

# The Pirate Bay & Box Office Buccaneers

Klaus Ackermann\*   Wendy Bradley<sup>†</sup>   Jack Francis Cameron<sup>‡</sup>

June 30, 2020

A key concern about digital piracy and weak intellectual property rights protection is that they may reduce innovation by displacing legitimate sales. As most piracy is neither reported nor documented, research on the topic faces significant challenges in identifying changes to piracy and relies on event-studies to aid causal analysis. By introducing a unique dataset measuring piracy directly, and leveraging a positive shock to the supply of piracy, this study solves previous identification issues. It finds that the piracy of films, on average, displaces gross daily box office revenue by as much as 16%. However, the effect of piracy becomes heterogeneous when films are separated by unique characteristics, including genre and budget. At the extreme, whereas comedy films experience a 21% decrease in box office revenue, horror films experience a 34% increase in box office revenue with the introduction of a good quality copy of the film. We conclude that, for specific product characteristics and under certain value-capture strategies, piracy may increase box office sales. We believe these results shed light on potential incentives and directions for growth and innovation for industries confronted with piracy, with academic, managerial, and policy-related implications.

---

\*Department of Econometrics & Business Statistics and SoDa Laboratories, Monash Business School, Monash University, Monash University; email: klaus.ackermann@monash.edu

<sup>†</sup>Strategy, SMU Cox School of Business; email: wabradley@mail.smu.edu

<sup>‡</sup>Department of Econometrics & Business Statistics and SoDa Laboratories, Monash Business School, Monash University, Monash University

# 1 Introduction

Knowingly or unknowingly, at any given moment an Internet user is just a few clicks away from accessing copyright-infringing material. This concern has prompted substantial resource allocation and policy intervention on the part of firms and governments. Academic studies investigating the effects of piracy on paid consumption find evidence of revenue displacement (Danaher and Waldfoegel 2012, Smith and Telang 2012), aggravating concerns about the future of incentives for innovation. However, literature on the counterfeiting of physical goods and piracy of digital goods have highlighted the ability of certain products to benefit from illegal consumption through trial versions (Gosline 2009, 2010) or a discovery effect (Liebowitz 2006, Oberholzer-Gee and Strumpf 2010). The emergence of conflicting empirical results about whether illegal copies increase or lower the consumption of their legal duplicates raises new questions for management scholars operating in the digital economy. To what extent does piracy displace legitimate sales, and do all firms suffer equally?

This study analyzes if and under what conditions digital piracy cannibalizes sales in the context of the film industry. Few studies have examined the impact of counterfeit products on sales directly (Grossman and Shapiro 1988, Qian 2008). Qian (2014) examines the effect of counterfeit products by quality tier on sales for physical products in the fashion industry. However, little is known about the extent to which such models apply to the counterfeiting of digital goods. This study furthers the discussion by directly measuring the effect of piracy on sales of digital products across heterogeneous categories. Similar to the heterogeneity in the quality tiers of fashion products, divided by brand recognition, brand age, and the average age of consumers, movies differ widely in popularity when categorized, such as by genre and budget. This study contributes to the growing literature on piracy by identifying what product characteristics lead piracy to negatively impact box office sales through a substitution effect and under which circumstances piracy acts as a complement through a sampling or discovery effect. To investigate the heterogeneous impact of piracy on box office returns for film, this study compiles a unique dataset that matches data on pirated movies from a dominant piracy website, The Pirate Bay, between 2004–2018 to box office revenue and film data from the Internet movie database (IMDb), which includes film-specific characteristics. This study aims to disentangle the effect of piracy as a substitute or a complement to box office sales by classifying films into genre and budget and pirated copies into two quality tiers—bad and good quality. The two quality tiers are defined based on screen resolution. Like the quality of a counterfeit physical product, digital pirated copies of a film also differ in their substitutability to

the authentic version shown in cinemas; they are similar to copycat products in other industries such as mobile apps (Wang et al. 2018). To causally identify the impact of piracy on revenue, this study leverages random time variations before the online upload of a good quality copy as a shock to the supply of piracy. The base methodology leverages the panel data structure and a fixed-effects approach in order to isolate the impact of the release of a good quality pirated copy of a movie on the box office revenue of that film. The random variation in timing before such an upload serves as our natural experiment to causally identify the marginal impact of a positive shock to the supply of piracy on the box office revenue.

Consistent with current research, our findings show at a minimum a 16% decline in daily gross revenue because of the online release of a good quality pirated copy of a film. Surprisingly, however, this study finds heterogeneity in the effects of piracy on box office revenue by both genre and budget. First, given the film's genre, on average, crime and fantasy films remain relatively unaffected by piracy, horror films experience a 34% increase in box office revenue, and comedy films experience a 10% to 21% decline in revenue, according to the quality of the digital copy. Given the film's budget, high-budget films experience a sharp decline, at a significance level of 1%, in daily gross revenue following the upload of a pirated copy. Low-budget films experience an increase in daily gross revenue after the release of a good quality pirated copy, which provides some evidence for the discovery effect (Liebowitz 2006, Oberholzer-Gee and Strumpf 2010). We conclude that film studio strategies to create value exert intentional and unintentional effects on value capture. On the one hand, films that create value by emphasizing story, such as comedies and dramas, suffer the most from piracy, regardless of the quality of the digital copy. On the other hand, films marketed as experiential goods, where special post-production effects enhance their cinematic experience, such as horror or adventure genres, benefit the most from piracy; it provides a preview that raises an individual's willingness-to-pay for the box office experience. In other words, the movie theater experience is worth more for some genres and budget categories than for others, which has strategic implications for film studio executives. Finally, films of the same genre with overlapping box office release dates vary in the level of competition they face among each other. While comedies and animated films suffer in the box office after the release of the pirated copy, irrespective of its quality, the same-week release of a good quality pirated copy of another film of the same genre increases their box office sales by 8% and 34%, respectively. This can be attributed to the fact that some films compete in the box office as well as in the piracy medium because the piracy of one appears to lead consumers to view the other

in theaters. High-budget films, on average, are negatively affected by the release of a good quality pirated copy, but this loss may be offset by a slight benefit received from the release of a good quality copy of a competing film in the box office the same week. Low-budget films receive an additional boost in box office sales when another film of the same genre is pirated. One interpretation of this result is that low-budget films appeal to a different audience, typically a more cine-file one exploiting the "long tail" of movies, who may see many films in theaters and at home, and be more susceptible to the discovery effect induced by piracy.

Since film studios vary in their product mix, for example, the number of high- and low-budget films produced each year, and their distribution strategies, the heterogeneous impact of piracy on box office revenue by film by genre and budget categories yields additional managerial insights. For instance, studios producing low-budget films, or studios specializing in certain genres unaffected by or benefiting from piracy, may choose to eliminate the resources allocated to intellectual property rights (IPR) enforcement and invest them in producing more films. Alternatively, movie studios may choose to vary enforcement levels according to the susceptibility of a genre or budget category to revenue displacement by piracy. Most recently, with the COVID-19 pandemic, studios investigated the viability of premium video on demand (PVOD) release strategies, whereby film studios forgo theatrical releases and allow consumers to pay a rental fee to online platforms, such as Apple TV, to view newly released films at home. Three weeks after the PVOD release of the sequel 'Trolls World Tour' (Schwartzel 2020), Universal Studios earned \$100 million in PVOD rental fees, generating more revenue in the first three weeks of the film's release than that earned by the original film in its first five months at the box office. Movie studios are just beginning to experiment with such release strategies, and this study contributes to these strategies by identifying what genre and budget films are most popular for at-home viewing. These new release strategies by film studios may potentially encourage new and innovative products, such as uncommon or previously unpopular cross-genre movies, as less revenue is lost to piracy. This study has implications for creative industries with products available both physically and digitally, such as hard-cover books versus e-book versions and console video games versus online ones; these products have heterogeneous quality tiers and discrete distribution strategies, as well as a growing place in management literature (Hu and Smith 2018, Smith and Telang 2009).

A key contribution of this study is that it resolves the identification issues in previous research. Existing efforts make use of events causing a negative shock to the supply of piracy, such as the 2012 shutdown of the popular peer-to-peer website Megaupload

(Danaher and Smith 2014). By failing to observe other similar websites affected by the shutdown, such studies investigate only a partial restriction in consumption, and therefore act as a lower-bound for the impact of piracy. Our research overcomes this issue through a novel approach to identification; in this approach, we use the timing of the release of a good quality copy of a film to signify a random positive shock to the supply of piracy. Through this methodology, we infer the wider impact of piracy on sales, as we measure the change in daily revenue of a film after a substitute to its cinematic release becomes available. Thus, our event study and treatment capture the entire impact of digital piracy on box office revenue by using a proxy to the availability of content across the Internet. Finally, by focusing on identification, this study’s findings will unearth the potential for new incentives and directions for innovation, as well as inform IPR enforcement strategies for firms in creative industries, with implications for competition, management, and policy.

This paper is structured as follows. In section 2, we outline the background of piracy as a phenomenon. In section 3, we review the literature and the contribution of this paper. In section 4, we outline the methodology for collecting data and identify the treatment effect. In section, 5 we describe the results based on the empirical models. Finally, in section 6, we discuss the broader industry effects as well as the implications for management and policy.

## 2 Background

The increase in Internet speeds has allowed individuals with online access to immediately—or with only a short wait—download or stream content for consumption. Such developments have allowed widespread content access without permission or payment, otherwise known as digital piracy. One of the most popular methods of piracy<sup>1</sup> is torrenting, which uses peer-to-peer technology allowing computers to transfer files among one another. In this activity, an individual torrent acts like a key to accessing a downloadable file, and popular websites act as catalogues for all available torrents.

Piracy has catalyzed the evolution of numerous industries. The music industry evolved through the early days of Internet piracy stemming from Napster and has since continually improved the accessibility of music in its fight against piracy, with streaming services such as Spotify, Apple Music, and YouTube now being the most popular ways to listen to songs<sup>2</sup>. Although streaming services for films and TV exist at similar price-points to

---

<sup>1</sup>Statista: Media Piracy - Statistics & Facts 2018, see <https://www.statista.com/topics/3493/media-piracy/>

<sup>2</sup>Nielson Insights: US Music Mid-Year Report 2019, see <https://www.nielsen.com/us/en/insights/report/2019/u->

music, such as Netflix, Amazon, and Stan in Australia, film piracy continues to be a large and persistent illegal industry<sup>3</sup>. The emergence of the Internet has challenged, in general, many established business models (Greenstein 2015).

One possible explanation could be that while music become instantly accessible, with new albums being released directly onto streaming services, there is often a considerable delay between a box office film release and on-demand access. Though Netflix and Amazon immediately stream movies they produce themselves, titles made for the box office experience a much longer journey toward online availability. The resulting gap between the legal supply of films and their demand is filled by illegal piracy. There are several revenue points in a film’s supply chain that can be affected by piracy, namely cinema revenue, digital and physical distribution, and royalties through paid streaming services.

This study centers around the effects of piracy on cinema, or box office revenue, an issue that has been subject to much discussion in both the industry and the media. In 2012, theatrical box office revenue accounted for 36% of the overall industry revenue, second to home video at 42% and electronic home video at 21%<sup>4</sup>. However, even today, cinema revenue remains a vital and healthy component of the industry’s revenue, with 2018 setting a new record for the box office at \$41 billion gross earnings<sup>5</sup>.

## 3 Literature Review

### 3.1 The Effects of Piracy

The fundamental challenge facing researchers assessing the effect of piracy is identification; this is because most piracy is unreported and the association between physical sales and pirated downloads is predominately driven by unobserved heterogeneity across goods. To overcome this issue, studies utilize exogenous shocks to the supply of piracy, between country and city variations, product-level instrumental variable analysis, and survey data methodologies.

Peukert et al. (2017) employ the 2012 Megaupload shutdown as an exogenous shock to the supply of piracy, with an identification strategy comparing revenue of films available

---

s-music-mid-year-report-2019/

<sup>3</sup>The Cost of Movie Piracy (2005), see [https://www.wired.com/images\\_blogs/threatlevel/files/MPAstudy.pdf](https://www.wired.com/images_blogs/threatlevel/files/MPAstudy.pdf)

<sup>4</sup>PWC Global Entertainment and Media Outlook 2013, see <https://www.pwc.com/gx/en/global-entertainment-media-outlook/assets/2013/pwc-global-entertainment-and-media-outlook-2013-2017-a-guided-tour.pdf>

<sup>5</sup>MPAA 2018, see <https://www.motionpictures.org/wp-content/uploads/2019/03/MPAA-THEME-Report-2018.pdf>

on Megaupload to those not available. The study shows that the shutdown negatively affected the box office revenue, with the direction of the effect on a film's revenue depending on the breadth of the release; it also concludes that some niche films see increased revenue through piracy-induced word-of-mouth and discovery effects. In a previous approach, Peukert et al. (2013) finds similar results by comparing box office revenues to a matched control group of movies unaffected by the shutdown. Danaher et al. (2010) utilizes the removal of NBC content from iTunes, and its subsequent restoration, as an exogenous shock to the legal supply of the content; it finds that the removal increased the demand for pirated content via BitTorrent. Smith and Telang (2009) studies the impact of piracy on DVD sales using free-to-air broadcasts as an exogenous shock; the results show that pirated content does not cannibalize sales.

Danaher and Waldfogel (2012) and Smith and Telang (2012) leverage cross-country variation when looking at the impact of differential release times on DVD sales and box office revenue, respectively; both studies find that increased release times correspond to high piracy levels and lower revenue. Danaher and Smith use cross-country analysis for looking at the Megaupload shutdown, finding evidence of sales displacement from piracy. Hui and Png (2003), Peitz and Waelbroeck (2004), and Zentner (2005) leverage international panel data on music sales and pirated downloads and find a displacement of paid consumption. Aguiar and Martens (2016) uses cross-country variation for assessing sales displacement from online music consumption, finding no evidence of such displacement.

Survey data of the paid and unpaid consumption habits of university students in the US and China by Rob and Waldfogel (2007) and Bai and Waldfogel (2012), respectively, show the unpaid consumption of films displaces paid consumption.

### **3.2 Response to Piracy**

Studies assessing the impact of various response measures predominantly focus on governmental implementations rather than on competitive strategies at the firm level. Poort et al. (2014) examine the effectiveness of Dutch ISPs in blocking The Pirate Bay and related sites; the study showed the effectiveness of the shutdown in reducing net piracy. Danaher and Smith (2014) also reviews the impact of website blockings from ISPs, finding a small reduction in piracy and evidence of user migration to other unblocked piracy websites. In a revisited study the following year, Danaher et al. (2016) finds that website blocking has a tangible but diminishing impact as more sites become blocked. Watson et al. (2017) examine an increase in publicized litigation and copyright law enforcement via a two-part survey, which finds that the benefit of piracy outweighs the perceived risk of

pirates and that illegal consumption is not affected by an increased risk perception. Bhattacharjee et al. (2007) studies the impact of legal threats on individual pirates throughout 2003 and finds a decline in file-sharing activity, but no change in the supply. Vernik et al. (2011) argues that the use of digital rights management (DRM) controls may increase piracy as legal users find it more difficult to access content or software.

The literature on response measures to piracy illustrate mixed results, with the restriction of the supply of pirated goods presented as the most effective measure. The literature largely ignores firm-specific responses, such as innovation in business models, new product offerings, or alternative value-capture strategies as solutions to revenue displacement.

## 4 Methodology

### 4.1 Data

The central dataset used in this study was collected by The Internet Archive; it contains all available torrents from The Pirate Bay (TPB) between January 2004 and December 2018<sup>6</sup>, with over 11 million individual torrents. Although not the only one, TPB is one of the oldest BitTorrent sites and was the most visited torrent directory from 2004 to 2018<sup>7</sup>. TPB's popularity and dominance can be attributed to the ease at which it allows the downloads and uploads and evades a shutdown by authorities; these aspects make it a convincing proxy for the wider dissemination of illegal digital content.

For each unique upload of a torrent, unstructured information is collected on:

- List of files included in torrent
- Details of the listing (Including size/type/language)
- Description (Written by uploader)
- Comments (Written by other TPB users)

In combination, structured data on box office revenues across time was collected from the Comprehensive R Archive Network (CRAN) package 'boxoffice' built for R. All Box

---

<sup>6</sup>The data until 2013 can readily be downloaded, for the subsequent dates a script is provided by the archive.

<sup>7</sup>Alexa The Web Information Company: Top Ranked Torrent Directories and Trackers, see [https://www.alexa.com/topsites/category/Top/Computers/Internet/File\\_sharing/BitTorrent](https://www.alexa.com/topsites/category/Top/Computers/Internet/File_sharing/BitTorrent)



Office information comes neatly packaged; the daily data was provided by [www.the-numbers.com](http://www.the-numbers.com) for our time-period to match our Pirate Bay data. We record daily information that includes the following:

- Film title
- Distributor
- Gross earnings
- theaters
- Total gross earnings (cumulative)
- Days since the premiere

Full data tables can be found in the appendix.

Our data is stored in structured and unstructured text files with varying degrees of linkability. There exists no unique identifier of a film, as every user is free to upload files, resulting in many different spellings and iterations. A significant challenge of this research project is data identification. This study correctly matches films to a unique identifier allowing across time comparisons as well as facilitates joint analysis with the box office revenue dataset.

As such, the methodology for this project is necessarily broken into two primary sections—data collection and identification strategy.

## 4.2 Empirical Strategy

### 4.2.1 Data Collection

As our data source and approach are novel, we have invested significant efforts in preparing the data for use. The unstructured nature of the bulk of our information forces a careful and strict approach to cleaning in order to ensure the robustness and integrity of our study. The preparation process is detailed below.

The 'Pirate Bay' archival project has been ongoing for a number of years at The Internet Archive. Methods and tools have been developed to scrape the platform and strip torrent listings of the website code to get clean metadata on the aforementioned variables. To this end, the team has been successful. This dataset entailed some command-line

procedures to parse through and extract the relevant information into a database, and bash scripting was used for this purpose.

An example of the information contained in one of the detailed BitTorrent metadata files is shown below.

*Title, Type, Files, Size, IMDB, Spoken Languages, Texted Languages, Tags, Quality (+), Quality (-), Uploaded, By, User Type, Seeders, Leechers, Info Hash, Picture, Capture Date Along. Came. Polly. DVDRiP. XviD-BRUTUS, 201,55,721.13 MiB (756163764 Bytes),http://www.imdb.com/title/tt0343135/,,,1,0,2004-05-18T20:11:30Z,Ph4z3,,4,3,F7CAFCB88A025AC57D6B30ED73E89712BA5FB1B7,,2012-09-15T13:54:12Z*

Our intention is to map the effect of good quality torrents on the box office revenue and note the heterogeneity across films by genre and budget. Such data is available on the IMDb website. We use the IMDb identifier to link different torrents to the corresponding movie available in the torrent description. The IMDb dataset has the international title as a field, which we use to merge torrents with the box office data. As the IMDb link was primarily used for matching films to torrents, this analysis focuses on the subset of TPB's archive, which includes IMDb links in the torrent description. This procedure also removes all TV series or other movie material not designed for public consumption in cinemas.

Extending this core dataset, we motivate two additional variables: Good Quality Dummy and Bad Quality Dummy. Good quality copies of a movie are high resolution versions that can be considered close, if not perfect substitutes to the box office version of the same movie. Good Quality Dummy is 0, until a copy of the film is uploaded with a particular keyword ("720p", "1080p", "x264", "bluray") in the title, and it is 1 thereafter to denote the upload of a good quality copy of the film. Bad quality copies of a movie are not high-definition, may have foreign language dubbing or subtitles, may be ripped from a lower resolution screener copy, or consist of grainy footage captured from a camcorder smuggled into the movie theater. Bad Quality Dummy is motivated simply as the first upload of a film to TPB. These definitions are fairly restrictive. An example of the condensed Pirate Bay metadata after filtering for IMDb links can be found in the appendix in Table 14.

Budget data were available for most movies, and hence we restricted our sample to movies with quoted USD budget estimates for consistency. On occasion, film genres yield a description of more than one genre. In case of multiple genres, no single genre can be considered the primary descriptor of the film. Therefore, we include the film in multiple genre classifications. Hence, for example, IMDb classifies Tom Cruise's movie The Edge

of Tomorrow (2014) as action and sci-fi. The current study will use this film to investigate the effect of piracy on both action and sci-fi movies.

A possible limitation to our approach is that it is impossible to know if the contents of each are as described and not, in fact, a different film or product altogether. However, an uploader has little to gain from misrepresenting a torrent; this is because TPB community upholds community-enforced standards for online behavior via voting and commenting. Therefore, it is reasonable to assume that this is unlikely to occur at a large-enough scale to confound our results, and hence we make no attempt to address it.

### 4.2.2 Identification Strategy

Our unique approach to identification exploits the variation in timing between the upload of a bad vs. good quality torrent. A timeline is shown in Figure 1, in weeks, illustrating a theoretical example of the time taken for films to transition from the box office release to wide consumer release domestically. An extension to this timeline might integrate international releases, which precede or follow a US domestic release. In today’s time, a film runs in the cinemas for an average of 4 weeks<sup>8</sup>, with successful films running for a longer period. Notwithstanding a leak or other online screening events, good quality copies of films appear online in the weeks approaching the official DVD release, after the box office release and the start of online sales of the film. As such, it is an unexpected random event during which a good quality copy is uploaded online while it is still being shown in the box office. We use such a random event as a positive shock to the supply of piracy, and thus treat it as a natural experiment to be analyzed.

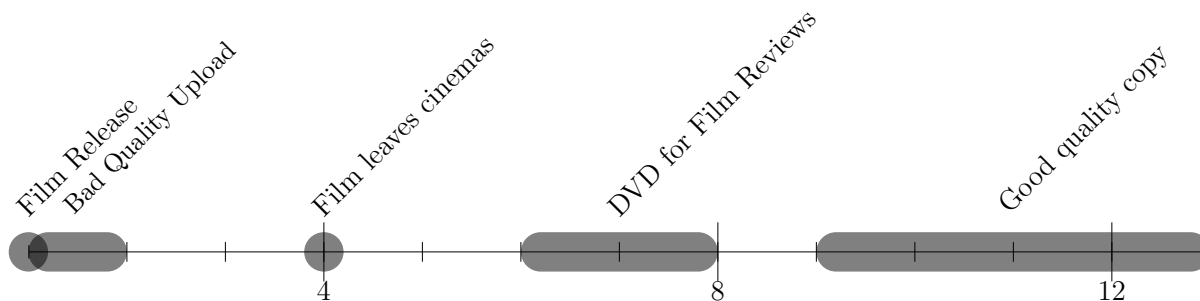


Figure 1: Theoretical timeline for the expected upload of a torrent

To this end, in our reduced dataset, we view the trends shown in Table 1.

<sup>8</sup>CNBC: Why movies are sometimes here and gone in theaters, see <https://www.cnbc.com/2015/11/17/why-movies-are-sometimes-here-and-gone-in-theaters.html>

Table 1: Average days to upload by torrent quality

	Good Quality	Bad Quality
Average Days	29	17
Median Days	15	9

Our central assumption is that the timing of the upload of a good quality torrent is a random event. This assumption is grounded in the idea that individuals are unable to obtain good quality copies of films, except through chance, be it leaks or otherwise, and the timing of the occurrence of this chance is unpredictable. A relevant concern is whether an individual obtains a copy through other means, which would allow for latent film demand to drive the uploads of good quality copies of torrents. As such, we investigate the relationship between the timing of a torrent upload of a film and its underlying demand. If such a relationship existed, and more popular films were uploaded faster, any findings would be biased due to the omission of a variable capturing this demand. We motivate budget as a proxy for a film’s demand; this is because a larger budget leads to a higher marketing spend, which grows potential viewership, awareness, and demand for the product. To this end, we find no relationship between the demand for a film and the timing of its release, as viewed in Figure 2, which indicates that more popular films have an equal likelihood as less popular titles of being uploaded at a given time since release.

In our empirical strategy, we extend our central analysis to allow for differential effects from film characteristics. In proceeding with this identification, each torrent, and thus film, is classified as belonging to a set of genre categories.

We can view the proportion of films with a good quality torrent as a function of days since film release, as in Figure 3. This illustration implies that while there is a small disparity in the proportions of good quality films available with respect to the total number of films in the box office at a particular day since release, there are no dramatic differences between genres in regard to slope and there are minor differences in intercept in the action genre. The proportion plot illustrates the heterogeneity between films, which will be explored and accounted for in model extensions.

## 5 Results

A consistent feature of box office revenue is its concentration in the early weeks of a film’s release. Time series plots of average daily gross revenue across films in Figure 4 illustrates

## Days to upload of good quality film by budget

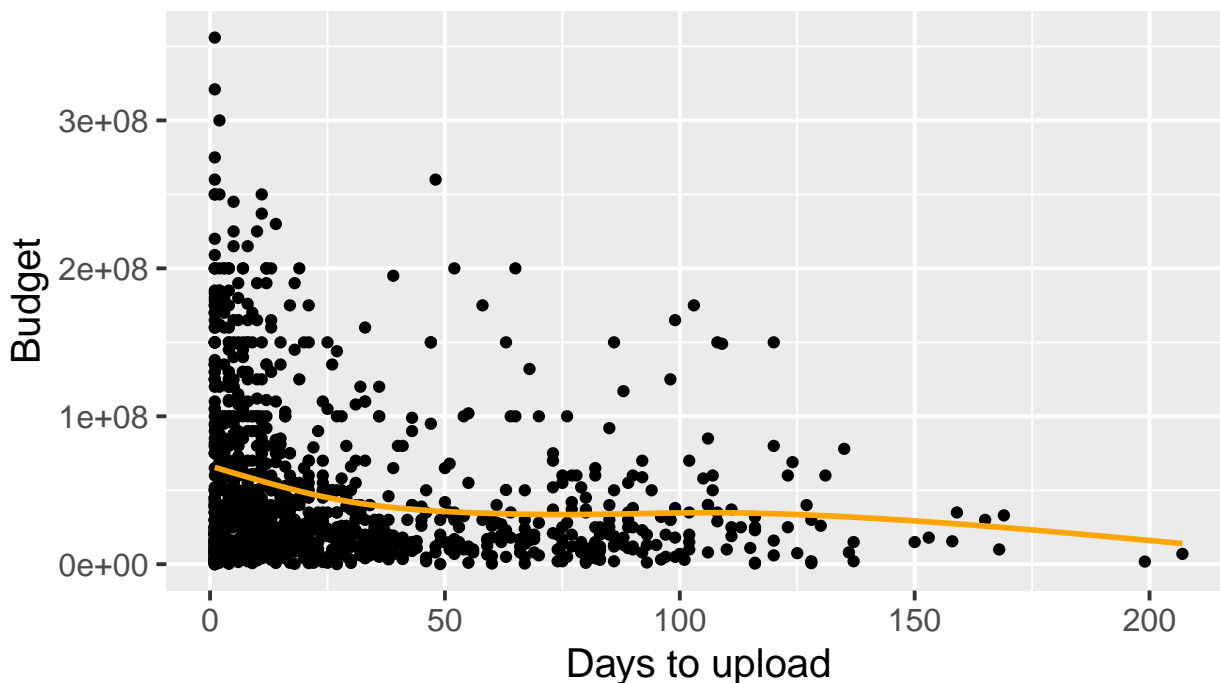


Figure 2: Average days to good quality uploads by budget

the spike and decay of a film's revenue in the days following the initial opening date.

Among the features of this graph, we discern seasonal characteristics of the time series, oscillating at a weekly frequency. This is expected when analyzing the box office because films are mostly viewed over the weekends. The differential peaks and troughs observed are a result of films released on different days of the week, though we can see consistencies among certain films, which reflect common release strategies employed by studios. Such seasonality is smoothed when we view film revenue as weekly aggregates, as in Figure 5.

We now see differences in trends among films. Those with lower opening revenues may experience subsequent weeks of rising revenues, whereas those with the highest subset of revenues mostly track downward after the initial week. By aggregating weekly film revenues by genre in Figure 6, we find a more distinct, almost linear trend.

After capturing the heterogeneity across genres, we examine what approximates a linear decline in the change in earnings across the weeks since a film's release.

In this regard, our basic empirical strategy asks whether such a return profile is shifted lower by the availability of good quality torrents for a given film, on a given day since release, at a given date. A visualization of this phenomenon is shown in Figure 7, where

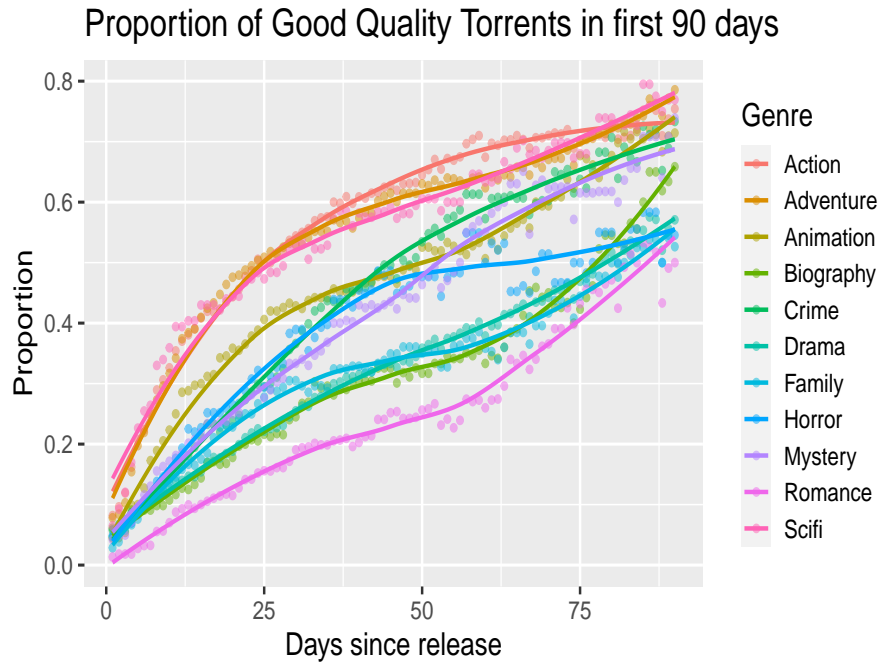


Figure 3: Average proportion of good torrents over time by genre

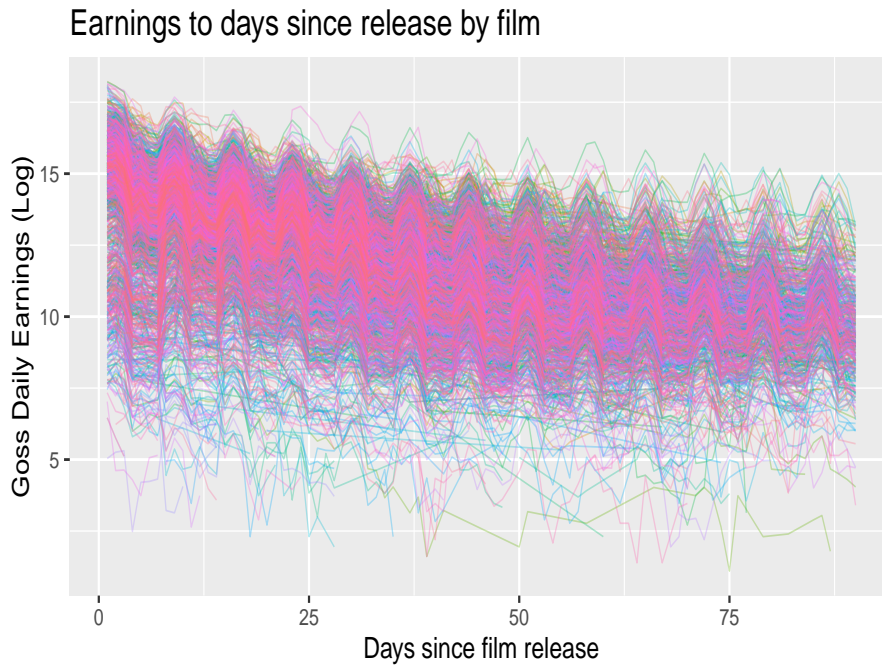


Figure 4: Individual film daily revenue decay over time

we view a subset of films experiencing a good quality upload at the same day since release. Through this event-study style approach, we clearly see a discontinuity in the form of a

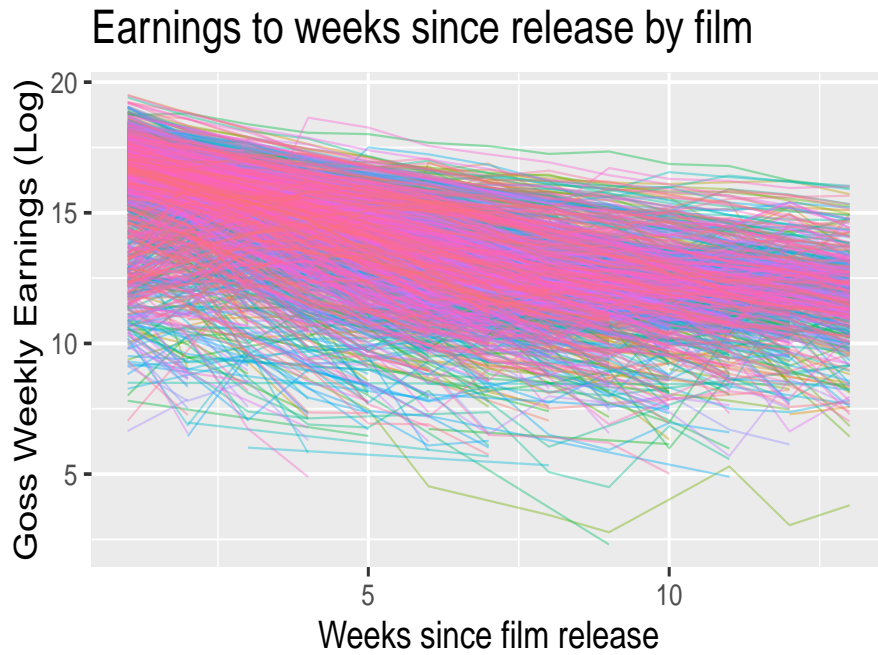


Figure 5: Individual film's weekly revenue decay over time

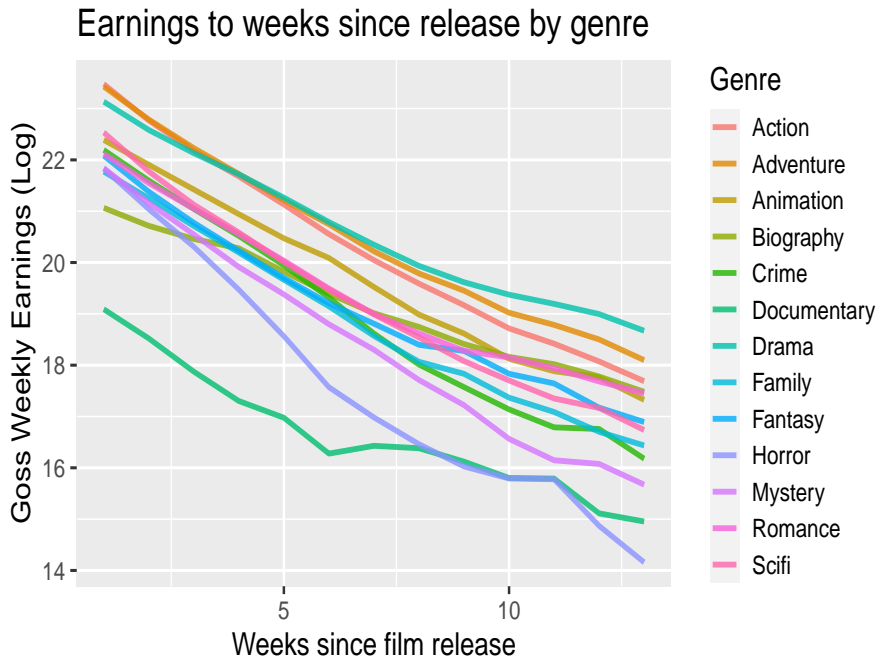


Figure 6: Average genre weekly revenue decay over time

downward shift of the gross daily revenue decay.

We also ask whether the upload of a bad quality torrent of a film has any effect on

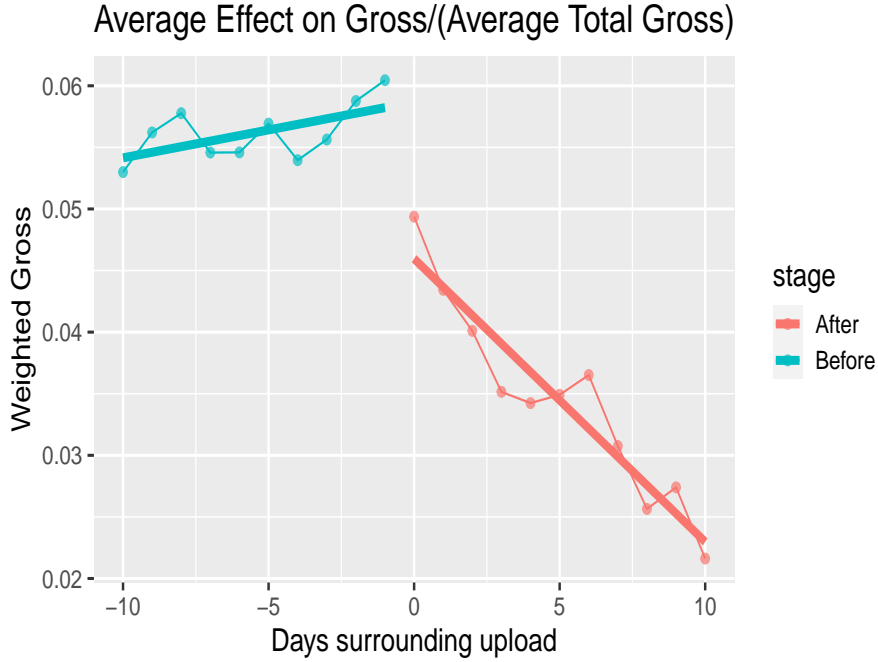


Figure 7: 10 days before and after the arrival of a good quality torrent

the gross daily revenue. To answer this question, we develop an initial fixed effect model, as estimated in Model 1. In this case, the key fixed effects are film and days, with an extension for time factors. The film- and day-fixed effects should remove variation in film- or time-specific characteristics. Estimation involves the following two steps: (1) coefficients are estimated by centering on all the group means and (2) OLS estimation. Our estimation method is equivalent to utilizing a large set of dummy variables accounting for the numerous nuanced observations and characteristics thereof.

$$InRevenue_{ijt} = \beta_1 Bad\_Quality_{ijt} + \beta_2 Good\_Quality_{ijt} + X_{ijt} + \mu_i + \Upsilon_j + \phi_t + \varepsilon_{ijt} \quad (1)$$

Where  $InRevenue$  represents the log of box office returns for film  $i$  on the day of the week  $j$  on the day after release  $t$ .  $Bad\_Quality$  is a dummy indicating the availability of a bad quality torrent of the film on TPB;  $Good\_Quality$  is a similar dummy for a good quality torrent of the film.  $X_{ijt}$  includes any other film, country, or time controls affecting the box Office performance;  $\mu_i$  is a film-specific fixed effect;  $\Upsilon_j$  is the time-fixed effect;  $\phi_t$  is the day-since-release fixed effect; and  $\varepsilon_{ijt}$  is an idiosyncratic day-movie-date shock.

Our test for whether online piracy cannibalizes box office returns proceeds by asking whether the daily box office returns change as a result of the availability of a piratable



copy. As previously mentioned, we take a forked approach to the availability of a torrent by separating film torrents into bad and good quality: largely differentiated resolution and pixel characteristics, as listed in the torrent’s title.

Although there are several strategies along this line, we must first ask whether the arrival of a good or bad quality torrent of a film affects the daily revenue earned at the box office. Results can be found in Table 2

Table 2: Fixed effect modeling strategies

VARIABLES	(1) All Movies
bad_quality_dummy	0.0320** (0.0153)
good_quality_dummy	-0.161*** (0.0160)
Observations	110,792
$R^2$	0.869
Movie FE	Yes
Days since release FE	Yes
Year-Week FE	Yes
Number of movies	2145

Robust standard errors in parentheses

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

In order to observe the impact of torrent availability, Column (1) reports our baseline approach, a linear fixed effects model, holding film, and days-since-release as fixed effects. This baseline model finds the release of a good quality torrent causes an average 16% reduction in gross daily revenue, significant at the 1% level, with an  $R^2$  of 87%.

Our next strategy investigates the variation between film budgets, using a similar approach to sub-setting, before running the fixed effect models. We categorize film budgets into mutually exclusive groups at the 25th inter quantile range and 75th percentile—low-budget films  $\leq \$15million$ , mid-budget films  $> \$20$  and  $\leq \$42million$ , and high-budget films  $\geq \$42million$ .

Table 3 displays the fixed effects output for this subset, again using our principle modeling approach with film, day-since-release and date-fixed effect groups.

We find heterogeneity among budgets, with some results contrary to our central find-

Table 3: Budget subset

VARIABLES	(1) Movies <P25	(2) Movies IQ	(3) Movies >P75
bad_quality_dummy	-0.344*** (0.040)	0.027 (0.021)	0.149*** (0.027)
good_quality_dummy	0.227*** (0.044)	-0.124*** (0.022)	-0.048** (0.022)
Observations	28,775	54,736	27,281
$R^2$	0.845	0.870	0.926
Movie FE	Yes	Yes	Yes
Days since release FE	Yes	Yes	Yes
Year-Week FE	Yes	Yes	Yes
Number of movies	712	1026	407

Robust standard errors in parentheses

\*\*\* p&lt;0.01, \*\* p&lt;0.05, \* p&lt;0.1

ings. While low-budget film revenues decrease by 34% because of a bad quality torrent, significant at the 1% level, they increase by 22% upon the release of a good quality torrent, significant at the 1% level.

Mid-budget films experience a 12% decline in daily revenue after the release of a good quality torrent. Interestingly, the cohort effect is negative, removing the significance effect of the torrents. We relate these results to the strong heterogeneity between films, which we will investigate further in our Discussion section.

Finally, high-budget films experience a 15% increase in daily revenue from the time of the upload of a bad quality torrent and a 4% decrease from the time of the upload of a good quality torrent, significant at the 1% and 5% levels, respectively. Low- and high-budget films benefit from cohort leaks by 18% and 10%, respectively, significant at the 1% level.

In addition to film budget, to capture other film-specific characteristics, we look at the film genre. In this way, we can observe variation in the effect of piracy on box office revenue across different styles of film.

Table 4,5, and 6 display the results by genre. Robust standard errors are used for each estimate. For each subset, we list the number observations as well as unique films (N Obs, N Unique).

Table 4: Genre subset 1

VARIABLES	(1) Action	(2) Adventure	(3) Animation	(4) Biography	(5) Comedy
bad_quality_dummy	0.063** (0.027)	0.092*** (0.030)	-0.116** (0.050)	0.493*** (0.070)	-0.106*** (0.022)
good_quality_dummy	-0.050** (0.021)	-0.046* (0.026)	-0.005 (0.054)	-0.646*** (0.078)	-0.210*** (0.024)
Observations	32,842	26,300	9,309	12,051	41,615
$R^2$	0.927	0.915	0.921	0.855	0.882
Movie FE	Yes	Yes	Yes	Yes	Yes
Days since release FE	Yes	Yes	Yes	Yes	Yes
Year-Week FE	Yes	Yes	Yes	Yes	Yes
Number of movies	589	412	143	220	770

Robust standard errors in parentheses

\*\*\* p&lt;0.01, \*\* p&lt;0.05, \* p&lt;0.1

Table 5: Genre subset 2

VARIABLES	(1) Crime	(2) Drama	(3) Family	(4) Fantasy	(5) Horror
bad_quality_dummy	-0.020 (0.041)	0.014 (0.025)	0.081* (0.046)	0.304*** (0.060)	-0.143*** (0.035)
good_quality_dummy	0.018 (0.039)	-0.158*** (0.028)	-0.186*** (0.061)	-0.006 (0.056)	0.344*** (0.037)
Observations	17,333	55,535	7,433	9,213	9,510
$R^2$	0.904	0.830	0.941	0.929	0.965
Movie FE	Yes	Yes	Yes	Yes	Yes
Days since release FE	Yes	Yes	Yes	Yes	Yes
Year-Week FE	Yes	Yes	Yes	Yes	Yes
Number of movies	377	1142	133	175	225

Robust standard errors in parentheses

\*\*\* p&lt;0.01, \*\* p&lt;0.05, \* p&lt;0.1

Table 6: Genre subset 3

VARIABLES	(1) Mystery	(2) Romance	(3) Sci-Fi
bad_quality_dummy	-0.146*** (0.040)	0.071* (0.041)	-0.187*** (0.044)
good_quality_dummy	0.096** (0.046)	-0.314*** (0.044)	0.150*** (0.042)
Observations	9,686	15,827	9,917
$R^2$	0.954	0.890	0.960
Movie FE	Yes	Yes	Yes
Days since release FE	Yes	Yes	Yes
Year-Week FE	Yes	Yes	Yes
Number of movies	212	332	167

Robust standard errors in parentheses

\*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ 

This approach yields surprising results—the revenue of some films is aided by the release of a bad or good quality torrent and others are completely unscathed by piracy.

We find that crime films are unaffected by both bad and good quality torrents at all listed significance levels. Horror and sci-fi films record a significant decline of 14% and 18%, respectively, in daily revenues after the release of a bad quality torrent; this decline is significant at the 1% level; however, these genres record an increase of 34% and 15%, respectively, after the upload of a good quality torrent; this increase is significant at the 1% level. Comedy films see a statistically significant decline in daily revenue after the release of both bad and good quality torrents, at 10% and 21%, respectively, at the 1% significance level. Biographies witness a significant increase of 49%, at the 1% significance level, in daily revenue after the release of a bad quality torrent. However, they witness a 64% decline, at the 1% significance level, in daily revenue from the upload of a good quality copy.. Adventure and romance films benefit from the upload of a bad quality torrent, with a 9% and 7% increase in daily revenue, respectively. However adventure films witness a decline of 4%, at a significance level of 10%, in daily revenue with the arrival of a good quality copy. Romance films' daily revenue decreases by 31% after the release of a good quality copy, at the 1% significance level. Animation titles see a statistically significant decline of 11% in daily revenue from the upload of a bad quality

Table 7: Cohort analysis

VARIABLES	(1) All Movies	(2) Movies P25	(3) Movies IQ	(4) Movies <i>P75</i>
bad_quality_dummy	0.031** (0.015)	-0.354*** (0.040)	0.033 (0.021)	0.151*** (0.027)
good_quality_dummy	-0.178*** (0.020)	0.104** (0.050)	-0.031 (0.029)	-0.105*** (0.030)
cohort_dummy	0.026 (0.019)	0.187*** (0.042)	-0.152*** (0.028)	0.100*** (0.029)
Observations	110,792	28,775	54,736	27,281
$R^2$	0.869	0.845	0.870	0.926
Movie FE	Yes	Yes	Yes	Yes
Days since release FE	Yes	Yes	Yes	Yes
Year-Week FE	Yes	Yes	Yes	Yes
Number of movies	2145	712	1026	407

Robust standard errors in parentheses

\*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ 

copy, at the 5% significance level, and are unaffected by the upload of a good quality torrent. Family and fantasy movies both see an increase in daily revenue from the upload of a bad quality torrent by 8% and 30%, respectively. Family movies see a decrease of 18% in daily revenue from the upload of a good quality torrent, at the 1% significant level, whereas the coefficient and significance for fantasy movies are zero. Crime films are unaffected by the upload of bad and good quality torrents.

## 5.1 Cohort Analysis

As an additional empirical test, we include a cohort dummy with our main specifications. The cohort dummy becomes 1 when for a given week, in the same genre (defined as the set of up to three genres), there is a good quality torrent; it is 0 otherwise. We believe that the cohort effect measures the extent to which movies compete not only with their own pirated copies but also with the pirated copies of similar substitutable movies. For example, the movie *The Second Act* was released on December 21, 2018 and was categorized under three genres—comedy, drama, and romance. The movie *Zero* was also

Table 8: Cohort Genre subset 1

VARIABLES	(1) Action	(2) Adventure	(3) Animation	(4) Biography	(5) Comedy
bad_quality_dummy	0.063** (0.027)	0.093*** (0.030)	-0.109** (0.050)	0.497*** (0.070)	-0.109*** (0.022)
good_quality_dummy	0.028 (0.027)	-0.146*** (0.033)	-0.141** (0.059)	-0.509*** (0.090)	-0.261*** (0.029)
cohort_dummy	-0.135*** (0.027)	0.170*** (0.033)	0.348*** (0.073)	-0.209*** (0.070)	0.081*** (0.028)
Observations	32,842	26,300	9,309	12,051	41,615
$R^2$	0.927	0.915	0.922	0.855	0.882
Movie FE	Yes	Yes	Yes	Yes	Yes
Days since release FE	Yes	Yes	Yes	Yes	Yes
Year-Week FE	Yes	Yes	Yes	Yes	Yes
Number of movies	589	412	143	220	770

Robust standard errors in parentheses

\*\*\* p&lt;0.01, \*\* p&lt;0.05, \* p&lt;0.1

Table 9: Cohort Genre subset 2

VARIABLES	(1) Crime	(2) Drama	(3) Family	(4) Fantasy	(5) Horror
bad_quality_dummy	-0.018 (0.041)	0.017 (0.025)	0.081* (0.046)	0.299*** (0.061)	-0.145*** (0.035)
good_quality_dummy	0.067 (0.045)	-0.120*** (0.035)	-0.176** (0.076)	-0.053 (0.059)	0.327*** (0.039)
cohort_dummy	-0.096** (0.048)	-0.057* (0.032)	-0.018 (0.079)	0.080 (0.069)	0.039 (0.049)
Observations	17,333	55,535	7,433	9,213	9,510
$R^2$	0.904	0.831	0.941	0.929	0.965
Movie FE	Yes	Yes	Yes	Yes	Yes
Days since release FE	Yes	Yes	Yes	Yes	Yes
Year-Week FE	Yes	Yes	Yes	Yes	Yes
Number of movies	377	1142	133	175	225

Robust standard errors in parentheses

\*\*\* p&lt;0.01, \*\* p&lt;0.05, \* p&lt;0.1

Table 10: Cohort Genre subset 3

VARIABLES	(1) Mystery	(2) Romance	(3) Sci-Fi
bad_quality_dummy	-0.147*** (0.041)	0.072* (0.041)	-0.193*** (0.045)
good_quality_dummy	0.137*** (0.049)	-0.202*** (0.053)	0.106** (0.048)
cohort_good_quality_dummy	-0.080 (0.063)	-0.183*** (0.050)	0.093* (0.052)
Observations	9,686	15,827	9,917
$R^2$	0.954	0.891	0.960
Movie FE	Yes	Yes	Yes
Days since release FE	Yes	Yes	Yes
Year-Week FE	Yes	Yes	Yes
Number of movies	212	332	167

Robust standard errors in parentheses

\*\*\* p&lt;0.01, \*\* p&lt;0.05, \* p&lt;0.1



released on December 21, 2018 and was categorized as comedy, drama, and romance. The cohort effect captures the extent to which the box office revenue for The Second Act movie benefitted, hurt, or had no impact from the release of a good quality copy of Zero or the leak of its competitor’s product.

Column (1) in Table 7 adds the cohort effect to our main specification. On average, across all movies, the cohort dummy is insignificant. However, the cohort effect gains significance when analyzed across different film categories of budget and genre.

Low- and high-budget films see a significant positive cohort effect, as reported in Columns (2) and (4) of the Table 8. Movies stuck in the middle see a significant negative cohort effect. Certain genres receive a boost in daily revenue from the leak of competitor’s movie; cohort dummy is significant and positive for adventure, animation, and comedy. Other genres suffer from the leak of a competitor’s movie; the cohort dummy is significant and negative for action, biography, crime, and drama. The remaining genres have no significant cohort effect. Coefficients and signs for the main variables are robust to the cohort analysis and are reported in Table 7 and Table 8.

Finally, it is useful to compare these results to the two most relevant papers in the piracy literature on the effect of piracy on the box office. As depicted in Table 11, Danaher and Smith (2014) find that the shutdown of Megaupload led to a 6.5-8.5% weekly revenue increase for films in the box office, while Peukert et al. (2017) find a 4% decrease in weekly revenue for films with narrow or niche release, highlighting the discovery benefit of piracy.

Table 11: Impact of piracy on Box Office as found in literature

<b>Study</b>	<b>Event</b>	<b>Result</b>
Danaher & Smith	Megaupload Shutdown	6.5-8.5% Rev Inc
Peukert et. al	Megaupload shutdown	4% Rev Decr for Niche Release

We find results in both directions, with greater magnitudes. A much stronger impact of piracy on box office revenue is found in our central model than that in Danaher and Smith (2014). This is as expected, and previously justified, from the partial effect observed in studies on the negative supply shock to an unknown proportion of supply. An equivalent explanation is given for the existence and magnitude of the discovery effect for films with a narrower release, and our budget subset analysis is consistent with these findings. Unlike other research in this area, our results demonstrate the capacity of both low- and high-budget films to benefit from a bad quality copy.

## 6 Discussion and Conclusion

### 6.1 Limitations

Our research is not without limitations. First, we consider only those torrents with IMDb links listed. As a result, we likely experience greater data loss than necessary, which results in small sample size issues. However, the cost to overcome this issue would be significant, as much work would be required to accurately identify and match torrents to films, a process simplified with the IMDb link. If an accurate tool could be implemented, the torrents dataset could grow significantly.

Second, our identification strategy hinges on accurate classification of both poor quality and good quality torrent uploads. Good quality torrents are classified using particular keywords that are present in the title of the upload, whereas bad quality torrents are simply the first uploaded torrent per film. When the date of good quality and bad quality uploads match, the observation is removed to prevent double classification. This is a strict approach to classification, which can be improved by adding additional keywords or through natural language processing. However, these methods also result in their own set of biases and fuzzy matches. Finally, we cannot directly inspect the contents of each torrent to confirm the quality tier, and therefore rely on the accuracy of each uploader.

### 6.2 Discussion

Empirical research and theoretical justifications for piracy have often led to ambiguous results in terms of the effects on box office revenue. The majority of empirical research on piracy focuses on specific events, such as the introduction of new laws and the shutdown of piracy websites, to conduct quasi-experiments measuring the impact of a negative shock to the supply of piracy on movie sales. Results from these studies have limited interpretations for the holistic effect of piracy on box office revenue as they study a partial restriction in consumption and act as a lower-bound for the impact of piracy. Our research overcomes this issue through a novel approach to identification; this approach uses the timing of a good quality release of a film to signify a random positive shock to the supply of piracy. This methodology allows us to measure the change in daily revenue with the availability of a substitute to cinematic release, and hence helps us to infer the wider impact of piracy. In this respect, we find strong evidence of an average negative impact on the Box Office of 16%, as seen from our central models in Table 2. The immediate implication is that, on average, films experiencing an upload of a good quality torrent between 2003-2018 lost

about a fifth of their daily revenue to piracy.

From the breakdown in genres in Table 4, we find that film genres represent different value propositions. For example, the value of horror, mystery, or sci-fi films may be enhanced through a cinematic experience. Conversely, films with a heavier concentration on story, which are potentially less experiential, may hold no additional value to consumers when seen in theaters. This may help explain why story-based genres suffer the negative effects of piracy; this is evident from a decline the daily revenue of comedy movies, which fall by 10% and 21% upon the arrival of bad and good quality copies, respectively. However, the release of a good quality copy of a horror or sci-fi film increases its daily box office revenue by 34% and 15%, respectively. For experiential film genres, a good quality copy cannot be viewed as an exact substitute; instead, the appearance of pirated version amplifies the discovery or word-of-mouth effects, thereby increasing the demand for cinematic consumption. In this case, the at-home consumption of a good quality copy pales in comparison to the in-theater experience. For film genres relying heavily on stories, both low- and good quality copies are viewed as exact substitutes, which negatively affects the box office revenue. The value captured by consumers through an at-home viewing of a story-based film is equivalent to the value captured through an in-theater viewing.

In regard to the breakdown in budget as seen in Table 3, we find similar evidence for the discovery or word-of-mouth effects highlighted in the literature (Peukert et al. 2017). Low-budget films benefiting from the release of a good quality copy of a film by 22.7% may imply an increase in awareness. This is because viewers discover a title they otherwise would not have known due to low-budget films' low marketing spend. If the film had low awareness, there may not have been enough consumption to be displaced by the copy. Conversely, high-budget films benefit from the arrival of a bad quality torrent with a 14.9% increase in revenue. Rather than a discovery effect, this effect may be an indicator of a high demand for the release of high-budget films. Bad quality torrents may add to the marketing campaign of a high-budget film by acting as an extended movie trailer for consumers who are eager to see it in theaters. Such films, however, are negatively affected by the upload of a good quality torrent with a 4.8% decline in revenue. This finding is consistent with the theory that, owing to the high demand for these films, consumption can be tangibly affected by the arrival of a substitute to the theatrical version. It is worth noting that the coefficient for the effect of good quality copies is smaller than that of the coefficient for the effect of bad quality copies on films. One interpretation is that the positive discovery effect of piracy on films is economically more important than that of the negative substitution effect of piracy on films. Finally, mid-budget films seem to be

negatively affected by the arrival of bad and good quality torrents at 2.7% and 12.4%, respectively, significant at the 1% level. One explanation is that mid-budget films have sufficient consumption to be displaced and sufficient marketing budgets where they do not benefit from a discovery effect.

### 6.3 Management Implications

Our results have several implications for management and strategy, in that they guide firms to select the right business model depending on the situations that might arise (Casadesus-Masanell and Zhu 2010). If films of different genres and budgets react variably to piracy, decision-makers at film studios can choose film release strategies that soften the impact of piracy or use it to their advantage. Results from this study may also help managers experiment with film distribution timing in terms of box office versus digital release schedules, as discussed for other products with complementary distribution methods, such printed books and e-books, which are at risk of cannibalizing one another (Hu and Smith 2018).

Given that film studios produce a diverse product mix, it may be possible for these firms to counteract the negative effects of piracy while amplifying the positive by altering the number of a particular genre it produces each year or changing the manner in which it is released. For example, large film studios produce several movies each year that vary by genre and budget allocated. Hence, for genres that benefit from piracy, film studios can reduce or eliminate resources allocated toward enforcement efforts and re-allocate them toward films of genres or budgets most at risk of piracy. For genres most affected by piracy, film studios can adopt more stringent enforcement efforts, from restricting access to film footage to limiting screeners sent to film critics, journalists, producers, and other film industry insiders. Low-budget films may consider eliminating resources toward enforcement altogether and, instead, consider piracy as unpaid advertisement or word-of-mouth publicity.

A bold implication from our genre subset’s fixed effects estimation, as displayed in Table 4,5 and 6, suggests that studios could artificially leak a bad quality copy, or particular scenes, of a film to boost box office revenue. For example, we find that genres such as biography and fantasy benefit as a result of an early arrival of bad quality torrents; they witness an increase in daily revenue at 49.3% and 30.4%, respectively.

Certain genres experience an increase in revenue from the release of a bad quality copy of the film and a decrease in revenue from the release of a good quality copy of the film, such as adventure, family, and romance. This result implies the sampling or preview effect

is at play, where consumers use a bad quality copy of the film as an extended trailer to decide whether they would like to see a movie in theaters. For these films, consumption is displaced from good quality release of pirated copies. Hence, film studios should consider pushing for faster wide releases of such films in the box office to maximize revenue before a good quality copy of the film is uploaded. For firms where revenue is completely displaced by piracy, such as comedy and drama, another strategy could be increasing access to such films, whether by paid streaming or otherwise, in conjunction with cinematic release, to capture revenue from piracy motivated by accessibility issues.

Our budget-fixed effects models, shown in Table 3, further justify such strategies. Consider the effect from horror films, which are especially known to subvert trends between budget and gross earnings, with notable shoe-string level budgets yielding all-time high earnings for the genre. A few examples of such titles include *The Blair Witch Project*, which grossed 248.6 million off a 60,000 budget, and *Paranormal Activity*, which grossed 193.4 million off an 11,000 budget. Our budget-fixed effects approach could explain such results or the contagion spread of awareness of these films. This is because we find that low-budget films (<16m) benefit from the arrival of a bad quality torrent, with a revenue surge of 18.1%. Whether purely from a discovery effect or otherwise, low-budget films can benefit from a leak of a bad quality torrent. This strategy can be employed as a guerrilla marketing technique to increase awareness, and thus ticket sales. A similar conclusion can be made for high-budget films (>40m), wherein bad quality copies or scenes can be leaked to increase marketing awareness or simply to alert viewers that the film is available in cinemas.

Finally, from the cohort analysis, it appears that piracy affects competition among films that are out in cinemas at the same time. While box office revenue of films of certain genres, like comedy and animation, decline as a result of the emergence of a pirated copy (good or bad), they experience a boost in sales when another film of the same genre in their release-date cohort is pirated. This implies that these films are in direct competition with both their pirated copy as well as same-genre films in their release-date cohort. Low- and high-budget films appear to benefit from the piracy of other films of a similar budget released in theaters the same week. We interpret this result broadly to mean that the piracy of films, including blockbuster ones, does not necessarily imply that audiences forgo the cinema. Audiences may, instead, alter the films they choose to see in theaters.

We find that movie studios should largely be concerned by the upload of good quality copies of their films to the Internet, as, on average, they should expect to lose 16% of daily revenue for that film. However, it is unclear the extent to which consumers

substitute the cinema experience with a piracy experience, or substitute their first-choice film with another film showing in cinemas. Our findings of differential effects between both budget and genre suggest that studios may benefit from bespoke release and enforcement strategies that take into account film characteristics and in-theater cohorts. Our results are generalizable to industries where piracy appears in quality tiers, such as music or video games. Taking the latter as an example, a pirated video game may lack the full integration or features of the authentic version; thus, we would expect different quality tiers of piracy to produce heterogeneous impacts on sales for video games. Our results also apply to digital industries where firms produce highly differentiated products; here, in the video game industry, firms may profit from the piracy of certain product categories over others, such as cooperative versus non-cooperative video games. Like the cinematic experience, cooperative video games may lose value when limited to a non-cooperative pirated version and serve, simply, as a preview. While this study has implications for a policy based response to piracy, it is likely that a strategic response from production companies is both faster to implement and more effective, through the adaptation of existing value-capture and distribution strategies, than the development of any single anti-piracy law.

## References

- Aguiar L, Martens B (2016) Digital music consumption on the internet: evidence from click-stream data. *Information Economics and Policy* 34:27–43.
- Bai J, Waldfogel J (2012) Movie piracy and sales displacement in two samples of chinese consumers. *Information Economics and Policy* 24(3-4):187–196.
- Bhattacharjee S, Gopal RD, Lertwachara K, Marsden JR, Telang R (2007) The effect of digital sharing technologies on music markets: A survival analysis of albums on ranking charts. *Management Science* 53(9):1359–1374.
- Casadesus-Masanell R, Zhu F (2010) Strategies to fight ad-sponsored rivals. *Management Science* 56(9):1484–1499.
- Danaher B, Dhanasobhon S, Smith MD, Telang R (2010) Converting pirates without cannibalizing purchasers: The impact of digital distribution on physical sales and internet piracy. *Marketing science* 29(6):1138–1151.
- Danaher B, Smith MD (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 MD, Telang R (2016) Website blocking revisited: The effect of the uk november 2014 blocks on consumer behavior. *Available at SSRN 2766795* .
- Danaher B, Waldfogel J (2012) Reel piracy: The effect of online film piracy on international box office sales. *Available at SSRN 1986299* .
- Gosline R (2009) *Simultaneous distinction, democratization and omnivorism effects: A longitudinal analysis of dynamic symbolic boundaries in counterfeit consumption networks*. Ph.D. thesis, Dissertation, Harvard Business School.
- Gosline R (2010) Counterfeit labels: Good for luxury brands. *Forbes* 2(12):2010.
- Greenstein S (2015) *How the internet became commercial: Innovation, privatization, and the birth of a new network* (Princeton University Press).
- Grossman GM, Shapiro C (1988) Foreign counterfeiting of status goods. *The Quarterly Journal of Economics* 103(1):79–100.
- Hu YJ, Smith M (2018) The impact of ebook distribution on print sales: Analysis of a natural experiment. *Management Science, forthcoming* .
- Hui KL, Png I (2003) Piracy and the legitimate demand for recorded music. *Contributions in Economic Analysis & Policy* 2(1).
- Liebowitz SJ (2006) Economists examine file-sharing and music sales. *Industrial organization and the digital economy* 145–174.
- Oberholzer-Gee F, Strumpf K (2010) File sharing and copyright. *Innovation policy and the economy* 10(1):19–55.

- Peitz M, Waelbroeck P (2004) The Effect of Internet Piracy on Cd Sales: Cross-Section Evidence. SSRN Scholarly Paper ID 511763, Social Science Research Network, Rochester, NY, URL <https://papers.ssrn.com/abstract=511763>.
- Peukert C, Claussen J, Kretschmer T (2013) Piracy and movie revenues: Evidence from megauupload. a tale of the long tail? *Kiel und Hamburg: ZBW-Deutsche Zentralbibliothek für ...*
- Peukert C, Claussen J, Kretschmer T (2017) Piracy and box office movie revenues: Evidence from megauupload. *International Journal of Industrial Organization* 52:188–215.
- Poort J, Leenheer J, van der Ham J, Dumitru C (2014) Baywatch: Two approaches to measure the effects of blocking access to the pirate bay. *Telecommunications Policy* 38(4):383–392.
- Qian Y (2008) Impacts of entry by counterfeiters. *The Quarterly Journal of Economics* 123(4):1577–1609.
- Qian Y (2014) Counterfeiters: Foes or friends? how counterfeits affect sales by product quality tier. *Management Science* 60(10):2381–2400.
- Rob R, Waldfogel J (2007) Piracy on the silver screen. *The Journal of Industrial Economics* 55(3):379–395.
- Schwartzel E (2020) WSJ News Exclusive | ‘Trolls World Tour’ Breaks Digital Records and Charts a New Path for Hollywood. *Wall Street Journal* ISSN 0099-9660, URL <https://www.wsj.com/articles/trolls-world-tour-breaks-digital-records-and-charts-a-new-pa>
- Smith MD, Telang R (2009) Competing with free: The impact of movie broadcasts on dvd sales and internet piracy. *mis Quarterly* 321–338.
- Smith MD, Telang R (2012) Assessing the academic literature regarding the impact of media piracy on sales. *Available at SSRN 2132153* .
- Vernik DA, Purohit D, Desai PS (2011) Music downloads and the flip side of digital rights management. *Marketing Science* 30(6):1011–1027.
- Wang Q, Li B, Singh PV (2018) Copycats vs. original mobile apps: A machine learning copycat-detection method and empirical analysis. *Information Systems Research* 29(2):273–291.
- Watson SJ, Zizzo DJ, Fleming P (2017) Risk, benefit, and moderators of the affect heuristic in a widespread unlawful activity: Evidence from a survey of unlawful file-sharing behavior. *Risk Analysis* 37(6):1146–1156.
- Zentner A (2005) File sharing and international sales of copyrighted music: An empirical analysis with a panel of countries. *The BE Journal of Economic Analysis & Policy* 5(1).



# Appendices

Table 12: Pirate Bay Data

Variable	Type	Description
Description		Description included on torrent page
Details		
<i>Title</i>	Character	<i>Title of torrent listing</i>
<i>Tags</i>	Character	<i>Tags of torrent for categorization</i>
<i>Type</i>	Numeric	<i>Type of Number</i>
<i>Files</i>	Numeric	<i>List of files</i>
<i>Size</i>	Character	<i>File Size</i>
<i>Quality (+)</i>	Numeric	<i>Listed Quality</i>
<i>Quality (-)</i>	Numeric	<i>Listed Quality</i>
<i>IMDB</i>	Character	<i>IMDB Link</i>
<i>Spoken Languages</i>	Character	<i>Spoken Languages</i>
<i>Texted Languages</i>	Character	<i>Languauge of any Text</i>
<i>Uploaded</i>	Character	<i>Date of Upload</i>
<i>By</i>	Character	<i>Username of Uploader</i>
<i>User</i>	Character	<i>User Type</i>
<i>Seeders</i>	Numeric	<i>Number of Seeders</i>
<i>Leechers</i>	Numeric	<i>Number of Leechers</i>
<i>Info Hash</i>	Character	<i>Unique hash of torrent for download</i>
<i>Picture</i>	Image	<i>Whether a picture was uploaded</i>
<i>Capture Date</i>	Numeric	<i>Date metadata was captured</i>
FileList		
<i>Filename</i>	Character	<i>Name of file</i>
<i>Size</i>	Numeric	<i>Size of file</i>
<i>Unit</i>	Character	<i>Unit of file size</i>
Comments	Character	Lists all comments on torrent

Table 13: Box Office Data: Source CRAN

<b>Variable</b>	<b>Type</b>	<b>Description</b>
Movie	Character	Film Title
Distributor	Character	Production Company
Gross	Numeric	Daily Gross Earnings
Percentage_Change	Numeric	Daily % Change
Threatres	Numeric	Number of Theatres Showing
Per_Threatre	Numeric	Revenue per Threatre (Av)
Total_Gross	Numeric	Cumulated Gross
Days	Numeric	Days showing
Date	Numeric	Current Date

Table 14: Raw Torrents

Title	Size	IMDB	Uploaded	By
1 The Dark Knight 2008 TS XVID - STG	1.28 GiB	imdb.com/title/tt0468569/	2008-08-02	circlessess
2 The.Dark.Knight.PROPER.DVDSCR.XViD-m Vs	1.41 GiB	imdb.com/title/tt0468569/	2008-09-04	circlessess
3 The Dark Knight PROPER DVDSCR m Vs[Single AVI File]	1.38 GiB	imdb.com/title/tt0468569/	2008-09-05	circlessess
4 The Dark Knight DVDSCR Norsub.Lfc.4Ever	1.37 GiB	imdb.com/title/tt0468569/	2008-09-07	Lfc_4Ever
5 The Dark Knight 2008-DVDSCR-Eng-DPG{E5P0}	333.86 MiB	imdb.com/title/tt0468569/	2008-09-30	E5P0
6 The Dark Knight DvDrip Norsub.Lfc.4Ever	1.36 GiB	imdb.com/title/tt0468569/	2008-11-18	Lfc_4Ever
7 The Dark Knight KLAXXON	1.47 GiB	imdb.com/title/tt0468569/	2008-12-08	KLAXXON
8 The.Dark.Knight.(2008).SWESUB.DVDRip.XviD-Sevaron	1.95 GiB	imdb.com/title/tt0468569/	2009-03-14	Sevaron
9 The Dark Knight in HINDI by mIKKImAX	1.44 GiB	imdb.com/title/tt0468569/	2009-04-11	mikkimax
10 The Dark Knight [2008] dvd rip nlx	699.11 MiB	imdb.com/title/tt0468569/	2009-06-20	nlx_78
11 The dark knight (2008) Retail Multi Audio+Subs NLUPPER	4.32 GiB	imdb.com/title/tt0468569/	2010-06-25	NLUPPER
12 The.Dark.Knight.2008. - PAL RoSub - by Laur	4.37 GiB	imdb.com/title/tt0468569/	2010-12-01	bylaur
13 The Dark Knight 2008 BRrip 720 <b>x264</b> [Herakler]	697.51 MiB	imdb.com/title/tt0468569/	2010-12-29	Herakler
14 The Dark Knight 6 (2008)(Multisubs)(BlackAnchor) TBS	4.34 GiB	imdb.com/title/tt0468569/	2011-01-12	Black-Anchor
15 The dark knight(2008).bdrip.HD. <b>x264</b> -ISAS	595.69 MiB	imdb.com/title/tt0468569/	2012-03-29	isasbay
16 The.Dark.Knight>Returns.Part.1.2012.STV.480p.DVDRip. <b>x264</b> .	620.98 MiB	imdb.com/title/tt2313197/	2012-09-09	loadstone2k12
17 The.Dark.Knight>Returns.Part.1.2012.STV.DVDRip.XviD.MP3	699.89 MiB	imdb.com/title/tt2313197/	2012-09-09	loadstone2k12
18 The.Dark.Knight>Returns.Part.1.2012.DVDRip.XviD-ViP3R	709.53 MiB	imdb.com/title/tt2313197/	2012-09-11	scenebalance
19 The.Dark.Knight>Returns.Part.1.2012.DVDRiP.XViD.AC3-LEGi0	1.43 GiB	imdb.com/title/tt2313197/	2012-09-12	scenebalance
20 The Dark Knight Returns Part I <b>x264</b> - Ganool	349.78 MiB	imdb.com/title/tt2313197/	2012-09-13	Anonymous