Management Science

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

The Effect of Subscription Video-on-Demand on Piracy: Evidence from a Household-Level Randomized Experiment

Miguel Godinho de Matos, Pedro FerreiraMichael D. Smith,

To cite this article:
Miguel Godinho de Matos, Pedro FerreiraMichael D. Smith, (2017) The Effect of Subscription Video-on-Demand on Piracy: Evidence from a Household-Level Randomized Experiment. Management Science

Published online in Articles in Advance 22 Nov 2017

. https://doi.org/10.1287/mnsc.2017.2875

MANAGEMENT SCIENCE
Articles in Advance, pp. 1–21
ISSN 0025-1909 (print), ISSN 1526-5501 (online)
http://pubsonline.informs.org/journal/mnsc/

The Effect of Subscription Video-on-Demand on Piracy: Evidence from a Household-Level Randomized Experiment

Miguel Godinho de Matos,[a] **Pedro Ferreira,**[b] **Michael D. Smith**[c]

[a] Católica Lisbon School of Business and Economics, Universidade Católica Portuguesa, 1649-023 Lisbon, Portugal; [b] Heinz College and Department of Engineering and Public Policy, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213; [c] Heinz College, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213

Contact: miguel.godinhomatos@ucp.pt, http://orcid.org/0000-0002-6333-2753 (MGM); pedrof@cmu.edu (PF); mds@cmu.edu, http://orcid.org/0000-0001-6844-7923 (MDS)

Received: July 22, 2015
Revised: July 23, 2016; January 5, 2017; March 8, 2017
Accepted: June 13, 2017
Published Online in Articles in Advance: November 22, 2017

https://doi.org/10.1287/mnsc.2017.2875

Abstract. We partner with a major multinational telecommunications provider to analyze the effect of subscription video-on-demand (SVoD) services on digital piracy. For a period of 45 consecutive days, a group of randomly selected households who used BitTorrent in the past were gifted with a bundle of TV channels with movies and TV shows that could be streamed as in SVoD. We find that, on average, households that received the gift increased overall TV consumption by 4.6% and reduced Internet downloads and uploads by 4.2% and 4.5%, respectively. However, and also on average, treated households did not change their likelihood of using BitTorrent during the experiment. Our findings are heterogeneous across households and are mediated by the fit between the preferences of households in our sample for movies and the content available as part of the gifted channels. Households with preferences aligned with the gifted content reduced their probability of using BitTorrent during the experiment by 18% and decreased their amount of upload traffic by 45%. We also show using simulation that the size of the SVoD catalog and licensing window restrictions limit significantly the ability of content providers to match SVoD offerings to the preferences of BitTorrent users. Finally, we estimate that households in our sample are willing to pay at most \$3.25 USD per month to access a SVoD catalog as large as Netflix's in the United States. Together, our results show that, as a stand-alone strategy, using legal SVoD to curtail piracy will require, at the minimum, offering content much earlier and at much lower prices than those currently offered in the marketplace, changes that are likely to reduce industry revenue and that may damage overall incentives to produce new content while, at the same time, curbing only a small share of piracy.

History: Accepted by Chris Forman, information systems.
Funding: This research was partially funded by the Carnegie Mellon University Initiative for Digital Entertainment Analytics (IDEA), which receives unrestricted (gift) funding from the Motion Picture Association of America. This research was conducted independently without any oversight or editorial control. M. Godinho de Matos was supported by the Portuguese Foundation of Science and Technology [Grant UID/GES/00407/2013].
Supplemental Material: The online appendix is available at https://doi.org/10.1287/mnsc.2017.2875.

Keywords: randomized experiment • digital piracy • movies • BitTorrent • downloads • Internet

1. Introduction

Digitization is transforming the entertainment industries: books, music, and video. On the supply side, digitization reduced the fixed cost of content creation (Waldfogel 2012) as well as the marginal costs of both reproduction and distribution (Varian 2005). On the demand side, digitization provided end users with easy access to digital piracy on a global scale (Belleflamme et al. 2014, Danaher et al. 2014a, Waldfogel 2012). For example, according to Price (2013), in January 2013, 26% of the Internet users in North America, Europe, and Asia-Pacific—that is, 327 million unique users—sought copyrighted content from illegal online sources, a 9.9% increase from 2011. The widespread growth of digital piracy led researchers to study its impact on the entertainment industry. A number of theoretical models show that piracy may hurt firm profit but increase short-term social welfare (Peitz and Waelbroeck 2006). Researchers have also looked at strategies that may reduce piracy, such as litigation against digital theft, enactment of more restrictive intellectual property laws (Boag 2004, Groennings 2005, Belleflamme et al. 2014), and development of technologies that make copying more difficult (Sinha et al. 2010). The empirical research finds that these approaches generally reduce piracy and increase sales through legal channels (Danaher and Smith 2014; Danaher et al. 2014b; Danaher et al. 2015a, b).

Less effort has been put into analyzing how new legal channels to distribute media affect piracy (Belleflamme et al. 2014, Waldfogel 2012, Danaher et al. 2010). However, such research is important given the number of new technologies and business models developed to distribute digital content (e.g., iTunes selling songs individually as opposed to full albums Waldfogel 2010). Using these technologies, content providers and content distributors have added subscription-based services to existing pay-per-view services (e.g., LastFM and Spotify Aguiar and Waldfogel 2017), which changed the way people consume and relate to media, and which in many cases have had strong implications for how people perceive media ownership. Examples in the entertainment industry include the introduction of MovieLink in 2002 by a consortium of movie studios (Sony, Universal, MGM, Paramount, and Warner) (Ulin 2013), which allowed for downloading movies from the Internet, and later in 2007 the use of online streaming services such as Netflix.

In this paper, we study how the legal distribution of free content to pirates changes their behavior. The first contribution of our paper is that we analyze the impact of piracy in the context of subscription video-on-demand content, an increasingly important source of video consumption. This contrasts with much of the prior literature that has analyzed the impact of piracy on a la carte sales of media products such as songs books and movies (Smith and Telang 2009, 2010; Danaher et al. 2010; Waldfogel 2010; Danaher and Smith 2014). The second contribution of our paper is that we use data from a randomized field experiment at the household level. This contrasts with previous studies that relied mostly on aggregate data obtained from natural experiments (Danaher et al. 2014a) and allows us to measure heterogeneous treatment effects. During this experiment, households that had previously used BitTorrent were given free access to 10 premium TV channels that broadcast movies and TV shows. The content broadcast on these channels could be consumed as subscription-based video-on-demand (SVoD) using time-shift TV. We find that, on average, giving pirates access to these channels increased their consumption of TV by 4.6%, decreased their consumption of Internet, both download and upload traffic by 4.2% and 4.5%, respectively, but did not change their likelihood of using BitTorrent during the experiment. These results show that it is hard to curtail piracy using SVoD.

The third contribution of our work is to identify a mechanism behind this result. We develop a recommender system for media content based on the observed behavior of pirates and compute the fit between what this system recommends to households in our sample and the content offered as part of the SVoD catalog. We find that the likelihood of reducing piracy is mediated by the fit between the latter and the household's preferences for media content. In particular, our models show that pirate households with a theoretical 100% fit with the free content offered by the SVoD service would reduce their likelihood of using BitTorrent by only 18% during the experiment. The fourth contribution of our paper is to use the data from the randomized experiment described above to estimate the pirates' willingness to pay for the SVoD service. Using a multinomial logit model in which households can watch media using transactional VoD (TVoD), SVoD, or piracy (or a combination of these channels), we show that households in our sample would likely pay at most \$3.25 USD per month to access a SVoD catalog as large as that offered by Netflix in the United States (5,600 titles). We also show empirically that such a catalog would likely have only 50% fit with the preferences of pirates.

Taken together, our results show that, as a stand-alone strategy, the use of legal channels to reduce piracy will face significant challenges from three main sources. First, and most importantly, our results show that even when legal SVoD channels are offered for free, and even when there is a very high fit between the content they offer and the interests of pirates, only very few pirates choose to shift their consumption from the pirated content to (free) legal SVoD content. This suggests that to convince users to stop piracy, firms will first need to increase the marginal cost of consuming pirated content (through, for example, higher search costs to discover pirated content, introduction of technological inconvenience in consuming pirated content, or imposition of other legal risks).

Second, even for the relatively small number of pirates who do choose to shift their consumption from pirate to legal SVoD, our recommender system analysis shows a relatively low fit between the content currently licensed to SVoD services and the interests of a typical pirate. This suggests that to attract pirates, content owners will need to make significant changes to their established business models of delayed licensing across various sales channels. Finally, our results show that the consequence of the low levels of fit and marginal cost of piracy is that even if pirates choose to adopt SVoD channels, they have a very low willingness to pay for content catalogs (\$3.25 USD per month for a Netflix-scale catalog of SVoD material), which is well below current market prices for legal SVoD services.

In short, our modeling, experimental, and simulation results all show that, absent significant changes in the marginal cost of consuming pirated content, using video-on-demand to reduce piracy is likely to be effective only if content distributors can offer content much earlier than they do today and at prices much lower than those currently offered in the marketplace. These changes, however, are likely to reduce industry revenue streams and may damage overall incentives for creators to produce new content.

2. Literature Review

The introduction of Napster in 1999 led to a boom in digital piracy, which in turn led to research in digital piracy in various fields including information systems, economics, and marketing (Danaher et al. 2014a). This body of work addresses two fundamental research questions. (1) Does the availability of illegal digital copies of copyrighted content hurt firms and/or societal welfare? (2) Which strategies can be pursued to limit the availability and the impact of pirated copies?

The empirical work analyzing the first question generally finds that piracy hurts firms by reducing legal sales. In the context of music, researchers have analyzed individual-level survey data (Rob and Waldfogel 2006, Zentner 2006, Waldfogel 2010), as well as city-level cross-sectional data (Liebowitz 2008), concluding that illegal music file sharing is harmful for industry profits. The exception is the work of Oberholzer-Gee and Strumpf (2007), who find no effect. With respect to the movie industry, Bounie et al. (2006), Rob and Waldfogel (2007), and Bai and Waldfogel (2012) use survey data to find that file sharing is harmful for the industry. Other studies reach similar conclusions using panel data (Ma et al. 2014) and data from natural experiments (Danaher et al. 2010, Danaher and Smith 2014). Smith and Telang (2009) is a notable exception in this context. The authors show that the availability of pirated copies of movies shown on broadcast TV, typically two to three years after their initial release in theaters, has no impact on subsequent DVD sales. However, and as the authors acknowledge, they do not look at the impact of piracy in the early stages of the movie's life cycle when it may have a significant negative impact on sales.

With respect to the welfare effect of piracy, two papers are worth noting. Rob and Waldfogel (2006) use survey data on consumers' valuations of music albums to estimate that one-third of the gain in consumer surplus that arises from piracy comes at the expense of producers, while the remainder comes from deadweight loss. Another paper is Telang and Waldfogel (2014), where the authors analyze the impact of VCR-based piracy on the production of movies in India. They find that piracy caused a significant decline in the number and quality of new movies produced by Bollywood studios.

Another stream of literature explores how product quality, search costs, and legal prosecution impact piracy behavior. A number of theoretical models assume that pirate copies have lower quality than the originals, and that firms can limit the effects of piracy through product differentiation. For example, Geng and Lee (2013) develop a model of sequential search where consumers obtain digital goods from a legal channel or from the piracy channel. The authors study how different piracy controls—increasing search costs, decreasing the quality of pirate copies, reducing the availability of illegal alternatives, and reducing the number of piracy sites—affect market prices, consumer surplus, and firm profit. They show that reducing the quality of pirate copies and increasing their search costs may be effective strategies to limit the harmful impact of piracy but may also yield heterogeneous effects across consumers. In a similar vein, Wu and Chen (2008) show that under certain conditions, versioning can be an effective tool to fight piracy in digital information goods and that it can both substitute or complement other instruments that may increase the cost of piracy. In addition, see Sundararajan (2004), Chellappa and Shivendu (2005), Belleflamme and Picard (2007), and Johar et al. (2012) for other examples of related theoretical work.

Empirical work in digital piracy has also analyzed extensively the effects of digital rights management (DRM) and litigation. Bhattacharjee et al. (2007) find that increasing litigation threats decreases piracy but does not reduce the availability of illegal content. Sinha et al. (2010) show that DRM may actually increase piracy because it decreases the usability of digital content, which, in turn, reduces the consumers' willingness to pay for content. Zhang (2016) studies the effect of removing DRM from digital music sales, Reimers (2014) studies the effectiveness of private copyright protection in the book industry, Danaher et al. (2014b) evaluate the effects of the introduction of the HADOPI law on digital music sales, Danaher and Smith (2014) analyze the impact of the shutdown of Megaupload on movie sales, Aguiar et al. (2015) study the effect of taking down copyright-infringing websites in Germany, and Danaher et al. (2015a, 2016) study the impact of website blocking in the United Kingdom on piracy and on consumption through legal channels. In general, these papers find that antipiracy efforts can reduce piracy and, in some cases, increase legal consumption in the short term.

Relatively less attention has been devoted to studying how new digital distribution channels, such as transactional video-on-demand (TVoD)—also called pay-per-view—and subscription-based video-on-demand (SVoD), affect digital piracy (Belleflamme et al. 2014, Waldfogel 2012, Danaher et al. 2010). These new distribution channels increase the amount and variety of content that consumers can watch at their leisure. For example, subscription-based services allow consumers to watch as much content as they want when they want. Technologies such as time-shift TV allow consumers to watch their preferred TV shows and movies whenever they want (Wilbur 2008, Belo et al. 2016). If consumers are able to more conveniently find content on legal channels that better match their preferences, then they may use them more often instead of illegally downloading content from the Internet.

So far, only Danaher et al. (2010, 2015b) have studied the effect of channel competition on piracy. In Danaher et al. (2010), the authors use data from a natural experiment to show that removing TV content from iTunes increased piracy. Similarly, Danaher et al. (2015b) analyze how adding content to Hulu.com reduces piracy on that content. These studies provide preliminary evidence of substitution between piracy and legal channels but they do so using: (1) outcomes from natural experiments, which limits the ability to ascertain causal effects, and (2) country-level data, which does not allow for analyzing the household-level drivers of the observed changes. Our paper addresses both of these shortcomings by using a field experiment observed at the household level. Further, the richness of our data allows us to generate an estimate of the pirates' willingness to pay for the legal distribution channel.

3. Model

Consider a representative consumer who can choose to consume content from three different sources: (i) transactional video-on-demand (TVoD) (e.g., iTunes), (ii) subscription video-on-demand (SVoD) (e.g., Netflix), and (iii) piracy (e.g., using BitTorrent). These channels are characterized by the following features.

- TVoD: Consumers choose the content that matches their preferences and pay an average price p per piece of content purchased. TVoD catalogs are generally very large (Resnick and Varian 1997). Hence, we assume that consumers using TVoD watch the content that they want to watch the most and thus there is no misfit between supply and demand under this channel.
- SVoD: Consumers can watch any content they want from a predetermined catalog negotiated between the content distributor and content providers and pay an average access fee A to access all content in the library. The catalog changes over time but at any point in time it may prevent consumers from watching the content that they most want to watch. Therefore, consumers bear an average misfit cost m per piece of content watched under this channel.
- Piracy: Consumers must bear a fixed learning cost F to pirate; for example, they need to learn how to use BitTorrent. Once they bear this cost, pirates are likely to find the content that best matches their preferences, and thus the marginal cost associated to using this channel, call it c, does not arise from unmatched preferences but, for example, from differences between the quality of the downloaded copy and that of the original version, from moral qualms associated with illegal behavior, or from expected litigation costs.

Assume that the utility derived from consuming q pieces of content is given by $V(q)$, with $V(\cdot)$ increasing and concave. Let q_T, q_S, and q_P represent the number of pieces of content consumed by the representative consumer using TVoD, SVoD, and piracy, respectively. This consumer solves

$$\begin{aligned} \underset{q_T, q_S, q_P}{\text{Max}} \quad & V(q_T + q_S + q_P) - pq_T - mq_S - cq_P \\ \text{subject to} \quad & L = q_T + q_S + q_P, \end{aligned}$$

where L represents the amount of time allocated to watching content (the total number of movies households can consume in the time allocated to watching content). Furthermore, assume that each channel has enough content to cover L time. In the experiment we describe in the next sections, we have both $A = 0$ and $F = 0$ because our industrial partner offered SVoD to pirates (who have already learned how to use BitTorrent) at no cost. In this setting, the consumer chooses to allocate her time to consume media from the channels with the lowest marginal cost, that is

$$(q_T^*, q_S^*, q_P^*) = \begin{cases} (L, 0, 0) & \text{if } p < m \text{ and } p < c \\ (0, L, 0) & \text{if } m < p \text{ and } m < c \\ (0, 0, L) & \text{if } c < p \text{ and } c < m \\ (x, L - x, 0) & \text{if } p = m \text{ and } p < c \\ (x, 0, L - x) & \text{if } p = c \text{ and } p < m \\ (0, x, L - x) & \text{if } m = c \text{ and } m < p \\ (x, y, L - x - y) & \text{if } p = m = c, \end{cases}$$

where $x \in [0, L]$ and $y \in [0, L]$. A consumer that uses only the SVoD channel bears a misfit cost of mL because she spends all of her time watching movies that do not completely match her preferences. However, this consumer may, for example, shift some of her consumption, say z, to the TVoD channel. In this case, her misfit cost decreases to $m(L - z)$, but a price pz has to be paid for the consumption in the TVoD channel. Therefore, for consumers using only SVoD and TVoD, $A + m(L - z) + pz$ is a measure of the cost they pay to watch z from TVoD and $L - z$ from SVoD compared to what would happen in a world where the best content for them would be available for free. Two interesting features arise from this simple model. First, consumers may choose to consume from different channels simultaneously if the marginal costs associated to them are similar. Second, when the cost of piracy is zero, consumers will only consume from the TVoD channel if the latter charges zero price. It is, however, unreasonable to believe that content providers and content distributors can sustainably offer TVoD content for free. Likewise, when the cost of piracy is zero, consumers will only consume from the SVoD channel if the latter has no misfit. However, and as we show later in this paper, it is extremely hard to build SVoD catalogs with low misfit costs even if one takes into consideration the observed preferences of pirates. Therefore, if a pirate uses TVoD or SVoD, it must be that the cost of piracy is not zero. We use this fact later in the paper to estimate how much pirates are willing to pay for SVoD. Finally, the model above also shows that no substitution away from the piracy channel towards the SVoD channel may arise from either a low marginal cost of the former or a high misfit cost of the latter. Later in this paper, we use a proxy for the fit of the SVoD channel to empirically disambiguate these effects. This, in turn, allows us to test whether the marginal cost of piracy is zero. If this cost is not zero, then we expect consumers with better fit with the free SVoD offer to be more likely to reduce their piracy consumption.

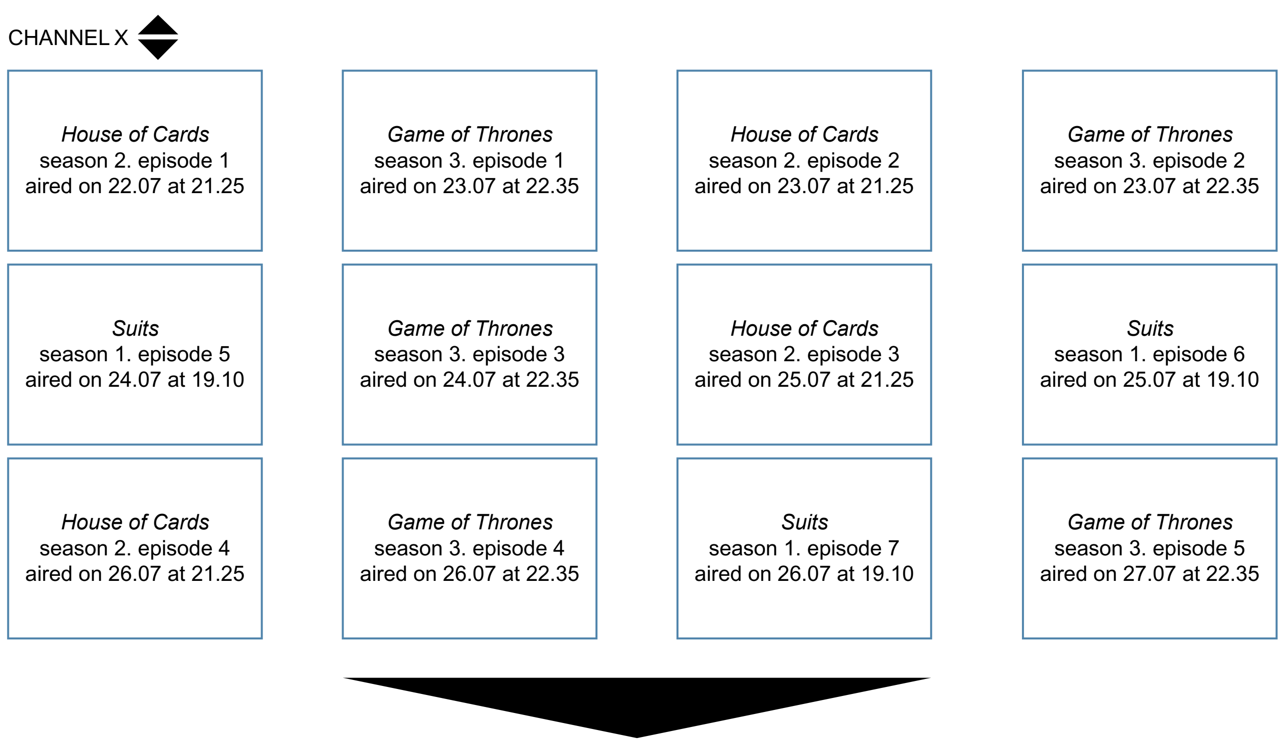
4. Experimental Context and Design

4.1. Experimental Context

Our industrial partner, hereinafter called TELCO, is a large multinational telecommunications provider. It is the market leader in pay-TV services in the country we analyze, serving more than 1 million households. It also offers high-speed Internet service, both transactional (TVoD) and subscription-based (SVoD) video-on-demand, and automated cloud recording (ACR) with a 1-week time window. The basic television service offered by TELCO includes 100 TV channels and access to a TVoD library with more than 2,000 movies and TV shows. In addition, TELCO sells access to packs of thematic TV channels such as documentaries, music, sports, movies, and TV shows that households can subscribe to à la carte. Our study focuses on the Cinema Pack, which is a bundle of 10 TV channels—8 with movies and 2 with TV shows—that can be purchased for $13 USD per month. These channels broadcast attractive content, including popular TV shows, such as *House of Cards*, *Suits*, and *Fargo*, just a few days after they premier in the United States. The Cinema Pack is covered by TELCO's ACR service, and therefore households that subscribe to it can watch any movie or TV show that aired in these channels for a week after the broadcast date without the need to prerecord them. Effectively, accessing the Cinema Pack with time-shift capabilities resembles a SVoD service, as depicted in Figure 1.

TELCO's fixed Internet service allows for download speeds between 3 and 300 Mbps. Upload speeds vary between 1 and 10 Mbps. There are no monthly restrictions on the amount of data that households can exchange with the Internet. Overall, 37% of the traffic in TELCO's network is associated with peer-to-peer file sharing, 35% pertains to web browsing, 19% is video, and the remaining 9% is associated with a number of other services such as video-gaming and voice-over-IP. BitTorrent accounts for 20% of the download traffic and 68% of the upload traffic that can be identified. Peer-to-peer file sharing originates in a small fraction of TELCO's clients, less than 10% of the households. BitTorrent accounts for 95% of the Internet upload

Figure 1. (Color online) Interface to Access the Content of the Cinema Pack with Time-Shift TV, Which Resembles a SVoD Service in Which Consumers Can Choose Movies and TV Shows to Watch Based on Broadcast Channel, Season, and Episode



traffic that can be identified and that pertains to peer-to-peer file sharing.

4.2. Data set

We analyze an anonymized daily panel of household-level data for the months of April, May, and December 2014, and January and February 2015. The data for the months of April and May 2014 were used by TELCO to set up the randomized experiment described in detail in Section 4.3), which took place between December 2014 and February 2015. This data set contains information on the services subscribed, the time spent watching each TV channel, overall download and upload traffic, and identifiable BitTorrent streams.

TELCO collects only aggregate Internet-usage statistics. It does not hold data at the household level on URLs accessed. Therefore, we relied on a third-party provider that regularly monitors the most popular active online BitTorrent swarms (all peers uploading and downloading a given torrent) and keeps a log of all IP addresses that were seen sharing content. This service also registers identifiers for the content that is shared. These IP addresses were matched to IP addresses at TELCO by another entity, who provides data escrow services. This process ensured that, as a research party, we only dealt with anonymized data and that no party could independently revert the encryption keys matching household accounts to their BitTorrent activity.

We did not disclose to TELCO the entity providing the BitTorrent logs. This entity claims that its logs are a representative sample of online BitTorrent streams. However, this entity does not observe all BitTorrent traffic that traverses the Internet. Therefore, the IP addresses that it finds sharing content are likely to be among the most active BitTorrent users. This introduces a limitation in our study: we can know for sure that a household is participating in piracy if it shows up in the BitTorrent logs, but a household that does not show up in these logs may still be illegally sharing copyrighted content. As such, we acknowledge that our results generalize only to the subpopulation of the most active pirates (more precisely, those at TELCO that used BitTorrent during April and May 2014).

4.3. Experimental Setup

We worked with TELCO to run an experiment to give free access to the Cinema Pack to a set of households for a period of 45 consecutive days. Treated households received an email and a text message announcing the activation of the Cinema Pack as a Christmas gift and were told that access would last until the end of January. Activation did not require any intervention from the households. After the experimental period, treated households received a follow-up email and a text message notifying them of the end of the free sampling period and inviting them to subscribe the Cinema Pack for the usual fee. The activation of the Cinema Pack occurred between December 15 and 18, 2014. Households with access to the Cinema Pack could use time-shift TV to consume the content broadcast in these channels through the interface depicted in Figure 1, which resembles that of a SVoD service such as, for example, Netflix or Hulu.

TELCO used stratified sampling to learn whether offering the new TV content would lead households to use less Internet and reduce piracy. With stratified sampling, the units of observation are split into stratum and randomly assigned to treatment and control in each stratum separately (Simon 1979). This design allows TELCO to increase statistical power, in particular to the subpopulation of pirates (Assmann et al. 2000) of which we are interested in this paper. Online Appendix A provides the full details of the experiment run by TELCO and the sampling strategy used. Households observed in the BitTorrent logs from April and May 2014 are analyzed in this paper. They download an average of 3.5 GB/day from the Internet (with a standard deviation of 5.2) and watch on average 4.4 hours of TV per day (with a standard deviation of 2.5). In the absence of priors for the potential effect of treatment, TELCO assumed that, on average, treated households would watch their preferred TV show on TV rather than download it illegally from the Internet. Identifying a smaller effect is arguably uninteresting from an economic point of view. According to Netflix, the average TV show is 450 MB. According to YouTube, this corresponds to 15 minutes of video at 1080p. Therefore, TELCO planed this experiment to identify changes of 15 minutes in TV consumption (which is a worst-case scenario because the average Netflix show is likely longer than 15 minutes) and changes of 450 MB in download traffic, with a confidence level of 95% and with a power of 80%. This entails a random sample of at least 4,034 households that use BitTorrent, half of which should be gifted the Cinema Pack.

TELCO's initial goal was to use a random sample of 18,000 households that did not subscribe to the Cinema Pack in April and May 2014 (before the experiment). However, the sample suffered from attrition; namely, some households (i) had legacy set-top boxes (which do not allow for tracking TV and Internet usage accurately), (ii) had opted out of promotional campaigns before the experiment started, (iii) did not register a single day of TV and Internet usage throughout the whole experiment, or (iv) churned from TELCO during the experiment. We note that sample attrition is orthogonal to treatment assignment and, therefore, the results obtained from this experiment still have a causal interpretation. The caveat is that they generalize only to the population of households that have up to date set-top boxes, use TV and Internet regularly, do not churn, and do not opt out from marketing campaigns—which is

Table 1. Description of Covariates Used in Table 2

Variable name	Variable description
Pirate score	Score from a machine-learning algorithm used to predict BitTorrent use (additional detail in Online Appendix A)
Flag torrent	Indicator for whether household used BitTorrent during the period
TV tenure	Months since household acquired TV service
Internet tenure	Months since household acquired Internet service
Telephone tenure	Months since household acquired telephone service
Active contract	Indicator for whether household must pay a financial penalty to churn
Download	Daily download traffic from the Internet in megabytes
Upload	Daily upload traffic to the Internet in megabytes
TV zapping	Number of distinct TV channels zapped per day
TV	Hours per day spent watching TV
CPTV	Hours per day spent watching channels in the Cinema Pack

Note. All of these covariates are averages computed during the preexperimental period (December 1–13, 2014).

still an interesting subpopulation of users in which to study piracy. From the original sample of 18,000 households, TELCO was left with 10,225 households, half of which were gifted the Cinema Pack.

4.4. Sample Balance

Table 1 describes key observed household-level characteristics, and Table 2 shows that the experimental design described in the previous section achieved good balance with respect to these covariates across treatment and control households. Balance for each covariate is assessed using a t-test for the difference in means between treated and control households. In all cases, we cannot reject the null hypothesis that treated and control households are statistically similar with respect to observable covariates at the 95% confidence level. Covariates included and described in Tables 1 and 2 are the ones that more strongly discriminate households that use BitTorrent according to a machine learning algorithm that we describe in Online Appendix A.

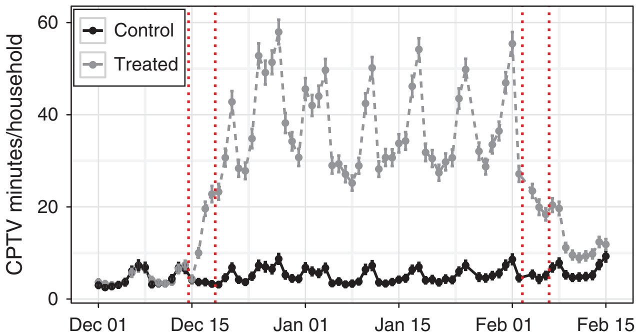
4.5. Usage of the Cinema Pack

Figure 2 shows the daily usage of the Cinema Pack per household in our sample before (December 1–13, 2014), during (December 19, 2014, to February 2, 2015), and immediately after (February 8–15, 2015) the experiment. The dashed vertical lines show the activation and deactivation periods of the offer. This figure shows that, on average, treated households started using the Cinema Pack immediately after they received it and that they stopped using it once the offer was over. In other words, learning effects are unlikely a significant factor in our setting. Moreover, a number of control households subscribed to the Cinema Pack on their own during the experiment. In fact, 19% of them used the Cinema Pack for more than 90 consecutive minutes at least once during the experiment. Similarly, a number of treated households did not use the Cinema Pack despite having access to it for free. In our sample, only 65% of the treated households used the Cinema Pack for more than 90 consecutive minutes at least once during the experiment. We choose 90 minutes of usage to define compliance because this is the average duration of a program broadcast in the TV channels offered as part of the Cinema Pack (results are qualitatively similar for other thresholds and are available on request). These statistics show that we have noncompliance on both sides of the experiment, and therefore

Table 2. Balance in Observed Household Covariates Across Control and Treatment Conditions

	Treated		Control		t-Test		
	Avg.	Std. dev.	Avg.	Std. dev.	Std. effect	t-Statistic	p-Value
Pirate score	0.566	0.251	0.569	0.249	−0.010	−0.492	0.623
Flag torrent	0.580	0.494	0.579	0.494	0.003	0.137	0.891
TV tenure	95.273	56.999	93.193	55.886	0.037	1.859	0.063
Internet tenure	69.437	33.981	68.734	34.094	0.021	1.042	0.297
Telephone tenure	56.923	17.900	56.787	17.959	0.008	0.381	0.703
Active contract	0.799	0.401	0.799	0.401	−0.001	−0.039	0.969
Download (MB/day)	3,498.210	4,652.621	3,548.545	5,181.386	−0.010	−0.516	0.606
Upload (MB/day)	2,098.286	4,794.800	2,152.785	4,987.432	−0.011	−0.562	0.574
TV zapping	11.791	8.438	11.760	8.019	0.004	0.191	0.849
TV	4.368	2.499	4.416	2.520	−0.019	−0.964	0.335
CPTV	0.067	0.282	0.068	0.295	−0.003	−0.142	0.887

Figure 2. (Color online) Daily Usage of the Cinema Pack Across Pirate Households in Our Sample



comparing treated and control households yields the effect of the intention to treat (ITT) (Frangakis and Rubin 1999, Hollis and Campbell 1999), which provides a lower bound for the average treatment effect (ATE). In addition, we compute the local average treatment effect (LATE) using treatment assignment as an instrument for treatment compliance. The latter is the average effect across compliers, the subpopulation of households whose behavior can be modified by random assignment (Angrist et al. 1997). We include our LATE estimates in Online Appendix C, which are in line with the findings reported throughout Section 5.

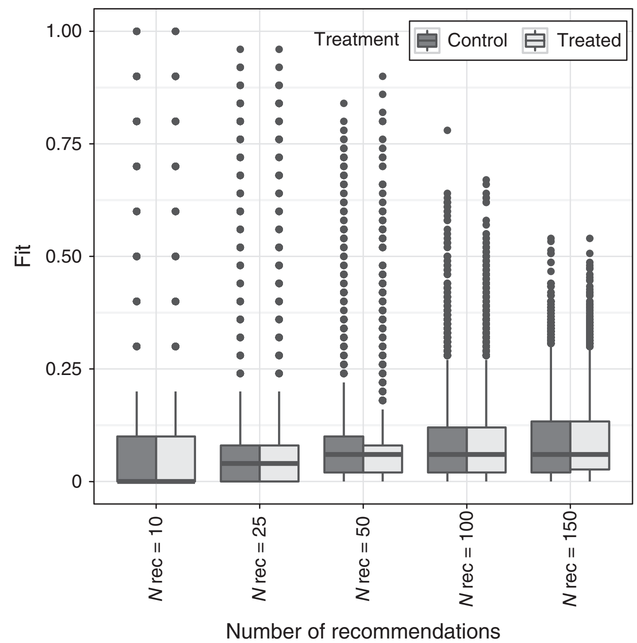
4.6. Estimating the Fit of the Cinema Pack

We use the BitTorrent logs between June and November 2014 to build a recommender system that allows us to generate personalized recommendations of movies to TELCO's pirate households. We note that this recommender system works very well for movies but not for TV series because we do not have episode-level information in our torrent logs. The consequence is that we know whether a household watched a particular series but are unable to issue recommendations for specific seasons or episodes of that series.

We use an item-based collaborative filtering (IBCF) algorithm to build this recommender system. IBCF is among the top technologies to design recommender systems and is used by leading firms such as Amazon's Marketplace (Linden et al. 2003). Online Appendix B describes the parameters used in our system, as well as its out-of-sample performance. In short, the performance of our algorithm is in line with that obtained elsewhere for data sets of similar complexity. In particular, our algorithm performs as well as those reported in Cremonesi et al. (2010) for the cases of the Netflix and MovieLens data sets, which have been repeatedly used to benchmark the performance of recommendation technologies in several academic and industry competitions.

Figure 3 shows how the fit between the content offered as part of the Cinema Pack and that recommended to pirate households in our sample using the

Figure 3. Overlap Between the Top-N Recommendations and the Content Offered as Part of the Cinema Pack



IBCF recommender system described above changes with the number of titles recommended. The average fit with 150 recommendations is 10%. This means that the Cinema Pack includes, on average across households in our sample, 15 out of the top 150 titles that this system recommends. We note that this statistic varies significantly. Its range spans from 0% to 54% with 150 recommendations, and its variation is even larger with fewer recommendations. In Section 6.1, we show that this seemingly low average level of fit is expected for a catalog as large as the Cinema Pack. Figure 3 also shows that the distribution of fit between the content offered as part of the Cinema Pack and that recommended to pirates by our algorithm is similar for treated and control households. This provides evidence that treated and control households have similar preferences for content, reinforcing the quality of the balance achieved by TELCO's randomized schedule. We note that this level of fit does not introduce any (small-sample) bias in our setup. If anything, and from an empirical point of view, it may only limit our ability to find the effect that fit may have on piracy.

5. Empirical Results

5.1. Household Level Effect of the Cinema Pack

Table 3 describes the covariates used throughout our results sections. Figure 4 provides model-free evidence of the effect of the Cinema Pack on TV usage, download traffic, upload traffic, and the likelihood of using BitTorrent during the experiment. The top row summarizes the average daily household behavior before and during the experiment, and the bottom row presents

Table 3. Description of Covariates Used in Tables 4–6, 10, and 11

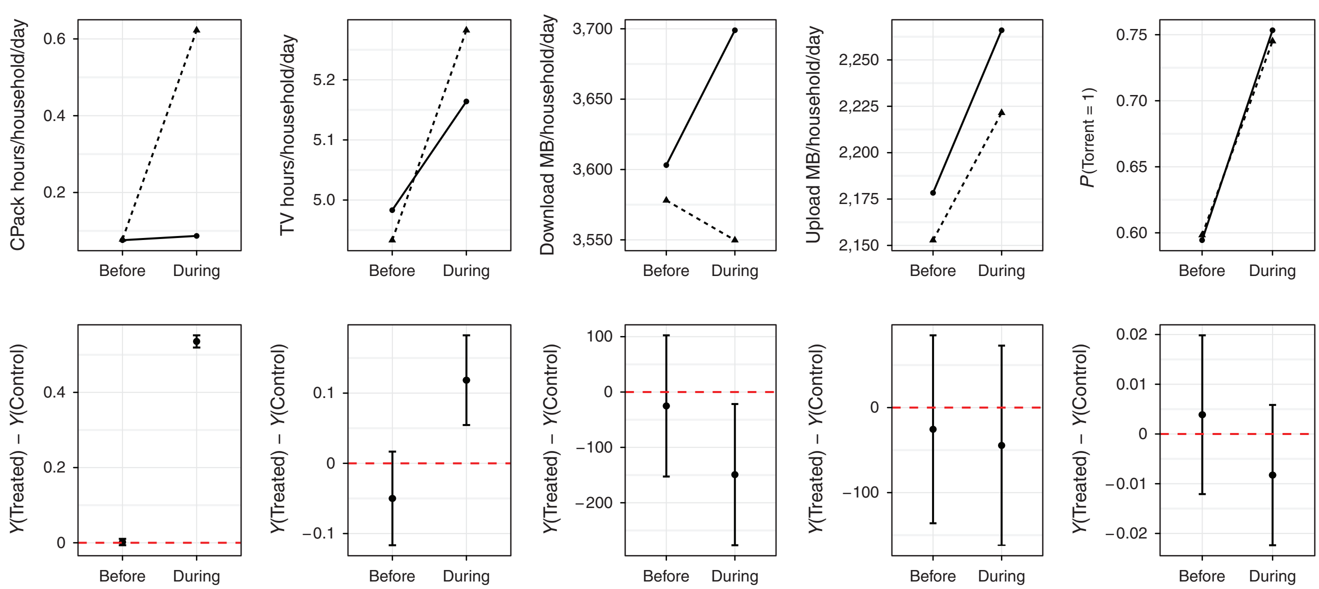
Variable name	Variable description
Treated	Whether access to the Cinema Pack was offered
TV	Time spent watching TV in hours per day
CPTV	Hours per day spent watching the Cinema Pack
Download	Download traffic from the Internet in megabytes per day
Upload	Upload traffic to the Internet in megabytes per day
Flag torrent all	Whether BitTorrent was used to download movies or TV shows
Flag torrent movie	Whether BitTorrent was used to download movies
Flag torrent TV shows	Whether BitTorrent was used to download TV shows
BExp. TV time	Hours per day spent watching TV
BExp. download	Download traffic in megabytes per day
BExp. upload	Upload traffic in megabytes per day
BExp. torrents	Number of torrent downloads per day
Flag no recs	Whether it was possible to generate recommendations
Offer fit	Share of titles suggested by our recommender system that were offered as part of the Cinema Pack during the experiment
Used	Whether the Cinema Pack was watched at least once for more than 90 consecutive minutes during the experiment
Wait time	Time elapsed in days since the beginning of experiment until the content was available as part of the Cinema Pack

Notes. Independent covariates starting with BExp. are averages computed during the preexperimental period (December 1–13, 2014). All other covariates are averages computed during the experiment unless otherwise indicated.

90% confidence intervals for the difference in means between treated and control households across these covariates. As expected, treated and control households do not differ in any of these covariates during the preexperimental period. During the experiment, treated households watched more TV and downloaded less traffic. However, this figure provides no evidence that treated households changed the amount of upload traffic or their likelihood of using BitTorrent when gifted with the Cinema Pack.

Table 4 shows ordinary least-squares (OLS) results for the effect of offering the Cinema Pack on the logarithm of the dependent variables (logarithm is used to adjust variable skewness), as well as time dummies and preexperimental controls to increase the precision of our estimators. Column (1) shows that, on average, treated households increased TV consumption by 4.6% ($p < 0.01$). Column (2) shows that this increase is associated with additional usage of the Cinema Pack. Columns (3) and (4) show that treated

Figure 4. (Color online) Model-Free Comparison of Key Outcome Variables Across Treatment and Control Households



Notes. The plots in the top row summarize (1) the time households spent watching the Cinema Pack, (2) the total time spent watching television, (3) download traffic, (4) upload traffic, and (5) the proportion of households using BitTorrent. Plots in the bottom row compute the t-test for the difference in means. Error bars are for the 90% confidence intervals.

Table 4. Effect of Treatment Assignment on TV Time (Overall and Cinema Pack), Download Traffic, and Upload Traffic During the Experiment (Intention to Treat) (Columns (1)–(4)), and Effect of Treatment on the Probability of Using BitTorrent During the Experiment (Intention to Treat) (Columns (5)–(7))

	Dependent variable						
					Flag torrent		
	Log(*TV*) Pooled OLS (1)	Log(*CPTV*) Pooled OLS (2)	Log(*Download*) Pooled OLS (3)	Log(*Upload*) Pooled OLS (4)	All OLS (5)	Movie OLS (6)	TV Show OLS (7)
Treated	0.046***	0.254***	−0.042**	−0.045**	−0.005	−0.007	−0.002
	(0.007)	(0.005)	(0.016)	(0.022)	(0.008)	(0.009)	(0.009)
Log(*BExp. TV time*)	0.796***	0.066***	−0.003	−0.054***	0.028***	0.041***	0.014**
	(0.015)	(0.004)	(0.014)	(0.018)	(0.006)	(0.006)	(0.006)
Log(*BExp. download*)	−0.004	0.003	0.875***	0.319***	0.003	−0.031***	−0.011**
	(0.005)	(0.003)	(0.014)	(0.016)	(0.005)	(0.005)	(0.005)
Log(*BExp. upload*)	−0.002	0.001	0.006	0.552***	0.080***	0.097***	0.072***
	(0.003)	(0.002)	(0.006)	(0.009)	(0.003)	(0.003)	(0.003)
BExp. torrents	0.002	−0.0005	0.004*	0.031***	0.005**	0.019***	0.024***
	(0.001)	(0.001)	(0.002)	(0.005)	(0.002)	(0.005)	(0.007)
Constant	−0.059	−0.142***	0.307***	−0.583***	0.201***	0.030	−0.058**
	(0.039)	(0.018)	(0.098)	(0.099)	(0.029)	(0.030)	(0.028)
Day dummies	Yes	Yes	Yes	Yes	No	No	No
Observations	372,080	372,080	372,080	372,080	10,225	10,225	10,225
R^2	0.194	0.105	0.333	0.347	0.145	0.135	0.111
Adjusted R^2	0.194	0.105	0.333	0.347	0.144	0.135	0.111
F-statistic	1,795.230***	873.793***	3,708.585***	3,957.534***	346.217***	319.728***	255.385***

Notes. Analysis pertains to the period during the experiment. Columns (1)–(4) have cluster-robust standard errors in parentheses. Clustering is at the household level. Columns (5)–(7) have heteroskedastic-consistent standard errors in parentheses.
*$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

households decreased download and upload traffic by 4.2% and 4.5%, respectively ($p < 0.05$). These statistics provide evidence of a small substitution effect away from Internet consumption toward TV consumption induced by the introduction of the Cinema Pack.

Columns (5)–(7) provide no evidence that, on average, the introduction of the Cinema Pack changed the likelihood of using BitTorrent for pirate households during the experiment. This is true overall, as well as for movies and TV shows separately. Results in columns (5)–(7) are obtained by transforming our daily panel into a cross section. We do so to minimize measurement error in our dependent variable—an indicator of whether the household uses BitTorrent during the experiment. Our concern is not related to bias, given the randomized treatment assignment, but arises because our torrent logs are obtained from a random sample of all torrent traffic on the Internet. Therefore, the uncertainty associated with knowing whether a household used BitTorrent during the experiment may increase the standard errors in our regression (Wooldridge 2010). Converting the daily panel into a cross section reduces the likelihood of incorrectly classifying a household as a nonuser of BitTorrent during the experiment.

Taken together, these results suggest that, on average, pirate households in our sample kept pirating during the experiment. However, this result may arise because of the low fit between the content offered as part of the Cinema Pack and their preferences, or because the marginal cost of piracy is low. We disambiguate the role of these two mechanisms in the next subsection.

5.2. Heterogeneous Effects of the Cinema Pack

We add to our regressions an interaction term measuring the level of fit between the content offered as part of the Cinema Pack and that recommended to each pirate household as identified by the recommender system described in Section 4.6. Table 5 shows the results obtained using the logarithm of TV usage, Cinema Pack usage, and download and upload traffic as dependent variables. For each case, we show results using 50, 100, and 150 recommendations to each household. Columns (1)–(3) show that offering the Cinema Pack to households in our sample increased the total time spent watching TV, irrespective of the fit, by about 5%. Columns (4)–(6) show that this increase is associated with a shift in TV consumption toward the channels offered as part of the Cinema Pack, and that this shift is larger for higher levels of fit. Columns (7)–(9) show a decrease in the amount of download traffic of about 4%, but we do not find evidence that this statistic

Table 5. Effect of Treatment Assignment on TV Usage, Cinema Pack Usage, Download Traffic, and Upload Traffic

	Dependent variable											
	Log(*TV*) Pooled OLS			Log(*CPTV*) Pooled OLS			Log(*Download*) Pooled OLS			Log(*Upload*) Pooled OLS		
	50 Recs. (1)	100 Recs. (2)	150 Recs. (3)	50 Recs. (4)	100 Recs. (5)	150 Recs. (6)	50 Recs. (7)	100 Recs. (8)	150 Recs. (9)	50 Recs. (10)	100 Recs. (11)	150 Recs. (12)
Treated	0.053***	0.048***	0.046***	0.253***	0.248***	0.245***	−0.037**	−0.039*	−0.043**	−0.016	−0.014	−0.015
	(0.008)	(0.009)	(0.009)	(0.006)	(0.006)	(0.006)	(0.019)	(0.020)	(0.021)	(0.024)	(0.026)	(0.027)
Treated × *Offer fit*	−0.117	−0.037	−0.0001	0.019	0.093*	0.136**	−0.074	−0.032	0.027	−0.445**	−0.432*	−0.430
	(0.075)	(0.091)	(0.090)	(0.045)	(0.053)	(0.059)	(0.172)	(0.196)	(0.207)	(0.198)	(0.245)	(0.272)
Offer fit	0.054	0.020	−0.007	0.012	0.015	0.029	−0.117	−0.226*	−0.308**	−0.105	−0.180	−0.079
	(0.054)	(0.064)	(0.065)	(0.019)	(0.023)	(0.027)	(0.114)	(0.134)	(0.144)	(0.148)	(0.183)	(0.200)
Flag no recs	−0.012	−0.011	−0.012	−0.001	0.003	0.006	−0.017	−0.025	−0.029	−0.115***	−0.123***	−0.112***
	(0.009)	(0.009)	(0.009)	(0.006)	(0.007)	(0.007)	(0.021)	(0.022)	(0.022)	(0.027)	(0.028)	(0.028)
Log(*BExp. TV time*)	0.796***	0.796***	0.796***	0.066***	0.066***	0.066***	−0.003	−0.003	−0.003	−0.051***	−0.052***	−0.052***
	(0.015)	(0.015)	(0.015)	(0.004)	(0.004)	(0.004)	(0.014)	(0.014)	(0.014)	(0.018)	(0.018)	(0.018)
Log(*BExp. download*)	−0.004	−0.004	−0.004	0.003	0.003	0.003	0.874***	0.873***	0.873***	0.317***	0.317***	0.318***
	(0.005)	(0.005)	(0.005)	(0.003)	(0.003)	(0.003)	(0.014)	(0.014)	(0.014)	(0.016)	(0.016)	(0.016)
Log(*BExp. upload*)	−0.003	−0.003	−0.003	0.001	0.001	0.001	0.006	0.006	0.007	0.547***	0.548***	0.549***
	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)	(0.006)	(0.006)	(0.006)	(0.010)	(0.010)	(0.010)
BExp. torrents	0.001	0.001	0.001	−0.0005	−0.0004	−0.0004	0.004	0.004	0.004*	0.030***	0.030***	0.030***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)	(0.005)	(0.005)	(0.005)
Constant	−0.056	−0.055	−0.053	−0.143***	−0.146***	−0.148***	0.331***	0.343***	0.348***	−0.505***	−0.500***	−0.520***
	(0.040)	(0.040)	(0.040)	(0.018)	(0.018)	(0.018)	(0.101)	(0.101)	(0.101)	(0.103)	(0.104)	(0.104)
Day dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
F-test:												
Treated × *OfferFit* + *Treated* = 0	0.799	0.0169	0.290	40.615***	46.204***	47.220***	0.455	0.151	0.00721	6.028**	3.726*	3.011*
Observations	372,080	372,080	372,080	372,080	372,080	372,080	372,080	372,080	372,080	372,080	372,080	372,080
R^2	0.194	0.194	0.194	0.105	0.105	0.106	0.333	0.333	0.333	0.348	0.348	0.348
Adjusted R^2	0.194	0.194	0.194	0.105	0.105	0.105	0.333	0.333	0.333	0.348	0.348	0.347
Residual std. error	0.994	0.994	0.994	0.401	0.401	0.401	1.547	1.547	1.547	1.921	1.921	1.921

Notes. Robust standard errors are shown in parentheses. Standard errors are clustered by household. Analysis pertains to the period during the experiment. TV time is in hours per day. CPTV time is in hours per day added the constant 1. Downloads and uploads are in megabytes per day.

$^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

Table 6. Effect of Treatment Assignment on the Likelihood of Showing Up in BitTorrent Logs

	Dependent variable					
	Flag torrent			*Flag movie torrent*		
	Linear probability model					
	50 Recs. (1)	100 Recs. (2)	150 Recs. (3)	50 Recs. (4)	100 Recs. (5)	150 Recs. (6)
Treated	0.007	0.008	0.008	0.004	0.008	0.011
	(0.009)	(0.010)	(0.011)	(0.010)	(0.011)	(0.011)
Treated × *Offer fit*	-0.184^{**}	-0.175^{*}	-0.182^{**}	-0.154^{*}	-0.198^{**}	-0.240^{**}
	(0.082)	(0.090)	(0.092)	(0.083)	(0.099)	(0.104)
Offer fit	-0.232^{***}	−0.027	0.152^{**}	−0.043	0.394^{***}	0.795^{***}
	(0.058)	(0.064)	(0.065)	(0.062)	(0.075)	(0.079)
Flag no recs	-0.213^{***}	-0.195^{***}	-0.179^{***}	-0.295^{***}	-0.257^{***}	-0.224^{***}
	(0.011)	(0.012)	(0.012)	(0.011)	(0.011)	(0.011)
Log(*BExp. TV time*)	0.033^{***}	0.032^{***}	0.032^{***}	0.047^{***}	0.047^{***}	0.046^{***}
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
Log(*BExp. download*)	0.002	0.004	0.005	-0.029^{***}	-0.025^{***}	-0.022^{***}
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
Log(*BExp. upload*)	0.070^{***}	0.071^{***}	0.070^{***}	0.081^{***}	0.080^{***}	0.078^{***}
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
BExp. torrents	0.002^{*}	0.003^{**}	0.003^{**}	0.016^{***}	0.017^{***}	0.017^{***}
	(0.001)	(0.001)	(0.001)	(0.005)	(0.005)	(0.005)
Constant	0.329^{***}	0.296^{***}	0.272^{***}	0.174^{***}	0.113^{***}	0.069^{**}
	(0.030)	(0.031)	(0.031)	(0.030)	(0.031)	(0.031)
F-test:						
Treated × *Offer fit* + *Treated* = 0	5.193^{**}	3.992^{**}	4.192^{**}	3.653^{*}	4.181^{**}	5.420^{**}
Observations	10,225	10,225	10,225	10,225	10,225	10,225
R^2	0.182	0.177	0.176	0.192	0.194	0.202
Adjusted R^2	0.181	0.176	0.176	0.192	0.194	0.202
Residual std. error	0.392	0.393	0.394	0.447	0.446	0.444

Notes. Analysis pertains to the period during the experiment. Robust standard errors are shown in parentheses.
$^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

changes with fit. Columns (10)–(12) show a decrease in upload traffic of roughly 45% for households with a 100% fit. The *F*-statistics to test *Treated* + *Treated* ∗ *Offer fit* = 0 are shown in the table. We believe that our failure to identify a heterogeneous effect on downloads arises from a lack of power to measure the interaction between treated and offer fit. Torrent download traffic accounts only for 20% of all download traffic in TELCO's network, while torrent upload traffic accounts for 68% of all upload traffic. Therefore, the effect of the Cinema Pack on Internet traffic is likely to be stronger on uploads, and we would need more households in our sample to measure a statistically significant heterogeneous effect on downloads.

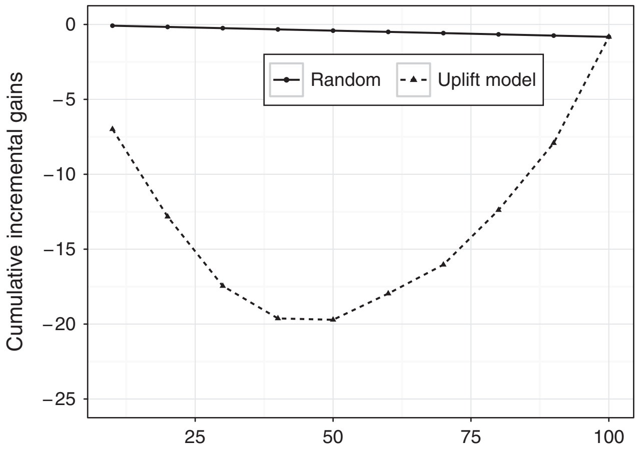
Table 6 shows the effect of the fit between the content offered as part of the Cinema Pack and that recommended to each pirate household on piracy. Columns (1)–(3) use the likelihood of using BitTorrent during the experiment as dependent variable. Columns (4)–(6) use the likelihood of using BitTorrent to share streams associated to movies during the experiment as the dependent variable. We reject the null hypothesis *treated* + *treated* ∗ *offer fit* = 0 for the likelihood of using BitTorrent in columns (1)–(3). The corresponding *F*-statistics are shown in the table.

In sum, we find that a household's decision to adopt the free legal SVoD offer is mediated by the fit between the offered content and household preferences for content. Nonetheless, our results show that even where there is a 100% predicted fit between the offered content and a household's preferences, there is only an 18% reduction in the likelihood of a household using BitTorrent during the experiment. This suggests that for the vast majority of pirate households, the marginal costs of consuming pirated content is essentially zero, a result that has significant implications for antipiracy policy, as we discuss in more detail below.

5.3. Robustness Checks

5.3.1. Household-Level Uplift Analysis. We use machine-learning techniques to develop an incremental response model (Rzepakowski and Jaroszewicz 2012a, b) as a robustness analysis for our results. These models have been increasingly used by economists and

Figure 5. Qini Curves Showing Heterogeneous Treatment Effects at the Household Level



econometricians (Crump et al. 2008, Imai et al. 2013, Athey 2015, Athey and Imbens 2016, Wager and Athey 2017) to predict heterogeneous treatment effects using results from randomized experiments. In this context, researchers are interested in computing uplift, defined as

$$P(Outcome_i \mid Treatment_i = 1, X_i) - P(Outcome_i \mid Treatment_i = 0, X_i) \quad (1)$$

for each subject i in the sample. In our case, the "outcome" of interest is whether a household uses BitTorrent during the experiment and "treatment" indicates whether she was given access to the Cinema Pack. We estimate expression 1 using a random forest incremental response model as described in (Guelman et al. 2015), which we parameterize using fivefold cross validation repeated 20 times. Figure 5 shows the obtained Qini curve. The horizontal axis shows the proportion of households targeted, while the vertical axis shows the resulting cumulative decrease in piracy behavior (in percentage points). The solid black line depicts the effect of targeting households at random. The dashed U-shaped curve shows the effect of targeting households on the basis of uplift starting from the households with the most negative uplift to the households with the most positive uplift.

In our specific case, this Qini curve helps us understand that while, on average, the treatment had no effect, there were heterogeneous responses to the treatment across households. For example, Figure 5 shows that treatment would be most effective if one targeted about 50% of the households; that is, if one ordered households according to uplift and treated the 50% of them with the most negative uplift, then the probability of using BitTorrent would reduce by 20% across our sample of households. As one adds households to the treatment group from the one with most negative uplift to the one with the most positive uplift, the effect of treatment represented by the U-shaped curve deviates from that obtained if one were to add households to the treatment group at random represented by the solid black line. For example, if one were to treat 25% of the households at random, the average effect of treatment would be roughly zero. If instead one were to treat the 25% households with the most negative response to treatment, then the average effect of treatment would be roughly a decrease of 15% in the likelihood of BitTorrent use.

Clearly, the solid black line and dashed U-shaped curve converge to each other when no households are treated (0% in the horizontal axis) and when all households are treated (100% on the horizontal axis). Hence, Qini curves are U-shaped relative to the line that represents treatment at random when treatment effects are heterogeneous, as they are in our case. Table 7 shows the characteristics of households across quintiles of predicted uplift. We observe that the introduction of the Cinema Pack was more effective at reducing the likelihood of using BitTorrent during the experiment for households with moderate Internet use (both downloads and uploads), moderate use of BitTorrent (before the experiment started), relatively more use of BitTorrent to exchange movies vis-à-vis TV shows, and most importantly, higher fit with the content offered as part of the Cinema Pack. Therefore, this analysis provides additional evidence of the mediating role that fit plays in deterring piracy.

5.3.2. Content-Level Analysis. We aggregate the data from the randomized experiment at the content level. Each observation now pertains to a title offered as part of the Cinema Pack. The covariate *Flag torrent* indicates whether the title was observed in the BitTorrent logs. The covariate *During* indicates whether the observation pertains to the period prior to the experiment or to the experimental period. The covariate *Treated* indicates whether *Flag torrent* is computed across treated or control households. More precisely, we select treated households and all movies and TV shows available in the Cinema Pack during the experiment to create a list of available titles. Then, for each title and each period (before and during the experiment), we set *Flag torrent* to one if at least one of the treated households used BitTorrent to download or upload this title during that period. We then repeat this procedure for control households setting *Flag treated* to zero in this case. This procedure provides us with *Flag torrent* computed over two distinct markets of households—the market with access to the Cinema Pack and the market without access to the Cinema Pack. The only difference between these two markets of households is that one comprises only households treated with the Cinema Pack and the

Table 7. Treatment Effect Across Quintiles of Predicted Uplift (Household Level)

	Dependent variable				
	Flag torrent				
	Q1 (1)	Q2 (2)	Q3 (3)	Q4 (4)	Q5 (5)
Treated × *During*	−0.303***	−0.099***	−0.018	0.029	0.324***
	(0.022)	(0.017)	(0.016)	(0.021)	(0.022)
Treated	−0.223***	−0.065***	−0.003	0.099***	0.219***
	(0.022)	(0.019)	(0.019)	(0.022)	(0.020)
During	0.314***	0.155***	0.104***	0.141***	0.059***
	(0.015)	(0.011)	(0.012)	(0.017)	(0.015)
Constant	0.608***	0.781***	0.768***	0.560***	0.230***
	(0.015)	(0.012)	(0.014)	(0.016)	(0.013)
Observations	4,090	4,090	4,090	4,090	4,090
R^2	0.200	0.044	0.015	0.044	0.224
Adjusted R^2	0.200	0.044	0.014	0.043	0.224
Residual std. error	0.439	0.386	0.387	0.451	0.439
F-statistic	340.900***	63.380***	20.810***	62.930***	393.500***
Household characteristics per uplift quintile before the experiment					
Quintile	1	2	3	4	5
Avg. uplift	−0.085	−0.035	−0.012	0.019	0.083
Avg. torrents /day	1.219	3.854	3.757	0.808	0.219
Frac. torrents unknown	0.460	0.491	0.490	0.490	0.474
Frac. torrents movies	0.214	0.178	0.166	0.141	0.117
Frac. torrents TV shows	0.111	0.139	0.146	0.174	0.184
Avg. download MB/day	3,916.000	5,406.000	4,228.000	2,695.000	1,690.000
Avg. upload MB/day	2,324.000	6,220.000	2,838.000	969.800	355.100
Avg. TV time h/day	4.351	4.477	4.560	4.504	4.534
Avg. offer fit (50 recs)	0.114	0.067	0.057	0.048	0.024
Avg. offer fit (100 recs)	0.113	0.082	0.068	0.053	0.027
Avg. offer fit (150 recs)	0.104	0.085	0.073	0.053	0.026

Notes. Robust standard errors are shown in parentheses. Standard errors cluster by household. Q1–Q5 are the quintiles of the "uplift" distribution. Frac. is shorthand for fraction and Avg. is shorthand for average.
$^{***}p < 0.01$.

other comprises only households without the Cinema Pack. Which households obtained the Cinema Pack was randomly determined in our setup. Therefore, differences in behavior across these two markets can only be attributed to the effect of treatment. Following Danaher and Smith (2014), we employ a differences-in-differences approach and cluster the errors at the title level. Table 8 shows the results obtained. We find that treated households pirated the titles included in the Cinema Pack less than did control households. As expected, this result is in line with our findings in Table 6 for households whose preferences fit well with the content offered as part of the Cinema Pack.

5.3.3. Content-Level Uplift Analysis. We apply the framework used in Section 5.3.1 to find heterogeneous effects at the content level. In this case, a subject in our analysis is a movie or a TV show. We use *Flag torrent* as our "outcome" of interest. In this case, *Treated* indicates whether this flag is computed over treated or control households. Therefore, the computation of *Flag torrent* in this subsection is similar to that employed in the previous subsection. Figure 6 shows the Qini curve obtained. The horizontal axis orders movies in the Cinema Pack from the most negative to the most positive in terms of estimated uplift. The "uplift" curve shows the cumulative incremental gains that would be obtained if TELCO changed the proportion of movies included in the Cinema Pack from the bottom to the top decile of the "uplift" distribution. The "random" curve plots the same information if the movies included in the Cinema Pack were selected at random. The fact that the Qini curve deviates from the random line provides evidence of heterogeneous effects across the population of titles included in the Cinema Pack. Table 9 shows the characteristics of the content across quintiles of predicted uplift. We observe that the introduction of the Cinema Pack was more effective at reducing the likelihood of using BitTorrent for more popular movies (IMDb votes), for higher-quality movies (IMDb rating), for younger movies, and relatively more for comedy and animation movies, and less so for drama, action, thriller, and horror movies. In the next section, we

Table 8. Effect of Treatment at the Content Level During the Experiment

	Dependent variable					
	Flag torrent Linear probability model					
	All (1)	Movie (2)	TV show (3)	All (4)	Movie (5)	TV show (6)
Treated × During	−0.036***	−0.035***	−0.048*	−0.036***	−0.035**	−0.048*
	(0.013)	(0.013)	(0.025)	(0.013)	(0.014)	(0.027)
Treated	0.033***	0.033***	0.019	0.033***	0.033***	0.019
	(0.010)	(0.011)	(0.014)	(0.011)	(0.012)	(0.014)
During	−0.001	−0.001	0.010	−0.001	−0.001	0.010
	(0.010)	(0.010)	(0.010)	(0.010)	(0.011)	(0.010)
Constant	0.353***	0.360***	0.250***	0.118***	0.118***	−0.002
	(0.016)	(0.016)	(0.060)	(0.007)	(0.007)	(0.007)
Movie fixed effects	No	No	No	Yes	Yes	Yes
Observations	6,392	5,976	416	6,392	5,976	416
R^2	0.001	0.001	0.001	0.657	0.641	0.910
Adjusted R^2	0.0005	0.0005	−0.006	0.610	0.592	0.897
Residual std. error	0.480	0.482	0.436	0.300	0.308	0.139

Notes. Robust standard errors are shown in parentheses. Standard errors are clustered by IMDb identifier. The analysis includes the periods before and during the experiment.
$^*p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

explore in more detail how movie age affects the ability of content distributors to build SVoD catalog tailored to the preferences of pirates.

5.4. The Effect of Household Impatience

The analyses in the previous sections show that the fit between the preferences of pirates and the content offered as part of the Cinema Pack mediates the effect of treatment assignment on piracy, measured by the amount of Internet traffic uploaded and the likelihood of using BitTorrent during the experiment. Another aspect of the misfit between what pirates would have liked to watch and the content offered to them as part of the Cinema Pack is the fact that households in our sample may be unaware of when a specific movie will be broadcast in the Cinema Pack. The previous results show that a pirate willing to watch a specific title is more likely to watch it using the Cinema Pack if this title is available there. However, if this title is not available in the Cinema Pack when the pirate browses for content, then she may pirate this title from the Internet or wait for it to show up in the Cinema Pack. This rationale leads us to hypothesize that the introduction of the Cinema Pack might reduce piracy more for the titles that are more readily available from the Cinema Pack and less so for the titles that households need to wait longer to consume from the Cinema Pack.

Figure 6. Qini Curves Showing Heterogeneous Treatment Effects at the Content Level

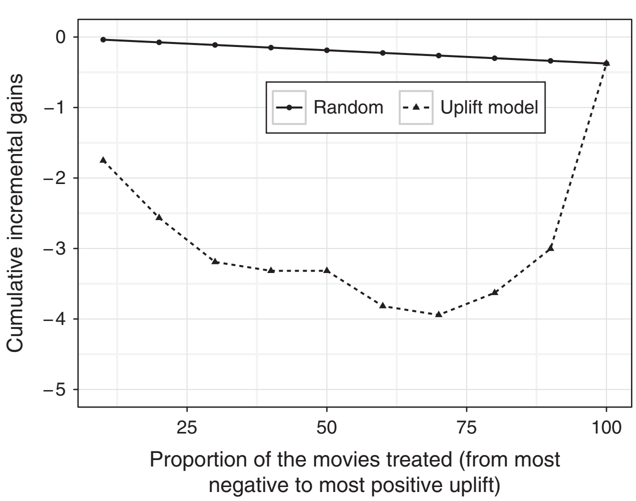


Table 10 shows the household-level results obtained when we interact treatment with the time that households had to wait for the content that matched their preferences to show up in the Cinema Pack. We compute the number of days that elapsed between the beginning of the experiment and the day in which a certain movie showed up in the Cinema Pack. We note that we perform this analysis for movies only using the recommender system described in Section 4.6. The movies offered as part of the Cinema Pack were all in their SVoD window and thus were available in the market much before the experiment started. Therefore, the time elapsed between the start of the experiment and the day they were available in the Cinema Pack is a good estimate for the time that households included in the experiment had to wait to watch them. For each household in our sample, *Wait time* in these regressions is the average of that statistic across the top

Table 9. Treatment Effect Across Quintiles of Predicted Uplift (Content Level)

	Dependent variable				
	Flag torrent Linear probability model				
	(1)	(2)	(3)	(4)	(5)
Treated × *During*	−0.205***	−0.042**	−0.059**	−0.010	0.143***
	(0.032)	(0.020)	(0.025)	(0.023)	(0.037)
Treated	−0.026	−0.007	0.042**	0.049**	0.101***
	(0.026)	(0.017)	(0.021)	(0.020)	(0.030)
During	0.101***	0.003	0.010	−0.020	−0.098***
	(0.024)	(0.019)	(0.018)	(0.017)	(0.029)
Constant	0.625***	0.313***	0.264***	0.206***	0.342***
	(0.034)	(0.033)	(0.030)	(0.029)	(0.030)
Observations	1,228	1,228	1,228	1,224	1,228
R^2	0.029	0.002	0.002	0.004	0.038
Adjusted R^2	0.026	−0.001	−0.001	0.001	0.035
Residual std. error	0.481	0.459	0.447	0.413	0.477
F-statistic	11.970***	0.747	0.723	1.538	16.040***
Quintile	1	2	3	4	5
Avg. uplift	−0.053	−0.022	−0.007	0.010	0.052
Avg. IMDb votes	151,309.000	56,188.000	50,790.000	36,593.000	41,594.000
Avg. IMDb rating	7.063	6.796	6.571	6.043	5.927
Avg. movie age	612.100	597.300	600.400	633.500	625.700
Avg. days to first TV broadcast	4.759	4.310	6.678	7.569	7.586
Frac. Oscar nominated	0.336	0.134	0.117	0.114	0.114
Frac. drama	0.495	0.570	0.687	0.542	0.612
Frac. comedy	0.440	0.391	0.287	0.242	0.205
Frac. action	0.062	0.094	0.179	0.170	0.342
Frac. thriller	0.052	0.088	0.111	0.216	0.336
Frac. animation	0.101	0.088	0.036	0.049	0.026
Frac. horror	0.020	0.042	0.023	0.101	0.114

Notes. Robust standard errors are shown in parentheses. Standard errors cluster by content IMDb identifier. Frac. is shorthand for fraction and Avg. is shorthand for average.
**$p < 0.05$; ***$p < 0.01$.

150 recommendations suggested by our recommender system. The first column shows the intention to treat (ITT), and the second column shows the local average treatment effect (LATE). In both cases, we observe that treated households that had to wait longer had a smaller reduction in their likelihood of using BitTorrent during the experiment. Households that did not have to wait for content that matched their preferences to show up in the Cinema Pack reduced their likelihood of using BitTorrent during the experiment by 3.2%. Each day that households had to wait for such content, this statistic decreased by 0.1%. For the average wait time of 6 days in our data set, the effect of impatience increases the likelihood of using BitTorrent during the experiment by 19% (0.00‘1/−0.032) relative to households that did not have to wait.

6. Estimating the Willingness to Pay for SVoD

6.1. Building a SVoD Catalog Tailored to Pirates

We discussed in the previous sections that the marginal cost of piracy is essentially zero for the vast majority of the households in our sample. However, households whose preferences align better with the content offered as part of the Cinema Pack reduce their likelihood of using BitTorrent during the experiment. Consequently, the marginal cost of piracy for these households cannot be zero because it must be higher than the marginal cost of consuming SVoD, and the latter carries a positive misfit cost. Thus, there must be a positive price that these households would be willing to pay for SVoD service (a price that would just undercut the positive marginal cost of piracy). In this section, we estimate this willingness to pay. The first challenge in this exercise is to build a SVoD catalog tailored to the preferences of pirates. We use the recommender system described in Online Appendix B to do so, but we need to take two constraints into account. One is the size of the catalog that can be recommended. The other is the titles that can be included in such a catalog because of restrictions in licensing windows.

A catalog including all titles in the top 150 recommendations for all pirate households in our sample would need to hold 12,616 different movies. For reference, and according to uNoGS.com, Netflix's catalog

Table 10. Evidence of Household Impatience

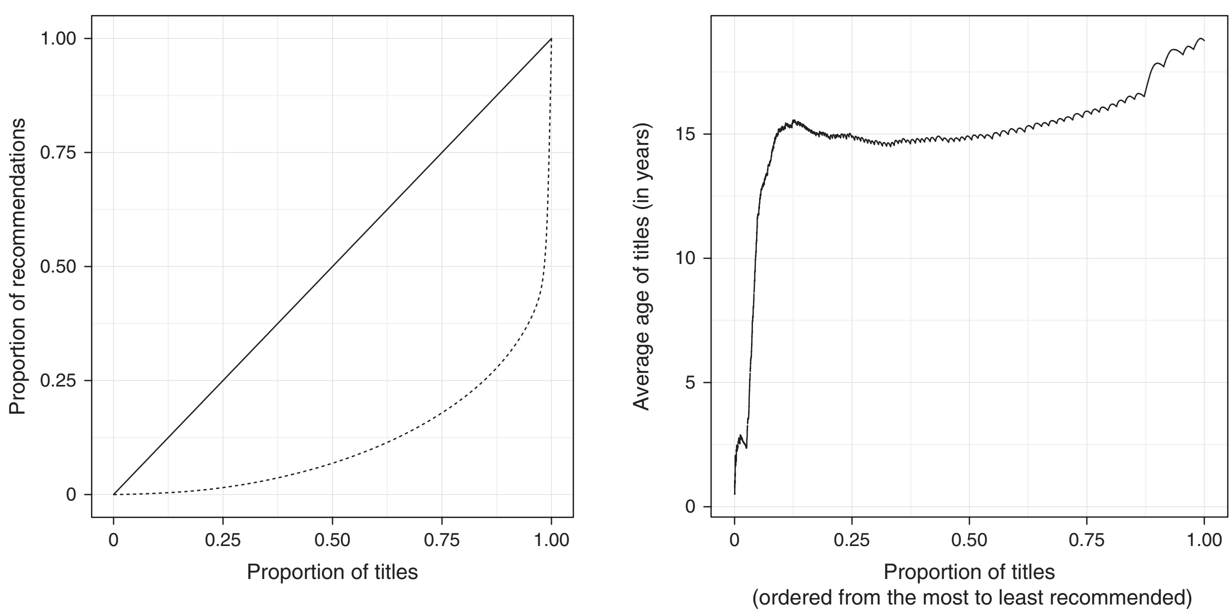
	Dependent variable:	
	Flag torrent (all)	
	OLS (1)	2SLS (2)
Treated	-0.032^{*} (0.017)	
Treated × *Wait time*	0.001^{*} (0.001)	
Used		-0.066^{*} (0.034)
Used × *Wait time*		0.002^{*} (0.001)
Wait time	-0.002^{***} (0.001)	-0.002^{***} (0.001)
Flag no recs	-0.152^{***} (0.016)	-0.153^{***} (0.016)
Log(*BExp. TV time*)	0.032^{***} (0.006)	0.033^{***} (0.006)
Log(*BExp. download*)	0.006 (0.005)	0.005 (0.005)
Log(*BExp. upload*)	0.070^{***} (0.003)	0.070^{***} (0.003)
BExp. torrents	0.003^{**} (0.001)	0.003^{**} (0.001)
Constant	0.324^{***} (0.031)	0.337^{***} (0.033)
Observations	10,225	10,225
R^2	0.177	0.175
Adjusted R^2	0.176	0.175

Notes. Robust standard errors are shown in parentheses. Wait time is measured in days.
$^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

in the United States carried 5,601 titles in May 2016, of which 79% were movies. At the time, the smallest Netflix catalogs carried about 300 titles and were offered in French Southern Territories. Some content has very broad appeal (Brynjolfsson et al. 2006), which may allow content distributors to put together attractive catalogs without the need to spend too much in licensing fees. Figure 7 shows that a small number of the 12,616 titles referred to above cover a significant portion of the recommendations issued to pirate households in our sample by our recommender system, but also that these titles are younger. For reference, according to Ulin (2013), the SVoD window occurs about 12–15 months after the theatrical release. In Figure 8, we vary the size of the catalog between 250 and 6,000 titles and the age of the titles that can be included in this catalog between zero and 15 months after theatrical release. In the left panel, we order the 12,616 movies referred above according to the number of households to which they are recommended and include the top ones as part of the catalog. In the right panel, we do the same for the 9,158 titles that cover the top 15 recommendations to all households in our sample. We observe that the ability to match household preferences expands logarithmically with the size of the catalog, because of the effect of the long tail, and decreases sharply for smaller licensing windows.

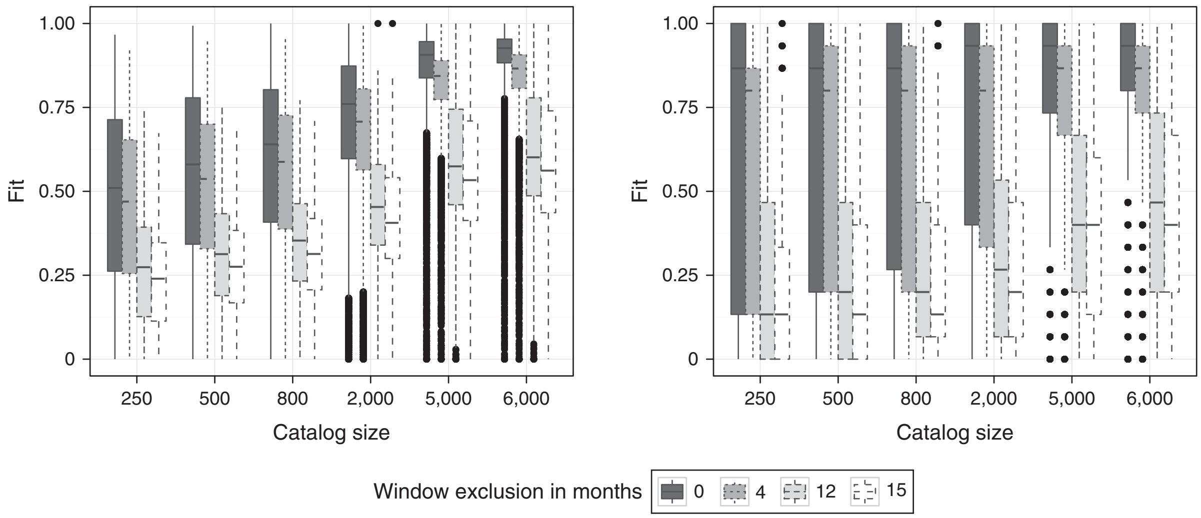
In sum, building catalogs that cater to the preferences of pirates may be challenging for content distributors. The left panel in Figure 8 shows that even a catalog as large as Netflix's in the United States (5,600 titles)

Figure 7. Concentration of Recommendations Over Popularity and Age



Notes. The left panel displays the Lorenz curve for the number of recommendations. The right panel displays the average age of titles as more of them are included in the list of recommendations from the most to the least recommended.

Figure 8. Distribution of the Overlap Between Household Recommendations and the Content of a SVoD Catalog as a Function of Catalog Size and Licensing Window



Notes. The left panel considers all recommended titles for all households in our sample. The right panel considers only the top 15 recommendations for each household in our sample.

would only yield an average fit of 50% with the top 15 recommendations to households in our sample, and that a catalog as large as the Cinema Pack (approximately 800 titles) is likely to only yield an average fit of 24%. These statistics are for an exclusion window of 15 months.

6.2. Estimating the Pirates' Willingness to Pay for SVoD

We use a standard discrete choice model (Train 2009) to estimate how much treated households would be willing to pay for SVoD. In line with the model presented in Section 3, we assume that households have eight alternatives to consume media; namely, TVoD, SVoD, piracy, TVoD + SVoD, TVoD + piracy, SVoD + piracy, TVoD + SVoD + piracy, and no consumption, which plays the role of the outside option. A dummy variable called *Piracy* is set to one for the alternatives that include piracy. Each alternative is characterized by its price. We set the price of TVoD at \$5.1 USD, which is the average monthly TVoD expenditure across households in our sample that used TVoD before the experiment started. We set the price of SVoD to \$13 USD for control households, which is the price charged by TELCO for the Cinema Pack, and to \$0 for treated households. We set the price of piracy to zero. The prices of TVoD + SVoD and TVoD + SVoD + piracy are set to \$5.1 + \$13 = \$18.1 USD, the price of TVoD + piracy is set to \$5.1 USD, and the price of SVoD + piracy is set to \$13 USD. Finally, we use treatment assignment as a characteristic of each household. Treatment was randomly assigned to each household in our setting, which allows us to immediately identify differences in willingness to pay between control and treated households. We interact treatment assignment with the fit between the content offered as part of the Cinema Pack and that recommended to pirate households by our recommender system to measure how willingness to pay changes with fit.

Table 11 shows the results obtained using a multinomial logit choice model for channel selection. The coefficients in this model represent marginal rates of substitution; that is, the dollar amounts that households would need to be paid to continue using piracy after being offered SVoD for free or, in other words, their valuation for the piracy channel. In line with previous results, column (1) shows that, on average, the treatment did not change the households' valuation of piracy. This lack of effect arises because the introduction of the Cinema Pack did not change the likelihood of using BitTorrent during the experiment for the average household. The introduction of the Cinema Pack induced only a change of $-0.044/-0.166 =$ \$0.27 USD per month in the valuation of the piracy channel across households in our sample, but this effect is not statistically different from zero. However, columns (2)–(4) show that households with a strong fit for the content offered as part of the Cinema Pack reduced piracy, and that the value associated to this change is statistically different from zero. For a household with 100% fit, this value is $-0.798/-0.166 =$ \$4.8 USD per month ($p < 0.1$), $-1.072/-0.166 =$ \$6.5 USD per month ($p < 0.1$), and $-1.081/-0.166 =$ \$6.5 USD per month ($p < 0.1$) when we use the recommender system with 50, 100, and 150 titles, respectively, to compute fit.

The estimates above measure the households' average valuation of the piracy channel and thus also their

Table 11. Results Obtained Using a Multinomial Logit Choice Model for Channel Selection

	Dependent variable			
	Media channel choice			
	(1)	50 Recs. (2)	100 Recs. (3)	150 Recs. (4)
Price	−0.166***	−0.166***	−0.166***	−0.166***
	(0.004)	(0.004)	(0.004)	(0.004)
Piracy	1.117***	1.112***	0.946***	0.830***
	(0.032)	(0.037)	(0.040)	(0.041)
Piracy × *Treated*	−0.044	0.007	0.020	0.014
	(0.046)	(0.053)	(0.057)	(0.058)
Piracy × *Treated* × *Offer fit*		−0.798*	−1.072*	−1.081*
		(0.417)	(0.567)	(0.654)
Piracy × *Offer fit*		0.073	2.715***	4.817***
		(0.305)	(0.418)	(0.476)
TVoD FE	Yes	Yes	Yes	Yes
SVoD FE	Yes	Yes	Yes	Yes
Delta piracy value treated 100% fit		−4.763	−6.330	−6.418
Predicted market share				
OUTSIDE	0.133	0.133	0.133	0.133
PIRACY	0.402	0.402	0.402	0.402
SVoD	0.099	0.099	0.099	0.099
TVoD	0.010	0.010	0.010	0.010
PIRACY-SVoD	0.293	0.293	0.293	0.293
TVOD-PIRACY	0.032	0.032	0.032	0.032
TVOD-SVoD	0.008	0.008	0.008	0.008
TVOD-PIRACY-SVoD	0.023	0.023	0.023	0.023
Observations	10,225	10,225	10,225	10,225
Log Likelihood	−14,136.590	−14,133.410	−14,102.890	−14,036.450

$^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

average willingness to pay for SVoD. For comparison, the Cinema Pack is offered by TELCO at $13 USD per month, and Netflix charges between $7.99 and $8.99 USD per month in the United States. As the previous subsection shows, on average, pirates are likely to obtain at most 50% fit with a SVoD catalog as large as Netflix's in the United States because of windowing restrictions. For this level of fit, the last column in Table 11 yields a valuation of $-0.1081/-0.166 * 0.5 =$ $3.25 USD per month for the piracy channel. Therefore, we conclude that attractive SVoD offers are only likely to convert pirates into lawful consumers of media content if SVoD services are offered at prices significantly lower than those currently offered in the market.

7. Conclusions

We partnered with a large telecommunications provider to obtain data from a real-world randomized experiment at the household level aimed at measuring the effect that offering SVoD television content to pirate households would have on piracy consumption. During this experiment, households that used BitTorrent in the past were given 45 days of free access to 10 new TV channels broadcasting movies and TV shows.

Using time-shift TV to watch these channels allows for consuming them as if they were part of an SVoD service. On average, we find that treated households increased TV consumption and decreased Internet use for both downloads and uploads, but did not reduce the likelihood of using BitTorrent during the experiment. We also show that the effect of treatment assignment on BitTorrent usage is mediated by the fit between what pirate households would have liked to watch and the content offered to them as part of the Cinema Pack. We build a state-of-the-art recommender system using item-based collaborative filtering technology, which we trained using BitTorrent logs for households in our sample prior to the experiment. We use this recommender system to develop a measure of fit between what households in our sample might have liked to have watched and the content offered as part of the Cinema Pack. The average and maximum fit of the Cinema Pack is 12% and 100%, respectively, when we use a recommender system that suggests 50 titles to each household.

We show that licensing windows impose significant restrictions on the content that can be included in SVoD catalogs, which hampers the ability of content distributors to offer catalogs that cater to the preferences of pirates. For example, a catalog as large as Netflix's in the United States (5,600 titles) would, at most, yield an average fit of 50% with the preferences of pirate households. Therefore, the seemingly low average level of fit between the content offered as part of the Cinema Pack and the preferences of pirate households in our sample is expected given its small size (only 800 titles). Still, we are able to use the random assignment of this SVoD service to pirate households to show heterogeneous effects. In particular, we show that households with 100% fit with the Cinema Pack reduced their likelihood of using BitTorrent by 18% and their amount of Internet upload traffic by 45% during the experiment. Finally, using a multinomial logit model, we estimate that pirate households whose preferences align 50% with the content offered by the Cinema Pack would be willing to pay $3.25 USD per month for it.

The policy and managerial implications of our results are significant. We show that using legal channels to curtail piracy will require more than just "making content available." Instead, it will likely first require increasing the marginal costs of using pirate channels, for example by increasing search costs to find content, the legal risks incurred when acquiring content, or the technological inconvenience of consuming content. Absent significant changes in the marginal costs of discovering, acquiring, or consuming pirated content, our results show that to be successful in the fight against piracy, content creators will need to make content available on digital channels much earlier than current industry practice and at much lower prices than those charged today. This will require fundamental changes to the current business model of the entertainment industry and will almost certainly lead to reduced revenue streams. Our paper provides a first estimate of how much of a reduction may be expected to eradicate piracy using legal channels. However, reducing revenue streams may also affect the production of content and the pace of innovation in the entertainment industry, which may ultimately also reduce consumer surplus.

Finally, we note several limitations of our results. First, our study analyzes household behavior in a specific country, and attitudes towards piracy might be different across countries based on local culture and regulation. Second, we can say for sure that a household in our sample used BitTorrent when we observe it in the BitTorrent logs; however, other households may also have used BitTorrent prior to the experiment. Third, we use differences in the use of TVoD, SVoD, and piracy induced by treatment to measure the pirates willingness to pay for SVoD. However, a better approach to identifying willingness to pay would have been to offer SVoD at random prices to a sample of pirates and track their behavior. However, for business reasons, this was infeasible to implement at TELCO.

Acknowledgments

The authors thank Ricardo Guizado, Antonio Moreira, Filipa Reis, and Chris Yang for excellent research assistantship; Rodrigo Belo, Brett Danaher, and Rahul Telang for valuable comments on earlier versions of this paper and for many brainstorming sessions; the industry partner for their support; and the organizers of and participants in the Workshop on Digitization and Economics of IT of the 2015 National Bureau of Economic Research Summer Institute for their comments and feedback. All remaining errors are entirely those of the authors.

References

Aguiar L, Waldfogel J (2017) As streaming reaches flood stage, does it stimulate or depress music sales? *Internat. J. Indust. Organ.*, ePub ahead of print July 5, http://www.sciencedirect.com/science/article/pii/S0167718717301753.

Aguiar L, Claussen J, Peukert C (2015) Online copyright enforcement, consumer behavior, and market structure. Technical report, Joint Research Centre of the European Commission.

Angrist JD, Imbens GW, Rubin DB (1997) Identification of causal effects using instrumental variables. *J. Amer. Statist. Assoc.* 91(434):444–455.

Assmann SF, Pocock SJ, Enos LE, Kasten LE (2000) Subgroup analysis and other (mis)uses of baseline data in clinical trials. *Lancet* 355(9209):1064–1069.

Athey S (2015) Machine learning and causal inference for policy evaluation. *Proc. 21th ACM SIGKDD Internat. Conf. Knowledge Discovery and Data Mining* (ACM, New York), 5–6.

Athey S, Imbens G (2016) Recursive partitioning for heterogeneous causal effects. *Proc. Natl. Acad. Sci. USA* 113(27):7353–7360.

Bai J, Waldfogel J (2012) Movie piracy and sales displacement in two samples of Chinese consumers. *Inform. Econom. Policy* 24(3–4): 187–196.

Belleflamme P, Picard PM (2007) Piracy and competition. *J. Econom. Management Strategy* 16(2):351–383.

Belleflamme P, Peitz M, et al. (2014) Digital piracy: An update. Technical report, Aix-Marseille Université, Provence, Southern France.

Belo R, Ferreira P, Godinho de Matos M, Reis F (2016) The impact of time-shift television on TV viewership behavior. NBER Summer Institute workshop paper, National Bureau of Economic Research, Cambridge, MA.

Bhattacharjee S, Gopal RD, Lertwachara K, Marsden JR, Telang R (2007) The effect of digital sharing technologies on music markets: A survival analysis of albums on ranking charts. *Management Sci.* 53(9):1359–1374.

Boag J (2004) The battle of piracy versus privacy: How the Recording Industry Association of America (RIAA) is using the Digital Millennium Copyright Act (DMCA) as its weapon against Internet users' privacy rights. *Calif. Western Law Rev.* 41(1):5.

Bounie D, Bourreau M, Waelbroeck P, others (2006) Piracy and the demand for films: Analysis of piracy behavior in French universities. *Rev. Econom. Res. Copyright Issues* 3(2):15–27.

Breiman L (2001) Random forests. *Machine Learn.* 45(1):5–32.

Breiman L, Friedman J, Stone CJ, Olshen RA (1984) *Classification and Regression Trees* (CRC Press, Boca Raton, FL).

Brynjolfsson E, Hu YJ, Smith MD (2006) From niches to riches: Anatomy of the long tail. *Sloan Management Rev.* 47(3):67–71.

Chellappa RK, Shivendu S (2005) Managing piracy: Pricing and sampling strategies for digital experience goods in vertically segmented markets. *Inform. Systems Res.* 16(4):400–417.

Cremonesi P, Koren Y, Turrin R (2010) Performance of recommender algorithms on top-N recommendation tasks. *Proc. Fourth ACM Conf. Recommender Systems* (ACM, New York), 39–46.

Crump RK, Hotz VJ, Imbens GW, Mitnik OA (2008) Nonparametric tests for treatment effect heterogeneity. *Rev. Econom. Statist.* 90(3):389–405.

Danaher B, Smith MD (2014) Gone in 60 seconds: The impact of the Megaupload shutdown on movie sales. *Internat. J. Indust. Organ.* 33(March):1–8.

Danaher B, Smith MD, Telang R (2014a) Piracy and copyright enforcement mechanisms. Lerner J, Stern S, eds. *Innovation Policy and the Economy*, Vol. 14 (University of Chicago Press, Chicago), 25–61.

Danaher B, Smith MD, Telang R (2015a) The effect of piracy website blocking on consumer behavior. (November), https://ssrn.com/abstract=2612063.

Danaher B, Smith MD, Telang R (2016) Website blocking revisited: The effect of the UK November 2014 blocks on consumer behavior. https://ssrn.com/abstract=2766795.

Danaher B, Dhanasobhon S, Smith MD, Telang R (2010) Converting pirates without cannibalizing purchasers: The impact of digital distribution on physical sales and Internet piracy. *Marketing Sci.* 29(6):1138–1151.

Danaher B, Dhanasobhon S, Smith MD, Telang R (2015b) Understanding media markets in the digital age: Economics and methodology. Goldfarb A, Greenstein SM, Tucker CE, eds. *Economic Analysis of the Digital Economy*, Chap. 13 (University of Chicago Press, Chicago), 385–406.

Danaher B, Smith MD, Telang R, Chen S (2014b) The effect of graduated response anti-piracy laws on music sales: Evidence from an event study in France. *J. Indust. Econom.* 62(3):541–553.

Frangakis CE, Rubin DB (1999) Addressing complications of intention-to-treat analysis in the combined presence of all-or-none treatment-noncompliance and subsequent missing outcomes. *Biometrika* 86(2):365–379.

Friedman JH (2001) Greedy function approximation: A gradient boosting machine. *Ann. Statist.* 29(5):1189–1232.

Geng X, Lee YJ (2013) Competing with piracy: A multichannel sequential search approach. *J. Management Inform. Systems* 30(2): 159–184.

Groennings K (2005) Costs and benefits of the recording industry's litigation against individuals. *Berkeley Tech. Law J.* 20(1):571–601.

Guelman L, Guillén M, Pérez-Marín AM (2015) Uplift random forests. *Cybern. Syst.* 46(3-4):230–248.

Hearst MA, Dumais ST, Osman E, Platt J, Scholkopf B (1998) Support vector machines. *IEEE Intelligent Systems Their Appl.* 13(4):18–28.

Hollis S, Campbell F (1999) What is meant by intention to treat analysis? Survey of published randomised controlled trials. *BMJ* 319(7211):670–674.

Imai K, Ratkovic M (2013) Estimating treatment effect heterogeneity in randomized program evaluation. *Ann. Appl. Statist.* 7(1): 443–470.

Johar M, Kumar N, Mookerjee V (2012) Content provision strategies in the presence of content piracy. *Inform. Systems Res.* 23(3, Part 2):960–975.

Kuhn M (2008) Building predictive models in *R* using the caret package. *J. Statist. Software* 28(5):1–26.

Liebowitz SJ (2008) Research note-testing file sharing's impact on music album sales in cities. *Management Sci.* 54(4):852–859.

Linden G, Smith B, York J (2003) Amazon.com recommendations: Item-to-item collaborative filtering. *IEEE Internet Comput.* 7(1): 76–80.

Ma L, Montgomery AL, Singh PV, Smith MD (2014) An empirical analysis of the impact of pre-release movie piracy on box office revenue. *Inform. Systems Res.* 25(3):590–603.

Oberholzer-Gee F, Strumpf K (2007) The effect of file sharing on record sales: An empirical analysis. *J. Political Econom.* 115(1): 1–42.

Peitz M, Waelbroeck P (2006) Piracy of digital products: A critical review of the theoretical literature. *Inform. Econom. Policy* 18(4):449–476.

Price D (2013) Sizing the piracy universe. White paper, NetNames, London.

Reimers I (2014) Can private copyright protection be effective? Evidence from book publishing. Technical report, Working paper.

Resnick P, Varian HR (1997) Recommender systems. *Comm. ACM* 40(3):56–58.

Rob R, Waldfogel J (2006) Piracy on the high C's: Music downloading, sales displacement, and social welfare in a sample of college students. *J. Law Econom.* 49(1):29–62.

Rob R, Waldfogel J (2007) Piracy on the silver screen. *J. Indust. Econom.* 55(3):379–395.

Rzepakowski P, Jaroszewicz S (2012a) Decision trees for uplift modeling with single and multiple treatments. *Knowledge Inform. Systems* 32(2):303–327.

Rzepakowski P, Jaroszewicz S (2012b) Uplift modeling in direct marketing. *J. Telecomm. Inform. Tech.* 2:43–50.

Sarwar B, Karypis G, Konstan J, Riedl J (2001) Item-based collaborative filtering recommendation algorithms. *Proc. 10th Internat. Conf. World Wide Web* (ACM, New York), 285–295.

Simon R (1979) Restricted randomization designs in clinical trials. *Biometrics* 35(2):503–512.

Sinha RK, Machado FS, Sellman C (2010) Don't think twice, it's all right: Music piracy and pricing in a DRM-free environment. *J. Marketing* 74(2):40–54.

Smith MD, Telang R (2009) Competing with free: The impact of movie broadcasts on DVD sales and Internet piracy. *MIS Quart.* 33(2):321–338.

Smith MD, Telang R (2010) Piracy or promotion? The impact of broadband Internet penetration on DVD sales. *Inform. Econom. Policy* 22(4):289–298.

Sundararajan A (2004) Managing digital piracy: Pricing and protection. *Inform. Systems Res.* 15(3):287–308.

Suykens JAK, Vandewalle J (1999) Least squares support vector machine classifiers. *Neural Processing Lett.* 9(3):293–300.

Telang R, Waldfogel J (2014) Piracy and new product creation: A Bollywood story. (August 6), https://ssrn.com/abstract=2478755.

Tibshirani R (1996) Regression shrinkage and selection via the lasso: A retrospective. *J. Roy. Statist. Soc. Ser. B Statist. Methodol.* 73(3):273–282.

Train KE (2009) *Discrete Choice Methods with Simulation* (Cambridge University Press, Cambridge, UK).

Ulin JC (2013) *The Business of Media Distribution: Monetizing Film, TV and Video Content in an Online World* (Focal Press, Oxon, UK).

Varian HR (2005) Copying and copyright. *J. Econom. Perspect.* 19(2): 121–138.

Wager S, Athey S (2017) Estimation and inference of heterogeneous treatment effects using random forests. *J. Amer. Statist. Assoc.* Forthcoming.

Waldfogel J (2010) Music file sharing and sales displacement in the iTunes era. *Inform. Econom. Policy* 22(4):306–314.

Waldfogel J (2012) Copyright research in the digital age: Moving from piracy to the supply of new products. *Amer. Econom. Rev.* 102(3):337–342.

Wilbur KC (2008) How the digital video recorder (DVR) changes traditional television advertising. *J. Advertising* 37(1):143–149.

Wooldridge JM (2010) *Econometric Analysis of Cross Section and Panel Data*, 2nd ed. (MIT Press, Cambridge, MA).

Wu S-y, Chen P-y (2008) Versioning and piracy control for digital information goods. *Oper. Res.* 56(1):157–172.

Zentner A (2006) Measuring the effect of file sharing on music purchases. *J. Law Econom.* 49(1):63–90.

Zhang L (2016) Intellectual property strategy and the long tail: Evidence from the recorded music industry. *Management Sci.*, ePub ahead of print November 11, https://doi.org/10.1287/mnsc.2016.2562.