

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

Miguel Godinho de Matos^{1,a}, Pedro Ferreira^{2,3,b}, and Michael D. Smith^{3,b}

¹Católica-Lisbon, School of Business and Economics

²Department Engineering and Public Policy, Carnegie Mellon University

³Heinz College, Carnegie Mellon University

^a*miguel.godinhomatos@clsbe.lisboa.ucp.pt*

^b*{pedrof,mds}@cmu.edu*

April 28, 2021

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

*The authors thank Ricardo Guizado, Antonio Moreira, Filipa Reis and Chris Yang for excellent research assistantship. They thank Rodrigo Belo, Brett Danaher and Rahul Telang for valuable comments to earlier versions of this paper and for many brainstorming sessions. The authors thank the Industrial Partner for their support and they thank organizers and participants in the Workshop on Digitization and Economics of IT of the 2015 NBER Summer Institute for their comments and feedback. This research was partially funded by Carnegie Mellon University's 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. The first author acknowledges the support from the Portuguese Foundation for Science and Technology with grant reference UID/GES/00407/2013. All remaining errors are entirely our own.

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 US. 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 may damage overall incentives to produce new content while, at the same time, might curb only a small share of piracy.

1 Introduction

Digitization is transforming the entertainment industries: books, music and video. On the supply side, digitization reduced the fixed cost of content creation (Waldfoegel, 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., 2013; Waldfoegel, 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, enacting more restrictive intellectual property laws (Boag, 2004; Groennings, 2005; Belleflamme et al., 2014) and developing 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., 2014, 2015; Danaher, 2015).

Less effort has been put into analyzing how new legal channels to distribute media affect piracy (Belleflamme et al., 2014; Waldfoegel, 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

(Waldfoegel, 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 Waldfoegel, 2015)), which changed the way people consume and relate to media, 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; Waldfoegel, 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., 2013) 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 broadcasted in 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/month to access a SVoD catalog as large as that offered by Netflix in the US (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, using legal channels to reduce piracy will face significant challenges from three main sources. First, and most important, 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, introducing technological inconvenience in consuming pirated content, or imposing 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/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 pirate 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., 2013). 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? and (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); 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 2-3 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 a 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 they 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 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); 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 (2014) studies the effect of removing DRM from digital music sales, Reimers (2016) study the effectiveness of private copyright protection in the book industry, Danaher et al. (2014) 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. (2015) and Danaher et al. (2016) study the impact of website blocking in the UK on piracy and on consumption through legal channels. In general, these papers find that

anti-piracy efforts can reduce piracy and, in some cases, increase legal consumption in the short term.

Relatively less attention has been devoted to study 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) and Danaher (2015) 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 (2015) 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 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 pre-determined 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:

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

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 a-la carte. Our study focus 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/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 US. 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 pre-record 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,

Figure 1: 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.



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

4.2 Dataset

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 dataset 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 sub-population 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 sub-population of pirates (Assmann et al., 2000), which we are interested in this paper. 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 4034 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; (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 still an interesting sub-population of users 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 Appendix A.

4.5 Usage of the Cinema Pack

Figure 2 shows the daily usage of the Cinema Pack per households in our sample before (2014-12-01 to 2014-12-13), during (2014-12-19 to 2015-02-02) and immediately after (2015-02-08 to 2015-02-15) the experiment. The dashed vertical lines show the activation and de-activation 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

Table 1: Description of covariates used in Table 2 of the paper. All these covariates are averages computed during the pre-experimental period (2014-12-01 to 2014-12-13).

Variable Name	Variable Description
Pirate Score	Score from ML algorithm used to predict BitTorrent use. Additional detail is provided in 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

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 upon request). These statistics show that we have non-compliance on both sides of the experiment and therefore 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 the sub-population of households “whose behavior can be modified by random assignment (compliers)” (Angrist et al., 1997). We include our LATE estimates in Appendix C, which are in line with the findings reported throughout section 5.

Table 2: Balance in observed household covariates across control and treatment conditions.

	Treated		Control		T-test		
	Avg.	StDev	Avg.	StDev	Std. Effect	T-stat	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

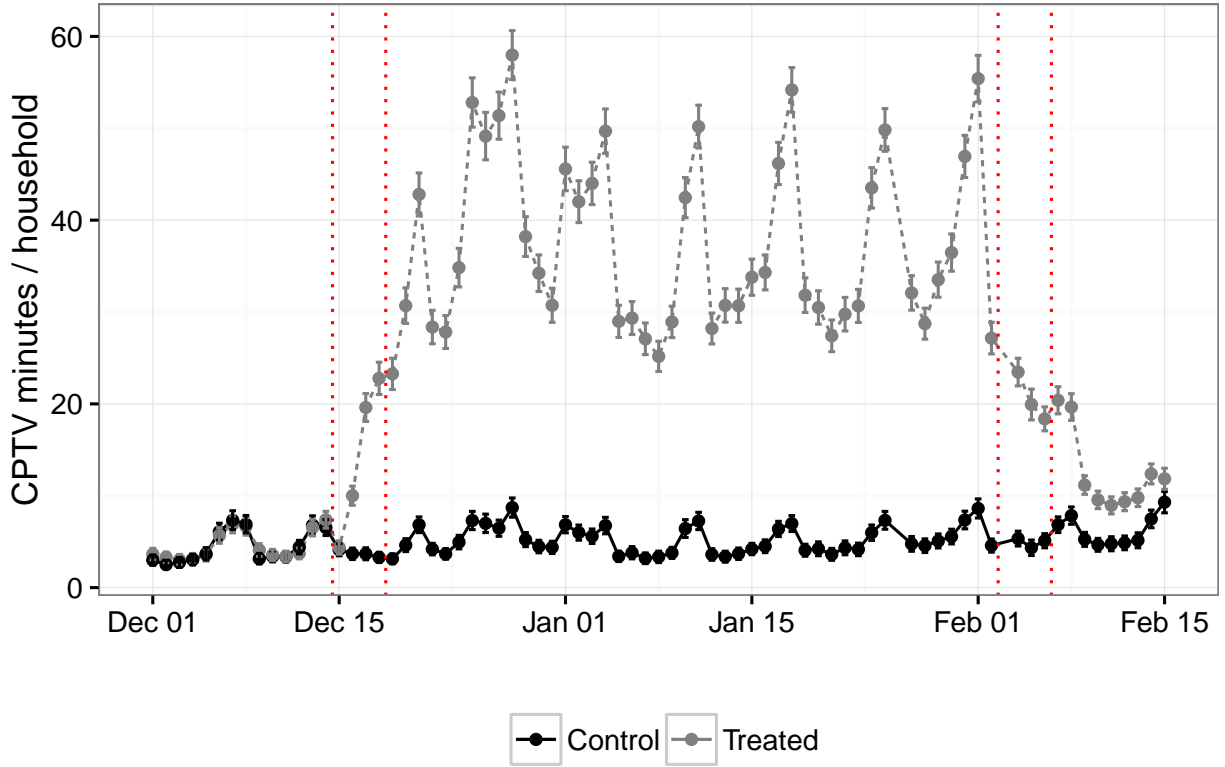


Figure 2: Daily usage of the Cinema Pack across pirate households in our sample.

4.6 Estimating the Fit of the Cinema Pack

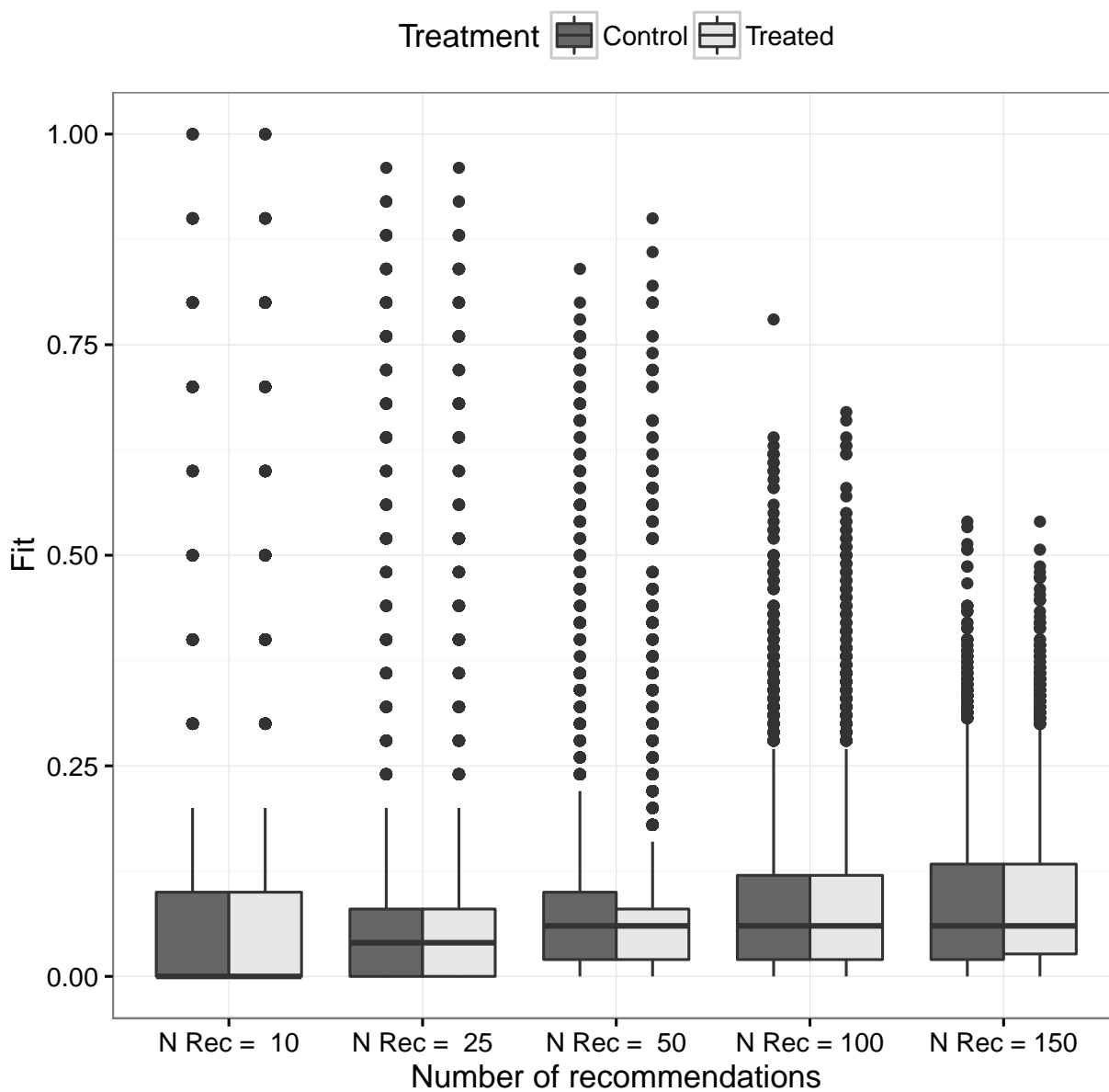
We use the BitTorrent logs between June and November 2014 to build a recommender system that allows us to generate personalized recommendations of movies to TELCO’s pirate households. We note that this recommender system works very well for movies but not for TV series because we do not have episode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

We use a 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). 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 datasets 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 datasets, 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 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 subsection 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.

Figure 3: Overlap between the top-N recommendations and the content offered as part of the Cinema Pack.



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 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 pre-experimental 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 3: Description of covariates used in Tables 4, 5, 6, 10 and 11 of the paper. Independent covariates starting with BExp. are averages computed during the pre-experimental period (2014-12-01 to 2014-12-13). All other covariates are averages computed during the experiment unless otherwise indicated.

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

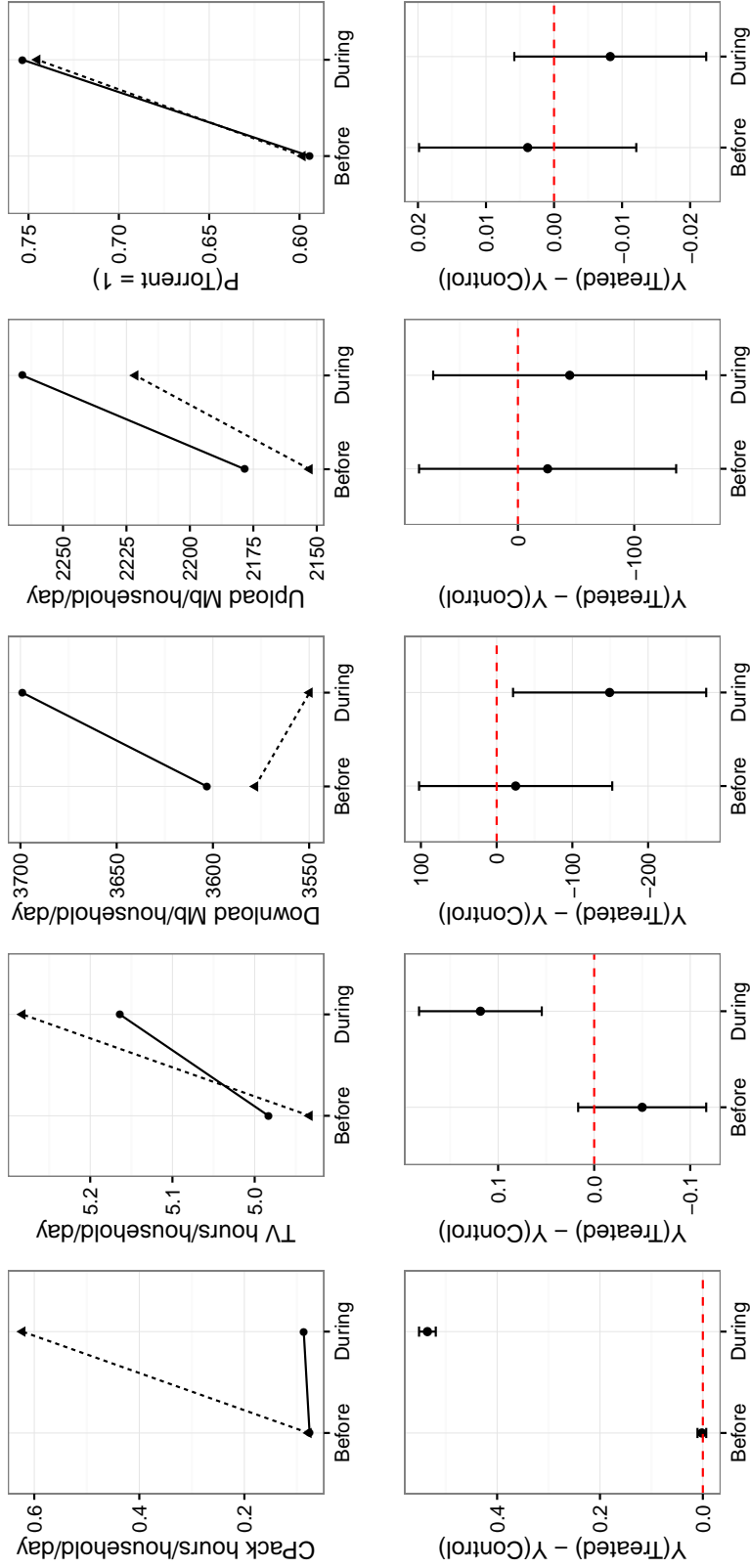


Figure 4: Model free comparison of key outcome variables across treatment and control households. The plots in the top row summarize (1) the time households spent watching the cinema pack, (2) the total time spent watching television, (3) download traffic, (4) upload traffic and (5) the proportion of 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 shows 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 pre-experimental controls to increase the precision of our estimators. Column (1) shows that, on average, treated households increased TV consumption by 4.6% ($p - value < 0.01$). Column (2) shows that this increase is associated with additional usage of the Cinema Pack. Columns (3) and (4) show that treated households decreased download and upload traffic by 4.2% and 4.5%, respectively ($p - value < 0.05$). 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 in 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 non user 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 due to 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 sub-section.

Table 4: Effect of treatment assignment on TV time, download and upload traffic during the experiment and on likelihood of using BitTorrent (Intention to Treat) in columns (1-4). Effect of treatment on the probability of using BitTorrent during the experiment (Intention to Treat) in columns (5-7)

	<i>Dependent variable:</i>						
	Log(TV)	Log(CPTV)	Log(Download)	Log(Upload)	All	Flag Torrent	TV Show
	<i>Pooled OLS</i>	<i>Pooled OLS</i>	<i>Pooled OLS</i>	<i>Pooled OLS</i>	<i>OLS</i>	<i>OLS</i>	<i>OLS</i>
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Treated	0.046*** (0.007)	0.254*** (0.005)	-0.042** (0.016)	-0.045** (0.022)	-0.005 (0.008)	-0.007 (0.009)	-0.002 (0.009)
Log(BExp. TV Time)	0.796*** (0.015)	0.066*** (0.004)	-0.003 (0.014)	-0.054*** (0.018)	0.028*** (0.006)	0.041*** (0.006)	0.014** (0.006)
Log(BExp. Download)	-0.004 (0.005)	0.003 (0.003)	0.875*** (0.014)	0.319*** (0.016)	0.003 (0.005)	-0.031*** (0.005)	-0.011** (0.005)
Log(BExp. Upload)	-0.002 (0.003)	0.001 (0.002)	0.006 (0.006)	0.552*** (0.009)	0.080*** (0.003)	0.097*** (0.003)	0.072*** (0.003)
BExp. Torrents	0.002 (0.001)	-0.0005 (0.001)	0.004* (0.002)	0.031*** (0.005)	0.005** (0.002)	0.019*** (0.005)	0.024*** (0.007)
Constant	-0.059 (0.039)	-0.142*** (0.018)	0.307*** (0.098)	-0.583*** (0.099)	0.201*** (0.029)	0.030 (0.030)	-0.058** (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: * p<0.1; ** p<0.05; *** p<0.01. Analysis pertains to the period during the experiment; Columns 1-4 have cluster Robust standard errors in (); Clustering at household level; Columns 5-7 have Heteroskedastic consistent standard errors in ()

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 towards 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 changes with fit. Columns (10)-(12) show a decrease in upload traffic of roughly 45% for households with a 100% fit. The $F - stats$ to test $treated + treated * offerfit = 0$ 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 5: Effect of treatment assignment on TV usage, Cinema Pack usage, download and upload traffic.

	<i>Dependent variable:</i>											
	Log(TV) <i>Pooled OLS</i>			Log(CPTV) <i>Pooled OLS</i>			Log(Download) <i>Pooled OLS</i>			Log(Upload) <i>Pooled OLS</i>		
	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*** (0.008)	0.048*** (0.009)	0.046*** (0.009)	0.253*** (0.006)	0.248*** (0.006)	0.245*** (0.006)	-0.037** (0.019)	-0.039* (0.020)	-0.043** (0.021)	-0.016 (0.024)	-0.014 (0.026)	-0.015 (0.027)
Treated×OfferFit	-0.117 (0.075)	-0.037 (0.091)	-0.0001 (0.090)	0.019 (0.045)	0.093* (0.053)	0.136*** (0.059)	-0.074 (0.172)	-0.032 (0.196)	0.027 (0.207)	-0.445** (0.198)	-0.432* (0.245)	-0.430 (0.272)
OfferFit	0.054 (0.054)	0.020 (0.064)	-0.007 (0.065)	0.012 (0.019)	0.015 (0.023)	0.029 (0.027)	-0.117 (0.114)	-0.226* (0.134)	-0.308** (0.144)	-0.105 (0.148)	-0.180 (0.183)	-0.079 (0.200)
Flag No Recs	-0.012 (0.009)	-0.011 (0.009)	-0.012 (0.009)	-0.001 (0.006)	0.003 (0.007)	0.006 (0.007)	-0.017 (0.021)	-0.025 (0.022)	-0.029 (0.022)	-0.115*** (0.027)	-0.123*** (0.028)	-0.112*** (0.028)
Log(BExp. TV Time)	0.796*** (0.015)	0.796*** (0.015)	0.796*** (0.015)	0.066*** (0.004)	0.066*** (0.004)	0.066*** (0.004)	-0.003 (0.014)	-0.003 (0.014)	-0.003 (0.014)	-0.051*** (0.018)	-0.052*** (0.018)	-0.052*** (0.018)
Log(BExp. Download)	-0.004 (0.005)	-0.004 (0.005)	-0.004 (0.005)	0.003 (0.003)	0.003 (0.003)	0.003 (0.003)	0.874*** (0.014)	0.873*** (0.014)	0.873*** (0.014)	0.317*** (0.016)	0.317*** (0.016)	0.318*** (0.016)
Log(BExp. Upload)	-0.003 (0.003)	-0.003 (0.003)	-0.003 (0.003)	0.001 (0.002)	0.001 (0.002)	0.001 (0.002)	0.006 (0.006)	0.006 (0.006)	0.007 (0.006)	0.547*** (0.010)	0.548*** (0.010)	0.549*** (0.010)
BExp. Torrents	0.001 (0.001)	0.001 (0.001)	0.001 (0.001)	-0.0005 (0.001)	-0.0004 (0.001)	-0.0004 (0.001)	0.004 (0.002)	0.004 (0.002)	0.004* (0.002)	0.030*** (0.005)	0.030*** (0.005)	0.030*** (0.005)
Constant	-0.056 (0.040)	-0.055 (0.040)	-0.053 (0.040)	-0.143*** (0.018)	-0.146*** (0.018)	-0.148*** (0.018)	0.331*** (0.101)	0.343*** (0.101)	0.348*** (0.101)	-0.505*** (0.103)	-0.500*** (0.104)	-0.520*** (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: * p<0.1; ** p<0.05; *** p<0.01; Robust standard errors in (). Standard errors clustered by household; Analysis pertains to the period during the experiment; TV time in hours per day, CPTV time in hours per day added the constant 1, Downloads and Uploads in Mega Bytes per day

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 $treated + treated * offerfit = 0$ for the likelihood of using BitTorrent in Columns (1)-(3). The corresponding $F - stats$ 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 anti-piracy 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 econometricians (Crump et al., 2008; Imai et al., 2013; Athey and Imbens, 2015; Athey, 2015; Wager and Athey, 2015) to predict heterogeneous treatment effects using results from randomized experiments. In this context, researchers are interested in computing uplift, defined as

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

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

Table 6: Effect of treatment assignment on the likelihood of showing up in BitTorrent logs

	<i>Dependent variable:</i>					
	Flag Torrent			Flag Movie Torrent		
	<i>LPM</i>					
	50 Recs.	100 Recs.	150 Recs.	50 Recs.	100 Recs.	150 Recs.
	(1)	(2)	(3)	(4)	(5)	(6)
Treated	0.007 (0.009)	0.008 (0.010)	0.008 (0.011)	0.004 (0.010)	0.008 (0.011)	0.011 (0.011)
Treated * Offer Fit	−0.184** (0.082)	−0.175* (0.090)	−0.182** (0.092)	−0.154* (0.083)	−0.198** (0.099)	−0.240** (0.104)
Offer Fit	−0.232*** (0.058)	−0.027 (0.064)	0.152** (0.065)	−0.043 (0.062)	0.394*** (0.075)	0.795*** (0.079)
Flg. No Recs	−0.213*** (0.011)	−0.195*** (0.012)	−0.179*** (0.012)	−0.295*** (0.011)	−0.257*** (0.011)	−0.224*** (0.011)
Log(BExp. TV Time)	0.033*** (0.006)	0.032*** (0.006)	0.032*** (0.006)	0.047*** (0.006)	0.047*** (0.006)	0.046*** (0.006)
Log(BExp. Download)	0.002 (0.005)	0.004 (0.005)	0.005 (0.005)	−0.029*** (0.005)	−0.025*** (0.005)	−0.022*** (0.005)
Log(BExp. Upload)	0.070*** (0.003)	0.071*** (0.003)	0.070*** (0.003)	0.081*** (0.003)	0.080*** (0.003)	0.078*** (0.003)
BExp. Torrents	0.002* (0.001)	0.003** (0.001)	0.003** (0.001)	0.016*** (0.005)	0.017*** (0.005)	0.017*** (0.005)
Constant	0.329*** (0.030)	0.296*** (0.031)	0.272*** (0.031)	0.174*** (0.030)	0.113*** (0.031)	0.069** (0.031)
F-Test:						
Treated×OfferFit+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:

*p<0.1; **p<0.05; ***p<0.01; Analysis pertains to the period during the experiment;
Robust standard errors in ();

using 5-fold 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 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 the 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, in which is 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-a-vis TV shows and, most importantly, with 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

Table 7: Treatment effect across quintiles of predicted uplift (household level).

	<i>Dependent variable:</i>				
	Flag Torrent				
	<i>Q1</i>	<i>Q2</i>	<i>Q3</i>	<i>Q4</i>	<i>Q5</i>
	(1)	(2)	(3)	(4)	(5)
Treated * During	−0.303*** (0.022)	−0.099*** (0.017)	−0.018 (0.016)	0.029 (0.021)	0.324*** (0.022)
Treated	−0.223*** (0.022)	−0.065*** (0.019)	−0.003 (0.019)	0.099*** (0.022)	0.219*** (0.020)
During	0.314*** (0.015)	0.155*** (0.011)	0.104*** (0.012)	0.141*** (0.017)	0.059*** (0.015)
Constant	0.608*** (0.015)	0.781*** (0.012)	0.768*** (0.014)	0.560*** (0.016)	0.230*** (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***

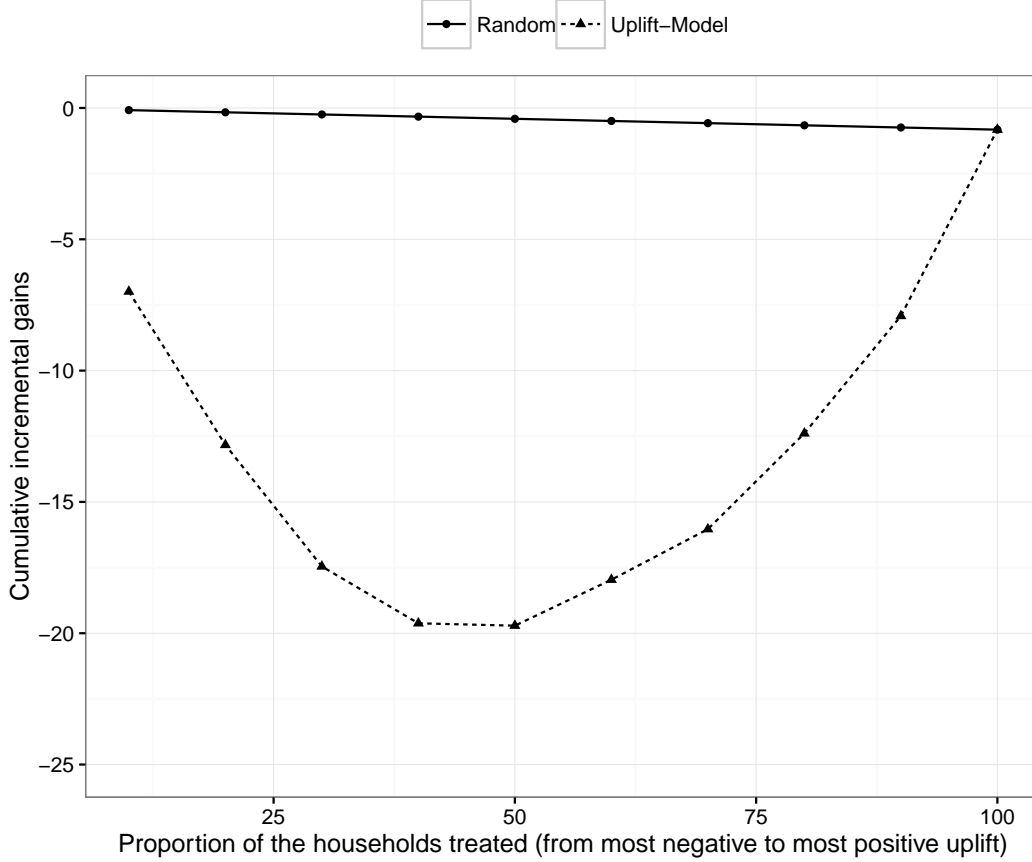
Household characteristics per uplift quintile before the experiment

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:

*p<0.1; **p<0.05; ***p<0.01; Robust standard errors in (); Standard errors cluster by household; Q1-5 are the quintiles of the “Uplift” distribution.

Figure 5: Qini curves showing heterogeneous treatment effects at the household level.



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 1 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 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 differ-

ences 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.

Table 8: Effect of treatment at the content-level during the experiment.

<i>Dependent variable:</i>						
Flag Torrent						
<i>Linear Probability Model (LPM)</i>						
	All	Movie	TV Show	All	Movie	TV Show
	(1)	(2)	(3)	(4)	(5)	(6)
Treated * During	−0.036*** (0.013)	−0.035*** (0.013)	−0.048* (0.025)	−0.036*** (0.013)	−0.035** (0.014)	−0.048* (0.027)
Treated	0.033*** (0.010)	0.033*** (0.011)	0.019 (0.014)	0.033*** (0.011)	0.033*** (0.012)	0.019 (0.014)
During	−0.001 (0.010)	−0.001 (0.010)	0.010 (0.010)	−0.001 (0.010)	−0.001 (0.011)	0.010 (0.010)
Constant	0.353*** (0.016)	0.360*** (0.016)	0.250*** (0.060)	0.118*** (0.007)	0.118*** (0.007)	−0.002 (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

Note: *p<0.1; **p<0.05; ***p<0.01; Robust standard errors in (); Standard errors clustered by IMDb id; Analysis includes the periods before and during the experiment

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 sub-section is similar to that employed in the previous sub-section. 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 6 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 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.

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

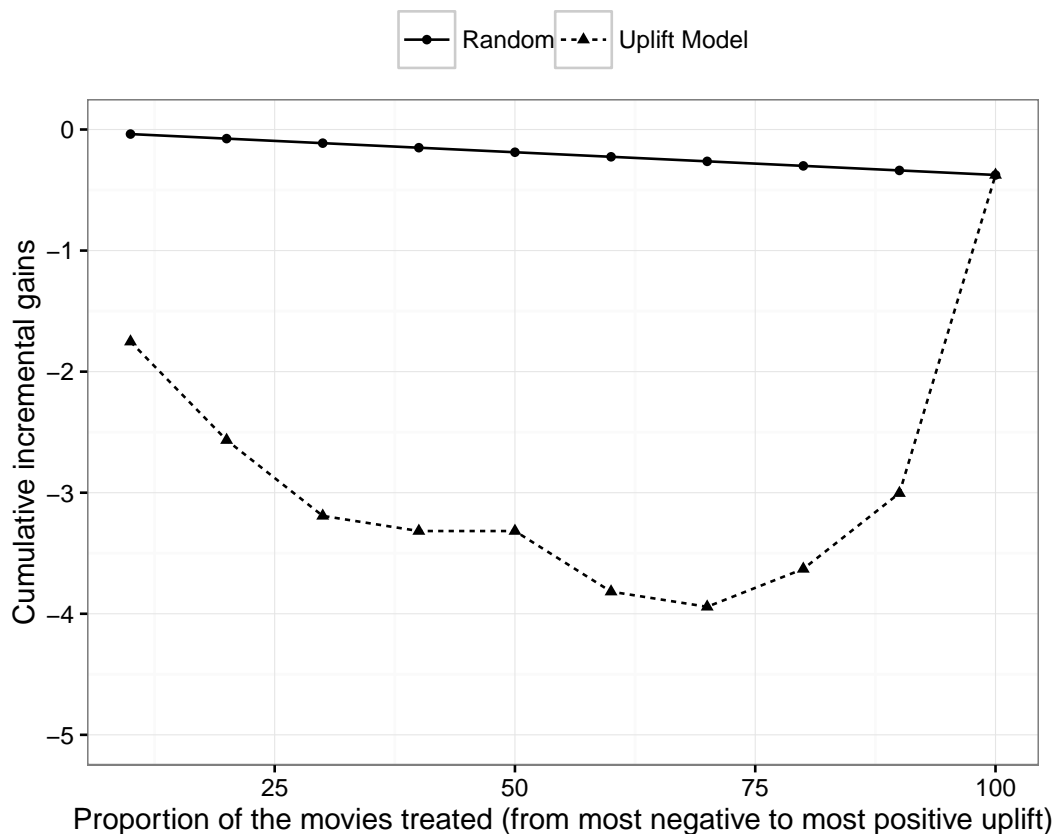
Table 9: Treatment effect across quintiles of predicted uplift (content-level).

	<i>Dependent variable:</i>				
	Flag Torrent				
	<i>Linear Probability Model (LPM)</i>				
	(1)	(2)	(3)	(4)	(5)
Treated * During	−0.205*** (0.032)	−0.042** (0.020)	−0.059** (0.025)	−0.010 (0.023)	0.143*** (0.037)
Treated	−0.026 (0.026)	−0.007 (0.017)	0.042** (0.021)	0.049** (0.020)	0.101*** (0.030)
During	0.101*** (0.024)	0.003 (0.019)	0.010 (0.018)	−0.020 (0.017)	−0.098*** (0.029)
Constant	0.625*** (0.034)	0.313*** (0.033)	0.264*** (0.030)	0.206*** (0.029)	0.342*** (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:

*p<0.1; **p<0.05; ***p<0.01; Robust standard errors in (); Standard errors cluster by content IMDb identifier

Figure 6: Qini curves showing heterogeneous treatment effects at the content level.



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 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 decreased this statistic by 0.1%. For the average wait time of 6 days in our dataset, 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.

Table 10: Evidence of household impatience.

	<i>Dependent variable:</i>	
	Flag Torrent (All)	
	<i>OLS</i>	<i>2SLS</i>
	(1)	(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***	-0.002***
	(0.001)	(0.001)
Flag No Recs	-0.152***	-0.153***
	(0.016)	(0.016)
Log(BExp. TV Time)	0.032***	0.033***
	(0.006)	(0.006)
Log(BExp. Download)	0.006	0.005
	(0.005)	(0.005)
Log(BExp. Upload)	0.070***	0.070***
	(0.003)	(0.003)
BExp. Torrents	0.003**	0.003**
	(0.001)	(0.001)
Constant	0.324***	0.337***
	(0.031)	(0.033)
Observations	10,225	10,225
R ²	0.177	0.175
Adjusted R ²	0.176	0.175

Note: *p<0.1; **p<0.05; ***p<0.01
Robust Standard Errors in ()
Wait time is measured in days

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 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 due to 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 in the US carried 5,601 titles in May 2016 of which 79% are movies. At the time, the smallest Netflix catalogs carried about 300 titles and are 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 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 0 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 due to the effect of the long tail, and decreases sharply for smaller licensing windows.

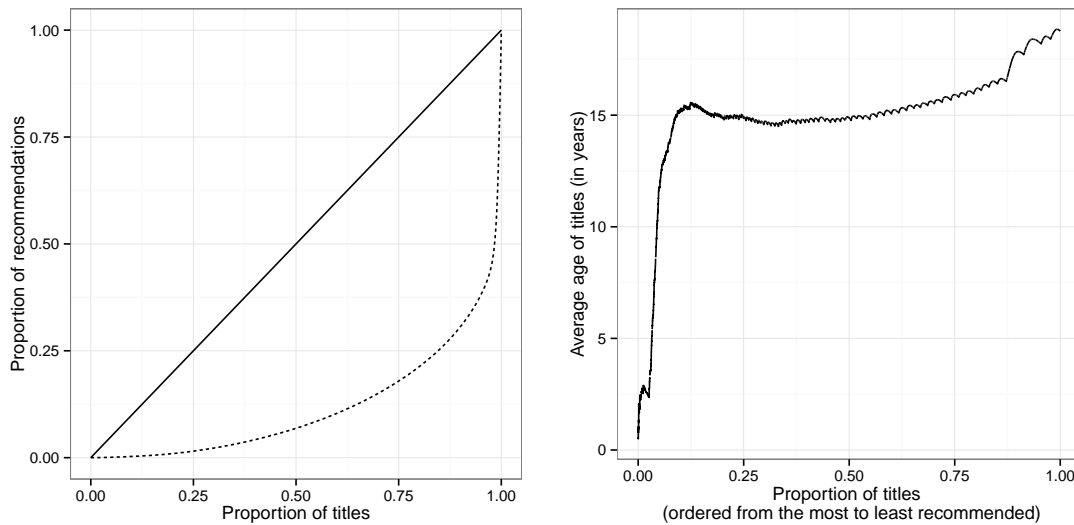


Figure 7: Concentration of recommendations over popularity and age. **Left Panel** displays Lorenz curve for the number of recommendations. **Righth Panel** displays the Average age of titles as more of them are included in the list of recommendations from the most to the least recommended

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 US (5,600 titles) would only yield an average fit of 50% with the top 15 recommendations to households in our sample and 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,

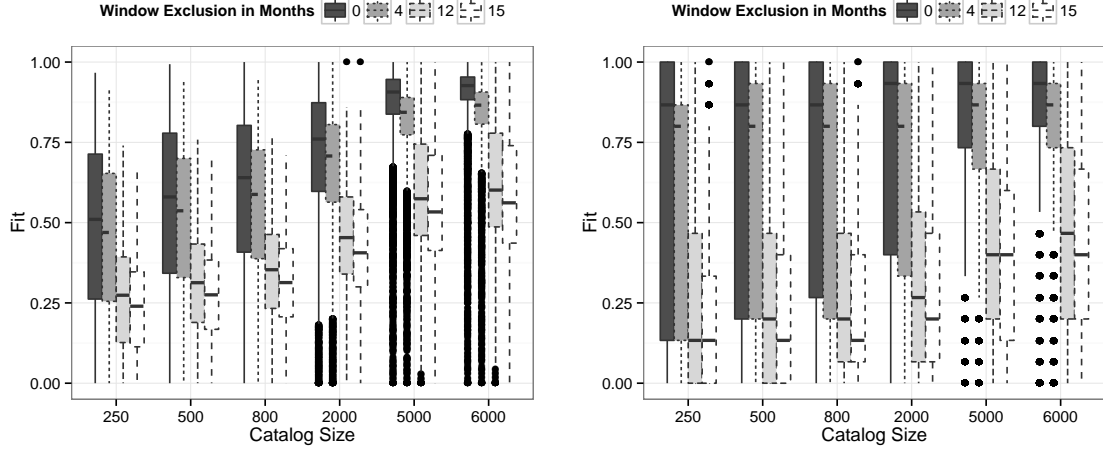


Figure 8: Distribution of the overlap between household recommendations and the content of a SVoD catalog as a function of catalog size and licensing window. **Left Panel** considers all recommended titles for all households in our sample; **Rigth Panel** considers only the top 15 recommendations for each household in our sample.

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, which is the price charged by TELCO for the Cinema Pack. We set the price of piracy to zero. The prices of TVoD+SVoD and of 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/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/month ($p\text{-value} < 0.1$), $-1.072 / -0.166 = \$6.5$ USD/month ($p\text{-value} < 0.1$) and $-1.081 / -0.166 = \$6.5$ USD/month ($p\text{-value} < 0.1$) 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 average willingness to pay for SVoD. For comparison, the Cinema Pack is offered by TELCO at \$13 USD/month and Netflix charges between \$7.99 and \$8.99 USD/month in the US. As the previous sub-section shows, on average pirates are likely to obtain at most 50% fit with a SVoD catalog as large as Netflix's in the US due to 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/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 will have on piracy consumption. During this

Table 11: Results obtained using a multinomial logit choice model for channel selection.

	<i>Dependent variable:</i>			
	Media Channel Choice			
	(1)	50 Recs (2)	100 Recs (3)	150 Recs (4)
Price	−0.166*** (0.004)	−0.166*** (0.004)	−0.166*** (0.004)	−0.166*** (0.004)
Piracy	1.117*** (0.032)	1.112*** (0.037)	0.946*** (0.040)	0.830*** (0.041)
Piracy * Treated	−0.044 (0.046)	0.007 (0.053)	0.020 (0.057)	0.014 (0.058)
Piracy * Treated * Offerfit		−0.798* (0.417)	−1.072* (0.567)	−1.081* (0.654)
Piracy * Offerfit		0.073 (0.305)	2.715*** (0.418)	4.817*** (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

Note:

*p<0.1; **p<0.05; ***p<0.01

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 a SVoD service. On average, we find that treated households increased TV consumption, 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 catered to the preferences of pirates. For example, a catalog as large as Netflix’s in the US (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/month for it.

The policy and managerial implications of our results are significant. We show that using legal channels to curtail piracy will not require 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 identify 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.

References

- Aguiar, L., Claussen, J., and Peukert, C. (2015). Online copyright enforcement, consumer behavior, and market structure. *Consumer Behavior, and Market Structure (March 30, 2015)*.
- Aguiar, L. and Waldfogel, J. (2015). Streaming reaches flood stage: Does spotify stimulate or depress music sales? Technical report, National Bureau of Economic Research.
- Angrist, J. D., Imbens, G. W., and Rubin, D. B. (1997). Identification of causal effects using instrumental variables. *Journal of the American statistical Association*, 91(434):444–455.
- Assmann, S. F., Pocock, S. J., Enos, L. E., and Kasten, L. E. (2000). Subgroup analysis and other (mis) uses of baseline data in clinical trials. *The Lancet*, 355(9209):1064–1069.
- Athey, S. (2015). Machine learning and causal inference for policy evaluation. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 5–6. ACM.
- Athey, S. and Imbens, G. (2015). Machine learning methods for estimating heterogeneous causal effects. *arXiv preprint arXiv:1504.01132*.
- Bai, J. and Waldfogel, J. (2012). Movie piracy and sales displacement in two samples of chinese consumers. *Information Economics and Policy*, 24(34):187–196.
- Belleflamme, P., Peitz, M., et al. (2014). Digital piracy: an update. Technical report, UCL.
- Belleflamme, P. and Picard, P. M. (2007). Piracy and competition. *Journal of Economics & Management Strategy*, 16(2):351–383.
- Belo, R., Ferreira, P., Godinho de Matos, M., and Reis, F. (2016). The impact of time-shift television on tv viewership behavior. *SI 2016 Economics of IT and Digitization*.
- Bhattacharjee, S., Gopal, R. D., Lertwachara, K., Marsden, J. R., and Telang, R. (2007). The effect of digital sharing technologies on music markets: A survival analysis of albums on ranking charts. *Management Science*, 53(9):1359–1374.

- Boag, J. (2004). 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, the. *Cal. WL Rev.*, 41:241.
- Bounie, D., Bourreau, M., Waelbroeck, P., and others (2006). Piracy and the demand for films: Analysis of piracy behavior in french universities. *Review of Economic Research on Copyright Issues*, 3(2):15–27.
- Breiman, L. (2001). Random forests. *Machine learning*, 45(1):5–32.
- Breiman, L., Friedman, J., Stone, C. J., and Olshen, R. A. (1984). *Classification and regression trees*. Chapman and Hall.
- Brynjolfsson, E., Hu, Y. J., and Smith, M. D. (2006). From niches to riches: Anatomy of the long tail. *Sloan Management Review*, 47(3):67–71.
- Chellappa, R. K. and Shivendu, S. (2005). Managing piracy: Pricing and sampling strategies for digital experience goods in vertically segmented markets. *Information Systems Research*, 16(4):400–417.
- Cremonesi, P., Koren, Y., and Turrin, R. (2010). Performance of recommender algorithms on top-n recommendation tasks. In *Proceedings of the fourth ACM conference on Recommender systems*, pages 39–46. ACM.
- Crump, R. K., Hotz, V. J., Imbens, G. W., and Mitnik, O. A. (2008). Nonparametric tests for treatment effect heterogeneity. *The Review of Economics and Statistics*, 90(3):389–405.
- Danaher, B., Dhanasobhon, S., Smith, M. D., and Telang, R. (2010). Converting pirates without cannibalizing purchasers: The impact of digital distribution on physical sales and internet piracy. *Marketing Science*, 29(6):1138–1151.
- Danaher, B. and Smith, M. D. (2014). Gone in 60 seconds: The impact of the megaupload shutdown on movie sales. *International Journal of Industrial Organization*, 33:1–8.
- Danaher, B., Smith, M. D., and Telang, R. (2013). Piracy and copyright enforcement mechanisms. Working Paper 19150, National Bureau of Economic Research.

- Danaher, B., Smith, M. D., and Telang, R. (2015). The effect of piracy website blocking on consumer behavior. *Available at SSRN 2612063*.
- Danaher, B., Smith, M. D., and Telang, R. (2016). Website blocking revisited: The effect of the uk november 2014 blocks on consumer behavior. *Available at SSRN 2766795*.
- Danaher, B., Smith, M. D., Telang, R., and Chen, S. (2014). The effect of graduated response anti-piracy laws on music sales: Evidence from an event study in france. *The Journal of Industrial Economics*, 62(3):541–553.
- Danaher, Brett, S. D. M. D. S. R. T. (2015). Understanding media markets in the digital age: Economics and methodology. In Goldfarb, A., Greenstein, S. M., and Tucker, C. E., editors, *Economic Analysis of the Digital Economy*, chapter 13, pages 385–406. University of Chicago Press.
- Frangakis, C. E. and Rubin, D. B. (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, J. H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of statistics*, pages 1189–1232.
- Geng, X. and Lee, Y.-J. (2013). Competing with piracy: A multichannel sequential search approach. *Journal of Management Information Systems*, 30(2):159–184.
- Groennings, K. (2005). Costs and benefits of the recording industry’s litigation against individuals. *Berkeley Technology Law Journal*, pages 571–601.
- Guelman, L., Guillaume, M., and Perrez-Marxed;n, A. M. (2015). Uplift random forests. *Cybern. Syst.*, 46(3-4):230–248.
- Hahsler, M. (2011). recommenderlab: A framework for developing and testing recommendation algorithms. *Working Paper*.
- Hearst, M. A., Dumais, S. T., Osman, E., Platt, J., and Scholkopf, B. (1998). Support vector machines. *Intelligent Systems and their Applications, IEEE*, 13(4):18–28.

- Hollis, S. and 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., et al. (2013). Estimating treatment effect heterogeneity in randomized program evaluation. *The Annals of Applied Statistics*, 7(1):443–470.
- Johar, M., Kumar, N., and Mookerjee, V. (2012). Content provision strategies in the presence of content piracy. *Information Systems Research*, 23(3-part-2):960–975.
- Kuhn, M. (2008). Building predictive models in r using the caret package. *Journal of Statistical Software*, 28(5):1–26.
- Liebowitz, S. J. (2008). Research note-testing file sharing’s impact on music album sales in cities. *Management Science*, 54(4):852–859.
- Linden, G., Smith, B., and York, J. (2003). Amazon. com recommendations: Item-to-item collaborative filtering. *Internet Computing, IEEE*, 7(1):76–80.
- Ma, L., Montgomery, A. L., Singh, P. V., and Smith, M. D. (2014). An empirical analysis of the impact of pre-release movie piracy on box office revenue. *Information Systems Research*, 25(3):590–603.
- Oberholzer-Gee, F. and Strumpf, K. (2007). The effect of file sharing on record sales: An empirical analysis. *Journal of political economy*, 115(1):1–42.
- Peitz, M. and Waelbroeck, P. (2006). Piracy of digital products: A critical review of the theoretical literature. *Information Economics and Policy*, 18(4):449–476.
- Price, D. (2013). Sizing the piracy universe. *NetNames, London*.
- Reimers, I. (2016). Can private copyright protection be effective? evidence from book publishing. *The Journal of Law and Economics*, 59(2):411–440.
- Resnick, P. and Varian, H. R. (1997). Recommender systems. *Communications of the ACM*, 40(3):5658.

- Rob, R. and Waldfogel, J. (2006). Piracy on the high cs: Music downloading, sales displacement, and social welfare in a sample of college students. *Journal of Law and Economics*, 49(1):29–62.
- Rob, R. and Waldfogel, J. (2007). Piracy on the silver screen. *The Journal of Industrial Economics*, 55(3):379–395.
- Rzepakowski, P. and Jaroszewicz, S. (2012a). Decision trees for uplift modeling with single and multiple treatments. *Knowledge and Information Systems*, 32(2):303–327.
- Rzepakowski, P. and Jaroszewicz, S. (2012b). Uplift modeling in direct marketing. *Journal of Telecommunications and Information Technology*, pages 43–50.
- Sarwar, B., Karypis, G., Konstan, J., and Riedl, J. (2001). Item-based collaborative filtering recommendation algorithms. In *Proceedings of the 10th international conference on World Wide Web*, pages 285–295. ACM.
- Simon, R. (1979). Restricted randomization designs in clinical trials. *Biometrics*, pages 503–512.
- Sinha, R. K., Machado, F. S., and Sellman, C. (2010). Don’t think twice, it’s all right: music piracy and pricing in a drm-free environment. *Journal of Marketing*, 74(2):40–54.
- Smith, M. D. and Telang, R. (2009). Competing with free: the impact of movie broadcasts on DVD sales and internet piracy 1. *MIS Quarterly*, 33(2):321338.
- Smith, M. D. and Telang, R. (2010). Piracy or promotion? the impact of broadband internet penetration on DVD sales. *Information Economics and Policy*, 22(4):289–298.
- Sundararajan, A. (2004). Managing digital piracy: Pricing and protection. *Information Systems Research*, 15(3):287–308.
- Suykens, J. A. and Vandewalle, J. (1999). Least squares support vector machine classifiers. *Neural processing letters*, 9(3):293–300.
- Swets, J. A. et al. (1988). Measuring the accuracy of diagnostic systems. *Science*, 240(4857):1285–1293.

- Telang, R. and Waldfogel, J. (2014). Piracy and new product creation: A bollywood story. *Available at SSRN 2478755*.
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 267–288.
- Train, K. E. (2009). *Discrete choice methods with simulation*. Cambridge university press.
- Ulin, J. (2013). *The business of media distribution: Monetizing film, TV and video content in an online world*. Focal Press, Taylor & Francis Group.
- Varian, H. R. (2005). Copying and copyright. *The Journal of Economic Perspectives*, 19(2):121–138.
- Wager, S. and Athey, S. (2015). Estimation and inference of heterogeneous treatment effects using random forests. *arXiv preprint arXiv:1510.04342*.
- Waldfogel, J. (2010). Music file sharing and sales displacement in the itunes era. *Information Economics and Policy*, 22(4):306–314.
- Waldfogel, J. (2012). Copyright research in the digital age: Moving from piracy to the supply of new products. *American Economic Review*, 102(3):337–342.
- Wilbur, K. C. (2008). How the digital video recorder (dvr) changes traditional television advertising. *Journal of Advertising*, 37(1):143–149.
- Wooldridge, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data*. The MIT Press, second edition edition.
- Wu, S.-y. and Chen, P.-y. (2008). Versioning and piracy control for digital information goods. *Operations Research*, 56(1):157–172.
- Zentner, A. (2006). Measuring the effect of file sharing on music purchases. *Journal of Law and Economics*, 49(1):63–90.
- Zhang, L. (2014). Intellectual property strategy and the long tail: Evidence from the recorded music industry. *Available at SSRN*.

A Sample Selection

The data used in this paper is a subset of a larger dataset that TELCO collected to study households' response to free trials of the cinema pack product. While our interest in this paper is to understand the behavior of the population of BitTorrent users, TELCO was also interested in learning if the average household would be more likely to subscribe to Cinema Pack after a full featured trial of the service.

To ensure their goals, TELCO used a stratified sampling to learn whether offering the new TV content would lead households to subscribe to the product afterwards, and to use less Internet data and reduce piracy.

With stratified sampling, the units of observation are split into stratum and are randomly assigned to the treatment and control in each stratum separately (Simon, 1979). This design allows TELCO to increase statistical power, in particular to the sub-population of pirates (Assmann et al., 2000) without compromising the generalizability of the analysis to the entire population of client households.

TELCO used data from April and May 2014 (before the experiment started) to build a classification algorithm to stratify a sample of households according to observable features that correlate to BitTorrent use. The Caret framework was used to train and evaluate the performance of different machine learning algorithms (Kuhn, 2008) on their ability to predict if a household would show up in future BitTorrent logs. All algorithms were trained and tested using 5 fold cross validation repeated 10 times. The outcome of this analysis is depicted in Figure 9. This figure shows that most models fit the data well. In particular, the Area Under the Curve (AUC) is near or above 80% in all cases. This threshold is usually used as rule of thumb to assume that a model is good for predictive purposes (Swets et al., 1988).

Variable selection is an integral part of gradient boosted model trees (GBM) (Friedman, 2001), random forests (RFOREST) (Breiman, 2001) and classification and regression trees (RPART) (Breiman et al., 1984). For Support Vector Machines with radial Kernel (SVM) (Hearst et al., 1998; Suykens and Vandewalle, 1999) and for the Logit model, feature selection is a separate step. These models were trained using LASSO (Tibshirani, 1996) for

feature selection. GBM was used to stratify the household sample because it yielded better performance scores in all the metrics usually used to gauge the predictive performance of these algorithms. Using the output of this model, the population classifier was constructed such that households with GBM scores above 50% were marked as pirates, while households with GBM scores below 50% were marked as non-pirates. Figure 10 plots the ROC curve we obtained. The black dot identifies the classifier used to stratify households.

After classifying households TELCO looked for whether they showed up in the BitTorrent logs. This allowed for creating four household strata. Households that were found in the BitTorrent logs were called “Confirmed Pirates” – C. Otherwise they were marked as non-pirates – NC. Households that the algorithm predicted as being pirates were called “Predicted Pirates” – P. Otherwise, they were marked as non-pirates – NP. Therefore, the four strata considered were (C,P), (C,NP), (NC,P) and (NC,NP). Figure 11 summarizes the features that the GBM algorithm used to classify households as pirates and non pirates. This figure shows that Internet upload traffic is the main determinant for this characterization, followed by how long ago households subscribed Internet service and by whether they have legacy or up-to-date equipment. We note that in this paper we end up analyzing only confirmed pirates, that is, strata (C,NP) and (C,P). Also we only focus on households with up-to-date equipment because households with legacy set-top-box devices cannot be tracked with respect to their television viewing habits.

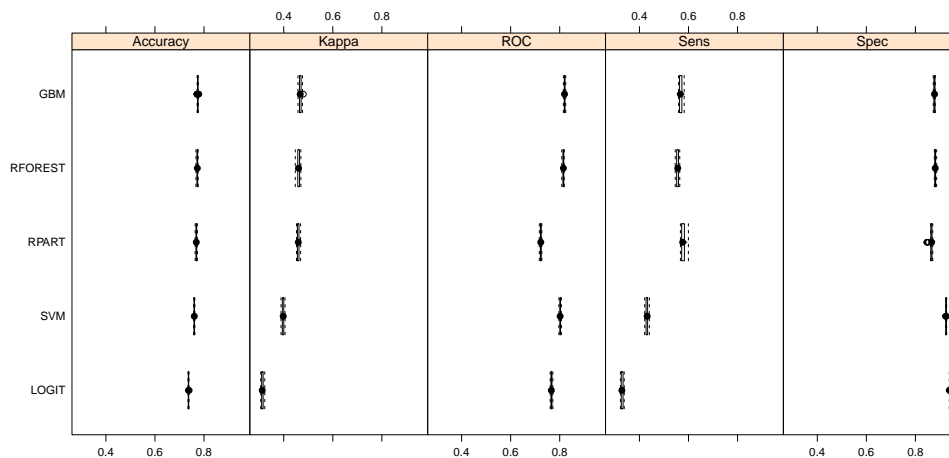


Figure 9: Performance of several machine learning algorithms used.

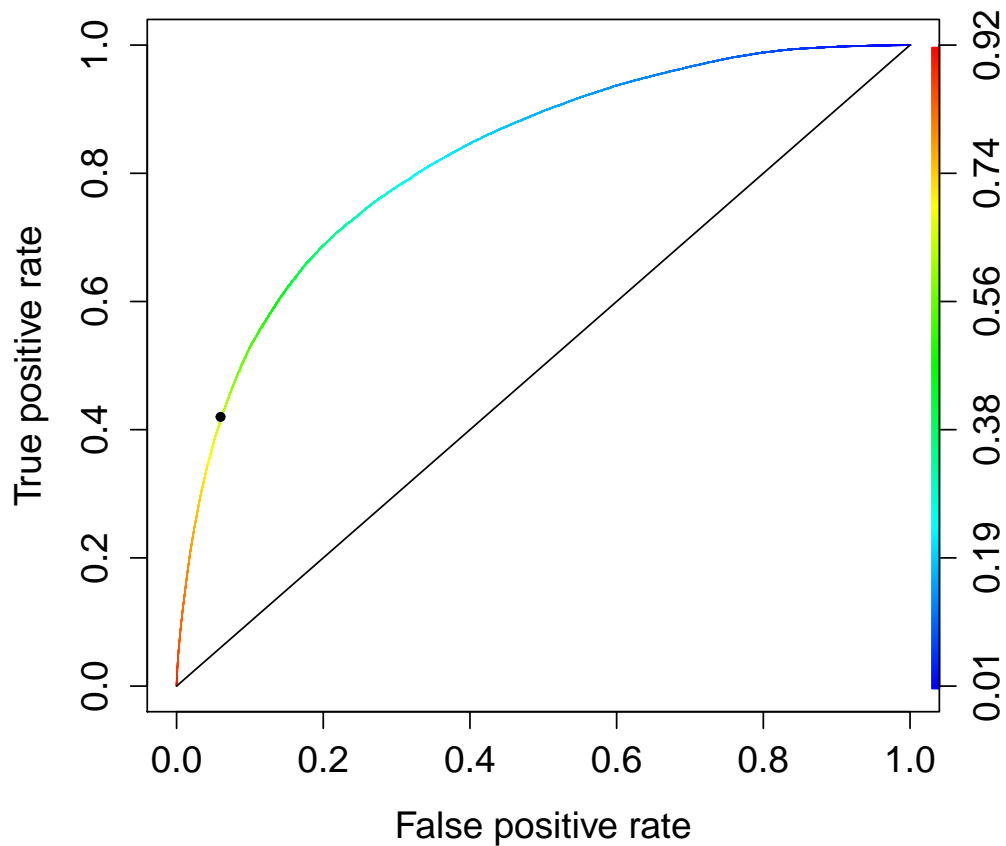


Figure 10: ROC chart for gradient boosted model trees algorithm.

Table 12 shows the average daily amount of traffic downloaded and TV usage per stratum in April and May 2014. 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 consumes 450 MB of bandwidth. 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%. Table 12 shows how many households would be needed in

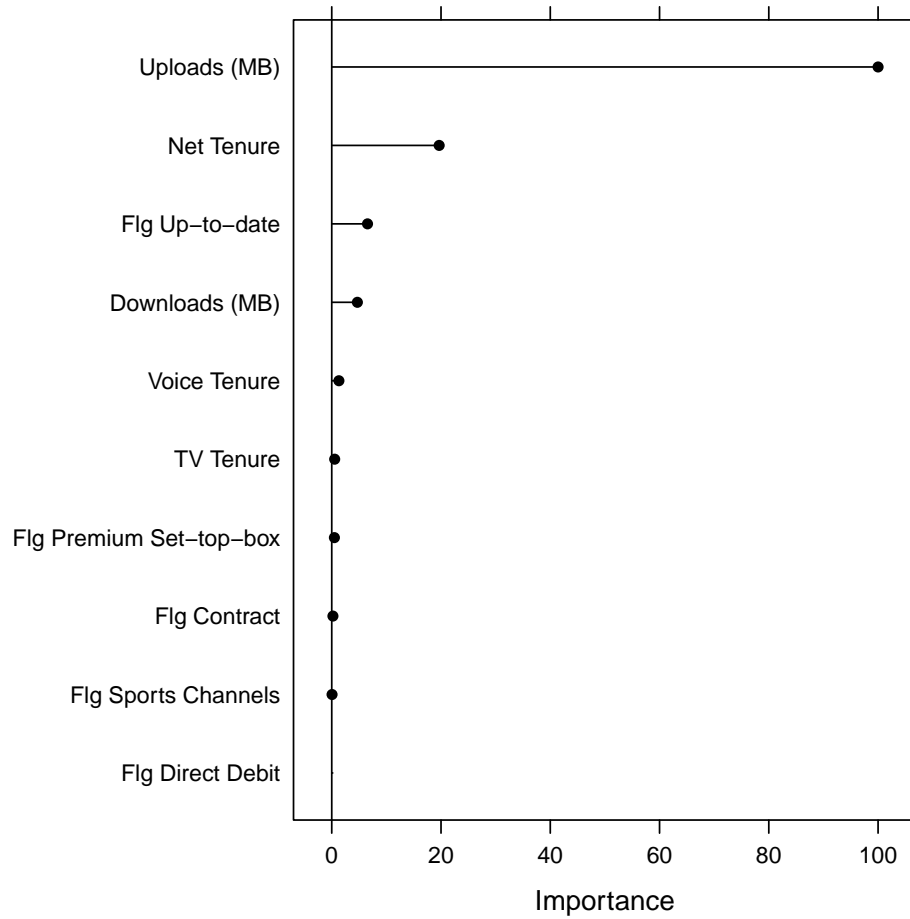


Figure 11: Variable importance plot for the gradient boosted model tree algorithm.

each stratum to obtain this level of power.

Table 12: Minimum sample size required to identify changes of 450MB in downloads per day and changes of 15 minutes of TV time per day with a 95% confidence level and 80% power, final sample size and number of treated households per stratum. Stratum statistics computed with data from April and May 2014

Stratum	Download (GB/Day)			TV (hours/Day)			Final Sample	
	Avg.	StDev.	Min Sample per treatment group	Avg.	StDev.	Min Sample per treatment group	All	Treated
NC,NP	1.2	2.0	311	4.4	2.4	1,329	6,107	3,077
NC,P	3.8	3.0	698	4.4	2.6	1,698	5,134	2,508
C,NP	2.2	2.4	447	4.4	2.9	2,113	4,307	2,153
C,P	4.4	6.1	2,885	4.4	2.5	1,570	5,918	2,963

To avoid running an underpowered experiment TELCO needed at least 3,057 treated households per stratum. In fact, and to account for potential practical problems that may arise during the experiment, TELCO randomly sampled 9,000 households in each stratum. Subsequently, treatment assignment within each stratum followed a simple randomized scheduled: half of TELCO’s households in each stratum were randomly assigned to receive the free gift.

From the initial sample of 36,000 households (9,000 per stratum), 7,590 households were removed from the because they had a legacy set-top-box equipment which can not be used to accurately track TV consumption. Another 3,270 households were removed from the sample because they opted out of marketing campaigns. 2,125 households were removed because they did not register a single day of TV or Internet usage during the experiment and another 1,549 households churned during the experiment.

A total of 21,466 households remained in the sample, distributed by strata as shown in the last two columns of Table 12.

The number of households in each strata was still well above the minimum threshold computed to identify the effect of interest.

In this paper we focus only on the population of confirmed pirates that include the 10,225 households in strata C,NP and C,P.

Table 13 shows that the experimental design described above achieved good balance in key observed household characteristics across treatment and control households in all strata. 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 at the 95% confidence level.

Table 13: Balance in observed household covariates across strata

	Treated		Control		T-test		
	Avg.	StDev	Avg.	StDev	Std. Effect	T-stat	p-value
NC,NP: Confirmed non-pirate, predicted non-pirate							
Pirate Score	0.181	0.135	0.183	0.135	-0.013	-0.492	0.623
TV tenure	86.878	60.930	87.588	60.957	-0.012	-0.454	0.650
Internet Tenure	56.289	35.253	57.165	36.007	-0.025	-0.959	0.338
Telephone tenure	49.037	21.923	49.151	21.792	-0.005	-0.204	0.839
Active Contract	0.824	0.381	0.818	0.386	0.016	0.633	0.527
Download (MB per day)	1, 217.236	1, 913.677	1, 242.850	2, 124.813	-0.013	-0.494	0.621
Upload (MB per day)	269.733	1, 531.998	229.026	862.270	0.033	1.280	0.200
TV Channels zapped	11.676	7.940	11.518	8.036	0.020	0.773	0.439
CPTV	0.055	0.256	0.056	0.240	-0.003	-0.125	0.901
TV	4.393	2.403	4.474	2.476	-0.033	-1.289	0.198
VoD streams	0.927	1.074	0.943	1.113	-0.015	-0.591	0.555
NC,P: Confirmed non-pirate, predicted pirate							
Pirate Score	0.763	0.065	0.764	0.064	-0.003	-0.121	0.904
TV tenure	99.347	55.084	99.905	55.204	-0.010	-0.362	0.717
Internet Tenure	75.900	34.593	75.829	33.590	0.002	0.074	0.941
Telephone tenure	60.441	17.468	60.789	17.241	-0.020	-0.717	0.474
Active Contract	0.808	0.394	0.809	0.393	-0.004	-0.151	0.880
Download (MB per day)	3, 770.558	4, 935.325	3, 738.642	5, 185.037	0.006	0.225	0.822
Upload (MB per day)	2, 331.399	5, 367.296	2, 229.316	4, 445.092	0.021	0.739	0.460
TV Channels zapped	11.809	8.427	12.094	8.339	-0.034	-1.213	0.225
CPTV	0.061	0.270	0.074	0.312	-0.046	-1.652	0.099
TV	4.402	2.601	4.390	2.429	0.005	0.181	0.857
VoD streams	0.983	1.112	0.950	1.085	0.030	1.064	0.288
C,NP: Confirmed pirate, predicted non-pirate							
Pirate Score	0.304	0.151	0.311	0.150	-0.044	-1.456	0.145
TV tenure	93.136	57.657	91.215	55.605	0.034	1.111	0.267
Internet Tenure	66.586	34.231	66.391	34.349	0.006	0.186	0.852
Telephone tenure	54.911	18.944	55.335	18.639	-0.023	-0.740	0.459
Active Contract	0.802	0.398	0.793	0.405	0.024	0.775	0.439
Download (MB per day)	2, 193.011	3, 156.323	2, 212.563	3, 544.914	-0.006	-0.191	0.849
Upload (MB per day)	806.465	2, 448.565	722.401	2, 329.261	0.035	1.153	0.249
TV Channels zapped	11.640	8.179	11.948	7.965	-0.038	-1.249	0.212
CPTV	0.063	0.280	0.067	0.311	-0.014	-0.452	0.651
TV	4.375	2.505	4.464	2.512	-0.035	-1.150	0.250
VoD streams	1.020	1.180	1.010	1.108	0.009	0.286	0.775
C,P: Confirmed pirate, predicted pirate							
Pirate Score	0.757	0.080	0.757	0.082	0.001	0.036	0.971
TV tenure	96.826	56.475	94.636	56.055	0.039	1.495	0.135
Internet Tenure	71.509	33.653	70.442	33.810	0.032	1.214	0.225
Telephone tenure	58.385	16.955	57.846	17.373	0.031	1.206	0.228
Active Contract	0.796	0.403	0.803	0.397	-0.019	-0.716	0.474
Download (MB per day)	4, 446.922	5, 292.187	4, 522.641	5, 917.408	-0.013	-0.518	0.605
Upload (MB per day)	3, 037.274	5, 766.615	3, 195.712	6, 039.995	-0.027	-1.030	0.303
TV Channels zapped	11.900	8.621	11.622	8.056	0.033	1.279	0.201
CPTV	0.070	0.284	0.069	0.284	0.005	0.211	0.833
TV	4.362	2.495	4.381	2.526	-0.007	-0.285	0.776
VoD streams	0.967	1.108	0.998	1.105	-0.028	-1.057	0.290

B IBCF Recommendation Technology

We adapted the Recommender Lab R package (Hahsler, 2011) to implement our Item-Based Collaborative Filtering (IBCF) algorithm (Sarwar et al., 2001). Item-to-item collaborative filtering matches each of the target users downloads to similar content, called items. Then it combines those similar items into a recommendation list. To determine the most-similar match for a given item, the algorithm builds an item-similarity table by finding items that customers tend to purchase together.

We note that we only use this algorithm to recommend movies and thus the proxy for fit that it provides applies only to movies. Unfortunately, we are unable to apply this algorithm to TV series because our torrent logs do not have episode level information. This means that we are able to know users preferences across series, but are unable to recommend specific seasons or episodes they should watch.

Figures 12 and 13 show the standard out-of-sample performance metrics used to evaluate the top-N recommendations for each household. As benchmarks we compare the output of our algorithm to that of: (1) a non-personalized algorithm based on item popularity which recommends the titles that are most shared using BitTorrent by our sample of households; (2) a personalized algorithm that provides recommendations at random.

The performance of our personalized IBCF algorithm is in line with the best results that are achieved in datasets of comparable complexity. In particular, they are in-line with the performance of the algorithms reported in Cremonesi et al. (2010) when applied to the Netflix and Movielens datasets. The latter have been repeatedly used to benchmark the performance of recommendation technologies in several academic and industry competitions. Similar to Cremonesi et al. (2010) we find that the performance of the non-personalized algorithm on the top-N recommendations is comparable to the performance of more sophisticated, personalized algorithms. However, the non-personalized popularity-based algorithm does not suit the goal of our exercise because it provides only (trivial) recommendations that capture the preferences of the average household. This, however, does not inform us about the preferences of each particular household in our sample, which is what we need in order to construct a measure of fit between each household’s preferences and the contents offered

as part of the Cinema Pack during the 45 days that it was available to treated households.

Figure 12: Precision and Recall of the Top-N recommendations generated by the models implemented.

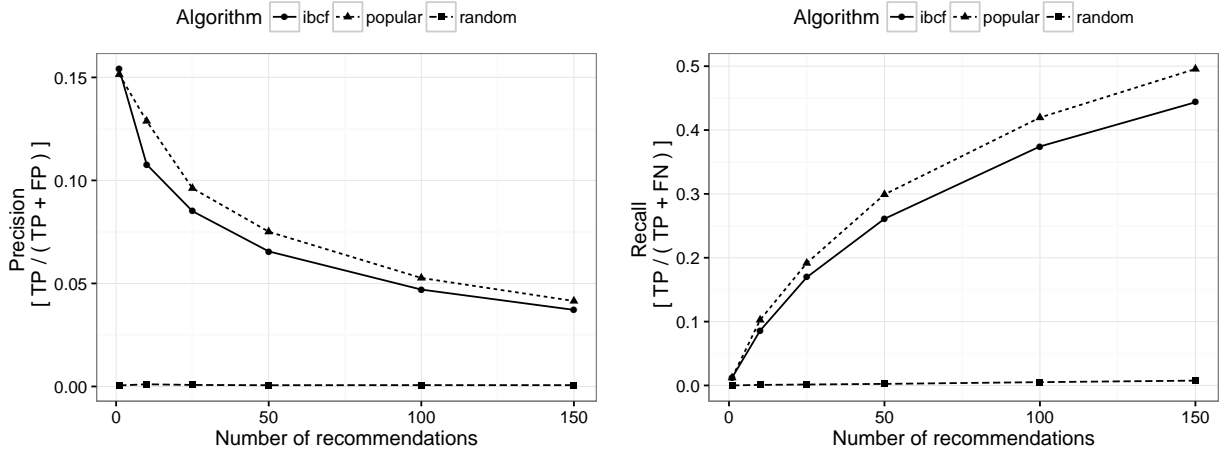
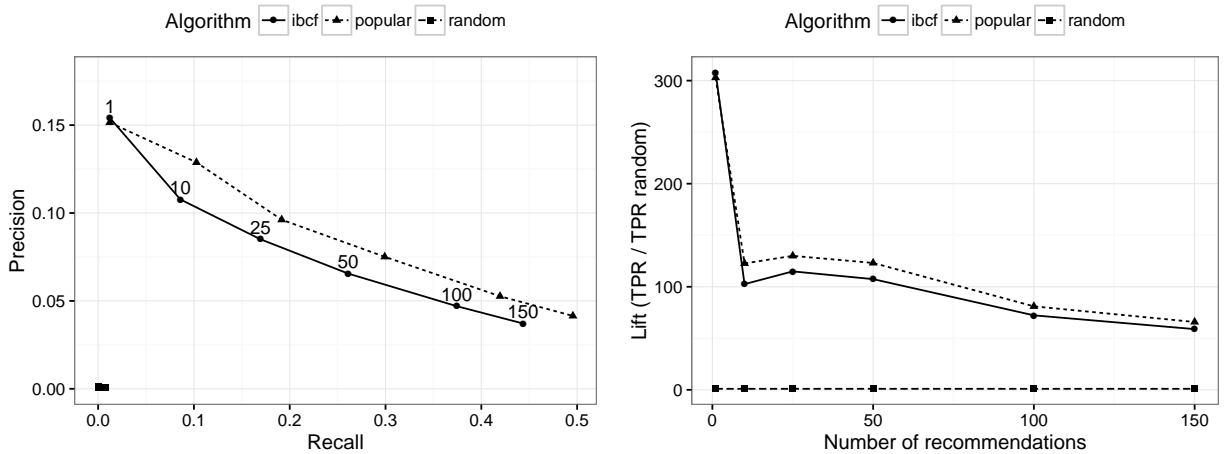


Figure 13: Left panel presents the Precision vs. Recall plot. The numbers on top of each point denote the size of the recommendation catalog issued. The right panel presents the Lift of the recommender systems implemented versus a random recommendations.



Finally, we determine the overlap between the set of titles recommended to households using our IBCF algorithm and the set of titles available via the Cinema Pack. Figure 14 shows the distribution of this overlap for the case of the recommender system that suggests popular items (those most shared content using BitTorrent by households in our sample

before the experiment took place). The average overlap is 10% for the case of a list with 100 recommendations. This means that, on average across households in our sample, the cinema pack includes 10 titles out of the 100 recommended by the recommender system. This figure highlights the very low variation in the overlap across households. In fact, the existing variation is solely explained by the fact that our recommender system removes titles that households downloaded before.

Figure 15 shows the distribution of the overlap for the case of our personalized IBCF recommender system. Similarly to the recommendation system based on content popularity, the average overlap is 10% for the case of a list with 100 recommendations. However, with the IBCF algorithm the range of the overlap is 0% to 78%, while with the non-personalized recommendation system the overlap never exceeds 25%. In short, the IBFC algorithm is able to recommend content to the tail of the distribution of preferences across households in our sample, while a popularity-based recommendation system does not.

Figure 14: Overlap between the top-N popular recommendations and the content offered as part of the Cinema Pack.

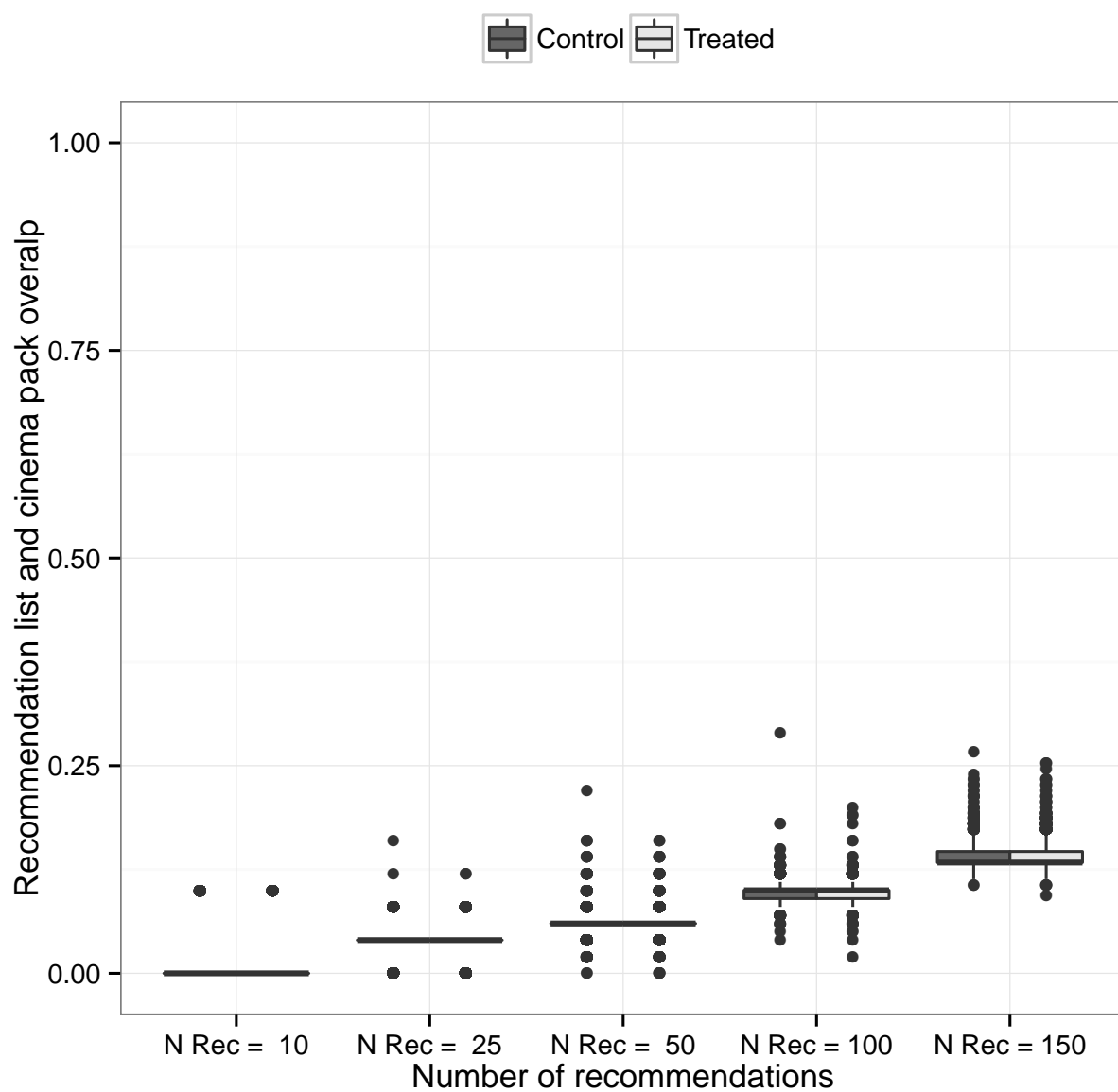
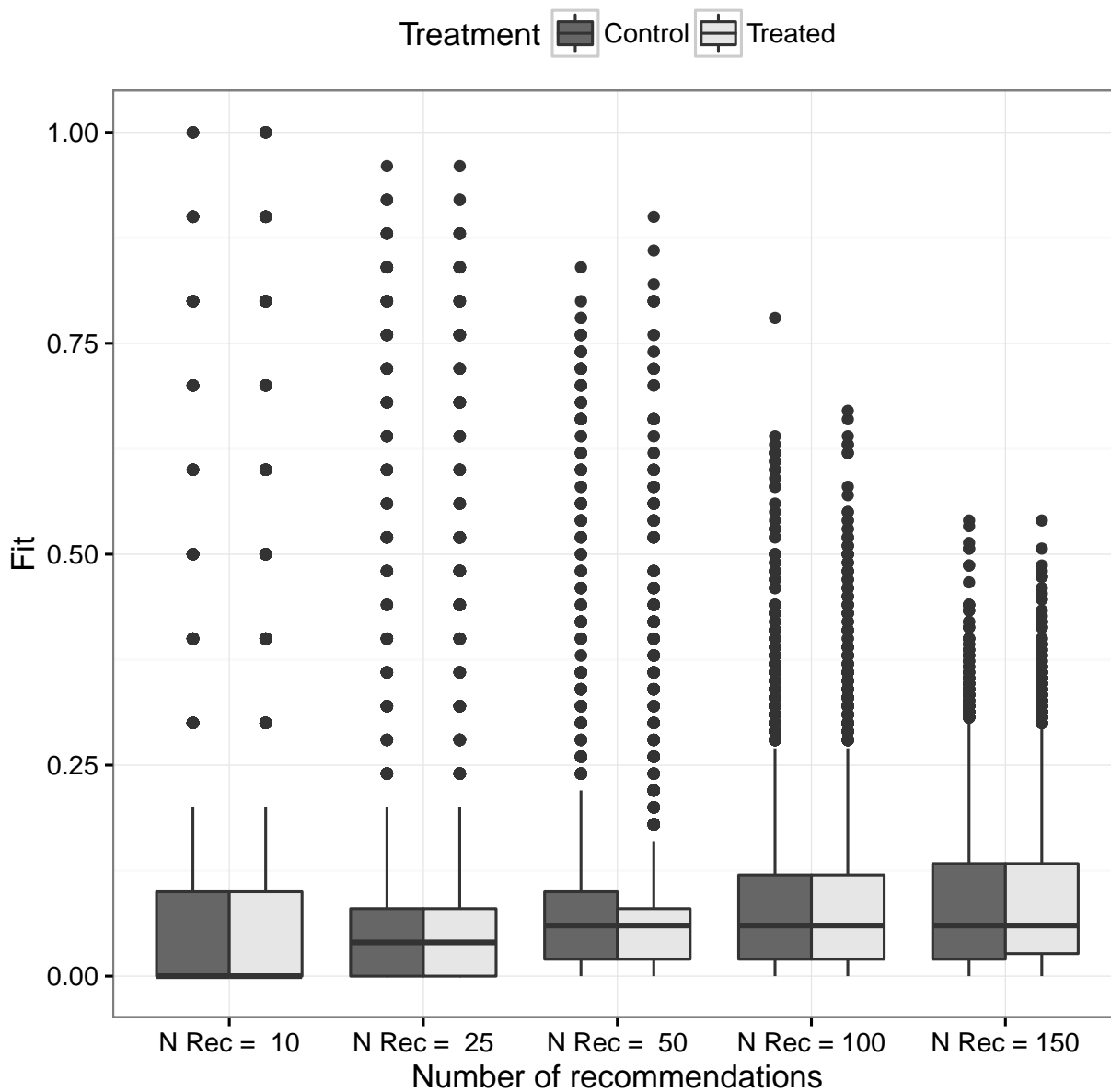


Figure 15: Overlap between the top-N IBCF recommendations and the content offered as part of the Cinema Pack.



C Local Average Treatment Effect

We define that a treated household complies with the treatment (and a control household does not) when the household uses the Cinema Pack for more than 90 minutes at least once during our experiment. Most movies in the Cinema Pack are 90 to 120 minutes long. Results using 20, 30, 60 and 120 minutes for the definition of compliance are qualitatively similar to those reported below and are available upon request. Figure 16 shows that across all strata in our sample, roughly 65% of the treated households used the cinema pack, compared to around 18% of the control households. Therefore, in our setting, using treatment assignment as an instrument for treatment compliance will yield the Local Average Treatment Effect (LATE) (Frangakis and Rubin, 1999; Hollis and Campbell, 1999). Table 14 shows the results obtained. In short, and in line with the main results in the paper, we find that the introduction of the Cinema Pack did not change the behavior of the average household in our sample, but that households whose preferences align well with the content offered as part of the cinema pack reduced their likelihood of using BitTorrent during the experiment. As expected, the magnitude of the effects reported in this table is larger than those reported in the main paper because the effect of the Intention-To-Treat averages out compliers and non-compliers. In this table, we report the effect for the sub-population of households in our sample that indeed changed their behavior due to using the Cinema Pack. We observe that among these households those whose preferences fit 100% with the content offered as part of the Cinema Pack reduce their likelihood of using BitTorrent during the experiment by more than 33%.

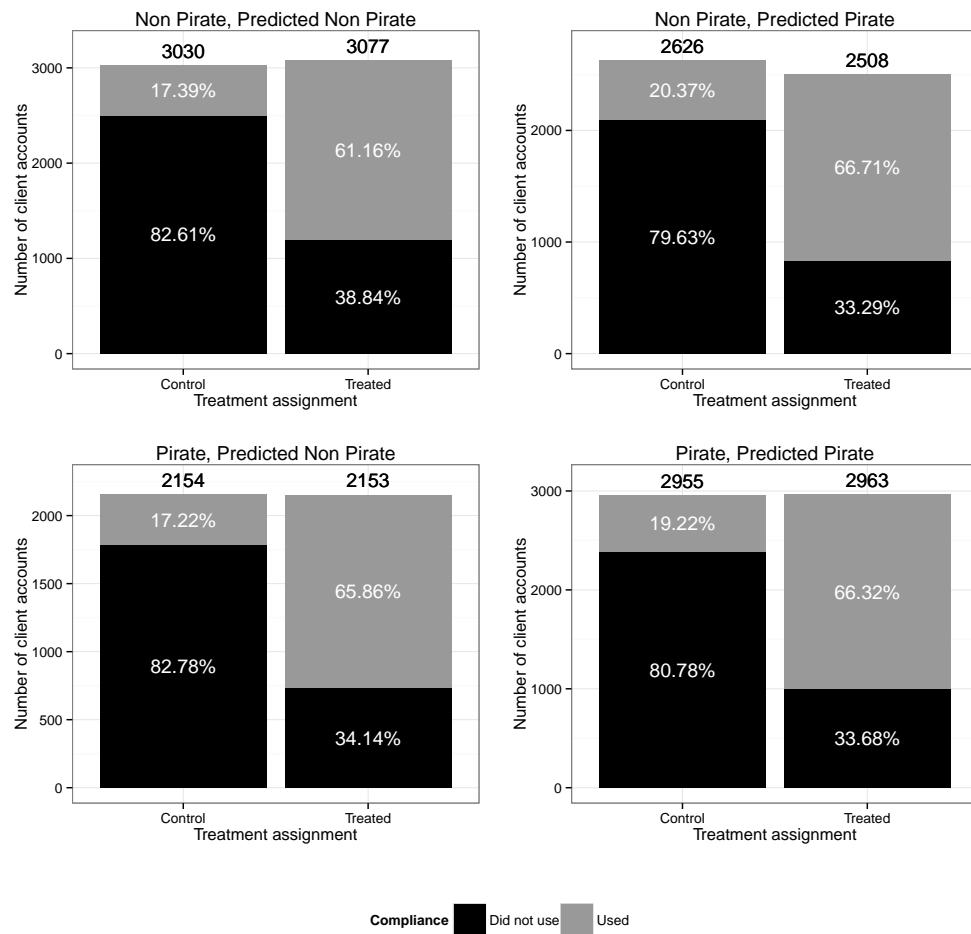


Figure 16: Compliance with treatment assignment in each stratum.

Table 14: The Local Average Treatment Effect (LATE) on BitTorrent usage.

	<i>Dependent variable:</i>					
	Flg. Torrent			Flg. Movie Torrent		
	<i>2SLS</i>			<i>2SLS</i>		
	50 Recs.	100 Recs.	150 Recs.	50 Recs.	100 Recs.	150 Recs.
	(1)	(2)	(3)	(4)	(5)	(6)
Used	0.016 (0.019)	0.016 (0.021)	0.017 (0.022)	0.008 (0.021)	0.016 (0.022)	0.022 (0.023)
Used * Offer Fit	-0.397** (0.179)	-0.352* (0.183)	-0.363* (0.186)	-0.334* (0.182)	-0.398** (0.201)	-0.476** (0.210)
Offer Fit	-0.154* (0.087)	0.038 (0.090)	0.220** (0.092)	0.023 (0.091)	0.467*** (0.103)	0.885*** (0.108)
Flg. No Recs	-0.213*** (0.011)	-0.195*** (0.012)	-0.179*** (0.012)	-0.295*** (0.011)	-0.257*** (0.011)	-0.224*** (0.011)
Log(BExp. TV Time)	0.034*** (0.006)	0.033*** (0.006)	0.033*** (0.006)	0.049*** (0.007)	0.047*** (0.007)	0.047*** (0.007)
Log(BExp. Download)	0.002 (0.005)	0.004 (0.005)	0.005 (0.005)	-0.029*** (0.005)	-0.026*** (0.005)	-0.023*** (0.005)
Log(BExp. Upload)	0.070*** (0.003)	0.071*** (0.003)	0.070*** (0.003)	0.081*** (0.003)	0.081*** (0.003)	0.078*** (0.003)
BExp. Torrents	0.002** (0.001)	0.003** (0.001)	0.003** (0.001)	0.016*** (0.004)	0.017*** (0.004)	0.017*** (0.004)
Constant	0.327*** (0.030)	0.294*** (0.031)	0.270*** (0.031)	0.173*** (0.031)	0.111*** (0.031)	0.066** (0.031)
Observations	10,225	10,225	10,225	10,225	10,225	10,225
R ²	0.180	0.175	0.175	0.190	0.192	0.200
Adjusted R ²	0.179	0.175	0.175	0.190	0.191	0.200
Residual Std. Error	0.393	0.394	0.394	0.447	0.447	0.444

Note:

*p<0.1; **p<0.05; ***p<0.01

Analysis pertains to the period during the experiment

Robust standard errors in ()