Using Markets to Measure the Impact of File Sharing on Movie Revenues

Koleman Strumpf*†
University of Kansas School of Business

Preliminary

July 11, 2014

Abstract

File sharing provides a useful laboratory for investigating the economic importance of intellectual property protection. There are two main empirical challenges: overcoming the non-random timing of the arrival date of illicit copies and dealing with low statistical power due to limited sample size. This paper uses markets to address these issues in the context of movies. I show forward-looking markets can be used to establish the unobserved counter-factual of how movie revenues would change on any possible file sharing release date, particularly those prior to the theatrical premier. Using movie-level tracking stocks in conjunction with the arrival date of illicit copies, I find that file sharing has only a modest impact on box office revenue.

*I thank a large BitTorrent index site and the Hollywood Stock Exchange for furnishing me with data. A portion of this paper was written while I was a Visiting Scholar at the Management & Strategy Department at the Kellogg School of Management. Correspondence should be sent to 1300 Sunnyside Avenue Lawrence, KS 66045 or cigar@ku.edu.

1 Introduction

Internet piracy remains an area of active interest in academia, policy and industry. Policy-makers and firms are primarily interested in how changes in the nature of intellectual property protection will influence the incentives for and creation of products such as movies, music and books. Academics are interested in these questions as well as broader lessons about the appropriate intellectual property policy regime. In this paper I add to the discussion by examining the economic consequences of the leading form of movie piracy, downloads of unauthorized copies on file sharing networks. This continues to remain a popular source of consumption, and by one measure unauthorized downloads consume about a quarter of internet bandwidth in developed countries (Price, 2013). Most previous work on the economic impact of internet piracy has focused on music (Oberholzer-Gee and Strumpf, 2009), and there are enough differences with movies that further attention is warranted. For example, with movies authorized consumption involves not just the content but a location which provides unique characteristics (large screens and social interaction) which are not bundled with a download. Also almost all consumption occurs over a relatively short period following the theatrical release.

This paper presents estimates of the economic impact of movie piracy over the period 2003-2009. My approach is based on a market in which stocks track the future US theatrical box office of specific movie titles. This market, the Hollywood Stock Exchange® (also known as HSX®), involves thousands of traders and has a track record of accurate forecasts and rapidly incorporating news.[1] I examine how the stock prices respond to releases of movies onto file sharing networks, which I date using internal records from one of the five largest file sharing sites during my study period. The econometric framework allows me to infer the market beliefs at any time of a movie's expected box office, both for the observed state (whether it has reached file sharing networks) and the unobserved state. Identification follows from shocks, namely how surprising it is that a movie did or did not reach file sharing networks on each day (expectations of release probability are derived from historical release patterns and from duration models). The displacement effect at each time is identified from price movements for all movies which have not yet reached file

[1] I will further establish HSX's ability to incorporate information by looking at how prices respond to the release of authorized movie trailers and how volume responds to unauthorized file sharing leaks.

sharing networks, which means the estimates are less likely to be biased for the pre-theatrical period when there are relatively few movies available illicitly. I can uncover the average impact of a file sharing release on each day, and use these estimates to appraise economic and statistical significance. I find these displacement effects to be of modest size and indicating only a small impact on aggregate revenues. This small effect is perhaps the result of low quality initial file sharing copies and the superiority of the theater viewing experience.

There are several advantages to using markets to quantify the impact of illicit movie downloads The first relates is establishing causality and creating appropriate counter-factuals. There is a small literature on the economics of movie piracy (Danaher and Waldfogel, 2012; Ma et al, 2013).[2] In most observational studies, the file sharing release pattern is treated as exogenous across movies. But this is unlikely to be true for both supply reasons (studios working harder to avoid releases for movies which they believe will be more negatively impacted) and demand reasons (file sharing participants work harder to release movies which they believe will have greater illicit demand and likely larger negative damages). **Table 1** shows that movies available on file sharing networks prior to their theatrical release are different across several observable characteristics including the releasing studio, genres and ratings by critics or users. Presumably there are many unobserved differences as well.[3] Finding earlier releases have higher or lower revenues could reflect sample selection and heterogeneous effects rather then a causal impact on revenues. Markets help alleviate this bias because both unreleased and newly released movies are used to infer the impact at every time. In particular this means in the period prior to the theatrical release, when relatively few films are available on file sharing networks, there is a rather full sample and the bias will be small. This is important since such "pre-releases" are of particular policy and academic interest.

A second advantage is improved statistical power. There is a substantial increase in the amount of variation used to measure economic damages. In typical observational studies, there is essentially one observation per movie (the time of the file sharing release and the total revenue). With markets, the price moves continuously

[2]The paper also related to the literature measuring the impact of piracy on the software industry. For example Athey and Stern (2013) show that restricting access to file sharing networks has little impact on piracy rates for operating systems.

[3]In principle this could be addressed using an instrumental variable approach, but it is hard to think of variables which capture the complex dynamics which govern this equilibrium process.

and each price change provides useful information. Market traders know there is some probability a movie will be released on file sharing networks at each time, and expectations about this release pattern are impounded in market prices. As time proceeds, observing whether a movie is or is not released provides new information relative to these expectations. Up until the release, the resulting price change can be used to infer traders beliefs about how a release at that moment would change box office revenues. Daily data available are available here, which provides several weeks of price changes rather than the the typical one observation per movie. This allows both greater precision and more flexible estimation of the impact of file sharing releases at each day relative to the theatrical release.

An additional advantage of this paper is the data. While previous studies have had rather imprecise information about the timing of a movie's file sharing release date, I have access to the complete database of a one of the leading index sites (the website users go to begin their movie downloads). This includes both the time at which each file was originally uploaded and the time at which each user download began. This allows me to eliminate fake files, which are rarely downloaded and/or quickly lose attention of users. I find that such fake files are common in the databases which have been used in other papers. The databases are also incomplete and omit some early release dates.

A final advantage is the generality of the approach. The econometric framework could be used to measure the displacement from illicit consumption of other goods, in other countries, and in other times. The key ingredient is the presence of futures markets, which are becoming more common due to the growth of online prediction markets.

The next section provides institutional details on file sharing and on HSX. **Section 3** presents the empirical framework and addresses identification. The data is described in **Section 4**, while the estimates are in **Section 5**. The last section concludes.

2 Institutional Background

Movie file sharing first came into vogue between 2001 and 2003 with the introduction of the BitTorrent protocol and the subsequent introduction of torrent index sites such as The Pirate Bay which provided links to clouds of users sharing specific titles. Between 2004 and 2009 BitTorrent was responsible of in excess of a third of internet traffic. Of particular interest is how movies reach file sharing networks (**Figure 1**). The activity is coordinated by scene release groups which obtain unauthorized copies, encode them in a usable form for viewing, and eventually post them to public sites such as torrent index sites. The release groups are private and membership is tightly regulated. They do not appear to generate any financial returns, but the large number of groups (perhaps as many as one hundred) compete fiercely to be the first to provide titles. The files first appear on a private server (topsite) and then eventually propagate to public sites. On the other side, movie studios and theaters have worked to slow these releases through greater security of the original film prints and the use of anti-piracy technologies such as night vision goggles to detect filming in theaters.

Movies are made available on file sharing networks near in time to their theatrical release. **Figure 2** shows the empirical distribution of release dates for a sample of movies discussed in **Section 4**. In the initial years of widespread movie file sharing, the typical movie was available on these networks prior to their U.S. theatrical release. Most of these pre-release copies came from insiders at movie studios (workprints from disgruntled employees or contractors in the distribution and production channel), the studios (advance copy screeners sent to industry personnel) and the remainder from copies premiering earlier outside the U.S. The workprints and screeners often lack post-processing, miss certain scenes and special effects, do not have the final audio, and have various watermarks, which list the studio name and elapsed time. Byers et al (2003) presents more details on the source of illict movie copies, and finds that 77% of movies in a sample of popular movies were leaked by industry insiders. As studios increased security and anti-piracy measures were implemented in theaters, this release was delayed to one to two weeks after the theatrical release. These early, post-theatrical releases were typically of low quality copies of the movie (coming from audience cams, projection-booth telesyncs, or copies missing special effects). Higher quality movies made from direct copies of the movie, such as Blu-Rays, DVDs or telecines typically are not available until after the movie's theatrical release. By the end of the 2000s, however, ten to twenty percent of movies were available in high quality during their theatrical run.

The other component of this study is the Hollywood Stock Exchange (HSX). HSX is a play money prediction market founded in 1996 in which stocks for particular movies are linked to their US theatrical box office gross. Such tracking stocks are available for every movie from major studios as well as many smaller releases. Movie stocks are introduced (IPOed) as early as the concept stage, and they are always traded many months before the theatrical opening. The stocks expire and pay off based on box office revenues four weeks after its wide release (or twelve weeks if the movie remains in limited release). Investors in the market start with a portfolio of play money which can be invested in stocks (shorts are permitted and there is a one percent commission on the value of each trade). By 2006 there were 623k active users making 42k trades daily. A growing literature has found that HSX prices accurately forecast the eventual box office (Pennock et al, 2000) and that information such as casting announcements is rapidly impounded in the stock price (Elberse, 2007). The latter in particular suggests that HSX investors are able to quickly interpret data, so tracking HSX stock prices might be used to ascertain the impact of unauthorized downloads on revenues.

A final point is to note that in principle the stock prices of movie studios could be used in place of HSX. This is difficult to do in practice since the major studios are all part of the larger conglomerates and so individual movies have a limited impact on stock prices. Still it is possible to look at smaller studios to show that HSX prices track stock prices. A leading example is Lionsgate whose stock has changed sharply in the wake of surprising movie outcomes. **Figure 3** shows that information is impounded in HSX prices at a comparable rate as Lionsgate's stock price both for positive (panel a) and negative (panel b) shocks. This provides some justification for using HSX, since traditional financial markets are generally considered to be efficient.

3 Empirical Framework

3.1 Empirical Model

The objective is to use HSX stock prices to measure the impact of a file sharing release on the movie box office. Ideally this would be accomplished by having state-contingent prices, where the state is whether a movie has reached file sharing networks. Then one could simply compare prices at any time to establish the market consensus on the impact of a file sharing release. HSX stocks, however, are not state-dependent. It is still possible to use observed prices to establish the impact on box office at *each day* relative to the theatrical release date.

Consider a movie which has not yet had a file sharing release at time t. Under efficient markets, the stock price will today will be the expected value of its price over all possible future states. The discrete time evolution of its stock price is then,

$$S_t = p_{t+1} E_t \left(\widetilde{S}_{t+1}(1) \right) + (1 - p_{t+1}) E_t \left(\widetilde{S}_{t+1}(0) \right) \tag{1}$$

where S_t is the stock price, p_{t+1} is the probability the movie will be released at time $t+1$ conditional on it not yet having been released by t, and $\widetilde{S}_{t+1}(s)$ is the price next period conditional on the state ($s = 1$ denotes a release and $s = 0$ no release). Next period prices are stochastic given the information available today, so an expectations operator is applied (I presume that the information set at time t of investors consists of all events up to that time).

Now suppose that next period's price and state is observed. The state-specific change in prices can be written as,

$$S_{t+1}(s) - S_t = (\mathbb{I}(t+1) - p_{t+1})\, \delta_{t+1} + \Omega_{t+1} \tag{2}$$

where $\mathbb{I}(t+1)$ is an indicator of whether the movie is released next period conditional on it not having yet been released, $\delta_{t+1} \equiv S_{t+1}(1) - S_{t+1}(0)$ is the difference in stock price in states where the movie is and is not released, and Ω_{t+1} is other price-relevant information which arrived since last period.[4] The interpretation of (2) is that the change in stock price (left hand side) is larger when next period's state of the world is more surprising (parenthesis term on right hands side) and when a file sharing release has a bigger impact on box office revenue (δ_{t+1}).

δ_{t+1} measures the causal impact of a file sharing release at time $t+1$. $\delta_{t+1} < 0$ indicates that file sharing displaces (expected) box office revenue. Note that while δ_{t+1} is never observed, all of the other terms in (2) in principle can be measured. It will be possible to estimate δ_{t+1} using the observed path of prices in each period up until the movie reaches file sharing networks.

3.2 Estimation and Endogeneity

Using the data described below I will estimate (2). The basic specification will be estimated using periods up until the movie is released on file sharing networks,

$$\triangle S_{i,t+1} = (\mathbb{I}_i(t+1) - p_{i,t+1})\, \delta_{t+1} + X_{i,t+1}\beta + \epsilon_{i,t+1} \tag{3}$$

[4] I presume $\widetilde{S}_{t+1}(s) = S_{t+1}(s) + \tilde{\Omega}_{t+1}$. For any state, all of the uncertainty about next period's price is contained in $\tilde{\Omega}_{t+1}$. $\tilde{\Omega}_{t+1}$ is presumed to be independent of whether a movie has a a file sharing release, and it is presumed to be an additively separable component of $\widetilde{S}_{t+1}(s)$.

where the i subscript has been added to denote movie-specific values, $\triangle S_{i,t+1} \equiv S_{i,t+1}(s_i) - S_{i,t}$, and Ω_{t+1} has been partitioned into an observed $X_{t+1}\beta$ and an unobserved ϵ_{t+1} component. In most specifications I will use movie and time fixed effects for the controls, $X_{i,t+1}\beta \equiv \mu_i + \upsilon_t$ (the dependent variable and the other right hand side variable vary over both i and t). The parameters δ_{t+1} will be estimated, with the sequence $\{\delta_{t+1}\}$ being the measure of file sharing impact at any t. One potential difficulty that may arise is discontinuities and jumps in the estimated $\{\delta_{t+1}\}$. A solution to this will be to use a smoothing estimator.

There are two main empirical issues. The first is that $p_{i,t}$ is unobserved. This represents the beliefs of market participants about how likely a movie is to be released in a particular period. I will use a variety of approaches to proxy for these beliefs and will discuss this in more detail in the next section.

A more pressing challenge is the potential endogeneity of the release date $\mathbb{I}_i(t)$. The main threat to identification would be that the file sharing timing arrival time is related to the extent of crowd-out.[5] That is there is heterogeneity in the crowd-out across movies in any period δ_{it}, and $Corr\left(\mathbb{I}_i(t), \delta_{it}\right)$ varies by t. For example, if movies which are less susceptible to crowd-out arrive later to file sharing sites then one might expect there to be a positive bias for δ_t for these later periods. Fixed effects need not eliminate this problem if there are interactions between the level of heterogeneity and time. Fortunately the δ_t estimates are based on the price changes in *all* movies which have not yet reached file sharing networks by t, so this is not a problem for earlier time periods (that is the estimated effect is roughly the average δ_{it} for all non-released movies, and almost all movies are unreleased in early t). Still the dynamic sample selection could lead to bias for later t. One solution I will use is to restrict the sample to movies which have later releases, and see if this changes the estimated δ_t in early t. Another solution is to restrict the sample to various subsets (such as grouping by movie genre or number of initial theatrical screens) which are expected to have similar file sharing release patterns and crowd-outs.

[5] Another issue is that the timing is related to box office revenue, say through the unobserved popularity of the movie. This is not a serious problem here since the dependent variable is differenced. Intuitively, identification stems from relatively high frequency events (i.e. whether a movie is released on file sharing networks in t rather than $t+1$) as well as longer term differences.

4 Data

To operationalize (3), empirical analogues of the variables are needed. This section discusses the data sources I use, and their advantage over other possible sources. The sample is the top one hundred and fifty highest grossing movies every year between 2003 and 2009 (for a total of N=1057 titles due to ties).[6] These movies account for at least ninety five percent of all domestic box office each year and include essentially all commercially relevant movies.

Some key data sources are listed in **Table 2**, and summary statistics are provided in **Table 3.**

4.1 File Sharing Variables

The file sharing data come from one of the five largest BitTorrent index sites during my study period, 2003-2009 (Zhang et al, 2011 shows there is substantial overlap in the most active files on the largest index sites). The site owner(s) gave me access to the original files for the site. The available data include a list of all files posted, the time-stamp of their posting, the user-name of the poster, the type of file (movie, music, tv show, software, etc), user votes on the quality of the file, meta-data about each file, and information about downloads. There are 7.6m files in the database of which 37% are movies. I matched the movies to my sample of titles using various fuzzy matching schemes particularly Levenshtein distance. A file is considered to be a high quality release if the meta-data contain certain keywords (DVDRip, BDRip, HD-DVD, R5, telecine, etc) or it indicates a high level of technical video or audio quality (such as use of certain video encodings and having high audio bitrates, though these data are often missing).

The file sharing release date is a key variable, which can be used to generate the movie-specific initial release indicator $\mathbb{I}_i(t+1)$ and the conditional release probability $p_{i,t+1}$ (there are also separate versions of these variables for high quality releases). An issue is that fake, corrupted or incomplete files can be posted. I eliminated these by only considering files which have at least one thousand downloads and which receive a minimum user rating score. For each title I then find the first date at which a file

[6] There are N=1047 movies with file sharing release dates. The titles without release data are multiple-movie showings or they are presented exclusively in 3-D or IMAX.

is available.[7] I then compare this to the movie's theatrical opening date to generate $\mathbb{I}_i(t+1)$.

These data can also be used to generate $p_{i,t+1}$. The simplest approach is to generate an empirical cdf of observed file sharing release dates relative to the theatrical release dates, $F(t)$ (see **Figure 2**). The desired measure is $p_{t+1} = (F(t+1) - F(t)) \,/\, (1 - F(t))$, though this does not vary across i. Preferred measures will tailor the probabilities to each movie. First I calculate the probability by the release year. Second I estimate a hazard model of movie release date and use the fitted values to create movie-specific probabilities (Let $H_i(t) \equiv 1 - F_i(t)$ be the fitted survivor function based on the hazard, then the fitted conditional probability is $p_{i,t} = 1 - (H_i(t+1)/H_i(t)))$. When selecting covariates for the hazard, it is important to not include factors which are related to box office revenue or the extent of crowd-out. Fortunately there are many plausibly exogenous factors influencing the release date. In addition to release date I will include the releasing studio of each movie. Studios vary in their security over early film prints and their policies regarding the circulation of screeners (both are common sources of file sharing releases, see **Section 2**), and the larger studios release a comparable and wide slate of movies with both high and low expected revenues. Another covariate is a measure of daily non-movie file sharing releases which is discussed next.

Figure 4 shows the number of files posted to the BitTorrent index site by day. The top panel is the total for movies and the bottom is for all other files such as music or software. The two series tend to move together, which is partly due to the secular growth in file sharing. But there are also common peaks and troughs as well as several differences between them. Much of this inter-temporal variation is due to the activity of certain posters who are often go dormant for periods and then return. In particular non-movie releases are a proxy for release group effort at bringing illicit content which is not driven either by the number of movies currently in release or demand for popular titles. A detrended version of this variable will be included in the hazard.

An important issue is how these data compare with other data sources which

[7] I constructed an alternative release date measure. It often takes time to download a movie, and at the beginning of the observation period it would take several days before the first completed download though by the end this fell to a few hours (Pouwelse et al, 2004 and Cuevas et al, 2010). Completed download times are not available in my data. I create an alternative release date measure based on when the number of downloads reaches a threshold that increases over the sample period. The main empirical results are qualitatively similar using this alternate measure.

have been used in other research (see **Table 2).** One approach has been to use public "Dupe Sites" which catalog file sharing releases. The leading movie sites over the observation period include NFOrce/NFOHump, VCDQuality, Scene Releases, and ORLYDB. There are several problems with using these sources. One is that they list when files are available on release group private servers available to a very limited number of users, with most of the content not becoming available on public sites until much later and in some cases never at all (see **Figure 1**). I collected the earliest release date from each of these sources for each title in my sample, and on average this is four days before they appear in my data from the BitTorrent index site. This effect is bigger for less popular movies: the correlation between the release date error and number of screens in the opening week (a measure of expected revenue) is -0.29. Using public dupe sites will over-state the revenue crowd-out from file sharing releases. A second problem is that are wide disagreements between the Dupe Sites. The range for initial release dates for the movies in my sample is 11 days across these sources, and there is no consistency in which site is first. This is likely due to different rules for omitting files and variation in which release groups are included in each site.[8]

4.2 Other Variables

Through a special arrangement, I have received daily HSX data for each of the movies in my sample. HSX data cover market activity starting with the stock IPO and continuing until the stock expires one to three months after its theatrical release. On average this includes almost three years of data, but a much smaller time window will be used since movies only reach file sharing networks relatively near the theatrical release data. The end-of-day HSX price will serve as the measure of S_t. I also consider two versions of daily trading volume, the number of shares and the dollar-weighted number of shares traded.

I also have data on movie characteristics, including genre and production studio, which come from `boxofficemojo.com` , `www.the-numbers.com`, and `www.imdb.com`.

[8] I was able to track down the source of these differences in certain cases. For example many files related to the movie Shrek 2 were removed from the VCDQuality due to a cease and desist order, and fake files related to the movie Bruce Almighty are present in NFOrce. I also used two databases leaked from scene release groups to confirm omissions and inconsistencies in the Dupe Sites (each leaked files is a "predb" or pre-database, which is an internal file the groups use to record the name, release times, and status of all files they have ever released). For example, VCDQuality is missing files posted over extended periods and incorrectly includes some files which have been listed as nuked meaning they are of poor quality or fake.

Critics rankings of the movies are from `www.rottentomatoes.com`. This are included in a few specifications which omit movie fixed effects.

Information on movie trailers come from the websites `www.hd-trailers.net`, `www.traileraddict.com`, and `trailers.apple.com` (iTunes). The coverage for these sites are 2007-2009 for HD-Trailers, 2007-2009 for Trailer Addict, and 2003-2007 for iTunes (for the latter I scraped `archive.org`, the Internet Archive, to get older versions of the page). The files include the list of trailers and their release date for each title; TrailerAddict includes number of online views and various measures of trailer quality (the latter are in terms of letter grades which are recoded as A=4, B=3, C=2, D=1, F=0). There are several trailers for each movie. Trailers are not available for about a fifth of the titles, largely reflecting gaps in the coverage of these sites. I am in the process of assembling data from two additional sites, `www.spike.com` and `www.moviefone.com`, which will improve coverage and also add additional measures of viewing.

5 Results

5.1 HSX and New Information: Impact of Trailer Releases

Before turning to the estimates it is important to show that HSX tracking stocks rapidly impound new information. I will see how the prices respond to the release of trailers, which are typically the first view of the movie and as such provide important information about the revenue which the movie might generate. If the markets are forward looking, prices should respond to this new information and (on average) become a better forecast of the revenue total at expiry. A secondary reason for looking at trailer releases is that they are roughly comparable to file sharing pre-releases: both give the first look at a version of the final product and so HSX response should be roughly of the same magnitude.[9]

Table 4 shows the impact of trailer releases on HSX prices (as discussed in Section 4 there is no trailer data for some movies, and this will be updated in the next draft).

[9] There are of course differences between trailers and pre-releases: the amount of the movie available (a snippet or the entirety), the quality of the copy (perfect or a low quality duplicate), and the timing (several months or a few weeks before the theatrical release). Note that Elbserse (2007) shows how news from early in production process (casting announcements) improves the forecast accuracy of HSX prices. The results here build on her results by looking at news with different timing and content.

I restrict the sample to the year prior to the theatrical release date, since the stocks are relatively lightly traded and volatile in earlier periods and almost no new trailers are released after a movie is released in theaters.. Panel (a) presents results for the first trailer released for each movie. The first column shows that prices tend not to move consistently up or down on the release date. The next two columns show that well-received trailers, as measured by user ratings and views, experience significant price increases. For example, a top rated trailer has an eleven point price increase on its release date while a worst rated one (the omitted category) has a six point drop. The next two columns further show that stocks respond to trailers through large price changes and trading volume. The last column looks at a measure of forecast error, mean absolute error (error in HSX price relative to its final expiry value),

$$MAE\ Price\ Error_{i,t} \equiv \|HSX\ Expiry\ Value_{i,t} - HSX\ Price_{i,t}\| \tag{4}$$

Smaller values indicate the market is doing a better job at forecasting. The estimates show that after a trailer is released the error is persistently smaller.

Panel (b) of **Table 4** repeats these estimates but focuses on subsequent trailer releases. At this point HSX traders have already seen the first trailer and so the new trailer provides limited information about the revenue potential of the movie. Consistent with this the parameter estimates for columns two through six are all smaller than (though of the same sign as) the ones in panel (a) and often statistically insignificant. In particular there is only a small reduction in forecast error accompanying a trailer release.

All these results are consistent with traders actively observing trailers and extracting from them information about the eventual box office success of the movies. Of interest for this paper, it indicates that expiry-relevant news is impounded in prices in a manner which improves their accuracy. The estimates here will also be helpful in benchmarking the impact of the magnitudes discussed below.

5.2 Motivating Example: Leak of *X-Men Origins: Wolverine*

While almost no high quality copies of movies arrive on file sharing networks prior to their theatrical release, in April 2009 a workprint version of the movie "X-Men Origins: Wolverine" was made available a full month before its release in theaters. The movie was downloaded about 1m times on BitTorrent in its first week of availability, 4.5m times by the time it was released in theaters, and 7.2m times over all 2009. The Wolverine case provides a unique opportunity for studying the impact of file sharing on movies. It provides a worst-case scenario in which a high quality copy of the movie was readily available prior to its theatrical showing. The extensive media coverage of the movie's leak probably meant virtually all the potential theatergoers were aware they could download it in advance. If individuals use file sharing to substitute for paying to see a film, it should be apparent in this case.

Figure 5 shows the evolution of HSX price for Wolverine's stock. There was little change in the stock price when the movie became available for illicit download, and a week later price began to rise continuously until the theatrical release date (trading volume, listed on the bottom of the figure, spiked on the day of the leak).[10] The movie also appears to have been relatively successful with a worldwide box office of $373m and 40-50m tickets sold. There is little evidence from this case study that an early release on file sharing networks has a significant negative effect on movie revenues.

5.3 Main Estimates

The first step is to show that HSX tracking stocks are responsive to file sharing releases.[11] **Table 5** displays the volume response. The first column shows that the daily number of shares traded increases by an amount which is double the typical average volume. The next column presents a placebo test by using file sharing dates from Dupe Sites which were shown to be noisy and imprecise. The effect is no longer economically or statistically significant. The third column shows that the volume response is reduced when the release was anticipates as measured by p_t the conditional release probability on that day. The interaction term suggests there would be no impact on volume when the release is fully anticipated $p_t = 1$ (though this is far out of sample since the largest p_t value is 0.28). p_t does not have a direct impact on volume. Columns 4-6 show that the same results hold on dollar-weighted volume. These results together suggest that file sharing releases, particularly unanticipated ones, lead to increases in trading activity.

[10]The drop in HSX price in the month prior to the leak was due to a poorly received trailer for the movie. There was also little price change around the leak date for contracts linked to the movie's box office futures on the real money Intrade site (though these markets were not very liquid).

[11]There is right censoring of the 7% of movies in the sample which do not have a file sharing release prior to their HSX stock expiry. This does not cause any issues for the empirical approach.

The other evidence about market responsiveness is in **Figure 6**. It shows the root mean square error of prices relative to their expiry,

$$RMSE_t \equiv \sqrt{N_t^{-1} \sum_{i=1}^{N_t} (HSX\ Expiry\ Value_{i,t} - HSX\ Price_{i,t})^2} \tag{5}$$

by days until the theatrical release, t. The top line is for all movies in the sample, and it shows that the average forecast error is relatively constant in the month until the theatrical date, and then it quickly falls as more of the box office returns are revealed. The bottom line shows that movies which are released on file sharing networks prior to the theatrical date have substantially less error (the differences is comparable to that associated with the first trailer release, **Table 4**). This error during the pre-theater period is relatively constant and is comparable to the error right after the theatrical release for the full sample. In the post-theater period, the two lines converge. This is consistent with the markets impounding the information from the file sharing release which is observing the actual movie but not any of the actual box office returns. This information is valuable in the pre-theater but not the post-theater period. These results should be viewed with some caution due to the possible dynamic sample selection as the set of movies with a file sharing release changes over time and early releases may differ in other ways. Still they suggest that market prices respond and become more informative after a file sharing release.

With evidence that markets respond in an informative way to file sharing releases, I now turn to the estimates of the displacement of revenues (3).[12] **Table 6** contains the initial results where I have assumed for simplicity that the effect is constant for all pre-theatrical periods and zero thereafter. This specification measures the displacement of a pre-release relative to having a post-theatrical release. The main result is that the effect is small, less than one percent of the average initial-run box office (HSX expiry value). Note that the estimates are relatively precise with small standard errors. This result holds whether time and movie fixed effects are included, and for various sub-samples such as restricting the observation period to the first few years, movies with wide initial releases, or movies in certain genres. Of particular interest is column three which focuses on a balanced panel of movies which are observed over the entire observation period. The comparable crowd-out effect here and with the

[12] I had wanted to look at the theatrical revenue of high quality (DVD-quality) file sharing releases. But only four percent of titles have high quality copies by the theatrical release date, and this rises to only twenty one percent by the typical four week HSX expiry date. I have been unable to obtain precise estimates given this release distribution. Future drafts will include more details on this point.

other sub-samples is evidence against the concern that the parameters are biased due to dynamic sample selection and heterogeneous effects (**Section 3.2**).

Table 7 presents the main results with separate displacement estimates for each date relative to the theatrical release, δ_t. Due to space limits only a subset of these parameters are presented. The first four columns show the estimates using the conditional release probabilities p_t for movies from that release year. In all cases the parameters are small in magnitude and precisely estimated. The most negative estimate, for movies reaching file sharing networks a month before their theatrical release (4.5% of titles are released by this time), implies a crowd out of \$2m or four percent of the average HSX expiry value of \$51m. The other parameters are much smaller and a null of no effect cannot be rejected in most cases. There is also a positive effect for movies released five days before the theatrical release. These results are robust to the presence of day and movie fixed effects and to looking at just the first half of the observation window. It is also not possible to reject a null that the parameters are unchanged if the sample is restricted to movies which do not have a file sharing release until at least four weeks after the theatrical release; there are some parameter changes for very early releases (when the restricted movie sample is notably smaller), but the difference flips sign. Consistent differences would have indicated heterogeneous effects.

The last three columns of **Table 7** use a more refined measure of release probabilities, $p_{i,t}$. **Table 8** presents estimates of a Cox proportional hazard model of the release time relative to the theatrical release date. There are three categories of covariates. There are significant differences across the producing studio, for example the hazard is 57% higher at Paramount than Buena Vista (the omitted category). The movie release year also has variation with the estimated hazard ratios falling almost monotonically (2003 is the omitted year). Finally the daily number of non-movie releases (in thousands and de-trended) has a positive effect. This covaiate varies across time, since the theatrical open date differs across movies. The hazard estimate is used to calculate a fitted measure of $p_{i,t}$ (**Section 4**). The resulting estimates of crowd-out in **Table 7** generally match those from using the observed $p_{i,t}$ with the main differences being the reduced magnitudes for early and late releases.[13] **Figure 7** plots the full sample parameter estimates for the observed and fitted release probabilities, along with a lowess smoothed curve. In both cases the values are close to the zero (no

[13]Since $p_{i,t}$ are generated regressors, standard errors are bootstrapped.

crowd-out) line, though there are negative effects for very early releases and positive effects for releases right before the theatrical open.

Table 9 uses the estimates to calculate the aggregate crowd-out due to file sharing. For each movie, I used the estimated δ_t (from the full sample, fitted $p_{i,t}$ specification) for its observed release date. I then sum these values up over all movies to get the aggregate crowd-out effect. I bootstrap across the movies to create a standard error. The effects are small. While the movies make \$7b-\$9b per year prior to the expiry of their HSX stocks, the annual crowd out is never more \$0.3b. Summed over all years, the crowd out is \$0.2b or about 0.3% of observed revenue. I cannot reject the null that this effect is zero.

6 Conclusion

This paper uses movie tracking stocks to measure the theatrical revenue displacement of file sharing. These stocks are forward looking and their forecasting performance improves when new information arrives such as the release of movie trailers. The empirical strategy considers how stock prices respond to news about file sharing, using both arrivals and non-arrivals as shocks. Because the approach exploits price variation for all movies which are unavailable on file sharing networks, the case for unbiased estimates is particularly strong for the period prior to the theatrical opening. The estimates indicate that the displacement effect is quite small, both on a movie-level and in aggregate. The effect is precisely estimated. This is perhaps not surprising given the low quality of early file sharing releases and the lack of amenities such as theater sound and video systems.

One consistent result is that file sharing arrivals shortly before the theatrical opening have a modest positive effect on box office revenue. One explanation is that such releases create greater awareness of the film. This is also the period of heaviest advertising. In conjunction with the main estimates, this suggests that free and potentially degraded goods such as the lower quality movies available on file sharing networks can have some beneficial effects on intellectual property.

References

[1] Athey, Susan and Scott Stern (2013). "The Nature and Incidence of Software Piracy: Evidence from Windows." In *Economics of Digitization*, University of Chicago Press.

[2] Byers, Simon, Lorrie Cranor, Dave Korman, Patrick McDaniel, and Eric Cronin (2003). "Analysis of Security Vulnerabilities in the Movie Production and Distribution Process." *DRM '03: Proceedings of the 3rd ACM workshop on Digital rights management*. p 1-12.

[3] Cuevas, R., M.Kryczka, A.Cuevas, S.Kaune, C.Guerrero, and R.Rejaie (2010). "Is Content Publishing in BitTorrent Altruistic or Profit Driven?" *ACM CoNEXT, Proceedings of the 6th International Conference.*

[4] Danaher, Brett and Joel Waldfogel (2012). "Reel Piracy: The Effect of Online Film Piracy on International Box Office Sales." SSRN WP #1986299.

[5] Elberse, Anita (2007). "The Power of Stars: Do Star Actors Drive the Success of Movies?" *Journal of Marketing*. 71:4. 102–120.

[6] Ma, Liye, Alan Montgomery, Param Singh, and Michael Smith (2013). "The Effect of Pre-Release Movie Piracy on Box-Office Revenue." SSRN WP #1782924.

[7] Oberholzer-Gee, Felix and Koleman Strumpf (2009). "File-Sharing and Copyright." In *NBER's Innovation Policy and the Economy series*, volume 10. ed. Joshua Lerner and Scott Stern. MIT Press.

[8] Pouwelse, J., P. Garbacki, D. Epema and H. Sips (2004). "A Measurement Study of the BitTorrent Peer-to-Peer File-Sharing System". Technical Report, Delft University of Technology

[9] Price, David (2013). "Sizing the Piracy Universe." Netnames. `http://copyrightalliance.org/sites/default/files/2013-netnames-piracy.pdf`.

[10] Pennock, David, S. Lawrence, C. Giles and F. Nielsen (2001). "The real power of artificial markets." *Science*. 291:5506. 987–988.

[11] Zhang, Chao, Prithula Dhungel, Di Wu, and Keith Ross (2011). "Unraveling the BiTorrent Ecosysytem." *IEEE Transactions in Parallel and Distributed Systems.* 22:7.

Table 1: **Pre-Release Movies Are Different**

Characteristic	Pre-Release	Not Pre-Release	Pearson χ^2 or Komogorov-Smirnov
Studio (major)			0.011
Theaters, opening (thousands)	2.569	2.494	0.072
Genre (main)			0.000
IMDB rating	6.35	6.14	0.014
IMDB rating: # votes (10^3 votes)	97.44	73.18	0.002
Rotten Tomatoes: tomatometer (critics)	52.11	46.80	0.042
Rotten Tomatoes: user meter (users)	63.69	61.95	0.012
Rotten Tomatoes: # user reviews (10^6users)	1.755	1.028	0.181

A pre-release movie is one which is available for un-authorized viewing on a file sharing network prior to its theatrical release in the United States. The sample is the top 150 grossing movies in the United States in each year over the period 2003-2009 (N=1057 movies). The first two columns show average values for each group of movie, while the last column is a test of the null that the distributions are identical for the two groups of movies; the Pearson test is used for categorical variables and the Kolmogorov-Smirnov for continuous variables (probability values are reported).

Table 2: **File Sharing Releases: Data Sources**

Source	Coverage	Features
Major BitTorrent Index Site	2003-2009	- all torrent uploads (filename, metadata, user rating) - all torrent downloads
Internal Release Group databases	2003-2009	initial posting date of all torrents on the group's private server
Public "Dupe Sites" NFOrce/NFOHump VCDQuality Scene Releases ORLYDB	2003-2009	initial posting date of all torrents on private scene release servers

The sources are listed in decreasing order of importance. The empirics are based just on using the BitTorrent Index site. The Internal Release Group databases and Public Dupe sites are only used to check the reliability of file sharing release dates used in the literature.

Table 3: Summary Statistics

Variable	Mean	Max	Min	Std Dev
Movie characteristics (N=1057 movies)				
Box office gross (millions $)	61.051	744.275	3.707	68.84
Box office, opening weekend (millions $)	17.545	158.412	0.002	19.85
Theaters, max (thousands)	2.494	4.455	0.021	0.95
Theaters, opening (thousands)	2.312	4.366	0.001	1.16
Major studio release?	0.711	1	1	0.45
IMDB rating	6.196	9	1.9	1.16
IMDB rating # votes (thousands)	78.327	96.346	0.045	90.36
Rotten Tomatoes: tomatometer (critics)	48.187	100	0	27.24
Rotten Tomatoes: user meter (users)	61.874	97	0	16.22
Rotten Tomatoes: # user reviews (millions)	1.139	3.465	0.000	0.51
File sharing release date (N=1047 movies)				
First copy (relative to theatrical)	5.223	778	-637	64.39
First high quality copy (relative to theatrical)	68.929	778	-637	72.12
First high quality copy (relative to home video)	-72.301	17	-1343	83.51
HSX (N=90214 title-days)				
price (daily close)	47.665	444.40	0.23	49.12
expiry price	51.972	441.63	1.62	55.51
volume traded (daily, millions shares)	4.640	363.234	0	8.325
volume dollars (daily, millions $ traded)	302.713	155531.20	0	1317.68
Trailer releases (N=1291 trailers)				
number trailers per movie	2.869	1.91	1	11
time until theatrical open (days, first trailer only)	134.899	77.33	0	365
ratings	2.807	0.89	0	4
number views (thousands)	13.443	34.83	1.574	1001.144

The sample is the top 150 grossing movies in the United States in each year over the period 2003-2009. All revenue and theater totals are for North American theaters.

Movie characteristics: IMDB (`http://www.imdb.com`) ratings are user submitted ratings of movie titles. Rotten Tomatoes (`http://www.rottentomatoes.com`) is an aggregator of professional critics as well as user ratings of movie titles.

File sharing release date: number of days relative to first US theatrical release or home video release. High quality is a copy of DVD-quality or higher.

HSX: Hollywood Stock Exchange data is for data 50 days before theatrical release though expiry (other data is available).

Trailer releases: some titles missing. restricted to trailers released in the year prior to the theatrical open. Ratings and number of views are from users at `www.traileraddict.com` and for 2007-2009 only.

Table 4: Impact of Trailer Releases

(a) initial trailer

Dependent Variables (HSX)	Price	Price	Price	$\|\triangle \text{Price}\|$	Trading Volume	MAE Price Error
$\mathbb{I}$(trailer in t)	1.057	-5.870	-6.487	11.016	8.784	
	(0.71)	(0.99)	(1.05)	(0.16)	(0.33)	
$\mathbb{I}$(trailer in t) $\times$Trailer Rating		4.206				
		(1.03)				
$\mathbb{I}$(trailer in t) $\times$Trailer Num Views			0.548			
			(0.26)			
$\mathbb{I}$(trailer in $s \leq t$)						-10.046
						(1.19)
N	276589	126789	126789	276589	276589	276589
R^2	0.72	0.82	0.79	0.51	0.71	0.84

(b) later trailers

Dependent Variables (HSX)	Price	Price	Price	$\|\triangle \text{Price}\|$	Trading Volume	MAE Price Error
$\mathbb{I}$(trailer in t)	0.789	-1.325	-1.008	1.578	1.256	
	(1.05)	(1.14)	(1.97)	(0.11)	(0.45)	
$\mathbb{I}$(trailer in t) $\times$Trailer Rating		0.512				
		(0.67)				
$\mathbb{I}$(trailer in t) $\times$Trailer Num Views			0.297			
			(0.19)			
$\mathbb{I}$(trailer in $s \leq t$)						-1.237
						(0.86)
N	276589	126789	126789	276589	276589	276589
R^2	0.69	0.73	0.73	0.41	0.62	0.70

The sample is the top 150 grossing movies in the United States in each year over the period 2003-2009 (except for cols 2 and 3 which are 2007-2009 only). The observation period is the year before theatrical release. All specifications include title and days until theatrical release fixed effects, and robust standard errors are in parentheses. In column 2 the omitted category is the lowest grade. Preliminary: some movies are omitted due to lack of trailer data; they will be added using additional sources listed in the main text.

Table 5: Impact of File Sharing Release on HSX Volume

Dependent Variables (HSX)	Trading Volume			Dollar Volume		
$\mathbb{I}(t)$	8.621		10.124	575.036		714.170
	(0.63)		(0.68)	(73.42)		(72.94)
$\mathbb{I}(t)_{Dupe\,Site}$		0.975			37.456	
		(0.84)			(94.16)	
p_t			0.648			49.961
			(1.46)			(243.61)
$\mathbb{I}(t)\times p_t$			-12.456			-1246.322
			(5.77)			(667.87)
μ_i	Y	Y	Y	Y	Y	Y
ν_t	Y	Y	N	Y	Y	N
N	83069	83069	83069	83069	83069	83069
R^2	0.40	0.34	0.30	0.51	0.45	0.38

The observation period is fifty days before theatrical release though expiry of HSX tracking stock. Robust standard errors in parenthesis. $\mathbb{I}(t)_{Dupe\,Site}$ is files sharing release date based on the Dupe Site (it is measured with noise, and is used a a placebo test). p_t is based the observed distribution of release dates. μ_i are movie fixed effects and ν_t are day fixed effects. Day fixed effects cannot be included in specifications with p_t, since it does not vary across movies.

Table 6: Initial Estimate (3)

The specifications have a constant pre-theatrical effect and no post-theatrical effect:

$$\delta_t = \begin{cases} \delta & t \leq 0 \\ 0 & t > 0 \end{cases}$$

Dep. Var.	$\triangle S_t$							
	All	All	$\mathbb{I}(t \geq 28)$	2003-2005	Wide	Action	Comedy	Drama
δ	0.457	0.362	0.658	0.592	0.542	0.438	0.688	-0.115
	(0.22)	(0.21)	(0.88)	(0.24)	(0.46)	(0.45)	(0.36)	(0.48)
μ_i	N	Y	Y	Y	Y	Y	Y	Y
ν_t	N	Y	Y	Y	Y	Y	Y	Y
N	58254	58254	10342	23892	17575	11762	17328	9443
R^2	0.10	0.19	0.28	0.22	0.19	0.16	0.25	0.31

The observation period is fifty days before theatrical release though the file sharing release date for each movie. Robust standard errors in parenthesis. In (3) $p_{i,t}$ is based the observed distribution of release dates by release year. μ_i are movie fixed effects and ν_t are day fixed effects. The last six columns are for sub-samples: movies which do not reach file sharing networks until at least four weeks after the theatrical release, releases in the first half of the observation period, wide opening (movies with at least three thousand screens in their first weekend), and titles in certain genres.

Table 7: Main Estimate (3)

Dep. Var.	$\triangle S_t$						
		Observed $p_{i,t}$			Fitted $p_{i,t}$		
	All	All	$\mathbb{I}(t\geq 28)$	2003-2005	All	$\mathbb{I}(t\geq 28)$	2003-2005
δ_{-30}	-1.387	-1.856	-2.037	-0.905	-0.679	-0.845	-0.689
	(1.74)	(1.38)	(2.41)	(0.11)	(1.52)	(2.31)	(0.55)
δ_{-25}	-0.673	-0.616	-1.834	-0.610	-0.379	-0.158	-0.354
	(0.25)	(0.12)	(0.53)	(0.13)	(0.25)	(0.79)	(0.27)
δ_{-20}	0.305	0.322	1.755	0.170	0.345	-0.573	-0.185
	(0.20)	(0.16)	(0.71)	(0.17)	(0.22)	(0.69)	(0.29)
δ_{-15}	-0.363	-0.533	-0.336	0.439	-0.379	-0.268	0.648
	(0.94)	(1.01)	(1.41)	(0.23)	(0.87)	(1.25)	(0.17)
δ_{-10}	0.372	0.315	0.631	-1.076	0.146	-0.246	-0.009
	(0.42)	(0.31)	(0.47)	(0.31)	(0.39)	(0.50)	(0.30)
δ_{-5}	3.056	2.896	2.704	3.029	1.975	1.360	2.546
	(0.63)	(0.59)	(1.42)	(0.50)	(0.89)	(2.41)	(0.46)
δ_0	1.138	0.998	1.284	1.611	0.942	0.798	1.346
	(0.48)	(0.46)	(0.89)	(0.59)	(0.39)	(0.99)	(071)
δ_5	-0.107	-0.108	-0.306	-0.458	0.278	0.189	-0.146
	(0.74)	(0.74)	(0.97)	(0.59)	(0.95)	(1.10)	(0.35)
δ_{10}	-0.076	0.145	0.287	-0.210	0.008	-0.169	0.002
	(0.33)	(0.33)	(0.65)	(0.60)	(0.45)	(0.56)	(0.71)
δ_{15}	-0.457	-0.345	-0.302	-0.352	-0.079	-0.198	-0.241
	(0.16)	(0.16)	(0.33)	(0.31)	(0.22)	(0.31)	(0.39)
δ_{20}	-0.934	-0.918	-1.168	0.276	0.115	-0.165	0.287
	(0.85)	(0.79)	(0.75)	(0.12)	(0.56)	(0.66)	(0.44)
μ_i	N	Y	Y	Y	Y	Y	Y
ν_t	N	Y	Y	Y	Y	Y	Y
N	58254	58254	10342	23892	58254	10342	23892
R^2	0.17	0.35	0.31	0.37	0.30	0.34	0.27

The observation period is fifty days before theatrical release though the file sharing release date for each movie. Robust standard errors (bootstrapped standard errors for Fitted specifications) in parenthesis. μ_i are movie fixed effects and ν_t are day fixed effects. There are separate parameters for $\delta_{-50}, \delta_{-49}, \ldots, \delta_{\max expiry\, time}$ and most are omitted due to space constraints. In the Observed columns $p_{i,t}$ is based the observed distribution of release dates by release year in (3), while in the Fitted columns fitted values from the hazard in Table 8 are used.

Table 8: Hazard Estimate of File Sharing Release Day

Coefficient Estimates					
Studios		**Year**		**Releases**	
Dreamworks	1.853	2004	0.820	# Non-Movies	1.235
	(0.42)		(0.11)		(0.20)
Focus	0.461	2005	0.663		
	(0.09)		(0.09)		
Fox	1.394	2006	0.737		
	(0.19)		(0.10)		
Lionsgate	1.627	2007	0.698		
	(0.35)		(0.09)		
MGM	1.067	2008	0.550		
	(0.23)		(0.07)		
Other/Minor	0.754	2009	0.495		
	(0.10)		(0.06)		
Miramax	0.952				
	(0.20)				
New Line	1.618				
	(0.28)				
Paramount	1.567				
	(0.23)				
Sony	1.451				
	(0.21)				
Universal	1.209				
	(0.16)				
Warner	1.103				
	(0.16)				
Weinstein	0.677				
	(0.17)				
N	1029				
$LogL$	-5395.85				

Cox proportional hazard model of time until movie is released on file sharing network, $\mathbb{I}(t)$. Movies are presumed to be at risk for release 50 days before their theatrical release. Coefficients reported as hazard ratio. Robust standard errors in parenthesis. Omitted categories: studio indicators=Buena Vista; release year=2003. Releases is a continuous, time-varying measure of the number of non-movie files posted to the BitTorrent index site each day (after de-trending, in thousands).

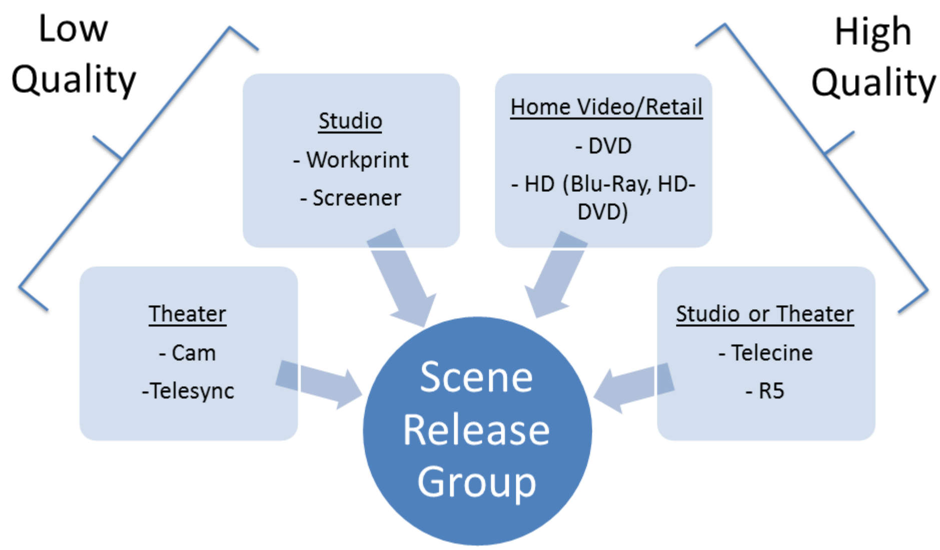
Table 9: Aggregate Impact of File Sharing Release (billions $)

Year	Observed Revenue	$\triangle$Revenue
2003	7.233	-0.256
		(0.35)
2004	7.294	-0.179
		(0.26)
2005	7.471	0.189
		(0.22)
2006	7.620	-0.095
		(0.39)
2007	7.969	0.041
		(0.17)
2008	8.163	0.084
		(0.65)
2009	8.851	0.058
		(0.53)
TOTAL	54.602	-0.158
		(1.23)

The observed revenues are the summed revenues for the movies at HSX expiry time. The change in revenues are the impact on revenues as based on the estimates in Table 7 using fitted $p_{i,t}$. Bootstrapped standard errors are in parenthesis below each change.

Figure 1: BitTorrent Ecosystem

(a) Release Groups



(b) Propagation of Files to Users

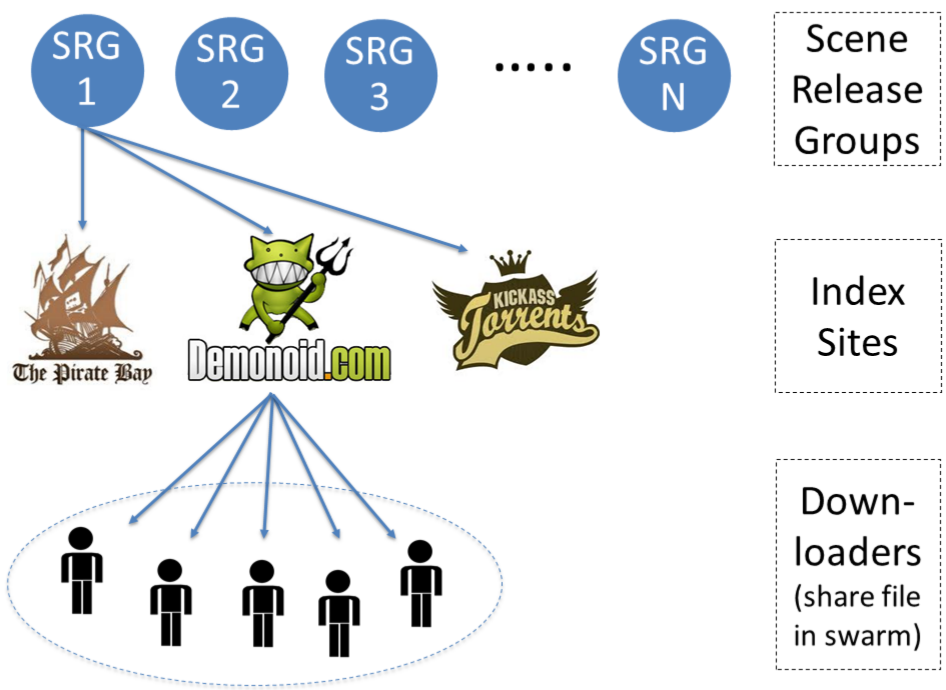
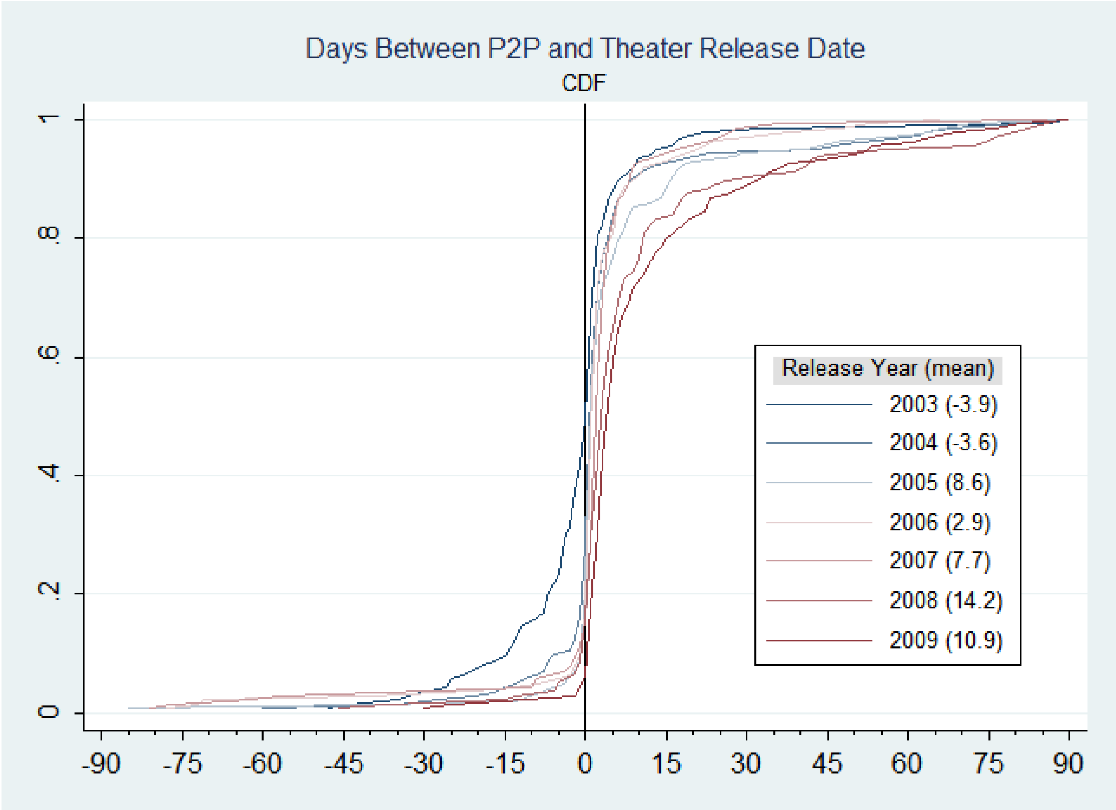


Figure 2: File Sharing Release Date Empirical CDF (Full Sample)

(a) Any Release



(b) High Quality

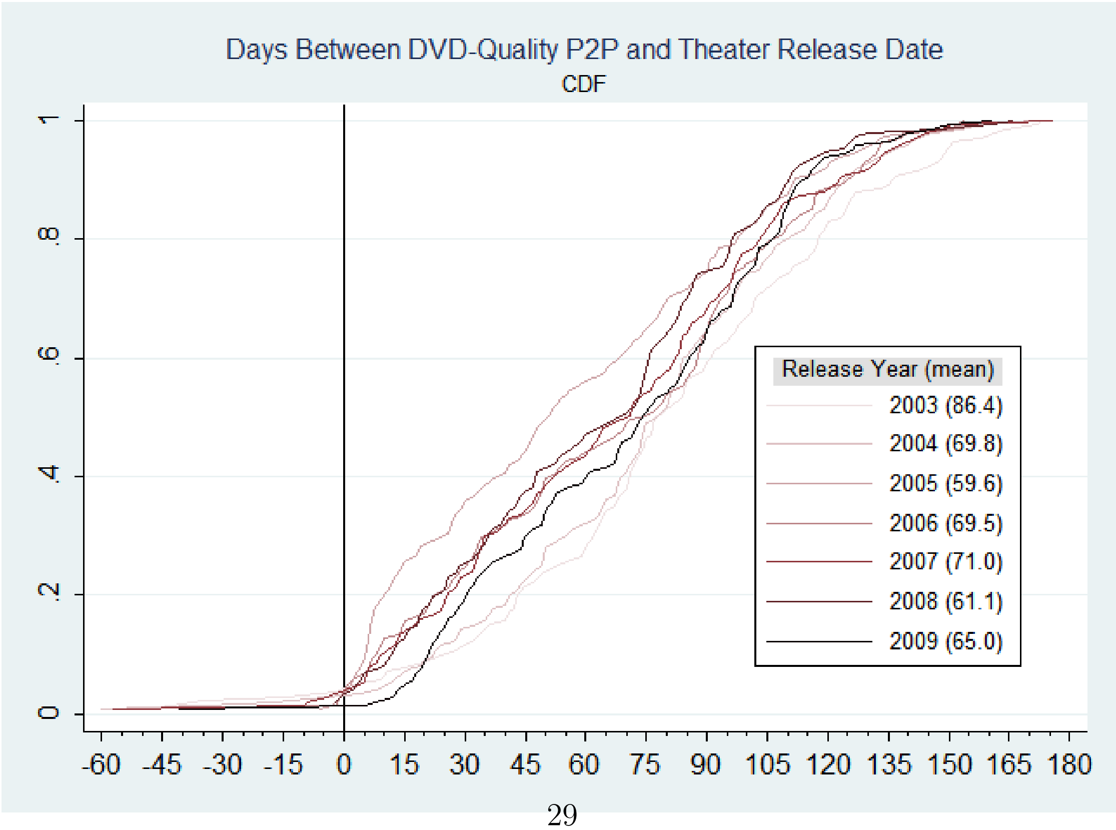
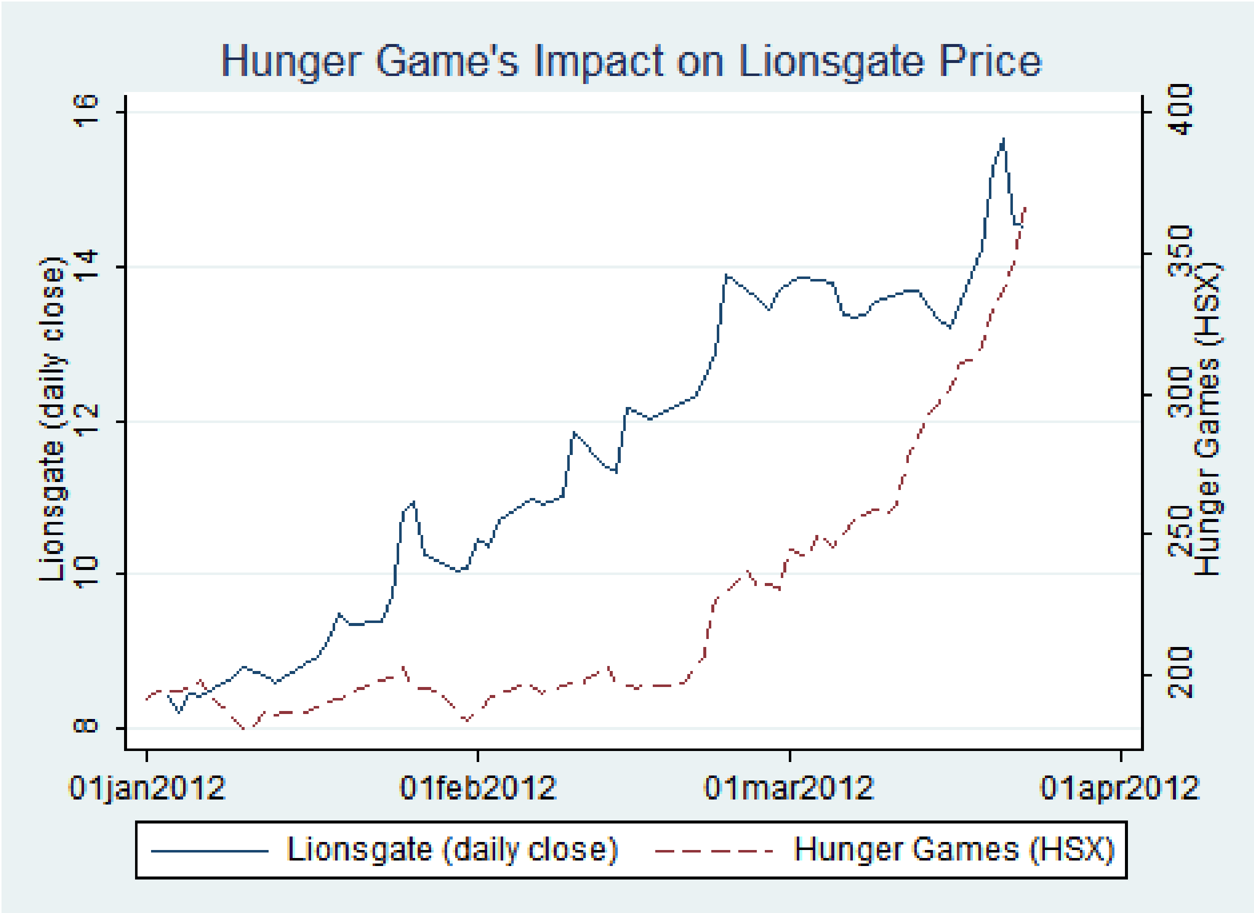


Figure 3: **HSX Prices Track Financial Stocks**

(a) Hunger Games



(b) Divergent

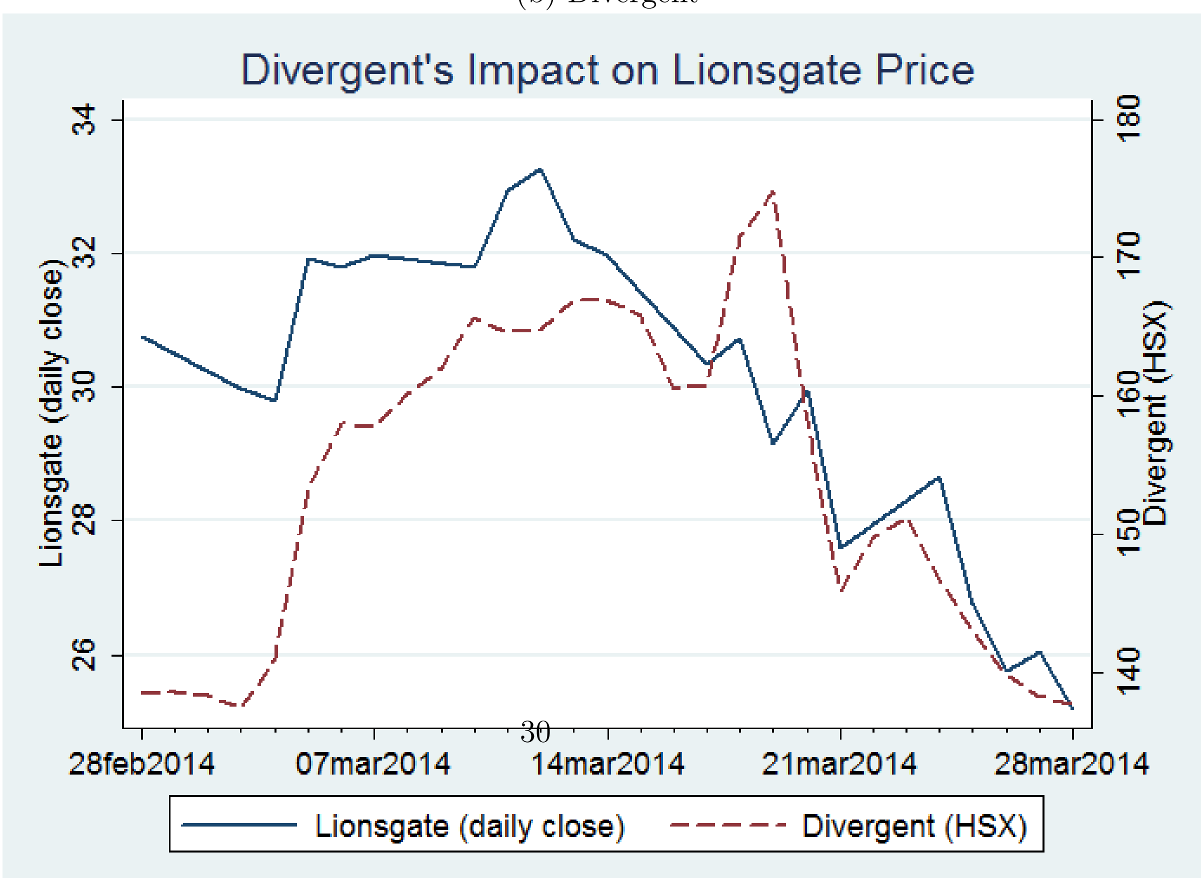
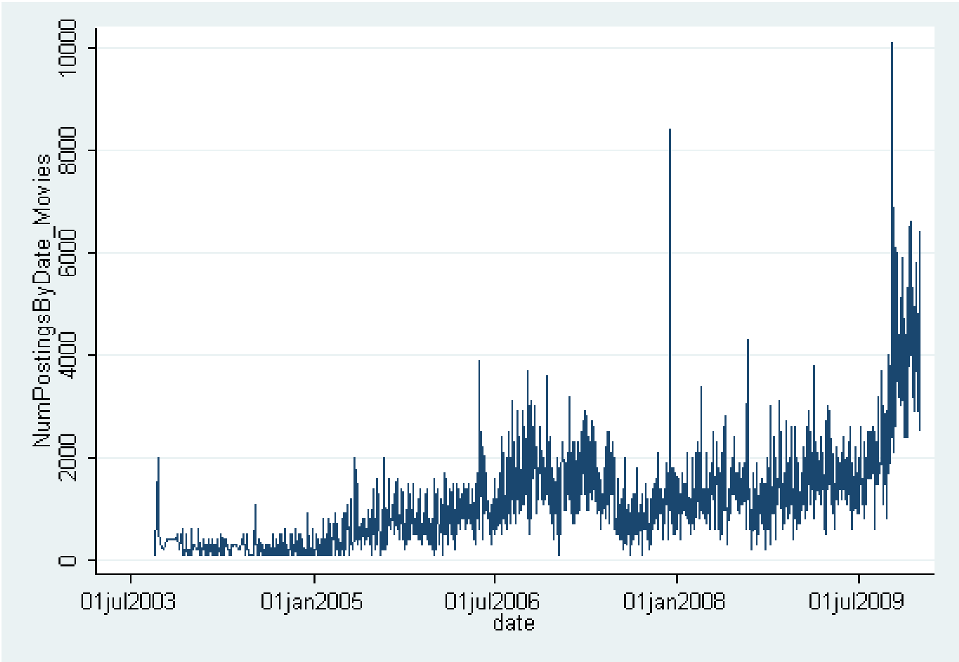


Figure 4: **Number Daily File Postings for Large BitTorrent Index Site**

(a) Movies



(b) Non-Movies

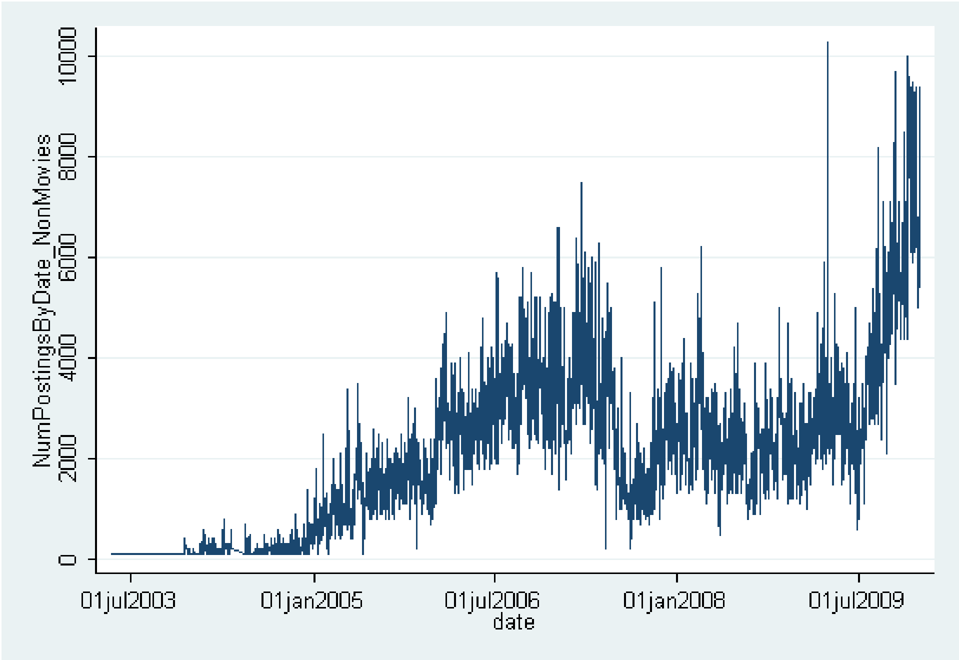
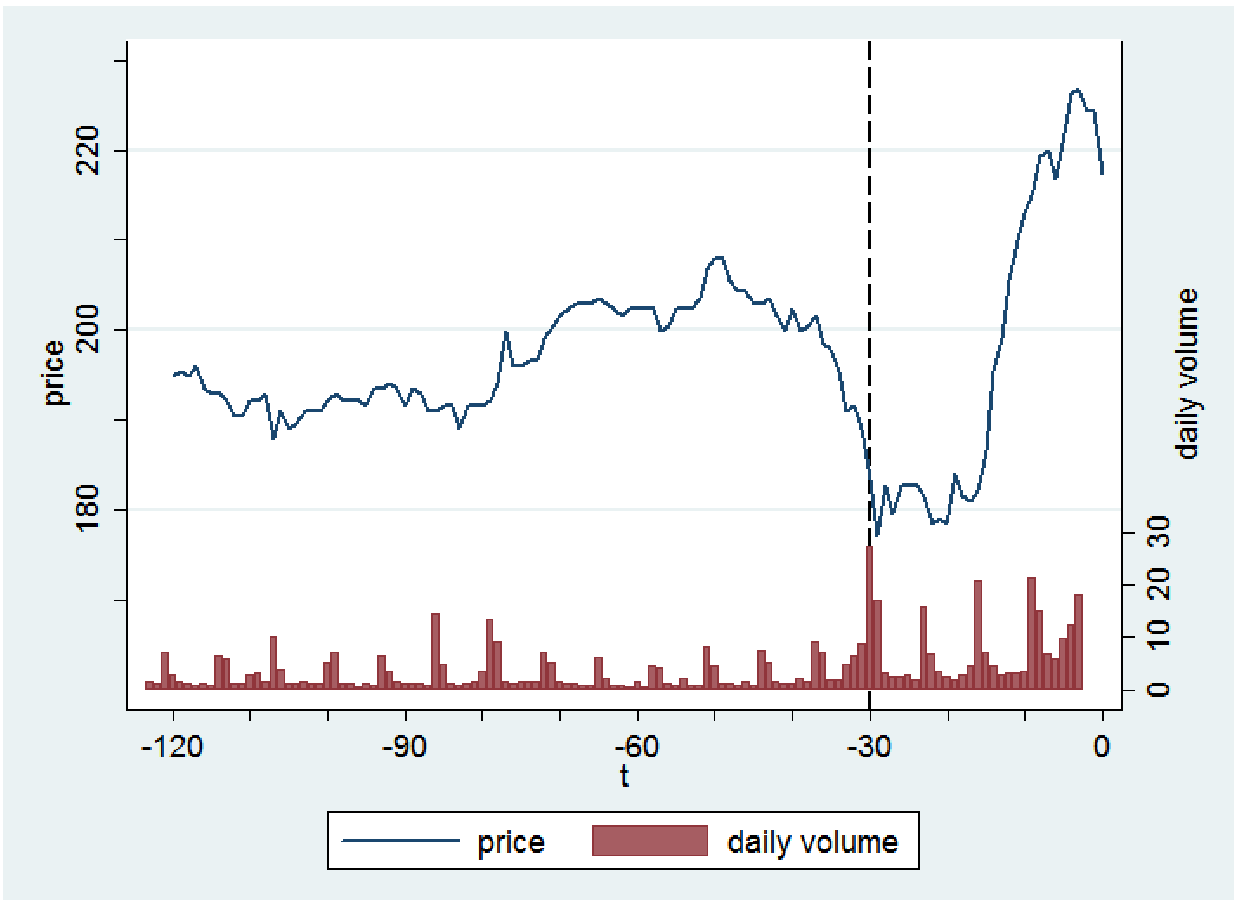


Figure 5: **Impact of *Wolverine* Leak on HSX**



The movie became available on file sharing networks on 1 April 2009 (t=-30), and the movie was released in theaters on 1 May 2009 (t=0).

Figure 6: **Impact of File Sharing Releases on HSX Price Forecast Error**

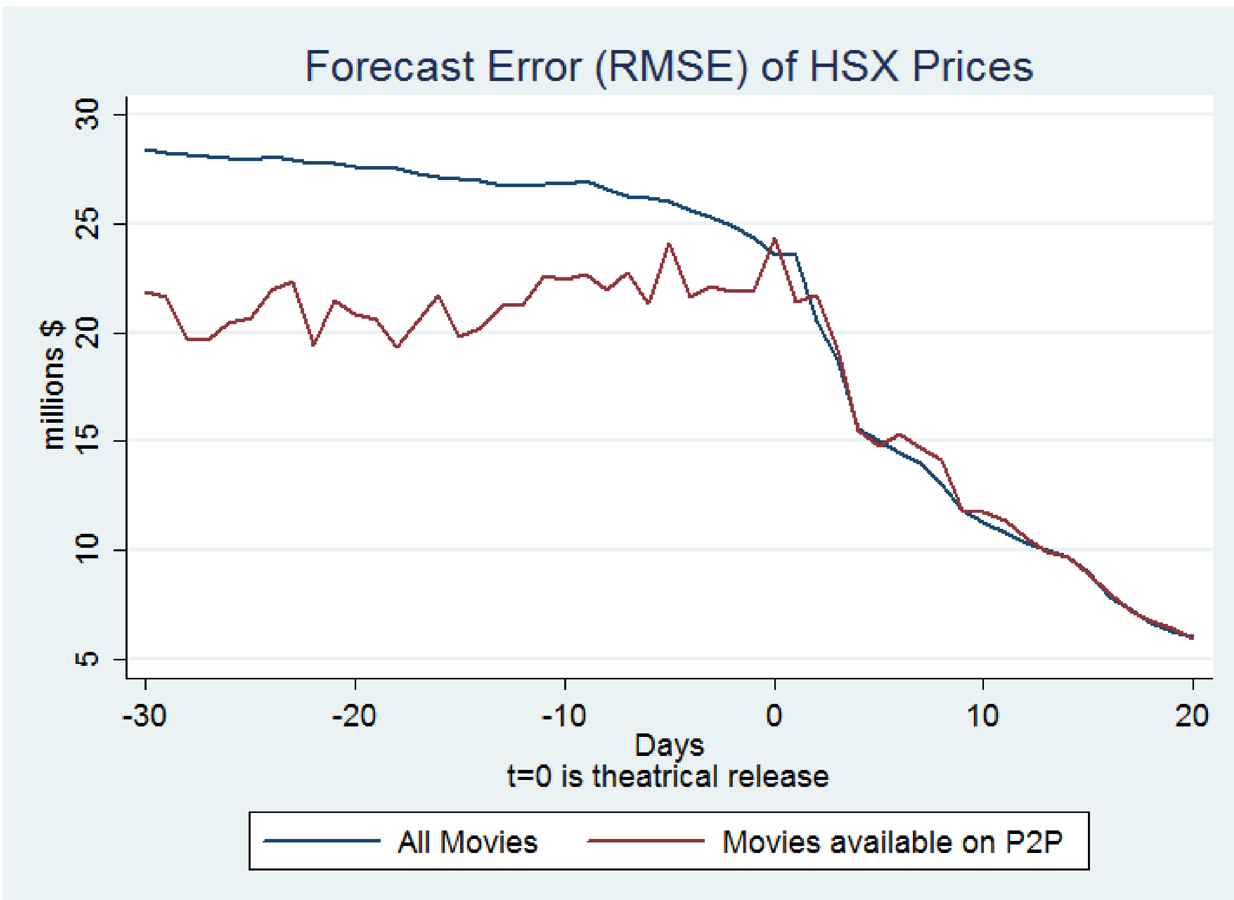
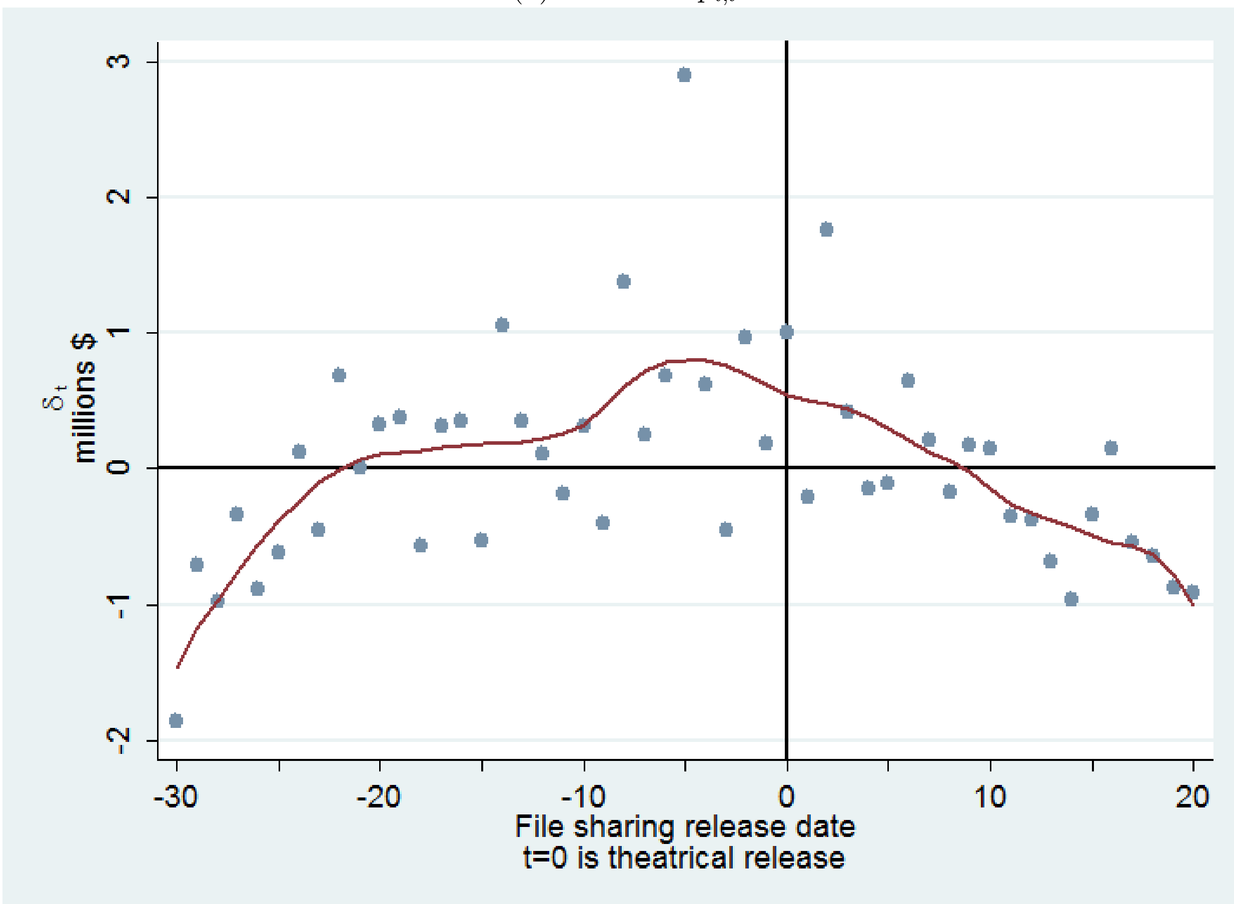


Figure 7: **Smoothed Parameter Estimates from Table 7**
Lowess (locally weighted scatterplot smoother) with bandwith = 0.25

(a) Observed $p_{i,t}$



(b) Fitted $p_{i,t}$