense of Their Piracy Paper." *SSRN Electronic Journal.* https://doi.org/10.2139/ssrn.2887122.

Liebowitz, S.J., Zentner, A. 2012. "Clash of the Titans: Does Internet Use Reduce Television Viewing?" *The Review of Economics and Statistics*, 94(1): 234-245.

Ma, L., Montgomery, A.L. Singh, P.V., Smith, M.D. 2014. "An Empirical Analysis of the Impact of Pre-Release Movie Piracy on Box-Office Revenue." *Information Systems Research*, 25(3): 443-666.

Oberholzer-Gee, F., Strumpf, K. 2007. "The Effect of File Sharing on Record Sales: An Empirical Analysis." *Journal of Political Economy*, 115(1): 1-42.

Smith, M., Telang, R. 2012. "Assessing the Academic Literature Regarding the Impact of Media Piracy on Sales." Working Paper, Carnegie Mellon University. http://ssrn.com/abstract=2132153.

Telang, R., Waldfogel, J. 2018. "Piracy and New Product Creation: A Bollywood Story." *Information Economics and Policy*, 43 (2018): 1-11.

Waldfogel, J. 2009. "Lost on the Web: Does Web Distribution Stimulate or Depress Television Viewing?" *Information Economics and Policy*, 21(2): 158-168.

Wallsten, S. 2015. "What Are We Not Doing When We Are Online?" in *Economic Analysis of the Digital Economy, National Bureau of Economic Research*, eds. Goldfarb, A., Greenstein, S., and Tucker, C.E., University of Chicago Press, 55-82, http://www.nber.org/chapters/c13001.

Zentner, A. 2006. “Measuring the Effect of File Sharing on Music Purchases.” *Journal of Law and Economics*, 49: 63-90.

Zentner, A. 2009. “Ten Years of File Sharing and Its Effect on International Physical and Digital Music Sales.” Working Paper, University of Texas at Dallas.

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.

Appendix A

Table A1. Descriptive Statistics

Variable	Obs.	Mean	SD	Min	Max
Pirated 10-Minutes	4,996,942	0.02	0.22	0	6
Non-Pirated 10-Minutes	4,996,942	0.78	1.53	0	6
Non-Streaming 10-Minutes	4,996,942	4.30	1.88	0	6
Kodi 10-Minutes	4,996,942	0.01	0.14	0	6
Google Drive 10-Minutes	4,996,942	0.02	0.20	0	6
Hulu 10-Minutes	4,996,942	0.01	0.26	0	6
YouTube 10-Minutes	4,996,942	0.23	0.87	0	6
Amazon 10-Minutes	4,996,942	0.13	0.73	0	6
Slingbox 10-Minutes	4,996,942	0.01	0.20	0	6
Netflix 10-Minutes	4,996,942	0.24	0.93	0	6
Other Non-Pirated 10-Minutes	4,996,942	0.23	0.79	0	6
Pirated Hour	4,996,942	0.01	0.09	0	1
Non-Pirated Hour	4,996,942	0.30	0.46	0	1
Non-Streaming Hour	4,996,942	1.00	0.06	0	1
TV Device	4,996,942	0.21	0.41	0	1
Household Size	4,996,942	3.02	1.50	1	10
Number of TVs	4,996,942	2.87	1.39	0	10
Head of Household Age	4,996,942	50.08	15.03	18	103
Presence of Children	4,996,942	0.42	0.49	0	1
Number of Children	4,996,942	0.58	0.97	0	13
Number of Teenage Boys	4,996,942	0.17	0.45	0	4
Number of Teenage Girls	4,996,942	0.15	0.42	0	5
Hispanic	4,996,942	0.14	0.35	0	1
African American	4,996,942	0.15	0.35	0	1
Antenna in Household	4,996,942	0.15	0.36	0	1
Cable in Household	4,996,942	0.51	0.50	0	1
Satellite in Household	4,996,942	0.25	0.43	0	1
Streaming Service in Household	4,996,942	0.38	0.49	0	1
OTT Household	4,996,942	0.82	0.38	0	1
Phone Device	4,834,193	0.30	0.46	0	1
Tablet Device	4,834,193	0.13	0.34	0	1
TV Device	4,834,193	0.04	0.19	0	1
Computer Device	4,834,193	0.19	0.39	0	1
DVR SetTop Box Device	4,834,193	0.09	0.28	0	1
Gaming Console Device	4,834,193	0.04	0.20	0	1
Streaming Box Device	4,834,193	0.12	0.33	0	1
iOS	4,834,193	0.18	0.38	0	1
Android OS	4,834,193	0.32	0.47	0	1
Chrome OS	4,834,193	0.02	0.15	0	1
Linux OS	4,834,193	0.13	0.34	0	1
Windows OS	4,834,193	0.17	0.37	0	1
Roku OS	4,834,193	0.06	0.24	0	1

Xbox OS	4,834,193	0.02	0.15	0	1
Amazon OS	4,834,193	0.00	0.03	0	1
Nintendo OS	4,834,193	0.01	0.10	0	1
Amazon Brand Device	4,833,988	0.09	0.29	0	1
Apple Brand Device	4,833,988	0.20	0.40	0	1
Google Brand Device	4,833,988	0.02	0.14	0	1
Roku Brand Device	4,833,988	0.06	0.23	0	1
Income Category	4,996,942	3.40	1.27	1	5
Education Category	4,996,942	4.81	1.35	1	8

Appendix B

These results, for streaming hours only, are generally smaller and less significant than the main set of results. However, they show the same effect direction, and roughly the same magnitude of effect. In particular, YouTube has nearly the same effect as in the other results, and "all other" sites have a larger effect.

Table B1. Estimated Crowding-Out Effects of Pirated Video on Non-Pirated Streaming Services, Streaming Hours Only

Variables	(1) All Non-Pirated Streaming	(2) Netflix	(3) Hulu	(4) Amazon	(5) YouTube	(6) Sling	(7) Other Non-Pirated Streaming
Pirated Streaming	-1.87	-3.48**	0.11	-1.56	6.89***	0.34	-5.61**
	(1.66)	(1.64)	(0.38)	(1.15)	(2.29)	(0.30)	(2.06)
Non-Streaming Activity	0.36***	0.03**	0.02***	0.12***	0.10***	0.01*	0.16***
	(0.02)	(0.02)	(0.00)	(0.01)	(0.02)	(0.00)	(0.02)
Head of Household Age	0.00	0.00	0.00**	0.00***	-0.00	-0.00*	-0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Household Size	0.02	0.00	-0.01**	0.05***	-0.07**	-0.00	0.07**
	(0.02)	(0.02)	(0.01)	(0.02)	(0.03)	(0.00)	(0.03)
Number of Televisions	0.02	0.04**	0.00	0.03**	-0.04	0.00	0.04*
	(0.02)	(0.02)	(0.00)	(0.01)	(0.03)	(0.00)	(0.02)
< $25,000	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)
$25,000 - $49,000	-0.10**	-0.10**	-0.03***	-0.05	0.02	0.01**	0.04
	(0.04)	(0.05)	(0.01)	(0.03)	(0.07)	(0.00)	(0.05)
$50,000 - $74,999	-0.17**	-0.25***	-0.04**	-0.15**	0.27**	0.01	-0.07
	(0.09)	(0.09)	(0.02)	(0.06)	(0.13)	(0.01)	(0.11)
$75,000 - $124,999	-0.15	-0.18*	-0.04	-0.15**	0.23	0.04**	-0.19
	(0.10)	(0.11)	(0.02)	(0.07)	(0.15)	(0.02)	(0.13)
> $125,000	-0.20*	-0.15	-0.07***	-0.05*	0.23	0.05**	-0.29**
	(0.11)	(0.11)	(0.03)	(0.07)	(0.12)	(0.02)	(0.13)
Presence of Children	0.07	0.04	0.01	0.02	-0.05	0.01	0.08
	(0.05)	(0.05)	(0.01)	(0.03)	(0.09)	(0.01)	(0.08)
Number of Teenage Boys	0.07**	0.05	-0.01	-0.08***	0.12*	-0.00	-0.02
	(0.03)	(0.04)	(0.00)	(0.02)	(0.05)	(0.00)	(0.05)
Number of Teenage Girls	0.11*	0.12*	-0.02	0.11***	-0.23***	-0.01	0.21***
	(0.06)	(0.06)	(0.01)	(0.04)	(0.08)	(0.01)	(0.07)
Observations	1,502,749	1,502,749	1,502,749	1,502,749	1,502,749	1,502,749	1,502,749
F Statistic	3008.66	30.62	200.96	880.61	331.15	81.47	186.31

Includes month fixed effects. Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1