Journal of Cultural Economics

Do consumer and expert reviews affect the length of time a film is kept on screens in the USA

Thaís L. D. Souza☆
School of Arts, Science and Humanities – University of Sao Paulo
donega.s@gmail.com

Marislei Nishijima
Institute of International Relations – University of São Paulo, Av. Prof. Lúcio Martins Rodrigues, s/n, travessas 4 e 5, Cidade Universitária- 05508-020 - São Paulo -SP – Brazil
marislei@usp.br

Ana C. P. Fava
Department of Economics, Federal University of ABC
ana.fava@ufabc.edu.br

Abstract: We evaluate the effect of critical reviews by consumers and experts on a film's running time at movie theaters in the United States using survival regression analysis. In addition to the usual expert critics' reviews, we employ the consumer reviews rating and their affectivity about films as proxies for the consumer influence effect. To provide measures for consumer affectivity, we perform affective computing using mining techniques of sentiment and emotion on consumer reviews. We build a very rich film dataset by collecting information from the Box-Office Mojo and the Rotten Tomatoes sites, including all matched films released between 2004 and 2015 that are available on these sites. We find evidences of consumer ratings matter in keeping a film running longer at the theaters, but experts' ratings have a larger influence on the movie market as a whole. Estimates by genre indicate that the influence of expert reviews on the length of run of widely opening film releases, which include blockbusters, is null, but that their influence on narrowly released films is large. Also, film running times of genres like foreign, drama and action films are greatly influenced by sentiments and emotions spread by consumers through their reviews.

JEL Codes: L82; D83;

Keywords: Consumer reviews; Movies; Film; Affective computing; Survival; Critic Reviews

☆ Corresponding author

We thank Ricardo Pires, Patrícia R. Oliveira, Adriana Schor, Marcelo Fantinato, Alexandre Brincalepe Campo, Bruno Faria Freitas, and the anonymous Reviewer of the JCE for their helpful comments and guidance. All errors are our own.

Do expert and consumer reviews affect the length of time a film is kept on screens in the U.S.?

1 Introduction

We investigate whether consumer reviews influence the length of the run of a film at U.S. movie theaters as much as expert reviews of films released between 2004 and 2015. While our outcome of interest is similar to McKenzie (2009) and Legouax et al. (2016) with regard to the length of time a film is kept on screens, we investigate the reduced demand form. Using a very rich weekly set of data for the period from the Box Office Mojo and Rotten Tomatoes websites, we estimate survival models to evaluate the influence of consumer and expert reviews on the number of weeks a movie is screened.

Online forums have reduced the cost of information searching, which increases the number of consumers who use them, both to grab information and to spread it by means of user reviews (Stigler, 1961). Thus, since consumers (or users as the literature refers to consumers who use the Internet to share their reviews) have the same channel of communication as expert reviewers, their opinion may matter for potential consumer choices (Chintagunta et al., 2010). In addition, as Internet access is widening and developing this channel of communication and influence among consumers, it may strengthen their relative importance vis-a-vis expert reviewers.

The main contribution of the paper is to perform an "affective analysis" for consumer reviews, where we employ both a "sentimental analysis" (connected to sentimental valence, good or bad evaluation) and an "emotional analysis" (connected to feelings like anger, joy, fear, disgust and sadness). To our knowledge, no other paper uses emotional analysis to investigate its effects on film success. The only previous works using sentimental analysis were found as a means to find polarity (in non-labeled texts (Rui, Liu, and Whinston, 2013; Vujic and Zhang, 2017).

To perform emotional analysis, we mine the online consumer/user review texts, using IBM Watson tools of the Nature Language Understanding set. This allows us to create variables able to represent the feelings a person spreads to his peers, or alternative measures for the effect of consumer reviews in addition to the traditional consumer rating/score that is mandatory for someone to write a critic review on the online forum investigated.[1]

To avoid the endogeneity problem in measuring the influence of expert and consumer reviews on screening length, where it is difficult to determine whether the review or the film quality is responsible for high demand (Reinstein and Snyder, 2005), we control our estimates for film quality by including both sources of reviews simultaneously. Since both experts and consumer critical reviews potentially suffer from the duality between "influence" and "prediction" as noted by Eliashberg and Shugan (1997) and both are quality information about films, including them jointly works as a mutual control for film not observable quality. In this sense, the consumer and expert reviews are mutually cleaned off the effect of the quality and their estimates may reflect the consumption influence. In addition, we include all observable variables available that bring quality information, as stars, genders and others.

Also, we control for bias of quality information provided by experts using the average of all expert reviews made available before and during the first weekend. Doing so, we avoid the bias of a particular expert reviewer to deviate from the unique subject of providing quality information for the less informed population, or the consumers, as pointed out by Reinstein and Snyder (2005)

---

[1] There is no full documentation of Watson tools available, but they employ deep learning techniques and became famous after winning the game of Jeopardy against human beings. https://www.ibm.com/watson/.

who investigate only two expert reviewers. We suppose that most of the experts are independent and committed to provide accurate information about film quality since bribing a huge number of reviewers would be very costly. The choice of time before and during the first week is also suitable and representative, since most of the experts' reviews are available before the film release or, as exceptions, when the review is done at the weekend release, and we control for endogeneity of just describing facts.

As distinct from Legouax et al. (2016), who study the effect of critical reviewers on theater owners where both sides have full information about the quality of a film, our concern is how reviews from informed people (expert and consumers who are supposed to have consumed the experience good) affect consumer decisions.

As the consumer reviews are available only after the film release, we use these reviews after the week of release to investigate their effect on screening time at theaters. To avoid the endogeneity problem, in addition to controlling for expert reviews' quality signs, we also employ exclusively quality measures in averages, since the measures that bring information of quantity (as volume of reviews) can be connected with the number of people going to the movies, and then with the success of the film and the length of time it is screened at the theaters.

Our findings indicate that consumer ratings matter to keep a film running longer at the theaters, but expert ratings have a much larger influence on film consumption as we investigate all genres of films at once. For specific genres such as action, adventure and animation that include blockbusters films, however, the consumer reviews matter for the length of time a film is kept on screens, while no significant influence of expert reviews is found.

In addition to this introduction, the second section discusses the literature on the effects of consumer and expert critical reviews on a movie's revenues and other correlated works. The third section presents the data and our empirical strategy, followed by section 4, where the main results are discussed. Finally, in section 5, we summarize our main conclusions and comments.

2. The importance of critic reviews for consumption and assumptions

Film is an information good that can be digitized, encoded, transmitted and stored as a stream of bits, and is also an experience good since its quality is revealed only after consumption (Shapiro and Varian, 1999). Thus, there is consumer uncertainty about a film's quality before its consumption, and this creates an information asymmetry between consumers and the more informed film producers and distributors, although consumers may have prior knowledge of quality based on genre, actors, director, ratings and budget.

Information theory argues that consumers will seek information about the product in order to reduce the risk associated with its purchase and consumption (Stigler, 1961). The role of this information is especially critical for experience goods, since the quality is unknown before consumption (Eliashberg and Shugan, 1997).

Historically, to deal with film's nature as an experience good, the motion picture industry relies on some signals of film quality spread to influence consumers at the time of a film's release, namely expert reviews. In general, an expert reviewer watches the film in advance, before its release in movie theaters, writes comments about the film and rates it according to a grade scale. Depending on the quantity of films released at a specific weekend, however, the expert critical reviews can arrive after the release, but in general no later than the first weekend.

There is some literature in different research areas about the effects of expert critical reviews on movie theater consumption (Boatwright, Basuroy and Kamakura, 2007; Kamakura, Basuroy and Boatwright, 2006; Reinstein and Snyder, 2005; Eliashberg and Shugan, 1997; Duan et al., 2008; Chintagunta et al., 2010; and Moon et al., 2010). These studies suggest that expert reviews have a positive and significant effect on consumption.[2]

In addition, there are studies that explore the effect of consumer reviews on word-of-mouth (WOM), which operates through peer effects, and studies that explore online consumer reviews, where consumers use open online forums to opine about films (Liu, 2006; Chintagunta, Gopinath and Venkataraman, 2010; Duan, Gu and Whinston, 2008b; and Duan, Gu and Whinston, 2008a).

Moretti (2011) has undertaken a comprehensive study about social learning and peer effects on movie sales and finds strong effects on movies, particularly on films with positive surprise, when the previous signals of quality of a film are lower than its actual quality revealed in the first week of release. The author uses a framework to deal with the phenomenon, which generates predictions for the observable variables (movie data from U.S. movie box offices from 1982 to 2000) under the influence of social learning and peer effects. Doing so, he avoids the need for consumer reviews and WOM variables.

According to Vujić and Zang (2017), the rapid rise of text-based social media and online WOM activity increases the importance of consumers in guiding other consumers. Performing a sentimental analysis, the authors use Twitter messages (tweets) to study the effect of WOM on cinema box office revenues. Based on static and dynamic panel data regression approaches, they conclude that negative sentiment is more damaging to box office revenues.

Regarding the influence of expert critical reviews, there are many papers addressing their effects on films' success measured by their revenues. More recently, this literature has focused on the potential endogeneity of the expert reviews, questioning whether they affect revenues (influence effect) or just reflect them (prediction effect). To address this concern, the general

[2] We perform an exploratory review in economics, business and computer areas in the main research machine learning, results available with authors.

procedure is to control the effects of critical reviews on film success measures for the time a film is kept on screens. The general understanding is that the influence is a short-run effect (Reinstein and Snyder, 2005; Eliashberg and Shugan, 1997; Basuroy et al. 2003).

Eliashberg and Shugan (1997), who conducted one of the first studies to examine this endogeneity problem, based on weekly correlations, argue that experts only have a prediction effect because they raise box offices for the entire screening time, as opposed to an influence effect, which can be seen in changes of first week revenues.

Basuroy et al. (2003) explore the hypothesis that the influence effect occurs in the short run and the prediction effect in the long run by estimating correlations of expert reviews, replicating Eliashberg and Shugan's (1997) method, and performing regressions. The authors find significant effects of both prediction and influence of experts' negative and positive evaluations. Regarding the influence effect, they argue the negative evaluations have a greater impact on revenues as compared with positive evaluations.

According to Reinstein and Snyder (2005), the causal effect of reviews on demand holding quality constant is the influence effect, while the spurious correlation between reviews and demand induced by their mutual correlation with quality is the prediction effect. The authors argue that there is a spurious correlation between expert reviews and demand for experience goods due to their common covariance with unobserved film quality signs. This may occur because positive reviews can be a sign of the high quality of a film, enhancing demand for it. Controlling for quality film information with variables such as WOM, publicity, and marketing allows the estimate of expert reviews measure the influence effect. Particularly, as WOM influences are not observable by econometricians, the authors use a difference-in-differences model and critical reviews only of the weekend of movie release to deal with the endogeneity between the revenues and expert reviews. Their results suggest different positive effects across categories of movies, the strongest for narrowly released movies and dramas.

Similarly, Gemser, Van Oostrum, and Leenders (2007), studying the Dutch film industry, find that expert reviews affect art movie revenues while no effect is found on mainstream movies. The authors focus on the effects of reviews on the opening weekend and on cumulative box office revenue. They use the number and size of film reviews in Dutch newspapers to measure the review effect.

The literature on the effect of consumer reviews on movies, which is connected to the WOM effect, is similar to the literature on expert reviews. Endogeneity concerns are the same, since consumers may just reflect film quality. The studies measure consumer reviews by their average rating, volume (number of reviews), variance and valence (if the critic is good or bad), while typical investigations use the volume of user critic reviews as the main measure. Evaluating these three effects on local geographic box office performance, Chintagunta et al. (2010) find significant results for valence in contrast with previous studies that usually get significant results only for the volume effect of critic reviews.

Duan et al. (2008b) investigate the effect of online consumer/user reviews on movies' daily box office performance using different lagged measures of user critic reviews to avoid endogeneity from reviews both influencing and being influenced by movie sales. Using 2003-2004 data from Box Office Mojo, Yahoo! Movies and Variety.com, their results point to no significant effect of the user review online ratings on movie box office revenues, but they do find a positive effect of the online posting volume on revenues, interpreting this as a WOM awareness effect.

McKenzie (2009) estimates a structural model for the Australian movie market to measure the success (defined as duration) of 360 films staying on screens using the duration/survival approach. Using a framework of survival analysis, the author controls for the WOM effect and concludes that screening time is affected by previewing, advertising, critical reviews and the U.S. box office.

We argue that WOM works as a very important channel of consumer influence, since it reduces uncertainty about film quality. The massive availability of Internet usage and the large proliferation of online consumer reviewer systems and social networks, provided by Web 2.0, enable widespread WOM. As a result of the Internet and increased Internet broadband usage, WOM, which represents a consumer-to-consumer connection, broadens the influence of one individual by connecting a unique person to a very large number of peers (Chen and Xie, 2008; Eliashberg, Elberse, and Leenders, 2006; Godes and Mayzlin, 2004; Godes et al. 2005; Mayzlin, 2006; Trusov, Bucklin, and Pauwels, 2009).

We go further, arguing that consumer reviews, both the online ratings and variables mined from online text reviews, obtained by means of sentimental and emotional analysis tools, can be good proxies for WOM among individuals' peers. According to Pollai, Hoelzl and Possas (2010), individuals often consider emotions during consumer decision-making.

Considering the literature, we make a number of assumptions to ensure that our model correctly estimates the effect of reviews on film running time. First, we assume that critical reviews on the Internet, both from experts and consumers, are good proxies for critical reviews in general, particularly for expert reviews, since printed content from magazines, newspapers and other media, which was traditionally the experts' channel of communication, is also available online. Second, we assume that the Internet has improved and increased the social network of an individual's influence among his peers (Chen and Xie, 2008; Chintagunta et al., 2010). Lastly, since critic reviews are also experience goods, and tradition matters as a signal of quality for the suppliers, our third assumption is that the consumer influence effects are lower than the experts.

3 Data and Methodological Strategy

3.1. Data

To build the dataset for this research, we created a Python scrapper script. Then, we collected all financial and technical information available on all films from 2004 to 2015 from the Box Office Mojo site. We merged this information with critic review information, from both consumers and experts, from the Rotten Tomatoes site based on movie name and year of release using another Python script. As result, we built a unique weekly database with around 42,000 observations for 4,700 movies. We choose this range because the Rotten Tomatoes user's area was launched in 2003, when consumer critical reviews became available.

We deflated the monetary values using the U. S. monthly consumer price index (CPI) from International Monetary Fund (IMF) statistics. In addition, we used information from Forbes magazine to classify the main actors in the films. We labeled an actor as a big star if she/he was on the list of highest paid actors in the two years prior to the film release. Our sample includes all genres, except documentary, due to its educational nature, and IMAX films because of the very long time they stay on screens, sometimes more than 6 years.

In order to obtain proxy variables for the WOM effect on consumer reviews, we conducted "affective analysis'" on the consumer – or "user" as the consumer is called on the review sites – and on critic reviews available on the Rotten Tomatoes site. We employed the cognitive computation tools of the Watson AlchemyLanguage, an application program interface (API) that uses deep learning, available on its developing platform, IBM Bluemix.[3]

For the purpose of dealing with data and building the affective variables, we created a script in Python language. Our "affective analysis" includes both "sentimental analysis" and "emotional analysis." The first is connected with sentiment in its polarity or valence regarding an opinionated text, which may be positive or negative, while the absence of opinion is classified as neutral or indifferent sentiment (Cattell, 1940). On the other hand, emotion is a state that goes beyond an opinion, and not does necessarily have a target. It involves feelings like anger, joy, disgust, fear and sadness (Fehr and Russell, 1984). For simplicity, we employ these five basic emotions to create our emotion variable set, although there is a broad range of theories about emotions, as suggested by Plutchik (2001), with a wider spectrum of feelings and diverse classifications. See Figure A1 in the appendix for more information.

Table 1 depicts our consumer and expert review variables. We create two sets of affective variables as proxies for WOM effects using IBM Watson: (i) performing sentimental analysis, we produce a valence set, where we classify a film as positive, negative, or absence of sentiment (or neutral) by counting the majority opinions; and (ii) performing the emotion analysis, we generate the 5 above-mentioned emotion variables, using grades from Watson, calculated according its algorithm of cognitive computation, and divided by the total of reviews available for a film in different periods. See Figure 1 for variables at different periods.

In addition to the consumer review variables from the affective analysis, which is our main contribution to the literature, we also include two more variables in our estimates. The first is a consumer review score, which ranges from zero to five stars and comes from the Rotten Tomatoes

[3] https://alchemy-language-demo.mybluemix.net/ and https://console.ng.bluemix.net.

website when the consumer posts his or her review (iii). This score is an average of consumer ratings for the film. The second is an expert review in binary variable that comes from the site's own film classification of good or bad, according to the positive or negative valence binary concept.

To measure the effect of the expert reviews, as they give the most traditional quality sign for consumers, we employ two variables of the top reviewers' average. The first is the top reviewers' positive evaluations divided by the total of top reviews, which can be positive or negative according to the valence binary concept, i.e., good or bad (iv). Note that each film receives many evaluations, sometimes hundreds of them, and all the information is available on Rotten Tomatoes. The second variable is a film classification of very good, good, bad, or neutral, and is available directly from the website[4] (v).

Table 1 - Critic Review Variables

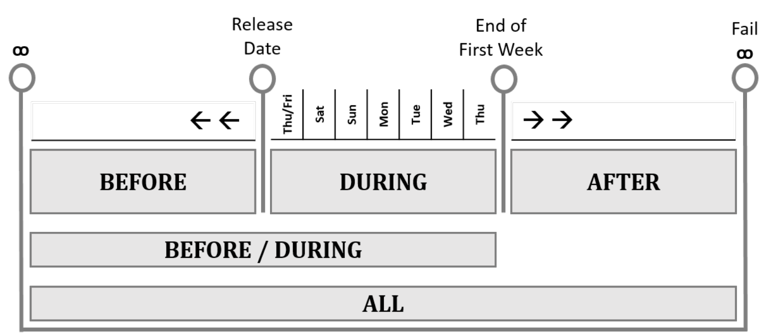
Variable	Obs	Means	SD	Min	Max
(i) Consumer Sentiment Variables					
Sentiment Positive	42.274	0,4938	0,4999	0	1
Sentiment Negative	42.274	0,1629	0,3693	0	1
(ii) Consumer Emotion Variables					
Anger Average	27.954	0.1205	0.0377	0.0025	0.5106
Joy Average	27.954	0.3271	0.0992	0.0019	0.9319
Fear Average	27.954	0.1348	0.0612	0.0027	0.7739
Sadness Average	27.954	0.4191	0.0702	0.0329	0.8833
Disgust Average	27.954	0.1102	0.0397	0.0044	0.5826
(iii) Audience Average (Consumers)	32.692	3,4191	0.5936	0.5	5
(iv) Consumer Evaluation -Values on the Site					
Good	42.274	0,5245	0,4994	0	1
Bad	42.274	0,3637	0,4811	0	1
(v) Top Critics Average (Experts)	32.497	0,5696	0,2723	0,025	0,98
(vi) Critics General (Experts)					
Top Very Good	42.274	0,3049	0,4604	0	1
Top Good	42.274	0,1511	0,3581	0	1
Top Bad	42.274	0,3657	0,4862	0	1

Source: Mining data from the Rotten Tomatoes website

[4] We follow Boatwright et al. (2007) in using only the top critics' reviews instead of all professional reviewers.

In order to build a dataset of critical reviews and avoid problems of biases of estimates, we use previous reviews or reviews during the first weekend of film release. All critical review variables are split into the following periods: before, during, before/during, and after the film's release (see Figure 1). The "before" period includes all reviews prior to release, as of 1996. The "during" period includes the first week of the release, including the release date, but before the second weekend, where the weekend starts on Friday. The "before/during" period includes both the previous periods; and finally, "after" means all critical reviews done from the second week to one year after the release.[5]

Figure 1: Timing of critical reviews used in survival analysis

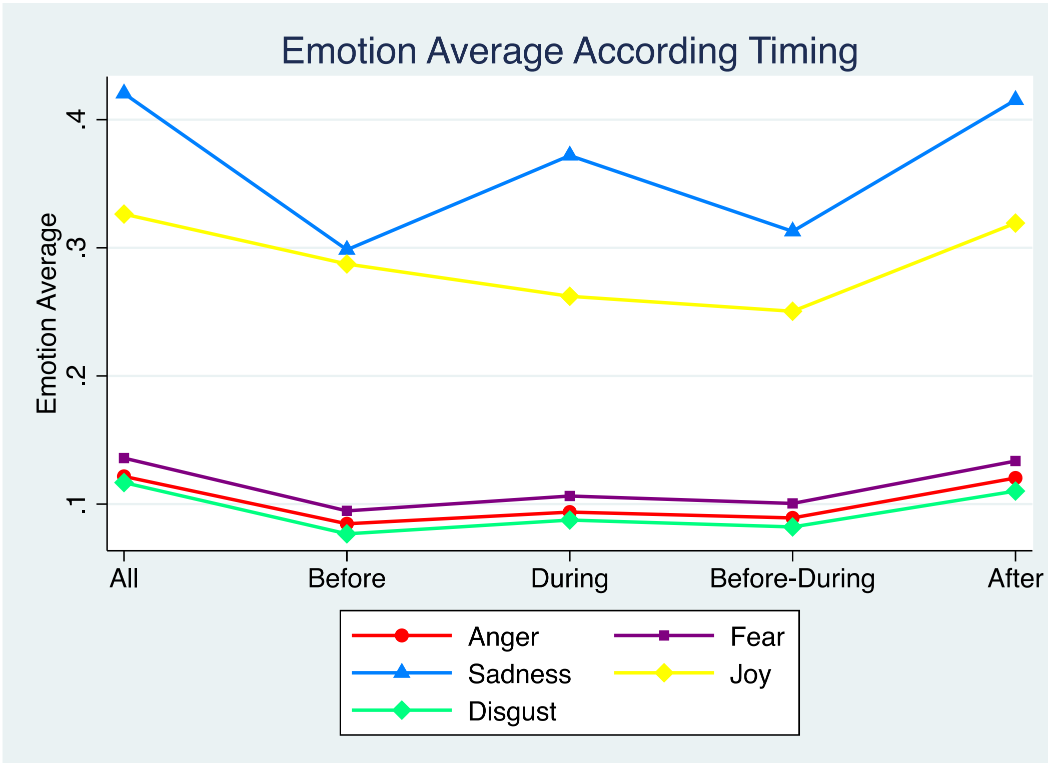


Source: Authors

To reinforce the need to control for the critics' reviews for different time periods, it is worthwhile to note that their measures vary accordingly. Figure 2 illustrates the emotion average by period, where in the after period, all emotion averages reach their highest value.

[5] We use this year as an average for "the date of broadband release in the U.S." based on Lee and Chan-Olmsted (2004), who indicate that Internet broadband in U.S. with DSL technology was developed by the early 1990s, but was not marketed until 1999.

Figure 2 – Emotions average by period of time



Source: authors using data from the Rotten Tomatoes site

3.2. Methodological Strategy

We use a survival analysis model (Hosmer, Lemeshow, and May, 2008) to study the effects of our interest variables on the time a film is screened at the movie theaters. We follow McKenzie (2009), Eliashberg and Shugan (1997), and Legouax et al. (2016), who study the effects of a set of variables on the duration time of screening, but we keep the general idea of explained variable regarding demand side. Focusing on movie lifetime provides an alternative to traditional gross metrics by considering the interaction between supply and demand in a dynamic configuration where the life expectancy of the film is the aggregation of the decisions to maximize the profits of the exhibitors in response to weekly demands. Thus, we focus on demand reduced form to investigate the reviews' effect on screening time of a film.

The traditional framework of survival analysis to study the effect of critical reviews on the screening time of a film at the theaters can be described as the following: *T* is the time an event occurs, with *T* being non-negative and continuous with a probability density function, $f(t)$, and cumulative density function $F(t)$. The survival model, $S(t)$, is the probability of survival timing be greater or equal to *t*, or $S(t) = \Pr(T > t) = 1 - F(t)$.

We estimate the non-parametric models without control variables. Additionally, following Allison (1984), we estimate parametric models with control variables, including our interest variables, evaluated by the estimates of β's. and controls.

$$ln\, S_i = \beta_0 + \beta_1 C_i + \beta_2 M_i + \beta_3 X_i + e_i \tag{1}$$

S_i is the timing a film $i = 1, \dots, I,$ is on the screens measured in the number of weeks until time T_i . Being $\beta_0, \beta_1, \beta_2$ and β_3 vectors of coefficients and e_i is an error term distributed according to Weibull distribution. β_1 refers to the average grades of expert reviews, C_i, while β_2 refers to the consumer's evaluations, M_i, which includes word of mouth measures. Thus, , M_i represents our proposed variables to measure consumer reviews.

According to Reinstein and Snyder (2005) and Eliashberg and Shugan (1997), the causal effect of reviews on demand is the influence only if quality of film is held constant. On the other hand, there is a spurious correlation between reviews and demand, which corresponds to the prediction effect. To circumvent this problem we introduce both C_i and M_i measures in the same regression, since both terms, which are also quality measures, can be used to control mutually by the quality of the film and allow their associated coefficients to represent the influence effect. Note that, differently from Reinstein and Snyder (2005), we do have measures for consumer reviews, then the estimated coefficient of C_i reflects the influence effect since the prediction effect, connected to the film quality, is controlled by the inclusion of the M_i variables, and the reverse is true for the coefficients of M_i.

In addition, we add many other observable film quality variables as control variables in the X_i matrix, which includes the following variables (see summary statistics presented in Table A1), where we try to cover all relevant controls to avoid omission biases:

i) Months: a set of dummy variables of the movie release month. This controls for advertising costs because movies released during summer and year-end vacations usually have higher ad expenditures. It also intends to control for eventual seasonality, when there are more cinematographic productions, like the periods that precede the awards (Reinstein and Snyder, 2005).

ii) Years: a set of dummy variables for the year of film release, from 2004 to 2015, aiming to control for household income and different policies for the sector and macroeconomic effects that affect the movie industry.

iii) Logarithm of the number of movie screens at which the movie is shown on its release weekend: this variable also controls for movie budget and advertising since these two variables are highly correlated. The number of theaters at the time of the premiere reveals the distributor's bet on the success of the film (Moretti, 2011), while still controlling for narrow releases and productions, whose number of screens is commonly less than 50, and for number of theaters.

iv) Big distributors: dummy variables for distributors that hold large amounts of the film production in the sample. This variable also controls for advertising and budget, since bigger studios have larger budgets for their productions.

v) Big stars: a binary variable to account for the presence of actors or actresses whose names and receipts were mentioned in the Forbes list of highest-paid actors. For each listed actor, his or her presence in the film was counted for up to two years following the year of the list.[6]

vi) MPAA rating: dummy variables of the indicative classification of the film. The variables were separated by up to 16 years (MPAA-PG classification) and more 16 years (MPAA-R).

vii) Genres: dummy variables for the genre according to Rotten Tomatoes. More specifically, we consider the following genres: action, comedy, drama, romance-family, foreign, crime-war-western, horror, adventure-animation, thriller and others. The last includes the remaining genres and is our reference group. Note that unique films may frequently be classified in two different genres; if so, we include the film in both genre variables accordingly.

viii) Dummy of cold opening: which has a value of 1 if the film is not reviewed before its release, but only after. This variable was eliminated from the model because of its low frequency in the data and high correlation with other controls.

ix) Logarithm of budget: which was eliminated because of being high collinear with control (iii), and we opted for keeping the last because we have budget information only for less than one third of the films.

We use measures of expert reviews in their averages including all available reviews in order to avoid bias of capture of some specific reviewers by film producers as Reinstein and Snyder (2005) observe. In our sample, the average of top reviewers classifying a film between good or bad is 29.8 experts (see the distribution of total top critics reviews in Figure A2 in the annex).

---

[6] https://www.forbes.com/actors/#5825be4040fe.

Brown, Camerer, and Lovallo (2012) discuss situations in which producers withhold movies from critics before their release, in what they refer to as cold openings. The authors find a positive effect on gross revenues using this strategy. Using a huge number of films, we believe to mitigate the effect of this mechanism in our results, since our analyses are aggregated to all films or to all films of a specific genre.

Regarding consumer reviews, we also employ measures of averages of many consumer reviews to avoid biases of specific consumers sharing their opinion on the Internet toward their preferences in these quality signals. Additionally, the large number of reviews eliminates the problem of weakness of computational affective analysis regarding the difficulty of dealing with circumstances like irony, for example. Yet, using average measures allows us to control for potential endogeneity between a number of reviews and the screening duration, since a large number of reviews may also reflect successful demand and a longer length of screening at the movies.

We employ consumer reviews only after films releases, see Figure 1, since we consider they have access to the film and full information predominantly after the release. Thus, we include those available only after the week of film release, since consumers are thought to be able to review a film only after consuming it, which is an information good and an experience good as well. Doing this, we avoid personal perspective biases of people who have not seen the film and do not know its quality, which is common in comments before and at the week of film release, or just a reflection of the expert reviews already made available.

Studying the effect of consumer/user critic reviews on a specific hotel demand, Mayzlin, Dover, and Chevalier (2014) argue the consumer/user reviews may be manipulated by the suppliers and may not be credible. Since we use an average of a large sample of critical reviews, however, we believe the manipulation that could bias our estimates is diluted due to the sizeable number of expert manifestations.

Authors such as King (2007) and Reinstein and Snyder (2005) find that expert reviews tend to have a greater influence on narrowly released films. Thus, we estimate the critic reviews on all movie markets and consider six genres – 3 of them wide release, including blockbuster films, and 3 narrow release, where we include drama that according our data is kind of medium release – in order to investigate potential differences of influence between expert and consumer reviews according to genre.

Summarizing our strategy to account for endogeneity problems: a) we include both expert and consumer reviews in regressions to control for film quality, besides other controls of quality such as big stars and budget proxies; b) we use expert reviews before and during the first weekend of film releases; and c) we use measures of quality in averages. It is important to note, however, that, since our subject is to compare the relative importance of expert and consumer reviews and they are under the same type and direction of bias, our relative results are consistent.

4. Results and discussion

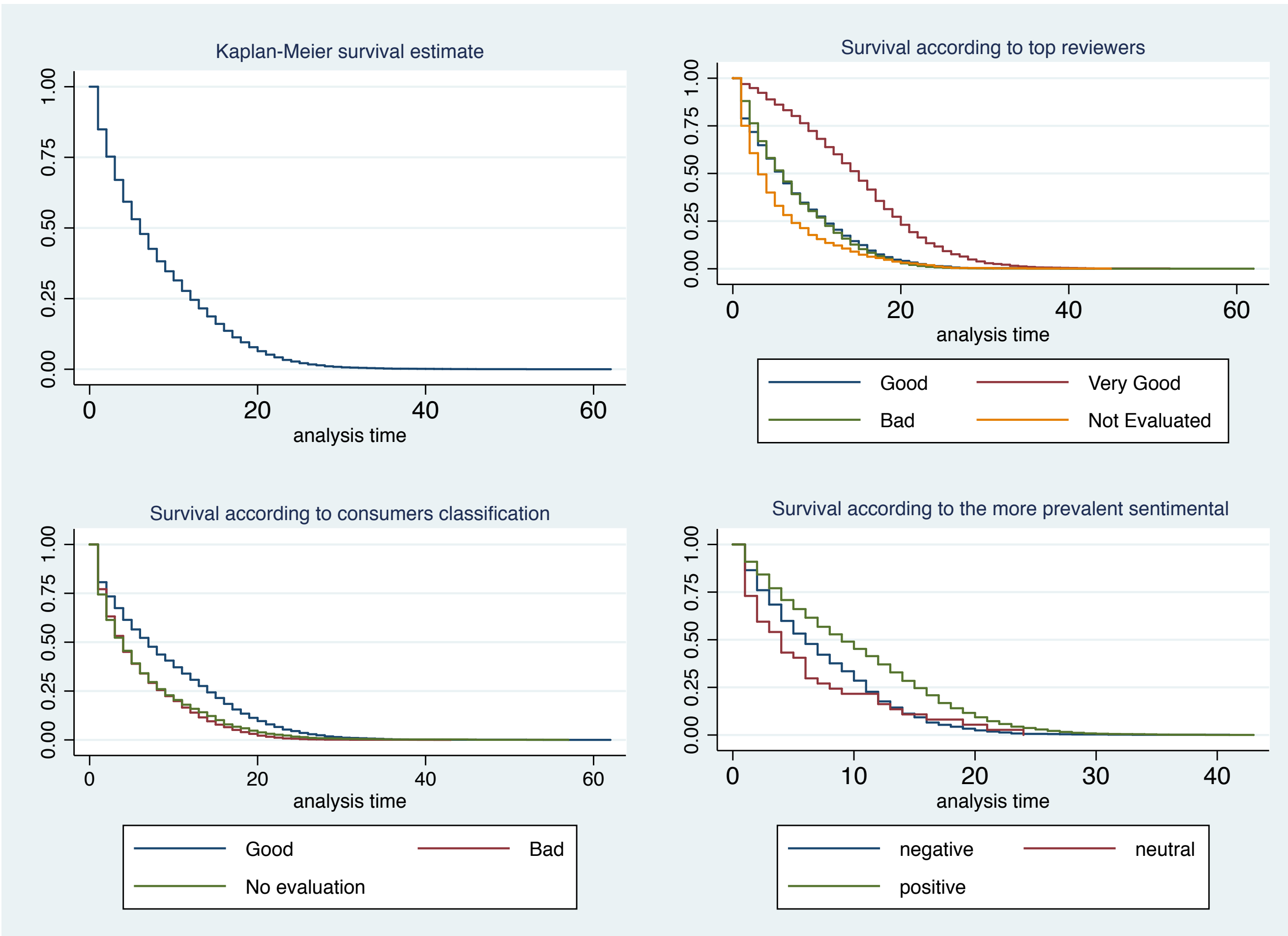
4.1. Non-Parametric Results

The non-parametric models without control variables can be seen below. Figure 3 shows the following films' survival graphics of Kaplan-Meier estimates:

Upper- left) Non-parametric survival estimates according the number of weeks a film is kept in the movie theaters. The longest time that a film is kept in the theaters is 62 weeks; in 7 weeks, however, 50% percent of the films are no longer on the screens.

Upper right) Estimates according to the general label from experts, extracted from Rotten Tomatoes. The non-evaluated movies have a shorter life expectancy compared to negative and positive reviews by experts. The results are consistent with Shrum (1991), who says “even a mediocre or negative evaluation is better than no evaluation at all…” (Shrum,1991, p. 368).

Figure 3: Survival models without controls according categorical critical reviews.



Source: Authors using data from the Box Office Mojo and Rotten Tomatoes sites

Lower left) Estimates according to the general label from consumers/users, extracted from Rotten Tomatoes. The non-evaluated movies and those with bad evaluations have a shorter life expectancy as compared to those with good reviews by consumers.

Lower right) Kaplan-Meier estimates according to the sentiment variables from our affective analysis. Note that movies with positive consumer reviews have longer life expectancies, while movies with neutral consumer reviews generally perform worse than those with negative reviews. Here again, the results are similar to those posted by Shrum (1991).

4. 2. Parametric Results

We estimate the model described in equation (1) in four specifications (I, II, III, and IV), according the type of critical reviews as shown in Table 1. Specifications (I) and (II) are our baseline models and use the directly observable icons or scores on the Rotten Tomatoes website as the critical reviews data. Our baseline models control for endogeneity since we include both expert and consumer reviews as discussed in section 3; however they are not controlled for potential endogeneity due to the timing of reviews.

These variables split in time are not available directly according the date of the review. Thus, to overcome the problem, we created other review variables by collecting every expert and consumer review and generating measures in averages of reviews according the timing each review was posted on the website (see Figure 1), as described in section 3.

Specifications (III) and (IV) are controlled by the time the review is posted, according the classification depicted in Figure 1. Thus, these specifications are free of potential bias due to circularity among variables expert reviews and dependent variable.

Regarding consumer reviews, we include those available only after the week of film release in order to avoid personal perspective biases of people who have not seen the film and do not know its quality and reviews that just reflect expert reviews.

Thus, in the last specifications we improve our estimates by just including expert reviews made before and during the week of film release. Since most experts review a film before or at the weekend of release, to include only reviews of this period is representative and avoids potential endogeneity due to double causality: expert reviews affecting duration, and duration affecting new reviews by experts.

For all specifications, (I, II, III, and IV), the vector $\boldsymbol{X}_i$ contains the control variables discussed in section 3. All the survival model estimates were performed in accelerated forms (AFT); thus positive coefficients mean longer life expectancy at the theaters, while negative coefficients imply the opposite. Table 2 resumes our 4 specifications and refers to the variables listed in Table 1. We estimate each specification for our whole sample and for six genres, where three can be considered narrow opening (comedy, drama and foreign) and the other three can be considered wide opening and include blockbuster films (action, adventure/animation, and romance/family).

Table 2 – Model specification

Model Specification	Expert reviews variable - C_i	Consumer reviews variable - M_i
(I) Baseline 1	(vi) in Table 1 with no control of timing	(iv) in Table 1 with no control of timing
(II) Baseline 2	(iii) in Table 1 with no control of timing	(v) in Table 1 with no control of timing
(III) Main 1	(iii) in Table 1 with control of timing – Before and During	(v) in Table 1 with control of timing - After
(IV) Main 2	(iii) in Table 1 with control of timing – Before and During	(i), (ii), (v) in Table 1 with control of timing – After

In our baseline models (I) and (II), Tables 3 and 4 respectively, we show that both experts and consumers affect the time a film is kept on screens among all genres, but the influence of consumers is much less than that of the experts. In addition, when we compared the influence between narrow and wide releases, we found bigger and more significant effects for experts and similar effects of consumer reviews between them.

In specification (I), we see that the top reviewers have a positive effect on the screening time when they evaluate a film as good or very good, except for the adventure, animation and action genres where only top experts' good evaluations are not significant. Using control variables in comparison with the non-parametric estimates, we see that bad evaluations by experts are not different from films that have no review, our references categories. Good evaluations by consumers as compared with bad evaluations are positive and statistically significant, similar to the professional evaluations; however, the coefficients size around 20% of the experts' coefficient.

Table 3 - Parametric specification (I) – Expert and consumer reviews according the Rotten Tomatoes site – all time critical reviews

	AFT Weib. Coeff						
VARIABLES		Narrowly opening releases			Widely opening releases		
	All movies	Comedy	Foreign	Drama	Romance-Family	Adventure Animation	Action
Top experts' very good evaluation	0.616***	0.631***	1.009***	0.607***	0.624***	0.325***	0.356***
	(0.039)	(0.062)	(0.080)	(0.052)	(0.109)	(0.069)	(0.087)
Top experts' good evaluation	0.199***	0.277***	0.320***	0.212***	0.354***	0.091	0.103
	(0.040)	(0.065)	(0.067)	(0.055)	(0.134)	(0.071)	(0.085)
Consumer good evaluation	0.127***	0.172***	0.049	0.191***	0.116*	0.108*	0.214***
	(0.028)	(0.047)	(0.062)	(0.050)	(0.068)	(0.056)	(0.064)
Observations	42,273	11,124	6,199	12,027	3,748	7,137	4,021

Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1
Variables included: stars, big distributers, logarithm of number of screens at release weekend, month dummies, year dummies, genre dummies, MPAA dummies.

Specification (II), which employs ratings variables for both expert and consumer reviews, suggests no influence of top expert reviews on action and adventure/animation films that are predominantly blockbuster films.[7] While consumer reviews positively affect the screening time of these films, people go to the movies to watch the blockbusters no matter what the expert reviews say. A possible explanation for this phenomenon is that people associate this type of film with

[7] Shrum (1991) studying the effects of critical reviews on shows in Edinburgh concludes for different effects in popular and highbrow genres.

good quality and fun in advance, but the experience of other consumers may be helpful regarding a decision about going to the movies.

Table 4 - Parametric specification (II) – Expert and consumer reviews according the Rotten Tomatoes site – all time critical reviews

	AFT Weib. Coeff						
VARIABLES		Narrowly opening releases			Widely opening releases		
	All movies	Comedy	Foreign	Drama	Romance-Family	Adventure Animation	Action
Top Critics' Average (Experts)	0.906***	0.841***	1.615***	1.155***	0.564***	0.227	0.290
	(0.061)	(0.110)	(0.199)	(0.108)	(0.182)	(0.195)	(0.202)
Audience Average (Consumers)	0.116***	0.140**	0.215***	0.141**	0.129	0.210**	0.291***
	(0.031)	(0.060)	(0.080)	(0.067)	(0.091)	(0.084)	(0.086)
Observations	28,058	7,985	3,646	8,020	2,729	3,244	2,623

Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1
Variables included: stars, big distributers, logarithm of number of screens at release weekend, month dummies, year dummies, genre dummies, MPAA dummies.

Specification (III) differs from (II) just by the time a review is made, as described in Table 2. Thus, cleaning potential endogeneity of expert reviews due to the circularity between critic reviews and demand by using reviews before-during the week of release, and controlling for consumer reviews made only after film consumption, we get reduction in the size of experts' influence and increase in the size of consumers' influence on the length of screening time. It is worth noting that, after controlling for the timing of reviews, films with wide opening releases are not influenced by experts, while those with narrow openings are highly influenced by experts.

Table 5 - Parametric specification (III) – Expert critical reviews before-during and consumer critical reviews after

	AFT Weib. Coeff						
VARIABLES		Narrowly opening releases			Widely opening releases		
	All movies	Comedy	Foreign	Drama	Romance-Family	Adventure Animation	Action
Top Critics' Average (Experts)	0.544***	0.518***	0.773**	0.836***	0.282	0.026	0.100
	(0.065)	(0.112)	(0.378)	(0.120)	(0.176)	(0.184)	(0.195)
Audience Average (Consumers)	0.191***	0.272***	0.222*	0.271***	0.099	0.243***	0.302***
	(0.031)	(0.060)	(0.121)	(0.058)	(0.096)	(0.066)	(0.096)

Observations	17,067	5,011	1,441	5,185	1,505	2,081	1,719

Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1
Variables included: stars, big distributers, logarithm of number of screens at release weekend, month dummies, year dummies, genre dummies, MPAA dummies.

Adding other metrics of consumer reviews to specification (III), acquired from mining texts and performing affective analysis, we come to specification (IV) in Table 6.[8] Then, for the genres, we obtained some changes connected to the inclusion of omitted variables, i.e., predominantly an increase in the coefficients of expert reviews variables of the narrow openings, especially for foreign films that increases greatly. On the other hand, the consumer rating coefficients decrease in the narrow opening genres and increase in the wide openings, where the experts tend to have less influence (Reinstein and Snyder, 2005). Since these variables strictly measure sentiments and emotions, we can say that the length of screening of foreign, drama, and action genres are influenced by consumer affectivity, while no effects are found for the length of screening of comedy and adventure/animation films.

A possible interpretation for these results is that sentiments or emotion spread by consumers on the Internet, regardless of being contradictory, affects some genres of film consumption.[9] Note that in our case positive and negative sentimental evaluation refers to the not included neutral sentiment variable. Meanwhile, disgust, anger or sadness averages increase the screening time of a film just as joy, except for disgust in foreign films. In this sense, the results corroborate Pollai, Hoelzl and Possas' (2010) findings, as they argue people are influenced by

[8] To compare the size of potential bias of expert reviews when not controlled for our consumer reviews, check Table A2 in the annex.
[9] See Table A3 in the appendix for examples of consumer critical reviews.

emotions in consumption decisions[10]. However, we got almost no effect of affective variables in the aggregate estimates, which suggests that the effect tends to be specific with these film genres.

Table 6 - Parametric specification (IV) – Expert critical reviews before-during and consumer critical reviews after

	AFT Weib. Coeff						
VARIABLES		Narrow opening			Wide opening		
	All movies	Comedy	Foreign	Drama	Romance-Family	Adventure Animation	Action
Top Critics' Average (Experts)	0.552***	0.518***	1.068**	0.802***	0.348	0.117	0.114
	(0.070)	(0.122)	(0.433)	(0.131)	(0.238)	(0.254)	(0.227)
Audience Average (Consumers)	0.204***	0.354***	0.138	0.235***	0.212*	0.264*	0.407**
	(0.040)	(0.080)	(0.163)	(0.073)	(0.127)	(0.154)	(0.162)
Sentiment Positive	-0.048	-0.316	0.855***	0.905**	0.064	0.158	dropped
	(0.162)	(0.352)	(0.300)	(0.362)	(0.152)	(0.359)	
Sentiment Negative	-0.050	-0.254	0.731**	0.784**	Dropped	0.165	0.248**
	(0.162)	(0.346)	(0.294)	(0.378)		(0.343)	(0.105)
Anger Average	0.903*	1.807	-0.237	3.597***	4.468**	-3.266	0.865
	(0.545)	(1.301)	(1.081)	(1.265)	(1.918)	(3.903)	(2.166)
Joy Average	0.190	0.281	0.729	2.564**	1.454	-3.038	2.358**
	(0.405)	(1.135)	(0.821)	(1.135)	(1.850)	(2.503)	(0.984)
Fear Average	0.203	1.270	0.411	1.525	0.800	-2.850	-0.134
	(0.376)	(1.208)	(0.609)	(0.978)	(1.719)	(2.772)	(1.362)
Sadness Average	-0.040	-0.832	1.482	2.215*	0.560	-3.493	1.641
	(0.416)	(1.030)	(1.003)	(1.172)	(1.854)	(2.651)	(1.358)
Disgust Average	0.204	1.227	-2.317**	2.362*	2.672	0.859	-0.800
	(0.517)	(1.183)	(1.150)	(1.378)	(2.668)	(2.964)	(1.974)
Observations	17,067	5,011	1,441	5,185	1,505	2,081	1,719

Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1
Variables included: stars, big distributers, logarithm of number of screens at release weekend, month dummies, year dummies, genre dummies, MPAA dummies.

[10] Note that the same results could be interpreted simply as people caring more about the average ratings rather than the actual (written) reviews, and the genres results could be due to spurious correlations since negative sentiments and negative emotions (anger and disgust) being significant and positive signed could support the results as well. However, since for the wide opening genres, where consumption is less connected with expert influence, the consumer rating coefficients became bigger after including sentiments and emotions variables in de models, and for narrow openings we got the opposite, we are sure this is not a spurious correlation. In addition, our sentiment positive and negative are regarding the neutral sentiment omitted in the models, and since the emotions are measured by the presence of a sentiment in a consumer review, in some sense they also can be measured as compared with the absence of emotion.

5. Final Remarks

To study the effects of critical reviews by experts and consumers on the number of weeks that a film is screened in the U.S. market, under survival models, we generate a very unique weekly dataset from the Box Office Mojo and Rotten Tomatoes websites and other sources of films shown from 2004 to 2015. Our main contribution to the literature is the utilization of emotion and sentiment variables, performed by computational affective analysis, in order to generate proxy variables to measure consumer reviews\WOM of films.

The inclusion of these proxy variables in the models allows us to avoid potential bias in the expert review coefficients, since the variables offer a control for quality information. In reference to the predictor versus influence effects discussed by Eliashberg and Shugan (1997), the authors argue that expert reviews are endogenous regarding film demand because the reviewers may have just been anticipating the quality of a film (prediction effect), which is not directly observable before the film's consumption. Then, as we include other variables of film quality information (our proxies for consumer reviews and WOM), we control for this effect, and the expert reviews coefficient reflects the influence effect. In addition, as expert reviews act in the same way regarding consumer/user reviews, we control for the consumer prediction effect since expert reviews are included in the model.

The results for the overall movie market point to good reviews by top experts and consumers positively affecting a film's lifespan at the movies very robustly through model specification, after controlling for omitted variables and timing of the critics. The consumer evaluation effects, however, are a small part of the top experts' effect, ranging from 30% to 40%. Contrary to Legoux et al. (2016),[11] who find a positive and significant effect only for the experts´ reviews on survival time of films at cinemas in Canada, we see positive and statistically significant effects of consumer reviews on the duration of screening.

These results are consistent with the hypothesis that a critic review is also an experience good, where tradition and reputation matters in influencing consumption as a trustworthy signal of quality. As reputed top experts have a very reliable signal of quality, we believe that some consumers are also increasing reliability in their quality signals.

The estimates by genres point to null influence of experts on widely opening film releases, and the opposite for narrowly released ones, which is similar to the results of King (2007) and Reinstein and Snyder (2005), which state that the influence of experts tends to be lower or null in films with wide openings as compared to those with drama and narrow openings.

Also, similar to Pollai, Hoelzl and Possas (2010), we find emotions and sentiments influence film consumption in genres like foreign, drama and action. In this case, good or bad emotions tend to increase the time of film on the screens, except for disgust in foreign films.

Overall both experts and consumer/users influence film consumption, and reviews have positive and robust effects. This result is very important since, with the Internet system reviews available (consumption sites, blogs and YouTube minimizing uncertainty about good quality), consumer reviews are becoming a powerful tool to influence consumption, and, more importantly, can also affect policy conductions.

[11] Even Legoux et al. (2016) control for potential endogeneity of the expert reviews using auxiliary regressions and average consumer critic reviews, but they do not control for consumer review endogeneity.

References

Allison, P. D., (1984). *Event history analysis: regression for longitudinal event data (Quantitative Applications in the Social Sciences).* California. Sage, 87p. ISBN-13: 978-0803920552 and ISBN-10: 9780803920552

Boatwright, P., Basuroy, S., Kamakura, W. (2007). Reviewing the Reviewers: The impact of Individual Film Critics on Box Office Performance. Quantitative Marketing and Economics, 5(4), 401-425.

Basuroy, S., Chatterjee, S., Ravid, S. A. (2003). How Critical Are Critical Reviews? The Box Office Effects of Film Critics, Star Power, and Budgets. Journal of marketing, 67(4), 103-117.

Brown, A. L., Camerer, C. F., Lovallo, D. (2012). To Review or Not to Review? Limited Strategic Thinking at the Box Office. American Economic Journal: Microeconomics, 4(2), 1–26.

Cattell, R. B. (1940). Sentiment or Attitude? The Core of a Terminology Problem in Personality Research. Journal of Personality, 9(1), 3–17.

Chen, Y., Xie, J. (2008). Online Consumer Review: Word-of-Mouth as a New Element of Marketing Communication Mix. Management Science, 54(3), 477–491

Chintagunta, P. K., Gopinath, S., Venkataraman, S. (2010). The Effects of Online User Reviews on Movie Box Office Performance: Accounting for Sequential Rollout and Aggregation Across Local Markets. Marketing Science, 29(5), 944–957.

Duan, W., Gu, B., Whinston, A. B. (2008a). Do Online Reviews Matter? - An Empirical Investigation of Panel Data. Decision Support Systems, 45(4), 1007–1016.

Duan, W., Gu, B., Whinston, A. B. (2008b). The Dynamics of Online Word-of-Mouth and Product Sales-An Empirical Investigation of the Movie Industry. Journal of Retailing, 84(2), 233–242.

Eliashberg, J., Elberse, A., Leenders, M. a. a. M. (2006). The Motion Picture Industry: Critical Issues in Practice, Current Research, and New Research Directions. Marketing Science, 25(6), 638–661.

Eliashberg, J., Shugan, S. M. (1997). Film Critics: Influencers or Predictors? Journal of Marketing, 61(2), 68.

Fehr, B., Russell, J. A. (1984). Concept of Emotion Viewed From a Prototype Perspective. Journal of Experimental Psychology: General, 113(3), 464–486.

Hosmer D.W., Lemeshow S., May S. (2008). Applied Survival Analysis: Regression Modeling of Time to-Event Data, Wiley.

Gemser, G., Van Oostrum, M., Leenders, M. A. A. M. (2007). The Impact of Film Reviews on the Box Office Performance of Art House Versus Mainstream Motion Pictures. Journal of Cultural Economics.

Godes, D., Mayzlin, D. (2004). Using Online Conversations to Study Word-of-Mouth Communication. Marketing Science, 23(4), 545–560.

Godes, D., Mayzlin, D., Chen, Y., Das, S., Dellarocas, C., Pfeiffer, B., Verlegh, P. (2005). The Firm's Management of Social Interactions. Marketing Letters, 16(3–4), 415–428.

Kamakura, W. A., Basuroy, S., Boatwright, P. (2006). Is silence golden? An Inquiry Into the Meaning of Silence in Professional Product Evaluations. Quantitative Marketing and Economics, 4(2), 119–141.

King, T. (2007). Does Film Criticism Affect Box Office Earnings? Evidence from Movies Released in the U.S. in 2003. Journal of Cultural Economics, 31(3), 171-186.

Lee, C., Chan-Olmsted, S.M. (2004). Competitive advantage of broadband Internet: a comparative study between South Korea and the United States. Telecommunications Policy. Volume 28, Issues 9–10, 645-766

Legoux, R., Larocque, D., Laporte, S., Belmati, S., Boquet, T. (2016). The Effect of Critical Reviews on Exhibitors' Decisions: Do Reviews Affect the Survival of a Movie on Screen? International Journal of Research in Marketing, 33(2), 357–374.

Liu, Y. (2006). Word of Mouth for Movies: Its Dynamics and Impact on Box Office Revenue. Journal of Marketing, 70(3), 74–89.

Mayzlin, D. (2006). Promotional Chat on the Internet. Marketing Science, 25(2), 155–163.

Mayzlin, D., Dover, Y., Chevalier, J. (2014). Promotional reviews: An Empirical Investigation of Online Review Manipulation. American Economic Review.

McKenzie, J. (2009). Revealed Word-of-Mouth Demand and Adaptive Supply: Survival of Motion Pictures at the Australian Box Office. Journal of Cultural Economics, 33(4), 279–299.

Moon, S., Bergey, P.K., Iancobucci, D. Dynamic Effects Among Movie Ratings, Movie Revenues, and Viewer Satisfaction. Journal of Marketing 74, 1, 108–121. 2010.

Moretti, E. (2011). Social Learning and Peer Effects in Consumption: Evidence from Movie Sales, *Review of Economic Studies*, Oxford University Press, Vol. 78(1), 356-393.

Plutchik, R. (2001). The nature of emotions: Human Emotions Have Deep Evolutionary Roots. American Scientist, 89(4), 344–350.

Pollai, M., Hoelzl, E., Possas, F. (2010). Consumption-Related Emotions Over Time: Fit Between Prediction and Experience. Marketing Letters, 21(4), 397–411.

Reinstein, D. A., Snyder, C. M. (2005). The Influence of Expert Reviews on Consumer Demand for Experience Goods: A Case Study of Movie Critics. Journal of Industrial Economics, 53(1), 27–51.

Rui, H., Liu, Y., Whinston, A. (2013). Whose and What Chatter Matters? The Effect of Tweets on Movie Sales. Decision Support Systems, 55(4), 863–870.

Shapiro, C., Varian, H. R. (1999). Information of Rules-a Strategic Guide To the Network Economy. Harvard Business School Press (Vol. 53).

Shrum, W. (1991). Critics and Publics: Cultural Mediation in Highbrow and Popular Performing Arts. American Journal of Sociology, 97(2), 347–375.

Stigler, G. J. (1961). The Economics of Information. The Journal of Political Economy, 69(3), 213–225.

Trusov, M., Bucklin, R. E., Pauwels, K. (2009). Effects of Word-of-Mouth Versus Traditional Marketing: Findings from an Internet Social Networking Site. Journal of Marketing, 73(5), 90–102.

Vany, A. S. De, Walls, W. D. (1997). The Market for Motion Pictures: Rank, Revenue, and Survival. Economic Inquiry, 35(4), 783–797.

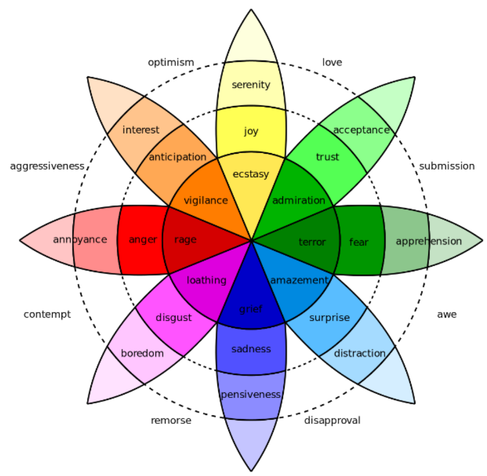
Vujić, S., Zhang, X. (2017). Does Twitter Chatter Matter? Online Reviews and Box Office Revenues. Applied Economics. DOI: 10.1080/00036846.2018.1436148.

Walls, W. D. (1998). Product Survival at the Cinema: Evidence from Hong Kong. Applied Economics Letters, 5(4), 215–219.

Wooldridge, J. M. (2002). Econometric Analysis of Cross Section and Panel Data. Booksgooglecom, 58(2), 752.

Appendix A

Figure A1: Wheel of feelings



Source: Plutchik (2001)

Figure A2: Distribution of the number of top critic reviews by film

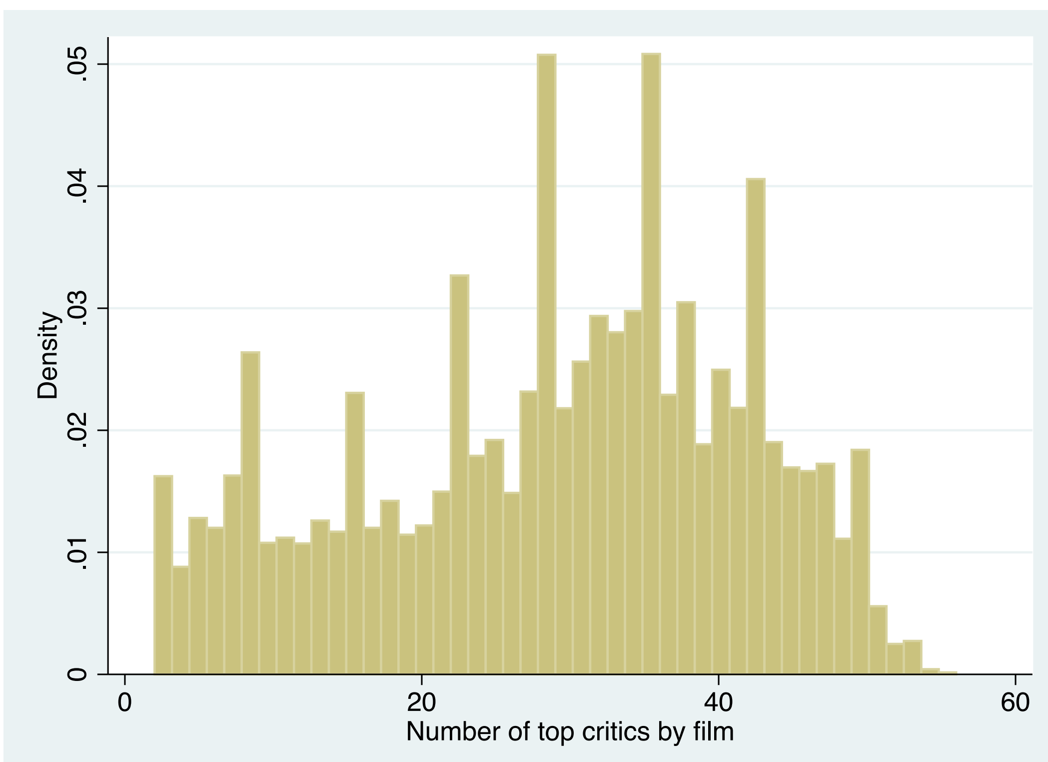
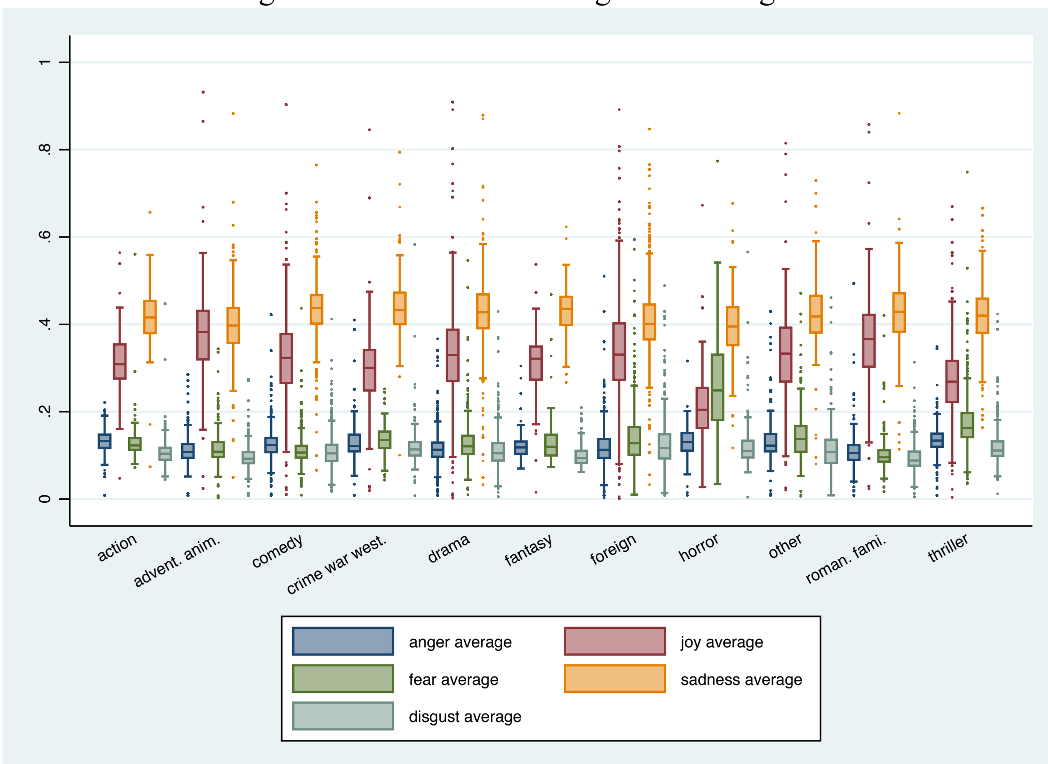


Figure A3 – Emotions average over film genre



Source: Authors using building dataset

Table A1 – Variables used on the mains estimates including controls

Variables	Obs.	Mean	Std. Dev.	Min.	Max.
Top expert reviews average - before and during period	22498	0.5456813	0.2305966	0.051	0.957
Audience Average (Consumers) - before and during period	21555	3.450435	0.5794844	0.375	5.000
Anger average - before and during period	19688	0.1178856	0.0361586	0.008	0.419
Joy average - before and during period	19688	0.3329321	0.1042836	0.007	0.856
Fear average - before and during period	19688	0.1327203	0.0604358	0.012	0.662
Sadness average - before and during period	19688	0.4141956	0.0726482	0.089	0.766
Disgust average - before and during period	19688	0.1074427	0.0410745	0.005	0.493
Sentiment positive (consumer) - before and during period	42274	0.4647064	0.4987587	0	1
Sentiment Negative (consumer) - before and during period	42274	0.1817666	0.3856566	0	1
Star dummy	42274	0.1226759	0.3280687	0	1
Big distributer dummy	42274	0.5803331	0.4935102	0	1
Logarithm number of theaters	42273	4.715597	3.167536	0	8.405
MPAA PG rating	42274	0.4892605	0.4998906	0	1
MPAA R rating	42274	0.3940247	0.488646	0	1
Action genre dummy	42274	0.0951176	0.2933807	0	1
Comedy genre dummy	42274	0.2631641	0.4403559	0	1
Drama genre dummy	42274	0.2845011	0.4511818	0	1
Romance, family genre dummy	42274	0.0886597	0.2842553	0	1
Foreign genre dummy	42274	0.1466386	0.3537495	0	1
Crime, war, western genre dummy	42274	0.0437385	0.204515	0	1
Horror genre dummy	42274	0.0560865	0.2300914	0	1
Adventure, animation genre dummy	42274	0.114231	0.3180953	0	1
Thriller genre dummy	42274	0.0859157	0.2802428	0	1

Source: Data set built by the authors from Rotten Tomatoes site, Box Office Mojo, IMF and Forbes Magazine

Table A2 – Estimates of top critics reviews with and without consumer reviews variables as controls for endogeneity biases

Regression without control for consumer reviews							
	AFT Weib. Coeff						
VARIABLES	Narrow opening				Wide opening		
	All movies	Comedy	Foreign	Drama	Romance-Family	Adventure Animation	Action
Top Critics Average (Experts)	2.155***	1.655***	2.247***	2.884***	1.415**	3.122***	1.587***
	(0.106)	(0.229)	(0.205)	(0.274)	(0.204)	(0.887)	(0.166)
Observations	22,498	2,169	6,888	6,720	2,217	1,939	2,789
Regression with control for consumer reviews - from specification (IV)							
	AFT Weib. Coeff						
VARIABLES	Narrow opening				Wide opening		

	All movies	Comedy	Foreign	Drama	Romance-Family	Adventure Animation	Action
Top Critics Average (Experts)	0.552***	0.518***	1.068**	0.802***	0.348	0.117	0.114
	(0.070)	(0.122)	(0.433)	(0.131)	(0.238)	(0.254)	(0.227)
Observations	17,067	5,011	1,441	5,185	1,505	2,081	1,719

Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1
Variables included: stars, big distributers, logarithm of number of screens at release weekend, month dummies, year dummies, genre dummies, MPAA dummies.

Table A3 - Examples of consumer critical reviews

Film: Titanic
Critics: 4 Stars
James Cameron might have seemed the incorrect choice to direct this film but he takes a step back from action films to deliver one of the better epics. The film is strengthened by Winslet, she alone carries the better half of the film, her arc is probably the strongest. Billy Zane is excellent as the jealous and conniving husband, while the ice berg might steal the show as the core villain, Billy is doing all the right things as the human villain. DiCaprio was at the beginning of his career and without this cementing him in the hearts of every person in the world, there wouldn't be the long career he has been granted. He isn't the best in the lead, he lacks that presence but to be honest, he doesn't have a lot to play with. Did the film deserve all the Oscar glory? Probably not. the filmmaking is flawless but the story and plotting are the weakest parts. It was a great film because Hollywood ignores these sweeping epics, but look at the box office, some gambles are worth taking.
Watson classification
Emotions: Joy 0.64; Anger 0.11; Disgust 0.11; Sadness 0.48; Fear 0.09.
Sentiment: Positive 0.14;

Film: Marvel's The Avengers
Critics: 5 Stars
WOW! The Avengers movie is amazing, one to be remembered. It brought characters that have had their own solo movies and brought them together with good humor, witty lines and a hell ton of action-packed scenes.
Watson classification
Emotions: Joy 0.79; Anger 0.05; Disgust 0.13; Sadness 0.04; Fear 0.01;
Sentiment: Positive: 0.95;

Film: Marvel's The Avengers
Critics: 2 Stars
Audiences and critics loved that the Avengers are all white. Thank God there are no minorities in the Avengers. All sarcasm aside, this was such a mediocre movie. A bunch of good guys get together to fight the bad guys and save the world. Predictably, the good guys don't get along at first but then they come together to defend Earth from some seriously bland and generic aliens who aren't intimidating at all. Coulson's death was supposed to be a pivotal moment in the film, but absolutely no one cared about him so that just totally fell flat. Tom Hiddleston gave a very limp-wristed performance as Loki. Apparently he reasoned that Loki sounds like "low-key" so why not play the character that way? The action sequences were all lackluster. But hey, at least the Avengers are all white, am I right?
Watson classification
Emotions: Joy 0.64; Anger 0.48; Disgust 0.17 Sadness 0.45; Fear 0.10;
Sentiment: Negative: -0.41

Film: Carrie
Critics: 1/2 Star
I'm severly disappointed with this remake of the 1976 classic. There scenes in it that made the film drag on and I wasn't too pleased with the casting or special effects, as soon as it hits the prom scenes, feels like watching final destination. A miga 2/10 from me.
Watson classification
Emotions: Joy 0.22; Anger 0.03; Disgust 0.22; Sadness 0.52; Fear 0.21;
Sentiment: Negative: -0.72;

Film: Sinister
Critics: 4 Stars
It can be difficult to make something visually scary these days, when everything has been done. With that said, Sinister does a great job on that front. I found its straightforward plot a terrifying one.
Watson classification
Emotions: Joy 0.13; Anger 0.04; Disgust 0.01; Sadness 0.06; Fear 0.86;
Sentiment: Positive: 0.33

Film: Sinister
Critics: 1 Star
bloody found footage crap. stop making these please.
Watson classification
Emotions: Joy 0.01; Anger 0.59; Disgust 0.38; Sadness 0.22; Fear 0.19;
Sentiment: Negative: -0.39