Pre-Release Movie Piracy and Box Office Sales: Estimates and Policy Implications

Liye Ma, Alan Montgomery, Param Vir Singh, and Michael D. Smith

WISE 2009: Research-In-Progress Extended Abstract

The impact of piracy on digital media sales is an important academic, managerial, and policy question. Answers to this question are particularly salient for theatrical piracy where rights-holders and policy-makers are struggling to determine both the size of the problem and the appropriate legislative and managerial responses, and where there are few if any rigorous academic studies to guide their decisions. The importance of this question became particularly clear earlier this year when a pirated copy of the movie "X-Men Origins: Wolverine" appeared on Internet weeks before its official theatrical release, generating many hundreds of thousands of downloads before the movie was available in theaters.

In this paper we analyze a new dataset collected from an Internet file-sharing site to examine the impact of pre-release theatrical piracy on box office revenues. Using a variety of empirical models and propensity score matching controls for endogeneity, we find that the presence of pre-release piracy reduces a movie's box office revenue by approximately 15%, that the vast majority of this reduction occurs in the movie's opening weekend, and that pirated copies of higher quality have a *less* severe impact on box office sales than lower quality pirated copies do.

Theory and Model Development

To test the impact of pre-release piracy across a movie's lifecycle we need to build a model to explain both the box office sales of movies and the distribution of sales to different time periods. Most movies see the highest level of sales in the opening weekend, with sales declining in subsequent weeks. Consistent with existing literature in marketing and information systems (e.g. Sawhney and Eliashberg 1996, Krider and Weinberg 1998), we model the box office revenue of movies using an exponentially declining function:

$$BO_{it} = m_i e^{-\lambda_i t + \varepsilon_{it}} = e^{\log m_i - \lambda_i t + \varepsilon_{it}} \quad (1)$$

where BO_{it} is the box office revenue of movie i at time t, and m_i and λ_i represent the market potential[1] and the rate of decay of the movie, respectively. Both the market potential and the rate of decay are determined by relevant movie characteristics:

$$\log m_i = X_i' \beta_i + \zeta_i \quad (2)$$

$$\lambda_i = Z_i' \gamma_i + \xi_i \quad (3)$$

where X_i is a $k \times 1$ vector of the characteristics of movie i and Z_i is a $l \times 1$ vector of the characteristics of movie i that influence the rate of decay.

[1] If we the first period is indexed by 0, then the total box-office revenue if the movie is played perpetually is $\sum_{t=0}^{\infty} m_i e^{-\lambda_i t} = \frac{m_i}{1 - e^{-\lambda_i}}$, which is proportional to m_i when the rate of decay, λ_i, is held constant.

We model the impact of pre-release piracy on market potential and the rate of decay, as follows:

$$\log m_i = X_i^{'}\beta_i + \rho Pir_i + \zeta_i \qquad (4)$$

$$\lambda_i = Z_i^{'}\gamma_i + \tau Pir_i + \xi_i \qquad (5)$$

where Pir_i is an indicator for the existence of pre-release piracy for the movie. Note that in this model $\rho < 0$ indicates that pre-release piracy harms box-office sales and $\tau < 0$ indicates that the harm from piracy is greater in the time periods closer to a movie's release date.

Data

We collected our data from four sources: IMDB.com, BoxOfficeMojo.com, Yahoo movies, and vcdquality.com. Our entire data set consists of all movies released within a three-year period between January 2006 and January 2009, according to IMDB.

We collected movie characteristics from both IMDB.com and BoxOfficeMojo.com, including weekly movie box office sales, distributor, genre, MPAA rating, director, star appeal, user rating, and critic rating. We collected pre-release movie piracy information from the website vcdquality.com. This is not an Internet file-sharing site itself. Instead, it monitors various Internet file sharing sites, and posts on its own website a message once a copy of certain movie becomes available for download at some of the sites. Each message is dated, allowing us to infer the date on which the copy was posted. Also for each such copy, the website tracks user ratings on the pirated content's video and audio quality.

Table 2: Movie Descriptive Statistics

		Total US Box Office	Production Cost	Opening Weekend Screens
With Pre-release Piracy	Mean	$41 million	$43 million	1421
	Std Dev	$60 million	$51 million	1293
	Min	$110 thousand	$2.5 thousand	11
	Max	$210 million	$200 million	3940
Without Pre-release Piracy	Mean	$55 million	$45 million	2324
	Std Dev	$70 million	$47 million	1138
	Min	$740 thousand	$100 thousand	13
	Max	$530 million	$260 million	4366

After eliminating movies where there is no production budget available or where the movie appeared in theaters for less than 6 weeks, our final dataset contains 194 movies. Among this sample, 21 movies had a pirated copy available prior to box office release. The descriptive statistics of movies in the final data set are reported in table 2. As is shown in the table, movies with pre-release piracy have on average lower production cost, number of opening weekend screens, and total US box office revenue. The difference between the box office sales is larger than that between the production costs.

Results and Analysis

We start with a parsimonious specification by assuming that the default rate of sales decline of all movies is the same, before accounting for the effect of pre-release piracy. With this assumption, the parameters to be estimated are coefficients for movie characteristics that impact the market potential of movies, the rate of

decline of box office revenue over time, the coefficient for pre-release piracy on market potential, and the coefficient for pre-release piracy on rate of decay. From equations (1) and (4), we can estimate all these parameters using the following model.

$$\log BO_{it} = X_i^{'}\beta_i - \lambda t + \rho Pir_i - \tau Pir_i t + \upsilon_{it} \qquad (6)$$

Most movie characteristics are expected to impact the market potential. Therefore, we included in X_i all the movie characteristics that are available to us. This includes movie distributor, genre, MPAA rating, director appeal, star appeal, budget, screen, user rating, and critic rating. We created dummy variables for different distributors, genres, and MPAA ratings. The result of the regression is reported in table 3.

Table 3: Result - Fixed Rate of Decay

Parameter	Estimate	Parameter	Estimate
Constant	7.4(***)	Warner	0.072
τ	-0.20(**)	Universal	0.59(***)
ρ	-0.65(**)	Paramount	0.48(***)
λ	0.65(***)	Fox	0.22(.)
Budget	0.11(**)	Sony	0.51(***)
Screen	0.96(***)	Newline	0.19
Director Appeal	0.61(***)	Lionsgate	0.44(*)
User Rating	0.089(*)	MGM	-0.54(***)
Critic Rating	0.073(**)	Action	-0.081
Star Appeal	0.46(***)	Comedy	-0.014
G	0.55	Drama	-0.19(*)
R	-0.69(*)	Adventure	0.022
PG13	-0.15	Horror	0.18
PG	0.056	Thriller	-0.11
		Animation	-0.23
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 Multiple R-squared: 0.8055, Adjusted R-squared: 0.8007			

These results show that the coefficient for piracy on market potential is -0.65, suggesting that pre-release piracy reduces the market potential of movies, and the coefficient for piracy on rate of decay is -0.20, suggesting that movies with pre-release piracy demonstrate a slower decay in revenue than other movies do. Since the rate of decay without pre-release piracy is 0.65, these parameter estimates imply a 18.2% revenue loss arising from pre-release piracy.

These results confirm that pre-release piracy negatively impacts box office revenue, and the impact is more significant to earlier weeks than later weeks. However, the assumption that all movies, piracy effect aside, have the same rate of decline in revenue, is quite restrictive. We can reasonably expect that the rate of decay varies from movie to movie based on quality. Thus, we extend the model in (6) by allowing the rate of sales decline to vary based on relevant movie characteristics as follows:

$$\log BO_{it} = X_i^{'}\beta_i - Z_i^{'}\gamma_i t + \rho Pir_i - \tau Pir_i t + \upsilon_{it} \qquad (7)$$

In this model, the movie characteristics that influence market potential (i.e. variables included in X_i) remain the same. Whereas most characteristics of movies are expected to play a role in determining the market size of viewers, the rate of decline in revenue should be primarily driven by quality-related characteristics – higher quality movies receives more positive word-of-mouth after release, and would see slower rate of revenue decline than lower quality ones. Among the movie characteristics we gathered, director appeal, star appeal, user ratings, and critic ratings can all be considered as indicators of quality, therefore, we include those variables in the vector Z_i.

Results for this model are reported in Table 4. The coefficients on market potential and rate of decay are both negative and significant as before. Moreover, based on the average movie characteristics in our data set, the coefficient estimates imply a 14.8% total reduction in box office revenue from pre-release piracy — slightly lower than our baseline results but still economically and statistically significant.

Table 4: Result – Variable Rate of Decay			
Piracy and Market Potential			
Parameter	Estimate	Parameter	Estimate
Constant	7.7(***)	Warner	0.072
τ	-0.16(**)	Universal	0.59(***)
ρ	-0.53(*)	Paramount	0.48(***)
λ	0.71(***)	Fox	0.22(.)
Budget	0.11(**)	Sony	0.51(***)
Screen	0.96(***)	Newline	0.19
Director Appeal	0.33(.)	Lionsgate	0.44(**)
User Rating	0.11	MGM	-0.54(***)
Critic Rating	-0.03	Action	-0.081
Star Appeal	0.01	Comedy	-0.014
G	0.54	Drama	-0.19(*)
R	-0.69(*)	Adventure	0.02
PG13	-0.15	Horror	0.18
PG	0.054	Thriller	-0.11
		Animation	-0.23
Rate of Revenue Decline			
User Rating	0.0073	Director Appeal	-0.080(.)
Critic Rating	-0.030(*)	Star Appeal	-0.13(**)
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 Multiple R-squared: 0.8101, Adjusted R-squared: 0.8047			

Impact of Piracy Quality

To evaluate the effect of piracy quality on box office sales, we extend equation (6) to include the piracy quality variable. This results in the following model

$$\log BO_{it} = X_i^{'}\beta_i - \lambda t + \rho_1 Pir_i + \rho_2 Pirqual_i - \tau_1 Pir_i t - \tau_2 Pirqual_i t + \upsilon_{it} \qquad (8)$$

which allows for the possibility that the quality of the pirated copy impacts both the market potential and the rate of revenue decline over time.

The result of the estimation is reported in Table 5. In this regression, the coefficient for piracy quality on market potential is 0.216 and is statistically significant at .10 level. This suggests that *ceteris paribus*, pirated versions of higher quality result in lower reduction in market potential. One possible explanation is that high piracy quality gives viewers a positive impression, which in turn may generate positive word-of-mouth. Meanwhile, the coefficient for piracy quality on rate of sales decline is -0.032, but is statistically insignificant.

Table 5: Impact of Piracy Quality	
Parameter	Estimate
τ_1	-0.195(***)
τ_2	-0.0322
ρ_1	-0.655(**)
ρ_2	0.216 (.)
λ	0.651(***)
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1	

Robustness

Because pre-release piracy pre-dates the movie's release date, simultaneity is not a dominant concern in evaluating the causal impact. Endogeneity, however, is a concern. We use pair-wise propensity score matching to address potential endogeneity concerns, and repeat our tests on the matched dataset. Propensity score matching is used to address possible selection bias by ensuring that pirated movies are compared with similar movies that are equally likely to be pirated but weren't. The propensity score can be calculated using classification models such as binary-probit or binary-logit, by regressing the piracy indicator onto other related variables.

Propensity score matching has been shown to be a powerful technique to address possible selection bias. However, when used on a small sample, care must be taken as the matching can exhibit high variance and make the result unstable (see, for example, Caliendo and Kopeinig 2005). Specifically, with a small sample, the number of variables used in the selection process, i.e. predicting piracy, must be small. To make the best use of propensity score matching in our study, we tested two specifications of the selection process. In the first specification, production budget and the number of opening weekend screens are used to predict piracy. These important supply-side characteristics, may be related to the pre-release piracy, which generally happens during the production process. In the second specification, we use production budget as well as star and director information are used to predict piracy. The rationale is that it is possible that movies with star actors or famous directors are more likely to be pirated than others[2].

For each specification, we paired each pirated movie with a similar movie that was not pirated. We then re-estimated the models in section 5.1 and 5.2 using the smaller paired data set. The results are shown in

[2] Note that we do not expect that certain selection process existed behind the pre-release piracy. Pre-release piracy is usually leaked by accident and it is highly unlikely there is a systematic pattern behind it. Nonetheless, we perform propensity-score-matching here just to validate the robustness of the result.

Tables 6-9 below. As is shown in the tables, the signs of coefficients remain the same in all configurations, and majority of the coefficients are statistically significant. This validates of the robustness of our findings.

Table 6: Propensity Score Matching - Fixed Rate	
Selection Criteria: Budget + Screen	
Parameter	Estimate
τ	-0.151(*)
ρ	-1.33(***)
λ	-0.598(***)
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1	

Table 7: Propensity Score Matching - Variable Rate	
Selection Criteria: Budget + Screen	
Parameter	Estimate
τ	-0.0735
ρ	-1.06(***)
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1	

Table 8: Propensity Score Matching - Fixed Rate	
Selection Criteria: Star + Director Appeal	
Parameter	Estimate
τ	-0.176(*)
ρ	-0.223
λ	-0.645(***)
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1	

Table 9: Propensity Score Matching - Variable Rate	
Selection Criteria: Star + Director Appeal	
Parameter	Estimate
τ	-0.127(.)
ρ	-0.0531
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1	

References

Caliendo, M. and Kopeinig, S., "Some Practical Guidance for the Implementation of Propensity Score Matching", 2005, Discussion Paper No. 1588, Institute for the Study of Labor, Germany

Krider, R.E. and Weinberg, C.B., "Competitive Dynamics and the Introduction of New Products: The Motion Picture Timing Game", Journal of Marketing Research, 1998, Vol. 35, No. 1

Sawhney, M.S. and Eliashberg, J., "A Parsimonious Model for Forecasting Gross Box-Office Revenues of Motion Pictures", Marketing Science, 1996, Vol. 15, No. 2