LATENT ESTIMATION OF PIRACY QUALITY AND
ITS EFFECT ON REVENUES AND DISTRIBUTION:
THE CASE OF MOTION PICTURES

Anthony Koschmann
Yi Qian

Latent Estimation of Piracy Quality and its Effect on Revenues and Distribution: The Case
of Motion Pictures
Anthony Koschmann and Yi Qian
NBER Working Paper No. 27649
August 2020
JEL No. K42,M21,M31,O3

ABSTRACT

Conventional wisdom holds that illegal copies cannibalize legitimate sales, even though previous research has found mixed effects, with illegal copies acting as both a substitute and complement. Yet, a relatively unexamined aspect to date is the quality of illegal copies. Building on product uncertainty and production quality, we propose that higher quality copies can benefit sales when product uncertainty is high, such as during the launch period. Using motion picture and online piracy data, we estimate piracy quality using a latent item response theory (IRT) model based on keyword signals in the copies. An interdependent system jointly estimates movie screens, revenues, downloads, and available illegal copies with piracy quality in both the launch and post-launch periods. We find that at launch, when rather little is known about the movie, higher quality illegal copies demonstrate a positive effect on revenues (sampling). In the post-launch period, however, higher quality illegal copies exhibit a negative effect on revenues (substitution). The findings suggest producers can alleviate product uncertainty through higher quality samples at product launch while diluting piracy quality post-launch.

Anthony Koschmann
Eastern Michigan University
akoschma@emich.edu

Yi Qian
Sauder School of Business
University of British Columbia
2053 Main Mall
Vancouver, BC V6T 1Z2
CANADA
and NBER
yi.qian@sauder.ubc.ca

INTRODUCTION

Piracy, or illegal copies of information goods such as movies, music, and books, pose a considerable threat to content creators' revenues. According to media trade group IBC, the cost of global online piracy as lost revenue will double from $26.7 billion in 2016 to an estimated $51.6 billion in 2022 (IBC 2017); for movie and television content creators, lost revenues from online piracy was estimated at 72% of real revenues ($37.0 billion) in 2016. Citing entertainment technology experts, the IBC report notes that digital rights management (DRM) and anti-piracy can only do so much to lower the quality of pirated copies, but providing a better legal consumer experience can reduce piracy losses. For example, a high-quality version of *The Expendables 3* surfaced prior to the film's launch; the film's underperformance at the box office was attributed to this piracy quality (Spangler 2014), even though some competing films had more illegal downloads but saw higher revenues. Together, these indicate that piracy quality presents an important facet regarding the effects of piracy.

Although piracy poses a threat to almost any information good, a paucity of empirical research exists on illegal copy quality and its impact on the legal market. In one study, piracy quality was assessed using subjective user ratings of the video and audio quality of pirated films (Ma et al. 2014). In other studies not using subjective measures, the binary coding of a high-definition keyword was used (Lu, Wang, and Bendle 2020; Ackermann, Bradley, and Cameron 2020). While helpful in understanding the impact of piracy quality, illegal copies convey more information beyond just its source type (e.g., camcorder or not). Indeed, product quality – even the illegal kind – is composed of many attributes that send signals to consumers (e.g., Zeithaml

1988). Also missing from these studies is consumer activity (i.e., downloads and uploads) of pirated copies, and its effects not just on revenues but the supply side as well.

To address these items, we study piracy quality in the context of motion pictures, a focal area in marketing research (e.g., Dhar and Weinberg 2016; Packard et al. 2016) for several reasons. First, the effect of piracy and consumption may be clearer in movies than in other information goods (Lu, Wang, and Bendle 2020), as few movies are seen by a consumer multiple times in theaters, but illegally downloaded music or software may be consumed repeatedly. Second, while makers of durable products are also concerned about the quality of illegal copies, social or aspirational elements can stimulate illegal consumption (Wilcox, Kim, and Sen 2009) typically not seen in information goods. Third, data collection on illegal activity is difficult to observe; physical products require finding markets with physical transactions. Piracy transactions, however, have some visibility online where users transact with less fear of being caught. Finally, movie piracy is an area that allows us to build on prior theories and findings.

Although piracy quality is of interest to marketers and presents a need for understanding, it faces similar measurement challenges as legal goods. Piracy quality, like product quality, is an important product aspect that affects consumer choice; but quality is often treated subjectively as consumers perceive certain signals (Zeithaml 1988). Indeed, piracy quality is rooted in "any valued attribute of a product" (Chellappa and Shivendu 2005, p.402), even though attribute valuation lends itself to open interpretation arising from heterogeneous tastes. Building on prior piracy research and production quality, we propose that high quality piracy provides information to consumers when product uncertainty is high, such as during new product launch. By more closely resembling the original good, high quality piracy provides more information to consumers; this reduces uncertainty and stimulates sales, particularly among enthusiastic

customers who seek out product information. Post-launch, however, more information is available (such as word of mouth) and product uncertainty is lower. The most willing consumers have likely purchased, making high quality piracy more substitutable for the original good.

Since product quality reflects the underlying attributes, we model piracy quality using the visible piracy file keywords to assemble the various dimensions of quality into a unidimensional latent index. An item response theory (IRT) model estimates the relation of each piracy keyword on this latent spectrum, where piracy files represent ranges of quality based on the ideal points (mean values) of those keywords. The data set consists of movies in wide release in the United States and Canada. Correspondingly, a daily panel tracks those movies' search results and activity on Pirate Bay while the films are first-run in theaters. The data includes box office revenues, screen availability, piracy, advertising, and movie characteristics. Since most research studies on piracy do not account for illegal supply, we use seemingly unrelated regressions (SUR) to jointly estimate the effects of piracy quality on both the legal and illegal sides of the market, using copulas to account for endogeneity.

We find that a 1% increase in the quality of the pirated copies, conditioning on a level of piracy downloads (leechers, or number of users downloading the illegal file), corresponds on average to a 0.52% increase in revenues in the launch period. Upon market introduction, movies lack some information for consumers, so higher quality copies function more like a sampling mechanism. Yet, post-launch shows a 1% increase in the quality of illegal copies, conditioning on a level of leechers, associates with a -0.38% decrease in revenues. As more information about the genuine good permeates the marketplace, higher quality illegal copies cannibalize sales.

This study makes several contributions. First, the study contributes to the piracy literature by proposing that higher quality piracy alleviates product uncertainty by providing product

information. We separate this into launch and post-launch periods as product uncertainty is higher in the former than the latter. Second, to the authors' knowledge, this is the first study to objectively assess the attributes of piracy quality and its effect on both the legal and illegal markets. We operationalize piracy quality through latent recovery of pirated copy attributes (keyword signals). This also includes the effect of piracy quality on illegal supply, which is often neglected in piracy research. Third, the findings highlight differential effects in the timing of piracy quality; in the launch period, higher quality piracy has a positive association with revenues, but this effect is negative in the post-launch period. Finally, the substantive results suggest producers have a unique and advantageous tool for fighting piracy: the legal good itself. Producers can create their own derivations of the genuine good; this creates an opportunity to encourage the right kind of sampling and discourage the wrong kind of cannibalism. These findings give useful meaning to both managers and policymakers regarding the quality nature of illegal variants, while also extending the piracy literature.

THEORETICAL DEVELOPMENT AND CONTRIBUTION

Piracy quality refers to how closely the illegal copy resembles the genuine product (Geng and Lee 2013). As a derived good, the pirated version presumably exhibits lower quality than the genuine product (Sundarajan 2004; Jain 2008; Geng and Lee 2013; Lahiri and Dey 2013; Machado et al. 2017; Dey, Kim, and Lahiri 2019). Piracy quality research has often approached the issue analytically, given the data collection challenges tied to illegal behaviors as well as quantifying the concept of quality. Studies that have empirically assessed piracy quality have focused on a particular signal (Lu, Wang, and Bendle 2020; Ackermann, Bradley, and Cameron

2020). Notwithstanding, information goods have many attributes, suggesting piracy quality is not an 'either-or' byproduct in terms of quality. Piracy quality studies have also lacked incorporation of consumer activity, such as whether consumers download higher or lower quality copies. Finally, studies to date have also not touched on how piracy quality affects both legal and illegal supply, core aspects of the piracy market that may induce omitted variable bias if not addressed (Koschmann and Bowman 2017). Table 1 presents key piracy quality studies to date.

Table 1. Prior Research into Piracy Quality

Year	Authors	Method	Objective Piracy Quality	Assess Quality Attributes	Piracy Quality Measure	Piracy Quality and Consumer Activity	Effect of Piracy Quality on Distribution
2004	Sundararajan	analytical	-	-	-	-	no
2008	Jain	analytical	-	-	-	-	no
2013	Lahiri and Dey	analytical	-	-	-	-	no
2013	Geng and Lee	analytical	-	-	-	-	no
2014	Ma et al.	empirical	no	no	binary	no	no
2017	Machado et al.	analytical	-	-	-	-	no
2019	Dey, Kim, and Lahiri	analytical	-	-	-	-	no
2020	Lu, Wang, and Bendle	empirical	yes	no	binary	no	no
2020	Ackermann, Bradley, and Cameron	empirical	yes	no	binary	no	no
2020	**This study**	**empirical**	**yes**	**yes**	**continuous**	**yes**	**yes**

In general, consumers prefer products of higher quality rather than lower quality given comparable prices. If the quality gap between the pirated copy and the genuine product is large, producers can expect minimal effects from pirated copies (Lahiri and Dey 2013). If the quality gap is small, producers can lower prices (Sundarajan 2004). However, we propose a more

complex relationship between piracy quality and product performance that speaks to product uncertainty and quality.

Piracy quality draws on two streams of research: piracy and production quality. Prior research on piracy has seen rigorous academic investigation, with a focal debate on whether illegal copies help or hurt legal sales. A tension exists in piracy research, as some studies have shown negative effects of illegal copies on legal demand (Hui and Png 2003; Bae and Choi 2006; Yoon 2007; Liebowitz 2008; Hong 2013; Waldfogel 2012; Belleflamme and Peitz 2014), with varying estimates on the sales displacement effect (Aguiar and Waldfogel 2018; Li, Liao, and Xie 2019; Yue 2019). Yet, other piracy research has found positive effects of piracy on legal demand (Fader 2000; Jain 2008; Mortimer, Nosko, and Sorensen 2012; Lu, Wang, and Bendle 2020). In many of these studies, a common belief is a trade-off in sampling versus cannibalization, with willingness to pay as a common explanation. Despite these studies, there is a dearth of empirical evidence regarding piracy quality and its effect on legal sales.

Production quality is the ability to meet tolerances and targets, or conformance, as determined by the production design (Reid and Sanders 2001). From a production standpoint, replication with minimal defect is desirable by both manufacturers (in waste reduction) and by consumers (in consistent expectations). This aligns with a general definition of quality as satisfying four conventions: value, excellence, specification conformance, and exceeding customer expectations (Reeves and Bednar 1994). As such, the ability to reproduce copies as close to the original intent represents high quality.

In the digital age, a challenge for producers is the creating of exact (or seemingly exact) copies of the genuine good from an original source (such as a DVD for movies, CD/legal download for music, or e-book for books). With music piracy, 90% of respondents perceived the

conversion of a CD song to mp3 format to be as good as the original CD version (Bhattacharjee, Gopal, and Sanders 2003). By converting from a physical source to electronic, the music files became much smaller and portable (with some audio loss), resulting in an imperceptible difference to most consumers.[1] Another example is when a person creates a copy of a movie in theaters by using a hand-held video camera to record the movie. This copy might capture comments by other audience members and jittery video from camera movement, resulting in noticeable differences with the film and hence a lower quality copy. As a closer approximation to the genuine product, the higher quality copy might be more substitutable for sales. But as we discuss next, this might not always be the case when product uncertainty exists.

Product Uncertainty, Quality, and Customer Enthusiasm

For many products, especially information goods, consumers are uncertain how well a product will perform until it is actually purchased or experienced (Nelson 1970). If uncertainty is large enough, consumers are less inclined to purchase. To alleviate uncertainty, consumers may seek out information in the market, such as movie previews or word of mouth. However, movie previews or 'film trailers' often depict the best scenes from a movie and are viewed as biased by consumers (Moul 2005). Likewise, consumers are wary of early consumer product reviews (Li and Hitt 2008), as early reviews might come from the producer or competitors and generate bias.

Consumers can also reduce product uncertainty through the illegal market. Consumers are inclined to search for more information when there is less information, especially as the number of attributes increases (Moorthy, Ratchford, and Talukdar 1997). Since information goods are experiential and more subjective in value, product quality is harder to determine before release (Hennig-Thurau, Houston, and Sridhar 2006). For complex or unknown products, illegal

versions provide information to consumers (Peitz and Waelbroeck 2006), which reduces product uncertainty.

Given that illegal copies provide information, high quality copies exhibit characteristics that more closely resemble the genuine good, further reducing product uncertainty. Product quality – for both the genuine and illegal versions – arises from attributes such as brand name, advertising, price, and product features (Zeithaml 1988). While higher quality denotes more features and/or greater degrees of a feature, quality can be objective or perceived. Objective quality describes a measurable, technical difference (e.g., Monroe and Krishnan 1985). For instance, a video with 1080P resolution provides more clarity than one with 720P resolution, and is therefore higher quality. Perceived quality is a subjective judgment, i.e., differences in taste arising from heterogeneous preferences, such as whether DTS or Dolby provides better audio.

High quality copies, then, exhibit both more features and greater objective quality. At product launch, the most enthusiastic customers may be in search of piracy to fill in this missing information (Ma et al. 2014). Since illegal copies can work as buzz agents to increase word of mouth (Qian 2015), higher quality copies more closely approximate the genuine good and foster accurate word of mouth. Because higher quality copies better resemble the genuine good, product uncertainty declines. As product uncertainty drops, the propensity to buy increases, raising revenues in the launch period.

After launch, product uncertainty diminishes as the product better permeates the market. One reason for this is that the most enthusiastic customers have likely purchased, which leaves the less enthusiastic customers remaining. When the quality gap is large between illegal copies and the genuine good, producers should not be concerned about the less enthusiastic customers, who were unlikely to purchase anyways (Qian 2014). As the post-launch period comprises less

enthusiastic consumers, there is less desire for information search, so higher quality piracy would not complement purchasing.

Underlying this timing difference is that enthusiastic customers seek information and serve as social agents for the launch period. The most interested consumers reinforce social intent and group behaviors, fostering desired consumer behaviors like purchasing and loyalty (Bagozzi and Dholakia 2006). The social interest shared by enthusiastic consumers then encourages consumption of the legal good and creates a stigma from illegal consumption. Post-launch, however, there is less social motivation to purchase and less social stigma in consuming illegal copies. The combination of less need for reducing product uncertainty, less enthusiastic consumers, and less social pressures suggest higher quality piracy should negatively affect legitimate sales post-launch.

METHODOLOGY

Since observing illegal behavior is difficult, we use a product category where the legal and illegal markets can be observed concurrently: motion pictures. We first describe the data sources and measures, then elaborate on the modeling and estimation procedures.

Data Sources

To examine the effect of piracy quality on the legal and illegal markets, we collect motion picture data from six data sources. First, a list of impending wide release movies in the U.S. and Canada was gathered from BoxOfficeMojo.com, which posts revenue and theater/screen information, from September 2013 to December 2014. All movies that opened or expanded to at

least 200 theaters were tracked for both piracy and performance (wide release movies typically open on 2,000 or more theaters: Koschmann and Bowman 2017). This yielded 173 movies which were tracked daily until weekend revenues fell below 1% of opening/expansion revenues (i.e., the motion picture had effectively reached the end of its theatrical run). Hereafter we use launch period and opening week synonymously.

Second, the Hollywood Stock Exchange (HSX: hsx.com) is a prediction market that estimates opening week revenues. Online users buy and sell 'stocks' of movies to reflect the estimated box office revenues for the first four weeks of wide release (opening/expansion). The closing 'stock price' of each film was collected prior to release and adjusted for the opening week. In this manner, the users' prediction of opening week revenues represents a proxy for demand (e.g., Elberse and Eliashberg 2003).

Third, product information comes from the Internet Movie Database (IMDB: imdb.com) daily for film attributes such as production studio, actors, production budget, genre, critical reviews, number of users rating the film, user reviews, buzz generated, release dates in other market, and Motion Picture Association of America (MPAA) rating. If the production budget data was not listed on IMDB, it was gathered from other websites.

Fourth, piracy data was observed daily at set time intervals from Pirate Bay (Pirate Bay.se), the most visited website for pirated content. Piracy searches for a film in the data set were collected using 'video' as the file type (to reduce unintended search results of 'music', 'tv shows', 'movie clips', or 'other'). The film's year of release was also part of the search to exclude similarly named motion pictures or remakes. The search results display the pirated file name and keyword signals, along with number of user downloads (leechers) and number of users with the file to share (seeders) at that time.

Fifth, advertising costs for each film come from Kantar Media's Ad$pender. The advertising expenses encompassed the twelve months leading up to and including the first week of release.

Finally, the sixth data source is actor/actress star power from the 2009 Forbes Star Power Index, the most recent survey available prior to data collection. The index surveys Hollywood executives, agents, and producers to assess how valuable a given actor/actress is for name recognition and box office revenue. Since motion pictures can take several years to develop, produce, and finish prior to launch, this data was still meaningful to the films in the data set.

Measures

Table 2 elaborates on the variables, descriptions, measures (operationalization), and data sources. *Ex ante*, the legal supplier decides how much product to supply (i.e., movie theaters decide screen allocations for a film) just prior to launch. To estimate the opening weekend revenues for a given film, the HSX prediction market serves as a market sentiment for expected demand (*Revenue_Est*). Because theater owners are unsure of demand at product launch, screen availability is allocated based on anticipated demand from the HSX. After the launch period, suppliers can adjust supply based on prior weeks' performance; week 2 is estimated with an industry average 30% drop-off in opening week revenues, while weeks 3 and onward use a double exponential smoothing model (i.e., Holt-Winters forecasting method). Since revenue decay is curved rather than linear, one parameter smooths and another accounts for the trend, giving more weight to more recent weeks, as done in prior research (Elberse and Eliashberg 2003; Koschmann and Bowman 2017). Web Appendix A further explains revenue estimation.

Table 2. Variables, Descriptions, Measures, and Sources of the Data

Variable	Description	Measure	Source
$Revenue_{it}$	Weekly revenues	Weekly box office, in $(000)	Boxoffficemojo
$Screens_{it}$	Weekly number of screens	Weekly number of screens	Boxoffficemojo
$Revenue\_Est_{it}$	Expected weekly revenues	Launch: HSX stock price two days before opening, divided by HSX multiplier, multiplied by 000,000; Post-Launch: double exponential smoothing	HSX, Boxofficemojo
$Prod\_Budget_i$	Production budget	in $(000)	IMDB, Wikipedia
$Actor\_Power_i^a$	Actor star power	Sum of actor power in a film	Forbes Star Power
$Advertising_i$	Advertising expense	Total advertising expense prior to and including launch, in $(000)	Kantar
$Critics_i$	Reviews from film critics	Metacritic rating from 1-100, divided by 20 (to get to 1-5 scale)	IMDB
$Screen\_Comp\_New_{it}^{a,b}$	Competition for screens from new releases	New releases, weighted by production budget, for every $10 million each week	Boxoffficemojo
$Screen\_Comp\_Ong_{it}^c$	Competition for screens from ongoing films	Average age, in weeks, of ongoing films of the top 25 films in the prior week	Boxoffficemojo
$Revenue\_Comp_{it}^d$	Competition for audience revenues from other films	Competitive similarity of other films based on MPAA rating and genre, weighted by week	Boxoffficemojo
WOM_{it}	Word of mouth	User rating	IMDB
$Leechers_{it}^a$	Leechers	Number of leechers, as a weekly average	Pirate Bay
$Seeders_{it}^a$	Seeders	Number of seeders, as a weekly average	Pirate Bay
$Seasonality_t$	Demand seasonality	Weekly U.S. total cinema revenues relative to the average U.S. week, based on prior 5-year average	Boxofficemojo

	Online users who rated the film	Number of online users rating the film, as a weekly average	IMDB
Num_Users$_{it}$			
Major_Studio$_i$	Distribution by a major U.S. film studio	Dummy coded if the film was released by Lions Gate, Warner Brothers, Universal, Sony/Columbia/TriStar, Fox, Paramount, or Disney	IMDB
Previous_Days$_i$	Days of prior market release	Number of days the film was released in another market prior to the U.S.	IMDB
Quality$_{it}^e$	Quality of piracy files	Average quality of film pirated files	Pirate Bay
Num_Files$_{it}$	Number of piracy files	Average number of unique film piracy files	Pirate Bay

Notes:
[a] Variable had 1 added to it, so that the log transformation was not undefined.
[b] In a given week, if movie X faces two new releases, movie Y with a budget of $50 million and movie Z with a budget of $115 million, movie X is assigned a score of 5 + 11.5 = 16.5.
[c] A higher number represents older (and presumably weaker) competition.
[d] Since many films have multiple genre and sub-genre appeal, a weighting system was used for each film. For example, *21 Jump Street* is listed as 3 genres: action, comedy, and crime. Its genre is then .33 for each, where all competing films in the top 25 that week that have any of those genre components are also weighted. When *21 Jump Street* (rated R) was in week 10 of its release and *Dark Shadows* (rated PG-13) was in week 2 of its release, *Dark Shadows* is .5 comedy and .5 fantasy, so only the .5 comedy part competes with *21 Jump Street*, so the competition score is genre/weeks (or .5/2) for .25. When *21 Jump Street* in week 10 was screening opposite week 6 of *The Cabin in the Woods* (rated R), which had genres of .33 each for Thriller, Horror, and Mystery genres (so no genre overlap with 21 Jump Street), but the MPAA rating was the same (R), then the value here is 1/6 (1 for matching genre, divided by its age, 6). Both genre and MPAA ratings were added together to get a total competition score.
[e] Standardized variable with minimum set to 0, then had 1 added to it, so that the log transformation was not undefined. A Kolmogorov-Smirnov test was conducted to test that the transformed variable distribution is statistically indifferent from the original variable distribution.

Legal supply (*Screens*) is the number of screens showing a film in a given week while

legal demand (*Revenue*) is the weekly box office revenue of a particular film. On the illegal side

of the market, illegal supply *(Seeders)* is the total number of pirated copies of a given film by

Pirate Bay users, averaged for that week. A pirated film can have different versions of varying

quality, which we subsequently describe. Since piracy can occur before product launch, we account for this as number of days the film was released in another major market before the U.S./Canada *(Previous_Days)*. Illegal demand *(Leechers)* reflects observed incidence of illegal behavior as downloads of pirated copies across all seeded versions, consistent with prior piracy research (e.g. Oberholzer-Gee and Strumpf 2007; Danaher et al. 2010).

Additional control variables used in movie research are included. Seasonality can affect motion picture demand (Vogel 2015), particularly in the summer or during holidays *(Seasonality)*. Production budget *(Prod_Budget),* film critic ratings *(Critics),* and actor star power *(Actor_Power)* speak to product quality while advertising costs (*Advertising*) pertain to promotion. Release by a major studio (*Major_Studio*) can influence distribution. Consumer sentiment as word of mouth reflects both online user ratings for valence (*WOM*) and number of online raters for volume (*Num_Users*). Several variables measure screen competition from other new releases (*Screen_Comp_New*) and existing releases (*Screen_Comp_Ong*), as well as competition for legal demand (*Revenue_Comp*) from other movies.

Measuring Piracy Quality through Observed Signals

While Table 2 presents the measures, we further address the focal variable of interest, piracy quality *(Quality)*. An issue with defining quality is the subjective nature of the construct. Despite this challenge, the pirated copies convey signals for how closely the illegal copies match the genuine product. For instance, in luxury goods such as handbags, the quality of the stitching, leather, and attention to logo can affect how similar the counterfeit matches the genuine good (e.g., Han, Nunes, and Drèze 2010). Although experts can assess these signals, a concern is that

expert opinions may differ. We measure piracy quality from observable signals as detailed below, making the distinction between perceptive objective quality and perceived quality.

With the Pirate Bay data, the illegal copies present features that meaningfully suggest quality to consumers, or online users. Figure 1 presents sample search results for the film *Edge of Tomorrow*. The first result indicates the file type is 'Movies', with keywords 'CAM' (video captured through a handheld camera), 'MKV' (a particular file container format), and 'NL.Subs' (for Dutch subtitles). The 'SE' column points to 1 seeder (one user who has that particular file), and 'LE' is the number of leechers (at that moment there were 58 users downloading that particular file). Other indicators include the time the file was uploaded (the prior day in this case), that the file size is 1.12 GB, and who uploaded the file (user 'purplefig'). Another example is the third search result; it is a 'Movies DVDR' file type, suggesting it came from a higher quality source. Signals here include '720P' (video resolution at 720 lines with progressive scan), 'TS' (telesync transfer, which is usually a handheld video recording with the film reel audio as a direct input), and 'DD2.0' (Dolby Digital surround sound with two audio channels).

Figure 1. Screenshot Sample from Pirate Bay Search Results

Type	Name (Order by: Uploaded, Size, ULed by, SE, LE)	View: Single / Double	SE	LE
Video (Movies)	Edge.of.Tomorrow.2014.CAM.MKV.NL.Subs.NLU002.purplefig Uploaded Y-day 19:54, Size 1.12 GiB, ULed by purplefig		1	58
Video (Movies)	Edge Of Tomorrow [2014] TS LiNE Audio x264-CPG Uploaded Y-day 09:17, Size 910.38 MiB, ULed by TvTeam		178	136
Video (Movies DVDR)	Edge of Tomorrow (2014) 720pTS-2-DVD DD2.0 NL Subs NLU002 Uploaded 06-21 12:11, Size 4.33 GiB, ULed by NLUPPER002		964	441
Video (Movies)	Edge.of.Tomorrow.2014.TS.Avi-Omifast[Greek Subs] Uploaded 06-24 18:41, Size 1.42 GiB, ULed by Omifast		48	14
Video (Movies)	Edge of Tomorrow 2014 TS x264 AC3 TiTAN + LEGENDA (PT-BR) Uploaded 06-23 19:30, Size 1.07 GiB, ULed by leg.ofi		1	134
Video (Movies)	No Limite Do Amanhã(Edge of Tomorrow)[2014	TS][Dublado] Uploaded 06-23 02:52, Size 823.8 MiB, ULed by messoradiola		81	28
Video (Movies)	Edge Of Tomorrow 2014 FRENCH CAM XViD - NoTag Uploaded 06-08 18:52, Size 1.37 GiB, ULed by OctoBlasT		2660	166
Video (Movies)	Edge of Tomorrow 2014 HDCAM x264 AC3 TiTAN Uploaded 06-07 19:25, Size 1.88 GiB, ULed by harks88		1301	138
Video (Movies)	Edge of Tomorrow 2014 HDCAM x264 AC3 TiTAN Uploaded 06-07 05:21, Size 1.88 GiB, ULed by makintos13		52	4
Video (Movies)	Edge Of Tomorrow 2014 Cam XviD MP3 - Idiocracy Uploaded 06-05 22:34, Size 1 GiB, ULed by Satafakap		18	2
Video (Movies)	Edge.of.Tomorrow.2014.TS.v2.XviD.AC3.2.0-RARBG Uploaded 06-12 18:27, Size 3.15 GiB, ULed by Drarbg		315	48
Video (Movies)	Edge of Tomorrow 2014 TS XVID AC3-VAiN Uploaded 06-11 03:23, Size 1.43 GiB, ULed by hotpena		108	48
Video (Movies)	Edge of Tomorrow 2014 TS x264 AC3 TiTAN Uploaded 06-10 18:08, Size 1.07 GiB, ULed by hotpena		23	7
Video (Movies)	Edge of Tomorrow 2014 TS x264 AC3 TiTAN Uploaded 06-10 18:07, Size 1.07 GiB, ULed by xxxlavalxxx		7820	1423
Video (Movies)	Edge.of.Tomorrow.2014.TS.XviD.MP3-RARBG Uploaded 06-10 17:57, Size 1.67 GiB, ULed by Drarbg		594	159
Video (Movies)	Edge Of Tomorrow 2014 HQTS x264 AC3 - SiNDK8 Uploaded 06-10 17:49, Size 1.61 GiB, ULed by hotpena		22	9
Video (HD - Movies)	Edge of Tomorrow 2014 720p TS x264 AC3-EVE Uploaded 06-10 17:09, Size 1.86 GiB, ULed by hotpena		555	310
Video (Movies)	Edge of Tomorrow 2014 TS XVID AC3-EVE Uploaded 06-10 16:36, Size 1.33 GiB, ULed by hotpena		18	9
Video (Movies)	Edge.of.Tomorrow.2014.CAM.XviD-VAiN Uploaded 06-08 23:06, Size 1.38 GiB, ULed by Drarbg		38	9

Other signals of quality include the skull icons (if the file was uploaded by a trusted or VIP user) and word balloon to denote the file has user comments. However, these are less consequential as the same file can be uploaded by different users (the eighth and ninth search results illustrate this). Web Appendix B describes further the initial sources of piracy files and

the inherent quality of each (e.g., copies made from a handheld camera are generally believed to be lower quality than those made directly from a film reel or DVD transfer).

Some keywords should suggest greater objective quality (e.g., '1080P' video is higher resolution than '720P' video), but other keywords are rooted in perceived quality (e.g., 'DTS' and 'DD' are competing multichannel audio technologies that both support 5.1 channel surround sound). We dichotomize thirty-four of the most common piracy keyword signals as present or not in each pirated file (a file count of all piracy keywords initially considered is in Web Appendix C). Furthermore, the combination of keywords can jointly signal quality, such as whether a pirated file with 720P video and DTS audio is a better quality copy than one with 1080P video and AC3 audio.

The piracy keywords represent specific attributes that together suggest overall quality in a piracy file. Prior research treats quality as a higher order global assessment (Olshavsky 1985; Holbrook and Corfman 1985). At this higher level, quality is a composite of elements consumers perceive, such as price and attributes in products (Zeithaml 1988). While quality might have several underlying factors, our focus is scale construction where estimates of each piracy keyword map onto that quality scale.

As each piracy keywords is either present or not, these jointly manifest as a latent, continuous spectrum of quality. Factor analysis may be inappropriate for unidimensional ideal point estimation (e.g., Van Schuur and Kiers 1994; Spector et al. 1997), and factor analysis of dichotomous data may induce artificial factors (e.g., Kubinger 2003).[2] Thus, the preferred model for this is an item response theory (IRT) model (Bartholomew et al. 2002). IRT models uncover latent relationships by inferring from the observed responses (Lord 2012). Uses include measuring student ability given exam difficulty in education, or legislators' liberal/conservative

views given voting patterns in political science (Jackman 2008). In marketing, the model has received not as much attention, but an example is consumers' willingness to redeem coupons conditioned on discount levels (Swaminathan and Bawa 2005). Given this, the IRT model uncovers each piracy keyword's propensity for quality.

The quality of each pirated keyword (and by extension, the quality of each pirated file as the presence or absence of those keywords), arises from the underlying correlations of the keywords relative to each other. The dichotomous nature of the presence/absence of keywords yields a tetrachoric correlation matrix (provided in Web Appendix D). For example, file types 'FT_HDMOVIES' and 'FT_HHELD' have a completely negative relationship ($r = -0.95$) as these two file types should never overlap. Audio quality keyword '5.1' for five channel audio correlates well with '1080P' ($r = 0.67$) but not '720P' ($r = -0.04$). The underlying relative relationships of keywords to each other allow the model determine the dimension of quality.

To estimate the IRT, the model assumes each piracy file i has an unobserved (latent) quality Θ_i. Across files, each piracy keyword j has an unobserved appeal that corresponds to keyword parameter b_j (that is, higher b_j suggests higher quality). Piracy quality is jointly determined by Θ_i and b_j, such that each b_j represents an ideal point on the piracy quality spectrum (i.e., if $b_1 < b_2$, then b_2 signifies higher quality). The parameters are estimated by measuring the probability (x) that piracy file i contains keyword j, as Θ_i and b_j are estimated simultaneously from these piracy keyword probability distributions. The model includes one more parameter, a_j, as the slope of this logistic regression, which is the ability of the keyword to discriminate between high and low piracy quality (i.e., higher a_j means easier separation in quality). The resulting IRT model is represented by Equation 1:

$$(1) \qquad p_{ij}\left(x_{ij} = 1 | \theta_i, a_j, b_j\right) = \frac{e^{a_j(\theta_i - b_j)}}{1 + e^{a_j(\theta_i - b_j)}}$$

To identify the model, one piracy keyword is marked at one of the far ends of the latent spectrum, such as very low or very high quality. From this, the quality of all the other piracy keywords should lie either to the left or the right of this item on the spectrum. In the coupon redemption study (Swaminathan and Bawa 2005), model identification comes from setting the coupon with the greatest dollar amount at the far left end of the spectrum (e.g., most consumers, will redeem a $4.00 coupon, but not a $0.25 coupon).

The Pirate Bay data encompasses 8,701 unique piracy files across the films studied. The IRT estimates each keyword's ideal point to quality using a Bayesian Markov Chain Monte Carlo (MCMC) process with 40,000 iterations, 5,000 iteration burn-in, and thinning every fifth draw. This results in 7,000 posterior draws per piracy keyword. For identification, we set 'FT_HHELD' at the far left end (i.e., very low quality) as piracy files that are captured through handheld recording devices are believed to be low quality. The estimated mean quality ideal points and posterior standard deviations are shown in Table 3.

Table 3. IRT Ideal Point Results of Piracy Keywords

Keyword	Description	Mean	SD
FT_MOVIES	file type is movie	1.836	0.026
FT_3D	file type is 3D	-2.066	0.510
FT_HHELD	file type is handheld	-1.842	0.356
FT_HDMovies	file type is high definition movie	0.652	0.016
FT_DVDR	file type is DVD movie	-0.722	0.105
F_2.0	audio (2 channels)	-1.445	0.234
F_5.1	audio (5.1 channels)	-0.284	0.047
F_AAC	audio (AAC format)	0.582	0.019
F_AC3	audio (AC3 format)	0.827	0.012
F_DTS	audio (DTS channels)	0.114	0.040

F_264	container (264 type)	0.803	0.012
F_MKV	container (MKV type)	-0.507	0.086
F_MP3	audio (MP3 format)	0.446	0.025
F_MP4	audio (MP4 format)	0.080	0.045
F_XVID	container (XVID type)	1.274	0.012
F_SUB	has subtitles	0.534	0.020
F_720P	video (720P resolution)	0.750	0.014
F_1080P	video (1080P resolution)	0.217	0.033
F_CAM	source (camcorder transfer)	0.813	0.013
F_TC	source (telecine transfer)	-1.693	0.435
F_TS	source (telesync transfer)	0.447	0.024
F_SCR	source (screener transfer)	0.663	0.015
F_DVD	source (DVD transfer)	1.157	0.011
F_BR	source (Blu-Ray transfer)	0.822	0.012
F_DIVX	container (DIVX type)	0.138	0.048
F_AVI	container (AVI type)	0.038	0.045
F_HQ	"high quality"	-0.167	0.070
F_V2	second version of a file	-0.556	0.108
F_V3	third version of a file	-1.855	0.405
F_R5	source (region 5 DVD)	-0.914	0.154
F_R6	source (region 6 DVD)	-0.105	0.064
F_RIP	source (ripped from a physical copy)	1.266	0.012
F_LINE	source (line input)	-1.326	0.238
F_BD	source (Blu-Ray disc transfer)	0.027	0.045

Note: results are MCMC posterior draws of 40,000 iterations, with 5,000 burn-in iterations and thinning every 5th draw (thus $N = 7,000$ per keyword). Prefix 'FT' denotes file type and 'F' denotes file keyword.

To aid interpretation, the IRT results are analogous to standard deviations from a mean of zero. Figure 2 plots the IRT mean and posterior standard deviations of the piracy keywords from lowest to highest mean quality. The results confirm some prior beliefs regarding piracy quality. File keywords720P video (F_720P: *Mean* = 0.750, *SD* = 0.014), DVD quality (F_DVD: *Mean* = 1.157, *SD* = 0.011), AC3 audio (F_AC3: *Mean* = 0.827, *SD* = 0.012), and high definition movie sources (FT_HDMovies: *Mean* = 0.652, *SD* = 0.016) each suggest higher quality. Some items also expectedly suggest lower quality. Telecine copies that come from film reel transfers (F_TC:

Mean = -1.693, *SD* = 0.435), adding in a separate English audio track if the audio is not in English (F_LINE: *Mean* = -1.326, *SD* = 0.238), and copies that come from region 5 DVD sources such as Russia and most of Asia (F_R5: *Mean* = -0.914, *SD* = 0.154) exhibit low quality.

Some quality results were also surprising. Camcorder sources were positive (F_CAM: *Mean* = 0.813, *SD* = 0.013), designations of 'high quality' were negative (F_HQ: *Mean* = -0.167, *SD* = 0.070), and files sourced from DVDs (FT_DVDR: *Mean* = -0.722, *SD* = 0.105) were also negative. Possible explanations for this are that pirated files often contain multiple keywords, so handheld sources can be supplemented with higher quality audio and file containers. Additionally, the films in the sample were not legally released on the secondary market for home video consumption (i.e., DVD and Blu-Ray), so suggesting a DVD source exists when it legally does not might generate skepticism of its quality. This can be reconciled with the DVD keyword above (R5) where a file can describe itself as having DVD-like quality (with the right keywords) yet not originate from a DVD source.[3] As a robustness check, we re-estimated the IRT to identify '720P' as higher quality. The results did not materially change from those reported in Table 3. The resulting quality estimates across keywords and pirated copies then enters as a covariate in the launch and post-launch models.

Figure 2. IRT Results of Piracy Quality Signals and Ideal Points (Mean and Standard Deviation)
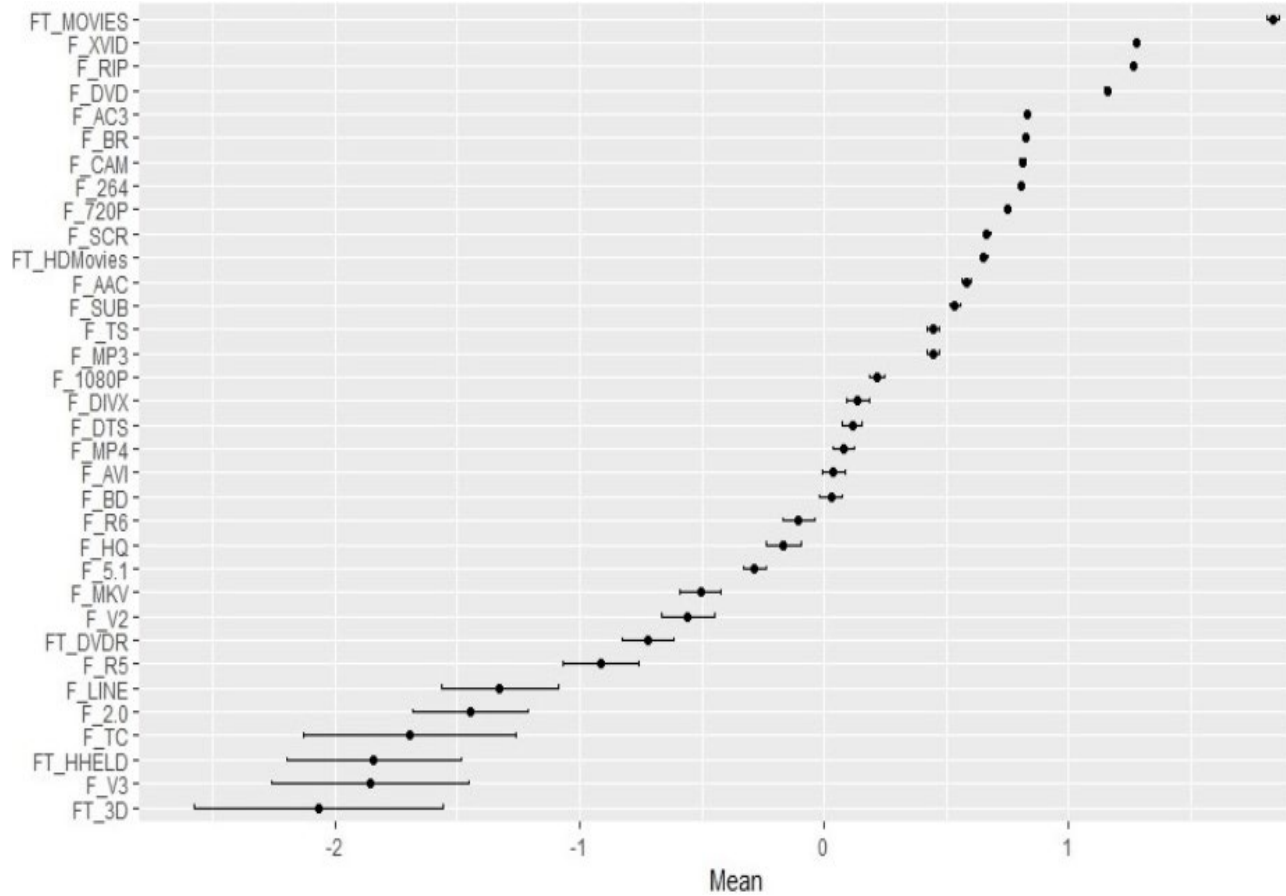
Launch Model

To estimate the effect of piracy quality on the market, we model an interdependent system of equations with legal supply and demand plus illegal supply and demand. For many products, and especially information goods like motion pictures, the launch period is different from the post-launch period. The system has four equations:

(2) $\ln(\text{Screens}_{i1}) = \alpha_0 + \alpha_1\ln(\text{Revenue\_Est}_{i1}) + \alpha_2\ln(\text{Prod\_Budget}_{i1}) + \alpha_3\ln(\text{Actor\_Power}_{i1}) + \alpha_4\ln(\text{Advertising}_{i1}) + \alpha_5\ln(\text{Critics}_{i1}) + \alpha_6\text{Major\_Studio}_{i1} + \alpha_7\ln(\text{Screen\_Comp\_New}_{i1}) + \alpha_8\ln(\text{Screen\_Comp\_Ong}_{i1}) + \alpha_9\ln(\text{Previous\_Days}_{i1}) + \varepsilon_{Si1}$

(3) $\ln(\text{Revenue}_{i1}) = \beta_0 + \beta_1\ln(\text{Screens}_{i1}) + \beta_2\ln(\text{Prod\_Budget}_{i1}) + \beta_3\ln(\text{Actor\_Power}_{i1})$

$$+ \beta_4 \ln(Advertising_{i1}) + \beta_5 \ln(Critics_{i1}) + \beta_6 Major\_Studio_{i1} + \beta_7 \ln(Revenue\_Comp_{i1}) +$$

$$\beta_8 Seasonality_{i1} + \beta_9 \ln(WOM_{i1}) + \beta_{10} \ln(Num\_Users_{i1}) + \beta_{11} \ln(Quality_{i1}) + \beta_{12} \ln(Seeders_{i1}) +$$

$$\beta_{13} \ln(Leechers_{i1}) + \beta_{14} \ln(Quality_{i1})*\ln(Seeders_{i1}) + \beta_{15} \ln(Quality_{i1})*\ln(Leechers_{i1}) + \varepsilon_{Ri1}$$

$$(4) \qquad \ln(Seeders_{i1}) = Ɣ_0 + Ɣ_1 \ln(Screens_{i1}) + Ɣ_2 \ln(Revenue_{i1}) + Ɣ_3 \ln(Prod\_Budget_{i1}) +$$

$$Ɣ_4 \ln(Actor\_Power_{i1}) + Ɣ_5 \ln(Advertising_{i1}) + Ɣ_6 \ln(Critics_{i1}) + Ɣ_7 Major\_Studio_{i1} +$$

$$Ɣ_8 \ln(Previous\_Days_{i1}) + Ɣ_9 \ln(WOM_{i1}) + Ɣ_{10} \ln(Num\_Users_{i1}) + Ɣ_{11} \ln(Quality_{i1}) +$$

$$Ɣ_{12} \ln(Leechers_{i1}) + Ɣ_{13} \ln(Quality_{i1})*\ln(Leechers_{i1}) + +\varepsilon_{Pi1}$$

$$(5) \qquad \ln(Leechers_{i1}) = \lambda_0 + \lambda_1 \ln(Screens_{i1}) + \lambda_2 \ln(Revenue_{i1}) + \lambda_3 \ln(Prod\_Budget_{i1}) +$$

$$\lambda_4 \ln(Actor\_Power_{i1}) + \lambda_5 \ln(Advertising_{i1}) + \lambda_6 \ln(Critics_{i1}) + \lambda_7 Major\_Studio_{i1} + \lambda_8 Seasonality_{i1} +$$

$$\lambda_9 \ln(WOM_{i1}) + \lambda_{10} \ln(Num\_Users_{i1}) + \lambda_{11} \ln(Quality_{i1}) + \lambda_{12} \ln(Seeders_{i1}) +$$

$$\lambda_{13} \ln(Quality_{i1})*(Seeders_{i1}) + \varepsilon_{Li1}$$

The model uses a multiplicative framework, log-transforming variables in accord with prior movie research (Elberse and Eliashberg 2003; Clement, Wu, and Fischer 2014; Koschmann and Bowman 2017). The system of equations treats legal supply as the starting point: illegal copies enter the market after the legal product has launched.[4] For motion pictures, theaters allocate screens in advance of a film's release in order to arrange show times to meet expected demand. Subscripts i denote the film and t for the launch period (here week $t = 1$). The error term of each equation, ε, is additionally subscripted S, R, P, L to denote the screens, revenue, seeders, and leechers equations, respectively. Equations (2)-(5) use typical motion picture control variables: production budget, star power, advertising, critic ratings, and an indicator for release by a major studio. Additionally, word of mouth includes not only the valence but also the volume; consumer sentiment as well as number of consumers talking about a given film (e.g.,

You, Vadakkepatt, and Joshi 2015) may affect demand and supply. Release by a major studio (binary coded) and seasonality (average week relative percentage) are not log transformed.

Post-Launch Model

The post-launch system of Equations (6)-(9) is similar to Equations (2)-(5), where $t > 1$, and Greek uppercase letters distinguish post-launch coefficients from the launch period:

(6) $\ln(Screens_{it}) = A_0 + A_1\ln(Revenue\_Est_{it}) + A_2\ln(Screen\_Comp\_New_{it}) +$

$A_3\ln(Screen\_Comp\_Ong_{it}) + A_4\ln(WOM_{it}) + A_5\ln(Num\_Users_{it}) + A_6\ln(Quality_{it-1}) +$

$A_7\ln(Seeders_{it-1}) + A_8\ln(Leechers_{it-1}) + A_9\ln(Quality_{it-1})*(Seeders_{it-1}) +$

$A_{10}\ln(Quality_{it-1})*(Leechers_{it-1}) + A_S D_{Sit} + \varepsilon_{Sit}$

(7) $\ln(Revenue_{it}) = B_0 + B_1\ln(Screens_{it}) + B_2\ln(Revenue\_Comp_{it}) + B_3 Seasonality_{it} +$

$B_4\ln(WOM_{it}) + B_5\ln(Num\_Users_{it}) + B_6\ln(Quality_{it}) + B_7\ln(Seeders_{it}) + B_8\ln(Leechers_{it}) +$

$B_9\ln(Quality_{it})*\ln(Seeders_{it}) + B_{10}\ln(Quality_{it})*(Leechers_{it}) + B_R D_{Rit} + \varepsilon_{Rit}$

(8) $\ln(Seeders_{it}) = \Gamma_0 + \Gamma_1\ln(Screens_{it}) + \Gamma_2\ln(Revenue_{it}) + \Gamma_3\ln(WOM_{it}) + \Gamma_4\ln(Num\_Users_{it})$

$+ \Gamma_5\ln(Quality_{it}) + \Gamma_6\ln(Leechers_{it}) + \Gamma_7\ln(Quality_{it})*(Leechers_{it}) + \Gamma_P D_{Pit} + \varepsilon_{Pit}$

(9) $\ln(Leechers_{it}) = \Lambda_0 + \Lambda_1\ln(Screens_{it}) + \Lambda_2\ln(Revenue_{it}) + \Lambda_3 Seasonality_{it} + \Lambda_4\ln(WOM_{it})$

$+ \Lambda_5\ln(Num\_Users_{it}) + \Lambda_6\ln(Quality_{it}) + \Lambda_7\ln(Seeders_{it}) + \Lambda_8\ln(Quality_{it})*(Seeders_{it}) +$

$\Lambda_L D_{Lit} + \varepsilon_{Lit}$

In the post-launch period, time-invariant variables are excluded and time dummies (*D*) are added to account for time-specific fixed effects. The coefficients on *D* in each equation (A_S, B_R, Γ_P, Λ_L) are a vector of estimates for each week. As legal supply is again the initial starting

point for the week, Equation (6) also includes seeders, leechers, and quality from the prior week, as previous research finds that supply follows demand (Krider et al. 2005), legal or illegal.

Correction for Endogeneity

A modeling concern is whether the dependent measures, as regressors, may be correlated with the error terms. For instance, a studio might counter lower than expected piracy by seeking an increase in screen allocation. Although studios typically want more screen availability to increase distribution for consumers, more showings also increase opportunities for in-theater piracy. Quality may also be correlated with the error terms; more higher-quality copies might entice more consumers to download, which affects pirates' incentives to create and share copies.

To address endogeneity concerns, we model the correlation between the error terms and potentially endogenous regressors (screens, revenues, seeders, leechers, and quality) using Gaussian copulas. Like instrumental variables, copulas parcel out the exogenous variation from an endogenous regressor, becoming more common in marketing research (e.g., Park and Gupta 2012; Schweidel and Knox 2013; Datta, Foubert, and Van Heerde 2015). Following prior work (Park and Gupta, 2012; Papies, Ebbes, and Van Heerde 2017), we generate copula-transformed terms:

(10) $$\ln\left(\widetilde{\text{Screens}}_{it}\right) = \Phi^{-1}[H_{\ln(\text{Screens})} \ln(\text{Screens})]$$

(11) $$\ln\left(\widetilde{\text{Revenue}}_{it}\right) = \Phi^{-1}[H_{\ln(\text{Revenue})} \ln(\text{Revenue})]$$

(12) $$\ln\left(\widetilde{\text{Seeders}}_{it}\right) = \Phi^{-1}[H_{\ln(\text{Seeders})} \ln(\text{Seeders})]$$

(13) $$\ln\left(\widetilde{\text{Leechers}}_{it}\right) = \Phi^{-1}[H_{\ln(\text{Leechers})} \ln(\text{Leechers})]$$

(14) $$\ln\left(\widetilde{\text{Quality}}_{it}\right) = \Phi^{-1}[H_{\ln(\text{Quality})} \ln(\text{Quality})]$$

where Φ^{-1} is the inverse normal cumulative distribution function and H(•) are the empirical cumulative distribution functions of the log transformed terms screens, revenues, seeders, leechers, and quality, respectively. Copula-transformed terms for the interactions of quality and seeders, and quality and leechers, are similarly created. Since the copulas are estimated densities, Park and Gupta (2012) suggest bootstrapping the data by sampling with replacement N draws of M samples, where N is the original sample size. In order to identify the model, the endogenous regressor must be non-normal in its distribution (Park and Gupta 2012). We generate 500 bootstrap samples. A Shapiro-Wilk test of the opening week finds only revenues are normally distributed ($W > .984$, $p > .05$) in 11 draws, so these samples are excluded from analysis. In the post-launch period, all endogenous regressor draws are non-normal ($W < .992$, $p < .01$), allowing all 500 draws to be included for analysis.

Descriptive Evidence

The estimates for each piracy keyword, as reported in Table 3, manifest in each piracy file as overall piracy quality measure. For the 8,701 unique piracy files, average quality is positive (*Mean* = 4.365, *SD* = 1.425, *Median* = 4.732), as the quality ranges from -2.066 to 7.588. Negative piracy quality means the pirated file exhibited much lower quality signals than higher quality signals in its file description. We average the quality for each film-week and then standardizes quality with a minimum of zero.

Table 4 presents descriptive statistics of the variables. Since motion pictures release weekly, we average daily piracy measures across all files for a given film to get a weekly figure. A total of 249,440 film-day-file observations were collected. Average quality for a given film in

the opening week is *Mean* = 4.66 (*SD* = 3.17). In total, 90.9% of the films in the sample

exhibited illegal copies during the theatrical run (i.e., 9% of the movies in the sample had no

pirated files on Pirate Bay). Although the piracy data is collected globally (i.e. users can

download illegal copies anywhere), the correlation of global revenues with U.S. and Canada

revenues is $r = 0.92$, suggesting global revenues may be similarly impacted by piracy.

Table 4. Summary Statistics by Product Period

Launch Period (N = 173)

	Mean	Median	SD	Min	Max
Screens	3,609.62	3,200.00	2,723.37	210.00	12,600.00
Revenue	$26,509.79	$14,366.97	$33,143.10	$289.61	$222,116.06
Seeders	215.47	9.75	329.36	0.00	1,795.53
Leechers	147.44	29.50	237.83	0.00	1,468.11
Quality	4.66	5.99	3.17	0.00	8.84
Prod_Budget	$47,688.48	$28,000.00	$52,273.61	$1,000.00	$255,000.00
Advertising	$13,274.80	$12,355.85	$9,692.61	$0.32	$37,901.70
Actor_Power	5.77	6.53	2.87	0.00	10.00
Critics	50.70	49.57	16.75	13.57	97.00
Previous_Days	7.25	2.00	19.24	0.00	223.00
Major_Studio	0.57	1.00	0.50	0.00	1.00
Revenue_Comp	3.38	3.17	1.50	0.30	9.32
Screen_Comp_New	10.34	8.80	7.95	0.00	41.00
Screen_Comp_Ong	5.61	5.56	0.94	3.60	8.20
Seasonality	0.98	0.90	0.30	0.56	1.82
WOM	6.70	6.80	1.25	1.36	8.90
Num_Users	6,794.74	2,016.79	11,889.72	57.00	61,343.29

Post-Launch Period (N = 1,204)

	Mean	Median	SD	Min	Max
Screens	1,542.84	775.00	1,806.48	5.00	11,500.00
Revenue	$4,694.45	$1,263.80	$8,877.06	$4.43	$87,548.90
Seeders	261.29	225.63	252.35	0.00	3,124.90
Leechers	87.69	55.43	123.46	0.00	1,480.50
Quality	5.41	6.02	2.23	0.00	8.84
Revenue_Comp	3.71	3.49	2.20	0.11	56.00
Screen_Comp_New	14.23	13.70	8.51	0.50	41.00
Screen_Comp_Ong	5.57	5.36	1.03	3.60	8.20
Seasonality	0.97	0.90	0.28	0.56	1.82
WOM	6.89	7.00	1.18	1.46	8.90

| Num_Users | 30,258.76 | 9,479.29 | 46,009.72 | 107.86 | 297,047.71 |

Notes. Dollars are in thousands (000).

EMPIRICAL RESULTS

Estimation of both launch and post-launch systems of equations utilizes seemingly unrelated regression (SUR), allowing the error terms of the equations to correlate for efficiency (Zellner and Theil 1962). The error terms may be correlated across equations for other exogenous factors that could "shock" both the legal and illegal sides of the market (e.g., award nominations may affect both supply and demand: Elberse and Eliashberg 2003).

Launch Estimation Results

Table 5 reports the SUR model estimates for the launch period system of Equations (2)-(5). The system weighted R^2 is 0.979, indicating high fit among the four interdependent parts of the market. Since both sides of the equation are log transformed, the coefficients are interpreted as elasticities. Although we focus on the effects of piracy quality, the control variables are consistent with those reported in existing motion picture research. Notably, in the screens equation, anticipated revenues ($\alpha_1 = 0.490, p < .01$), advertising expense ($\alpha_4 = 0.164, p < .01$), and film critics reviews ($\alpha_5 = -0.371, p < .01$) are significant and in the same direction as those found elsewhere (Elberse and Eliashberg 2003; Clement, Wu, and Fischer 2014; Koschmann and Bowman 2017). In the revenues equation, screens ($\beta_1 = 1.335, p < .01$) and film critics reviews ($\beta_5 = 0.441, p < .01$) are significant and in the same direction as those prior studies.

Table 5. Launch Period SUR Estimation Results

Variable	DV:ln(Screens) Estimate		DV:ln(Revenue) Estimate		DV:ln(Seeders) Estimate		DV:ln(Leechers) Estimate	
Intercept	0.895	***	-1.073		4.596	*	-2.643	
	(0.279)		(2.033)		(2.552)		(1.856)	
ln(Revenue)[a]	0.490	***			-0.097		0.212	
	(0.044)				(0.331)		(0.255)	
ln(Screens)			1.335	***	-0.643	*	0.233	
			(0.262)		(0.364)		(0.259)	
ln(Prod_Budget)	0.153	***	-0.204	***	0.074		-0.013	
	(0.032)		(0.078)		(0.103)		(0.078)	
ln(Actor_Power)	-0.014		0.009		0.087		-0.052	
	(0.025)		(0.045)		(0.067)		(0.044)	
ln(Advertising)	0.164	***	-0.020		0.057		-0.060	
	(0.052)		(0.095)		(0.075)		(0.054)	
ln(Critics)	-0.371	***	0.441	***	0.195		-0.205	
	(0.066)		(0.165)		(0.237)		(0.159)	
Major_Studio	0.002		0.230	**	0.093		-0.120	
	(0.050)		(0.115)		(0.157)		(0.113)	
ln(Screen_Comp_New)	-0.099	***						
	(0.020)							
ln(Screen_Comp_Ong)	0.015							
	(0.130)							
ln(Previous_Days)	-0.039	**			0.062	**		
	(0.020)				(0.025)			
ln(Revenue_Comp)			-0.075					
			(0.088)					
Seasonality			0.485	***			-0.118	
			(0.151)				(0.091)	
ln(WOM)			0.287		-0.938	**	0.716	**
			(0.280)		(0.391)		(0.276)	
ln(Num_Users)			0.058		0.102		-0.112	**
			(0.054)		(0.072)		(0.051)	
ln(Quality)			-0.205		-0.593		0.620	*
			(0.290)		(0.466)		(0.358)	
ln(Seeders)			0.878	*			0.349	
			(0.466)				(0.276)	
ln(Leechers)			-0.814		1.837	***		
			(0.529)		(0.412)			
ln(Quality)*ln(Seeders)			-0.454	**			0.127	
			(0.227)				(0.137)	
ln(Quality)*ln(Leechers)			0.519	**	-0.225			
			(0.263)		(0.210)			
ln(Revenue) copula					0.074		0.040	
					(0.368)		(0.273)	
ln(Screens) copula			0.011		0.457	*	-0.321	
			(0.205)		(0.266)		(0.201)	
ln(Quality) copula			0.091		0.219		-0.239	
			(0.170)		(0.271)		(0.215)	
ln(Seeders) copula			-0.604	*			0.658	***
			(0.356)				(0.205)	

ln(Leechers) copula		0.442	-0.731 **	
		(0.437)	(0.308)	
ln(Quality)*ln(Seeders) copula		0.406		-0.252
		(0.280)		(0.199)
ln(Quality)*ln(Leechers) copula		-0.419	0.333	
		(0.395)	(0.321)	
System Weighted R^2	0.979			

Notes. Standard errors in parentheses. [a] is expected value in Screens equation.
*** $p < 0.01$; ** $p < 0.05$;* $p < 0.10$.

Focusing on piracy, revenues are marginally affected by illegal supply ($\beta_{12} = 0.878$, $p < .07$), but not influenced by illegal demand ($\beta_{13} = -0.814$, $p > .12$). The first order effect of piracy quality ($\beta_{11} = -0.205$, $p > .48$) is not particularly interpretable. That is, the mere presence of piracy quality is not of interest, only when it manifests into consumer activity (such as downloads or uploads of high-quality copies) that is important. As more high-quality copies are downloaded, this has a positive effect on revenues ($\beta_{15} = 0.519$, $p < .05$). This positive effect aligns with our expectation that consumption of higher quality illegal copies in the launch period works as a sampling mechanism. Yet, pulling in the opposite direction is that higher quality supply hurts revenues ($\beta_{14} = -0.454$, $p < .05$). Thus, while downloads of higher quality copies provides an average positive effect on revenues, the availability of too many high-quality copies has an average negative effect.

Although piracy quality has an effect on revenues, of additional interest is its effect on the illegal side of the market. In the seeders equation, the main effect of downloading ($\gamma_{12} = 1.837$, $p < .01$) exhibits a significant relationship with seeding, yet downloading higher quality copies has no significant effect ($\gamma_{13} = -0.225$, $p > .28$). This indicates that demand for higher quality copies does not spur piracy sharing. In theory, consumers would seek out higher quality copies, which would incentivize pirates to increase more copies. We find no evidence of this in

the opening week. One explanation for this may be that high quality copies are limited at first, as the product just recently entered the market.

In the leechers equation, the main effect of available piracy ($\lambda_{12} = 0.349$, $p > .20$) is not significant. Also not significant is the interaction term for the presence of high quality copies ($\lambda_{13} = 0.127$, $p > .35$). Thus, higher quality seeding does not correspond to higher downloading incidence, albeit an overall effect of piracy quality on revenue is a function of seeders and leechers. Figure 3 illustrates the effect of quality on revenue with respect to seeders and leechers using the first-order and interaction terms from Table 5; the effect is greatest with few seeders, but many leechers. But, quality's effect is most negative with many seeders and few leechers.

Figure 3. Launch Effect of Quality on Revenue as a Function of Seeders and Leechers
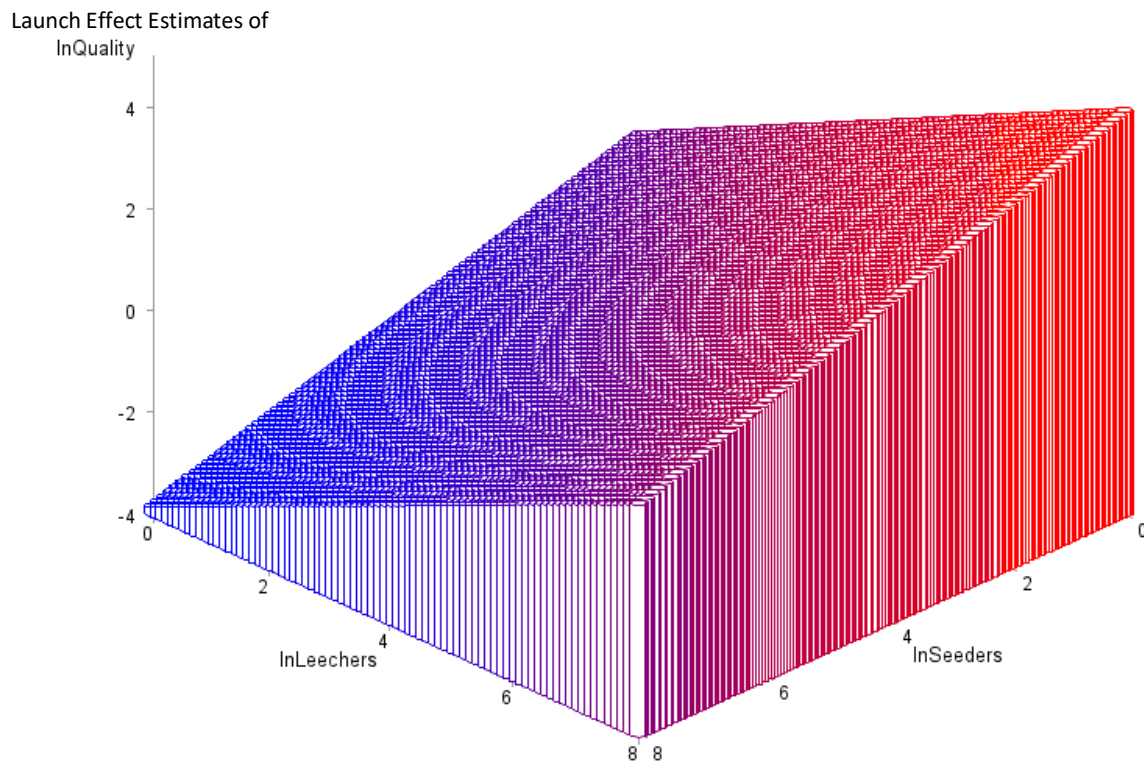
Post-Launch Estimation Results

The post-launch system of equations, Equations (6)-(9), is arranged like the opening week, with legal supply the starting point for the week. Table 6 presents the SUR post-launch estimates. Like the opening week, the system weighted R^2 of 0.982 exhibits similarly high fit. Whereas piracy presumably had little effect on screen allocation prior to a film's release, piracy effects from the prior week after launch might influence screen allocation in the current week. Prior week piracy availability ($A_7 = -0.344, p > .35$) showed no significant effect as a substitute for legal supply, although downloading had a marginally positive effect ($A_8 = 0.717, p < .10$). Together with online activity, an increase in higher quality supply ($A_9 = 0.201, p > .28$) had no effect on screens. As such, higher quality copies did not act as a substitute for legal supply. However, downloads of higher quality copies ($A_{10} = -0.364, p < .10$) had a marginally negative impact on screen availability. One explanation for these results, like our managerial interviews, is that theater owners are not overly concerned about piracy supply (including high quality piracy files). These findings align with their belief. Increased demand for illegal copies in the prior week may help screen allocation through demand as sampling, but the trade-off with higher quality downloads is that these might substitute sales through expected revenues.

Table 6. Post-Launch Period Estimation Results

Variable	DV:ln(Screens) Estimate		DV:ln(Revenue) Estimate		DV:ln(Seeders) Estimate		DV:ln(Leechers) Estimate	
Intercept	1.532	***	-0.585		-2.588	***	1.867	***
	(0.257)		(0.394)		(0.460)		(0.419)	
ln(Revenue)[a]	0.538	***			-0.053		-0.003	
	(0.018)				(0.121)		(0.105)	
ln(Screens)			1.037	***	0.068		-0.063	
			(0.037)		(0.099)		(0.087)	
ln(Screen_Comp_New)	-0.066	***						
	(0.022)							
ln(Screen_Comp_Ong)	0.287	***						
	(0.091)							
ln(Revenue_Comp)			-0.006					

	(1)	(2)	(3)	(4)
		(0.018)		
Seasonality		0.540 ***		-0.003
		(0.037)		(0.017)
ln(WOM)	-0.346 **	0.817 ***	-0.308 ***	0.225 **
	(0.148)	(0.074)	(0.093)	(0.088)
ln(Num_Users)	0.162 ***	0.003	0.042 ***	-0.037 ***
	(0.020)	(0.012)	(0.014)	(0.013)
ln(Quality)[b]	0.117 ***	-0.427 ***	0.678 ***	-0.374 **
	(0.035)	(0.117)	(0.102)	(0.146)
ln(Seeders)[b]	-0.344	0.385		1.297 ***
	(0.372)	(0.403)		(0.368)
ln(Leechers)[b]	0.717 *	0.753 **	1.142 ***	
	(0.432)	(0.360)	(0.314)	
ln(Quality)*ln(Seeders)[b]	0.201	-0.225		-0.311 *
	(0.187)	(0.201)		(0.183)
ln(Quality)*ln(Leechers)[b]	-0.364 *	-0.380 **	0.208	
	(0.219)	(0.183)	(0.156)	
ln(Revenue) copula			0.077	0.038
			(0.172)	(0.151)
ln(Screens) copula		0.280 ***	-0.100	0.091
		(0.056)	(0.101)	(0.090)
ln(Quality) copula		0.173 ***	-0.299 ***	0.214 ***
		(0.058)	(0.050)	(0.071)
ln(Seeders) copula		0.113		0.260 **
		(0.147)		(0.120)
ln(Leechers) copula		-0.879 ***	-1.253 ***	
		(0.303)	(0.222)	
ln(Quality)*ln(Seeders) copula		0.892 ***		-0.063
		(0.294)		(0.127)
ln(Quality)*ln(Leechers) copula		0.059	0.773 ***	
		(0.152)	(0.218)	
System Weighted R^2	0.982			

Notes. Standard errors in parentheses. Weekly time dummies not shown. [a] is expected value in Screens equation. [b] is lagged in Screens equation.
*** $p < 0.01$; ** $p < 0.05$;* $p < 0.10$.

With revenues, the availability of illegal copies has no significant effect ($B_7 = 0.385$, $p > .34$), although illegal demand is positive ($B_8 = 0.753$, $p < .04$). This suggests illegal copy downloads act more as a sampling mechanism after a movie release. However, high quality downloads inhibit revenues ($B_{10} = -0.380$, $p < .05$). The availability of high-quality copies ($B_9 = -0.225$, $p > .26$) shows no effect on revenues post-launch. Unlike the launch period, post-launch

consumption of high-quality copies exhibits a negative effect on revenues. This aligns with our belief that post-launch, copies more closely resembling the legal good act as substitutes.

Although seeders showed no effect on revenues, piracy downloads ($\Gamma_6 = 1.142$, $p < .01$) encourage illegal supply. However, downloads of higher quality copies ($\Gamma_7 = 0.208$, $p > .18$) does not affect seeding. Given the linkage of file sharing between seeders and leechers, we expected demand for higher quality copies to facilitate more piracy copies. One possibility for this is that consumers may be less inclined to seek high quality copies post-launch as there is more information available about a movie (such as through word of mouth). If post-launch consumers are not information seeking, they might accept any copy to supplant willingness to pay; lower quality copies are likely smaller in file size, making downloading faster. Additionally, there might be less 'reward' (i.e., street credibility) for a pirate in creating a high quality copy if the illegal market has other high quality copies available.

Demand for illegal copies post-launch is positively affected by seeders ($\Lambda_7 = 1.297$, $p < .01$). The availability of higher quality copies ($\Lambda_8 = -0.311$, $p < .10$) has a marginally negative effect on downloading. This indicates that post-launch, consumers looking for illegal copies are more inclined to download lower quality copies. Although there should be more high-quality copies available after launch, this might be explained through information seeking. In the opening week, consumers sought high quality copies for more information about the legal good, but post-launch there is more information available. Another possibility is that demand for a film (both legally and illegally) just naturally declines over time. Figure 4 highlights the post-launch effect of quality on revenue as a function of the seeders and leechers. Unlike the opening week, the effect of piracy quality on revenue is greatest here when both seeders and leechers are few.

Figure 4. Post-Launch Effect of Quality on Revenue as a Function of Seeders and Leechers
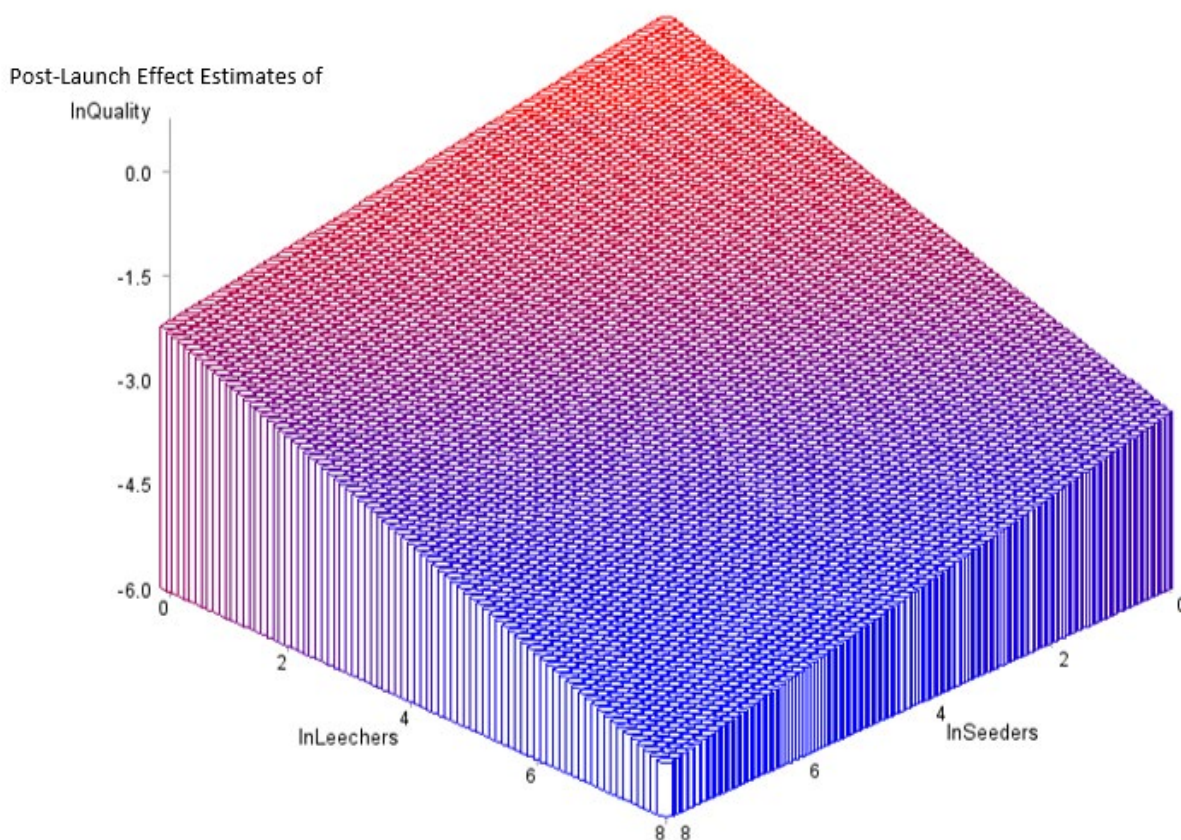
As an additional consideration, is whether piracy quality evolves over time. The theoretical belief is that higher quality versions should evolve over time, suggesting piracy quality increases monotonically. A unit root test of piracy quality with an intercept approaches stationarity ($\rho = 0.813$, $\tau = -2.805$, $p < .06$). Neither the inclusion of one or two lag periods was significant, precluding the need for an Augmented Dickey-Fuller (ADF) test. As such, piracy quality, on average, is relatively stable during a movie's theatrical run.

Managerial Implications Simulation

To highlight the managerial impact of piracy quality on revenues, we use the launch and post-launch estimates to explore how changes of +/-1 standard deviation (SD) from average in seeders, leechers, and piracy quality might alter expected revenues. While we found that higher quality piracy functions more like sampling in the launch period and cannibalism in the post-launch period, piracy quality needs to be interpreted in conjunction with consumer activity for seeders and leechers.

To demonstrate effect changes, we created a 3x3x3 grid for lower, average, or higher levels of quality, seeders, and leechers. A total of 30,000 simulated draws were randomly assigned across the twenty-seven cells. We use the results from Tables 5 and 6 to calculate the revenue change from one of the cell conditions. Then, we subtract the revenues when the draw was set at average levels. A draw in the middle of the table, or 'average' for all three piracy measures would have no revenue change, or 0%. Table 7 displays average percent change in revenue relative to average levels for both launch and post-launch. For instance, one specific draw might be piracy quality remained average, seeders increased +1 SD, and leechers decreased -1 SD (sixth row, first column), which shows a 1.081% increase in launch period revenues relative to when all three piracy measures are average.

Table 7. Simulated Elasticity Differentials of Piracy Quality Changes on Revenue

| | | Launch Period | | | | Post-Launch Period | | |
| | | Leechers | | | | Leechers | | |
Quality	Seeders	-1 SD	Average	+1 SD		-1 SD	Average	+1 SD
-1 SD	-1 SD	-0.338	-1.775	-3.197		0.852	1.541	2.328
-1 SD	Average	1.528	0.067	-1.462		1.231	1.966	2.411
-1 SD	+1 SD	3.483	2.151	0.398		1.508	2.131	3.026
Average	-1 SD	-0.383	-0.642	-0.691		-0.180	-0.007	0.284
Average	Average	0.249	0.000	-0.134		-0.211	0.000	0.419
Average	+1 SD	1.081	0.748	0.458		-0.276	-0.063	0.046
+1 SD	-1 SD	-0.441	0.574	1.674		-1.354	-1.552	-2.074
+1 SD	Average	-0.897	-0.058	1.144		-1.479	-2.032	-2.180
+1 SD	+1 SD	-1.524	-0.462	0.337		-1.760	-2.358	-2.478

Note: +1 (-1) SD refers to one standard deviation above (below) mean for all films in sample.

A few items stand out from the launch period simulation. First, brand managers want to avoid the upper right cell (lower levels of seeders and quality but higher leechers), where expected revenues decline -3.197%. The best scenario for managers is when quality and downloading are low given ample piracy supply (3.483%). Our theory and results point to benefits of higher quality downloads by providing consumers with more information about the legal good. This helps explain why more high-quality downloads exhibit positive revenues changes (lower right cells) while more low-quality downloads negatively affect revenues (upper right cells).

In the post-launch period, expected revenue changes follow a clearer pattern where higher quality piracy generally hurts revenues (lower rows). We postulated that higher quality piracy acts more like a substitute after the product has entered the market, as other information in the marketplace can inform consumers. Consumers seeking higher quality piracy here are likely not information seeking, but looking for the product without paying for it. Hence, the bottom rows, which feature higher quality piracy, exhibit the most negative changes on revenue. In

particular, when all three of piracy quality, seeders, and leechers are higher, the expected decline in revenues is -2.478%.

CONCLUSION

Piracy represents a considerable threat to revenues of both producers and distributors (in the case of movies, studios and theaters, respectively). Extant piracy research has found mixed findings for whether piracy encourages sampling versus cannibalization. This study examined the role of piracy quality and its effect on the market. In particular, we theorize that higher quality copies can both hurt and help sales. We contribute to the piracy literature by theorizing that when product uncertainty is high, namely during the launch period, enthusiastic consumers will search out more information to reduce this uncertainty. By better approximating the genuine good, higher quality copies lower product uncertainty, which better aligns consumer expectations with purchasing. Yet, product uncertainty is lower post-launch as information spreads in the market. As the most enthusiastic consumers have likely purchased, the less interested customers remain, but for these customers higher quality piracy merely cannibalize sales.

This research also makes several substantive contributions. To address the subjective nature of piracy quality, an item response model uncovered latent estimates of quality using keyword signals as attributes of the pirated copies. The impact of piracy quality was estimated using panel data on motion picture supply and demand (screens and revenues, respectively), in conjunction with observed illegal supply and demand (seeders and leechers, respectively). As the legal and illegal sides of the market are interdependent, the model uses seemingly unrelated regression with copulas to address endogeneity. Certain piracy keywords signal higher quality

(e.g., XVID and AC3) while others indicate lower quality copies (e.g., TC and DVDR). We find that a 1% increase in downloads of higher quality piracy corresponds to a 0.52% increase in revenues in the launch period. Yet, the post-launch period shows a 1% increase in downloads of higher quality piracy yields a -0.38% decline in revenues. These differential effects in timing, in addition to the role of quality, help alleviate prior research tensions as to whether piracy acts as a sampling mechanism or substitution. A managerial simulation of the findings highlights the interplay between piracy quality and related consumer activity (downloads and uploads).

The findings point to two key managerial implications, especially because enforcement resources are limited even in the most resourceful nations (Fink et. al. 2016). First, studios could afford to be less stringent on higher quality piracy in the opening week. Second, since piracy derives from the original product, the genuine good represents a powerful tool for managers. By owning the film, studios can release their own sampling variations. Since high quality piracy helps during launch, managers can use this to their advantage by providing more information to consumers with some degree of high quality (but not full) versions. Post-launch, managers can reduce the available piracy quality by releasing more low quality (and still not full) versions. Studio enforcement efforts could focus on the higher quality copies to turn consumers to theaters, the only channel with a guaranteed full version of the film.

Along with the contributions are some limitations. First, while we use data collected from the leading piracy network, we can only speak to the data on this particular website. Second, while we observe piracy quantity and quality online, piracy can still exist in physical forms (i.e., an illegal copy burned to a DVD). We consider our estimates as a lower bound for the total piracy effect. Although observing online activity has advantages over observing it in the physical world, we cannot account for the degree of piracy (and its quality) in this physical form;

physical copies may even follow different quality distributions as well. Our approach could be extended to that realm once data becomes available. Third, although we focus on information goods (specifically motion pictures), illegal versions in other product categories might exhibit different consumption patterns, consumer responses, and efforts by illegal suppliers. This last point in particular serves as an inherent data limitation, but presents a potential direction for future research avenues. As such, this study serves as a stepping stone in the broader literature in piracy by assessing illegal quality as part of the new agenda for the economics of digitization.

REFERENCES

Ackermann, Klaus, Wendy Bradley, and Jack Francis Cameron (2020), "The Pirate Bay & Box Office Buccaneers," working paper, Monash University.

Aguiar, Luis and Joel Waldfogel (2018), "As Streaming Reaches Flood Stage, Does It Stimulate or Depress Music Sales?" *International Journal of Industrial Organization*, 57, 278-307.

Bae, Sang Hoo and Jay Pil Choi (2006), "A Model of Piracy," *Information Economics and Policy*, 18 (3), 303-320.

Bagozzi, Richard P. and Utpal M. Dholakia (2006), "Antecedents and Purchase Consequences of Customer Participation in Small Group Brand Communities," *International Journal of Research in Marketing*, 23 (1), 45-61.

Bartholomew, David J., Fiona Steel, Irini Moustaki, Jane I. Galbraith (2002), *The Analysis and Interpretation of Multivariate Data for Social Scientists.* Boca Raton, FL: Chapman & Hall/CRC.

Belleflamme, Paul and Martin Peitz (2014), "Digital Piracy: An Update," CORE Discussion Paper (2014/19), Université Catholique de Louvain.

Bhattacharjee, Sudip, Ram D. Gopal, and G. Lawrence Sanders (2003), "Digital Music and Online Sharing: Software Piracy 2.0?" *Communications of the ACM*, 46 (7), 107-111.

Chellappa, Ramnath K. and Shivendu Shivendu (2005), "Managing Piracy: Pricing and Sampling Strategies for Digital Experience Goods in Vertically Segmented Markets," *Information Systems Research*, 16 (4), 400-417.

Clement, Michel, Steven Wu, and Marc Fischer (2014), "Empirical Generalizations of Demand and Supply Dynamics for Movies," *International Journal of Research in Marketing*, 31 (2), 207-223.

Danaher, Brett, Samita Dhanasobhon, Michael D. Smith, and Rahul Telang (2010), "Converting Pirates Without Cannibalizing Purchasers: The Impact of Digital Distribution on Physical Sales and Internet Piracy," *Marketing Science*, 29 (6), 1138-1151.

Datta, Hannes, Bram Foubert, and Harald J. Van Heerde (2015), "The Challenge of Retaining Customers Acquired with Free Trials," *Journal of Marketing Research*, 52 (2), 217-234.

Dey, Debabrata, Antino Kim, and Atanu Lahiri (2019), "Online Piracy and the "Longer Arm" of Enforcement," *Management Science*, 65 (3), 1173-1190.

Dhar, Tirtha and Charles B. Weinberg (2016), "Measurement of Interactions in Non-Linear Marketing Models: The Effect of Critics' Ratings and Consumer Sentiment on Movie Demand," *International Journal of Research in Marketing*, 33 (2), 392-408.

Elberse, Anita and Jehoshua Eliashberg (2003), "Demand and Supply Dynamics for Sequentially Released Products in International Markets: The Case of Motion Pictures," *Marketing Science*, 22 (3), 329-354.

Fader, Peter (2000), "Expert report: A&M Records, Inc. v. Napster, Inc."

Fink, Carsten, Keith Maskus, and Yi Qian (2016), "The Economic Effects of Counterfeiting and Piracy: A Review and Implications for Developing Countries," *The World Bank Research Observer*, 31 (1), 1-28.

Geng, Xianjun and Young-Jin Lee (2013), "Competing with Piracy: A Multichannel Sequential Search Approach," *Journal of Management Information Systems*, 30 (2), 159-184.

Han, Young Jee, Joseph C. Nunes, and Xavier Drèze (2010), "Signaling Status with Luxury Goods: The Role of Brand Prominence," *Journal of Marketing*, 74 (4), 15-30.

Hennig-Thurau, Thorsten, Mark B. Houston, and Shrihari Sridhar (2006), "Can Good Marketing Carry a Bad Product? Evidence from the Motion Picture Industry," *Marketing Letters*, 17 (3), 205-219.

Holbrook, Morris B. and Kim P. Corfman (1985), "Quality and Value in the Consumption Experience: Phaedrus Rides Again," in *Perceived Quality*. Lexington, MA: Lexington Books, 31-57.

Hong, Seung-Hyun (2013), "Measuring the Effect of Napster on Recorded Music Sales: Difference-in-Differences Estimates under Compositional Changes," *Journal of Applied Econometrics*, 28 (2), 297-324.

Hui, Kai-Lung and Ivan Png (2003), "Piracy and the Legitimate Demand for Recorded Music," *The BE Journal of Economic Analysis & Policy*, 2 (1), 1-22.

IBC (2017), "Cost of Online Piracy to Hit $52bn," https://www.ibc.org/publish/cost-of-online-piracy-to-hit-52bn/2509.article.

Jain, Sanjay (2008), "Digital Piracy: A Competitive Analysis," *Marketing Science*, 27 (4), 610-626.

Jackman, Simon (2009). *Bayesian Analysis for the Social Sciences*. Hoboken, NJ: Wiley.

Kaiser, Henry F. (1974), "An Index of Factorial Simplicity," *Psychometrika*, 39 (1), 31-36.

Koschmann, Anthony and Douglas Bowman (2017), "Simultaneous Estimation of Legal and Illegal Supply and Demand: The Case of Motion Pictures," *International Economic Journal*, 31 (4), 555-577.

Krider, Robert E., Tieshan Li, Yong Liu, and Charles B. Weinberg (2005), "The Lead-Lag Puzzle of Demand and Distribution: A Graphical Method Applied to Movies," *Marketing Science*, 24 (4), 635-645.

Kubinger, Klaus D. (2003), "On Artificial Results Due to Using Factor Analysis for Dichotomous Variables," *Psychology Science*, 45 (1), 106-110.

Lahiri, Atanu and Debabrata Dey (2013), "Effects of Piracy on Quality of Information Goods," *Management Science*, 59 (1), 245-264.

Liebowitz, Stan J. (2008), "Testing File Sharing's Impact on Music Album Sales in Cities," *Management Science*, 54 (4), 852–859.

Li, Xinxin and Lorin M. Hitt (2008), "Self-Selection and Information Role of Online Product Reviews," *Information Systems Research*, 19 (4), 456-474.

Li, Xiaolin, Chenxi Liao, and Ying Xie (2019), "Digital Piracy, Extrinsic Incentives, and Writer Efforts," working paper, University of Texas at Dallas.

Lord, Frederic M. (2012), *Applications of Item Response Theory to Practical Testing Problems*. New York: Routledge.

Lu, Shijie, Xin Wang, and Neil Bendle (2020), "Does Piracy Create Online Word of Mouth? An Empirical Analysis in the Movie Industry," *Management Science*, 66 (5), 2140-2162.

Luo, Hong and Julie Holland Mortimer (2019), "Infringing Use as a Path to Legal Consumption: Evidence from a Field Experiment," working paper 971, Boston College.

Ma, Liye, Alan L. Montgomery, Param Vir Singh, and Michael D. Smith (2014), "An Empirical Analysis of the Impact of Pre-Release Movie Piracy on Box Office Revenue," *Information Systems Research*, 25 (3), 590-603.

Machado, Fernando S., T. S. Raghu, Preethika Sainam, and Rajiv Sinha (2017), "Software Piracy in the Presence of Open Source Alternatives," *Journal of the Association for Information Systems*, 18 (1), 1-21.

Monroe, Kent B. and R. Krishnan (1985), "The Effect of Price on Subjective Product Evaluations," in *Perceived Quality*. Lexington, MA: Lexington Books, 209-232.

Moorthy, Sridhar, Brian T. Ratchford, and Debabrata Talukdar (1997), "Consumer Information Search Revisited: Theory and Empirical Analysis," *Journal of Consumer Research*, 23 (4), 263-277.

Mortimer, Julie Holland, Chris Nosko, and Alan Sorensen (2012), "Supply Responses to Digital Distribution: Recorded Music and Live Performances," *Information Economics and Policy*, 24 (1), 3-14.

Moul, Charles C. (2005), *A Concise Handbook of Movie Industry Economics*. Cambridge: University Press.

Nelson, Phillip (1970), "Information and Consumer Behavior," *Journal of Political Economy*, 78 (2), 311-329.

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

Olshavsky, Richard W. (1985), "Perceived Quality in Consumer Decision Making: An Integrated Theoretical Perspective," in *Perceived Quality*. Lexington, MA: Lexington Books, 3-29.

Packard, Grant, Anocha Aribarg, Jehoshua Eliashberg, and Natasha Z. Foutz (2016), "The Role of Network Embeddedness in Film Success," *International Journal of Research in Marketing*, 33 (2), 328-342.

Papies, Dominik, Peter Ebbes, and Harald J. Van Heerde (2017), "Addressing Endogeneity in Marketing Models," in *Advanced Methods for Modeling Markets*. Berlin: Springer, 581-627.

Park, Sungho and Sachin Gupta (2012), "Handling Endogenous Regressors by Joint Estimation Using Copulas," *Marketing Science*, 31 (4), 567-586.

Peitz, Martin and Patrick Waelbroeck (2006), "Piracy of Digital Products: A Critical Review of the Theoretical Literature," *Information Economics and Policy*, 18 (4), 449-476.

Qian, Yi (2014), "Brand Management and Strategies Against Counterfeits. *Journal of Economics & Management Strategy*, 23 (2), 317–343.

Qian, Yi (2015), "Counterfeiters: Foes or Friends? How Counterfeits Affect Sales by Product Quality Tier," *Management Science*, 60 (10), 2381-2400.

Reid, R. Dan, and Nada R. Sanders (2001), *Operations Management: An Integrated Approach*. Hoboken, NJ: John Wiley & Sons.

Reeves, Carol A. and David A. Bednar (1994), "Defining Quality: Alternatives and Implications," *Academy of Management Review*, 19 (3), 419-445.

Schweidel, David A. and George Knox (2013), "Incorporating Direct Marketing Activity into Latent Attrition Models," *Marketing Science*, 32 (3), 471-487.

Spangler, Todd (2014), "'Expendables 3' Illegally Downloaded 5 Million Times, but Still Isn't Top Hit for Pirates," *Variety* (August 18), https://variety.com/2014/digital/news/expendables-3-illegally-downloaded-5-million-times-but-still-isnt-top-hit-for-pirates-1201285179/

Spector, Paul E., Paul T. Van Katwyk, Michael T. Brannick, and Peter Y. Chen (1997), "When Two Factors Don't Reflect Two Constructs: How Item Characteristics Can Produce Artifactual Factors," *Journal of Management*, 23 (5), 659-677.

Sundararajan, Arun (2004), "Managing Digital Piracy: Pricing and Protection," *Information Systems Research*, 15 (3), 287-308.

Swaminathan, Srinivasan and Kapil Bawa (2005), "Category-Specific Coupon Proneness: The Impact of Individual Characteristics and Category-Specific Variables," *Journal of Retailing*, 81 (3), 205-214.

Van Schuur, Wijbrandt H. and Henk AL Kiers (1994), "Why Factor Analysis Often Is The Incorrect Model For Analyzing Bipolar Concepts, And What Model To Use Instead," *Applied Psychological Measurement*, 18 (2), 97-110.

Vogel, Harold L. (2015), *Entertainment Industry Economics*. New York: Cambridge University Press.

Waldfogel, Joel (2012), "Digital Piracy: Empirics," in *The Oxford Handbook of the Digital Economy*. Oxford: Oxford University Press, 512-546.

Wilcox, Keith, Hyeong Min Kim, and Sankar Sen (2009), "Why Do Consumers Buy Counterfeit Luxury Brands?" *Journal of Marketing Research*, 46 (2), 247-259.

Yoon, Kiho (2007), "On the Impact of Digital Music Distribution," working paper, Hitotsubashi University.

You, Ya, Gautham G. Vadakkepatt, and Amit M. Joshi (2015), "A Meta-Analysis of Electronic Word-of-Mouth Elasticity," *Journal of Marketing*, 79 (2), 19-39.

Yue, Yang (2019), "How Does Online Piracy Affect Film Revenue in China?" working paper, Xiamen University.

Zeithaml, Valarie A. (1988), "Consumer Perceptions of Price, Quality, and Value: A Means-End Model and Synthesis of Evidence," *Journal of Marketing*, 52 (3), 2-22.

Zellner Arnold and H. Theil (1962), "Three-Stage Least Squares: Simultaneous Estimation of Simultaneous Equations," *Econometrica*, 30 (1), 54–78.

FOOTNOTES

1. We clarify that copying an electronic file and not altering it will result in an exact replication. Copying or replicating from an analog form to either an electronic or another analog form will result in some loss of quality.

2. A factor analysis was conducted to ideally reduce the keywords to several underlying factors. However, the 34 keywords combine to 14 factors (with eigenvalue > 1), cumulatively explaining 62.03% of the variance. A KMO test of the correlations (*IFS* < 0.50) indicates the data is not suitable for factor analysis (Kaiser 1974). Together, this suggests factor analysis is inappropriate for uncovering latent quality from the piracy keywords here.

3. A file that is of seemingly high quality yet smaller in file size might trigger suspicion among users. However, high quality video might come with reduced frames, lower quality audio, or newer file compression techniques that could shrink file size. There is some 'honor among thieves' in that pirates may be seeking social capital, yet the piracy files allow for user comments; files with mislabeled keywords can be pointed out quickly by other users.

4. Discussions with executives of a major theater chain indicated that piracy supply prior to a new film's release was not material in its screen allocation decision, since little piracy was anticipated in the market prior to release.

<u>**Web Appendix A**. Smoothing Forecast for Estimated Revenues</u>

Movie theater managers anticipate demand (revenues) for film i in week t, and adjust screens accordingly. Opening week (launch period) screen allocation is a function of the HSX closing price prior to release. After observing demand, theater managers update revenue estimates for week 2 onwards. Two time periods are required for parameter smoothing; week 2 is estimated by averaging the opening week actual and estimated revenues then multiply by .70 (presuming industry revenue declines 30%). Single exponential smoothing estimates this in equation A1, and prior week expected revenues are updated by part of the prediction error (Elberse and Eliashberg 2003):

$$\text{Revenue\_Est*}_{it} = \text{Revenue\_Est*}_{i,t\text{-}1} + \lambda_{i,t}(\text{Revenue}_{i,t\text{-}1} - \text{Revenue\_Est*}_{i,t\text{-}1}) \quad \text{for t>2} \qquad \text{(A1)}$$

Revenue_Est*$_{it}$ is the expected revenues from simple smoothing and λ is the smoothing parameter (between 0 and 1). As revenues decline over time, a double exponential smoothing procedure is applied with a trend, T_{it}, and a second smoothing parameter, π_{it} in Equation A2. Weeks 3 and on use the minimized sum of squared differences between actual and expected revenues to update the smoothing parameters. Note, $T_{i1} = 0$, since no trend has formed yet.

$$T_{it} = \pi_{i,t}(\text{Revenue\_Est*}_{it} - \text{Revenue\_Est*}_{i,t\text{-}1}) + (1 - \pi_{i,t})T_{i,t\text{-}1} \qquad \text{for t>2} \qquad \text{(A2)}$$

$$\text{Revenue\_Est}_{it} = \text{Revenue\_Est*}_{it} + T_{it}(1 - \pi_{i,t})/\pi_{i,t} \qquad \text{for t>2} \qquad \text{(A3)}$$

The *Revenue_Est$_{it}$* used in the launch and post-launch models come from the double smoothing process in Equation A3.

Web Appendix B. Source Types for Piracy

Type	Quality	Common Pirate Signals	Description
Workprints	low	"WP", "WORKPRINT"	The "dailies" (rough-cut production from the studio lot, without editing or effects) that happen to get out. Rare. Often need color correction and audio mixing to resemble the finished product.
Camcorder	low	"CAMRip", "CAM"	Audio and video captured in the theater from a camcorder or mobile phone.
Pay-Per-View	low-medium	"PPV", "PPVRip"	Viewings in hotels, usually through a camcorder
Telesync	low-medium	"TS", "TELESYNC", "PDVD"	Camcorder footage (often done in an empty theater) but direct audio input from the film track, or synced with the film audio track.
Telecine	medium	"TC", "TELECINE"	Machine conversion of the film reel to a digital form; not as good as DVDs due to jittering of the reel in process and color quality.
Screener	low-high	"SCR", "SCREENER", "DVDSCR", "DVDSCREENER", "BDSCR"	Advance copies sent to movie critics, MPAA members, executives, or studio business affiliates (such as advertising agencies or post-production houses). Not full DVDs, as some scenes may be missing or film mastering not complete. A digital version only meant for download/FTP can be labeled "DDC".
R5	medium-high	"R5" and variations such as "R5.LINE" and "R5.AC3.5.1.HQ"	Denotes the regional DVD coding: region 5 for India, Africa, Russia, North Korea, and Mongolia. Often not a DVD copy, but a very good Telecine transfer. If the original audio is non-English, English audio is synced and "LiNE" is used in the file description. Split audio tracks enable multiple channels, such as Dolby 5.1 surround sound capability.
DVD	high	"DVDRip", "DVDR", "DVD-Full", "Full-Rip", "ISO rip", "lossless rip", "untouched rip", "DVD-5" or "DVD-9"	The film copy from a DVD. Full copies of DVDs, including extras, bonus scenes and the like may be listed as regional encoding (i.e. "DVD-5" for region 5). File sizes usually range from 4 to 8GB.
HDTV	high	"DSR", "DSRip", "DTHRip", "DVBRip", "HDTV", "PDTV", "TVRip", "HDTVRip"	Captured from satellite or television broadcasts, often through the digital receiver and additional equipment and not a camcorder. Quality can be better than DVD. Also, Video on Demand copying ("VODRip", "VODR").
Blu-Ray	high	"BDRip", "BRRip", "Blu-Ray", "BluRay", "BLURAY", "BDR", "BD5", "BD9"	Blu-Ray format discs, the highest quality available in both picture and sound, with disc space for more extras than DVDs. File sizes can range from 8 to 60GB, but smaller sizes exist as compressed files with reduced resolution. These are similarly coded for region like DVDs.

Source: https://pirates-forum.org/Thread-Movie-Sources-Movie-Formats

Web Appendix C. List of Piracy Keywords/Signals in the Illegal Copy Files

Keyword	Observations	In IRT	Keyword	Observations	In IRT
MOVIES	203,714	Yes	7.1	254	no
3D	545	Yes	265	69	no
HHELD	465	Yes	1080i	30	no
HDMovies	40,165	Yes	4K	48	no
DVDR	4,551	Yes	dolby	68	no
2.0	1,694	Yes	FLAC	136	no
5.1	5,046	Yes	FLV	169	no
264	92,997	Yes	HDTV	295	no
1080P	11,525	Yes	HEVC	12	no
720P	36,944	Yes	mpeg	36	no
AAC	30,835	Yes	mpeg4	0	no
AC3	65,569	Yes	PPV	56	no
AVI	14,050	Yes	print	2	no
BD	5,665	Yes	telecine	0	no
BR	32,436	Yes	telesync	158	no
CAM	87,846	Yes	VC1	0	no
DIVX	391	Yes	VC-1	0	no
DTS	11,766	Yes	vcd	0	no
DVD	54,939	Yes	VP9	0	no
HQ	8,674	Yes	wmv	0	no
LINE	1,564	Yes	work	70	no
MKV	3,617	Yes	WP	0	no
MP3	14,457	Yes			
MP4	9,851	Yes			
R5	1,533	Yes			
R6	4,472	Yes			
RIP	66,821	Yes			
SCR	46,745	Yes			
SUB	21,747	Yes			
TC	9,125	Yes			
TS	36,626	Yes			
V2	5,691	Yes			
V3	453	Yes			
XVID	102,960	Yes			

*Note*s. Italicized items are file types, as a separate search parameter in Pirate Bay. All items are prefixed with 'F' for file or 'FT' for file type in Figure 2.

Web Appendix D. Tetrachoric Correlations of Piracy Keyword Signals

	FT_MOVIES	FT_3D	FT_HHELD	FT_HDMovies	FT_DVDR	F_2.0	F_5.1	F_AAC	F_AC3	F_DTS	F_264	F_MKV	F_MP3	F_MP4	F_XVID	F_SUB	F_720P	F_1080P	F_CAM	F_TC	F_TS	F_SCR	F_DVD	F_BR	F_DIVX	F_AVI	F_HQ	F_V2	F_V3	F_R5	F_R6	F_RIP	F_LINE	F_BD
FT_MOVIES	1.00	-0.98	-1.00	-1.00	-0.99	-0.54	-0.66	-0.23	-0.05	-0.55	-0.58	-0.29	0.58	-0.16	0.83	-0.37	-0.78	-0.89	0.60	-0.02	0.29	0.22	0.59	-0.66	0.34	0.35	-0.05	0.12	0.07	0.22	-0.04	0.29	-0.13	-0.27
FT_3D		1.00	-0.95	-0.93	-0.81	-0.80	0.03	-0.17	-0.30	0.46	0.26	-0.84	-0.97	-1.00	-0.94	0.02	-0.31	0.60	-0.95	-0.77	0.14	-0.85	-0.93	0.33	-0.88	-0.80	0.07	-0.99	-0.77	-0.79	-0.82	-0.29	-0.81	0.16
FT_HHELD			1.00	-0.95	-0.85	-0.78	-0.82	-0.97	0.16	-0.05	0.30	-1.00	0.53	-0.93	0.06	-0.09	-0.85	0.06	-0.83	0.14	0.06	-0.15	-0.01	-0.98	-0.78	-0.82	0.35	-0.83	-0.83	-0.83	-0.26	-0.88	-0.79	-0.79
FT_HDMovies				1.00	-0.94	-0.81	-0.54	-0.82	-0.97	0.53	0.63	0.28	-0.56	0.17	-0.80	0.14	0.82	0.86	-0.64	0.04	-0.34	-0.30	-0.76	0.68	-0.36	-0.32	0.05	-0.16	-0.05	-0.91	0.07	-0.22	0.11	0.28
FT_DVDR					1.00	0.67	0.50	0.27	0.10	0.53	0.10	0.28	-0.91	-0.13	-0.68	0.77	-0.32	-0.34	-0.18	-0.76	0.04	0.21	0.54	-0.03	-0.10	-1.00	0.01	0.04	-0.76	0.23	-0.91	0.23	-1.00	-0.17
F_2.0						1.00	0.01	0.19	0.06	0.03	0.07	0.08	-0.92	-1.00	-0.33	0.49	0.08	0.27	-0.22	-0.74	0.14	0.24	0.03	0.10	-0.84	-1.00	-0.80	0.25	-0.74	-0.77	-0.80	-0.23	0.25	-1.00
F_5.1							1.00	-0.13	0.05	0.44	0.22	0.17	-0.88	-0.90	-0.54	0.47	-0.04	0.67	-0.47	-0.19	0.01	-0.23	0.36	-0.83	-0.12	0.10	-0.85	-1.00	0.11	-1.00	-0.33	-1.00	0.13	0.13
F_AAC								1.00	-0.97	-0.06	0.44	0.05	0.03	0.10	0.23	0.21	0.03	0.13	0.21	-0.76	-0.11	-0.14	0.13	-0.15	-1.00	0.32	0.01	0.22	-0.83	-0.01	0.13	0.58	0.22	-0.10
F_AC3									1.00	-0.12	-0.16	0.29	-0.16	-1.00	-0.02	-0.03	0.30	-0.20	-0.02	-0.03	0.02	0.18	0.18	0.35	-0.88	-0.28	0.01	0.03	0.11	0.03	0.11	0.05	-0.03	0.12
F_DTS										1.00	0.42	0.09	0.04	0.11	-0.52	0.09	0.22	0.60	-1.00	-0.03	-0.98	-0.49	-0.64	0.34	-0.96	-0.22	0.11	0.19	0.10	0.10	-0.87	-0.07	0.36	0.04
F_264											1.00	-0.05	-0.30	0.47	-0.96	-0.02	0.44	0.50	0.14	0.10	0.10	-0.09	0.13	0.44	-0.44	-0.18	-0.06	0.25	0.10	0.01	0.21	-0.33	0.04	0.23
F_MKV												1.00	-0.21	-1.00	-0.72	0.05	0.17	0.19	-0.13	0.25	0.04	0.24	-0.21	-0.04	-0.88	-0.14	0.10	-0.03	0.25	0.07	0.28	-0.31	0.28	0.33
F_MP3													1.00	-0.88	0.44	-0.13	-0.26	-0.95	0.31	0.42	0.30	0.09	-0.38	-0.08	0.01	0.35	-0.05	0.00	0.25	0.05	0.18	-0.24	0.29	-0.89
F_MP4														1.00	0.10	0.30	0.04	0.35	0.15	0.17	0.06	-0.40	-0.11	-1.00	-0.91	0.26	0.51	0.51	-0.98	-0.21	-0.30	0.10	-0.92	-0.92
F_XVID															1.00	-0.42	-0.57	-0.85	-0.15	-0.11	-0.10	-0.17	0.71	-0.45	0.28	-0.26	-0.24	-0.30	-0.05	-0.27	0.60	-0.14	-0.34	-0.34
F_SUB																1.00	0.02	0.10	0.13	0.10	0.10	0.21	-0.17	0.13	-0.23	0.31	0.29	-0.05	0.04	0.06	-0.08	0.18	0.07	0.07
F_720P																	1.00	-0.60	-0.34	0.13	-0.22	-0.26	-0.73	0.63	-0.34	-0.37	-0.02	0.06	0.13	-0.33	0.09	-0.08	0.15	0.11
F_1080P																		1.00	-0.50	-0.47	-0.47	-0.55	-0.81	0.58	-0.89	-1.00	0.02	-0.16	-0.81	-0.84	-0.20	-0.37	-0.03	0.26
F_CAM																			1.00	-0.14	-0.34	-0.69	-0.81	-0.61	-0.38	0.20	-0.21	0.56	0.50	0.25	0.21	-0.68	-0.93	-0.05
F_TC																				1.00	-0.83	-1.00	-1.00	-0.88	-0.81	-1.00	-0.21	-0.76	-0.89	-0.71	-0.74	-0.77	-1.00	-0.75
F_TS																					1.00	-0.34	-0.52	-0.56	-0.93	0.41	0.51	-0.11	-0.83	-0.86	-0.28	-0.60	0.42	-0.06
F_SCR																						1.00	0.72	-0.62	-0.88	0.05	0.35	-0.14	0.13	-1.00	-0.78	0.00	-0.19	-0.19
F_DVD																							1.00	-0.83	-0.41	0.35	-0.39	-0.23	0.01	-0.30	-0.20	0.58	-0.36	-0.54
F_BR																								1.00	-0.40	-0.15	-0.32	-0.88	-0.32	-0.30	-0.16	0.13	-0.09	-0.25
F_DIVX																									1.00	0.24	0.30	-1.00	-0.81	-0.83	-0.09	0.32	-0.85	-0.84
F_AVI																										1.00	0.41	0.01	0.32	0.01	-0.01	0.13	0.22	-0.10
F_HQ																											1.00	0.11	0.22	-0.80	-0.83	-0.47	-0.80	0.08
F_V2																												1.00	-0.89	-0.89	-0.05	-0.83	0.14	-0.08
F_V3																													1.00	-0.74	-0.77	-1.00	-0.75	-1.00
F_R5																														1.00	-0.80	-0.15	0.24	-1.00
F_R6																															1.00	-0.36	-0.81	-1.00
F_RIP																																1.00	-0.26	0.31
F_LINE																																	1.00	0.15
F_BD																																		1.00