Assessing Piracy Quality and its Effect on the Legal Market for Information Goods

Anthony Koschmann, Eastern Michigan University
Yi Qian, University of British Columbia

Abstract

A common perception about piracy, or illegal copies of information goods such as movies and books, is that it acts as a substitute for the legitimate product, reducing revenues for the producer. Previous research into piracy has shown mixed effects, with piracy acting as both a substitute and complement. These findings have been explained by moderating factors such as release timing, network effects, and quality of the legal offering. However, a relatively unexamined aspect to date is the role of piracy quality, and whether lower (or higher) quality copies exert differential effects on legal sales. Using motion picture and online piracy data, we estimate piracy quality using a latent item response theory (IRT) model based on keyword signals from illegal copies. We apply an interdependent system of simultaneous equations models to jointly estimate 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 (i.e., sampling). However, in the post-launch period, higher quality illegal copies exhibit a negative effect on revenues (i.e., substitution). Additionally, we find empirical evidence that piracy quality is endogenous only in the post-launch period, which we account for using instrumental variables. The findings suggest that producers can nudge the market by alleviating product uncertainty through higher quality samples at product launch while securing enforcement post-launch.

Keywords piracy quality; legal sales; launch and post-launch; motion pictures

* The authors would like to thank Fred Feinberg, Shane Greenstein, Avi Goldfarb, Xixi Hu, Ginger Jin, Josh Lerner, Xiaolin Li, Julie Mortimer, Kevin Ryan, Marie Thursby, Hui Xie, Joel Waldfolgel, Zining Wang, and Chuck Weinberg for helpful comments.

1. Introduction

Illegal copies are a key issue facing creators and producers, affecting many types of products. Like the genuine good, illegal variants convey quality signals to consumers. Counterfeit automobile parts cost manufacturers an estimated $12 billion globally in 2010 (Kimbrough 2014), although fake parts are more readily detected through flimsy packaging, discolored or mislabeled pieces, lack of branding, and substandard assembling. The World Health Organization estimates counterfeit pharmaceuticals cost manufacturers $75 billion globally in 2010 (WHO 2010); notably, counterfeit pharmaceuticals often have less (or even none) of the active ingredients (Bassat et al. 2016) and may be substituted with inert or possibly dangerous ingredients. Although legal pharmaceuticals are increasing consumer protection with tamper-proof packaging, authentication labels, and hologram stamping, counterfeiters' increasing sophistication sends signals to more closely resemble the genuine good.

Piracy, or illegal copies of information goods such as movies, music, and books, also encounter this qualitative aspect. 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 industry technologists, 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. Together, these indicate that piracy quality, in relation to the quality of the legal version, present an important facet of the effects of piracy.

Identifying the impact of piracy quality is important to both marketing academics and practitioners. A prevailing belief among producers of information goods is that piracy is largely negative, particularly higher quality piracy. 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 films in other genres had higher downloads but saw higher revenues. While prior piracy research has shown both positive and negative effects of piracy, respectively treated as sampling and cannibalization, a deeper understanding of piracy quality can shed light on not just when but also how positive effects might occur.

Although piracy poses a threat to almost any information good, a dearth of research exists on the quality of illegal copies and its role on the legal market. To address how piracy might generate word-of-mouth, Lu et al. (2019) treat piracy quality as the ratio of piracy files that are high definition relative to camcorder recordings. Other research examined piracy prior to a film's release using subjective user ratings of the video and audio quality of pirated films (Ma et al. 2014). Although helpful in understanding the impact of piracy quality, illegal copies convey more information beyond just its source type (e.g., camcorder or not), and may have timing effects on legal demand as well as legal supply. Indeed, Ma et al. (2014) noted as a future research direction that as illegal copies infiltrate the marketplace, the quality might evolve during both the launch and post-launch periods.

We use motion pictures as the specific context for studying piracy quality. As a focal area for marketing research (e.g., Dhar and Weinberg 2016, Packard et al. 2016), the movie industry is used here for several reasons. First, the effect of piracy and consumption may be clearer in movies than in other information goods (Lu et al. 2019), as few movies are seen by a consumer multiple times in theaters, but illegally downloaded music or software may be consumed repeatedly. Second, while durable products also exhibit illegal quality concerns, social or aspirational elements can stimulate illegal consumption (Wilcox et al. 2009) typically not seen in information goods. Third, data collection on illegal activity is difficult to find; physical products would 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 as an area 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 the similar challenge in legal goods of adequate measurement. That is, piracy quality – like product quality in general – is an important product feature that affects consumer choice, but is often treated subjectively by consumers perceiving certain signals (Zeithaml 1988). Piracy quality is rooted in "any valued attribute of a product" (Chellappa and Shivendu 2005, p.402), yet attribute valuation lends itself to open interpretation arising from heterogeneous tastes. In their examination of piracy prior to a movie's release, online users rated the quality of movie piracy files for audio and video (Ma et al. 2014), which were then averaged into one measure. Although an attempt was made to define piracy quality, this definition is still rooted in the subjectivity of users, many of whom did not actually view the pirated copies. While Lu et al. (2019) used objective file types, additional information that conveys piracy quality such as file size, subtitles, and audio/video descriptors were not incorporated.

To model piracy quality and its effect on both the legal market (theaters) and illegal market (online), we assemble a data set of movies in wide release in the United States and Canada from 2013-2014. Correspondingly, a daily panel of those movies' search results and activity on Pirate Bay is tracked while the films are first-run in theaters. The data includes box office revenues, screen availability, piracy, advertising, and movie characteristics. To estimate piracy quality, we use the visible piracy file keywords to objectively 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. Since many research studies on piracy do not account for illegal supply, we follow prior movie research to use a system of simultaneous equations (e.g., Elberse and Eliashberg 2003, Koschmann and Bowman 2017) to estimate piracy quality effects on both the legal and illegal sides of the market, and instrumental variables to address endogeneity. We model this system of equations for both the launch and post-launch periods separately, as movies (like most information goods) particularly emphasize the launch period, and piracy quality may evolve when the movie has had ample time in the legal market.

We find that a 1% increase in the quality of the pirated copies, conditioning on a level of piracy downloads (leechers, or those users downloading the illegal file on the network), corresponds to a 0.18% increase in revenues in the launch period, on average. As an information good, less is known about a movie upon market introduction, so higher quality copies function more like a sampling mechanism. However, since the quality of illegal copies might improve after release, we find that post-launch a 1%

increase in the quality of illegal copies, conditioning on a level of leechers, associates with a 0.79% 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, to the authors' knowledge, this is the first study to objectively measure the effect of piracy quality, as well as its effect on both the legal and illegal markets for motion pictures. Second, 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. Third, piracy quality was determined to be exogenous in the launch period, but endogenous post-launch. We believe a core driver of this is the lack of legal supply, from which the illegal supply derives from, up to and including launch. Finally, the substantive results suggest producers have a unique and advantageous tool for fighting piracy: the legal good itself, and the ability to encourage the right kinds of sampling. We believe these findings give useful meaning to both managers and policymakers regarding the quality nature of illegal variants, while also extending the piracy literature.

2. Related Literature

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. Among information goods like movies, a tension exists where some research has shown negative effects of illegal copies on legal demand (Hui and Png 2003, Bae and Choi 2006, Yoon 2007, Liebowitz 2008, Hong 2013, Belleflamme and Peitz 2014, Waldfogel 2012), with varying estimates on the sales displacement effect (e.g. Aguiar and Waldfogel 2018, Yang 2019, Li et al. 2019). Yet, other research on piracy has found positive effects of piracy on legal demand (Fader 2000, Jain 2008, Mortimer et al. 2012, Lu et al. 2019). For many of these studies, a common underlying belief is a trade-off in sampling versus cannibalization, with willingness to pay as a common explanation; users weigh the financial cost as well as the social cost of being caught pirating.

While many of these prior piracy studies have also examined moderators of piracy effects on legal revenues, few studies have sought to explore the effect of the quality of the pirated copies. Indeed, a challenge for information goods in the digital age is the ability to generate exact copies of the genuine good from an original source (such as a DVD for movies, CD or legal download for music, or e-book for books). In their research on music piracy, Bhattacharjee et al. (2003) found that 90% of respondents perceived the quality of a CD song converted and compressed to the mp3 format was the same or as good as the original. 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]

Such imperceptible differences are a foundation of production quality, which speaks to 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). Products of high quality should meet both these needs of manufacturers and consumers. Unsurprisingly, 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.

Although these qualities point more to actual product use, consumers might not encounter conformance quality at all without first searching for and buying a product based on quality signals. As Zeithaml (1988) points out, consumers perceive quality from the seller's reputation (such as brand name and advertising), price perceptions (stemming from the actual observable price), and intrinsic attributes. Together these speak to the value, excellence, and ability to exceed expectations (i.e., greater performance for the same price will demonstrate both value and surpassing expectations). However, these qualities – and the implication that greater conformance and excellence indicate higher quality – face heterogeneous

[1] We wish to 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.

tastes. Consumers have differing opinions on what makes one seller more reputable than another, as well as varying thresholds for what constitutes a high-quality attribute from a low-quality attribute.

For information goods, the degree of performance uncertainty (and ultimately quality) looms large as performance is often unknown until the consumer actually consumes the product. Furthermore, many information goods derive value from the experience and are subjectively assessed by consumers. This subjective nature adds complexity in evaluating information goods. Indeed, for products that are complex or uncertain, illegal versions provide information to consumers (Peitz and Waelbroeck 2006). Through the illegal version, consumers absorb information that might better judge the quality of the genuine good. If the legal good is of uncertain quality, its illegal counterpart might provide more certainty. A key assumption of piracy quality has been that, as a derivative product, it is inherently the same or lower quality than that of the legal version (e.g., Lahiri and Dey 2014). At face value, when piracy is of high quality – and more closely approximates the quality of the legal good – this makes it seemingly more substitutable, cannibalizing legal sales. Conversely, when piracy quality is deemed low, this is less likely to be substitutable for the legal good. Although this might be the case, prior research found piracy quality had no significant effect on revenues (Ma et al. 2014, Lu et al. 2019).

Yet, no effect of piracy quality might not entirely be the case. Information goods provide weak signals prior to launch, particularly pertaining to quality (Nelson 1970). Consumers seeking a legal sample for information goods run into the concern that some samples misrepresent the product experience. For instance, movie previews or 'film trailers' are seen as biased by consumers and often reflect only the movie's best scenes (Moul 2005). Likewise, consumers are wary of early product reviews of the legal good (Li and Hitt 2008), either from postings generated by the firm, or competitors seeking to foster distrust. While consumers with a low willingness to pay may be more prone to consuming lower quality piracy, Ma et al. (2014) suggest that at the time of product release, the most enthusiastic consumers (and perhaps more likely to convert into paying customers) may be most interested in low quality piracy as a means of searching out product information. From this view, low quality piracy might have a positive effect on sales, at least for initial product launch. However, higher quality copies should more closely resemble the genuine good, giving consumers more accurate product information. Since illegal copies can work as buzz agents (Qian 2015) to increase word-of-mouth, higher quality copies may generate more buzz among consumers, leading to more interest in the good and potentially more revenue.

An additional complication of piracy quality on sales arises with consumer response over time. For most information goods, general interest and revenues decline over time as consumers have learned about and experienced the product. Over time, the legal offering becomes more diffuse in marketplace distribution. Prior research has shown that while some piracy permeates the market prior to launch, it is usually rather limited (Koschmann and Bowman 2017, Ma et al. 2014). As such, the piracy at the time of product launch is presumably of lower quality, whereas higher quality piracy emerges much later after the product has launched and had more time to permeate the market. After product launch, the consumers who have not experienced the information good are less likely to be its most enthusiastic and willing to pay. If piracy quality increases post-launch, this will reduce the quality gap between the legal and pirated versions. With a smaller difference in quality between the pirated and legal version, and the consumers most willing to pay have exited the market, higher piracy quality should have its largest impact on those less willing to pay. These less-attached consumers are more likely to be satisfied with the pirated version. As such, demand for higher quality copies should exhibit a negative effect on revenues in the post launch period. However, since general demand for piracy and revenues decline over time (e.g., Koschmann and Bowman 2017), it is unclear if higher quality piracy will adversely impact sales in the post-launch period.

3. Methodology

Since observing illegal behavior is difficult, we use a product category where the legal and illegal markets can be observed concurrently: motion pictures. In exploring the impact of the quality of the illegal market, we first describe the data and measures. Then we elaborate on the modeling and estimation procedures.

3.1. 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 motion pictures 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, opening, and release synonymously.

Second, the Hollywood Stock Exchange (HSX: hsx.com) is a prediction market that estimates opening week revenues. Here, 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).

Fifth, advertising costs for each film was gathered 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. This index is a survey of 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.

3.2. Measures

Table 1 elaborates on the variables, descriptions, operationalizations (measures), and data sources. *Ex ante*, the legal supplier decides on 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 declines are curved rather than linear, one parameter smooths and another accounts for the trend, giving more weight to more recent weeks (additional details are described in Elberse and Eliashberg 2003, and Koschmann and Bowman 2017).

*** Insert Table 1 about here ***

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 describe further in Section 3.3. Since piracy can occur before product launch, we account for this as number of days 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 that account for supply and demand 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 is captured online through IMDB ranking the film's popularity in search results *(Online_Buzz),* the user rating of the film as word-of-mouth (*WOM*), and the number of users rating the film (*Num_Users*). These last two reflect the valence and volume, respectively, for consumer word-of-mouth. Several variables also capture competition: competition for screen allocation 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.

3.3. Measuring Piracy Quality through Observed Signals

While Table 1 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 market for illegal goods also sends signals regarding how close the illegal version matches the legal good. 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 et al. 2010). Although these signals can be assessed by experts, a concern is that expert opinions may differ.

With the Pirate Bay data, however, the illegal copies send signals that meaningfully suggest quality to consumers, or online users. Figure 1 presents a screenshot with 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' refers to number of leechers (at that point in time, there were 58 users in the network 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'). As another example, the third search result is a 'Movies DVDR' file type, suggesting it came from a higher quality source. The 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). 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 A1 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).

*** Insert Figure 1 about here ***

The keywords in each pirated copy signal qualitative aspects to the user. While some keywords should suggest greater quality than others (e.g., '1080P' video is higher resolution than '720P' video), other keywords are more subjective in quality (e.g., 'DTS' and 'DD' are competing multichannel audio technologies that both support 5.1 channel surround sound). As such, the 34 keyword signals are dichotomized as being present or not in each piracy file.

At a higher level, quality is a composite of elements consumers perceive, such as price and attributes in products (Zeithaml 1988), as well as expectations relative to performance in services (Parasuraman et al. 1985). While quality might have several underlying factors, our focus is scale construction with ideal point estimates of each piracy keyword on that quality scale. Prior research has shown 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]

Given this, we use an item response theory (IRT) model to estimate each piracy keyword's propensity for quality. IRT models are used to uncover latent measures (Lord 2012), such as 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). Similarly, our focus here is to uncover piracy quality provided a set of piracy keywords.

The model assumes each piracy file i has an unobserved (latent) quality Θ_i, where 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 the 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 model is represented by Equation 1:

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

[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.

For identification, one piracy keyword is marked at one of the far ends of the latent spectrum. By doing so, the quality of all the other piracy keywords should be either to the left or the right of this item. In the coupon redemption study (Swaminathan and Bawa 2005), the model could be identified by setting the coupon with the greatest dollar amount at the far left end of the spectrum (e.g., most consumers, even those who normally do not use coupons, would redeem a $4.00 coupon, but not a $0.25 coupon).

The Pirate Bay data encompasses 8,701 unique piracy files (*Num_Files*) across the films studied. We estimate the quality of the 34 most common piracy keyword signals using a Bayesian Markov Chain Monte Carlo (MCMC) process with 40,000 iterations, 5,000 iteration burn-in, 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 results for the mean quality ideal points and posterior standard deviations are shown in Table 2 (a file count of all piracy keywords initially considered is in Web Appendix A2).

*** Insert Table 2 about here ***

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 34 piracy keywords from lowest to highest mean quality. The estimation results confirm some prior beliefs regarding piracy quality, namely that files purporting to be in 720P 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. Furthermore, some items 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), or copies that come from region 5 DVD sources such as Russia and most of Asia (F_R5: *Mean* = -0.914, *SD* = 0.154). Some quality results were also surprising, such as camcorder sources being 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, while merely suggesting high quality is not enough for consumers. 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 the DVD source exists when it legitimately does not could 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 2. The resulting inferred underlying quality across keywords and pirated copies then enters as a covariate in the simultaneous equations models detailed in the following sections.

*** Insert Figure 2 about here ***

3.4. Launch Model

To estimate the effect of piracy quality on the legal market, we model an interdependent system of simultaneous equations with legal supply and demand plus illegal supply and demand. For many products, and especially motion pictures, the launch period is different from the post-launch period. The system has four equations, where *Y* is a vector of endogenous variables, *X* is a vector of time-varying variables, *Z* is a vector of time-invariant variables, and Greek capital letters for a vector of parameters (A fully-specified model appears in Web Appendix A3):

$$\ln(\text{Screens}_{i1}) = A_0 + A_1\ln(Y_{Si1}) + A_2\ln(X_{Si1}) + A_3\ln(Z_{Si1}) + \varepsilon_{Si1} \tag{2}$$

[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.

$$\ln(\text{Revenue}_{i1}) = B_0 + B_1\ln(Y_{Ri1}) + B_2\ln(X_{Ri1}) + B_3\ln(Z_{Ri1}) + \varepsilon_{Ri1} \quad (3)$$

$$\ln(\text{Seeders}_{i1}) = \Gamma_0 + \Gamma_1\ln(Y_{Pi1}) + \Gamma_2\ln(X_{Pi1}) + \Gamma_3\ln(Z_{Pi1}) + \varepsilon_{Pi1} \quad (4)$$

$$\ln(\text{Leechers}_{i1}) = \Lambda_0 + \Lambda_1\ln(Y_{Li1}) + \Lambda_2\ln(X_{Li1}) + \Lambda_3\ln(Z_{Li1}) + \varepsilon_{Li1} \quad (5)$$

The model uses a multiplicative framework, log transforming variables in accord with prior movie research (e.g., Elberse and Eliashberg 2003; Koschmann and Bowman 2017), allowing the coefficients to be interpreted as elasticities. Subscripts *i* denote the film and *1* for the launch period (week 1). Legal supply (subscript *S*) is treated as the starting point for the system of market components.[4] In the legal supply equation, *X* contains screen competition and *Z* contains production budget, star power, advertising, critic ratings, and an indicator for major studio release. Since release by a major studio is binary coded, it is not log transformed. In the legal demand equation (subscript *R*), *X* is composed of competition for revenue, seasonality, and piracy quality (subscript *Q* in the Results); *Z* contains star power, advertising, and critic reviews. Since seasonality is a percentage relative to the average week, it is not log transformed. Lastly, ε is an error term with multivariate normal distribution

In the illegal supply equation (subscript *P*), *X* contains online buzz and piracy quality while *Z* is similar to legal supply: production budget, star power, advertising, and critic reviews. Also, *Z* includes the prior days the film had been released previously in a major market. Online buzz is not log transformed. In the illegal demand equation (subscript *L*), *X* includes online buzz and piracy quality; *Z* is similar to legal demand by including star power, advertising, and critic reviews. Additionally, *Z* incorporates the prior days of release. Word-of-mouth is excluded in the launch period since consumers are not particularly trusting of early consumer reviews arising from product biases (Li and Hitt 2008), as competitors might manipulate online reviews (Dallarocas 2006).

[4] Discussions with executives of a major chain of film theaters indicated that piracy supply prior to a new film premiering was not considered material in its decision for screen allocations, since little piracy was anticipated in the market prior to a film's release.

3.5. Post-Launch Model

The post-launch system of equations is similar to Equations (2)-(5), where $t > 1$ and Greek lowercase letters denote a vector of parameters:

$$\ln(\text{Screens}_{it}) = \alpha_0 + \alpha_1 \ln(Y_{Sit}) + \alpha_2 \ln(X_{Sit}) + \alpha_3 D_{Sit} + \varepsilon_{Sit} \quad (6)$$

$$\ln(\text{Revenues}_{it}) = \beta_0 + \beta_1 \ln(Y_{Rit}) + \beta_2 \ln(X_{Rit}) + \beta_3 D_{Rit} + \varepsilon_{Rit} \quad (7)$$

$$\ln(\text{Seeders}_{it}) = \gamma_0 + \gamma_1 \ln(Y_{Pit}) + \gamma_2 \ln(X_{Pit}) + \gamma_3 D_{Pit} + \varepsilon_{Pit} \quad (8)$$

$$\ln(\text{Leechers}_{it}) = \lambda_0 + \lambda_1 \ln(Y_{Lit}) + \lambda_2 \ln(X_{Lit}) + \lambda_3 D_{Lit} + \varepsilon_{Lit} \quad (9)$$

Here, time-invariant variables (the *Z* vector) are dropped and time dummies (*D*) account for time-specific fixed effects. Since legal supply is the initial starting point from which legal demand and illegal supply/demand emanate, we treat *Y* as lagged effects in Equation 6. In the legal supply and legal demand equations, word-of-mouth is now incorporated into *X*. The illegal supply equation now includes the effect of online user volume as a proxy for online interest, and the illegal demand equation includes a squared word-of-mouth term to account for curvilinear effects. The legal supply equation also includes legal demand, seeders, and leechers from the prior week, in light of prior findings that movie supply changes follow observed demand (e.g., Krider et al. 2005).

Estimation of both launch and post-launch systems of equations utilizes a three-stage least squares (3SLS) regression. Endogenous regressors yield inconsistent elements in ordinary least squares (OLS), and allowing the error terms of the equations to correlate is more efficient than two-stage least squares (Zellner and Theil 1962) given common endogenous regressors. Additionally, the error terms may be correlated across equations for other factors (such as awards nominations, which may affect supply and demand).

3.6. Endogeneity Checks

To account for endogeneity, we seek exogenous instrumental variables (IV), as outlined by prior movie research (e.g., Elberse and Eliashberg 2003). In particular, for each of the key endogenous variables we apply the relevant IV that could relate to the outcome measure exclusively through its relevance to the endogenous independent variable being instrumented. Unlike classic supply-demand estimation, there is no price variable to establish (and instrument for) equilibrium since movies use uniform pricing (i.e., ticket prices are the same for same-time showings, with few exceptions for IMAX or 3D films). While piracy has been instrumented using legislative efforts (e.g., Andrés and Asongu 2013) or network effects instrumented using changes in state legislatures (e.g., Miller and Tucker 2009), both present challenges given the illegal online market traversing borders and the international nature of the legal market.

Rather, we exploit variables in the illegal market with the legal market, namely how the online environment behaves differently from the legal (e.g., technology shocks such as high speed downloading and BitTorrent protocol). We follow prior movie research (e.g. Elberse and Eliashberg 2003) in instrumenting for screens and revenues. Consider the launch period; screens may be instrumented by production budget as expensive films need more distribution to cover costs, but production budget does not always translate into demand as low-budget films can also succeed. An additional instrument for screens is screen competition; screen space is limited, but the online space is relatively not (analogously, brick-and-mortar retailers face shelf space constraints while an online retailer like Amazon does not). Revenue is instrumented with competition, based on competing films’ genres, weeks in release, and MPAA rating (e.g., G, PG, PG-13, R), as competition more likely affects willingness to pay – and revenues – but not downloads (which are treated as no monetary cost to consumers).

Likewise, instrumenting the illegal market also follows prior movie research (e.g., Koschmann and Bowman 2017), as the illegal market may be instrumented using plausibly exogenous factors other than prices or policies, such as individual factors (e.g., French and Popovici 2011). Seeders are instrumented with number of piracy files as an ease of downloading; while more piracy files should result in more downloads, there needs to be more seeders with that particular file to enable network effects of downloading. This can be thought of as a 'technical ease': many piracy files with few seeders each will have different effects on downloads than few piracy files with many seeders each. Both the overall number of piracy files and the average number of piracy files in the same movie category (e.g. genre, origin of country, etc.), excluding the focal movie, are constructed as IVs. The exclusion restriction relies on the rationale that the numbers of files on other movies are not expected to directly relate to downloading of the focal movie. For leechers, given the different nature of the illegal online environment, the number of online users rating the film should affect downloads as a proxy for general interest. However, this is not necessarily the same for revenues; an online user does not have to see the actual movie in order to rate it. Therefore, we use the number of IMDB users rating the film to instrument for leechers (i.e., the exclusion that rating a movie does not directly lead to downloading it).

In the post-launch period, since production budget is time invariant, screens are instrumented using online buzz; as a proxy for general interest the searches on IMDB for a film should influence screen allocation, but not necessarily translate into actual revenues. Additionally, seasonality will affect screen supply (such as summer blockbusters) that might not directly affect realized demand for a given film.[5] Like the opening week, revenues are instrumented using competition, and seeders are instrumented using number of piracy files. For leechers, we consider not just the number of online users but also word-of-mouth as an instrument. Although higher rated movies may be more in demand (for both the legal and illegal markets), this might not affect illegal supply; piracy users gain 'street cred' from being the first to provide a movie (Kravets 2012), even for lower rated films that are hard to find.[6] All the aforementioned IVs are shown to have statistically significant correlations with the treatment variables being instrumented for (Web Appendices A4 and A7).

[5] Discussions with movie theater managers confirm seasonality in screen allocations and showings, but this is largely attributed to availability of the movies. The supply of movies incorporates expected demand due to seasonality. Seasonality is unlikely to directly affect realized demand for a particular film and piracy.

[6] Since legal supply is emphasized as each week's starting point, it is fair to wonder if screen allocation is truly endogenous in the other equations. Show times and ticket offerings are set by theaters for the coming week for operational reasons (advance ticket ordering, minimizing switching film reels among screens). Thus, we might consider seasonality and major studio releases as possible instruments for screens (Elberse and Eliashberg 2003).

3.6. Descriptive Evidence

The estimates in Table 2 for mean quality of the piracy signals leads each piracy file to be calculated with an 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. This was averaged for each film-week and standardized with a minimum of zero. Here, negative piracy quality means the file exhibited more lower quality signals than higher quality signals in its file description.

Table 3 presents descriptive statistics of the variables. Since motion pictures are released weekly, daily piracy measures across all files for a given film are averaged 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.

*** Insert Table 3 about here ***

4. Results

4.1. Launch Results

Since legal supply represents the starting point for each week, this equation utilizes estimated revenues (the HSX data) and no piracy considerations in the launch period (i.e., seeders and leechers are excluded from the screens estimation). Allowing for the full set of instruments of excluded variables in each equation yields more instruments than necessary; a Basmann (1960) over-identification test rejects that there are not enough instruments at $p < .05$ in each equation. Eight films did not have a production budget available, which was imputed using a joint multivariate normal distribution of the other observed variables, which is reliable given sufficient sample size (Demitras et al. 2008, Lee and Carlin 2010).

The 3SLS launch period results appear in Table 4 (the first-stage results appear in Web Appendix A4). The system-weighted R^2 is 0.940, indicating reasonably high fit among the four interdependent parts of the market. We particularly focus on the illegal side of the market, namely the quality of the illegal copies (the effects of the set of control variables are consistent with those reported in Elberse and Eliashberg 2003, and Koschmann and Bowman 2017). Since both sides of the equation are log transformed, the coefficients are interpreted as elasticities. In the revenues equation, neither illegal supply (B_{1P} = -0.288, $p > .20$) nor illegal demand (B_{1L} = 0.494, $p > .10$) are significant. Piracy quality has a negative average effect on revenues in the launch period (B_{1Q} = -0.483, $p < .02$). However, as more high quality copies are downloaded, this interaction has a positive effect on revenues (B_{1QL} = 0.175, $p < .01$). The positive effect suggests that in the launch period, consumption of higher quality illegal copies works as a sampling mechanism. Yet, pulling in the opposite direction is that higher quality supply hurts revenues (B_{1QP} = -0.105, $p < .01$). That is, while downloads of higher quality copies exhibit an average positive effect on revenues, too many high-quality copies available has an average negative effect.

*** Insert Table 4 about here ***

In the illegal supply equation, illegal demand is not significant (Γ_{1L} = -0.082, $p > .59$), but piracy quality is marginally significant (Γ_{1Q} = -0.369, $p < .10$), indicating that as higher quality copies are available, there is less incentive for pirates to generate more illegal supply. For illegal suppliers, though, the lure is that users seek higher quality illegal copies (the coefficient on the interaction term of quality and the number of leechers, Γ_{1QL} = 0.582, $p < .01$). Therefore, existing higher quality copies deter generating illegal supply, but consumer demand for higher quality copies spurs the illegal supply.

With illegal demand, illegal supply is not significant (Λ_{1P} = 0.179, $p > .13$) and piracy quality is also not significant (Λ_{1Q} = 0.065, $p > .67$). However, the interaction of quality and illegal supply is significant ($\Lambda_{1QP} = 0.410$, $p < .01$). Prior studies have drawn the correlation between greater illegal supply and greater illegal demand (Bhattacharjee et al. 2003). Our findings are not inconsistent with this, rather, it is that higher quality supply (demand) exerts the effect on demand (supply). Additionally, there is an overall effect of quality of piracy on revenue as a function of the seeders and leechers. Figure 3 illustrates this effect, which is greatest when there are few seeders, but many leechers.

*** Insert Figure 3 about here ***

In estimating the launch period system of equations, it is critical that certain conditions are satisfied. One, if the coefficients for each equation represent a matrix, the rank condition is necessary and sufficient to identify the model (Gujarati 2004). To check for linear dependence, a sub-matrix of the excluded equation should exist with a non-zero determinant (i.e., there are different significant variables in each equation so that a given equation is not a linear combination of the others). This condition holds here. Two, we test for weak instruments; a concern for using instrumental variables is whether the instrument is not correlated enough with the endogenous term. A rule of thumb is that if the F statistic in the reduced form regression is greater than 10, then weak instruments are not an issue (Staiger and Stock 1997). However, in the case of multiple endogenous regressors, this rule of thumb may be insufficient (Stock and Yogo 2002). Allowing a conservative estimate of no more than 5% bias relative to OLS, an equation with three endogenous regressors and six instruments requires an F statistic greater than 12.20. Regressing each equation on the endogenous and exogenous variables shows each dependent variable readily meets this test for ruling out weak instruments: screens ($F = 79.99$), revenues ($F = 55.04$), seeders ($F = 92.94$), and leechers ($F = 113.01$).

Two final considerations merit exploration here. The first is whether quality itself is endogenous. A Hausman (1978) specification test examines whether quality is endogenous in the revenues, seeders, and leechers equations of the launch period system (excluded from the screens equation as illegal copies are not presumed to affect screen allocation). In all three equations, piracy quality is not endogenous (statistical tests of all the residuals are $p > .05$). Intuitively, the piracy in our setting, as mentioned before, are not for profit and is plausibly exogenously determined by the technology equipment of the seeders.

A second consideration with the opening week is that 52 films (29.5% of the sample) exhibited no piracy. The results from Table 4 might be inflated from films with zero piracy observations. We estimate a logistic regression for a pirate/no pirate outcome based on predictors in the screens equation in Web Appendix A5, and re-estimate the launch period system of equations to exclude the observations with zero piracy (Web Appendix A6). The substantive results as reported in Table 4 do not change.

4.2. Post-Launch Results

The post-launch model, Equations (6)-(9), is similar to that of the opening week. The estimation excludes lagged endogenous variables as possible instruments due to potential autocorrelation concerns (Greene 2008). Like the launch period, a Basmann test finds there are more than enough instruments (each equation $p < .05$). To account for time effects in panel data, t-1 binary variables are used for each week of a film's theatrical run. Unlike the launch period, quality is endogenous in the seeders equation ($p < .01$). This is perhaps expected as the existence of higher quality copies might deter individuals from creating higher quality illegal copies. Since quality is endogenous, it is excluded from instrumentation, and its interaction with other endogenous variables (seeders and leechers) presents a challenge. To address this, instrument interactions are used (Wooldridge 2010), as well as screen competition, which is unlikely to affect seeders directly but which should impact piracy quality (i.e., many screens are allocated for the expectation of many customers, which may inhibit efforts to create copies, let alone high-quality copies).

Table 5 presents the 3SLS estimates for the post-launch period of $N = 1{,}202$ film-week observations (Web Appendix A7 has the first-stage results). The system-weighted R^2 is 0.958, similar to the launch period. As the starting point each week, legal supply (screen allocation) is impacted by the illegal market in the prior week. Surprisingly, seeders have a positive impact ($\alpha_{1P} = 2.523$, $p < .01$) to reinforce supply, while leechers have a negative effect ($\alpha_{1L} = -1.533$, $p < .01$). Piracy quality itself has no effect (α_{IQ} = -0.038, p > .38), but higher quality illegal supply has a negative effect on legal supply (α_{IQP} = -1.316, p < .01) while demand for higher quality illegal copies helps legal supply (α_{IQL} = 0.921, p < .01). Taken together, the supply of illegal copies in general reinforces legal supply of the product, but higher quality illegal copies negatively impact legal supply.

*** Insert Table 5 about here ***

On the revenue side, illegal demand is positive (β_{IL} = 3.745, p < .01), but piracy quality itself is not significant (β_{IQ} = 0.095, p > .54). Whereas higher quality copies exhibited a positive effect in the launch period (as a sampling mechanism), higher quality illegal copies exert a negative effect on revenues in the post-launch period (β_{IQL} = -0.788, p < .01). This is in part driven by the illegal supply, which is negative both at the baseline level (β_{IP} = -0.674, p < .01) and for higher quality illegal supply (β_{IQP} = -0.543, p < .01).

From the illegal market standpoint, illegal supply is positively influenced by illegal demand (γ_{IL} = 3.495, p < .01), just like in the launch period. However, piracy quality is not significant here (γ_{IQ} = 0.153, p > .18). Whereas demand for higher quality copies was positive in the launch period, it is negative here (γ_{IQL} = -1.161, p < .01). One possibility for this might be the presence of more higher-quality copies available in the post-launch period than during launch.

In the illegal demand equation, illegal supply has a negative effect (λ_{IP} = -0.652, p < .01), which was positive in the launch period. Piracy quality itself is still not significant (λ_{IQ} = -0.131, p > .15), but like the launch period, the supply of higher quality illegal copies encourages downloading (λ_{IQP} = 0.751, p < .01). One explanation for why the seeders coefficient is now negative may be that demand for a film (both legally and illegally) naturally declines over time. Figure 4 highlights the effect of quality on revenue as a function of the seeders and leechers in the post-launch period, similar to Figure 3 for the launch period. Unlike the opening week, the effect is greatest here when both seeders and leechers are few, which might be due to the decaying demand over time for information goods.

*** Insert Figure 4 about here ***

To identify the model, the rank condition again examines whether a matrix of significant results prevents the system from collapsing into linear dependence. This is satisfied by the results in Table 5. In evaluating weak instruments, the F statistics are large enough to rule out weak instrumentation concerns: screens ($F = 118.12$), revenues ($F = 636.24$), seeders ($F = 388.50$), and leechers ($F = 384.84$).

As a final consideration, we consider whether piracy quality evolves over time. While the time dichotomous variables account for some of this, 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.

To summarize the findings, Table 6 presents the effect of each of the endogenous variables on the four market components (screens, revenues, seeders, and leechers) for both the launch and post-launch periods. In the opening week, none of the four market components exhibits a significant effect on the illegal side of the market. Post-launch, however, each of the four market components exhibit significant effects on the other three. We put these findings in the context of a film that might have an opening box office of $100M in revenues. Although neither seeders nor leechers are significant in the launch period, a 10% increase in high quality downloads would correspond to an almost $2M gain on average (to $102M), holding quality and seeders constant. However, in the post-launch period, where the same hypothetical film might have box office revenues of $50M, a 10% increase in high quality downloads would become nearly a $4M decrease (to $46M), on average, keeping quality and seeders constant.

*** Insert Table 6 about here ***

5. Conclusion

This research examined the quality of illegal, pirated copies and its effect on the legal market. To address the subjective nature of piracy quality, an item response model uncovered latent estimates of piracy keyword signals as indicated in 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 a system of simultaneous equations with instrumental variables to address endogeneity. The empirical results are robust in regards to related research, and the intertemporal effects of piracy quality are compatible with theories in the literature. For instance, pirated movies take on an inferior quality to the original version, and the identified substitution effect resembles those predicted in Qian (2014).

This study makes several contributions to the piracy literature. First, it fills a gap in the existing literature by addressing and presenting evidence of the role of piracy quality, namely in the illegal market. This speaks to important managerial and policy questions for legal products, in particular information goods such as books, video games, and music. In light of the recent finding in Luo and Mortimer (2019) on the importance of simplifying the search and transaction process for digital goods and other small-value transactions, it is somewhat surprising that piracy has some initial positive impacts on legal sales. Prior literature has noted the aspirational nature that illegal copies might have on the legal good (e.g., Han et al. 2010), and piracy quality exhibits some similar outcomes. Second, we use a model that accounts for objective data to define quality, rather than subjective perceptions. This also speaks to the aspect of illegal supply (and its quality), a key market component often omitted from prior piracy research. Third, the analysis addresses two different timing aspects, launch and post-launch. In doing so, it speaks to effects inherent in product lifecycles, and sets a realistic expectation on the sampling impact of piracy.

A particular contribution of this study is letting the data speak to piracy quality to objectively assess quality, whereas prior research has previously relied on subjective measures. Certain piracy keywords signal higher quality (e.g., XVID and AC3) while others indicate lower quality copies (e.g., TC and DVDR). A key question for this research was whether piracy quality acted as a sampling or cannibalization mechanism. Specifically, we find that higher quality illegal copies exhibit a positive effect on revenues in the launch period; a 1% increase in higher quality illegal downloads corresponds to a 0.18% increase in revenues. We attribute this to a "liability of newness" where upon initial launch a product or brand is not well known, so consumers seek out ways to sample the product. However, once more information is available in the marketplace post-launch, higher quality copies no longer draw in samplers, but instead those consumers who are willing to forsake the genuine good. Post-launch, a 1% increase in high quality piracy downloads yields approximately a 0.79% reduction in revenues. The sampling effect could be further coupled with consumer network effects in the launch period and dwindle post-launch, as discussed in relation to prior theories. 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 cannibalism.

From these results, some managerial applications emerge. Since downloads of higher quality copies shows a positive effect on revenues in the launch period, studios should be more inclined to engage in piracy during the opening week. Although studios and theaters both want higher revenues, motion pictures represent an information good whose quality is somewhat ambiguous prior to release, with consumers wary of reviews prior to release and movie trailers only capturing the key scenes for a film. By owning the film, studios can release their own piracy quality variations, potentially as high quality but shorter versions (e.g., the first twenty minutes of a film to better capture the story and production values without putting the entire film online). The challenge from this lies in the post-launch period, where higher quality copies negatively affect revenues. By creating higher quality copies in the opening week, these pirated versions can persist longer in the online market, having a negative effect on revenues. As such, studios should be hesitant to release full-length films as pirated copies, let alone higher quality copies. One way to blunt the impact of higher quality copies post-launch would be to flood the online market with lower quality piracy. In doing so, the studio makes searching for higher quality copies more challenging for online pirates, which may lead to turning to the theaters as the only channel with a guaranteed full version of the film. In our discussion of piracy with theater managers, many consider piracy prior to and during the first week of a film's release as less a concern. Our study provides empirical generalization of such intuition and findings that higher quality piracy has a positive effect on revenues at launch. With post-launch piracy being more a concern, theater owners are likely to aim for greater enforcement of piracy creation post-launch, given that self-enforcement resources are often constrained.

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). Although observing online activity (even in its illegal form) has advantages over observing it in the physical world, we cannot account for the degree of piracy (and its quality) in this physical form. Third, while we have focused on motion pictures – which can be extended to other information goods such as books, music, and software – illegal versions of 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 evaluating the impact of piracy and use of digital information as part of the new agenda for the economics of digitization (Greenstein et al. 2010, 2013).

References

Aguiar L, Waldfogel J (2018). As streaming reaches flood stage, does it stimulate or depress music sales? *International J. of Industrial Organization*. 57:278-307.

Andrés AR., Asongu SA (2013). Fighting software piracy: Which governance tools matter in Africa? *J. of Business Ethics*. 118(3):667-682.

Bae SH, Choi JP (2006). A model of piracy. *Information Economics and Policy*. 18(3):303-320.

Basmann RL (1960). On finite sample distributions of generalized classical linear identifiability test statistics. *J. of the Amer. Statistical Association*. 55(292):650-659.

Bassat OQ, Tanner M, Guerin PJ, Stricker K, Hamed K (2016). Combating poor-quality anti-malarial medicines: A call to action. *Malaria J*. 15(302):1-12.

Belleflamme P, Peitz M. (2014). Digital piracy: An update. No. CORE Discussion Paper (2014/19), UCL.

Bhattacharjee S, Gopal RD, Sanders GL (2003). Digital music and online sharing: Software piracy 2.0? *Communications of the ACM*. 46(7):107–111.

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

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

Dhar T, Weinberg CB (2016). Measurement of interactions in non-linear marketing models: The effect of critics' ratings and consumer sentiment on movie demand. *International J. of Research in Marketing*. 33(2):392-408.

Demirtas H, Freels SA, Yucel RM (2008). Plausibility of multivariate normality assumption when multiply imputing non-Gaussian continuous outcomes: A simulation assessment. *J. of Statistical Computation and Simulation*. 78(1):69-84.

Elberse A, Eliashberg J (2003). Demand and supply dynamics for sequentially released products in international markets: The case of motion pictures. *Marketing Science*. 22(3):329-354.

Fader PS (2000). Expert report: A&M Records, Inc. v. Napster, Inc. Retrieved from http://news.findlaw.com/hdocs/docs/napster/napster/fader_070300.pdf

French MT, Popovici I (2011). That instrument is lousy! In search of agreement when using instrumental variables estimation in substance use research. *Health Economics*. 20(2):127-146.

Greene WH (2008). *Econometric Analysis*. Pearson Prentice Hall, Upper Saddle River, NJ.

Greenstein S, Goldfarb A, Tucker C (2013). *The Economics of Digitization*. Edward Elgar Publishing.

Greenstein S, Lerner J, Stern S (2010). The economics of digitization: An agenda for NSF. Greenstein, American Economic Association, Ten Years and Beyond: Economists Answer NSF's Call for Long-Term

Research Agendas. Available at SSRN: https://ssrn.com/abstract=1889153 or http://dx.doi.org/10.2139/ssrn.1889153

Gujarati D (2004) *Basic Econometrics*. McGraw-Hill, New York.

Han YJ, Nunes JC, Drèze X (2010). Signaling status with luxury goods: The role of brand prominence. *J. of Marketing Research.* 74(4):15-30.

Hausman JA (1978). Specification tests in econometrics. *Econometrica*. 46(6):1251–1271.

Hong SH (2013). Measuring the effect of Napster on recorded music sales: Difference-in-differences estimates under compositional changes. *J. of Applied Econometrics*. 28(2): 297-324.

Hui K, Png I (2003). Piracy and the legitimate demand for recorded music. *Contributions to 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 S (2008). Digital piracy: A competitive analysis. *Marketing Science*. 27(4):610–626.

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

Kaiser HF (1974). An index of factorial simplicity. *Psychometrika*. 39(1):31-36.

Kimbrough B (2014). Counterfeit automotive parts: Knock it off (October 1), https://www.dragzine.com/tech-stories/other-tech/counterfeit-automotive-parts-knock-it-off/

Koschmann A, Bowman D (2017). Simultaneous estimation of legal and illegal supply and demand: The case of motion pictures. *International Economic J.* 31(4):555-577.

Kravets D (2012). Busted camcording piracy group sought street cred. *Wired* (May 12), http://www.wired.com/2012/05/camcording-piracy-cred/

Krider R, Li T, Liu Y, Weinberg C (2005). The lead-lag puzzle of demand and distribution: A graphical method applied to movies. *Marketing Science*. 24(4):525-648.

Kubinger KD (2003). On artificial results due to using factor analysis for dichotomous variables. *Psychology Science*. 45(1):106-110.

Lahiri A, Dey D (2014). Effects of piracy on quality of information goods. *Management Science*. 59(1):245-264.

Liebowitz SJ (2008). Testing file sharing's impact on music album sales in cities. *Management Science*. 54(4):852–859.

Lee KJ, Carlin JB (2010). Multiple imputation for missing data: Fully conditional specification versus multivariate normal imputation. *Amer. J. of Epidemiology*. 171(5):624-632.

Li X, Hitt LM (2008). Self-selection and information role of online product reviews. *Information Systems Res.* 19(4):456–474.

Li X, Liao C, Xie Y (2019). Digital piracy, extrinsic incentives, and writer efforts. Working paper, University of Texas at Dallas.

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

Lu S, Wang X, Bendle N (2019). Does piracy create online word-of-mouth? An empirical analysis in movie industry. Forthcoming in *Management Science*, https://pubsonline.informs.org/doi/abs/10.1287/mnsc.2019.3298

Luo H, Mortimer J (2019). Infringing use as a path to legal consumption: Evidence from a field experiment. Working paper 971, Boston College.

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

Miller AR, Tucker C (2009). Privacy protection and technology diffusion: The case of electronic medical records. *Management Science.* 55(7):1077-1093.

Mortimer JH, Nosko C, Sorensen A (2012). Supply responses to digital distribution: Recorded music and live performances. *Inform. Econom. Policy*. 24(1):3–14.

Moul CC (2005). *A Concise Handbook of Movie Industry Economics*. Cambridge University Press, Cambridge, UK.

Nelson P (1970). Information and consumer behavior. *J. of Political Economy*. 78(2):311-329.

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

Packard G, Aribarg A, Eliashberg J, Foutz NZ (2016). The role of network embeddedness in film success. *International J. of Research in Marketing*. 33(2):328-342.

Parasuraman A, Zeithaml VA, Berry LL (1985). A conceptual model of service quality and its implications for future research. *J. of Marketing*. (4):41-50.

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

Qian Y (2014). Brand Management and Strategies against Counterfeits. *J. of Economics and Management Strategy*. 23(2):317–343.

Qian Y (2015). Counterfeiters: Foes or friends? How counterfeits affect sales by product quality tier. *Management Science*. 60(10):2381-2400.

Reid RD, Sanders NR (2001). *Operations Management*. Wiley & Sons, Hoboken, NJ.

Reeves CA, Bednar DA (1994). Defining quality: Alternatives and implications. *Academy of Management Review*. 19(3):419-445.

Spangler T (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 PE, Van Katwyk PT, Brannick MT, Chen PY (1997). When two factors don’t reflect two constructs: How item characteristics can produce artifactual factors. *J. of Management*. 23(5):659-677.

Staiger D, Stock JH (1997). Instrumental variables regression with weak instruments. *Econometrica*. 65(3):557-586

Stock JH, Yogo M (2002). Testing for weak instruments in linear IV regression, NBER technical working paper no. 284.

Swaminathan S, Bawa K (2005). Category-specific coupon proneness: The impact of individual characteristics and category-specific variables. *J. of Retailing*. 81(3):205-214.

Van Schuur WH, Kiers HA (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 HL (2015). *Entertainment Industry Economics*. Cambridge University Press, New York.

Waldfogel J (2012). Digital piracy: empirics. In M. Peitz and J. Waldfogel (Eds.), *The Oxford Handbook of the Digital Economy*, pp. 512–46. Oxford University Press.

Wilcox K, Kim HM, Sen S (2009). Why do consumers buy counterfeit luxury brands? *J. of Marketing Research*. 46(2):247-259.

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

World Health Organization (2010). Growing threat from counterfeit medicines. *Bulletin of the World Health Organization*. 88(4):247-248.

Yang Y (2019). How does online piracy affect film revenue in China? Working paper, Xiamen University.

Yoon K (2007). On the impact of digital music distribution. Working paper, Hitotsubashi University.

Zeithaml VA (1988). Consumer perceptions of price, quality, and value: A means-end model and synthesis of evidence. *J.* of *Marketing*. 52(3):2-22.

Zellner A, Theil H (1962). Three-stage least squares: Simultaneous estimation of simultaneous equations. *Econometrica*, 30(1):54–78.

Table 1. 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

Num_Users$_{it}$	Online users who rated the film	Number of online users rating the film, as a weekly average	IMDB
Online_Buzz$_{it}$	Broad interest	IMDB ranking of the film based on user search and interest	IMDB
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.

Table 2. 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 subtittles	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.

Table 3. Summary Statistics by Product Launch 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
Online_Buzz	401.69	50.36	968.32	1.43	5,000.00
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
Num_Files	2.00	10.92	0.00	97.71	2.00

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
Online_Buzz	560.03	54.50	1,254.63	1.00	5,000.00
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
Num_Files	29.11	15.71	37.76	0.00	289.43

Notes. Dollars are in thousands (000).

Table 4. Launch Period Estimation Results (N = 173)

	DV: ln(Screens)		DV: ln(Revenue)		DV: ln(Seeders)		DV: ln(Leechers)	
Intercept	0.944	***	-0.051		-1.742		2.100	
	(0.334)		(1.332)		(1.867)		(1.667)	
ln(Screens)			0.919	***	-0.585		-0.350	
			(0.301)		(0.641)		(0.464)	
ln(Revenue)[a]	0.519	***			0.414		0.005	
	(0.037)				(0.334)		(0.277)	
ln(Seeders)			-0.288				0.179	
			(0.228)				(0.119)	
ln(Leechers)			0.494		-0.082			
			(0.298)		(0.155)			
ln(Quality)			-0.483	**	-0.369	*	0.065	
			(0.188)		(0.219)		(0.157)	
ln(Quality)*ln(Seeders)			-0.105	***			0.410	***
			(0.031)				(0.017)	
ln(Quality)*ln(Leechers)			0.175	***	0.582	***		
			(0.037)		(0.024)			
ln(Prod_Budget)	0.125	***			0.275	**		
	(0.031)				(0.117)			
ln(Actor_Power)	0.001		-0.002		-0.076		-0.124	**
	(0.021)		(0.047)		(0.068)		(0.056)	
ln(Advertising)	0.141	***	0.082		0.011		0.163	
	(0.025)		(0.125)		(0.168)		(0.136)	
ln(Critics)	-0.416	***	0.516	**	-0.142		-0.114	
	(0.071)		(0.224)		(0.390)		(0.308)	
ln(Screen_Comp_New)	-0.111	***						
	(0.030)							
ln(Screen_Comp_Ong)	0.076							
	(0.146)							
Major_Studio	0.035							
	(0.058)							
ln(Revenue_Comp)			-0.021					
			(0.113)					
Seasonality			0.804	***				
			(0.228)					
Online_Buzz					0.000		0.000	
					(0.000)		(0.000)	
ln(Previous_Days)					0.174	**	-0.230	***
					(0.087)		(0.079)	
System Weighted R^2	0.940							

Notes. 3SLS Frequentist estimates shown, standard errors in parentheses. [a] is estimated in Screens equation.
*** $p < 0.01$; ** $p < 0.05$;* $p < 0.10$.

Table 5. Post-Launch Period Estimation Results (*N* = 1,202)

	DV: ln(Screens)		DV: ln(Revenue)		DV: ln(Seeders)		DV: ln(Leechers)	
Intercept	0.639	**	-3.150	***	-1.606	***	1.220	***
	(0.292)		(0.377)		(0.281)		(0.247)	
ln(Screens)			1.336	***	0.696	***	-0.578	***
			(0.063)		(0.087)		(0.074)	
ln(Revenue)[a]	0.542	***			-0.498	***	0.413	***
	(0.014)				(0.058)		(0.048)	
ln(Seeders)[b]	2.523	***	-0.674	***			-0.652	***
	(0.425)		(0.257)				(0.037)	
ln(Leechers)[b]	-1.533	***	3.745	***	3.495	***		
	(0.480)		(0.417)		(0.156)			
ln(Quality)[b]	-0.038		0.095		0.153		-0.131	
	(0.044)		(0.158)		(0.116)		(0.092)	
ln(Quality)*ln(Seeders)[b]	-1.316	***	-0.543	***			0.751	***
	(0.215)		(0.003)				(0.001)	
ln(Quality)*ln(Leechers)[b]	0.921	***	-0.788	***	-1.161	***		
	(0.244)		(0.176)		(0.077)			
ln(Screen_Comp_New)	-0.060	**						
	(0.029)							
ln(Screen_Comp_Ong)	0.697	***						
	(0.113)							
ln(Revenue_Comp)			-0.023					
			(0.025)					
Seasonality			0.174	***				
			(0.052)					
Online_Buzz					0.000			
					(0.000)			
ln(WOM)	0.310	***	0.219				0.096	
	(0.113)		(0.143)				(0.062)	
WOM^2							-0.002	**
							(0.001)	
ln(Num_Users)					0.003			
					(0.009)			
System Weighted R^2	0.958							

Notes. 3SLS Frequentist estimates shown, standard errors in parentheses. Weekly time dummies not shown. [a] is estimated in Screens equation. [b] is lagged in Screens equation.
*** $p < 0.01$; ** $p < 0.05$; * $p < 0.10$.

Table 6. Summary of Significant Results in Launch and Post-Launch Systems of Equations

Launch Period

Effect of: \ on:	Screens	Revenue	Seeders	Leechers
Screens		+	n.s.	n.s.
Revenue	+		n.s.	n.s.
Seeders		n.s.		n.s.
Leechers		n.s.	n.s.	
Quality		-	n.s.	n.s.
Quality*Seeders		-		+
Quality*Leechers		+	+	

Post-Launch Period

Effect of: \ on:	Screens	Revenue	Seeders	Leechers
Screens		+	+	-
Revenue	+		-	+
Seeders	+	-		-
Leechers	-	+	+	
Quality	n.s.	n.s.	n.s.	n.s.
Quality*Seeders	-	-		+
Quality*Leechers	+	-	-	

Notes. n.s. denotes not significant at the 5% level, while +/- denote significantly positive or negative effects, respectively. For example, in the launch period, screen availability has a positive effect on revenues, but no significant effect on seeders or leechers.

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

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

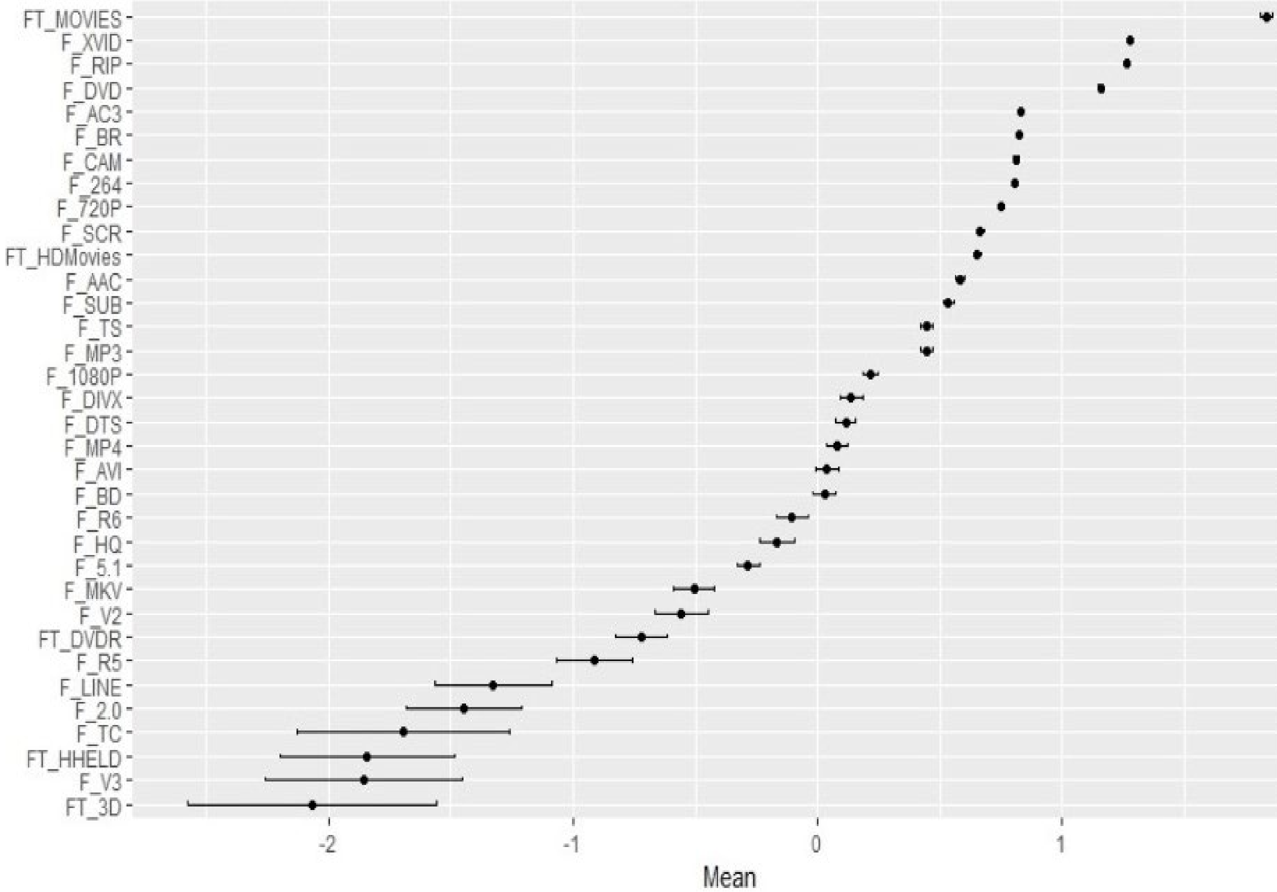


Figure 3. Effect of Quality on Revenue as a Function of Seeders and Leechers (Launch)

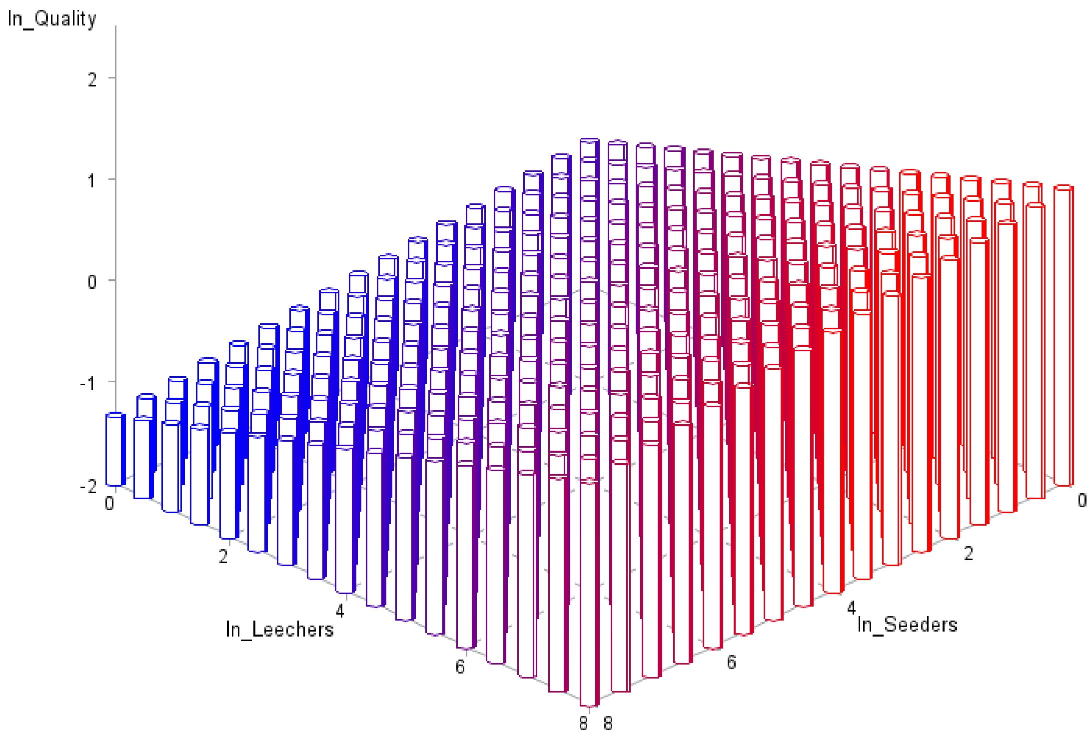
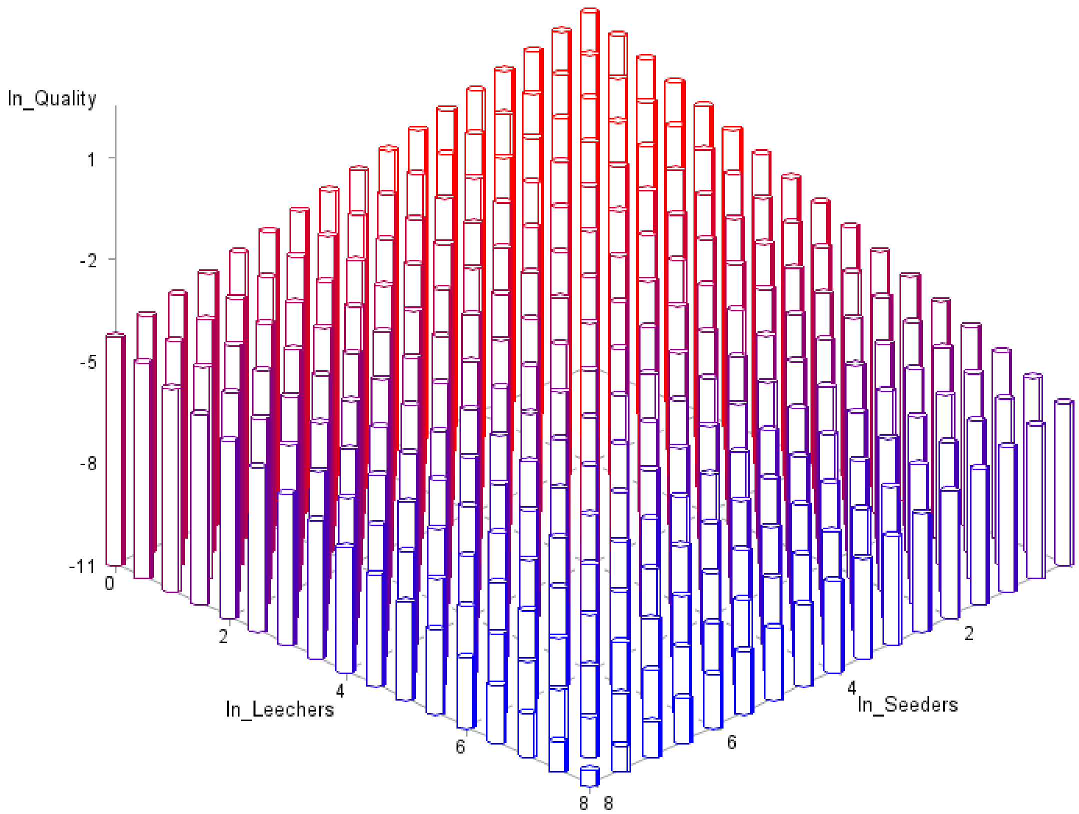


Figure 4. Effect of Quality on Revenue as a Function of Seeders and Leechers (Post-Launch)



Web Appendix A1. 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 A2. 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 A3. Fully-specified model equations for Launch and Post-Launch

$$\ln(Screens_{i1}) = A_0 + A_{1R}\ln(Revenue\_Est_{Si1}) + A_{2a}\ln(Screen\_Comp\_New_{Si1}) + A_{2b}\ln(Screen\_Comp\_Ong_{Si1}) + A_{3a}\ln(Prod\_Budget_{Si1}) + A_{3b}\ln(Actor\_Power_{Si1}) + A_{3c}\ln(Advertising_{Si1}) + A_{3d}\ln(Critics_{Si1}) + A_{3e}\ln(Major\_Studio_{Si1}) + \varepsilon_{Si1} \quad (2)$$

$$\ln(Revenue_{i1}) = B_0 + B_{1S}\ln(Screens_{Ri1}) + B_{1P}\ln(Seeders_{Ri1}) + B_{1L}\ln(Leechers_{Ri1}) + B_{1Q}\ln(Quality_{Ri1}) + B_{1QP}\ln(Quality_{Ri1})*\ln(Seeders_{Ri1}) + B_{1QL}\ln(Quality_{Ri1})*\ln(Leechers_{Ri1}) + B_{2a}\ln(Revenue\_Comp_{Ri1}) + B_{2b}\ln(Seasonality_{Ri1}) + B_{3a}\ln(Actor\_Power_{Ri1}) + B_{3b}\ln(Advertising_{Ri1}) + B_{3c}\ln(Critics_{Ri1}) + \varepsilon_{Ri1} \quad (3)$$

$$\ln(Seeders_{i1}) = \Gamma_0 + \Gamma_{1S}\ln(Screens_{Pi1}) + \Gamma_{1R}\ln(Revenue_{Pi1}) + \Gamma_{1L}\ln(Leechers_{Pi1}) + \Gamma_{1Q}\ln(Quality_{Pi1}) + \Gamma_{1QL}\ln(Quality_{Pi1})*\ln(Leechers_{Pi1}) + \Gamma_{2a}\ln(Online\_Buzz_{Pi1}) + \Gamma_{3a}\ln(Prod\_Budget_{Pi1}) + \Gamma_{3b}\ln(Actor\_Power_{Pi1}) + \Gamma_{3c}\ln(Advertising_{Pi1}) + \Gamma_{3d}\ln(Critics_{Pi1}) + \Gamma_{3e}\ln(Previous\_Days_{Pi1}) + \varepsilon_{Pi1} \quad (4)$$

$$\ln(Leechers_{i1}) = \Lambda_0 + \Lambda_{1S}\ln(Screens_{i1}) + \Lambda_{1R}\ln(Revenue_{i1}) + \Lambda_{1P}\ln(Seeders_{i1}) + \Lambda_{1Q}\ln(Quality_{i1}) + \Lambda_{1QP}\ln(Quality_{i1})*\ln(Seeders_{i1}) + \Lambda_{2a}\ln(Online\_Buzz_{i1}) + \Lambda_{3a}\ln(Actor\_Power_{i1}) + \Lambda_{3b}\ln(Advertising_{i1}) + \Lambda_{3c}\ln(Critics_{i1}) + \Lambda_{3d}\ln(Previous\_Days_{i1}) + \varepsilon_{Li1} \quad (5)$$

$$\ln(Screen_{it}) = \alpha_0 + \alpha_{1R}\ln(Revenue\_Est_{Sit}) + \alpha_{1P}\ln(Seeders_{Sit-1}) + \alpha_{1L}\ln(Leechers_{Sit-1}) + \alpha_{1Q}\ln(Quality_{Sit-1}) + \alpha_{1QP}\ln(Quality_{Sit-1})*\ln(Seeders_{Sit-1}) + \alpha_{1QL}\ln(Quality_{Sit-1})*\ln(Leechers_{Sit-1}) + \alpha_{2a}\ln(Screen\_Comp\_New_{Sit}) + \alpha_{2b}\ln(Screen\_Comp\_Ong_{Sit}) + \alpha_{2c}\ln(WOM_{Sit}) + \alpha_3 D_{Sit} + \varepsilon_{Sit} \quad (6)$$

$$\ln(Revenue_{it}) = \beta_0 + \beta_{1S}\ln(Screens_{Rit}) + \beta_{1P}\ln(Seeders_{Rit}) + \beta_{1L}\ln(Leechers_{Rit}) + \beta_{1Q}\ln(Quality_{Rit}) + \beta_{1QP}\ln(Quality_{Rit})*\ln(Seeders_{Rit}) + \beta_{1QL}\ln(Quality_{Rit})*\ln(Leechers_{Rit}) + \beta_{2a}\ln(Revenue\_Comp_{Rit}) + \beta_{2b}\ln(Seasonality_{Rit}) + \beta_{2c}\ln(WOM_{Rit}) + \beta_3 D_{Rit} + \varepsilon_{Rit} \quad (7)$$

$$\ln(Seeders_{it}) = \gamma_0 + \gamma_{1S}\ln(Screens_{Pit}) + \gamma_{1R}\ln(Revenue_{Pit}) + \gamma_{1L}\ln(Leechers_{Pit}) + \gamma_{1Q}\ln(Quality_{Pit}) + \gamma_{1QL}\ln(Quality_{Pit})*\ln(Leechers_{Pit}) + \gamma_{2a}\ln(Online\_Buzz_{Pit}) + \gamma_{2b}\ln(Num\_Users_{Pit}) + \gamma_3 D_{Pit} + \varepsilon_{Pit} \quad (8)$$

$$\ln(Leechers_{it}) = \lambda_0 + \lambda_{1S}\ln(Screens_{it}) + \lambda_{1R}\ln(Revenue_{it}) + \lambda_{1P}\ln(Seeders_{it}) + \lambda_{1Q}\ln(Quality_{it}) + \lambda_{1QP}\ln(Quality_{it})*\ln(Seeders_{it}) + \lambda_{2a}\ln(WOM_{it}) + \lambda_{2b}WOM_{it}^2 + \lambda_3 D_{Lit} + \varepsilon_{Lit} \quad (9)$$

Web Appendix A4. Launch Period First-Stage Least Squares Estimates

	DV: ln(Screens)		DV: ln(Revenue)		DV: ln(Seeders)		DV: ln(Leechers)	
Intercept	3.138	***	1.980	**	-7.458	***	-5.262	**
	(0.516)		(0.892)		(2.674)		(2.213)	
ln(Quality)	0.071	*	0.146	**	1.252	***	1.395	***
	(0.036)		(0.062)		(0.187)		(0.155)	
ln(Prod_Budget)	0.202	***	0.112		0.544	**	0.403	**
	(0.043)		(0.074)		(0.220)		(0.182)	
ln(Actor_Power)	-0.022		-0.038		-0.086		-0.114	
	(0.025)		(0.043)		(0.130)		(0.107)	
ln(Advertising)	0.332	***	0.412	***	0.043		0.057	
	(0.029)		(0.050)		(0.150)		(0.124)	
ln(Critics)	-0.410	***	-0.124		-0.152		-0.347	
	(0.111)		(0.192)		(0.576)		(0.477)	
ln(Screen_Comp_New)	-0.085	**	0.008		0.285		0.248	
	(0.040)		(0.070)		(0.210)		(0.173)	
ln(Screen_Comp_Ong)	-0.133		-0.025		-0.803		-1.038	
	(0.183)		(0.316)		(0.946)		(0.783)	
Major_Studio	0.153	**	0.393	***	-0.040		-0.070	
	(0.073)		(0.127)		(0.379)		(0.314)	
ln(Revenue_Comp)	-0.254	***	-0.392	***	-0.332		-0.414	*
	(0.056)		(0.096)		(0.288)		(0.238)	
Seasonality	-0.138		0.155		-1.355	**	-1.457	***
	(0.112)		(0.193)		(0.579)		(0.479)	
Online_Buzz	0.000		0.000		0.001	***	0.000	**
	(0.000)		(0.000)		(0.000)		(0.000)	
ln(Previous_Days)	-0.127	***	-0.292	***	-0.176		-0.297	**
	(0.030)		(0.053)		(0.158)		(0.131)	
ln(WOM)	-0.268		-0.060		-0.917		-0.191	
	(0.173)		(0.299)		(0.897)		(0.742)	
ln(Num_Users)	0.183	***	0.411	***	0.943	***	0.781	***
	(0.036)		(0.063)		(0.188)		(0.156)	
ln(Num_Files)	0.139	***	0.309	***	1.648	***	1.526	***
	-0.034		-0.060		-0.161		-0.142	
Adjusted R^2	0.843		0.779		0.516		0.592	
F-statistic	63.540		38.630		16.460		17.810	

*** $p < 0.01$; ** $p < 0.05$;* $p < 0.10$.

Web Appendix A5. Log-Odds Ratio that a Film will be Pirated in the Launch Period

	DV: Is Pirated	
Intercept	9.361	***
	(2.646)	
ln(Screens)	-1.406	***
	(0.428)	
ln(Prod_Budget)	-0.302	
	(0.273)	
ln(Actor_Power)	0.074	
	(0.164)	
ln(Advertising)	0.371	
	(0.227)	
ln(Critics)	-0.873	
	(0.577)	
Major_Studio	0.146	
	(0.456)	
Online_Buzz	0.000	
	(0.000)	
Seasonality	0.935	
	(0.648)	

Notes. Standard errors in parentheses.
*** $p < 0.01$; ** $p < 0.05$;* $p < 0.10$.

Web Appendix A6. Launch Period Estimation Results Excluding Films with No Piracy (*N* = 121)

	DV: ln(Screens)		DV: ln(Revenue)		DV: ln(Seeders)		DV: ln(Leechers)	
Intercept	1.076	***	1.207		5.732	**	3.505	*
	(0.353)		(1.610)		(2.697)		(2.101)	
ln(Screens)			0.652	*	-0.470		-0.179	
			(0.378)		(0.736)		(0.464)	
ln(Revenue)[a]	0.490	***			0.195		0.421	
	(0.043)				(0.365)		(0.264)	
ln(Seeders)			0.094				-0.183	
			(0.160)				(0.152)	
ln(Leechers)			0.150		0.190			
			(0.325)		(0.336)			
ln(Quality)			-0.496		-3.559	***	-1.465	***
			(0.334)		(0.495)		(0.397)	
ln(Quality)*ln(Seeders)			-0.141	***			0.348	***
			(0.029)				(0.019)	
ln(Quality)*ln(Leechers)			0.286	***	0.597	***		
			(0.036)		(0.032)			
ln(Prod_Budget)	0.152	***			0.063			
	(0.032)				(0.229)			
ln(Actor_Power)	-0.026		0.019		0.076		-0.045	
	(0.021)		(0.065)		(0.104)		(0.078)	
ln(Advertising)	0.120	***	0.113		0.048		-0.061	
	(0.023)		(0.116)		(0.185)		(0.129)	
ln(Critics)	-0.260	***	0.334		-0.340		-0.017	
	(0.072)		(0.240)		(0.408)		(0.309)	
ln(Screen_Comp_New)	-0.132	***						
	(0.029)							
ln(Screen_Comp_Ong)	0.150							
	(0.137)							
StudioMajor	-0.021							
	(0.060)							
ln(Revenue_Comp)			0.061					
			(0.118)					
Seasonality			0.476	**				
			(0.234)					
Online_Buzz					0.000		0.000	
					(0.000)		(0.000)	
ln(Previous_Days)					0.334	***	-0.169	*
					(0.114)		(0.095)	
System Weighted R^2	0.933							

Notes: 3SLS estimates shown. [a] is estimated in Screens equation.
*** $p < 0.01$; ** $p < 0.05$;* $p < 0.10$.

Web Appendix A7. Post-Launch Period First-Stage Least Squares Estimates

	DV: lnScreens		DV: ln(Revenue)		DV: ln(Seeders)		DV: ln(Leechers)		DV: ln(Quality)	
Intercept	5.901	***	9.314	***	13.429	***	10.950	***	4.031	***
	(1.906)		(2.385)		(2.882)		(2.366)		(0.852)	
ln(Screen_Comp_New)	-0.041		-0.075		-0.054		-0.027		0.038	*
	(0.043)		(0.054)		(0.065)		(0.053)		(0.019)	
ln(Screen_Comp_Ong)	0.155		-0.030		0.038		0.013		-0.012	
	(0.153)		(0.192)		(0.232)		(0.190)		(0.068)	
ln(Revenue_Comp)	-0.447	***	-0.565	***	-0.146	*	-0.086		-0.043	*
	(0.050)		(0.063)		(0.076)		(0.062)		(0.022)	
Seasonality	-0.221	**	0.336	**	-0.122		-0.028		0.026	
	(0.107)		(0.134)		(0.162)		(0.133)		(0.048)	
Online_Buzz	0.001	***	0.001	***	0.000		0.000		0.000	***
	(0.000)		(0.000)		(0.000)		(0.000)		(0.000)	
ln(WOM)	-0.899		-3.200	**	-10.020	***	-8.041	***	-2.837	***
	(1.014)		(1.269)		(1.533)		(1.259)		(0.454)	
ln(Num_Users)	0.176		-0.337		-1.106	***	-0.936	***	-0.359	***
	(0.222)		(0.277)		(0.335)		(0.275)		(0.099)	
ln(Num_Files)	0.650	***	1.088	***	3.890	***	3.168	***	1.411	***
	(0.156)		(0.195)		(0.236)		(0.194)		(0.070)	
d3	-0.623	***	-0.958	***	-0.223		-0.328	***	-0.092	**
	(0.102)		(0.127)		(0.154)		(0.126)		(0.046)	
d4	-1.181	***	-1.713	***	-0.504	***	-0.785	***	-0.122	***
	(0.105)		(0.131)		(0.159)		(0.130)		(0.047)	
d5	-1.626	***	-2.280	***	-0.599	***	-0.965	***	-0.110	**
	(0.108)		(0.136)		(0.164)		(0.134)		(0.048)	
d6	-1.943	***	-2.739	***	-0.560	***	-1.039	***	-0.118	**
	(0.114)		(0.143)		(0.172)		(0.142)		(0.051)	
d7	-2.141	***	-3.033	***	-0.595	***	-1.117	***	-0.149	***
	(0.123)		(0.154)		(0.186)		(0.152)		(0.055)	
d8	-2.250	***	-3.130	***	-0.630	***	-1.178	***	-0.135	**
	(0.136)		(0.171)		(0.206)		(0.169)		(0.061)	
d9	-2.402	***	-3.270	***	-0.608	***	-1.104	***	-0.154	**
	(0.150)		(0.188)		(0.228)		(0.187)		(0.067)	
d10	-2.594	***	-3.480	***	-0.519	**	-0.947	***	-0.152	**
	(0.172)		(0.216)		(0.261)		(0.214)		(0.077)	
d11	-2.813	***	-3.728	***	-0.629	**	-0.890	***	-0.251	***
	(0.197)		(0.246)		(0.297)		(0.244)		(0.088)	
d12	-3.071	***	-4.034	***	-1.104	***	-1.258	***	-0.415	***
	(0.226)		(0.282)		(0.341)		(0.280)		(0.101)	
d13	-3.159	***	-4.123	***	-1.116	***	-1.349	***	-0.392	***
	(0.270)		(0.338)		(0.408)		(0.335)		(0.121)	
d14	-3.486	***	-4.541	***	-0.914	*	-1.249	***	-0.218	
	(0.313)		(0.392)		(0.474)		(0.389)		(0.140)	
d15	-3.415	***	-4.411	***	-1.104	**	-1.454	***	-0.272	*
	(0.346)		(0.433)		(0.523)		(0.429)		(0.155)	
d16	-3.493	***	-4.404	***	-1.271	**	-1.721	***	-0.302	*
	(0.396)		(0.496)		(0.599)		(0.492)		(0.177)	
d17	-3.747	***	-4.733	***	-1.447	**	-1.903	***	-0.293	
	(0.434)		(0.543)		(0.657)		(0.539)		(0.194)	

d18	-3.400	***	-4.531	***	-1.398	*	-2.068	***	-0.270	
	(0.483)		(0.605)		(0.731)		(0.600)		(0.216)	
d19	-2.934	***	-4.181	***	-1.621		-2.350	***	-0.311	
	(0.669)		(0.837)		(1.012)		(0.831)		(0.299)	
d20	-3.418	***	-4.996	***	-1.766	*	-2.595	***	-0.308	
	(0.670)		(0.838)		(1.013)		(0.831)		(0.300)	
d21	-3.904	***	-5.669	***	-1.786	*	-2.687	***	-0.286	
	(0.670)		(0.838)		(1.013)		(0.832)		(0.300)	
d22	-4.394	***	-6.212	***	-1.338		-2.185	*	-0.125	
	(0.942)		(1.179)		(1.425)		(1.170)		(0.422)	
d23	-5.148	***	-6.814	***	-1.433		-2.261	*	-0.104	
	(0.942)		(1.180)		(1.425)		(1.170)		(0.422)	
ln(Num_Users) x ln(WOM)	0.151		0.535	***	1.070	***	0.890	***	0.326	***
	(0.116)		(0.145)		(0.175)		(0.144)		(0.052)	
ln(Num_Users) x Online_Buzz	0.000		0.000		0.000	***	0.000	*	0.000	*
	(0.000)		(0.000)		(0.000)		(0.000)		(0.000)	
ln(Num_Users) x ln(Num_Files)	-0.055	***	-0.100	***	-0.308	***	-0.255	***	-0.117	***
	(0.017)		(0.021)		(0.025)		(0.021)		(0.008)	
ln(WOM) x Online_Buzz	-0.001	***	-0.001	***	0.000		0.000		0.000	***
	(0.000)		(0.000)		(0.000)		(0.000)		(0.000)	
Online_Buzz x ln(Num_Files)	0.000		0.000		0.000	***	0.000	***	0.000	***
	(0.000)		(0.000)		(0.000)		(0.000)		(0.000)	
Adjusted R^2	0.596		0.628		0.607		0.577		0.621	
F-statistic	53.1		60.62		55.47		49.27		58.83	

Notes. 'd' variables denote time dummies (e.g., d3 is 3rd week of film release), standard errors in parentheses. Week 2 omitted is the base week and omitted as a dummy variable.
*** $p < 0.01$; ** $p < 0.05$;* $p < 0.10$.