NBER WORKING PAPER SERIES

ASSESSING THE QUALITY OF ILLEGAL COPIES AND ITS IMPACT ON REVENUES AND DISTRIBUTION

Anthony Koschmann
Yi Qian

Working Paper 27649
http://www.nber.org/papers/w27649

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
August 2020, Revised February 2025

The authors would like to thank Fred Feinberg, Chris Forman, Shane Greenstein, Avi Goldfarb, Xixi Hu, Ginger Jin, Josh Lerner, Xiaolin Li, Julie Mortimer, Kevin Ryan, Marie Thursby, Hui Xie, Ying Xie, Zining Wang, Joel Waldfolgel, and Chuck Weinberg for helpful comments. Yi would like to acknowledge grant support from the Canadian SSHRC IG 435-2018-0519, and thank Bowen Zhang for excellent RA work. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Assessing the Quality of Illegal Copies and its Impact on Revenues and Distribution
Anthony Koschmann and Yi Qian
NBER Working Paper No. 27649
August 2020, Revised February 2025
JEL No. K42, M21, M31, O3

ABSTRACT

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

Anthony Koschmann
Eastern Michigan University
121 Hill Hall
College of Business
Ypsilanti, MI 48178
akoschma@emich.edu

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

1. Introduction

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

Although piracy poses a threat to almost any information good, prior empirical research of piracy quality has encountered three key limitations. First, the quality of pirated copies has been assessed using subjective user ratings, such as the video and audio quality of pirated films (Ma et al., 2014). While useful in thinking about the effect of piracy quality, subjective interpretations encounter validity concerns. Second, even non-subjective measures, like the binary coding of a high-definition (HD) keyword (Lu et al., 2020; Ackermann et al. 2020) treats piracy quality as an either-or proposition. Albeit helpful in understanding the impact of piracy quality, illegal copies convey more information beyond just its source type (e.g., camcorder, film reel, or digital copy). Indeed, product quality – both legal and illegal – is composed of many attributes that send signals to consumers (e.g., Zeithaml, 1988). Third, missing from prior studies is the impact of consumer activity (i.e., downloads and uploads) of pirated copies in relation to piracy quality, particularly its effect on distribution (i.e., both legal and illegal supply).

To address these gaps, we use objective measures to define piracy quality and its effect on supply and demand in both the legal and illegal markets using the context of motion pictures. Motion pictures are a focal area in marketing research (e.g., Dhar & Weinberg, 2016; Packard et al., 2016, McKenzie, 2023) for several reasons. First, the effects of piracy may be clearer in movies than in other information goods (Lu et al., 2020); few movies are seen by a consumer multiple times in theaters but illegally downloaded music or software may be consumed repeatedly at home. Second, while makers of durable products are also concerned about the quality of illegal copies, 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 observe; physical products require finding markets with physical transactions. Piracy transactions, however, have some visibility online where users transact with less fear of being caught. Finally, movie piracy is an area that allows us to build on prior theories and findings.

Although piracy quality is of interest to producers and presents a need for understanding, it faces similar measurement challenges as quality of legal goods. Piracy quality, like product quality, is an important product aspect that affects consumer choice, but quality is often treated subjectively as consumers perceive certain signals (Zeithaml, 1988). Indeed, piracy quality is rooted in "any valued attribute of a product" (Chellappa & Shivendu, 2005, p.402), even though attribute valuation lends itself to open interpretation arising from heterogeneous tastes. Building on prior piracy research and production quality, we propose that high-quality piracy provides information to consumers when product uncertainty is high, such as during new product launch. Piracy that is of higher quality should provide more information to consumers by more closely resembling the original good; this reduces uncertainty and stimulates sales, particularly among enthusiastic customers who seek out product information. Post-launch, however, more information is available (such as word of mouth) and product uncertainty is lower. The most willing consumers have likely purchased, leaving more consumers who are less interested and willing to pay, making high-quality piracy more substitutable for the original good.

Since product quality can be treated in two ways, perceived and objective (Monroe & Krishnan, 1985), and reflects the underlying product attributes, we model piracy quality using the visible piracy file keywords to assemble the various aspects of quality into a unidimensional index. We estimate this piracy quality using item response theory (IRT). IRT models estimate the relation of items on a latent spectrum, where piracy quality is estimated from the relative ideal points (mean values) of piracy keywords across illegal files. An advantage here over prior methods is the treatment of quality as a continuous measure by utilizing common, observable piracy attributes rather than one or two salient features; this creates a measure more concise and precise for the main estimations.

The dataset consists of movies in wide release in the United States and Canada. Correspondingly, a daily panel tracks those movies' search results and activity on Pirate Bay while the films are first-run in theaters. The data includes box office revenues, screen availability, piracy, advertising, and movie characteristics. Since most research studies on piracy do not account for illegal supply, we use seemingly unrelated regressions (SUR) to jointly estimate the effects of piracy quality on both the legal and illegal sides of the market, using copulas to account for endogeneity.

We find that a 1% increase in the quality of the pirated copies, conditioned on piracy downloads (leechers, or downloads of the illegal file), corresponds on average to a 0.143% increase in revenues in the launch period. Upon market introduction, movies lack some information for consumers, so higher quality copies act more like a sampling mechanism. Post-launch, however, our findings show that a 1% increase in the quality of illegal copies, conditioning on a level of leechers, yields a -0.075% decrease in revenues. As more product information permeates the marketplace, higher quality illegal copies cannibalize sales.

This study makes several contributions to the illegal copy literature. First, we propose that higher quality piracy alleviates product uncertainty by providing product information. We build on prior research of information goods to separate the launch and post-launch periods, as product uncertainty is higher at launch than post-launch. Second, to the authors' knowledge, this is the first study to use an objective, continuous measure of piracy quality, by operationalizing piracy quality through latent recovery of pirated copy attributes (keyword signals). Third, we assess the effect of piracy quality on supply (both legal and illegal) which is often omitted in piracy research. Fourth, 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. Fifth, we further test moderation effects through stratification analyses. In particular, we find that higher quality downloads of major studio releases have no differential effect on revenues, but higher quality downloads of films with greater production budgets, star power, and advertising expense show negative effects on revenue at launch. These findings differ from prior studies which found piracy quality had no significant effect on revenues (Ma et al., 2014, Lu et al., 2020), in part because of the more nuanced definition of piracy quality and separate period estimations. The substantive results suggest studios and producers have a unique and advantageous tool for fighting piracy: the legal good itself. Producers can create their own derived copies of the genuine good to reduce uncertainty, encouraging the right kind of sampling and discouraging the wrong kind of cannibalism. The expansion effects of sampling echo theoretical predictions (Bawa and Shoemaker 2024) in the piracy context. These findings give useful meaning to both managers and policymakers regarding the quality nature of illegal variants, while also extending the piracy literature.

2. Theoretical Development and Contribution

In this section, we first motivate the importance of piracy quality, then discuss its effect regarding product uncertainty. This includes customers' information seeking as well as customer enthusiasm.

2.1 Piracy Quality and Production Quality

Piracy quality has been treated as how closely the illegal copy resembles the genuine product (Geng & Lee, 2013). As a derived good, the pirated version presumably exhibits lower quality than the genuine product (Sundarajan, 2004; Jain, 2008; Geng & Lee, 2013; Lahiri & Dey, 2013; Machado et al., 2017; Dey et al., 2019). In general, consumers prefer products of higher quality rather than lower quality given comparable prices. If the quality gap between the pirated copy and the genuine product is large, producers can expect minimal effects from pirated copies (Lahiri & Dey, 2013). If the quality gap is small, producers can lower prices (Sundarajan, 2004). However, the quality gap between the copy and genuine good is unknown to consumers unless they have consumed both. We nuance Geng & Lee’s view of piracy quality to then mean that piracy copies of higher quality should be more informative and more resembling the original good than lower quality piracy copies. Hence, a more complex relationship between piracy quality and product performance exists that speaks to product uncertainty and quality.

Research into the aspects and effects of piracy quality have primarily approached the issue analytically, owing to two challenges in collecting data. One, observing illegal activity is not easy. Two, quality is difficult to measure, as it has two related aspects: objective quality (where quantifiably more of something is better and is easier to assess), and perceived quality (the subjective judgment of quality: Mitra & Golder, 2006). More recently, research has sought to overcome these two challenges to empirically address piracy quality. One such approach has been to try and treat piracy quality as being high or low quality depending on its source. Lu et al. (2020) code piracy files as either high-quality (from a high-definition source) or low quality (from a camcorder) to examine the effect on revenues and user ratings of films. In Liu (2019) and Ackermann et al. (2020), piracy is similarly coded as high or low quality if a particular keyword is present, such as ‘720P’, ‘1080P’, ‘x264’, or ‘Blu-Ray’.

Such designations of high or low quality in film piracy is useful in understanding piracy effects. However, pirated copies exist that cannot be classified simply as either high or low quality, or may even be presented to the market as being both high and low quality. For instance, Danaher et al. (2019) look at the effect of high-quality piracy copies stemming from early release (i.e., legal streaming video) on theatrical revenues of 33 Chinese movies. These were distinguished from camcorder copies captured in theaters. Yet, other sources may still generate piracy that are neither theatrical camcorders nor at-home recording, such as film reel transfers or pre-release copies (for film critics or pre-release festival screeners). Additionally, using a low- or high-quality indicator does not account for some combination of the two. For instance, camcorder copies can be re-mixed with high quality audio afterwards, while high-quality film transfer copies might have audio mixed onto only one track. In this scenario, the camcorder quality moves closer to higher quality, while the film transfer copy has degraded quality. Ma et al. (2014) look at the effect of pre-release piracy on box office revenues for 52 movies, using subjective user ratings of pirated files’ audio and video as one combined measure. Aside from the subjective interpretation of quality, such scoring does not indicate whether a piracy file with high-quality video (but low-quality audio) is better than another file with low-quality video (but high-quality audio). As such, information goods have many attributes, whose combinations signal to consumers a more refined view that piracy quality is not an ‘either-or’ byproduct.

While the aforementioned studies have advanced our understanding of piracy quality, three limitations persist. One, piracy quality includes both objective and perceived quality, rather than one or the other. A file having greater video resolution, or five-channel surround sound (versus two channels) has higher objective quality. Yet the subjective nature of perceived quality is important, such as whether the audio format AAC is better than AC3, and how this indication can be incorporated into a definition of piracy quality. Two, pirated copies, as will be shown, impart considerable information to consumers, and many pirated copies do not fit easily into a 'low-or-high' definition (e.g., Danaher et al. 2019; Lu et al. 2020). While select keywords like 'HD', '1080P', or 'Blu-Ray' may signal high quality sourcing (e.g., Ackermann et al. 2020), such keywords may be neither necessary nor sufficient for establishing high quality. Is 1080P or HD more indicative of quality? For instance, a pirated copy with 1080P has greater video resolution than one with 720P. However, if the pirated file with 720P has five channel surround sound while the 1080P has two audio channels, is the 1080P file really the one with higher quality? We empirically account for such inter-relatedness and tradeoffs among the set of observable attributes. Three, research into piracy quality has focused on the impact to legal revenues. Nonetheless, prior piracy quality studies have also lacked incorporation of consumer activity, such as whether consumers download higher or lower quality copies. Furthermore, studies to date have largely not touched on how piracy quality affects both legal and illegal supply, core aspects of the piracy market that may induce omitted variable bias if not addressed (Koschmann & Bowman, 2017). This study generates contributions by investigating these three limitations in prior research, as highlighted in Table 1.

Table 1. Key Points from Prior Piracy Quality Research

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

Piracy quality draws on two streams of research: piracy and production quality. Prior research on piracy has seen rigorous academic investigation, with a focal debate on whether illegal copies help or hurt legal sales. A tension exists in piracy research, as some studies have shown negative effects of illegal copies on legal demand (Hui & Png, 2003; Bae & Choi, 2006; Yoon, 2007; Liebowitz, 2008; Waldfogel, 2012; Hong, 2013; Belleflamme & Peitz, 2014), with varying estimates on the sales displacement effect

(Hennig-Thurau et al., 2007; Danaher & Smith, 2014; Godinho de Matos et al., 2017; Aguiar & Waldfogel, 2018; Li et al., 2019; Yue, 2019; Tyrowicz et al. 2020). Yet, other piracy research has found positive effects of piracy on legal demand (Fader, 2000; Jain, 2008; Mortimer et al., 2012; Peukert et al., 2017; Zhang, 2018; Lu et al., 2020; Kretschmer & Peukert, 2020). Piracy of television dramas has shown mixed effects on TV viewership: the negative direct effect is mediated by positive online buzz, which indirectly dilutes the negative impact of digital piracy (Kim et. al., 2022). Many of these studies present a sampling versus cannibalization trade-off, with willingness to pay as a common explanation. Still other research has found no association between piracy and legal sales (Smith & Telang, 2009; Andersen & Frenz, 2010; Aguiar & Martens, 2016; Martikainen, 2014; McKenzie, 2017). The lack of consistent guidance from previous studies warrants the need to consider the effects of piracy in the detailed light of piracy quality, given the dearth of empirical evidence regarding piracy quality and its effect on legal sales.

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

In the digital age, a challenge for producers is creating exact (or seemingly exact) copies of the genuine good from an original source (such as a DVD for movies or CD/legal download for music). With music piracy, 90% of respondents perceived the conversion of a CD song to mp3 format to be as good as the original CD version (Bhattacharjee et al., 2003). By converting from a physical source to electronic, the music files become much smaller and portable (with some audio loss), resulting in an imperceptible difference to most consumers. When a person copies a movie in theaters by using a hand-held video camera to record the movie, this copy captures comments by other audience members and may have jittery video from camera movement, resulting in noticeable differences with the film and hence a lower quality copy. Copying an electronic file and not altering it will result in an exact replication, but copying from an analog source to either an electronic or other analog form will result in some loss of quality. As a closer approximation to the genuine product, the higher quality copy might be more substitutable for sales. Yet as we discuss next, this might not always be the case when product uncertainty exists.

2.2 Product Uncertainty, Information Search, and Customer Enthusiasm

For many products, especially information goods, consumers are uncertain how well a product will perform until it is actually purchased or experienced (Nelson, 1970). If uncertainty is large enough, consumers are less inclined to purchase. To alleviate uncertainty, consumers may seek out information in the market, such as movie previews or word of mouth. However, movie previews or ‘film trailers’ often depict the best scenes from a movie and are viewed as biased by consumers (Moul, 2005). Likewise, consumers are wary of early consumer product reviews (Li & Hitt, 2008), although early word of mouth can have an effect (Liu, 2006; Dellarocas et al., 2007; Gopinath et al., 2013) on reducing uncertainty.

When a product is new to market and uncertainty is high, consumers may engage in greater information search particularly as the number of product attributes increases (Moorthy et al., 1997). Since information goods are experiential and more subjective in value, product quality is harder to determine before market release (Hennig-Thurau et al., 2006). For complex or unknown products, pirated versions provide information to consumers (Peitz & Waelbroeck, 2006), which reduces product uncertainty. Thus, consumers may search the illegal market to acquire product information and reduce product uncertainty.

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

At product launch, the interested and more enthusiastic customers are likely searching for more information, and may turn to piracy for this missing information (Ma et al., 2014). Demonstrating the product by providing more information makes the product more valuable to interested consumers, increasing purchase likelihood (Yi et al., 2022). For information goods like movies, higher quality copies provide better information than low-quality piracy by giving consumers a better sense of the genuine good: better video resolution to see the actors’ facial expressions, better audio to capture ambient scene noises (or even omit distractions of the audience around them), and even trustworthiness of the pirates who uploaded the file. As higher quality copies better approximate the original good, it gives consumers a better sense of the product and alleviate uncertainties surrounding the quality of the original good. Consumers are more likely to pay for a product or service if they have prior knowledge of its value and quality (Luca, 2011). Reducing product uncertainty may facilitate group consumption: in determining a movie is worth seeing, it will lower social risk of seeing it with friends or recommending it to others. Additionally, illegal copies can work as buzz agents to increase word of mouth (Qian, 2015); higher quality copies can more closely approach the genuine good and foster more accurate word of mouth. Altogether, higher quality copies (by better approximating the genuine good) provide more information and reduce product uncertainty better than lower quality copies; as product uncertainty drops, consumer propensity to buy increases, raising revenues in the launch period.

After launch, product uncertainty diminishes as the product better permeates the market, such as through word of mouth. More positive reviews primarily correlate with higher opening revenue (e.g., Elberse & Eliashberg, 2003, Liu, 2006; Zhang & Dellarocas, 2006). An explanation for this is that the most enthusiastic customers likely purchased at product launch and shared their experiences; more information may reassure semi-interested consumers and inform other consumers to decide to purchase.

Even though there is more information in the market post-launch, less enthusiastic customers seek illegal copies less for information seeking and more for substitution; a market composed of mostly lower quality illegal copies appeals primarily to less enthusiastic customers who were less inclined to purchase anyways (Qian, 2014). Since the post-launch period comprises more of these less enthusiastic consumers, higher quality copies should serve more as substitutes for the genuine good rather than a means to reduce information uncertainty. With less enthusiastic consumers making up more of the post-launch market, higher quality downloads should have a negative effect on purchasing the legal good.

Underlying this timing difference is that enthusiastic customers seek information and serve as social agents for the launch period. The most interested consumers reinforce social intent and group behaviors, fostering desired consumer behaviors like purchasing and loyalty (Bagozzi & Dholakia, 2006). The social interest shared by enthusiastic consumers then encourages consumption of the legal good and creates a stigma from illegal consumption for group members. Additionally, illegal copies consumed on one's own or at home cannot entirely replicate the experience of the legal good, particularly when it is shared with friends as a social experience. Post-launch, however, there is less social motivation to purchase and less social stigma in consuming illegal copies. The combination of less need for reducing product uncertainty, less enthusiastic consumers, and less social pressures suggest higher quality piracy should negatively affect legitimate sales post-launch.

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. We first describe the data sources and measures, then 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; this threshold captures almost all wide release movies, which typically open on 2,000 or more theaters. This yielded 173 movies which were tracked daily until weekend revenues fell below 1% of opening/expansion revenues (i.e., the motion picture had effectively reached the end of its theatrical run). Hereafter we use launch period and opening week synonymously.

Second, the Hollywood Stock Exchange (HSX: www.hsx.com) is a prediction market that estimates opening week revenues. Online users buy and sell movie 'stocks' to reflect the estimated box office revenues for the first four weeks of wide release (opening or 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 & Eliashberg, 2003).

Third, product information comes from the Internet Movie Database (IMDB: www.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 the same time from Pirate Bay (www.piratebay.se), the most visited website for pirated content.[1] 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, keyword signals, number of user downloads (leechers), and number of users with that pirated file to share (seeders) at that time. Of the 173 movies tracked daily, 122 (70.5%) had piracy files in the opening week, and 157 (90.8%) had any piracy at all during launch or post-launch. Only 20 films (11.6%) had piracy files pre-launch (the week prior to release), similar to the 10% found by Ma et al. (2014), with a median of 4 pirated files. Pre-launch piracy was a tenth (10.6%) of the number of pirated files in the launch week, suggesting pre-launch piracy was relatively small compared to launch piracy, suggesting piracy had negligible impact prior to launch, and any persistence of these files would be reflected in the launch period.

Fifth, advertising costs for each film come from Kantar Media's Ad$pender. The advertising expenses encompassed the twelve months leading up to and including the first week of release. On average, only 3.7% of a film's advertising expense came after the opening week.

Finally, the sixth data source is actor/actress star power from the 2009 Forbes Star Power Index, the most recent survey available prior to data collection. The index surveys Hollywood executives,

[1] As piracy is global, hourly downloads for a separate sample found downloads were close to uniform, which may be unsurprising since the countries with the most piracy (U.S., China, Russia) span 20 time zones (Go-Globe 2020).

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

Ex ante, the legal supplier decides how much product to supply (i.e., movie theaters decide screen allocations for a film) just prior to launch. 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 available pirated copies of a given film on Pirate Bay, across number of unique piracy files and number of users with that pirated copy, averaged for that week. Pirate Bay facilitates file sharing through BitTorrent protocol; rather than downloading one large file from a single source, a user may download the same file in pieces from multiple users at the same time ('swarming'). This creates a network effect where having few files with many users is more valuable than having many files with few users (Qiu & Srikant, 2004), reducing download times and risk of an incomplete download. Additionally, number of seeders has been emphasized over number of pirated files available because of this effect (Koschmann & Bowman, 2017). Since piracy can occur before product launch, we account for this with number of days the film was released in another major market before the U.S./Canada *(Previous_Days)*. Illegal demand *(Leechers)* reflects observed incidence of illegal behavior as downloads of pirated copies across all seeded versions, consistent with prior piracy research (e.g. Oberholzer-Gee & Strumpf, 2007; Danaher et al., 2010).

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 audience demand. We estimate expected demand from the HSX prediction market. After the launch period, theaters can adjust screen allocation based on prior weeks' performance; week 2 is estimated with an industry average 30% drop-off in opening week revenues, while weeks 3 and onward use a double exponential smoothing model (i.e., Holt-Winters forecasting method). Since revenue decay is curved rather than linear, one parameter smooths and another parameter accounts for the trend, giving more weight to more recent weeks, as done in prior movie research (Elberse & Eliashberg, 2003; Koschmann & Bowman, 2017).

Additional control variables used in movie research are included. Time of year seasonality (*Seasonality*) can affect both legal and illegal motion picture demand (Vogel, 2015), particularly in the summer or during holidays. Production budget *(Prod_Budget),* film critic ratings *(Critics),* and actor star power *(Actor_Power)* speak to product quality while advertising costs (*Advertising*) pertain to promotion. Release by a major studio (*Major_Studio*) can influence distribution. Consumer sentiment as word of mouth reflects online user ratings for both valence (*WOM*) and volume in online raters (*Num_Users*) (Liu, 2006, Chintagunta et al., 2010, Zhu & Zhang, 2010). Competition is accounted for both in legal supply and demand with screen competition from other new releases (*Screen_Comp_New*) and existing releases (*Screen_Comp_Ong*), as well as competition for revenues (*Revenue_Comp*) from other movies. Appendix A1 further explains the variable operationalizations and the expected revenue estimation.

3.3 Assessing Piracy Quality through Observed Signals

Of focal interest is piracy quality (Quality). An issue with defining quality is the subjective nature of the construct. Despite this challenge, pirated copies convey signals that demonstrate higher or lower quality. Piracy files that exhibit greater quality signals should close the production quality gap to the genuine 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 experts can assess these signals, a concern is that expert opinions may differ. We measure piracy quality from observable signals, using both objective quality and perceived quality (Monroe & Krishnan, 1985). In the Pirate Bay data, the illegal copies present features that meaningfully suggest quality to consumers, such as 'CAM' for copies captured with a handheld camera in the theater or 'DD2.0' for two channel Dolby Digital surround sound audio. A screenshot example with further keyword elaboration appears in Appendix A2, and the individual piracy keywords altogether suggest overall quality in a piracy file.

Quality is often treated as a higher order global assessment (Olshavsky, 1985; Holbrook & Corfman, 1985). At this higher level, quality is a composite of elements consumers perceive, such as price and product attributes (Zeithaml, 1988), as different product features should speak to underlying quality. This is not unlike the construction of product quality from total quality management surveys (Garvin 1987), where estimates of each piracy keyword map onto that quality scale. Here, the presence of piracy keywords should jointly manifest as a latent, continuous spectrum of quality. Because quality is a latent continuous measure, estimating latent continuous measures are often estimated by either factor analysis or item response theory (IRT) models. We utilize the IRT model, which is more appropriate for unidimensional ideal point estimation (e.g., Van Schuur & Kiers, 1994; Spector et al., 1997) and preferred to factor analysis when using categorical data (Bartholomew et al., 2002). IRT models uncover latent relationships by inferring from the observed data (see Lord,2012; De Jong, et al., 2007). Instead of including all the keywords and observed characteristics of the pirated movies in a regression, which could introduce noise and make the regression surface high-dimensional and unstable, the IRT method summarizes the information presented by the copies' characteristics to latent quality. This contrasts with a traditional multi-attribute model (e.g., Fishbein, 1963), which requires either *a priori* knowledge of the weight of a given attribute or the weighted outcome, neither of which is known here. The IRT model and keyword ideal point estimates, an analog to loading scores in factor analysis, appear in Appendix A3.

3.4 Model-Free Evidence: Descriptive Statistics

Across 12,710 unique piracy files, average quality is positive (*Mean* = 4.365, *SD* = 1.425, *Median* = 4.732), as the quality ranges from -2.066 to 7.588. Negative piracy quality means the pirated file exhibited more lower quality signals than higher quality signals in its file description. These averages for each file were then scaled to be non-negative.

Table 2 presents descriptive statistics of the variables. Since motion pictures release weekly, we average daily piracy measures across all files for a given film to get weekly figures. A total of 249,440 film-day-file observations were collected. Weekly piracy quality for each film is averaged daily weight of each piracy file for that film. For instance, a higher quality piracy file that appears on Friday (the start of a film week) and remains on PirateBay each day through Thursday (the end of a film week) will have seven days of data; a low-quality piracy file that appears only Wednesday and stays on through the next day will only have two days of data. In this case, the higher quality copy has more weight than the lower quality copy. Average piracy quality for a given film in the opening week is *Mean* = 4.31 (*SD* = 3.17). Table 2 also shows that average piracy quality increased to 5.19 (SD = 5.90) post-launch, which may be expected that over time higher quality copies should permeate the market. Most films had high quality copies at launch, with more higher quality copies arising post-launch. In total, 90.8% of the films in the sample had illegal copies during the theatrical run (i.e., 9% of the movies had no piracy on Pirate Bay). Although the piracy data is collected globally, the correlation of global revenues with U.S./Canada revenues is $r = 0.92$, suggesting global revenues may be similarly impacted by piracy.

Table 2. Summary Statistics by Product Period

Launch Period ($N = 173$)

	Mean	Median	SD	Min	Max
Screens	3,609.62	3,200.00	2,723.37	210.00	12,600.00
Revenue	$26,509.79	$14,366.97	$33,143.10	$289.61	$222,116.06
Seeders	215.47	9.75	329.36	0.00	1,795.53
Leechers	147.44	29.50	237.83	0.00	1,468.11
Quality	4.31	5.72	2.90	0.00	8.26
Prod_Budget	$47,688.48	$28,000.00	$52,273.61	$1,000.00	$255,000.00
Advertising	$13,274.80	$12,355.85	$9,692.61	$0.32	$37,901.70
Actor_Power	5.77	6.53	2.87	0.00	10.00
Critics	50.70	49.57	16.75	13.57	97.00
Previous_Days	7.25	2.00	19.24	0.00	223.00
Major_Studio	0.57	1.00	0.50	0.00	1.00
Revenue_Comp	3.38	3.17	1.50	0.30	9.32
Screen_Comp_New	10.34	8.80	7.95	0.00	41.00

Screen_Comp_Ong	5.61	5.56	0.94	3.60	8.20
Seasonality	0.98	0.90	0.30	0.56	1.82
WOM	6.70	6.80	1.25	1.36	8.90
Num_Users	6,794.74	2,016.79	11,889.72	57.00	61,343.29

Post-Launch Period (*N* = 1,204)

	Mean	Median	SD	Min	Max
Screens	1,542.84	775.00	1,806.48	5.00	11,500.00
Revenue	$4,694.45	$1,263.80	$8,877.06	$4.43	$87,548.90
Seeders	261.29	225.63	252.35	0.00	3,124.90
Leechers	87.69	55.43	123.46	0.00	1,480.50
Quality	5.19	5.90	2.17	0.00	9.65
Revenue_Comp	3.71	3.49	2.20	0.11	56.00
Screen_Comp_New	14.23	13.70	8.51	0.50	41.00
Screen_Comp_Ong	5.57	5.36	1.03	3.60	8.20
Seasonality	0.97	0.90	0.28	0.56	1.82
WOM	6.89	7.00	1.18	1.46	8.90
Num_Users	30,258.76	9,479.29	46,009.72	107.86	297,047.71

Notes. Dollars are in thousands (000).

3.5 Launch Model

To estimate the effect of piracy quality on the market, we model an interdependent system of equations with legal supply and demand plus illegal supply and demand. For many products, especially information goods like motion pictures, the launch period differs from the post-launch period. The movie industry emphasizes a large opening week, often constituting one third of total ticket sales for a film's run (Eller & Friedman, 2008). As such, we separate the two time periods – launch and post-launch – in a multiplicative framework model that log-transforms variables like prior movie research (Elberse & Eliashberg, 2003; Somlo et al., 2011; Clement et al., 2014; Koschmann & Bowman, 2017),[2] where the supply and demand sides of the market are modeled differently. The launch period system of equations is:

$$\ln(\text{Screens}_{i1}) = \alpha_0 + \alpha_1\ln(\text{Revenue\_Est}_{i1}) + \alpha_2\ln(\text{Prod\_Budget}_{i1}) + \alpha_3\ln(\text{Actor\_Power}_{i1}) + \alpha_4\ln(\text{Advertising}_{i1}) + \alpha_5\ln(\text{Critics}_{i1}) + \alpha_6\text{Major\_Studio}_{i1} + \alpha_7\ln(\text{Screen\_Comp\_New}_{i1}) + \alpha_8\ln(\text{Screen\_Comp\_Ong}_{i1}) + \alpha_9\ln(\text{Previous\_Days}_{i1}) + \varepsilon_{Si1} \quad (1)$$

$$\ln(\text{Revenue}_{i1}) = \beta_0 + \beta_1\ln(\text{Screens}_{i1}) + \beta_2\ln(\text{Prod\_Budget}_{i1}) + \beta_3\ln(\text{Actor\_Power}_{i1}) + \beta_4\ln(\text{Advertising}_{i1}) + \beta_5\ln(\text{Critics}_{i1}) + \beta_6\text{Major\_Studio}_{i1} + \beta_7\ln(\text{Revenue\_Comp}_{i1}) + \beta_8\text{Seasonality}_{i1} + \beta_9\ln(\text{WOM}_{i1}) + \beta_{10}\ln(\text{Num\_Users}_{i1}) + \beta_{11}\ln(\text{Quality}_{i1}) + \beta_{12}\ln(\text{Leechers}_{i1}) + \beta_{13}\ln(\text{Quality}_{i1})*\ln(\text{Leechers}_{i1}) + \varepsilon_{Ri1} \quad (2)$$

$$\ln(\text{Seeders}_{i1}) = \gamma_0 + \gamma_1\ln(\text{Screens}_{i1}) + \gamma_2\ln(\text{Prod\_Budget}_{i1}) + \gamma_3\ln(\text{Actor\_Power}_{i1}) + \gamma_4\ln(\text{Advertising}_{i1}) + \gamma_5\ln(\text{Critics}_{i1}) + \gamma_6\text{Major\_Studio}_{i1} + \gamma_7\ln(\text{Previous\_Days}_{i1}) + \gamma_8\ln(\text{WOM}_{i1}) + \gamma_9\ln(\text{Num\_Users}_{i1}) + \gamma_{10}\ln(\text{Quality}_{i1}) + \gamma_{11}\ln(\text{Leechers}_{i1}) + \gamma_{12}\ln(\text{Quality}_{i1})*\ln(\text{Leechers}_{i1}) + \varepsilon_{Pi1} \quad (3)$$

[2] We also conducted a panel vector autoregressive (PVAR) model pooling all periods together (Appendix A4).

$$\ln(Leechers_{i1}) = \lambda_0 + \lambda_1\ln(Revenue_{i1}) + \lambda_2\ln(Prod\_Budget_{i1}) + \lambda_3\ln(Actor\_Power_{i1}) + \lambda_4\ln(Advertising_{i1}) + \lambda_5\ln(Critics_{i1}) + \lambda_6 Major\_Studio_{i1} + \lambda_7 Seasonality_{i1} + \lambda_8\ln(WOM_{i1}) + \lambda_9\ln(Num\_Users_{i1}) + \lambda_{10}\ln(Quality_{i1}) + \lambda_{11}\ln(Seeders_{i1}) + \lambda_{12}\ln(Quality_{i1})*\ln(Seeders_{i1}) + \varepsilon_{Li1} \quad (4)$$

The system of equations treats legal supply as the starting point: informational interviews with theater managers suggest that illegal copies enter the market after the legal product has launched.[3] For motion pictures, the starting point is that theaters allocate screens in advance of a film's release in order to arrange show times to meet expected demand. Subscript *i* denotes the film and *t* for the launch period (here week $t = 1$). The error term of each equation, ε, is additionally subscripted S, R, P, L to denote the screens, revenue, seeders, and leechers equations, respectively. Equations (1)-(4) use typical motion picture control variables that suggest a film's qualitative belief to consumers: production budget, star power, advertising, critic ratings, and an indicator for release by a major studio. The revenue equation also accounts for competing films' genre and MPAA rating (e.g., Elberse & Eliashberg, 2003), and the speed of word of mouth that week; i.e., the 'Bruno' effect where consumer reviews on Friday can quickly affect demand for the rest of the opening weekend and week (Wasow et al., 2010). Additionally, word of mouth includes not only the valence (consumer sentiment) but also the volume in number of consumers talking about a given film (e.g., You et al., 2015), which may affect demand and supply. Release by a major studio (binary coded) and seasonality (average week relative percentage) are not log-transformed.

3.6 Post-Launch Model

The post-launch system of Equations (5)-(8) is similar to Equations (1)-(4), where $t > 1$, and Greek uppercase letters distinguish post-launch coefficients from the launch period:

$$\ln(Screens_{it}) = A_0 + A_1\ln(Revenue\_Est_{it}) + A_2\ln(Screen\_Comp\_New_{it}) + A_3\ln(Screen\_Comp\_Ong_{it}) + A_4\ln(WOM_{it}) + A_5\ln(Num\_Users_{it}) + A_6\ln(Quality_{it-1}) + A_7\ln(Seeders_{it-1}) + A_8\ln(Quality_{it-1})*\ln(Seeders_{it-1}) + A_S D_{Sit} + \varepsilon_{Sit} \quad (5)$$

$$\ln(Revenue_{it}) = B_0 + B_1\ln(Screens_{it}) + B_2\ln(Revenue\_Comp_{it}) + B_3 Seasonality_{it} + B_4\ln(WOM_{it}) + B_5\ln(Num\_Users_{it}) + B_6\ln(Quality_{it}) + B_7\ln(Leechers_{it}) + B_8\ln(Quality_{it})*\ln(Leechers_{it}) + B_R D_{Rit} + \varepsilon_{Rit} \quad (6)$$

$$\ln(Seeders_{it}) = \Gamma_0 + \Gamma_1\ln(Screens_{it}) + \Gamma_2\ln(WOM_{it}) + \Gamma_3\ln(Num\_Users_{it}) + \Gamma_4\ln(Quality_{it}) + \Gamma_5\ln(Leechers_{it}) + \Gamma_6\ln(Quality_{it})*\ln(Leechers_{it}) + \Gamma_P D_{Pit} + \varepsilon_{Pit} \quad (7)$$

$$\ln(Leechers_{it}) = \Lambda_0 + \Lambda_1\ln(Revenue_{it}) + \Lambda_2 Seasonality_{it} + \Lambda_3\ln(WOM_{it}) + \Lambda_4\ln(Num\_Users_{it}) + \Lambda_5\ln(Quality_{it}) + \Lambda_6\ln(Seeders_{it}) + \Lambda_7\ln(Quality_{it})*\ln(Seeders_{it}) + \Lambda_L D_{Lit} + \varepsilon_{Lit} \quad (8)$$

---

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

The post-launch period excludes time-invariant variables; for instance, critic ratings do not change after release and we find an average of 96.3% of advertising for a given film is spent leading up to and into the launch period. Time dummies (*D*) are added to account for time-specific fixed effects (i.e., for how long the film has been in theaters). The coefficients on *D* in each equation (A_S, B_R, Γ_P, Λ_L) are a vector of estimates for each week. Legal supply is again the initial starting point for the week, theater owners see revenues in the prior week and adjust screen allocations. Thus, Equation (5) includes not only anticipated revenue based on the prior week, but seeders and piracy quality from the prior week, as previous research finds that supply follows demand (Krider et al., 2005), legal or illegal. As legal supply sets the stage for the current week, the revenue, seeders, and leechers equations use contemporaneous rather than lagged observations given the speed of digital piracy (Koschmann & Bowman, 2017), and WOM is not lagged as it may affect demand in the current week (Wasow et al., 2010).[4]

3.7 Correction for Endogeneity

A modeling concern is whether the dependent measures, as regressors, may be correlated with the error terms. In movie research, key endogenous variables like screens and revenue may be correlated with the error term through an award nomination (Elberse & Eliashberg, 2003). A studio might counter lower than expected piracy by seeking an increase in screen allocation. Although studios typically want more screen availability to increase distribution for consumers, more showings also increase opportunities for in-theater piracy. Quality may also be correlated with the error terms; more higher quality copies might entice more consumers to download, which affects pirates' incentives to create and share copies.

To address endogeneity concerns, we model the correlation between the error terms and potentially endogenous regressors (screens, revenues, seeders, leechers, and quality) using Gaussian copulas. Like instrumental variables, copulas "absorb the correlation between potentially endogenous marketing-mix variables and the normally distributed error term" (Datta et al. 2022, p.259). By partitioning the endogenous from the exogenous components, the copula model enables researchers to construct a flexible multivariate joint distribution that captures the correlation between the endogenous regressor and the error term, gaining popularity in marketing applications. Following prior work (Park & Gupta, 2012; Papies et al., 2017), we generate copula-transformed terms:

$$\ln\left(\widetilde{\mathrm{Screens}_{\imath t}}\right) = \Phi^{-1}\left[\mathrm{H}_{\ln(\mathrm{Screens})}\ln(\mathrm{Screens})\right] \quad (9)$$

[4] Elberse & Eliashberg (2003) used lagged revenue per screen as word of mouth, in a time where there was no social media and instead office 'cooler talk' after the weekend which might affect film demand next weekend.

$$\ln\left(\widetilde{\text{Revenue}}_{\text{ıt}}\right) = \Phi^{-1}\left[\text{H}_{\ln(\text{Revenue})}\ln(\text{Revenue})\right] \quad (10)$$

$$\ln\left(\widetilde{\text{Seeders}}_{\text{ıt}}\right) = \Phi^{-1}\left[\text{H}_{\ln(\text{Seeders})}\ln(\text{Seeders})\right] \quad (11)$$

$$\ln\left(\widetilde{\text{Leechers}}_{\text{ıt}}\right) = \Phi^{-1}\left[\text{H}_{\ln(\text{Leechers})}\ln(\text{Leechers})\right] \quad (12)$$

$$\ln\left(\widetilde{\text{Quality}}_{\text{ıt}}\right) = \Phi^{-1}\left[\text{H}_{\ln(\text{Quality})}\ln(\text{Quality})\right] \quad (13)$$

where Φ^{-1} is the inverse normal cumulative distribution function and H(•) is the empirical cumulative distribution functions of the log-transformed terms screens, revenues, seeders, leechers, and quality, respectively.[5] In line with prior research, copulas are used only for the endogenous variables in a given equation, and not created for the interactions of quality with seeders, and quality with leechers, as these terms can bias the results (Qian et al., 2022).[6]

To identify the model, the errors are presumed normal and the endogenous regressor must be non-normal in its distribution (Park & Gupta, 2012; Rutz & Watson, 2019). A Shapiro-Wilk test of the opening week log-transformed screens (W = 0.921), revenues (W = 0.973), seeders (W = 0.789), leechers (W = 0.832), and quality (W = 0.667) are each not normally distributed (each p < .01). Similarly for the post-launch period, these endogenous regressors are again not normally distributed (each p < .01). After the copula adjustment, each equation becomes free of endogeneity.[7] Copulas can also be applied to a system of equations just the same as single equation regressions (Trivedi & Zimmer, 2006; Pastpipatkul et al., 2016a; Pastpipatkul et al., 2016b).

4. Empirical Results

Estimation of both launch and post-launch systems of equations utilizes seemingly unrelated regression (SUR), allowing the error terms of the equations to correlate for efficiency (Zellner & Theil, 1962). The error terms may be correlated across equations for other exogenous factors that could "shock" both the legal and illegal sides of the market (e.g., award nominations). Estimation includes movies with no piracy (setting leechers, seeders, and piracy quality to zero), which made up a small part of the sample.

4.1 Launch Estimation Results

Table 3 reports the SUR model estimates for the launch period system of Equations (1)-(4). The system weighted R^2 = 0.978, indicating high fit among the four interdependent parts of the market. Here as well,

[5] We separately test word of mouth for endogeneity (Appendix A5), finding it is not endogenous.
[6] If the copula is correlated with the endogenous regressors, the 2sCOPE approach should be used over Park & Gupta (Yang et al., 2022). Here, the focal piracy variables are not particularly correlated with the endogenous regressors, so we use the Park & Gupta copula approach here.
[7] We examine whether there may be reverse causality, i.e., if screens, revenues, seeders, or leechers drive piracy quality. A Granger causality test showed none of these four variables Granger causes piracy quality.

both sides of the equation are log-transformed, so the coefficients are elasticities, and a visual inspection of the residuals showed an approximately normal distributions of the error terms.

Although we focus on the effects of piracy quality, the control variables are consistent with those reported in existing motion picture research. Notably, in the screens' equation, anticipated revenues (α_1 = 0.498, p < .01), advertising expense (α_4 = 0.145, p < .01), and film critic reviews (α_5 = -0.382, p < .01) are significant and in the same direction as those found elsewhere (Elberse & Eliashberg, 2003; Clement et al., 2014; Koschmann & Bowman, 2017). Star power of actors (α_3 = -0.027, p < .01) is negative for screen allocation, but one reason for this may be that leading actors sometimes choose smaller budget, independent films for increased creative expression and control over the project (Casting Networks, 2023), which may get fewer screens.

Table 3. Launch Period SUR Estimation Results with Quality Interactions

	DV:ln(Screens)		DV:ln(Revenue)		DV:ln(Seeders)		DV:ln(Leechers)	
Variable	Estimate		Estimate		Estimate		Estimate	
Intercept	0.913	***	-2.021	***	2.709	***	-0.691	
	(0.134)		(0.622)		(0.793)		(0.623)	
ln(Revenue)[a]	0.498	***					0.247	***
	(0.016)						(0.060)	
ln(Screens)			1.458	***	-0.537	***		
			(0.076)		(0.093)			
ln(Prod_Budget)	0.170	***	-0.180	***	0.110	***	-0.066	**
	0.014		(0.028)		(0.042)		(0.028)	
ln(Actor_Power)	-0.027	***	0.023		0.053	**	-0.035	*
	(0.010)		(0.017)		(0.027)		(0.019)	
ln(Advertising)	0.145	***	-0.035	*	0.057	**	-0.051	***
	(0.010)		(0.021)		(0.029)		(0.018)	
ln(Critics)	-0.382	***	0.469	***	0.305	***	-0.255	***
	(0.032)		(0.070)		(0.109)		(0.074)	
Major_Studio	0.004		0.202	***	0.084		-0.111	**
	(0.024)		(0.042)		(0.064)		(0.045)	
ln(Screen_Comp_New)	-0.098	***						
	(0.012)							
ln(Screen_Comp_Ong)	-0.012							
	(0.058)							
ln(Previous_Days)	-0.039	***			0.047	**		
	(0.009)				(0.019)			
ln(Revenue_Comp)			-0.153	***				
			0.033					
Seasonality			0.543	***			-0.083	*
			(0.061)				(0.046)	
ln(WOM)			0.269	***	-0.790	***	0.602	***
			(0.101)		(0.164)		(0.116)	
ln(Num_Users)			0.072	***	0.091	***	-0.095	***
			(0.019)		(0.031)		(0.021)	
ln(Quality)			-0.155	*	-0.827	***	0.629	***

		(0.089)		(0.137)		(0.092)	
ln(Seeders)						0.530	***
						(0.076)	
ln(Leechers)		-0.164	*	1.609	***		
		(0.084)		(0.123)			
ln(Quality)*ln(Seeders)						0.047	
						(0.039)	
ln(Quality)*ln(Leechers)		0.143	***	-0.121	*		
		(0.042)		0.062			
ln(Revenue) copula						-0.095	
						(0.084)	
ln(Screens) copula		-0.163	**	0.280	***		
		(0.064)		(0.080)			
ln(Quality) copula		0.075		0.287	***	-0.196	***
		(0.048)		(0.075)		(0.051)	
ln(Seeders) copula						0.354	***
						(0.026)	
ln(Leechers) copula		-0.057		-0.323	**		
		(0.042)		(0.050)			
System Weighted R^2	0.978						

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

In the revenues' equation, screens ($\beta_1 = 1.458$, $p < .01$) and film critic reviews ($\beta_5 = 0.469$, $p < .01$) are significant and in the same direction as the previously mentioned studies. Regarding piracy, revenues are negatively affected by illegal demand ($\beta_{12} = -0.164$, $p < .06$). The first-order effect of piracy quality alone is not particularly interpretable, because the mere presence of piracy quality is not of interest, but only when it manifests into consumer activity (as downloads or uploads of high-quality copies) that is important. As more high-quality copies are downloaded, there is a positive effect on revenues ($\beta_{13} = 0.143$, $p < .01$).[8] This positive effect aligns with our expectation that consumption of higher quality illegal copies in the launch period works as a sampling mechanism to alleviate product uncertainty. Thus, while downloads of higher quality copies provides an average positive effect on revenues, too many downloads in general have an average negative effect.

Although piracy quality influences revenues, also of interest is its effect on the illegal side of the market. In the seeders' equation, several control variables are worth noting: fewer screens corresponded to more piracy supply ($\gamma_1 = -0.537$, $p < .01$), while production budget, actor power, advertising, and critic rating all exhibited positively significant effects on seeding. Interestingly, WOM had a negative effect ($\gamma_8 = -0.790$, $p < .01$); one possibility here is that strong word of mouth may inhibit piracy from cinema attendance, as strong (weak) word of mouth would lead to more (few) attendees. Low or zero attended

[8] A separate estimation of main effects only without the quality interactions (Appendix A6) yields similar results here except with leechers in the revenues equation, which is positive. We attribute this difference in the need to explore and disentangle moderating effects of quality, warranting the interaction term.

screenings yield less chance of being caught making copies. The main effect of downloading (γ_{11} = 1.609, $p < .01$) exhibits a significant relationship, yet downloading higher quality copies has a marginally negative effect (γ_{12} = -0.121, $p < .06$). This indicates that demand for higher quality copies does not spur piracy sharing. In theory, consumers would seek out higher quality copies, which would incentivize pirates to add more copies. We find no evidence of this in the opening week. One explanation for this may be that high-quality copies are limited at first, as the product just recently entered the market.

In the leechers equation, an increase in revenues corresponded to an increase in downloads (λ_1 = 0.247, $p < .01$), reinforcing the legal demand side of the market. Yet, key movie control variables had negative correspondence on downloading: production budget, actor power, advertising, and critic rating. One plausibility is that films that are more 'mediocre' and perhaps not worth the regular price of admission may be more pirated because they lack strong market signals (i.e., big budgets, big name actors, and critic adoration suggest product quality). This buttresses our view that when information goods lack strong market signals, pirated copies – and the quality of those copies – play a role in informing consumers. The main effects of available piracy (λ_{11} = 0.530, $p < .01$) and desire for piracy quality (λ_{10} = 0.629, $p < .01$) are significant. Yet, the interaction term for the presence of high-quality copies (λ_{12} = 0.47, $p > .22$) is not significant. A possible answer for this is that, like in the seeders' equation, high-quality copies may be hard to come by initially given the recent market introduction, or induce skepticism of its proposed high-quality.

4.2 Post-Launch Estimation Results

Like the launch period, the post-launch system of Equations (5)-(8) uses legal supply as the starting point. The SUR post-launch estimates (Table 4) show a similarly high fit (system weighted R^2 = 0.983). A visual inspection of the error terms showed an approximately normal distribution. Whereas piracy presumably had little effect on screen allocation prior to a film's release, piracy effects from the prior week might influence screen allocation in post-launch weeks. Prior week piracy availability (A_7 = 0.499, $p < .01$) exhibited a positive effect on legal supply, although the supply of higher quality illegal copies (A_8 = -0.210, $p < .01$) had a negative effect on screen allocation. As such, greater supply of illegal copies appears to act like a signal for continued demand, suggesting complementarity for legal supply; however, greater availability of higher quality copies negated this effect as a substitute. One explanation for these results is that theater owners are not overly concerned about piracy supply, including high-quality piracy: piracy effects should manifest in demand for tickets, which is of interest to theater owners. Increased demand for illegal copies in the prior week may help screen allocation through demand as sampling, but the trade-off with higher quality downloads is that these might substitute sales through expected revenues.

Table 4. Post-Launch Period SUR Estimation Results with Quality Interactions

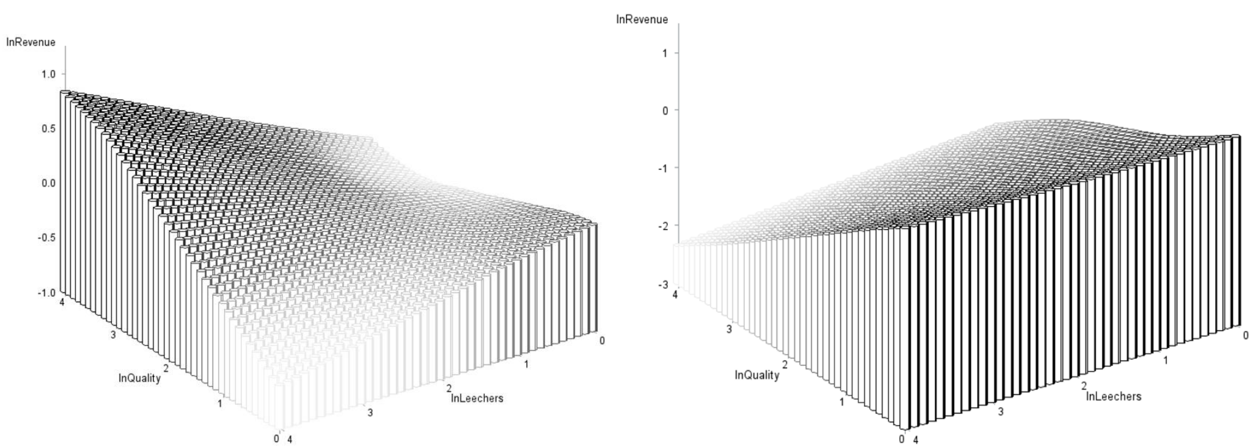
Variable	DV:ln(Screens) Estimate		DV:ln(Revenue) Estimate		DV:ln(Seeders) Estimate		DV:ln(Leechers) Estimate	
Intercept	-0.909	*	-0.832	**	-1.378	***	0.971	**
	(0.257)		(0.385)		(0.472)		(0.437)	
ln(Revenue)[a]	0.526	***					-0.114	***
	(0.009)						(0.040)	
ln(Screens)			0.936	***	0.043			
			(0.035)		(0.035)			
ln(Screen_Comp_New)	-0.077	***						
	(0.018)							
ln(Screen_Comp_Ong)	0.357	***						
	(0.068)							
ln(Revenue_Comp)			0.022	*				
			(0.013)					
Seasonality			0.540	***			-0.016	
			(0.025)				(0.017)	
ln(WOM)[b]	-0.044		0.736	***	-0.287	***	0.195	***
	(0.081)		(0.045)		(0.061)		(0.051)	
ln(Num_Users)[b]	0.151	***	0.013	*	0.030	***	-0.024	***
	(0.013)		(0.008)		(0.010)		(0.008)	
ln(Quality)[b]	0.144	***	-0.374	***	0.607	***	-0.325	***
	(0.030)		(0.049)		(0.060)		(0.005)	
ln(Seeders)[b]	0.499	***					0.814	***
	(0.045)						(0.052)	
ln(Leechers)			0.074		1.341	***		
			(0.076)		(0.085)			
ln(Quality)*ln(Seeders)[b]	-0.210	***					-0.070	***
	(0.023)						(0.026)	
ln(Quality)*ln(Leechers)			-0.075	**	0.107	**		
			(0.035)		(0.042)			
ln(Revenue) copula							0.246	***
							(0.079)	
ln(Screens) copula			0.443	***	-0.085			
			(0.055)		(0.054)			
ln(Quality) copula			0.155	***	-0.230	***	0.168	***
			(0.025)		(0.030)		(0.025)	
ln(Seeders) copula							0.179	***
							(0.012)	
ln(Leechers) copula			0.137	***	-0.451	***		
			(0.028)		(0.027)			
System Weighted R^2	0.983							

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

On the revenue side, downloads overall did not have a significant effect ($B_7 = 0.074$, $p > .32$), but high-quality downloads weaken revenues ($B_8 = -0.075$, $p < .04$). Underlying this is that an increase in piracy quality exhibited a greater negative effect ($B_6 = -0.374$, $p < .01$). Unlike the launch period, post-launch consumption of higher quality copies exhibits a negative effect on revenues; separating the two time periods may explain why prior research found no overall significant effect of piracy quality (Ma et al., 2014, Lu et al., 2020). While downloads generally do not influence revenues post-launch, increasing the quality of pirated copies magnifies the problem for revenues. This aligns with our belief that higher quality copies act like substitutes post-launch, as there is less need for uncertainty reduction. Although prior studies using piracy quality found no significant effect on revenues (Ma et al., 2014; Danaher et al., 2019; Lu et al., 2020), or a negative effect from high quality piracy (Ackermann et al., 2020), we attribute the findings here to two factors. One, we distinguish between the launch and post-launch periods, rather than pooling the two, with our belief that product uncertainty in the launch period affects piracy consumption differently than post-launch. Second, the effect of piracy quality is likely more complex in its measurement and impact on revenues that might not be captured by a high versus low dichotomy.

These results pose a challenge to studios: encourage some beneficial piracy (high-quality downloads) against the detriment of too many downloads overall. Figure 1a (left graph) shows the impact of quality and leechers on revenue using the first-order and interaction term results from Table 3: when piracy downloads and piracy quality are both high or both low, revenues are higher (far left and far right). When one of piracy downloads or piracy quality is higher, revenues are lowest (the ‘white’ bars in corners closest and furthest away from the reader). Although consumers who wish to see the movie in theaters may turn to high-quality piracy copies if showtimes are sold out, interviews with theater managers note an excess of show times beyond expected demand in the first week to avoid missed consumers.

Figure 1. Effect of Leechers and Quality on Revenues in Launch (a) and Post-Launch (b)



(a) (b)

Figure 1b (right graph) highlights the post-launch effect of quality and leechers on revenue. Unlike the opening week, the effect of piracy quality on revenue is greatest here when both leechers and piracy quality are low (darker bars closest to the reader and further right in Figure 1b), and most negative on revenue, where both leechers and piracy quality are high (‘white’ bars to the far left).

In the seeders' equation, piracy downloads ($\Gamma_5 = 1.341$, $p < .01$) encourage illegal supply, and downloads of higher quality copies ($\Gamma_6 = 0.107$, $p < .02$) also affects seeding. Given the file sharing linkage between seeders and leechers, we expected demand for higher quality copies to facilitate more piracy copies. Even after launch, consumers should still have some continued information seeking, as well as demand for higher quality illegal copies. While there might be less 'reward' (i.e., street credibility) for a pirate to create high-quality copies if the illegal market has other high-quality copies available, consumer demand feeds the supply.

Demand for illegal copies post-launch is positively affected by seeders ($\Lambda_6 = 0.814$, $p < .01$), but the availability of higher quality copies ($\Lambda_7 = -0.070$, $p < .01$) had a negative effect on downloading. This indicates that post-launch, consumers looking for illegal copies are more inclined to download lower quality copies. Although there should be more high-quality copies available after launch (and consumers would want higher quality copies) this might be explained through information seeking. In the opening week, consumers sought high-quality copies for more information about the legal good. Post-launch, there is more information available, but if post-launch consumers are not information seeking, they might accept any copy to supplant willingness to pay. The number of pirated copies available post-launch should increase, making search costs for an ideal copy harder. Also, lower quality copies are likely smaller in file size, making downloading faster. Still another possibility is that demand for a film (both legal and illegal) naturally declines over time.

An additional consideration is whether piracy quality evolves over time. The theoretical belief is that higher quality versions should appear over time, suggesting piracy quality increases monotonically. For an unbalanced panel, a Fisher-type panel unit root test (e.g., Choi, 2001) of piracy quality finds the panels are stationary (four test measures, P, Z, L^*, P_m are each significant at $p < .01$). As such, piracy quality is relatively stable during a movie's theatrical run after the launch period, and is unlikely to explain the effects alone in the equations. The results echo past findings on the significant effects of WOM (WOM contributes $68.7 million for a 40-week period to the box office revenue: Liu, 2019). We further show WOM has a larger impact on revenues post-launch as compared to the opening week.

4.3 Additional Moderation Considerations

As developed in Section 2.2, we motivated and discussed the mechanism of customer enthusiasm. This gives rise to potential sampling effects of piracy for the original movies. With this mechanism, we conjecture that the consumer enthusiasm may be more important for certain movies than others, especially in the launch period when information is not yet well disseminated. We stratify the sample in the launch period to estimate the effects on revenues where consumers may use other cues: releases by major studios, production budget, star power of the actors, and advertising expense. To do so, we interact these key variables piracy downloads and quality to see if any of these variables show differential piracy effects on revenues. The results appear in Appendix A7. While the key findings from Table 3 do not change in significance, we find no differential effect on revenues from higher quality downloads among major studio releases. In the other three variables, moderation between downloads and production budget, star power, and advertising budget were each positively associated with greater revenues; however, higher quality downloads on each of these three variables was significantly negative. This presents a trade-off, where some sampling helps such films' revenues, but too high a quality in the illegal market hinders their financial performance. This serves as an effective test of the moderation effects of consumer enthusiasm and further corroborates the theory we propose. Finally, a simulation of the findings highlights this interplay between piracy quality and consumer activity of downloading (Appendix A8).

4.4 Robustness Checks

In addition to the reported results, several supplemental analyses and robustness checks were carried out. First, a mediation test (Appendix A9) was conducted to gauge whether all endogenous variables should belong in each equation. For example, revenues might be affected by seeders through leechers, rather than directly. The results indicate mediation exists, and some endogenous variables likely do not have direct effects on other endogenous variables (e.g., the effect of seeders on revenue in the launch period happens through leeching, so seeders are excluded in the revenues equation).

Second, a correlation analysis (Appendix A10) of total downloads and movie traits explored whether leeching could be driven by certain movie features (e.g., action films, actor star power, or those rated 'PG-13'). No correlation exceeded $r > .50$, suggesting total downloads are not driven by a particular movie trait.

Third, since the IRT model might include a stochastic component, quality was regressed on the piracy keywords, where the resulting estimates (Appendix A11) and observed values could be substituted into the quality term in Equations (1)-(8). This was run with both comments as a binary variable and number of comments across pirated files. With $R^2 = 0.999$, inserting these values into the model suggests any stochastic effect would be negligible to the reported results. Furthermore, the predicted quality from this hedonic regression provides a valuable cross-validation for the estimated quality from the IRT model with a high correlation of $r = 0.972$.

Lastly, several additional robustness tests were carried out on the current model form, Equations (1)-(8). Both launch and post-launch models were estimated: as separate ordinary least squares (OLS) equations, using number of pirated files in place of seeders, with a word of mouth copula, with no control variables, with online 'buzz' variable reflecting search interest in the film (i.e., search popularity of the film on IMDB). The substantive results did not change. Additionally, the post-launch model was also estimated with film fixed effects and with robust standard errors to account for possible serial correlation. The findings on the focal variables of interest are largely the same as those reported in Tables 3 and 4. Finally, although most advertising occurs up to and including the week of launch, a re-estimation of the post-launch model with advertising found no significant effect of advertising after launch.

5. Conclusion

Piracy represents a considerable threat to revenues of both producers and distributors (in the case of movies, studios and theaters, respectively). Extant piracy research has found mixed findings for whether piracy encourages sampling versus cannibalization. This study examined the role of piracy quality and its effect on the market. In particular, we theorize that higher quality copies can both hurt and help sales. We contribute to the piracy literature by proposing that when product uncertainty is high, namely during the launch period, enthusiastic consumers will search out more information to reduce this uncertainty. Higher quality copies should be more informative, lowering product uncertainty, and better align consumer expectations for purchasing. Yet, product uncertainty is lower post-launch as information spreads in the market. As the most enthusiastic consumers have likely purchased, the less interested customers remain; these customers are drawn more toward unwillingness to pay than information search, cannibalizing sales.

This research makes several contributions to the piracy literature. First, while piracy is a well-researched area, less has been said about piracy quality, which we theorized may ease product uncertainty in the marketplace. Second, the subjective nature of piracy quality was addressed using observed signals from pirated copies, where an item response model uncovered a continuous measure of quality, rather than merely treat the presence of a given keyword as either high or low quality. Third, the impact of piracy quality was assessed to account for the supply side (both legal screens and illegal seeders), which is often omitted in piracy research. As the legal and illegal sides of the market are interdependent, the model uses seemingly unrelated regression, with copulas to address endogeneity. Fourth, the interaction effects show the piracy quality effects in relation to consumer downloads, or ln(Quality)*ln(Leechers) in our model. In particular, a 1% increase in higher quality piracy downloads corresponds to a 0.143% increase in revenues in the launch period, although a 1% increase in higher quality piracy downloads post-launch yields a -0.075% decline in revenues, controlling for everything else. These differential effects in timing, in addition to the role of quality, help alleviate prior research tensions as to whether piracy acts as a sampling mechanism or substitution.

The findings point to two key managerial implications, especially because enforcement resources are limited even in the most resourceful nations (Fink et. al., 2016), tighter regulation may have negative effects (Chen et al., 2024), and the low efficiency of public policy against digital piracy (Bourreau et. al., 2021). First, studios could afford to be less stringent on higher quality piracy in the opening week, as higher quality copies help reduce information uncertainty among consumers. However, higher quality copies may persist over time, hurting post-launch sales. This second implication is not a foregone conclusion: studios could flood the illegal market with their own copies after launch. Although this seems counter-intuitive, as studios work with authorities to remove illegal copies, putting many copies on the market will make it difficult for pirates to search for a good copy. Our findings speak to how studios can better optimize anti-piracy efforts: it is not possible to get rid of all piracy, but to prioritize enforcement against the competitive quality piracy at the opening week and post-launch differently.

Since piracy derives from the original product, the genuine good represents a powerful tool for managers; by owning the film, studios can release their own sampling variations. Studios can use this to their advantage by providing more information to consumers with some degree of high-quality (but not full) versions. For instance, releasing a high-quality copy of a film's first 30 minutes but making the remaining 90 minutes a black screen will provide information on the story, characters, and pacing to interest consumers, but frustrate pirates. In this way, studios can create 'honest trailers' to similar effect: get viewers hooked by showing how a movie starts (such as the first 10-20 minutes), or even creating longer trailers to put together more scenes to better convey the plot. Post-launch, managers can reduce piracy quality overall by releasing their own low quality (and still not full) versions to preempt low-quality entrants (Dawande et. al. 2010). Strong brands should enforce their own IP protection (e.g., Pun & Hou, 2022) rather than leave this to government, and studio enforcement efforts could focus on the higher quality copies to turn consumers to theaters, the only channel with a guaranteed full version of the film.

Along with this study's 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, in a sampling period that is relatively immune to recent streaming services; our results are still relevant given the ease with which streaming brings a digital copy into consumers' homes for possible pirating (Jain et. al. 2020). 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). Our approach could be extended to that realm once data becomes available. Third, our framework is not a general equilibrium one. If we assume that producers and theaters have already optimized based on all information on the legal market, we focus on proposing novel methods in estimating the effect of piracy shocks from the illegal market. However, we acknowledge the changing nature of the film industry – due to streaming and theater closures from COVID – as altering how consumers watch movies. Once studios make full-length high-quality copies available for home consumption, it becomes easy for consumers to create copies with less risk of being caught. This merits further investigation on how timing home releases affect piracy, and what an optimal window for streaming consumption may be; McKenzie et al. (2019) and Frick et. al. (2023) present useful investigations of the Netflix and Subscription Video on Demand (SVoD) effects and discussions on future research on streaming impacts. The methodologies and findings here have long-lasting implications. Finally, while we focus on information goods, illegal versions in other product categories might exhibit different consumption patterns, presenting a potential direction for future research avenues. As such, this study serves as a stepping stone in the broader piracy literature by assessing copy quality as part of the new agenda for the economics of digitization.

REFERENCES

Ackermann, K., Bradley, W., & Cameron, J. F. (2020). The pirate bay & box office buccaneers. Working paper, Monash University, Australia.

Aguiar, L., & Martens, B. (2016). Digital music consumption on the internet: Evidence from clickstream data. *Information Economics and Policy*, *34*, 27-43.

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

Andersen, B., & Frenz, M. (2010). Don't blame the P2P file-sharers: the impact of free music downloads on the purchase of music CDs in Canada. *Journal of Evolutionary Economics*, *20*(5), 715-740.

Bae, S. H., & Choi, J. P. (2006). A model of piracy. *Information Economics and Policy*, *18*(3), 303-320.

Bagozzi, R. P., & Dholakia, U. M. (2006). Antecedents and purchase consequences of customer participation in small group brand communities. *International Journal of research in Marketing*, *23*(1), 45-61.

Bartholomew, D. J., Steele, F., Moustaki, I., & Galbraith, J. I. (2002). *The analysis and interpretation of multivariate data for social scientists*. Chapman and Hall/CRC, Boca Raton, FL.

Bawa, K., & Shoemaker, R. (2004). The effects of free sample promotions on incremental brand sales. Marketing Science, 23(3), 345-363.

Belleflamme, P., & Peitz, M. (2014). Digital piracy: An update. Working paper, Université Catholique de Louvain, Belgium.

Bhattacharjee, S., Gopal, R. D., & Sanders, G. L. (2003). Digital music and online sharing: software piracy 2.0? *Communications of the ACM*, *46*(7), 107-111.

Bourreau, M., Lumeau, M., Moreau, F. & Viotto da Cruz, J. (2021). Attitude face au risque et piratage de films en ligne. *Revue économique*, 72, 633-666.

Casting Networks (2023). From Low Budget to High Impact: Mastering the Craft of Indie Film Acting. Retrieved January 10, 2024 from https://www.castingnetworks.com/news/from-low-budget-to-high-impact-mastering-the-craft-of-indie-film-acting/

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

Chen, J., Gao, Y., & Ke, T. T. (2024). Regulating digital piracy consumption. *Journal of Marketing Research*, forthcoming.

Chintagunta, P. K., Gopinath, S., & Venkataraman, S. (2010). The effects of online user reviews on movie box office performance: Accounting for sequential rollout and aggregation across local markets. *Marketing Science*, *29*(5), 944-957.

Choi, I. (2001). Unit root tests for panel data. *Journal of International Money and Finance*, *20*(2), 249-72.

Clement, M., Wu, S., & Fischer, M. (2014). Empirical generalizations of demand and supply dynamics for movies. *International Journal of Research in Marketing*, *31*(2), 207-223.

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

Danaher, B., & Smith, M. D. (2014). Gone in 60 seconds: The impact of the Megaupload shutdown on movie sales. *International Journal of Industrial Organization*, *33*(1), 1-8.
Danaher, B., Smith, M.D. and Tu, X. (2019). The impact of early international digital release of films on US box office revenues. *25th Americas Conference on Information Systems*, *Cancun*.
Datta, H., van Heerde, H. J., Dekimpe, M. G., & Steenkamp, J. B. E. (2022). Cross-National Differences in Market Response: Line-Length, Price, and Distribution Elasticities in 14 Indo-Pacific Rim Economies. *Journal of Marketing Research*, *59*(2), 251-270.
Dawande, M., Drobouchevitch, I., Rajapakshe, T., & Sriskandarajah, C. (2010). Analysis of Revenue Maximization Under Two Movie-Screening Policies. *Production and Operations Management*, *19*(1), 111-124.
De Jong, M. G., Steenkamp, J. B. E., & Fox, J. P. (2007). Relaxing measurement invariance in cross-national consumer research using a hierarchical IRT model. *Journal of Consumer Research*, *34*(2), 260-278.
Dellarocas, C., Zhang, X., & Awad, N. F. (2007). Exploring the value of online product reviews in forecasting sales: The case of motion pictures. *Journal of Interactive Marketing*, *21*(4), 23-45.
Dey, D., Kim, A., & Lahiri, A. (2019). Online piracy and the "longer arm" of enforcement. *Management Science*, *65*(3), 1173-1190.
Dhar, T., & Weinberg, C. B. (2016). Measurement of interactions in non-linear marketing models: The effect of critics' ratings and consumer sentiment on movie demand. *International Journal of Research in Marketing*, *33*(2), 392-408.
Eckert, C., & Hohberger, J. (2023). Addressing endogeneity without instrumental variables: An evaluation of the gaussian copula approach for management research. *Journal of Management*, *49*(4), 1460-1495.
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.
Eller, C., & Friedman, J. (2008, June 14). *Release date almost as important as good film*. SF Gate. Retrieved October 31, 2022, from http:// www.sfgate.com/entertainment/article/Release-date-almost-as-important-as-good-film-3209001.php
Fader, P. S. (2000). *Expert report: A&M Records, Inc. v. Napster, Inc*. Retrieved October 31, 2022, from http://news.findlaw.com/hdocs/docs/napster/napster/fader_070300.pdf
Fink, C., Maskus, K. E., & Qian, Y. (2016). The economic effects of counterfeiting and piracy: A review and implications for developing countries. *The World Bank Research Observer*, *31*(1), 1-28.
Fishbein, M. (1963). An investigation of the relationships between beliefs about an object and the attitude toward that object. *Human Relations*, *16*(3), 233-239.
Frick, S. J., Fletcher, D., & Smith, A. C. (2023). Pirate and chill: The effect of netflix on illegal streaming. *Journal of Economic Behavior & Organization*, *209*, 334-347.
Garvin, D. A. (1987). Competing on the eight dimensions of quality. *Harvard Business Review*, *65*(6), 101-109.
Geng, X., & Lee, Y. J. (2013). Competing with piracy: A multichannel sequential search approach. *Journal of Management Information Systems*, *30*(2), 159-184.
Godinho de Matos, M., Ferreira, P., & Smith, M. D. (2018). The effect of subscription video-on-demand on piracy: Evidence from a household-level randomized experiment. *Management Science*, *64*(12), 5610-5630.
Go-Globe (2020). Eye-Opening Statistics of Online Piracy in Numbers. Retrieved October 10, 2023 from https://www.go-globe.com/online-piracy-in-numbers-facts-and-statistics-infographic/
Gopinath, S., Chintagunta, P. K., & Venkataraman, S. (2013). Blogs, advertising, and local-market movie box office performance. *Management Science*, *59*(12), 2635-2654.
Han, Y. J., Nunes, J. C., & Drèze, X. (2010). Signaling status with luxury goods: The role of brand prominence. *Journal of Marketing*, *74*(4), 15-30.
Hennig-Thurau, T., Henning, V., & Sattler, H. (2007). Consumer file sharing of motion pictures. *Journal of Marketing*, *71*(4), 1-18.

Hennig-Thurau, T., Houston, M. B., & Sridhar, S. (2006). Can good marketing carry a bad product? Evidence from the motion picture industry. *Marketing Letters*, *17*(3), 205-219.
Holbrook, M. B., & Corfman, K. P. (1985). Quality and value in the consumption experience: Phaedrus rides again. In J. Jacoby & J. C. Olsen (Eds.), *Perceived quality* (pp. 31-57). Lexington Books, Lexington, MA.
Hong, S. H. (2013). Measuring the effect of napster on recorded music sales: difference-in-differences estimates under compositional changes. *Journal of Applied Econometrics*, *28*(2), 297-324.
Hui, K. L., & Png, I. (2003). Piracy and the legitimate demand for recorded music. *Contributions in Economic Analysis & Policy*, *2*(1), 1-22.
IBC (2017). *Cost of online piracy to hit $52bn*. IBC.org. Retrieved October 31, 2022, from 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.
Jain, T., Hazra, J., & Cheng, T. E. (2020). Illegal content monitoring on social platforms. *Production and Operations Management*, *29*(8), 1837-1857.
Kim, D., Park, K., & Bang, Y. (2022). The effect of TV drama piracy: An analysis of digital piracy users, internet buzz, and TV drama viewership. *Information & Management*, *59*(2), 103599.
Koschmann, A., & Bowman, D. (2017). Simultaneous estimation of legal and illegal supply and demand: The case of motion pictures. *International Economic Journal*, *31*(4), 555-577.
Kretschmer, T., & Peukert, C. (2020). Video killed the radio star? Online music videos and recorded music sales. *Information Systems Research*, *31*(3), 776-800.
Krider, R. E., Li, T., Liu, Y., & Weinberg, C. B. (2005). The lead-lag puzzle of demand and distribution: A graphical method applied to movies. *Marketing Science*, *24*(4), 635-645.
Lahiri, A., & Dey, D. (2013). Effects of piracy on quality of information goods. *Management Science*, *59*(1), 245-264.
Li, X., & Hitt, L. M. (2008). Self-selection and information role of online product reviews. *Information Systems Research*, *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.
Liebowitz, S. J. (2008). Research note—Testing file sharing's impact on music album sales in cities. *Management Science*, *54*(4), 852-859.
Liu, Y. (2006). Word of mouth for movies: Its dynamics and impact on box office revenue. *Journal of Marketing*, *70*(3), 74-89.
Liu, Z. (2019). Quantifying the Heterogeneous Effects of Piracy on the Demand for Movies. Working Paper, University of Western Ontario.
Lord, F. M. (2012). *Applications of item response theory to practical testing problems*. Routledge, New York.
Lu, S., Wang, X., & Bendle, N. (2020). Does piracy create online word of mouth? An empirical analysis in the movie industry. *Management Science*, *66*(5), 2140-2162.
Luca, M. (2011). Reviews, reputation, and revenue: The case of yelp.com. Working paper. Harvard University.
Ma, L., Montgomery, A. L., Singh, P. V., & Smith, M. D. (2014). An empirical analysis of the impact of pre-release movie piracy on box office revenue. *Information Systems Research*, *25*(3), 590-603.
Machado, F. S., Raghu, T. S., Sainam, P., & Sinha, R. (2017). Software piracy in the presence of open source alternatives. *Journal of the Association for Information Systems*, *18*(1), 3.
Martikainen, E. (2014). Does file-sharing reduce DVD sales? *NETNOMICS: Economic Research and Electronic Networking*, *15*(1), 9-31.
McKenzie, J. (2017). Graduated response policies to digital piracy: Do they increase box office revenues of movies? *Information Economics and Policy*, *38*, 1-11.
McKenzie, J. (2023). The economics of movies (revisited): A survey of recent literature. *Journal of Economic Surveys*, *37*(2), 480-525.

McKenzie, J., Crosby, P., Cox, J., & Collins, A. (2019). Experimental evidence on demand for "on-demand" entertainment. *Journal of Economic Behavior & Organization*, *161*, 98-113.

Mitra, D., & Golder, P. N. (2006). How does objective quality affect perceived quality? Short-term effects, long-term effects, and asymmetries. *Marketing Science*, *25*(3), 230-247.

Monroe, K. B., & Krishnan, R. (1985). The effect of price on subjective product evaluations. In J. Jacoby & J. C. Olsen (Eds.), *Perceived Quality* (pp. 209-232). Lexington Books, Lexington, MA

Moorthy, S., Ratchford, B. T., & Talukdar, D. (1997). Consumer information search revisited: Theory and empirical analysis. *Journal of Consumer Research*, *23*(4), 263-277.

Mortimer, J. H., Nosko, C., & Sorensen, A. (2012). Supply responses to digital distribution: Recorded music and live performances. *Information Economics and Policy*, *24*(1), 3-14.

Moul, C. C. (2005). *A concise handbook of movie industry economics*. Cambridge University Press, New York.

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

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

Olshavsky, R. W. (1985). Perceived quality in consumer decision making: an integrated theoretical perspective. In J. Jacoby & J. C. Olsen (Eds.), *Perceived Quality* (pp. 3-29). Lexington Book, Lexington, MA.

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

Papies, D., Ebbes, P., & Van Heerde, H. J. (2017). Addressing endogeneity in marketing models. In P. S. H. Leeflang, J. E. Wieringa, T. H. A. Bijmolt , & K. H. Pauwels (Eds.), *Advanced methods for modeling markets* (pp. 581-627). Springer, Berlin.

Park, S., & Gupta, S. (2012). Handling endogenous regressors by joint estimation using copulas. *Marketing Science*, *31*(4), 567-586.

Pastpipatkul, P., Maneejuk, P., Wiboonpongse, A., & Sriboonchitta, S. (2016). Seemingly unrelated regression based copula: an application on Thai rice market. In V. N. Huynh, V. Kreinovich, & S. Sriboonchitta (Eds.) *Causal inference in econometrics* (pp. 437-450). Springer, Switzerland.

Pastpipatkul, P., Panthamit, N., Yamaka, W., & Sriboochitta, S. (2016, November). A copula-based Markov switching seemingly unrelated regression approach for analysis the demand and supply on sugar market. In V. N Huynh, M. Inuiguchi, B. Le, B. N. Le, & T. Denoeux (Eds.) *International Symposium on Integrated uncertainty in knowledge modelling and decision making* (pp. 481-492). Springer, Vietnam.

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

Peukert, C., Claussen, J., & Kretschmer, T. (2017). Piracy and box office movie revenues: Evidence from Megaupload. *International Journal of Industrial Organization*, *52*, 188-215.

Pun, H., & Hou, P. (2022). Combating copycatting from emerging market suppliers in global supply chains. *Production and Operations Management*, *31*(8), 3304-3319.

Qian, Y. (2014). Brand management and strategies against counterfeits. *Journal of Economics & 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.

Qian, Y., Xie, H., & Koschmann, A. (2022). Should copula endogeneity correction include generated regressors for higher-order terms? No, it hurts. Working paper (No. w29978). National Bureau of Economic Research.

Qiu, D., & Srikant, R. (2004). Modeling and performance analysis of BitTorrent-like peer-to-peer networks. *ACM SIGCOMM Computer Communication Review*, *34*(4), 367-378.

Reid, R. D., & Sanders, N. R. (2001). *Operations management.* Wiley & Sons, Hoboken, NJ.

Reeves, C. A., & Bednar, D. A. (1994). Defining quality: alternatives and implications. *Academy of Management Review*, *19*(3), 419-445.

Rutz, O. J., & Watson, G. F. (2019). Endogeneity and marketing strategy research: An overview. *Journal of the Academy of Marketing Science*, *47*(3), 479-498.
Smith, M. D., & Telang, R. (2009). Competing with free: The impact of movie broadcasts on DVD sales and Internet piracy. *MIS Quarterly*, 321-338.
Somlo, B., Rajaram, K., & Ahmadi, R. (2011). Distribution planning to optimize profits in the motion picture industry. *Production and Operations Management*, *20*(4), 618-636.
Spangler, T. (2014, April 18). *'Expendables 3' illegally downloaded 5 million times, but still isn't top hit for pirates*. Variety. Retrieved October 31, 2022, from https://variety.com/2014/digital/news/expendables-3-illegally-downloaded-5-million-times-but-still-isnt-top-hit-for-pirates-1201285179/
Spector, P. E., Van Katwyk, P. T., Brannick, M. T., & Chen, P. Y. (1997). When two factors don't reflect two constructs: How item characteristics can produce artifactual factors. *Journal of Management*, *23*(5), 659-677.
Sundararajan, A. (2004). Managing digital piracy: Pricing and protection. *Information Systems Research*, *15*(3), 287-308.
Trivedi, P. K., & Zimmer, D. M. (2007). *Copula modeling: An introduction for practitioners*. Now Publishers Inc., Hanover, MA.
Tyrowicz, J., Krawczyk, M., & Hardy, W. (2020). Friends or foes? A meta-analysis of the relationship between "online piracy" and the sales of cultural goods. *Information Economics and Policy*, *53*, 100879.
Van Schuur, W. H., & Kiers, H. A. (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, H. L. (2015). *Entertainment industry economics*. Cambridge University Press, New York.
Wasow, O., Baron, A., Gerra, M., Lauderdale, K., & Zhang, H. (2010). Can tweets kill a movie? An empirical evaluation of the Bruno effect. CHI 2010 Workshop on Microblogging (Atlanta).
Waldfogel, J. (2012) Digital piracy: Empirics. In M. Peitz & J. Waldfogel (Eds.), *The Oxford Handbook of the Digital Economy* (pp. 512-546). Oxford University Press, England.
Wilcox, K., Kim, H. M., & Sen, S. (2009). Why do consumers buy counterfeit luxury brands? *Journal of Marketing Research*, *46*(2), 247-259.
Yang, F., Qian, Y., & Xie, H. (2022). Addressing endogeneity using a two-stage copula generated regressor approach. *Journal of Marketing Research*, forthcoming.
Yi, Z., Cao, Z., & Cheung, K. L. (2022). Managing digital piracy under consumer valuation uncertainty: The roles of product demonstrations and antipiracy measures. *Information & Management*, 59(2), 103601.
Yoon, K. (2007). On the impact of digital music distribution. Working paper, Hitotsubashi University, Japan.
You, Y., Vadakkepatt, G. G., & Joshi, A. M. (2015). A meta-analysis of electronic word-of-mouth elasticity. *Journal of Marketing*, *79*(2), 19-39.
Yue, Y. (2020). The effects of movie piracy on box-office revenue: an empirical analysis of the Chinese movie market. *Journal of Applied Economics*, *23*(1), 618-655.
Zeithaml, V. A. (1988). Consumer perceptions of price, quality, and value: a means-end model and synthesis of evidence. *Journal of Marketing*, *52*(3), 2-22.
Zellner, A., & Theil, H. (1962). Three-stage least squares: Simultaneous estimation of simultaneous equations. *Econometrica*, *30*(1), 54–78.
Zhang, L. (2018). Intellectual property strategy and the long tail: Evidence from the recorded music industry. *Management Science*, *64*(1), 24-42.
Zhang, X. M., & Dellarocas, C. (2006). The lord of the ratings: How a movie's fate is influenced by reviews. *Proc. 27th International Conference on Information Systems (ICIS)*, 117.
Zhu, F., & Zhang, X. (2010). Impact of online consumer reviews on sales: The moderating role of product and consumer characteristics. *Journal of marketing*, *74*(2), 133-148.