Decision Support Systems 94 (2017) 77–84

Contents lists available at ScienceDirect

Decision Support Systems

journal homepage: www.elsevier.com/locate/dss

The power of the “like” button: The impact of social media on box office

Chao Ding [a,*], Hsing Kenneth Cheng [b], Yang Duan [c], Yong Jin [d]

[a] Innovation and Information Management, Faculty of Business and Economics, The University of Hong Kong, Pok Fu Lam, Hong Kong
[b] Department of Information Systems and Operations Management, Warrington College of Business Administration, University of Florida, Gainesville, FL 32611-7150, USA
[c] Department of Finance and Decision Sciences, School of Business, Hong Kong Baptist University, Kowloon, Hong Kong
[d] School of Accounting and Finance, The Hong Kong Polytechnic University, Kowloon, Hong Kong

ARTICLE INFO

Article history:
Received 8 March 2016
Received in revised form 7 November 2016
Accepted 8 November 2016
Available online 12 November 2016

Keywords:
Box office
Facebook “like”
Social impact
Social marketing

ABSTRACT

The mainstream research of social factors and box office performance has concentrated on post-consumption opinion mining and sentiment analysis, which are difficult to operationalize to the benefits of the industry practitioners whose objective is to maximize box office sales. In this study, we propose the Facebook “like” as an effective social marketing tool before the release of movies for several reasons. Firstly, people's prerelease “liking” of movies can be influenced by marketing campaigns. Secondly, the clicks of “likes” create social impact, as suggested by the Social Impact Theory, on moviegoers' consumption behaviors. And thirdly, Facebook “like” provides practitioners with real-time visible updates. By studying the impact of prerelease “likes” on box office sales, we not only contribute to the literature by offering a new social metric to evaluate the box office performance, but also provide the industry practitioners with quantitative support for the effectiveness of their social marketing activities. Our empirical results indicate that the prerelease “likes” exert a significantly positive impact on box office performance. More specifically, 1% increase in the number of “likes” in the one week prior to release is associated with an increase of the opening week box office by about 0.2%. As it approaches the release date, the prerelease “like” impact becomes stronger, suggesting that the latest prerelease “likes” are more effective in driving box office performance.

1. Introduction

Movie studios typically run week-long or even month-long marketing campaigns of their movie productions before releasing to theaters. Traditionally, these marketing practices include advertising through such channels as TV, newspapers, cinemas, and public transit systems. McKinsey & Company [43] reports that the global cinema advertising expenditure reached roughly US$2.1 billion in 2014, and is expected to grow to US$2.8 billion in 2018. These traditional marketing activities, however, are costly and sometimes not as effective as expected. Elberse and Anand [18] estimate that, on average, every dollar increase in advertising increases box office sales expectation by only up to $0.65, a bad news for the decision maker. Another downside of traditional marketing is the limited coverage of audience. Because of the costly nature, studios allocate a majority of the marketing budget to potentially more lucrative movie productions (blockbusters) and to more mature markets. As Eliashberg et al. [19] indicates, it is still unclear to what extent marketing affects box office performance.

The advent of social platforms, for instance, Facebook, Twitter, Google+ and LinkedIn, has offered studios new marketing opportunities. Businesses have been employing social platforms as a low-cost marketing venue to increase brand awareness [3,40,52], attract web traffic [57], grow demand [8,9,10,11,62], discover product information [1,23] and enhance firm value [41,61]. Among the many social marketing tools, the Facebook “like” button is overwhelmingly popular in the business world [38,57], evidenced by the ubiquitous “like” campaigns both online and offline.

In the motion picture industry, the “like” button is also widely embedded in most movie-related promotions and marketing campaigns long before movies are released, so that people could tap on the “likes” and help spread the words. People's prerelease “liking” of movies is generally based on one's preference, which is largely influence by external factors like advertisements and marketing campaigns. Unlike traditional marketing efforts that have limited coverage of audience, the clicks of “likes” create social impact that could become viral [26]. To study the social impact, we borrow from Latané [36]'s Social Impact Theory (to introduce in Section 2) in Psychology and examine the number of prerelease “likes”, which is also in line with prior studies like Kuan et al. [34]. We develop our empirical analyses in three stages. We first run cross-sectional regression to understand the basic relationship between prerelease “likes” and box office performance. We examine the

* Corresponding author.
E-mail addresses: chao.ding@hku.hk (C. Ding), kenny.cheng@warrington.ufl.edu (H.K. Cheng), yangduan@hkbu.edu.hk (Y. Duan), jimmy.jin@polyu.edu.hk (Y. Jin).

http://dx.doi.org/10.1016/j.dss.2016.11.002
0167-9236/

prerelease “likes” up to one month and box office performance up to one month. Then we construct a panel data and apply the Fama-MacBeth regression [22] to estimate the prerelease “like” impact over time. Lastly, to address the potential endogeneity issues, we employ Two-Stage Least Squares (2SLS) method and Instrument Variable – Generalized Method of Moments (IV-GMM) in the empirical test.

Our empirical results indicate that the prerelease “likes” exert a significantly positive impact on box office performance. More specifically, 1% increase in the number of “likes” in the one week prior to release is associated with an increase of the opening week box office by about 0.2%. As it approaches the release date, the prerelease “like” impact becomes stronger, suggesting that the latest prerelease “likes” are more effective in driving box office performance.

It is our belief that scholarly outputs should benefit both the academic community and the industry practitioners. Therefore, it is equally important to contribute to the academia effective metrics/methodologies/theories, and to offer to the motion picture industry insightful marketing strategies/plans. By studying the impact of prerelease “likes” on box office sales, we do not only contribute to literature by offering a new social metric to evaluate the box office performance, but also we are able to provide the industry practitioners with quantitative support for the effectiveness of their social marketing activities. We demonstrate that the prerelease “like” impact is significantly positive, encouraging practitioners to employ such social marketing tools in campaigning endeavors. Compared to traditional marketing efforts whose outcomes are difficult to observe and manage in a timely manner, Facebook “like” provides real time updates that are instantly visible by both the moviegoers and the practitioners, so that the latter can take immediate decision and actions. Our empirical results suggest that, to improve box office performance, practitioners should invest more in Facebook “like” marketing as movies approach the release dates since the “like” impact is stronger.

The remainder of the paper is organized as follows. In Section 2, we present a review of related literature. In Section 3, we describe data collection and summary statistics. We then demonstrate empirical results in Section 4, and conclude with discussions and limitations in Section 5.

2. Literature review

There is a rich literature in the information systems and marketing areas that explores the various factors associated with box office performance. We categorize them as social metrics and non-social metrics. Non-social metrics are those that create little or no word-of-mouth over social platforms. They include ads expenditure [18,25,63], prerelease piracy [12,13,42], reviews and ratings [6,7,11,15,16,20,32,39], prerelease search activities [35], the number of concurrent movie showings [2], political views of the moviegoers [53], Wikipedia status [44], and Hollywood Stock Exchange [18,55]. Delen et al. [14] have devised a web-based decision support system to make forecast on box office sales. The system incorporates many of the measurements mentioned.

On the social metrics side, the mainstream research of social metrics and box office performance has concentrated on extracting and evaluating Twitter sentiments [5,28,54,59], and blog reviews [25,45,48]. Hennig-Thurau et al. [28] test the “Twitter Effect”, which generates word-of-mouth that potentially influences moviegoers, and find support for a negativity bias. Rui et al. [54] also use public tweets to evaluate box office revenue. They find that the effect of Twitter chatters is significant, however, the magnitude and direction of the effect depend on the content and sources. Jansen et al. [31] show that Twitter is a good tool for brand management. Although these studies all provide effective social metrics to evaluate or predict the box office performance, three issues may arise concerning sample bias and the lack of applications to the industry practitioners. First, there is potential sample bias with Twitter users. Duggan and Brenner [17] report that “US-based Twitter users were disproportionately young, urban or suburban, and black.” Therefore, the extracted Twitter sentiments are only a partial representation of the public.

The second issue with the existing social metrics studies is the lack of practical use to the movie producers whose objective is to maximize box office sales. Currently, scholars have made a great effort to improve the algorithms, text mining, and methodologies. For examples, Pak and Paroubek [49] focus on performing linguistic analysis to create a corpus which can be used to build Twitter sentiment classifier. Khan et al. [33] adopt a hybrid approach to address sentiment classification problems. Ghiassi et al. [24] utilize *n*-gram and statistical analysis techniques to develop a Twitter sentiment lexicon. For a detailed summary of the opinion mining and sentiment analysis, see Pang and Lee [50]. They make meaningful contribution to the academic world but not as much to the industry.

The third issue is that the composing of the relevant tweets or reviews is fundamentally opinion-oriented, in which it serves to express one's *post*-consumption state of mind. Both prerelease and post-release movie related tweets are mainly moviegoers' opinions after movie consumption, and they are hardly the results of a studio's marketing endeavors. In addition, from a more practical perspective, it is highly unlikely that a moviegoer scrutinizes the aggregate Twitter data, as the scholars do, before going to a movie. Therefore, although such data analytics produce effective metrics that contribute to the literature, they provide little actionable guidance to producers, studios and filmmakers to improve the box office performance.

This study aims to not only provide an additional social metric, Facebook “like”, but more importantly, to help the practitioners gain useful insights that can be operationalized for their own benefits. To study the impact of Facebook “like”, we refer to the Social Impact Theory from the field of Psychology. Latané [36] describes in his Social Impact Theory (SIT) that the behaviors of people (the target) are impacted by other sources through three social forces. They are: number, immediacy, and strength. Number refers to the number of sources, immediacy refers to the distance between the target and the sources, and strength refers to the importance of the sources. Studies have confirmed that the larger the social size [4,47,56], the more important the sources [27,30], and the more immediate to the source [4,51], will lead to stronger social impact. For instance, Jackson [29] argues that a large number of strangers will be somewhat more effective than a small number of strangers in persuading a target to make a donation, given that they are of equal importance and immediacy. Therefore, a stronger social impact on a target may convert to more consumption.

3. Data collection and summary statistics

3.1. Data collection

Internet Movie Database (IMDb) is an online database of movies, television programs, and video games. It keeps records of a movie's trailer, production particulars, casting information, summary of plot, reviews, rankings, and more. Accessing IMDb is a common practice for potential moviegoers who would like to have a first glance at movies of interest. IMDb has also embedded the Facebook “like” button on each movie's pages. We created a web crawler to collect data from three sources: IMDb, Box Office Mojo and Facebook, throughout the 2013 calendar year. On IMDb, we obtained movie-specific characteristics, including the movie's name, genre, MPAA rating, production budget and release date. From Box Office Mojo, we collected time series data on movies' box offices and number of screens. We also linked to the Facebook API (Application Programming Interface) to fetch the time series data of “likes” activities on the movie's IMDb page. We then combined the three databases by matching a movie title's name and ID.

We construct the sample as follows: 1) we keep titles that have complete data from one month prior to and one month after the release date; 2) for each title, we compute the gross box office on the opening day, in the opening week, in the opening month and the final total revenue; 3) we record the total number of prerelease “likes”; 4) we calculate the corresponding incremental number of “likes” on the opening

Table 1
Definition of variables.

t	Number of days after a title is released ($t=1$ is the opening day)
REV_1	Gross box office on the opening day
REV_t	Gross box office on day t
REV_W	Gross box office in the opening week
REV_M	Gross box office in the opening month
REV_T	Total box office
$like_t$	Cumulative number of "likes" on day t
$like_T$	Total number of "likes"
$prelike_0$	Cumulative number of "likes" one day prior to release
$prelike_x$	Cumulative number of "likes" by the end of xth week prior to release, $x=1,2,3,4$
$\Delta prelike_x$	Incremental number of "likes" in the xth week prior to release
$\Delta like_t$	Incremental number of "likes" on day t
$\Delta like_1$	Incremental number of "likes" on the open day
$\Delta like_W$	Incremental number of "likes" in the opening week
$\Delta like_M$	Incremental number of "likes" in the opening month
$screen_t$	Number of screens on day t
$screen_1$	Number of screens on the open day
$screen_W$	Number of screens in the opening week
$screen_M$	Number of screens in the opening month
budget	Production budget of the movie
Friday	Dummy whether the movie opened on Friday
MPAA	Movie's MPAA rating
genre	Movie's genre

day, in the opening week, and opening month; 5) we synchronize the release dates of all titles as day 1. Table 1 shows the description of all the variables in our sample.

3.2. Summary statistics

There are a total of 64 movie titles in our sample. As described before, the dataset includes movie-specific characteristics, "like" activities and box office performance. Table 2 presents the summary statistics of the variables.

The total box office of one title is 77.4 million USD on average with a large standard deviation of 81.18 million USD. Among all the box office sales, the opening day box office accounts for 12% of the total, that is, 8.37 million USD. The revenue from opening week makes up 47% of the total revenue on average, with a dollar value of 3.22 million USD. Prior studies (e.g., Liu [39]) primarily use the opening week revenue as a proxy for the box office performance. The opening month box office more than doubles the opening week box office, reaching 67.7 million USD. >87% of the total box office is achieved within the first month of release. For each title, >9187 "likes" have been clicked by the end of the year 2013. More than half of the "like" clicks occur before a title is released, and over 10% of the total prerelease "likes" is contributed by the clicks in the one week prior to release. The total number of "likes" increases by 185.06 on the opening day on average, which is more than double the average number of "likes" in the week prior to release. The rapid surge in the number of "likes" persists through the first week after release, which reaches >1100 "likes". After the initial week, the rate of increase slows down and the next three weeks' total number of "likes" is slightly higher than that of the first week.

Table 3 presents the correlation matrix across the above-mentioned variables in our empirical analysis. The box office revenues across different time periods are highly correlated, which implies that the revenues are persistent after the title is released. The number of "likes" one day prior to release as well as one week prior to release have a strong correlation with the box office performance. The number of screens on the opening day is highly correlated with the opening day and opening week box office performances; however, it is weakly correlated with the total box office and that after the first week. It provides evidence that the number of screens on the opening day only influences the box office over a short period but has little effect in the long run. The budget is positively correlated with the total box office as well as the number of prerelease "likes".

Table 3
Spearman/Pearson Correlation matrix.

	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]
[1] REV_T	1							
[2] REV_1	0.644	1						
[3] REV_W	0.799	0.903	1					
[4] REV_M	0.997	0.691	0.825	1				
[5] $prelike_0$	0.479	0.411	0.362	0.493	1			
[6] $prelike_1$	0.470	0.403	0.358	0.485	0.998	1		
[7] $screen_1$	0.240	0.842	0.690	0.302	0.065	0.066	1	
[8] *budget*	0.653	0.383	0.540	0.652	0.442	0.439	0.158	1

4. Empirical analyses

Latané [36] suggests in Social Impact Theory that the social impact of other sources on a target is based on three social forces: number, immediacy and strength. Under the context of this study, the target is a potential moviegoer, and the other sources are people who have clicked Facebook "like" before the release of movies. Given the huge number of sources and that the sources are mainly anonymous on the Internet, we argue that the immediacy and strength of the sources are considered the same to any target. Then, we focus on the number of sources, which is equivalent to the number of prerelease "likes".

Table 2
Summary statistics.

Variables	Mean	Min	p25	p50	p75	Max	S.D.
REV_1 (in 1000)	8369.63	47.70	2896.42	6008.23	9881.30	70,950.13	10,475.62
REV_W (in 1000)	32,189.77	214.49	11,455.07	24,198.52	38,463.16	222,116.10	34,371.29
REV_M (in 1000)	66,696.17	1299.00	24,578.31	51,430.19	83,139.26	362,953.80	65,231.11
REV_T (in 1000)	77,381.17	1703.13	26,107.22	55,608.21	96,972.93	417,970.00	81,184.71
$like_T$	9187.80	236.00	2216.00	6243.00	12,737.50	45,333.00	9175.23
$prelike_0$	4798.86	124.00	1278.00	2417.00	6270.50	28,942.00	5562.21
$prelike_1$	4328.88	111.00	1104.50	2224.50	5814.50	28,542.00	5133.29
$prelike_2$	4073.17	99.00	981.50	2069.00	5503.00	28,242.00	4976.91
$prelike_3$	3871.05	93.00	900.00	1883.50	5262.00	28,142.00	4881.98
$prelike_4$	3698.95	88.00	825.00	1712.50	4988.50	28,042.00	4806.46
$\Delta prelike_2$	725.69	25.00	245.50	439.00	799.50	5000.00	895.73
$\Delta prelike_1$	469.98	13.00	135.00	260.00	508.50	3800.00	652.85
$\Delta like_1$	185.06	0.00	40.00	92.00	181.50	1600.00	300.36
$\Delta like_W$	1100.69	59.00	244.00	510.00	1288.50	9699.00	1611.72
$\Delta like_M$	2712.48	105.00	576.50	1263.50	3176.50	25,199.00	3859.90
$screen_1$	2745.36	1.00	2483.00	3031.00	3404.50	4163.00	1085.89
$screen_W$	2811.89	5.00	2509.00	3042.50	3449.00	4163.00	1039.68
$screen_M$	2513.77	98.25	1880.05	2685.11	3138.82	4013.00	928.98
budget	65,578.13	1500.00	20,000.00	40,000.00	102,500.00	225,000.00	58,052.67

Table 4
Cumulative prerelease “likes” and the box office performance.

	Panel A: opening day box office					Panel B: opening week box office					Panel C: opening month box office				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
Variables	$\ln(REV_1)$	$\ln(REV_1)$	$\ln(REV_1)$	$\ln(REV_1)$	$\ln(REV_1)$	$\ln(REV_w)$	$\ln(REV_w)$	$\ln(REV_w)$	$\ln(REV_w)$	$\ln(REV_w)$	$\ln(REV_M)$	$\ln(REV_M)$	$\ln(REV_M)$	$\ln(REV_M)$	$\ln(REV_M)$
$\ln(prelike_0)$	0.452***					0.297***					0.199***				
	(0.000)					(0.001)					(0.007)				
$\ln(prelike_1)$		0.432***					0.279***					0.187***			
		(0.000)					(0.001)					(0.010)			
$\ln(prelike_2)$			0.418***					0.268***					0.181**		
			(0.000)					(0.002)					(0.011)		
$\ln(prelike_3)$				0.409***					0.261***					0.176**	
				(0.000)					(0.002)					(0.012)	
$\ln(prelike_4)$					0.398***					0.253***					0.171**
					(0.000)					(0.002)					(0.012)
$\ln(screen_1)$	0.799***	0.799***	0.798***	0.797***	0.800***	0.801***	0.799***	0.797***	0.796***	0.797***	1.205***	1.207***	1.208***	1.211***	1.215***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
$\ln(budget)$	0.235**	0.244**	0.250**	0.250**	0.256**	0.311***	0.319***	0.324***	0.325***	0.329***	0.258***	0.262***	0.264***	0.264***	0.265***
	(0.021)	(0.018)	(0.016)	(0.016)	(0.013)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)
Friday	0.709***	0.727***	0.732***	0.736***	0.748***	0.205	0.217	0.220	0.223	0.230	0.147	0.154	0.156	0.156	0.160
	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)	(0.339)	(0.309)	(0.302)	(0.296)	(0.280)	(0.373)	(0.351)	(0.345)	(0.342)	(0.329)
genre	Controlled	Controlled	Controlled	Controlled	Controlled	Controlled	Controlled	Controlled	Controlled	Controlled	Controlled	Controlled	Controlled	Controlled	Controlled
MPAA	Controlled	Controlled	Controlled	Controlled	Controlled	Controlled	Controlled	Controlled	Controlled	Controlled	Controlled	Controlled	Controlled	Controlled	Controlled
Constant	1.061	1.111	1.178	1.278	1.268	2.826**	2.865**	2.910**	2.975**	2.973**	2.197	2.248*	2.271*	2.248	2.240
	(0.478)	(0.468)	(0.446)	(0.412)	(0.416)	(0.043)	(0.043)	(0.041)	(0.037)	(0.038)	(0.128)	(0.099)	(0.095)	(0.123)	(0.124)
Observations	64	64	64	64	64	64	64	64	64	64	64	64	64	64	64
Adjusted R^2	0.869	0.865	0.864	0.863	0.862	0.844	0.841	0.840	0.839	0.838	0.825	0.823	0.823	0.822	0.822

Notes: *z*- or *t*-statistics are in parentheses. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

4.1. Cumulative prerelease "likes" and the box office performance

In this section we employ the following cross-sectional regression model to study the relationship between the prerelease "likes" and the box office performance.

$$\ln(REV) = \ln(prelike) + \ln(screen) + \ln(budget) + friday + genre + MPAA + \epsilon$$

The dependent variable *REV* is the box office on the opening day (REV_1), in the opening week (REV_W) or the opening month (REV_M). The explanatory variable in interest, *prelike* is the number of prerelease "likes". We construct five different *prelike* variables to differentiate time periods before release: $prelike_0$, $prelike_1$, $prelike_2$, $prelike_3$, and $prelike_4$, which correspond to the cumulative number of prerelease "likes" one day, one week, two weeks, three weeks, and four weeks prior to release, respectively. We incorporate them separately into the regressions to avoid multicollinearity and more importantly, to examine the changes of prerelease "like" impact over time. We also include several control variables which are documented in the literature as correlated with box office performance. For instance, number of screens, production budget, Friday dummy (equals 1 if the releasing day is a Friday and 0 otherwise), genre and MPAA ratings. Table 4 presents OLS regression results of the above empirical model. Following White [58], we adjust standard errors for heteroskedasticity in all regressions. In general, the estimates of prerelease "like" impact are all significantly positive. We obtain relatively high Adjusted R^2 in all OLS regressions, suggesting that the empirical model has a strong explanatory power on total box office performance.

In particular, Panel A presents the opening day box office regression results. In Model (1), the coefficient of $prelike_0$ is significantly positive at 1% level. The magnitude is 0.452, which implies that other things equal, 1% more "like" clicks, that is, 480 more "likes" than average, before the opening day is related to a 0.452% increase in the opening day box office. Such prerelease "like" impact is monotonically increasing as it approaches the release day, which is evidenced by Models (2) to (5). Coefficients of additional control variables are consistent with our expectation. The number of screens on the opening day, total production budget and Friday dummy are all positively related with opening day box office in all models. Panels B and C suggest similar prerelease "like" impact when examining the opening week and opening month box office. The monotonic pattern still exists for more recent prerelease "likes". In other words, the latest prerelease "likes" exert a stronger effect on box office performance after controlling additional explanatory variables. However, the Friday effect disappears in the opening week and opening month regressions. Overall, our empirical model fits the data well. We find a significant and positive relationship between prerelease "likes" and box office performance. The positive prerelease "like" impact extends to even 4 weeks prior to release, although the impact becomes stronger as it approaches the releasing day.

4.2. Incremental prerelease "likes" and the box office performance

In this section, we study the relationship between the incremental number of prerelease "likes" ($\Delta prelike_x$) and the box office performance. We employ the following cross-sectional regression model to estimate the coefficients.

$$\ln(REV) = \ln(\Delta prelike) + \ln(screen) + \ln(budget) + friday + genre + MPAA + \epsilon$$

Both the dependent variables and control variables are the same as in the last subsection. The explanatory variable is $\ln(\Delta prelike_1)$, which is natural logarithm of incremental number of "likes" in the one week prior to release. The choice of one week is the result from our previous analyses that more recent prerelease "likes" have a stronger effect on the box office performance. Table 5 presents our empirical results. After controlling other explanatory variables, we find significantly positive relationships between the incremental number of "likes" one week prior to release and the opening day/week/month box office.

In Table 5, the coefficients of $\ln(\Delta prelike_1)$ are 0.517, 0.386 and 0.247 for the opening day, opening week and opening month box offices, respectively, all significant at 1% level. The results provide empirical evidence that the incremental number of "likes" in the one week prior to release has persistent impact on opening day, opening week and opening month box office performances. The impact is also diminishing over time.

4.3. Prerelease "likes" and box office performance: A panel data approach

We have so far conducted empirical analysis on the cross-sectional level, which means that all the differences are made across different movies. We are also interested in understanding how the prerelease "likes" affect an average movie's box office over time. In this subsection, we construct a panel and apply the Fama-MacBeth [22] regression method, which is widely used in the finance literature to estimate the premium rewarded to a risk factor. For a panel dataset, the Fama-MacBeth regression method generally considers two steps. The first step is to run a cross-sectional regression to obtain time-series coefficients. The second step is to calculate the coefficients' time series averages, and adjust the standard errors by serial autocorrelation, which provides an estimate of the average effect across time. Compared with panel regression with fixed effect, the Fama-MacBeth regression would not drop any independent variable that is determined. Another advantage is that we can correct for the dynamics of post-release "likes" by including time-series variations, which is not addressed in the preceding regressions. Though the Fama-MacBeth regression does not address the serial correlation issue which results in the biased *t* statistics, we further follow Newey and West [46], and compute the robust standard errors to adjust for the serial correlation. Table 6 presents the Fama-MacBeth regression results and the robust Newey-West standard errors are reported in brackets.

The coefficient of prerelease "likes" (one week prior) is 0.240, significant at 1% level, indicating that the effect of prerelease "likes" on daily box office revenue is consistently strong in the opening month. We also report the average coefficient on incremental number of prerelease "likes", which is 0.174 and significant at 10% level. We further conduct several robustness analyses in our empirical tests. We control the changes in post-release "likes", and the significant positive relationship between prerelease "likes" and the box office performance still holds.

Table 5
Incremental prerelease "likes" and the box office performance.

Variables	(1) $\ln(REV_1)$	(2) $\ln(REV_W)$	(3) $\ln(REV_M)$
$\ln(\Delta prelike_1)$	0.517***	0.386***	0.247***
	(0.000)	(0.000)	(0.003)
$\ln(screen_1)$	0.781***		
	(0.000)		
$\ln(screen_W)$		0.838***	
		(0.000)	
$\ln(screen_M)$			1.252***
			(0.000)
$\ln(budget)$	0.241**	0.265***	0.212**
	(0.018)	(0.002)	(0.020)
Friday	0.497**	0.043	0.034
	(0.045)	(0.848)	(0.833)
genre	Controlled	Controlled	Controlled
MPAA	Controlled	Controlled	Controlled
Constant	2.027	3.546***	2.797**
	(0.189)	(0.008)	(0.033)
Observations	64	64	64
Adjusted R^2	0.879	0.862	0.835

Notes: *z*- or *t*-statistics are in parentheses. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

Table 6
Prerelease "likes" and the box office performance (Fama-MacBeth regression).

Variables	(1) $\ln(REV_t)$	(2) $\ln(REV_t)$	(3) $\ln(REV_t)$
$\ln(prelike_1)$	0.240***		
	(0.054)		
$\ln(prelike_0)$		0.241***	
		(0.066)	
$\ln(\Delta prelike_1)$			0.174*
			(0.098)
$\ln(\Delta like_{t-1})$	0.294***	0.289***	0.271***
	(0.061)	(0.060)	(0.062)
$\ln(screen_t)$	1.037***	1.041***	1.074***
	(0.129)	(0.130)	(0.139)
Friday	−0.395	−0.407	−0.486
	(0.307)	(0.309)	(0.320)
$\ln(budget)$	0.311***	0.313**	0.360**
	(0.148)	(0.151)	(0.161)
genre	Controlled	Controlled	Controlled
MPAA	Controlled	Controlled	Controlled
Constant	−2.032	−1.954	−1.879
	(2.622)	(2.615)	(2.702)
Observations	1779	1779	1779
Avg R^2	0.607	0.607	0.610
Number of groups	28	28	28

Notes: *z*- or *t*-statistics are in parentheses. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

Therefore, the results from the Fama-MacBeth regression complement our earlier findings and further support the strong prerelease "like" impact.

4.4. Addressing endogeneity concerns

Two of the main endogeneity concerns in this study are reverse causality and omitted variables. In the previous section, we have controlled fixed effects to partially alleviate the omitted variables concerns. We further apply the generated instrument method to address the endogeneity issues. Lewbel [37] recently points out that testing the endogeneity in IV-GMM method is equivalent to testing whether a simultaneous equation system is triangular. Specifically, consider the structural model of the form in our study:

$$\begin{cases} \ln(\mathrm{REW_w}) = \gamma_1 \ln\left(\widehat{\mathrm{prelike}_1}\right) + X'\beta_1 + \epsilon_1 \\ \ln\left(\widehat{\mathrm{prelike}_1}\right) = \gamma_2 \ln(\mathrm{REW_w}) + X'\beta_2 + \epsilon_2 \end{cases},$$

where X is the vector of the other exogenous variables. The system is triangular (identified) when $\gamma_2 = 0$, otherwise the system is simultaneous and the previous results are endogenous. Lewbel [37]'s approach achieves the identification by restricting correlation of between ϵ_j and X, where $j = 1, 2$ and suggests the generated instruments, which are defined as

$$Z_j = (X_j - \overline{X}) \cdot \epsilon,$$

where ϵ is the residual vector from first stage regression based on the exogenous variables. The generated instrument Z_j will fulfill the identification requirement, that is, $cov(X, \epsilon_j^2) \neq 0, j = 1, 2$ and $cov(Z_j, \epsilon_1\epsilon_2) = 0$. For more details, see Lewbel [37] and Emran and Hou [21]. Following Wooldridge [60] and Lewbel [37], we construct a generated instrument variable and apply two-stage least square (2SLS) regression as well as instrument variable-generalized methods of moment (IV-GMM) to address the issues raised. Table 7 presents the regressions results.

Empirical evidence shows that the number of "likes" one week prior to release has a positive effect on the opening week box office after controlling for endogeneity. The coefficient of the $\ln(\widehat{prelike_1})$ (fitted natural logarithm of the number of "likes" one week prior to release) is 0.256 in the IV-GMM model and 0.272 in the 2SLS regression, both significant at 10% level. The coefficients of other controlled variables are consistent with previous regression results. The centered R^2 is as high as 0.86 and the identification test (Hansen J statistic) of the generated instrument variable shows that the generated instrument variable is valid and appropriate. Therefore, our earlier results do not appear to be driven by reverse causality or omitted variables.

5. Conclusion

Although box office performance has been investigated from different perspectives related to social media, for instance, blog reviews and Twitter sentiment, such investigations usually concentrate on *post*-consumption opinion mining and sentiment analysis that are difficult to operationalize to the benefits of the industry practitioners, as their objective is to maximize box office sales.

In this study, we propose the Facebook "like" as an effective social marketing tool before the release of movies for several reasons. Firstly, people's prerelease "liking" of movies is generally based on one's preference, which is largely driven by external factors like advertisements and marketing campaigns. Secondly, the clicks of "likes" create social impact, as suggested by the Social Impact Theory, on moviegoers' consumption behaviors. And thirdly, Facebook "like" provides real time updates that are instantly visible by the practitioners and decision makers, so that they can take immediate actions.

Table 7
Controlling for Endogeneity: IV regression.

Variables	(1) GMM estimation $\ln(REV_W)$	(2) 2SLS estimation $\ln(REV_W)$
, $\ln(\widehat{prelike_1})$	0.256*	0.272*
	(0.066)	(0.068)
$\ln(screen_W)$	0.782***	0.808***
	(0.000)	(0.000)
Friday	0.107	0.229
	(0.541)	(0.239)
$\ln(budget)$	0.333***	0.326***
	(0.008)	(0.007)
genre	Controlled	Controlled
MPAA	Controlled	Controlled
Constant	14.799***	14.570***
	(0.000)	(0.000)
Observations	64	64
Centered R^2 (RMSE)	0.861	0.865
Uncentered R^2	0.999	0.999
Hansen J statistic (overidentification test)	0.285	0.285

Notes: *z*- or *t*-statistics are in parentheses. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

In our empirical tests, we borrow from the Social Impact Theory to examine the social impact of prerelease "likes". We measure the number of prerelease "likes" up to one month and box office performance up to one month. Our results indicate that the prerelease "likes" exert a significantly positive impact on the box office performance. More specifically, 1% increase in the number of "likes" in the one week prior to release is associated with an increase of the opening week box office by about 0.2%. In addition, the prerelease "like" impact becomes stronger as it approaches the release date, which suggests that the latest prerelease "likes" are more effective in driving box office performance. These results strengthen marketers' confidence in utilizing the "like" marketing in their prerelease movie promotions. Compared to traditional marketing efforts whose outcomes are difficult to observe and manage in a timely manner, Facebook "like" provides real time updates that are instantly visible by the practitioners. Our empirical results suggest that, to improve box office performance, practitioners should invest more in Facebook "like" marketing as movies approach the release dates since the "like" impact is stronger. Examples of such practices could be Facebook "like" sweepstake events, friends' "like" referrals and free giveaways with "like" clicks. Although this research examines the motion picture industry, we believe the positive social impact can also be extended to a variety of e-businesses such as online crowdfunding and sharing economies.

Admittedly, box office performance can be a result of many unforeseen factors. For instance, celebrity effect where a famous director or actor/actress could make a movie title a big hit, social/political/economics environment, policies and regulations, and many others. The challenge of estimating box office has been tackled from a variety of viewpoints. Researchers have never stopped at discovering new ways to meet the end. The proposed prerelease "likes" in this study serve as a social metric of box office performance. We conduct both cross-sectional and panel data regressions to understand the relationship between prerelease "likes" and box office performance. We also generate instrument variables and employ 2SLS method and IV-GMM to address the potential reverse causality and omitted variables issues. We have concentrated on the social impact while controlling a number of relevant variables. Nevertheless, we are not able to incorporate all the necessary variables in the empirical model given limited availability of data sources and other uncontrollable factors. Although we have applied Social Impact Theory to study the social impact of prerelease Facebook "like", we only look at the number factor while assuming away differences in immediacy and strength. While we believe this is a reasonable assumption given the context of this study, it would be a good research direction to control all three factors. A possible solution would be to conduct a controlled experiment where the flow of social impact can be observed and traced.

Acknowledgements

Funding: This work was supported by The University of Hong Kong's Seed Funding Programme for Basic Research [project code 201409159011]; and PolyU AF Departmental Research Grant [project number 1-ZE6V].

References

[1] A.S. Abrahams, J. Jiao, G.A. Wang, W. Fan, Vehicle defect discovery from social media, Decis. Support. Syst. 54 (1) (2012) 87–97.
[2] A. Ainslie, X. Drèze, F. Zufryden, Modeling movie life cycles and market share, Mark. Sci. 24 (3) (2005) 508–517.
[3] S. Aral, D. Walker, Creating social contagion through viral product design: a randomized trial of peer influence in networks, Manag. Sci. 57 (9) (2011) 1623–1639.
[4] J.J. Argo, D.W. Dahl, R.V. Manchanda, The influence of a mere social presence in a retail context, J. Consum. Res. 32 (2) (2005) 207–212.
[5] S. Asur, B.A. Huberman, Predicting the future with social media, WI-IAT '10 Proceedings of the 2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, 1, 2010, pp. 492–499.
[6] P. Boatwright, S. Basuroy, W. Kamakura, Reviewing the reviewers: the impact of individual film critics on box office performance, Quant. Mark. Econ. 5 (4) (2007) 401–425.
[7] A. Chakravarty, Y. Liu, T. Mazumdar, The differential effects of online word-of-mouth and critics' reviews on pre-release movie evaluation, J. Interact. Mark. 24 (3) (2010) 185–197.
[8] Y. Chen, J. Xie, Online consumer review: word-of-mouth as a new element of marketing communication mix, Manag. Sci. 54 (3) (2008) 477–491.
[9] Y. Chen, Q. Wang, J. Xie, Online social interactions: a natural experiment on word of mouth versus observational learning, 2011. J. Mark. Res. 48 (2) 238–254, http://dx.doi.org/10.1509/jmkr.48.2.238.
[10] J.A. Chevalier, D. Mayzlin, The effect of word of mouth on sales: online book reviews, J. Mark. Res. 43 (3) (2006) 345–354.
[11] P.K. Chintagunta, S. Gopinath, S. Venkataraman, The effects of online user reviews on movie box office performance: accounting for sequential rollout and aggregation across local markets, Mark. Sci. 29 (5) (2010) 944–957.
[12] Danaher, B., & Waldfogel, J. (2012). Reel piracy: the effect of online film piracy on international box office sales. 10.2139/ssrn.1986299 (Available at SSRN 1986299).
[13] A.S. De Vany, W.D. Walls, Estimating the effects of movie piracy on box-office revenue, Rev. Ind. Organ. 30 (4) (2007) 291–301.
[14] D. Delen, R. Sharda, P. Kumar, Movie forecast guru: a web-based DSS for Hollywood managers, Decis. Support. Syst. 43 (4) (2007) 1151–1170.
[15] C. Dellarocas, X.M. Zhang, N.F. Awad, Exploring the value of online product reviews in forecasting sales: the case of motion pictures, J. Interact. Mark. 21 (4) (2007) 23–45.
[16] W. Duan, B. Gu, A.B. Whinston, Do online reviews matter? — An empirical investigation of panel data, Decis. Support. Syst. 45 (4) (2008) 1007–1016.
[17] M. Duggan, J. Brenner, The demographics of social media users – 2012, 2013. Pew Research Center's Internet & American Life Project, 2013 (http://www.pewinternet.org/files/old-media/Files/Reports/2013/PIP_SocialMediaUsers.pdf Accessed 16.10.28).
[18] A. Elberse, B. Anand, The effectiveness of pre-release advertising for motion pictures: an empirical investigation using a simulated market, Inf. Econ. Policy 19 (3) (2007) 319–343.
[19] J. Eliashberg, A. Elberse, M.A. Leenders, The motion picture industry: critical issues in practice, current research, and new research directions, Mark. Sci. 25 (6) (2006) 638–661.
[20] J. Eliashberg, S.M. Shugan, Film critics: influencers or predictors? J. Mark. 61 (2) (1997) 67–78.
[21] M.S. Emran, Z. Hou, Access to markets and rural poverty: evidence from household consumption in China, Rev. Econ. Stat. 95 (2) (2013) 682–697.
[22] E.F. Fama, J.D. MacBeth, Risk, return, and equilibrium: empirical tests, J. Polit. Econ. 81 (3) (1973) 607–636.
[23] R. Garg, M.D. Smith, R. Telang, Measuring information diffusion in an online community, J. Manag. Inf. Syst. 28 (2) (2011) 11–38.
[24] M. Ghiassi, J. Skinner, D. Zimbra, Twitter brand sentiment analysis: a hybrid system using n-gram analysis and dynamic artificial neural network, Expert Syst. Appl. 40 (16) (2013) 6266–6282.
[25] S. Gopinath, P.K. Chintagunta, S. Venkataraman, Blogs, advertising, and local-market movie box office performance, Manag. Sci. 59 (12) (2013) 2635–2654.
[26] T.W. Gruen, T. Osmonbekov, A.J. Czaplewski, eWOM: the impact of customer-to-customer online know-how exchange on customer value and loyalty, J. Bus. Res. 59 (4) (2006) 449–456.
[27] R.G. Hass, Effects of source characteristics on cognitive responses and persuasion, in: P. Richard, T.M. Ostrom, T.C. Brock (Eds.), Cognitive Responses in Persuasion, first ed.Lawrence Erlbaum Associates, Inc., New Jersey 1981, pp. 141–172.
[28] T. Hennig-Thurau, C. Wiertz, F. Feldhaus, Does twitter matter? The impact of microblogging word of mouth on consumers' adoption of new movies, J. Acad. Mark. Sci. 43 (3) (2015) 375–394.
[29] J.M. Jackson, Social impact theory: a social forces model of influence, in: B. Mullen, G.R. Goethals (Eds.), Theories of Group Behavior, Springer-Verlag New York Inc., New York 1987, pp. 111–124.
[30] J.M. Jackson, B. Latané, All alone in front of all those people: stage fright as a function of number and type of co-performers and audience, J. Pers. Soc. Psychol. 40 (1) (1981) 73.
[31] B.J. Jansen, M. Zhang, K. Sobel, A. Chowdury, Twitter power: tweets as electronic word of mouth, J. Am. Soc. Inf. Sci. Technol. 60 (11) (2009) 2169–2188.
[32] E.V. Karniouchina, Impact of star and movie buzz on motion picture distribution and box office revenue, Int. J. Res. Mark. 28 (1) (2011) 62–74.
[33] F.H. Khan, S. Bashir, U. Qamar, TOM: twitter opinion mining framework using hybrid classification scheme, Decis. Support. Syst. 57 (2014) 245–257.
[34] K.K.Y. Kuan, Y. Zhong, P.Y.K. Chau, Informational and normative social influence in group-buying: evidence from self-reported and EEG data, J. Manag. Inf. Syst. 30 (4) (2014) 151–178.
[35] G. Kulkarni, P.K. Kannan, W. Moe, Using online search data to forecast new product sales, Decis. Support. Syst. 52 (3) (2012) 604–611.
[36] B. Latané, The psychology of social impact, Am. Psychol. 36 (4) (1981) 343.
[37] A. Lewbel, Using heteroscedasticity to identify and estimate mismeasured and endogenous regressor models, J. Bus. Econ. Stat. 30 (1) (2012) 67–80.
[38] A. Lipsman, G. Mudd, M. Rich, S. Bruich, The power of "like", J. Advert. Res. 52 (1) (2012) 40–52.
[39] Y. Liu, Word of mouth for movies: its dynamics and impact on box office revenue, J. Mark. 70 (3) (2006) 74–89.
[40] M. Lovett, R. Peres, R. Shachar, On brands and word of mouth, J. Mark. Res. 50 (4) (2013) 1–47.

[41] X. Luo, J. Zhang, How do consumer buzz and traffic in social media marketing predict the value of the firm? J. Manag. Inf. Syst. 30 (2) (2013) 213–238.
[42] L. Ma, A.L. Montgomery, P.V. Singh, M.D. Smith, An empirical analysis of the impact of pre-release movie piracy on box office revenue, Inf. Syst. Res. 25 (3) (2014) 590–603.
[43] McKinsey & Company, Global media report 2015, 2015. http://www.mckinsey.com/industries/media-and-entertainment/our-insights/global-media-report-2015/ (Accessed 16.08.17).
[44] M. Mestyán, T. Yasseri, J. Kertész, Early prediction of movie box office success based on wikipedia activity big data, PLoS One 8 (8) (2013), e71226.
[45] G. Mishne, N.S. Glance, Predicting movie sales from blogger sentiment, Proceedings of AAAI Spring Symposium on Computational Approaches to Analyzing Weblogs 2006, pp. 155–158.
[46] W.K. Newey, K.D. West, A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix, Econometrica 55 (3) (1987) 703–708.
[47] B. Oc, M.R. Bashshur, Followership, leadership and social influence, Leadersh. Q. 24 (6) (2013) 919–934.
[48] H. Onishi, P. Manchanda, Marketing activity, blogging and sales, Int. J. Res. Mark. 29 (3) (2012) 221–234.
[49] A. Pak, P. Paroubek, Twitter as a corpus for sentiment analysis and opinion mining, Proceedings of the Seventh Conference on Language Resources and Evaluation, 10, 2010, pp. 1320–1326.
[50] B. Pang, L. Lee, Opinion mining and sentiment analysis, Found. Trends Inf. Retr. 2 (1–2) (2008) 1–135.
[51] R. Perez-Vega, K. Waite, K. O'Gorman, Social impact theory: an examination of how immediacy operates as an influence upon social media interaction in Facebook fan pages, Mark. Rev. 16 (3) (2016) 299–321.
[52] R. Rishika, A. Kumar, R. Janakiraman, R. Bezawada, The effect of customers' social media participation on customer visit frequency and profitability: an empirical investigation, Inf. Syst. Res. 24 (1) (2013) 108–127.
[53] J.M. Roos, R. Shachar, When Kerry met Sally: politics and perceptions in the demand for movies, Manag. Sci. 60 (7) (2014) 1617–1631.
[54] H. Rui, Y. Liu, A. Whinston, Whose and what chatter matters? The effect of tweets on movie sales, Decis. Support. Syst. 55 (4) (2013) 863–870.
[55] M. Spann, B. Skiera, Internet-based virtual stock markets for business forecasting, Manag. Sci. 49 (10) (2003) 1310–1326.
[56] G. Stasser, J.H. Davis, Group decision making and social influence: a social interaction sequence model, Psychol. Rev. 88 (6) (1981) 523.
[57] C. Trattner, F. Kappe, Social stream marketing on Facebook: a case study, Int J. Soc. Humanist. Comput. 2 (1) (2013) 86–103.
[58] H. White, A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity, Econometrica 48 (4) (1980) 817–838.
[59] F.M.F. Wong, S. Sen, M. Chiang, Why watching movie tweets won't tell the whole story? Proceedings of the 2012 ACM Workshop on Workshop on Online Social Networks 2012, pp. 61–66.
[60] J.M. Wooldridge, Introductory Econometrics: A Modern Approach, fourth ed. South-Western Cengage Learning, Mason, 2008.
[61] Y. Yu, W. Duan, Q. Cao, The impact of social and conventional media on firm equity value: a sentiment analysis approach, Decis. Support. Syst. 55 (4) (2013) 919–926.
[62] F. Zhu, X. Zhang, Impact of online consumer reviews on sales: the moderating role of product and consumer characteristics, J. Mark. 74 (2) (2010) 133–148.
[63] F.S. Zufryden, Linking advertising to box office performance of new film releases-a marketing planning model, J. Advert. Res. 36 (4) (1996) 29–42.

Dr. Chao Ding is an Assistant Professor in the group of Innovation and Information Management at The University of Hong Kong. Chao obtained his doctoral degree in Information Systems from the University of Florida in 2014. Chao also holds a B.S. in Optical Engineering from Sun Yat-sen University and an M.S. in Economics from Hong Kong University of Science and Technology. His primary research interests include Econ of IS, digital goods, e-commerce, crowdfunding and social media.

Dr. Hsing Kenneth Cheng is the John B. Higdon Eminent Scholar and Chair of Department of Information Systems and Operations Management of Warrington College of Business Administration at the University of Florida. Prior to joining UF, he served on the faculty at The College of William and Mary from 1992 to 1998. He received his Ph.D. in computers and information systems from William E. Simon Graduate School of Business Administration, University of Rochester in 1992. Professor Cheng teaches information technology strategy, object-oriented analysis and design, managerial statistics, electronic commerce, and supply chain management. He was awarded the Warrington College of Business Administration Teacher of the Year for 2000–2001, and Outstanding Faculty Award, for service and teaching excellence to Professional MBA Class of 2012, of Warrington College of Business Administration, University of Florida. Dr. Cheng's research interests focus on analyzing the impact of Internet technology on software development and marketing, and information systems policy issues, in particular, the national debate on network neutrality.

Dr. Yang Duan is an Assistant Professor in the Department of Finance and Decision Sciences at Hong Kong Baptist University. Yang obtained her doctoral degree with a concentration in Corporate Finance at the Chinese University of Hong Kong in 2013. Yang also holds a B.A. and M.A. in Economics from Nankai University. Her primary research interests include corporate finance, financial markets, social media and e-commerce.

Dr. Yong (Jimmy) Jin is an Assistant Professor in the School of Accounting and Finance at The Hong Kong Polytechnic University. Yong obtained his doctoral degree in Finance from the University of Florida in 2016. Yong also holds a B.S. (Honors) and an M. Phil. in Risk Management Science from the Chinese University of Hong Kong. His primary research interests include social media, asset pricing and financial markets.