Estimating the effect of movie piracy on box-office revenue essay

This paper addresses these issues directly.

We develop and estimate a statistical model of the effects of piracy on the box-of? ce performance of a widely-released movie. The model discredits the argument that piracy increases sales, showing unambiguously that Internet piracy diminished the box-of? ce revenues of a widely released motion picture. The model overcomes a major weakness of counterfactual or “ but for piracy” methods widely used to estimate damages. These counterfactual methods violate the “ nobody knows” principle because they forecast what the movie would have earned in the absence piracy.

The model we present does not violate this basic principle of motion picture uncertainty. We estimate that pre-release and contemporaneous Internet downloads of a major studio movie accelerated its box-of? ce revenue decline and caused the picture to lose about $40 million in revenue. Keywords estimation Movie piracy · Nobody knows principle · Forensic revenue loss A. S. De Vany Department of Economics, University of California Irvine, Irvine, CA 92717, USA e-mail: ca W.

D.

Walls School of Business and Management, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong 292 A. S. De Vany, W. D. Walls JEL Codes L820 · Z110 · C800 I Introduction Copyright infringement and other forms of theft of intellectual property have become a large and growing problem around the world.

Even a small amount of infringement can generate enormous absolute losses to property rights holders if one recognizes that the intellectual capital of the S&P 500 companies is worth an estimated US$3. 4 trillion (Bowers 2001). The problem is acute for the motion picture industry.

It relies on foreign markets for the bulk of its sales; these are the markets where piracy is most common (Walls in press a). The Motion Picture Association of America estimates that motion-picture industry losses due to piracy exceed $3 billion annually in potential worldwide revenue (MPAA 2005). 1 A $3 billion revenue loss to an industry that takes in less than $10 billion annually is a serious loss.

Revenues lost to piracy are dif? cult for the motion picture industry to absorb because most movie projects are not profitable and a handful of extraordinary successes drive total pro? (De Vany and Walls 2004b; Vogel 1998). Since these hits are the movies most likely to be pirated, the impact on pro? t is greater than the impact on revenues would suggest. In 2000, the average major studio ? lm cost $55 million to produce and $27 million to advertise and market (MPAA 2005). A small revenue leakage from piracy can be magni? ed when the movie is released sequentially, ? rst to cinemas and then to other windows of distribution. If piracy reduces revenues early in the theatrical run of a motion picture, these early losses are magni? d later in the run and in subsequent venues because there are increasing returns to information in motion pictures (De Vany and Walls 1996; Walls 1997; Hand 2001; Maddison 2004).

Studios have begun to release their major movies in international markets simultaneously with the domestic release in order to limit the damage of piracy. This form of release makes it less likely that a copy pirated from an early release venue can steal revenues from later venues. The motion picture we analyze in this paper was simultaneously released in the US, UK, Spain, and Argentina.

In this paper we develop a model of piracy and revenue dynamics that – permits early detection of piracy, – directly tests the competing hypotheses regarding the effect of piracy on box of? ce revenue dynamics (+, ? , 0), and – does not violate the " nobody knows" principle. 2 1 Our use of the word " potential" is deliberate and mirrors the MPAA's use. The reason is that these estimates are highly indirect and require the counterfactual conjecture: What would sales have been but for piracy? 2 Screenwriter William Goldman's (1983) famous statement that " nobody knows anything" about how a movie will turn out at the box of? e has been veri? ed and rigorously developed as the stable Paretian hypothesis by a variety of authors (Walls 2005; De Vany and Walls 2004b; De Vany 2004; McKenzie 2003; Lee 1999).

Goldman's statement has been re? ned and restated by Richard Caves as the nobody knows principle: " That is, producers and executives know a great deal about what has succeeded commercially in the past and constantly seek to extrapolate that knowledge to new projects. But their ability to predict at an early stage the commercial success of a new ? lm project is almost nonexistent" (Caves 2000, p. 371).

Estimating the Effects of Movie Piracy 293 II Piracy and the Theatrical Movie Market Conceptually, piracy has several effects on supply and demand in the motion picture theatrical market.

On the supply side, piracy expands the number of sources of supply relative to the number of booked exhibitions; this alternative supply would ordinarily drain revenue from contracted theaters and lower their revenue per screen. 3 If wider exposure through pirates induces some viewers of the pirated product to purchase an admission in the theater, then piracy would increase demand for the movie.

Though this positive effect of piracy may seem remote, the argument has been made and it is dif? cult to reject or accept within standard approaches. We develop a direct test of this hypothesis once we have speci? ed the model. If pirate copies take revenue from theaters, then revenue per screen would decline. This would lead theaters to drop the pirated ? lm for a better prospect.

In addition, a pirated ? lm may fail to meet the holdover amount required to extend the run for another week. As a result, other exhibitors will be less likely to exhibit a pirated ? lm because of its low grosses.

Through these mechanisms, piracy would accelerate the rate at which a movie’s theatrical gross revenues decline and increase the rate at which it loses theaters through the weeks of its run. On the demand side, if piracy accelerates the decline of a movie’s theatrical gross, it will fall more rapidly in the highly competitive and closely followed contest for box-of? ce rankings.

When a movie declines in rank and revenue that event may be used as a signal by those potential viewers who key on rankings to choose among movies. 4 To the extent that piracy accelerates a movie’s box-of? e revenue decline its revenues in subsequent weeks are adversely affected. A rapid loss of revenue speeds the loss of theaters which, in turn, causes a further decline in revenue. Because motion picture box-of? ce revenues follow a recursive and non-linear dynamic, adverse signals can promote a rapid and growing decline in revenue and in? uence revenues in later weeks of the run as well.

5 If the pirated copies are poor or show an un? nished movie or work print that contributes to a poor assessment of the movie, then word-of-mouth transmission of this (biased) information would have a negative effect on theatrical grosses.

So, theory suggests a relationship of the following kind: Rt = F(Pt , t) (1) where Rt is the change in the movie’s theatrical revenue in week t and Pt is the number of pirate sources of the movie in week t. We expect the function F(Pt , t) to be decreasing in both Pt and t, which is to say that piracy ought to accelerate the rate 3 See De Vany and Walls (1997) for a description of the exhibition contract and how revenue affects movie survival at the cinema. See Walls (in press b) for an overview of current economic modeling of the movie industry.

See De Vany and Walls (1996) for a model of the dynamic demand process. 5 De Vany and Walls (2004a) model and estimate the dynamic mapping of revenues and show there is an expansive and contractive path. De Vany and

Lee (2001) simulate a model of quality signals in information cascades that yields similar results. 294 A.

S. De Vany, W. D. Walls at which revenues decline. If, on the other hand, piracy increases box of? ce revenues, then total revenues will rise and the weekly decline in revenue will be lower.

If piracy has no effect, the rate of change of weekly revenue with respect to piracy will be zero.

To summarize the economics, piracy adversely affects the dynamics of demand and supply in the theatrical market. The immediate effect on the dynamics is to increase the rate at which revenue and the number of screens showing the movie decline. The longer-term effect is to alter the information dynamics in such a way that demand in later weeks is also reduced.

Thus, a pirated movie will play off more rapidly and lose revenue at an accelerated rate during its run. We test this economic hypothesis by empirically modeling the rate of decline in weekly revenues in relation to the number of pirated prints available for Internet download.

III Estimating the Effect of Piracy on Box-Of? ce Gross Having speci? ed the theoretical linkage between piracy and revenues we now proceed to estimate it with data. The empirical model we estimate relates the change in revenues to the number of pirate copies available and the week of the run. Thus, Rt ? Rt? 1 = ? + ? Pt? 1 + ? t + . (2) Our measures of US box of? ce revenues are those reported by the distributor and match publicly available ? gures from Box Of? ce Mojo.

Our measure of pirate sources is based on a study commissioned by a major studio. 6 Data crawlers were sent out over the Internet to detect sites that had a ? e of the movie available for download. Numerous protocols were included into the search, including eDonkey2000, BitTorrent, Gnutella, FastTrack, Hotline, FTP, Usenet, and Internet Relay Chat (IRC). 7 The source of the ? les was an incomplete working print sent by the studio to an advertising agency. The DVD was apparently given by an agency employee to a person who loaded it onto one or more of the sites set out above. Detailed weekly data on pirate supply were collected over a seven week period beginning one week before the movie's release and continuing for six weeks into its theatrical release.

During the week before the movie even opened, there were 18, 134 pirate site days during which the print was available for download. During the ? rst week of the run, there were 15, 700 active pirate sites. This fell to 12, 034 in the second week and then grew to a peak of 28, 332 sites by the third week. Successive weeks then had 15, 210, 15, 806, and ? nally 8159 pirate sites. Over the seven weeks for which we have data, pirate sites expanded the available sources beyond those theaters 6 The unique data that allow us to address this topic were provided to us on a con? dential basis. 7 The movie ? e was distributed electronically over the many alternative channels set out above.

These included the peer-to-peer networks of eDonkey2000, BitTorrent, Gnutella, FastTrack, and Hotline that connect users directly; users query the particular system for a list of available ? les and then individually download the ? les they select. Other channels were the FTP ? le transfer protocol in

which a client can upload or download ? les to a central server, IRC internet relay chat where multiple users “ chat” in groups or privately and can download ? les, and Usenet where individual messages (including ? e attachments) are posted to newsgroups and downloaded by users. Estimating the Effects of Movie Piracy 295 legitimately contracted to exhibit it by a total of 131, 509 download source-days. 8 This corresponds roughly to 2, 700 active pirate sites per day for each of the ? rst six weeks of the run, a large number when compared to the movie’s 3, 660 opening theaters and 1, 319 theaters in its ? fth week. The studio commissioned download study estimates that there were 750, 330 peer-to-peer downloads of the movie, 50 FTP downloads and 1, 386 IRC downloads.

Usenet downloads are estimated to be 240, 000. So, there were about 1 million downloads.

At a $6 average admission price, the movie would lose $6 million in box of? ce receipts if each download displaced one admission. Of course, it is not as simple as that. Each downloaded copy may have been seen multiple times or by several people. So, the loss might be a multiple of the downloads.

On the other hand, the downloaders may not have intended to see the movie in a theater and only got a copy because it was free or somehow fun to see it in this manner. Perhaps seeing the pirated copy induced the viewer to go to a theater to see the ? nished movie with all the surroundings that a theater offers.

So, it is not clear what impact the downloads had or whether it was positive or negative. A further complication is based on the non-linearity of motion

picture demand shown by De Vany and Walls (2004a). They show that the expectation of revenue in future weeks is conditional on revenue earned in prior weeks of the run.

Higher revenue in early weeks translates into higher expected revenue in future weeks. Thus, early success implies future success. Early disappointment implies growing disappointment in later weeks. The demand mapping is non-linear which implies that piracy in early weeks has some impact on demand in later weeks.

Both theory and evidence suggest that Internet downloads will not displace theatrical admissions one for one. For a rough visual of the impact of the available download sites, in Figure 1 we have plotted the number of download days and theater days where the opportunity to view the movie was available by week of the run.

The pirate supply was almost as large as the legitimate supply on opening week (and was available two weeks before the opening) and surpasses the legitimate supply in most weeks during the early (high revenue) weeks of the run. Both supply sources closely tracked one another during the weeks of the run as well.

The way these supply sources track one another is evidence that pirate and legitimate outlets were competitive with one another. Because the download estimates require assumptions about bandwidth and utilization of the Internet sites on which the illegal copies of the ? lm were posted, we prefer to measure directly the number of sites instead.

This is a direct measure of the alternative supply of pirate copies available on the Internet that competes with the legitimately licensed supply in theaters. This is comparable to counting the number of theaters where the movie might be seen.

The hypothesis is that this alternative source of supply is competitive with the legitimately licensed supply and will affect demand at these legitimate theatrical sources. The data show that the power of the Internet to expand supply is enormous. The Internet sites made the movie available throughout the world even though it had only been released to theaters in the US, UK, Spain, and Argentina. 8 The daily supply counts were estimated separately for each internet distribution protocol based on the number of unique users—de? ned as a unique IP address—detected and documented by the web crawler as offering for download a copy of the movie ? e or a derivative of it.

296 40000 movie theater supply A. S. De Vany, W. D. Walls internet pirate supply Theater? Days and Internet? Site? Days 30000 20000 10000 0 1 5 10 14 Week of Run Fig.

1 Pirate and Theater Supply of Viewing Opportunities Table 1 Revenue Decline and Pirate Download Sites variable constant pirate supply time estimate ? 4670494 (5436410) ? 621. 8818 (214. 846) 463808. 4 (524752) 0. 7799 13 Note: Estimated standard errors are in parentheses R2 Observations Table 1 shows the regression estimates for the model relating US box-of? ce revenue decline to the number of pirate download sites.

The coef? cient on download sites is negative, implying that piracy reduced theatrical demand.

The coef? cient was not constrained in the estimation and, so, could have been negative, positive, or zero implying that piracy decreased, increased, or did not affect theatrical demand. The sign and significance of the piracy coef? cient indicate that piracy decreased theatrical demand. The estimated model is a good ? t to the data with an R-squared of about 0. 78.

The coef? cient on the week of the run tells us that weekly revenue declines $463, 808 with each week of the run. But this decline is accelerated by $621. 8 for each active pirate download site in the previous week. The effect of an active site is comparable to taking about 100 admissions from the total US theatrical revenue of the ? lm.

The previous model uses observations for all 14 weeks of the run (since we are taking ? rst differences of weekly gross, we have 13 observations). But, for the last eight weeks of the run we lack pirate data. This has the effect of undercounting the Estimating the Effects of Movie Piracy Table 2 Revenue Decline and Pirate Download Supply Variable Constant Pirate supply Time Note: Estimated standard errors are in parentheses R2 Observations 97 Estimate ? 2. 93e+07 (8357117) ? 248. 0071 (240.

982) 5279188 (1328348) 0. 9448 5 number of pirate sites is undercounted. Consequently, this equation may overestimate the effect of a pirate site on the movie’s theatrical gross. When we take this into account by using only the 6 weeks for which we have an accurate count of pirate sites to estimate the model, our estimates still show a negative impact on revenue of piracy but it is smaller as shown in Table 2. The estimates reported in Table 2 con? m the result of the previous model that piracy accelerates revenue decline,

though the point estimate differs on the effect of a supply site on the rate at which revenue declined.

Instead of $621. 88 revenue loss per site, we now have an estimated loss of revenue of $248 per active download site. Our estimates show that the movie’s early revenue decline, which was precipitous, was accelerated by piracy. Early in the run, the effect of piracy was considerable because there were many sites.

As sites declined during the run, piracy had a smaller effect. The movie’s revenue declined at an average rate of $4. 7 million per week. According to the above estimates, the rate of revenue decline was from $242 to $621 greater per pirate supplier. To estimate the total theatrical loss from piracy, we sum the effect by each week of the run.

The cumulative loss of theatrical revenue due to pirate download sites over the 14 week run is between $29. 2 million (using $242 per active site) and $49. 5 million (using $621 per active site). While the damage from a single site is not large, there are so many sites that the total impact is large.

There are many caveats to these estimates, particularly the limited number of observations.

For the brief time that motion pictures run in theaters, there is no way around this problem. A larger sample of movies would be helpful, but that would require a major effort by the MPAA and their member studios to collect these data. Finally, note that nowhere in these estimates is it required that a forecast of the movie’s revenue must be made.

The model does not forecast anything about revenue and, thus, does not violate the “ nobody knows” principle. The model used only observed box of? ce revenue and empirical data on the number of pirate sites where the work print of the ? lm were available.

There was no attempt to forecast the effect of piracy on opening week revenue. The only variables that were used were directly observed empirical data. We did not have to pose the counterfactual conjecture asking what the movie might have earned but for piracy. Nor did we need to speculate as to how many viewers of the pirated version might have gone to see the legitimate version in a theater.

What we did do was to test directly the impact of pirate supply on the rate at which the movie’s theatrical revenues declined during the course of its run. 298 0 A. S. De Vany, W. D. Walls Change in Weekly Revenue 1.

0e+07 ? 2. 0e+07 ? 3. 0e+07 1 5 10 14 Week of Run Fig. 2 Change in Weekly Revenue versus Week of Run Table 3 Statistics on Revenue Change and Active Pirate Sites per Week Variable revenue change pirate supply obs 13 14 mean ? 4, 772, 879 8, 111.

286 std dev 8, 411, 162 1164 min ? 23. 7e+06 0 max ? 29, 650 33, 853 Table 4 Revenue Decline and Pirate Download Supply Variable constant pirate supply time time2 least-squares estimate ? 4. 35e+07 (5428957) ? 443. 8191 (115.

04) 1. 6e07 (3266542) ? 1422298 (427634) 0. 9954 5 median regression estimate ? 4. 15e+07 (1. 09e+07) ? 437. 9149 (31.

2635) 1. 48e07 (6508654) ? 287740 (712499) 0. 9563 5 Note: Estimated standard errors are in parentheses R2 Observations In Figure. 2 we plot the change in weekly revenue against the week of run. Clearly the trend is quadratic in time, so we should add a t 2 term to the regression model. The results of this regression are shown in the ? rst column of Table III.

The gain is an improved ? t, but the cost is a loss of degrees of freedom from adding another variable to the regression. There is a lot of noise in estimating the expected change in weekly revenue. We know that, with the heavy tails of motion picture revenue distributions, the estimator

Estimating the Effects of Movie Piracy 299 of the mean is going to be fragile in small samples. A robust alternative to estimating the mean change would be to estimate the median change in weekly revenue instead. The second column of Table III shows a quantile regression of the change in weekly revenue with a quadratic time trend and contemporaneous piracy.

This is the best regression to use. It has an excellent ? t to the data, and all of the parameters are estimated precisely. The coef? cient on piracy has a coef? cient of ? 437. 9149 with an estimated standard error of 31. 26.

The median loss per pirate site of 437. 91 is close to the mean loss and right in the middle of the estimates obtained with a linear revenue decline in the previous section. IV Loss Estimates There are three hypotheses regarding the effect of download sites on the rate at which revenues decline during the course of its theatrical run: piracy had no effect, a positive effect, or a negative effect. The model controls for the effect of the week of the run on the change in revenue.

The hypotheses that piracy had no effect or a positive effect (a slowing of the negative change) on the rate at which box-of? e revenue declined are strongly rejected. The hypothesis that piracy had a negative effect (an acceleration of the negative change) on the rate at which revenues declined is con? rmed. The results show that the magnitude of the decline in weekly box-of? ce revenues is positively related to the number of pirate download sites that are active before and during its run. The negative and highly significant value of the coef? cient of revenue change on pirate site-days provides evidence that piracy increased the magnitude of the decline in theatrical revenue.

The econometric model is an excellent ? t to the data and the estimated coef? cients are precisely estimated and highly significant. Moreover, the model parameter values do not change when a potential positive feedback from box-of? ce revenue to piracy is controlled for.

The assessment of revenue losses relies on the estimates of this dynamic model of box-of? ce revenues during the course of the ? lm’s run. In spite of the limited data one always has to deal with in the movies, owing to their brief product lives, the model captures the patterns in the data.

Because of the way we have modeled the problem, the estimates of revenue losses do not depend on counterfactual (but for piracy) assumptions as to what revenue would have been at the opening or in any week of its run. Rather, they depend on actual revenue and carefully developed measures of the number of pirate sites active before and during the run. We estimate the median rather than the mean loss because it is more conservative and is not

sensitive to the in? nite variance that research shows holds for movie grosses (De Vany and Walls 1999). Further, to avoid the in? ite variance problem, we do not predict opening week revenues and have excluded them as a source of damages from piracy.

As a result, these estimates do not assess how the availability of the work print for approximately 18, 134 pirate site-days during the week preceding the opening affected the ? rst week’s revenue. Finally, by directly controlling for the effect of the week of the run on weekly revenue decline, the model makes allowance for other factors, such as an adverse response to the ? lm and its intrinsic rate of play off, on its box-of? ce dynamics. Thus, the effect of piracy is estimated net of these other factors. 00 A.

S. De Vany, W. D. Walls The model estimates imply a median loss of revenue each week of $437. 91 per active pirate site. Integrating the median loss per site-day over the number of download sites active during the second through sixth weeks of the run, gives an estimate of the loss in US theatrical revenues from piracy of $41.

70 million. V Conclusion We develop and estimate a statistical model of the effects of piracy on the box-of? ce performance of a widely-released movie. The model resolves a number of contentious issues about piracy and permits on-going piracy to be detected early in the run of a motion picture.

The approach adopted in this paper overcomes a major weakness of customary counterfactual methods widely used to estimate damages of piracy. These counterfactual methods violate the “ nobody knows” principle by forecasting what the movie would have earned in the absence of piracy.

The model we present does not violate this basic principle of motion picture uncertainty and its estimates have the rigor that other estimates lack. Based on this study of a single widely-released movie, piracy has a substantial effect on motion picture revenues.