Journal of Cultural Economics (2020) 44:57–96
https://doi.org/10.1007/s10824-019-09350-7

ORIGINAL ARTICLE

Is everybody an expert? An investigation into the impact of professional versus user reviews on movie revenues

Suman Basuroy[1] · S. Abraham Ravid[2,3] · Richard T. Gretz[1] · B. J. Allen[4]

Received: 20 October 2017 / Accepted: 15 April 2019 / Published online: 6 May 2019

Abstract

This study is the first attempt to examine the effect of electronic word of mouth (user reviews) relative to expert reviews on moviegoing decisions. For the first time, we use time-varying data on expert reviews. We find that expert ratings matter much more for moviegoing decisions than user ratings and volume. Our data also show that experts tend to be more critical but more consistent in their reviews than users. We find that experts, but not eWOM, affect wide release moviegoing, contrary to industry thinking. Finally, we show that experts' reviews matter most when consumers and critics are in closer agreement about the quality of the film. The study uses OLS as well as instrumental variables analysis to account for possible endogeneity.

Keywords eWOM volume · eWOM valence · Expert reviews · Movies

✉ Suman Basuroy
suman.basuroy@utsa.edu

S. Abraham Ravid
ravid@yu.edu

Richard T. Gretz
richard.gretz@utsa.edu

B. J. Allen
ballen@walton.uark.edu

[1] Department of Marketing, College of Business, The University of Texas at San Antonio, One UTSA Circle, San Antonio, TX 78249-0631, USA

[2] Sy Syms School of Business, Yeshiva University, Belfer Hall, 500 W. 185 St., New York, NY 10033, USA

[3] Knut Wicksell Center, Lund University, Lund, Sweden

[4] Sam M. Walton College of Business, University of Arkansas, 220 N Mcllroy Ave, Fayetteville, AR 72701, USA

1 Introduction

Electronic word of mouth (eWOM) or Internet-mediated written communications between consumers (user reviews) seems to be playing an increasing role in the consumer decision-making process (Babic et al. 2016; Baker et al. 2015; You et al. 2015; Marchand et al. 2017; Mayzlin et al. 2014). Companies are allocating larger portions of their marketing budgets to generate and manage the eWOM process, and ZenithOptimedia estimates that such spending had reached $8.22 billion by 2015 indicating a growth rate of about 35% annually (Sass 2013).

Most of the literature to date analyzes either user reviews or expert reviews. This is the first paper to compare these two forms of evaluation in the context of moviegoing. We are also the first to follow both types of reviews over time.

The majority of eWOM research focuses on two key metrics, volume and valence as measures of the extent of eWOM. However, there are conflicting findings regarding the impact of these two eWOM metrics on consumer decisions.[1] Such conflicting findings prompted You et al. (2015) to perform a meta-analysis of 51 eWOM studies (13 of these articles cover the movie industry). Their meta-analysis contains 15 platforms where information regarding products and services is posted and exchanged by consumers such as Amazon, Yahoo! Movies, Yahoo! Games, CNET, GameSpot, and others. You et al. (2015) find that (1) both volume and valence have a positive and significant impact on market outcomes, (2) valence elasticity is generally higher (.417) than volume elasticity (.236), and that (3) impact varies depending on three contextual factors—product, industry, and platform characteristics. Importantly, to our knowledge, we are the first to include time-varying professional critics and eWOM in the same study.

A key characteristic of many internet platforms is that along with consumer opinions and reviews they simultaneously (on the same page) carry or display experts' opinions and reviews for the same product or service. For example, www.GameSpot.com carries reviews and opinions from consumers and experts on video games, www.CNET.com carries reviews and opinions on consumer electronic products from both consumers and experts, while www.rottentomatoes.com carries both consumer and expert reviews for movies.

In this study, we analyze user and expert reviews in the context of motion pictures. We have several interesting findings. First, we show that if we exclude expert reviews, we are able to replicate prior research and find that eWOM volume and valence significantly affect sales. However, if we include expert opinions, we find that while eWOM volume and valence are significant, expert opinions have a greater

[1] Conflicting findings occur even in the context of the same industry (e.g. movies): Liu (2006) and Duan et al. (2008) find that the volume but not the valence of consumer reviews is significantly associated with movie revenues. Chintagunta et al. (2010) find that it is valence of eWOM, rather than volume, that drives revenues. Also, the elasticities calculated vary a great deal. A few studies find negative elasticities, in particular in some movie and book studies, and You et al. (2015) suggest that "poor ratings can result in sales especially because the marginal cost of these products is so low" (ibid. p. 34). We are not sure how to reconcile these findings—our results seem to suggest a robust positive elasticity for volume and valence even in the presence of expert reviews.

impact on sales than either measure of eWOM. In doing so, we instrument for both eWOM measures and expert reviews.

The findings are qualitatively similar for OLS and IV regressions. We also show that our findings are not due to the aggregation bias discussed in Chintagunta et al. (2010). Finally, we find that user reviews do become relatively more important when experts and users disagree most and that, counter-intuitively, expert reviews matter most for wide release as opposed to platform release ("art house") movies. We also show that user reviews tend to be more positive than critical reviews.

The remainder of the article is organized as follows. The next section provides theory and background. Section 3 discusses the data and estimation methodology. Section 4 reviews estimation results. Section 5 concludes with a discussion of the academic and managerial implications of our findings.

2 Literature, theory and background

Product uncertainty can arise from consumers' lack of information about available product alternatives or from doubts as to whether or not the product matches his or her needs. Such uncertainty determines the search for information a consumer undertakes (Maity et al. 2014; Moorthy et al. 1997; Urbany et al. 1989). Traditional papers on consumer information search show that there are costs (e.g., monetary costs, psychological costs, etc.) and benefits (e.g., reduced product uncertainty, greater fit to user needs, etc.) associated with such a search (You et al. 2015). In the presence of various eWOM, company websites, and information aggregators (e.g., www.GameSpot.com), the costs of consumer search decline. Moreover, the numerous internet platforms make it easier to capture a range of diverse consumer and expert perspectives that better help consumers in judging the fit of the product with their own needs and preferences (You et al. 2015). As a result consumers are switching from offline to online information search (Klein and Ford 2003; Ratchford et al. 2007; Marchand et al. 2017).

A study conducted by KRC Research (2012)—who surveyed customers in the consumer electronics industry (smartphones, tablets, cameras, etc.)—shows that consumers routinely search for both consumer and expert reviews online before purchase: "65% of the subjects were influenced by a favorable consumer review to buy a consumer electronic product while 59% were influenced by professional critic review…" (p. 2). Subjects reported reading 11 reviews on average. Thus, it seems that today's consumers routinely search for both expert and user reviews to make informed decisions.

Holbrook (1999) finds that consumers and experts emphasize different criteria in their reviews (of movies) and discusses potential sources of such differences. He argues that professional critics who offer expert judgments are different from lay consumers because: (1) professional critics display familiarity and consistency with the values of the "cognoscenti", conventionally empowered to determine what passes for excellence; (2) professional critics receive extensive education and training, and (3) professional critics often serve in an institutionalized capacity sanctioned by official appointments or by the support of various scholars, authorities and editors. In contrast, consumer reviews stem from ordinary consumers who may or may not have extensive familiarity with the product, do not necessarily possess training and education in the area, usually do not serve in an institutional capacity and rarely have any official appointments. Given the significant differences between experts and users, we expect the two information sources to impact consumer decision processes differently.

A key issue that plagues eWOM is their authenticity. In the aforementioned KRC study, a full 80% of consumers polled reported concerns about the authenticity of consumer reviews online. A report from www.pbs.org recently reported that the restaurant review website, www.yelp.com, labels 25% of submitted reviews as suspicious or not recommended (PBS Newshour 2015). Mayzlin et al. (2014) document seemingly fake customer reviews of hotels on www.tripadvisor.com. The ubiquity of fake reviews has recently prompted the Federal Trade Commission in the USA to update its guidelines governing endorsements and testimonials to also include guidelines for online reviews.

Therefore, our hypothesis is simple:

H1 Expert reviews are more influential than user reviews in determining consumer behavior and purchases.

This follows the literature above suggesting that consumer reviews may be less credible and less informative.

The alternative hypotheses are that user reviews are more important and that both types of reviews have equal effects on consumers' behavior.

3 Data and methodology

There are two main objectives to our empirical study. First, we compare the relative influence of expert and non-expert ratings on weekly movie box office receipts. Second, we examine the conditions that affect the relative influence of expert and non-expert ratings on weekly box office changes.

We identify a random sample of 194 films that had theatrical release in the US market in 2007 and early 2008.[2] The data include 171 movies from 2007 and 23 movies released in January and February of 2008. In 2007 there were only 189 movies released in theaters domestically by the members of MPAA.[3] Thus, our sample is fairly representative and comparable in size to those used in other

[2] In our dataset, there are 194 movies, each followed for 10 weeks. Thus there are 1940 potential observations—movie-week data points. However, there are 1629 observations in the OLS regressions due to some movies leaving the market before the full 10 weeks as well as missing observations for other variables.

[3] Source- MPAA.org. We should note that the theatrical market has not changed much in the last 10 years in terms of admissions and real revenues (see Theatrical Market Statistics at MPAA.org) except the change to digital projection. The other change in distribution towards streaming does not affect the impact of user and professional critics on weekly revenues which we are studying here.

studies (Ravid 1999; Basuroy et al. 2003; You et al. 2015). It is larger than various datasets in several eWOM studies. Liu (2006), for example, uses a selected sample of 40 movies from the summer months of 2002, Chintagunta et al. (2010) use a sample of 148 movies released from November 2003 to February 2005, and Gopinath et al. (2013) use a sample of 75 movies released in 2004. Baseline provides information on weekly domestic box office revenues (*Box*), theater counts (*Screen*), and other revenues sources.

There are various sites which aggregate professional critical opinions and internet user reviews. We looked for a site which includes both types of reviews side-by-side, because consumers are more likely to view both user and expert reviews when they visit the site seeking information, and thus our tests can be cleaner. We use a single website, www.rottentomatoes.com (hereafter RT, see, Moon et al. 2010) that is ranked 3rd among top websites by category (movies) in the USA by the website ranking company, Alexa (http://www.alexa.com/topsites/category/Top/Arts/Movies), and which displays professional critics' ratings as well as user ratings next to each other on the same webpage. It is also commonly followed by industry insiders. We display a portion of the RT website for the movie *Shrek the Third* in Fig. 1 as an example.

Each user review in this website is dated. We collect the number of users who had reviewed the movie to date for each week (*UserVolume*). This variable is similar to the volume of reviews used by Liu (2006) or Chintagunta et al. (2010). We also collect the valence of user ratings, which is the average rating of users who have reviewed the movie through each week (*UserRating*). Chevalier and Mayzlin (2006) and Chintagunta et al. (2010) underscore the importance of the valence of user reviews. Also, recent studies control for variance of eWOM (e.g. Chintagunta et al. (2010) and Zhang and Dellarocas (2006)) in their estimations. We follow this trend and calculate the variance of average user ratings (*UserVariance*) for movie i during the tth week on the market as

$$UserVariance_{it} = \frac{\sum_{\tau=1}^{t} weekUserVolume_{i\tau}\left(weekUserRating_{i\tau} - UserRating_{it}\right)^2}{UserVolume_{it}}$$

where $weekUserVolume_{it}$ and $weekUserRating_{it}$ are the number of user reviews and average user rating for movie i during the tth week, respectively. This is a measure of variance using the average ratings within each week (weighted by the respective number of reviews within the week) relative to the average of all user ratings up to and including week t.

We also collect the valence of the critic rating, which is the average critic rating up to the week in question (*CriticRating*) provided by RT. It is important to note that the RT website aggregates and averages critical reviews daily. Therefore, the average critic rating for a film changes as soon as a new critical review is available and we can see how average critical evaluation changes over the run of a movie. Our paper is the first to offer a dynamic measure of critics' ratings (as opposed to the average rating prior to the opening of a movie).

Also, RT includes reviews only once even if reviews are reprinted in local papers. For example, *New York Times* reviews are often reprinted in local papers alongside reviews by local critics—RT will only include the *New York Times* review when it is originally published. Thus, our average critic rating measure avoids any bias due to counting a single review multiple times. The only other paper that we are aware of that leverages the different publication times of critical reviews in the movie industry is Chen et al. (2012). Chen et al. (2012) conduct an event study using stocks of publicly traded movie studios and show that prerelease expert reviews exert an impact on company stocks in the direction implied by their valence.

Recent research also controls for the impact of variation in critical opinion (Kupfer et al. 2018) and the number of critical reviews (Zhang and Dellarocas 2006) on box office success. With this in mind, and to ensure our measures of critical opinion parallel our measures of user opinion. We collect the number of critics who have reviewed the movie to date for each week (*CriticVolume*) and calculate the variance of average critic ratings (*CriticVariance*) with

$$CriticVariance_{it} = \frac{\sum_{\tau=1}^{t} weekCriticVolume_{i\tau}\left(weekCriticRating_{i\tau} - CriticRating_{it}\right)^2}{CriticVolume_{it}}$$

where $weekCriticVolume_{it}$ and $weekCriticRating_{it}$ are defined similar to the analogous measures for users.

We control for various other potential determinants of weekly movie revenues. Our advertising data (*Advs*) cover total television and print advertising expenditures for each film in each week as collected by Kantar Media (www.kantarmedia.com). The data are weekly—a common unit of analysis for the motion picture industry (see, Ho et al. 2009).

There are several other variables that have been shown to affect movie revenues, including genre variables—*action*, *comedy*, *drama*, *romance* and *thriller*—and MPAA ratings (De Vany and Walls 2002; Ravid 1999). Also, the impact of star power on movie revenues has been debated in the literature. In general, star participation, however defined, does not seem to affect revenues per se (see Ravid 1999; De Vany and Walls 1999; Elberse 2007). However, star power may have a competitive and "insurance" role (see Basuroy et al. 2003). Other determinants of film revenues can be the identity of the studio (see Elberse and Eliashberg 2003) and seasonal effects (see Radas and Shugan 1998). As these film-specific characteristics are constant throughout a film's run, we control for these effects by first differencing our variables in the estimations below.

As noted, part of our analysis concerns the release strategy of the film in question. Studios follow two distinct strategies. One strategy is wide release. The movie opens in as many theaters as possible around the country. Typically, the number of theaters and revenues decline steeply as time goes by. The alternative strategy is called "platform release" or limited release. In that case, the movie opens in a small number of theaters, and if it does well, theaters are added as time goes by. Industry believes that platform release is better for iffy, more "artistic" movies which may not be suitable for a wide audience. Einav (2007)

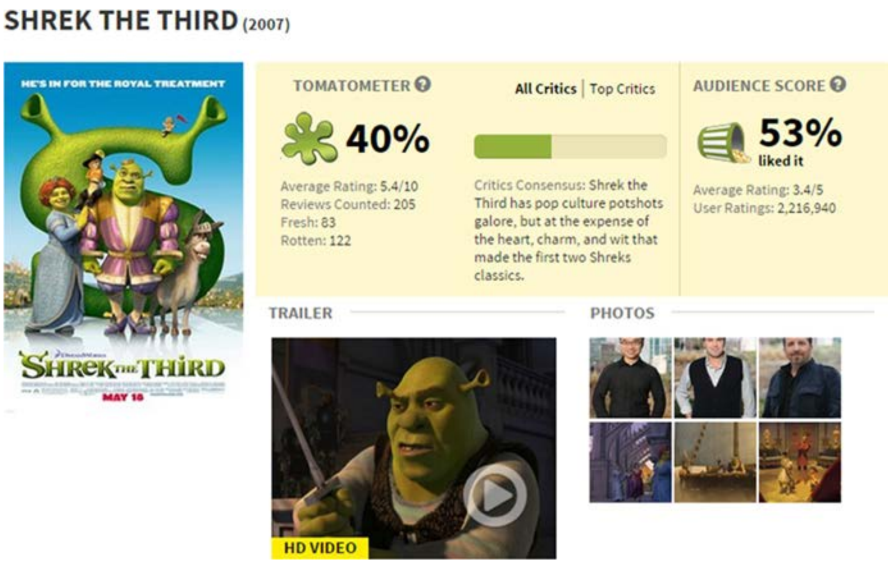

Fig. 1 Portion of *Shrek the Third* homepage on www.rottentomatoes.com. Summary of critic and user ratings displayed side-by-side

categorizes films opening on less than 600 screens as "platform release" movies. We follow that definition. $PlatformRelease_i$ is a dummy variable that takes a value of 1 if movie i is released on <600 screens, and 0 otherwise (Table 1).

Table 2 contains all the key variables, their definitions, references, where these variables have been previously used as well as the sources for our data. Table 3 provides descriptive statistics. Table 4 is the correlation matrix. Table 3 shows a significant variability in both user and expert ratings of our movies, but user ratings are on average higher than professional critics' ratings (7.08 vs. 5.82). This is interesting in itself, and may be because most users review movies they choose to see, and thus by definition like some features of the movie even before setting foot in the theater, whereas reviewers review all movies. User volume varies a great deal per week; movie revenues range from a high or over \$160 million per week to a low of \$220,000 per week. Table 4 shows that box office revenues are positively correlated with expert ratings, screens and advertising and negatively correlated with a movie's age (i.e. the number of weeks since the movie was released). We are going to explore these relationships further in our analysis.

Because of varying lengths of theatrical runs of movies in our sample, we follow the work of Basuroy et al. (2003) and Liu (2006) and restrict the empirical analyses to the first 10 weeks. The first 10 weeks typically account for more than 90% of the box office revenues. In the next subsection, we describe the model, identifying the effect of critic valence, and the instrumental variables we use to correct for possible endogeneity.

3.1 Model specification

We use the following key revenue Eq. (1):

$$\begin{aligned} Log\left(Box_{it}\right) = Const' &+ \beta_1 UserVolume_{it-1} + \beta_2 UserRating_{it-1} + \beta_3 UserVariance_{it-1} \\ &+ \beta_4 CriticVolume_{it-1} + \beta_5 CriticRating_{it-1} + \beta_6 CriticVariance_{it-1} \\ &+ \beta_7 Screen_{it} + \beta_8 Log\left(1 + Advs_{it-1}\right) + \beta_9 Age_{it} + \beta_{10} Age^2_{it} + \mu_i + \varepsilon_{it}. \end{aligned} \tag{1}$$

Note that we take the natural log of Box_{it} and $1 + Advs_{it-1}$ because the levels of these variables are highly skewed; we add 1 to $Advs_{it-1}$ before taking the natural log in order to ensure the natural log is defined when $Advs_{it-1} = 0$. Since $Log\left(Box_{it}\right)$ is the dependent variable in Eq. (1), β_8 is interpreted as the approximate elasticity of box office with respect to advertising dollars in the previous period; the βs on all the other variables in levels are interpreted as the percentage change in box office caused by a 1 unit change in the respective variable.

A key difference from the previous literature is the inclusion of expert ratings in the estimation equation along with user ratings. μ_i is a vector which captures any time invariant film-specific effects on box office revenues such as genre, rating, studio, star power, any unobserved time of release effects, etc. In addition, μ_i captures time invariant film-specific unobservables—failure to control for these may bias estimates as well if unobservables are correlated with the independent variables (Greene 2011a). Previous work has either relied on fixed effects estimation which makes it difficult to separately identify the effect of time invariant expert ratings and μ_i (e.g. Chintagunta et al. 2010) or has identified μ_i without controlling for time invariant unobservables directly in the estimation via fixed effects (e.g. Eliashberg and Shugan 1997; Ravid 1999; Basuroy et al. 2003; Gopinath et al. 2013[4]).

Our main independent variables of interest are (see Table 2): *UserVolume* = Number of user reviews for film *i* by (up to and including) week *t*; *UserRating* = Average user rating for film *i* by week *t* (i.e., the valence of users' comments); *CriticRating* = Average critic rating for film *i* by week *t* (i.e., the valence of critics' opinions). In addition, we include a number of control variables: *UserVariance* = Variance of average user ratings of film *i* by week *t*; *CriticVolume* = Number of critic reviews for film *i* by week *t*; *CriticVariance* = Variance of average critic ratings of film *i* by week *t*; *Screen* = Number of screens (in 100 s) per week for film *i* in week *t*; *Advs* = Total advertising dollars for film *i* by week *t*; and *Age* = Age of film *i* in week *t* (i.e., weeks since film *i*'s release) as well as age squared (Age^2) to control for a linear time trend and possible higher order trends, respectively. We lag *UserVolume*,

[4] We note that Gopinath et al. (2013) identify μ_i using a two-stage approach which controls for time invariant unobservables but their identification assumption is based on the standard random effects approach. First they estimate box office in different markets at different points in time incorporating movie fixed effects. Second, they regress the movie fixed effects coefficients from the first stage on time invariant regressors to recover μ_i. However, this approach to identifying time invariant regressors relies on the assumption that they are not correlated with any unobserved movie specific effects (Greene 2011a) which reduces to the standard assumptions of a random effects model (Greene 2011a, b).

Table 1 Sample of review websites that utilize both expert reviews and eWOM

Review website	Product type	Expert reviews included?	User reviews included?
All music	Music	Yes	Yes
Barnes and noble	Books	Yes	Yes
Cars.com	Automobiles	Yes	Yes
CNET	Electronics	Yes	Yes
Consequence of sound	Music	Yes	Yes
Digital trends	Electronics	Yes	Yes
Edmunds	Automobiles	Yes	Yes
Engadget	Electronics	Yes	Yes
Gamespot	Video games	Yes	Yes
IGN	Video games	Yes	Yes
IMDb	Movies	Yes	Yes
Kelly blue book	Automobiles	Yes	Yes
Metacritic	Various	Yes	Yes
Rotten tomatoes	Movies	Yes	Yes
Trusted reviews	Various	Yes	Yes

UserRating, *UserVariance*, *CriticVolume*, *CriticRating*, *CriticVariance* and *Advs* in Eq. (1) to limit the effect of contemporaneous correlation as reverse causality may bias estimates when using current values. For example, it is likely that current box office revenues influence the current volume of user reviews since moviegoers are likely to review films the same week they watch them. Greater revenues in a week mean a large number of moviegoers that week and, as a result, a larger number of potential user reviews. Lagging these variables limits this influence on coefficient estimates. Lastly, e_{it} is an error term.

As discussed, by first differencing Eq. (1), we abstract from the need to estimate μ_i. This also eliminates any bias in the estimates due to correlation of the explanatory variables with unobservable film characteristics. We can achieve the same goal by incorporating fixed film effects; however, first differencing offers some advantages for our instrumental variable strategy (Greene 2011a)—we discuss this further in the instrumental variables section below.

Equation (2) is the first difference equation:

$$\begin{aligned}\Delta Log(Box_{it}) = Const &+ \beta_1\left(\Delta UserVolume_{it-1}\right) + \beta_2\left(\Delta UserRating_{it-1}\right)\\ &+ \beta_3\left(\Delta UserVariance_{it-1}\right) + \beta_4\left(\Delta CriticVolume_{it-1}\right)\\ &+ \beta_5\left(\Delta CriticRating_{it-1}\right) + \beta_6\left(\Delta CriticVariance_{it-1}\right)\\ &+ \beta_7\left(\Delta Screen_{it}\right) + \beta_8\left(\Delta Log\left(1 + Advs_{it-1}\right)\right) + \beta_{10}\left(\Delta Age^2_{it}\right) + \Delta\varepsilon_{it}.\end{aligned} \tag{2}$$

Table 2 List of variables used in the analysis, their definitions, references, and sources

Variables	Variable definition	Literature support	Data source
Box_{it}	Main dependent variable. Box office revenue of film i in week t	Basuroy et al. (2003), Liu (2006)	Baseline
$UserVolume_{it}$	Endogenous variable. Number of user reviews for film i by week t	Liu (2006), Zhang and Dellarocas (2006), Chintagunta et al. (2010)	www.rottentomatoes.com
$UserRating_{it}$	Endogenous variable. Average user rating of film i by week t	Liu (2006), Zhang and Dellarocas (2006), Chintagunta et al. (2010)	www.rottentomatoes.com
$UserVariance_{it}$	Endogenous variable. Variance of average user ratings of film i by week t	Zhang and Dellarocas (2006), Chintagunta et al. (2010)	www.rottentomatoes.com
$CriticVolume_{it}$	Number of critic reviews for film i by week t	Zhang and Dellarocas (2006)	www.rottentomatoes.com
$CriticRating_{it}$	Endogenous variable. Average critic rating of film i by week t	Basuroy et al. (2003), Liu (2006), Zhang and Dellarocas (2006), Gopinath et al. (2013)	www.rottentomatoes.com
$CriticVariance_{it}$	Endogenous variable. Variance of average critic ratings of film i by week t	Kupfer et al. (2018)	www.rottentomatoes.com
$Screen_{it}$	Endogenous variable. The number of screens (in 100 s) for film i in week t	Basuroy et al. (2003), Liu (2006), Chintagunta et al. (2010), Gopinath et al. (2013)	www.rottentomatoes.com
$Advs_{it}$	Endogenous variable. Total advertising dollars for film i in week t	Chintagunta et al. (2010)	Kantar Media
$PlatformRelease_i$	Dummy variable = 1 if film i is a platform release, 0 otherwise	Chintagunta et al. (2010), Einay (2007), Ho et al. (2009)	Baseline
Main instrumental variables			
$UserVolume_{it-1}$	Number of user reviews for film i by week $t-1$	Greene (2011a)	www.rottentomatoes.com
$PrevUserVolumeDiffGen_{it-1}$	Average number of user reviews in week $t-1$ after release for all movies in a different genre than film i released before film i	Lee (2013)	www.rottentomatoes.com
$UserRating_{it-1}$	Average user rating of film i by week $t-1$	Greene (2011a)	www.rottentomatoes.com
$AvgPrevRev_{it-1}$	The average number of reviews critics who have reviewed film i by time $t-1$ have done prior to reviewing film i		www.rottentomatoes.com

Table 2 (continued)

Variables	Variable definition	Literature support	Data source
$CompScreen_{it}$	The average number of screens that show movies of the same genre as film i in week t	Neelamegham and Chintagunta (1999), Chintagunta et al. (2010)	www.boxofficemojo.com
$Advs_{it-1}$	Total advertising dollars for film i in week $t-1$	Greene (2011a)	Kantar Media

Table 3 Descriptive statistics

Variables	Mean	SD	Min	Max
Box_{it} (in \$1000s)	4613.1020	12016.5700	0.2240	160099.0000
$UserVolume_{it}$	75.5402	53.6253	0.0000	252.0000
$UserRating_{it}$	7.0829	0.9152	0.0000	10.0000
$UserVariance_{it}$	0.4795	0.5554	0.0000	6.8056
$CriticVolume_{it}$	3.2462	7.9658	0.0000	42.0000
$CriticRating_{it}$	5.8194	1.5243	0.0000	9.4792
$CriticVariance_{it}$	0.4561	0.5828	0.0000	3.8850
$Screen_{it}$(in 100s)	9.2313	11.3215	0.0100	43.6200
$Advs_{it}$ (in \$1000s)	466.8621	955.0027	0.0000	8347.1870
$Age_{it}(Week)$	5.1031	2.8124	1.0000	10.0000
$PlatformRelease_i$	0.3843	0.4866	0.0000	1.0000

$N = 1629$

Δ indicates the change in the variable from the previous period. Note our coefficients of interest, the βs, are not changed by first differencing.[5]

3.2 Identification of critics' valence

In papers which included critical reviews (Ravid 1999; Basuroy et al. 2003; Eliashberg and Shugan 1997; Gopinath et al. 2013), the valence of critical reviews is constant for each film. This makes it impossible to identify critic valence separately from time invariant film-specific effects (μ_i in Eq. (1)). The researcher is left with two options: (1) either include fixed effects without separately identifying the impact of critics on revenue (e.g. Chintagunta et al. 2010) or (2) include observable time invariant film-specific variables such as rating, genre, studio, star power, critical reviews, and impose restrictions that these variables are not correlated with the film-specific unobservables (Gopinath et al. 2013). The former strategy controls for all time invariant film-specific observables and unobservables, but does not allow for separate identification of time invariant film-specific effects. The latter strategy allows for identification of observable time invariant film-specific effects, but obtaining consistent estimates relies on the tenuous assumption that regressors are not correlated with unobservable film-specific effects (Greene 2011a).

In this paper, as discussed, we use for the first time the precise date when a critical review is released. Critic valence changes over the run of each film because of two reasons: (1) different outlets have different publication schedules (e.g., the *New York Times* is a daily publication, *Time* is weekly, *Rolling Stone* is bi-weekly) and (2) movies are released later in some local markets, and hence, some local reviews come out later. We exploit RT's convention of listing the publication date of every

[5] The effect of age is not separately identified from the constant in Eq. (2) when differenced.

Table 4 Correlation table of continuous variables

		1	2	3	4	5	6	7	8	9	10
1	Box_{it}	1.000									
2	$UserVolume_{it}$	−0.110	1.000								
3	$UserRating_{it}$	−0.038	0.016	1.000							
4	$UserVariance_{it}$	−0.074	−0.384	−0.154	1.000						
5	$CriticVolume_{it}$	0.396	−0.187	−0.029	−0.127	1.000					
6	$CriticRating_{it}$	0.017	−0.279	0.454	−0.003	0.015	1.000				
7	$CriticVariance_{it}$	−0.135	−0.170	0.129	0.191	−0.093	0.180	1.000			
8	$Screen_{it}$	0.670	0.024	−0.060	−0.184	0.253	−0.094	−0.228	1.000		
9	$Advs_{it}$	0.657	−0.172	0.083	−0.131	0.516	0.149	−0.131	0.504	1.000	
10	Age_{it}	−0.361	0.246	0.084	0.216	−0.521	0.056	0.189	−0.443	−0.462	1.000

critical review, and updating their average critic rating score accordingly, to obtain a time-varying measure of critic valence over the run of each film.

Although many reviews come out on the release date or slightly before that, every movie in our data set has critical reviews that come out after the movie is released. In total, there are 859 reviews in our sample which were posted between the second week and the tenth week the movie is on the market, and new reviews appear every week (though there is a greater percentage of review in early weeks). This provides us with enough data to identify the effect of time-varying critic valence. We provide a summary of the distribution of the number of critical reviews over time as well as the average critic and user rating through the opening week and after the opening week in Table 5. Note that, as discussed, users tend to rate movies higher than critics and that average critical and user valence is higher for reviews released after the opening week than reviews released during or before the opening week.

Table 6 shows how the average user valence and critic valence change over the weeks of a film's release. Note that both the change in average critic and user valence tend to decrease in absolute value over the run of films' release. However, there are still relatively large shifts in both even at the end of the 10-week period—in absolute terms, the largest change in critic and user valence for a film is .2206 and .6857, respectively (see Table 6). The fact that these variables change over time allows us to identify the effects of both critical and user valance on weekly box office revenue while including fixed effects to control for time invariant film-specific observables and unobservables. Interestingly, user reviews tend to show much greater changes over time than expert's ratings, which seems to suggest that there is some element of professional evaluation which is common to the professionals and may be lacking in user reviews.

Since several variables, including screens, user volume user rating and critic rating can be endogenous, we make extensive use of instrumental variables in our analysis. A very detailed "Appendix 1" describes the issues, the instrumental variables and the estimation involved.

4 Results

The availability of instruments enables us to address endogeneity and possible correlations between the error terms and variables we use in our analysis, including user volume, user valence, variance of user valence, critical valence, variance of critical valence, number of screens, weekly advertising expenditure. The generalized method of moments (GMM) allows us to explicitly deal with these endogeneity concerns. GMM is also preferred in the presence of heteroskedasticity (Cameron and Trivedi 2005; Greene 2011a). We use the continuously updating GMM estimator which has been shown to have better performance in finite samples (Imbens 2002). Importantly, the Hansen's J-statistic is never significant in any GMM estimation indicating the instruments appear to be orthogonal to the error terms. We use clustered robust standard errors control for any arbitrary correlation of errors for observations within the same movie; failure to control for this can result in severe downward bias of standard errors (Cameron and Trivedi 2005).

Table 7 shows models including eWOM volume and valence with and without expert reviews. The unit of observation is movie-week. Models 1 and 2 allow us to compare our results to previous work on eWOM in general (You et al. 2015) and in the movie industry specifically (Chintagunta et al. 2010; Gopinath et al. 2013; Liu 2006) where the explanatory variables of interest are user volume and/or user valence (the number of screens and advertising dollars are included in all the estimations as controls). We should mention that in every model in Table 7 advertising investment has a positive and significant effect on revenues, the number of screens has a positive and significant effect as well (see Elberse and Eliashberg 2003), and there is a significant negative time trend (i.e. the negative and significant effect of the squared term for movie age always dominates) as found in previous research (Chintagunta et al. 2010). Also, we show variance inflation factors (VIFs) to accompany each model in Table 7. The greatest mean VIF is 1.83 (for Models 3 & 4) and no individual VIF is above 5. Therefore, we believe multicollinearity is not unduly influencing the results.

Model 1 shows OLS estimation without any endogeneity corrections. Using this specification, we find that both user rating and user volume have a positive and significant impact on movie revenues. These results are consistent with those reported in the recent Meta-Analyses of the eWOM literature by You et al. (2015) and Babic et al. (2016). Our results hold when we control for endogeneity with our GMM estimation in Model 2.[6] Using estimates from Model 2, we find the marginal effect of user valence is significantly larger than the marginal effect of user volume ($\chi^2(1) = 132.07, p < 0.0000$) consistent with Chintagunta et al. (2010) and Gopinath et al. (2013).[7]

In Models 3 and 4 we include critic valence (ratings) and variance of critic valence, user valence and variance of user valence, user volume, critic volume, number of screens, and weekly advertising expenditure. Model 3 is a simple OLS while Model 4 addresses endogeneity concerns with GMM estimation.[8] Models 3 and 4 show that the effects of user volume are still significant, but the coefficient is significantly lower. User ratings have much lower coefficients and in model 3 user ratings are insignificant.

More importantly, the impact of critic valence on revenues is significantly larger than the corresponding effect of user valence in both Models 3 and 4.

[6] We reject the null hypothesis that our endogenous regressors are exogenous with $\chi^2(5) = 71.623$ and associated p value < 0.0000.

[7] We should note that Liu (2006) and Babic et al. (2016) find that box office is more sensitive to user volume than valence. However, one important difference between our approach and Liu's (2006) is that Liu (2006) uses measures only from the previous week and does not consider cumulative variables. While we do not display here for brevity, our results in Models 1 and 2 are similar to Liu's (2006) when we use the same weekly variables. The approach we take in the paper is similar to the research of Chintagunta et al. (2010) and Gopinath et al. (2013) in that we use cumulative measures for our variables (e.g. total user volume up to week $t-1$ rather than user volume only in week $t-1$).

[8] We reject the null hypothesis that our endogenous regressors are exogenous with $\chi^2(7) = 50.274$ and associated p value < 0.000.

Table 5 Distribution of the number of critical movie reviews over time in the dataset

	Total	Total (%)	Average critic rating for reviews released during respective time period	Average user rating for reviews released during respective time period
Before opening week	2366	39.63	5.8193	6.9889
Opening week	2745	45.98		
2nd week	392	6.57	5.9126	7.0744
3rd week	213	3.57		
4th week	117	1.96		
5th week	46	0.77		
6th week	27	0.45		
7th week	16	0.27		
8th week	11	0.18		
9th week	10	0.17		
10th week	27	0.45		

Table 6 Change in average critic and user valence over the weeks after films' opening week

Week after film's opening week	Variables	Mean	SD	Min	Max
2nd week	$\Delta CriticRating_{it}$	−0.0125	0.1191	−0.9356	0.3629
	$\Delta UserRating_{it}$	0.1376	1.0859	−1.1667	10.0000
3rd week	$\Delta CriticRating_{it}$	−0.0108	0.0649	−0.3351	0.2529
	$\Delta UserRating_{it}$	0.0198	0.4115	−1.2500	5.0000
4th week	$\Delta CriticRating_{it}$	−0.0146	0.0808	−0.5313	0.2371
	$\Delta UserRating_{it}$	0.0023	0.1696	−0.4286	1.5000
5th week	$\Delta CriticRating_{it}$	−0.0076	0.0606	−0.4003	0.2580
	$\Delta UserRating_{it}$	−0.0105	0.1456	−0.8571	0.5238
6th week	$\Delta CriticRating_{it}$	−0.0079	0.0410	−0.3281	0.0601
	$\Delta UserRating_{it}$	−0.0001	0.0708	−0.3036	0.3247
7th week	$\Delta CriticRating_{it}$	−0.0025	0.0203	−0.2227	0.0860
	$\Delta UserRating_{it}$	0.0068	0.0787	−0.2794	0.5000
8th week	$\Delta CriticRating_{it}$	−0.0037	0.0262	−0.1833	0.0647
	$\Delta UserRating_{it}$	0.0084	0.1697	−0.5308	2.0000
9th week	$\Delta CriticRating_{it}$	−0.0022	0.0197	−0.1525	0.0906
	$\Delta UserRating_{it}$	0.0096	0.0647	−0.3712	0.3500
10th week	$\Delta CriticRating_{it}$	−0.0012	0.0170	−0.2206	0.0494
	$\Delta UserRating_{it}$	0.0062	0.0757	−0.4074	0.6857

In sum, we find that expert opinion matters more than eWOM to movie revenues. Table 7 suggests that previous results, which fail to consider the impact of time-varying critic valence may overstate the impact of user volume and valence on market outcomes.

Table 7 OLS and instrument variables regression results using GMM estimations

	Model 1 OLS	Model 2 GMM	Models 1 & 2 VIFs	Model 3 OLS	GMM Model 4	Models 3 & 4 VIFs
Endogenous variables						
$\Delta UserVolume_{it-1}$	0.0710*** (0.0074)	0.1722*** (0.0182)	1.94	0.0163** (0.0066)	0.0617*** (0.0095)	2.17
$\Delta UserRating_{it-1}$	0.5540*** (0.1010)	1.2347*** (0.1024)	1.07	0.1102 (0.0738)	0.3433*** (0.0717)	1.16
$\Delta UserVariance_{it-1}$	1.4251*** (0.3011)	12.3171*** (2.1447)	1.09	0.5624*** (0.2018)	1.3388*** (0.4640)	1.11
$\Delta CriticRating_{it-1}$				0.8063*** (0.0774)	0.8552*** (0.0814)	1.70
$\Delta CriticVariance_{it-1}$				0.0634 (0.2235)	0.2386 (0.2539)	1.08
$\Delta Screen_{it}$	0.4000*** (0.0143)	0.2052*** (0.0514)	1.35	0.1478*** (0.0213)	0.2153*** (0.0304)	2.53
$\Delta Log\left(1 + Advs_{it-1}\right)$	0.1623*** (0.0250)	0.7064*** (0.1206)	1.18	0.0854*** (0.0158)	0.3116*** (0.0706)	1.20
Control variables						
$\Delta CriticVolume_{it-1}$				0.3038*** (0.0270)	0.1743*** (0.0318)	4.18
$\Delta Age^2_{it}(Week)$	−0.6982*** (0.0889)	−0.9476*** (0.2473)	1.01	−0.5171*** (0.0621)	−0.6460*** (0.0658)	1.02
Const^	0.4604*** (0.0955)	−0.7434*** (0.2862)	1.97	−0.2680*** (0.0731)	−0.1651 (0.1062)	2.17
Mean VIF			1.37			1.83
N	1629	1629		1629	1629	
F-value	330.78	185.68		370.51	385.30	
R-Sq.	0.6783			0.8653		
Adjusted R-Sq.	0.6769			0.8645		

Table 7 (continued)

	Model 1 OLS	Model 2 GMM	Models 1 & 2 VIFs	Model 3 OLS	GMM Model 4	Models 3 & 4 VIFs
F-Stat (p value)/χ^2 (p value) for testing CriticalRating > UserRating				36.27*** (< 0.0000)	17.38*** (< 0.0000)	
Tests for endogeneity: relevance and exogeneity of instruments						
First-stage F-statistics						
$\Delta UserVolume_{it-1}$		115.12***			49.02***	
$\Delta UserRating_{it-1}$		47.83***			34.65***	
$\Delta UserVariance_{it-1}$		56.70***			60.56***	
$\Delta CriticRating_{it-1}$					143.63***	
$\Delta CriticVariance_{it-1}$					67.01***	
$\Delta Screen_{it}$		91.40***			23.76***	
$\Delta Log\left(1 + Advs_{it-1}\right)$		53.99***			20.32***	
Hansen J-Statistic (p value)		4.482 (0.3446)			12.796 (0.1191)	

The dependent variable is the first difference of the natural log of weekly movie box office revenue. All estimations are with first differences of variable values to control for idiosyncratic movie specific effects

*, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively

(Cluster robust standard errors in parenthesis)

$\Delta \equiv$ first difference

^ *Const* not separately identified from $\Delta Age_{it}(Week)$

As an additional robustness check, we obtain review data on the same movies over the same timeframe from IMDB.com (for users) and Metacritic.com (for critics) and rerun the analysis. These data offer a more precise measure of the variance in user and critic reviews. Our results are qualitatively similar to Models 3 and 4 in Table 7 and presented in "Appendix 2".

4.1 Additional tests: aggregation bias

National data can be subject to an aggregation bias (Chintagunta et al. 2010). This may occur because inferences using national box office, aggregated across many heterogeneous markets, may reflect distribution strategy rather than the influence of user ratings or critic ratings. Chintagunta et al. (2010) highlight this concern: for movies with a sequential distribution strategy, premiering in some markets before others, the effect of user valance may be positively biased. This can occur if the movie is not well received in markets where it first opened, resulting in low user ratings, but box office increases anyway in the next week because the movie opens in additional markets. Indeed, a key contribution of Chintagunta et al. (2010) and Gopinath et al. (2013) is the use of data from individual geographic markets to deal with the aggregation problem.

Unfortunately we do not have individual data from geographic markets to ensure our results are robust to aggregation bias. However, aggregation bias is less likely to be an issue for wide release movies—these movies open simultaneously in most markets so aggregation bias caused by a sequential release strategy is not a concern. An uptick in box office the week following a wide release cannot be the result of entering new markets because the movie has already opened everywhere in the first week.

Figure 2 shows the average number of screens for wide release movies (i.e. non-platform release movies) and platform release movies by the number of weeks the movie is on the market. Wide release movies start on a large number of screens, but then drop precipitously as the movie ages. On the other hand, platform release movies on average are shown on more screens each week until the movie is five weeks old, then the number of screens declines from 6 weeks on. This reflects the sequential release strategy for platform release movies as they are released in more markets over time.

We repeat our analysis using only wide release movies. OLS and GMM results are shown in Table 8. GMM estimates are preferred because endogeneity is likely to be present.[9] Again, VIFs indicate that multicollinearity is not a concern in these estimations.

Our main result, that the impact of critic valance is significantly greater than the impact of user valence, holds when we estimate Eq. (2) with both OLS and GMM using only observations where aggregation bias is least likely to be present. Furthermore, for wide release movies, only experts' reviews matter, whereas there is

[9] Testing the null hypothesis that the endogenous regressors are exogenous in Model 6 yields $\chi^2(7) = 27.437$ and associated p value$=0.0003$.

no significant effect of user ratings on revenues. This is interesting, since "industry wisdom" suggests that wide release movies are carried by word of mouth whereas "art house" films are more influenced by professional critical reviews. Here we see that word of mouth valence does not seem to matter at all for wide release movies, whereas moviegoing is significantly affected by experts' reviews.

4.2 Additional tests: disagreement between critics and users

Our review of theory on the influence of experts and lay people suggests that there may be situations where experts' views matter more, whereas in other cases users' views matter more. When the level of disagreement between experts and laypeople is relatively high, lay people may rely on opinions of people they consider to be more like themselves. In order to test this idea, we estimate Eq. (2) using observations where critics and users disagree relatively more and compare the findings to a case where critics and users disagree relatively less.

Formally, we define $Disagreement\,Value_{it} = UserRating_{it} - CriticRating_{it}$ which captures the difference between user valence and critic valence for each movie in each period. In Table 9 we show the values of $Disagreement\,Value_{it}$ that break the data into quartiles.

We use the outer quartiles in Table 9 as our measure of a high level of disagreement between experts and consumers, whereas the inner quartiles proxy for more agreement. In other words, if $Disagreement\,Value_{it} < 0.3771$ or $Disagreement\,Value_{it} > 2.1938$, then we classify the observation as reflecting relatively large disagreement between users and critics. Similarly, observations with $0.3771 \leq Disagreement\,Value_{it} \leq 2.1938$ are classified as small disagreement observations.[10] This classification places roughly half of the observations in each category. Estimation results for Eq. (2) using both samples are displayed in Table 10.

Models 7 and 8 use OLS and GMM estimations, respectively, for cases when disagreement between users and critics is relatively small; Models 9 and 10 show results when disagreement is relatively large. VIFs for each set of models do not indicate an issue with multicollinearity. We present OLS results for comparison but focus on GMM estimations in Models 8 & 10 for our discussion—in both estimations we reject the null hypothesis that our endogenous regressors are exogenous.[11]

Model 8 shows that when users and critics are relatively close in their evaluation of a movie, the marginal impact of both a change in critic valence and a change in user valence are significant but critics matter more. In fact, in the OLS model (model 7) user ratings are insignificant. This model is similar in many respects to our base results shown in Model 4 from Table 7. However, there are a few notable differences

[10] As a robustness check we ran the estimations below using the less than or equal to constraint for classifying the relatively larger disagreement observations. Our results are qualitatively similar to those presented below.

[11] Testing the null hypothesis that the endogenous regressors are exogenous in Model 8 yields $\chi^2(7) = 23.893$ and associated p value $= 0.0012$; similarly, Model 10 yields $\chi^2(7) = 33.725$ and associated p value < 0.0000.

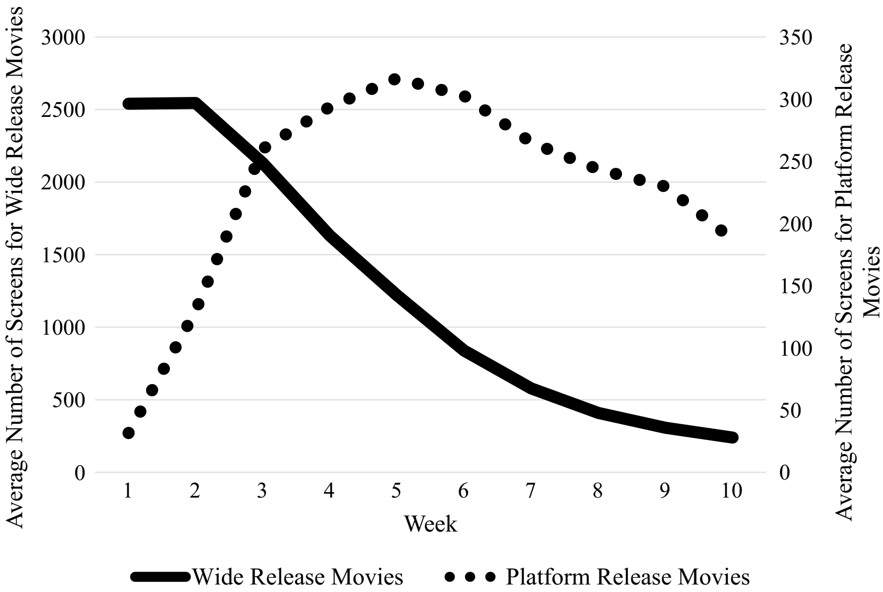

Fig. 2 Average number of screens by week for wide release and platform release movies

from Model 4. Specifically, the impact of user valence is lower and the impact of critic valence is greater in Model 8 compared to Model 4. This suggests that when professional critics and users are roughly in the same boat, experts' views matter a great deal and can sway the moviegoing public. This is a case where the distinctions drawn by Holbrook (1999) are less important—possibly there is less of a difference between "experts' opinions" and "popular appeal" and thus consumers are open to experts' recommendations.

In contrast, Model 10 shows that user opinions tend to become relatively more important when there is larger disagreement between users and critics. In model 9 (OLS) user reviews are significant. More importantly comparing Model 10 to Model 4, the coefficient on user valance is higher. This is even more pronounced when we compare Model 10 to Model 8. Indeed, in Model 10, we cannot reject the null hypothesis that user opinions have the same impact as critical opinions. The point estimates for experts' reviews are still higher, but the difference between the coefficient on experts' ratings and the coefficient of users' ratings is cut by more than half. In other words, when the moviegoing public feels very differently than the experts, then user ratings start to matter more. However, possibly in contrast with what we can expect from Holbrook (1999), experts' reviews still matter and may still matter more than users' reviews.

Table 8 Additional tests estimations for wide release movies

	Model 5 OLS	Model 6 GMM	Models 5 & 6 VIFs
Endogenous variables			
$\Delta UserVolume_{it-1}$	0.0224*** (0.0067)	0.0167** (0.0068)	2.50
$\Delta UserRating_{it-1}$	−0.0178 (0.1125)	0.1432 (0.1026)	1.21
$\Delta UserVariance_{it-1}$	0.9107** (0.3915)	2.0806*** (0.5459)	1.17
$\Delta CriticRating_{it-1}$	0.3742*** (0.1057)	0.4043*** (0.1040)	1.90
$\Delta CriticVariance_{it-1}$	−1.0985** (0.4331)	−0.9444** (0.4718)	1.16
$\Delta Screen_{it}$	0.2077*** (0.0284)	0.1481*** (0.0299)	3.95
$\Delta Log\left(1+Advs_{it-1}\right)$	0.0086 (0.0085)	0.0144 (0.0585)	1.28
Control variables			
$\Delta CriticVolume_{it-1}$	0.2860*** (0.0342)	0.3359*** (0.0299)	5.29
$\Delta Age^2_{it}(Week)$	−0.6500*** (0.1064)	−0.4033*** (0.0978)	1.08
Const ^	−0.1999** (0.0950)	−0.4213*** (0.0791)	2.48
Mean VIF			2.20
N	1003	1003	
F-value	435.16	477.85	
R-Sq.	0.9259		
Adjusted R-Sq.	0.9252		
F-Stat (p value)/ χ^2 (p value) for testing CriticalRating > UserRating	8.05*** (0.0054)	3.53* (0.0604)	
Tests for endogeneity: relevance and exogeneity of instruments			
First-stage F-statistics			
$\Delta UserVolume_{it-1}$		91.65***	
$\Delta UserRating_{it-1}$		141.28***	
$\Delta UserVariance_{it-1}$		14.92***	
$\Delta CriticRating_{it-1}$		163.26***	
$\Delta CriticVariance_{it-1}$		18.91***	
$\Delta Screen_{it}$		44.58***	
$\Delta Log\left(1+Advs_{it-1}\right)$		11.12***	
Hansen J-Statistic (p value)		13.113 (0.1080)	

The dependent variable is the first difference of the natural log of weekly movie box office revenue. All estimations are with first differences of variable values to control for idiosyncratic movie specific effects

*, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively

(Cluster robust standard errors in parenthesis)

$\Delta \equiv$ first difference

^ *Const* not separately identified from $\Delta Age_{it}(Week)$

Table 9 Distribution of disagreement between users and critics

$Disagreement\ Value_{it} = UserRating_{it} - CriticRating_{it}$	−8.4476	0.3771	1.1364	2.1938	7.0870
% of observations less than disagreement value	0	25	50	75	100

5 Conclusions

Electronic word of mouth (eWOM) seems to play a significant role in consumer decision-making processes and corporations allocate significant marketing budgets to manage their eWOM process. Several studies measure the effects of eWOM metrics (volume and valence) on sales. Two recent meta-analyses (Babic et al. 2016; You et al. 2015) document such endeavors within and outside of marketing. You et al. (2015) find that consumers search various online sources in order to understand and incorporate eWOM metrics in their decision processes and reduce uncertainty. A key characteristic of many such platforms is that along with eWOM they simultaneously (on the same page) carry or display experts' opinions and reviews for the same product or service (e.g., www.kbb.com for cars). However, most academic research focuses on eWOM, often excluding expert opinions that are accessible on the same platforms. However, earlier research (see Basuroy et al. 2003) points out the importance of expert reviews, and studies in the trade press document that consumers are influenced by both experts and eWOM and that expert opinions are playing a significant role in the purchasing decisions.

Our most important findings is that in a world where experts' opinions and user reviews are available side by side, experts' opinions matter more than eWOM for movie revenues. This finding is consistent with a recent paper (Rao et al. 2017) that shows that the most important element a studio can include in an ad is a review by a trusted critic (see also Ravid et al. 2006).

Second, surprisingly, we find that user ratings have no impact on wide release movies, whereas experts' views do matter. This runs counter to the conventional Hollywood view that blockbusters are sustained by word of mouth whereas "art house" (platform release) movies are carried by critical reviews. Our findings suggest that for any kind of film the importance of experts cannot be overstated.

We also show for the first time that expert reviews tend to be worse but more stable than users' reviews, possibly because users review only movies they chose to see and that choice already reflects a positive attitude toward the movie.

Our last finding is that the relative importance of experts' reviews versus eWOM is in determining revenues is affected by the degree of disagreement between the two. Thus, studios should worry more about critical reviews if a movie is likely to be equally liked (or disliked) by audiences and experts, whereas a movie that is likely to be panned by critics but please audiences (or the other way around) needs the support of eWOM as well.

These conclusions should matter to media outlets and to the studios. We show in this paper that at least at this stage of the digital revolution professional critics provide a very useful public service. Thus, media outlets can still benefit from hiring credible critics who can boost sales of the media in question. This flies in the face of

Table 10 Estimations where disagreement between users and critics are large versus small

	Relatively small disagreement between moviegoers and critics			Relatively large disagreement between moviegoers and critics		
	Model 7 OLS	Model 8 GMM	Models 7 & 8 VIFs	Model 9 OLS	Model 10 GMM	Models 9 & 10 VIFs
Endogenous variables						
$\Delta UserVolume_{it-1}$	0.0085 (0.0079)	0.0470*** (0.0126)	2.19	0.0256** (0.0103)	0.0852*** (0.0142)	2.20
$\Delta UserRating_{it-1}$	−0.0345 (0.0718)	0.2275** (0.0944)	1.12	0.2379** (0.1115)	0.5435*** (0.1044)	1.24
$\Delta UserVariance_{it-1}$	0.3985 (0.3557)	1.2233* (0.6408)	1.18	0.6003*** (0.2030)	1.0978*** (0.4074)	1.07
$\Delta CriticRating_{it-1}$	0.9052*** (0.0998)	1.0242*** (0.1023)	1.76	0.6848*** (0.1165)	0.6584*** (0.1264)	1.67
$\Delta CriticVariance_{it-1}$	−0.1903 (0.3728)	0.2362 (0.4889)	1.12	0.3371 (0.3047)	0.3959 (0.3307)	1.06
$\Delta Screen_{it}$	0.1432*** (0.0292)	0.2528*** (0.0386)	2.51	0.1521*** (0.0302)	0.1420*** (0.0430)	2.58
$\Delta Log\left(1 + Advs_{it-1}\right)$	0.0781*** (0.0256)	0.2207*** (0.0838)	1.22	0.0905*** (0.0206)	0.4479*** (0.1024)	1.18
Control variables						
$\Delta CriticVolume_{it-1}$	0.3150*** (0.0354)	0.1478*** (0.0394)	4.41	0.2908*** (0.0403)	0.2086*** (0.0453)	4.04
$\Delta Age^2_{it}(Week)$	−0.4471*** (0.0867)	−0.6177*** (0.0883)	1.03	−0.5486*** (0.0988)	−0.5685*** (0.1141)	1.01
Const^	−0.2721** (0.1155)	−0.0894 (0.1379)	2.20	−0.2854*** (0.0953)	−0.3574** (0.1510)	2.18
Mean VIF			1.87			1.82
N	815	815		814	814	
F-value	250.24	213.52		173.31	219.45	

Table 10 (continued)

	Relatively small disagreement between moviegoers and critics			Relatively large disagreement between moviegoers and critics		
	Model 7 OLS	Model 8 GMM	Models 7 & 8 VIFs	Model 9 OLS	Model 10 GMM	Models 9 & 10 VIFs
R-Sq.	0.8911			0.8445		
Adjusted R-Sq.	0.8897			0.8426		
F-Stat (p value)/χ^2 (p value) for testing CriticalRating > UserRating	65.81*** (< 0.0000)	32.95*** (< 0.0000)		5.90** (0.0167)	0.36 (0.5484)	
Tests for endogeneity: relevance and exogeneity of instruments						
First-stage F-statistics						
$\Delta UserVolume_{it-1}$		18.52***			48.02***	
$\Delta UserRating_{it-1}$		19.50***			23.68***	
$\Delta UserVariance_{it-1}$		35.69***			30.19***	
$\Delta CriticRating_{it-1}$		157.91***			46.73***	
$\Delta CriticVariance_{it-1}$		19.62***			31.80***	
$\Delta Screen_{it}$		11.02***			16.87***	
$\Delta Log\left(1 + Advs_{it-1}\right)$		12.72***			12.76***	
Hansen J-Statistic (p value)		9.053 (0.3379)		10.099 (0.2582)		

The dependent variable is the first difference of the natural log of weekly movie box office revenue. All estimations are with first differences of variable values to control for idiosyncratic movie specific effects

*, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively

(Cluster robust standard errors in parenthesis)

$\Delta \equiv$ first difference

^ *Const* not separately identified from $\Delta Age_{it}(Week)$

the belief in particular in the print media that the age of professional critics is over (see also Rao et al. 2017).

The other implication this work is for studios. As established in previous work, studios had been keenly aware of the importance of critics throughout the twentieth century (see Ravid et al. 2006). We suggest that studios should still pay attention to critics. We also show that far from blockbusters being "critics proof", they are very much influenced by critical reviews. For example, the Incredible Hulk, a presumed blockbuster from the Marvel series opened to tepid reviews and grossed $134 M in North America, well below its budget. At the other end of the Marvel universe, the Black Panther which opened to glowing reviews covered its budget of $200 M in the opening weekend alone.

Finally, we do find some support for Holbrook's (1999) idea in that movies where critics and audiences are sharply divided audiences views matter more. In such cases studios may want to spend more on audience development.

Acknowledgements Suman Basuroy thanks the Carl De Santis Center for Motion Picture Industry Studies for partially supporting this project with a grant. Avri Ravid thanks Rutgers Business School for a research grant partially supporting this work. We thank participants in the annual Business and Economics Scholars Workshop in Motion Picture Industry Studies as well as participants in the second behavioral economics workshop at Tel Aviv College for comments and suggestions. All errors remain our own.

Appendix 1: Instrumental variables and first-stage estimations

Instrumental variables

Rossi (2014) in a recent critical overview of instrumental variable analysis highlights the need for researchers to adequately identify the potential sources of omitted variable bias (i.e. endogeneity issues) and discuss why the chosen instruments should be considered exogenous from the estimation equation but related to the independent variables. We address these issues below. Then we discuss additional instruments generated leveraging heteroskedasticity in the first-stage estimations (Lewbel 2012) to further aid in identification.

There are seven potentially endogenous variables in our analysis: *UserVolume*, *UserRating*, *UserVariance*, *CriticRating*, *CriticVariance*, *Screen*, and *Advs*. We note that we are not concerned with correlation between the number of critic reviews (*CriticVolume*) and the error term because critics are typically assigned to review a film well in advance of the film's release date[12]—it is highly unlikely that editors when assigning critics to review films consider predictions of changes in a film's unobservable characteristics weeks after its release.

While we control for time invariant film characteristics with first differencing, we still need to control for possible time variant endogeneity. For example, both critical and user reviews may be driven by the characteristics of the focal film relative to other films on the market. The focal film's relative "quality" may change over the

[12] Private conversation with one of the authors and a New York Times movie reviewer.

10-week run as the set of other films on the market changes every week. Differential relative "quality" of the same movie may drive different average ratings from both critics and users as well as induce a higher volume of internet reviews which consequently can generate higher box office revenues. This results in endogeneity concerns for variance of critic and user ratings as well since they are a function of their respective averages. Similarly, different relative "quality" may impact studio distribution and advertising strategies. Chintagunta et al. (2010) and Gopinath et al. (2013) express similar concerns in their study and identify viable instruments for three of these variables, namely *UserVolume*, *UserRating*, and *Screen*.

We follow Neelamegham and Chintagunta (1999) in identifying an instrument for *Screen* using the average number of screens that show movies of the same genre as the focal movie i in week t ($CompScreen_{it}$). However, we deviate from Chintagunta et al. (2010) in our choice of instruments for *UserVolume* and *UserRating* since our data are national rather than local.

Instruments need to be correlated with the suspected endogenous variable but not with the error term in Eq. (2), $\Delta\varepsilon_{it}$ (Greene 2011a; Rossi 2014). One benefit of first differencing to account for time invariant film-specific effects is that we can use lagged levels of our endogenous variables as instruments (Greene 2011a): "without group effects, there is a simple instrumental variables estimator available. Assuming that the time series is long enough, one could use the lagged differences … or the lagged levels … (p. 308)". Lagged levels are appropriate as long as they are not correlated with $\Delta\varepsilon_{it}$ and influence the independent variable (Greene 2011a; Rossi 2014). For *UserVolume* and *UserRating* we believe users are highly unlikely to consider forecasts of $\Delta\varepsilon_{it}$ when (1) deciding to leave a review and (2) evaluating the film. Also, Moon et al. (2010) show that previous user ratings influence future user ratings. Additionally, in a recent article published in *Science*, Muchnik et al. (2013) show in a randomized experiment that social influence works on users in that prior ratings of others significantly affected individual ratings. Indeed, one problem of using lagged levels as instruments is that they are typically weak with little correlation to the first difference (Greene 2011a). However, as we show in our first-stage estimations below, this is not a concern in our dataset—the lagged levels of *UserVolume* and *UserRating* have significant explanatory power on their counterparts in the first-stage estimations.

We leverage the longitudinal aspect of our data set to obtain additional instruments which can help identify *UserVolume*. For film i observed in week t after release, we find the average number of user reviews in the tth week after release for all movies released before film i from a different genre. For example, the *UserVolume* of a movie observed in its 3rd week will be instrumented by the average 3rd week *UserVolume* of all movies from a different genre released before the focal movie. A similar approach is used in Lee's (2013) study of the video game industry when obtaining instruments for console and game prices. This measure will be uncorrelated with $\Delta\varepsilon_{it}$ by construction since users are very unlikely to consider forecasts of future films' $\Delta\varepsilon$, especially future films in different genres, when deciding to leave their reviews.

For each film in each week t, we obtain the average *UserVolume* in the tth week of all previously released movies in a different genre (*PrevUserVolumeDiffGen*). We use the change in average *t*th from the $t-1$th week to the *t*th week of all previously released movies in a different genre ($\Delta PrevUserVolumeDiffGen$). We expect a positive correlation with the change in *UserVolume* as the instrument likely captures industry wide changes in the amount of reviews consumers typically leave for films from week to week. Another concern highlighted by Rossi (2014) is the use of instruments that may not vary among groups (e.g. price indices). It is important to note that there is variation in this instrument by film and over time since very seldom we find two films of the same genre released at the same time in our dataset.

UserVariance is likely to be identified by instruments for *UserRating* because the latter is a nonlinear function of the former. To address this nonlinear relationship, we include the natural log of the lagged level of *UserRating*, $Log(1 + UserRating)$,[13] as an additional instrument for *UserValence*.

The valence of professional critic reviews, *CriticRating* may also be endogenous since critical reviews should be correlated with unobserved relative movie "quality." We use an instrument for critics' ratings designed to capture critic experience.[14] More experienced critics can be in a different position vis-a-vis corporate headquarters. There is an entire literature in finance and economics suggesting that people with more experience may require higher incentives to act in accordance with shareholders' values (Prendergast and Stole 1996; Scharfstein and Stein 1990; Zwiebel 1995). In line with this research, Ravid et al. (2006) argue that professional movie critics with a better reputation/more experience exhibit stronger corporate biases than others. As such, we expect average critics' experience to be negatively correlated with the change in average critical rating.

Reviewers' experience is almost by definition completely uncorrelated with the unobserved relative "quality" of the movie. It is difficult to obtain data on critics' tenure and experience, but as a proxy we are able to obtain previous reviews on RT.[15] We postulate that the greater the number of reviews, the greater the experience. For each critic, we find the number of reviews posted prior to reviewing the focal film. Then we find the level and the natural log of the average for all critics who review the focal film. Though we expect a positive impact of experience on the change in average critical rating, we include the natural log to address a possible nonlinear relationship. These values, $AvgPrevRev_{it}$ and $Log\left(1 + AvgPrevRev_{it}\right)$, change as new critical reviews become available throughout a movie's run.[16] Note that we do not use the lagged level of *CriticRating* as an instrument because it is unlikely that professional/well-trained critics coordinate their reviews taking into

13 We add 1 before taking the log to ensure the variable is defined when the level is 0.

14 While critics' ratings may be endogenous, we should emphasize that critics are assigned to review a movie by their editor, and studios do not have a role in that decision.

15 www.rottentomatoes.com keeps track of each critic's review history and can be found by clicking on the critic's name. The review history contains all previous reviews the critic has done including the date of the review.

16 We add 1 before taking the log to ensure the variable is defined when the level is 0.

account previous critical assessment of a film (although this remains an empirical question).[17]

As with *UserVariance* and *UserRating*, we expect *CriticVariance* is in part identified by instruments for *CriticRating*. However, we generate additional internal instruments leveraging heteroskedasticity in the first-stage estimation of *CriticVariance* (Lewbel 2012) to help separately identify the impact of the variance of critic rating on box office. We discuss this in greater depth below.

Another potential endogenous variable is advertising. We follow Chintagunta et al. (2010) and note that (a) prerelease advertising accounts for the vast majority of advertising spending in the movie industry (e.g., Elberse and Anand 2007 find that 88% of television advertising spending was spent prior to initial release), (b) prerelease advertising budgets are typically a fixed proportion of production budget (see Ravid 1999; Vogel 2007). Finally, any impact of prerelease advertising would be captured by μ_i since this value is unchanged over the run of the film—the estimations control for this through first differencing. However, changes in advertising expenditure over the course of a film's run may be influenced by the unobserved relative "quality" of competing films. As an instrument we leverage the panel structure of the data and use the level of advertising expenditure in the previous week. Since the endogenous variable in the estimation equation is the lagged first difference of advertising, the instrument is the 2-period lagged level of advertising. This is a valid instrument as it is unlikely that firms set advertising expenditures in a given week considering the change in forecasted future shocks two or more periods out, $\Delta\varepsilon_{it}$. In the movie industry, as is well established in the marketing literature, the key emphasis is on buzz creation and brand awareness (Houston et al. 2018). Advertising is less likely to have such long-term effects in this industry. For example, De Vries et al. (2017) find that empirically only one lag of advertising mattered in their VARX model exploring the influence of several variables (including traditional advertising) on their contemporaneous counterparts.[18]

Additional instruments leveraging heteroskedasticity in the first-stage estimations

We note that the empirical results below are robust to including only the traditional instruments we outline above. However, in our empirical setting it is difficult to

[17] We thank an anonymous referee for pointing this out.

[18] To provide additional evidence of this assertion we calculate the GMM distance statistic (Hayashi 2000) to directly test the validity of all instruments created using the lagged level of advertising. The GMM distance statistic compares the Hansen's J of two models—one including all instruments and one excluding the instruments created using lagged level advertising—where exogeneity of all instruments is the null hypothesis. Under the null, the test statistic is Chi square with degrees of freedom equal to the number of suspect instruments. The alternative hypothesis is that the suspect instruments are correlated with the error term and therefore invalid. This test relies on the assumption that the instruments not being tested are indeed exogenous. We believe this is likely given the theoretical arguments and the literature support we provide above. The GMM distance statistic testing the exogeneity of lagged level advertising is $\chi^2(1) = 0.557$ with a p value = 0.4554. Failure to reject the null provides additional support that lagged level advertising is exogneous.

strongly identify all endogenous variables using traditional external instruments alone. Identification improves when we leverage the procedure outlined in Lewbel (2012) and augment the traditional instruments with internal instruments that exploit heteroskedasticity in the first-stage estimations.[19]

Lewbel (2012) shows that if heteroskedasticity is present in the first-stage estimations,[20] then additional instruments can be created by interacting the residuals of the respective first-stage estimation with any (or all) demeaned independent variables included in the first-stage estimation. The key idea is that while the exogenous regressors in the first-stage estimation are uncorrelated with the error term in the first-stage regression by construction, there is no reason to believe that the residuals will be independent of the regressors in the reduced form estimation. If the residuals are heteroskedastic (i.e. dependent on the regressors), then this information can be used to further untangle the endogenous part of the offending variable from the exogenous part. In fact, more heteroskedasticity aids in identifying the endogenous regressor (Lewbel 2012).

In the extreme, Lewbel (2012) shows that models are identified using only heteroskedasticity in first-stage estimations without any additional instruments, though "[t]he resulting identification is based on higher moments and so likely to provide less reliable estimates than identification based on standard exclusion restrictions, but may be useful in applications where traditional instruments are not available or could be used along with traditional instruments to increase efficiency" (p. 67). We follow this advice and use any additional instruments along with the traditional instruments we discuss below.

Additional instruments are created first by estimating the first-stage regression including all exogenous variables and instruments and obtaining residuals. Then the residuals are interacted with demeaned values of the relevant regressors from the first-stage estimation to create new variables. Any new variable created using this method is then included like a standard instrument in instrumental variable analysis.

We do not include extra instruments generated from the residuals of all first-stage estimations interacted with all exogenous variables to avoid instrument proliferation which may weaken results (Roodman 2009). Rather, we focus on generating instruments for endogenous variables that may be difficult to identify using external instruments alone. Our main concern is with identifying $\Delta UserVolume$, $\Delta UserVariance$, $\Delta CriticRating$, $\Delta CriticVariance$, and $\Delta Screen$. We believe the impact of the variance of user opinion will be difficult to separately identify from *UserRating* and *UserVolume* (given *UserVariance* is a function of both) with external instruments only. It also may be difficult to separately identify critic rating and variance of critic rating for similar reasons; this is likely compounded by the fact that we have only two external instruments ($AvgPrevRev_{it}$ and $Log\left(1 + AvgPrevRev_{it}\right)$) for both endogenous variables. Lastly, we augment the single instrument for $\Delta Screen$,

[19] We thank an anonymous referee for suggesting this approach.

[20] We note that we find significant heteroscedasticity in each first-stage estimation using the Wald test for groupwise heteroskedasticity proposed by Greene (2011a, b) for panel data.

which is not based on lagged level of the variable as the instrument for advertising is, to aid in identification.

For $\Delta UserVariance$ we create an additional instrument by interacting the demeaned $\Delta PrevUserVolumeDiffGen$ ($DM(\Delta PrevUserVolumeDiffGen)$ where $DM(\bullet)$ is the demeaning operator) with the residuals from the preliminary first-stage estimation of lagged first differenced $UserVariance$ ($Res\Delta UserVariance_{it-1}$). Our expectation is that the size of the variation in the variance of user opinion changes as more users leave reviews—we expect $\Delta PrevUserVolumeDiffGen$ will capture this since it is a relevant instrument for $UserVolume$. We include the interaction of the first-stage residuals for $\Delta UserVolume$ with its demeaned lag level to help separately identify from $\Delta UserVariance$. For $\Delta CriticRating$ and $\Delta CriticVariance$, we create two additional instruments by interacting the residuals from both preliminary first-stage estimations with demeaned $Log\left(1 + AvgPrevRev_{it}\right)$ because it is likely that critic experience influences both critic rating and the size of the variation in critic rating.[21] Additionally, we create another Lewbel (2012) style instrument for $\Delta CriticRating$ using lagged level user rating. The key idea is that the level of disagreement among critics may be related to the popular appeal of the film. Also as the variance in critical opinion may be related to the popularity of the film, we include another instrument for $\Delta CriticVariance$ using lagged level user volume. Finally, we use lagged level user volume to create another instrument for $\Delta Screen$. This is a relevant instrument if studios alter their distribution intensity in response to online chatter and if the variation in the size of the response varies with the amount of online chatter.

First-stage estimations

We list the descriptive statistics for the traditional external instruments as well as the Lewbel (2012) style instruments in Table 11. We show the first-stage estimations for our full model in Table 12. The first-stage results indicate the instruments are relevant (all first-stage F-statistics are well above 10) and the signs generally conform to priors: $\Delta UserVolume$, $\Delta UserRating$, and $\Delta Log(1 + Advs)$ are significantly identified by using lagged levels; $\Delta PrevUserVolumeDiffGen$ has a positive and significant impact on $\Delta UserVolume$; critical experience captured in $AvgPrevRev_{it}$ and $Log\left(1 + AvgPrevRev_{it}\right)$ is significantly related $\Delta CriticRating$[22] and $\Delta CriticVariance$; $\Delta CompScreen$ significantly identifies $\Delta Screen$.

Finally, the additional instruments created leveraging heteroskedasticity in the preliminary first-stage estimations aid in identification. They significantly impact their respective endogenous variables, and they are all significant in other first-stage estimations. Note that we do not make predictions of the signs of these additional

[21] Additional instruments created using $AvgPrevRev_{it}$ for both critical valence and variance were never significant.

[22] We note that impact of critic experience on $\Delta CriticRating$ is consistent with our priors since the negative coefficient on $Log\left(1 + AvgPrevRev_{it}\right)$ dominates the positive coefficient on $AvgPrevRev_{it}$ for all values of average experience in our dataset.

instruments since we do not have theoretical guidance on the impact of higher moments (Lewbel 2012). However, the significance of these variables suggests they are related to heteroskedasticity in the first-stage estimations (which is what is required for identification).

Table 11 Descriptive statistics for instrumental variables

Variables	Mean	SD	Min	Max
Instrumental variables				
$UserVolume_{it-2}$	59.1473	51.7554	0.0000	243.0000
$\Delta PrevUserVolumeDiffGen_{it-1}$	7.8268	6.1679	0.0000	27.0563
$UserRating_{it-2}$	6.8291	1.6917	0.0000	10.0000
$Log\left(1 + UserRating_{it-2}\right)$	2.0009	0.4344	0.0000	2.3979
$\Delta CompScreen_{it}$	− 0.0286	1.7560	− 12.4300	24.5400
$Log\left(1 + Advs_{it-2}\right)$	10.4403	5.1960	0.0000	16.8053
$AvgPrevRev_{it-2}$	521.7126	220.0828	0.0000	1676.5000
$Log\left(1 + AvgPrevRev_{it-2}\right)$	5.7711	1.7562	0.0000	7.4251
Lewbel (2012) style instrumental variables				
$DM\left(UserVolume_{it-2}\right) \times Res\Delta UserVolume_{it-1}$	0.0000	516.4667	− 3635.2150	3131.4410
$\mathrm{DM}\left(\Delta PrevUserVolumeDiffGen_{it-1}\right) \times Res\Delta UserVariance_{it-1}$	0.0000	2.3041	− 19.4699	58.4152
$\mathrm{DM}\left(Log\left(1 + AvgPrevRev_{it-2}\right)\right) \times Res\Delta CriticRating_{it-1}$	0.0000	4.4473	− 25.1672	41.4343
$\mathrm{DM}\left(UserRating_{it-2}\right) \times Res\Delta CriticRating_{it-1}$	0.0000	4.0642	− 41.7535	45.5142
$\mathrm{DM}\left(Log\left(1 + AvgPrevRev_{it-2}\right)\right) \times Res\Delta CriticVariance_{it-1}$	0.0000	0.3886	− 10.6706	2.8255
$\mathrm{DM}\left(UserVolume_{it-2}\right) \times Res\Delta CriticVariance_{it-1}$	0.0000	11.0419	− 191.0933	84.8986
$\mathrm{DM}\left(UserVolume_{it-2}\right) \times Res\Delta Screen_{it}$	0.0000	241.5792	− 1931.5640	1152.4460

$N = 1629$

DM(•) ≡ Variable in parenthesis is demeaned

$ResX$ ≡ Residuals from the preliminary first-stage regression of X. For example, $Res\Delta Screen_{it}$ are the residuals from the preliminary first-stage regression of $\Delta Screen_{it}$

Table 12 First-stage regressions. The dependent variable in each estimation is listed across the first row

	$\Delta UserVolume_{it-1}$	$\Delta UserRating_{it-1}$	$\Delta UserVariance_{it-1}$	$\Delta CriticRating_{it-1}$	$\Delta CriticVariance_{it-1}$	$\Delta Screen_{it}$	$\Delta Log(1 + Advs_{it-1})$
Excluded instruments							
$UserVolume_{it-2}$	0.0444***	0.0005*	−0.0004***	0.0004	−0.0002***	−0.0084***	−0.0076***
	(0.0088)	(0.0003)	(0.0001)	(0.0003)	(0.0001)	(0.0030)	(0.0014)
$\Delta PrevUserVolumeDiffGen_{it-1}$	1.0214***	−0.0135*	0.0041***	0.0060	0.0068***	−0.0434*	0.1467***
	(0.0833)	(0.0079)	(0.0014)	(0.0079)	(0.0009)	(0.0248)	(0.0168)
$UserRating_{it-2}$	−1.4040**	0.2748***	−0.0179	0.1917***	−0.0031	0.7201***	0.5160***
	(0.6112)	(0.0616)	(0.0232)	(0.0514)	(0.0075)	(0.1971)	(0.1196)
$Log(1 + UserRating_{it-2})$	12.4239***	−3.6243***	0.1368*	−1.0659***	0.0237	−1.2489	−1.6201***
	(2.3286)	(0.3404)	(0.0808)	(0.2176)	(0.0268)	(0.7902)	(0.4284)
$\Delta CompScreen_{it}$	−0.1110	−0.0441*	0.0049	−0.0232***	−0.0003	0.2928***	0.0246
	(0.1214)	(0.0233)	(0.0031)	(0.0082)	(0.0013)	(0.0879)	(0.0461)
$Log(1 + Advs_{it-2})$	0.3458***	0.0072**	−0.0018	−0.0071*	−0.0007	−0.2050***	−0.1991***
	(0.0593)	(0.0029)	(0.0013)	(0.0037)	(0.0008)	(0.0230)	(0.0176)
$AvgPrevRev_{it-2}$	−0.0018	−0.0001	0.00000	0.0006**	−0.0001***	0.0024***	0.0003
	(0.0019)	(0.0002)	(0.00004)	(0.0003)	(0.0000)	(0.0008)	(0.0005)
$Log(1 + AvgPrevRev_{it-2})$	−0.3178	0.0763	−0.0181***	−0.4881***	0.0252***	−0.8101***	−0.0392
	(0.2940)	(0.0464)	(0.0057)	(0.0465)	(0.0031)	(0.1197)	(0.0520)
Lewbel (2012) style instrumental variables							
$\mathrm{DM}(UserVolume_{it-2}) \times Res\Delta UserVolume_{it-1}$	−0.0085***	−0.0001	0.00003***	0.0000	0.00000	−0.0009***	0.0002
	(0.0014)	(0.0001)	(0.00001)	(0.0001)	(0.00001)	(0.0003)	(0.0001)
$\mathrm{DM}(\Delta PrevUserVolumeDiffGen_{it-1}) \times Res\Delta UserVariance_{it-1}$	−0.3703***	0.0011	0.0645***	0.0075	−0.0003	0.0285	0.0184
	(0.1298)	(0.0155)	(0.0076)	(0.0128)	(0.0005)	(0.0511)	(0.0237)
$\mathrm{DM}(Log(1 + AvgPrevRev_{it-2})) \times Res\Delta CriticRating_{it-1}$	0.1298**	0.0121	−0.0003	−0.1712***	0.0011*	0.0181	0.0026
	(0.0660)	(0.0076)	(0.0010)	(0.0069)	(0.0006)	(0.0278)	(0.0085)
$\mathrm{DM}(UserRating_{it-2}) \times Res\Delta CriticRating_{it-1}$	−0.1309**	−0.0028	0.0003	−0.0489***	−0.0002	0.0572**	−0.0131**
	(0.0528)	(0.0172)	(0.0007)	(0.0170)	(0.0004)	(0.0254)	(0.0062)
$\mathrm{DM}(Log(1 + AvgPrevRev_{it-2})) \times Res\Delta CriticVariance_{it-1}$	0.4315	−0.1142**	0.0127	−0.0277	0.0531*	0.3213	−0.2185*
	(0.4314)	(0.0559)	(0.0159)	(0.0671)	(0.0305)	(0.2079)	(0.1322)

Table 12 (continued)

	$\Delta UserVolume_{it-1}$	$\Delta UserRating_{it-1}$	$\Delta UserVariance_{it-1}$	$\Delta CriticRating_{it-1}$	$\Delta CriticVariance_{it-1}$	$\Delta Screen_{it}$	$\Delta Log\left(1+Advs_{it-1}\right)$
$\mathrm{DM}\left(UserVolume_{it-2}\right) \times Res\Delta CriticVariance_{it-1}$	−0.0078 (0.0138)	0.0010 (0.0018)	0.0004 (0.0007)	−0.0019 (0.0071)	−0.0171*** (0.0013)	0.0154** (0.0065)	0.0008 (0.0048)
$\mathrm{DM}\left(UserVolume_{it-2}\right) \times Res\Delta Screen_{it}$	−0.0043*** (0.0016)	−0.0001 (0.0001)	−0.00003 (0.00004)	0.0003** (0.0001)	0.00003*** (0.00001)	−0.0098*** (0.0014)	0.0003 (0.0002)
Control variables							
$\Delta CriticVolume_{it-1}$	0.4104*** (0.0561)	0.0167** (0.0072)	0.0002 (0.0008)	0.0381*** (0.0084)	−0.0001 (0.0004)	0.6724*** (0.0215)	0.0175** (0.0076)
$\Delta Age_{it}^{2}(Week)$	0.5442 (0.3332)	−0.0207 (0.0437)	−0.0076 (0.0147)	−0.0876*** (0.0228)	0.0280*** (0.0085)	0.9107*** (0.1958)	−0.2329** (0.0949)
$Const$^	−18.5350*** (2.1565)	5.1519*** (0.4629)	0.0341 (0.0348)	3.5666*** (0.3040)	−0.0865*** (0.0184)	1.7300** (0.7604)	−0.0431 (0.3192)
N	1629	1629	1629	1629	1629	1629	1629
First-stage F-statistics on excluded instruments	49.02***	34.65***	60.56***	143.63***	67.01***	23.76***	20.32***

*, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively

(Cluster robust standard errors in parenthesis)

Δ $\equiv$ first difference

Appendix 2: Robustness check of main model using imdb.com and metacritic.com data

We obtain data on critical evaluation from Metacritic.com and consumer evaluation from IMDB.com for each movie included in the analysis in a similar fashion as the RT data. We use the Metacritic.com and IMDB.com data to calculate variance in critic and user valence within a week rather than the variance in average evaluation from week to week, something we are not able to do with the RT data. We estimate the main models in the paper (Table 7, Models 3 and 4) using this data, along with the more precise measures of variance. Descriptive statistics for these variables and for the instruments used in this robustness check are in Table 13 (note that we are able to identify the endogenous variables in instrumental variable estimation with fewer Lewbel style instruments); estimation results are in Table 14. The results are qualitatively similar to the findings in the text based on the RT data.

Table 13 Descriptive statistics for IMDB.com and Metacritic.com data

Variables	Mean	SD	Min	Max
$UserVolume - IMDB_{it}$	109.8300	191.5468	0.0000	1346.0000
$UserRating - IMDB_{it}$	6.7193	1.4166	0.0000	10.0000
$UserVariance - IMDB_{it}$	6.6657	3.0740	0.0000	17.3878
$CriticRating - Metacritic_{it}$	5.8943	2.1451	0.0000	9.7750
$CriticVariance - Metacritic_{it}$	1.8408	1.0872	0.0000	5.3469
Instrumental variables for IMDB.com and Metacritic.com data analogous to instrumental variables used for Rottentomatoes.com data				
$UserVolume - IMDB_{it-2}$	6.2477	2.3502	0.0000	10.0000
$Log\left(1 + UserRating - IMDB_{it-2}\right)$	1.8647	0.6033	0.0000	2.3979
Other instrumental variables (same as instruments used with Rottentomatoes.com data)				
$\Delta PrevUserVolumeDiffGen_{it-1}$	7.8268	6.1679	0.0000	27.0563
$\Delta CompScreen_{it}$	−0.0286	1.7560	−12.4300	24.5400
$Log\left(1 + Advs_{it-2}\right)$	10.4403	5.1960	0.0000	16.8053
$AvgPrevRev_{it-2}$	521.7126	220.0828	0.0000	1676.5000
$Log\left(1 + AvgPrevRev_{it-2}\right)$	5.7711	1.7562	0.0000	7.4251
Lewbel (2012) style instrumental variables for IMDB.com and Metacritic.com data				
$\mathrm{DM}\left(\Delta PrevUserVolumeDiffGen_{it-1}\right) \times Res\Delta UserVariance - IMDB_{it-1}$	0.0000	20.0818	−226.1392	263.2552
$\mathrm{DM}\left(UserVolume - IMDB_{it-2}\right) \times Res\Delta CriticVariance - Metacritic_{it-1}$	0.0000	50.0907	−689.0827	431.6583
$\mathrm{DM}\left(UserVolume - IMDB_{it-2}\right) \times Res\Delta Screen_{it}$	0.0000	668.3252	−9246.9270	9461.6350

$N = 1629$

$\mathrm{DM}(\bullet)$ $\equiv$ Variable in parenthesis is demeaned

$ResX$ $\equiv$ Residuals from the preliminary first-stage regression of X. For example, $Res\Delta Screen_{it}$ are the residuals from the preliminary first-stage regression of $\Delta Screen_{it}$

Table 14 OLS and instrument variables robustness check regression results using GMM estimations

	Robustness check OLS	Robustness check GMM	Robustness check VIFs
Endogenous variables			
$\Delta UserVolume - IMDB_{it-1}$	−0.0070*** (0.0018)	0.0256*** (0.0055)	1.18
$\Delta UserRating - IMDB_{it-1}$	0.0181 (0.0546)	0.3932*** (0.0892)	1.28
$\Delta UserVariance - IMDB_{it-1}$	−0.0168 (0.0393)	−0.0761 (0.0809)	1.20
$\Delta CriticRating - Metacritic_{it-1}$	0.5939*** (0.0754)	0.7982*** (0.1398)	2.08
$\Delta CriticVariance - Metacritic_{it-1}$	0.1818 (0.1589)	−0.1002 (0.1915)	1.56
$\Delta Screen_{it}$	0.1479*** (0.0252)	0.2359*** (0.0536)	2.51
$\Delta Log\left(1 + Advs_{it-1}\right)$	0.0918*** (0.0160)	0.4420*** (0.0796)	1.20
Control variables			
$\Delta CriticVolume_{it-1}$	0.3112*** (0.0250)	0.1839*** (0.0478)	4.34
$\Delta Age^2_{it}(Week)$	−0.4206*** (0.0795)	−0.8735*** (0.1551)	1.12
Const^	−0.0330 (0.0945)	0.2355* (0.1342)	1.81
Mean VIF			1.83
N	1629	1629	
F-value	346.98	276.60	
R-Sq.	0.8548		
Adjusted R-Sq.	0.8539		
F-Stat (*p* value)/χ^2 (*p* value) for testing CriticalRating > UserRating	32.62*** (< 0.0000)	5.49** (0.0192)	
Tests for endogeneity: relevance and exogeneity of instruments			
First-stage F-statistics			
$\Delta UserVolume - IMDB_{it-1}$		40.74***	
$\Delta UserRating - IMDB_{it-1}$		36.83***	
$\Delta UserVariance - IMDB_{it-1}$		95.43***	
$\Delta CriticRating - Metacritic_{it-1}$		24.04***	
$\Delta CriticVariance - Metacritic_{it-1}$		583.45***	
$\Delta Screen_{it}$		10.47***	
$\Delta Log\left(1 + Advs_{it-1}\right)$		14.26***	
Hansen J-statistic (*p* value)		3.723 (0.2930)	

The dependent variable is the first difference of the natural log of weekly movie box office revenue. All estimations are with first differences of variable values to control for idiosyncratic movie specific effects

*, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively

(Cluster robust standard errors in parenthesis)

Δ ≡first difference

^*Const* not separately identified from $\Delta Age_{it}(Week)$

References

Babic, A., Sotgiu, F., de Valck, K., & Bijmolt, T. H. (2016). The effect of electronic word of mouth on sales: A meta-analytic review of platform, product, and metric factors. *Journal of Marketing Research, 53*(3), 297–318.

Baker, A. M., Donthu, N., & Kumar, V. (2016). Investigating how word of mouth conversations about-brands influence purchase and retransmission intentions. *Journal of Marketing Research*, *LIII*, 225–239.

Basuroy, S., Subimal Chatterjee, S., & Ravid, A. (2003). How critical are critical reviews? The box office effects of film critics, star power, and budgets. *Journal of Marketing, 67*, 103–117.

Cameron, A. C., & Trivedi, P. K. (2005). *Microeconometrics: Methods and applications*. Cambridge: Cambridge University Press.

Chen, Y., Liu, Y., & Zhang, J. (2012). When do third-party product reviews affect firm value and what can firms do? The case of media critics and professional movie reviews. *Journal of Marketing, 76*(2), 116–134.

Chevalier, J. A., & Mayzlin, D. (2006). The effect of word of mouth on sales: Online book reviews. *Journal of Marketing Research, 43*(3), 345–354.

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.

De Vany, A., & Walls, W. D. (1999). Uncertainty in the movies: Can star power reduce the terror at the box office? *Journal of Cultural Economics, 23*(4), 235–318.

De Vany, A., & Walls, W. D. (2002). Does hollywood make too many R-rated movies? Risk, stochastic dominance, and the illusion of expectation. *Journal of Business, 75*(3), 425–451.

De Vries, L., Gensler, S., & Leeflang, P. S. H. (2017). Effects of traditional advertising and social messages on brand-building metrics and customer acquisition. *Journal of Marketing, 81*(5), 1–15.

Duan, W., Bin, G., & Whinston, A. B. (2008). The dynamics of online word-of-mouth and product sales—An empirical investigation of the movie industry. *Journal of Retailing, 84*(2), 233–242.

Einav, L. (2007). Seasonality in the U.S. motion picture industry. *Rand Journal of Economics, 38*(1), 127–145.

Elberse, A. (2007). The power of stars: Do star actors drive the success of movies? *Journal of Marketing, 71*(4), 102–120.

Elberse, A., & Anand, B. (2007). The effectiveness of pre-release advertising for motion pictures: An empirical investigation using a simulated market. *Information Economics and Policy, 19*(3), 319–343.

Elberse, A., & Eliashberg, J. (2003). Demand and supply dynamics behavior for sequentially released products in international markets: The case of motion pictures. *Marketing Science, 22*(3), 329–354.

Eliashberg, J., & Shugan, S. M. (1997). Film critics: Influencers or predictors? *Journal of Marketing, 61*(April), 68–78.

Gopinath, S., Chintagunta, P. K., & Venkataraman, S. (2013). Blogs, advertising, and local-market movie box office performance. *Management Science, 59*(12), 2635–2654.

Greene, W. H. (2011a). *Econometric analysis* (7th ed.). Upper Saddle River: Prentice Hall.

Greene, W. H. (2011b). Fixed effects vector decomposition: A magical solution to time-invariant variables in fixed effects models? *Political Analysis, 19*(2), 135–146.

Hayashi, F. (2000). Econometrics. Princeton University Press.

Ho, J. Y. C., Dhar, T., & Weinberg, C. B. (2009). Playoff-payoff: Superbowl advertising for movies. *International Journal of Research in Marketing, 26*(3), 168–179.

Holbrook, M. B. (1999). Popular appeal versus expert judgments of motion pictures. *Journal of Consumer Research, 26*(September), 144–155.

Houston, M., Kupfer, A.-K., Hennig-Thurau, T., & Spann, M. (2018). Pre-release consumer buzz. *Journal of the Academy of Marketing Science, 46*(2), 338–360.

Imbens, G. W. (2002). Generalized method of moments and empirical likelihood. *Journal of Business and Economic Statistics, 20,* 493–506.

Klein, L. R., & Ford, G. T. (2003). Consumer search for information in the digital age: An empirical study of prepurchase search for automobiles. *Journal of Interactive Marketing, 17*(3), 29–49.

KRC Research. (2012). *Buy it, try it, rate it*. Available at, https://www.webershandwick.com/uploads/news/files/ReviewsSurveyReportFINAL.pdf. Accessed August 2, 2015.

Kupfer, A. K., Pähler vor der Holte, N., Kübler, R. V., & Hennig-Thurau, T. (2018). The role of the partner brand's social media power in brand alliances. *Journal of Marketing, 82*(3), 25–44.

Lee, R. S. (2013). Vertical integration and exclusivity in platform and two-sided markets. *American Economic Review, 103*(7), 2960–3000.

Lewbel, A. (2012). Using heteroscedasticity to identify and estimate mismeasured and endogenous regressor models. *Journal of Business & Economic Statistics, 30*(1), 67–80.

Liu, Y. (2006). Word-of-mouth for movies: Its dynamics and impact on box office revenue. *Journal of Marketing, 70*(3), 74–89.

Maity, M., Dass, M., & Malhotra, N. K. (2014). The antecedents and moderators of offline information search: A meta-analysis. *Journal of Retailing, 90*(2), 233–254.

Marchand, A., Hennig-Thurau, T., & Wiertz, C. (2017). Not All digital word of mouth is created equal: Understanding the respective impact of consumer reviews and microblogs on new product success. *International Journal of Research in Marketing, 34*(2), 336–354.

Mayzlin, D., Dover, Y., & Chevalier, J. A. (2014). Promotional reviews: An empirical investigation of online review manipulation. *American Economic Review, 104*(8), 2421–2455.

Moon, S., Bergey, P. K., & Iacobucci, D. (2010). Dynamic effects among movie ratings, movie revenues, and viewer satisfaction. *Journal of Marketing, 74*(1), 108–121.

Moorthy, S., Ratchford, B. T., & Talukdar, D. (1997). Consumer information search revisited: Theory and empirical analysis. *Journal of Consumer Research, 23*(4), 263–277.

Muchnik, L., Aral, S., & Taylor, S. J. (2013). Social influence bias: A randomized experiment. *Science, 341*, 647–651.

Neelamegham, R., & Chintagunta, P. (1999). A Bayesian model to forecast new product performance in domestic and international markets. *Marketing Science, 18*(2), 115–136.

PBS Newshour. (2015). *Spotting the fakes among the five-star reviews*. Available at, http://www.pbs.org/newshour/bb/spotting-fakes-among-five-star-reviews/. Accessed September 7, 2015.

Prendergast, C., & Stole, L. (1996). Impetuous youngsters and jaded old-timers: Acquiring a reputation for learning. *Journal of Political Economy, 104*(6), 1105–1134.

Radas, S., & Shugan, S. M. (1998). Seasonal marketing and timing new product introductions. *Journal of Marketing Research, 35*(3), 296–315.

Rao, V. R., Abraham, S., Basuroy, S., Gretz, R., & Chen, J. (2017). The impact of advertising content on movie revenues. *Marketing Letters, 28,* 341–355.

Ratchford, B. T., Talukdar, D., & Lee, M.-S. (2007). The impact of the internet on consumers' use of information sources for automobiles: A re-inquiry. *Journal of Consumer Research, 34*(1), 111–119.

Ravid, S. A. (1999). Information, blockbusters, and stars: A study of the film industry. *Journal of Business, 72*(4), 463–492.

Ravid, S. A., Wald, J. K., & Basuroy, S. (2006). Distributors and critics: Does it takes two to Tango? *Journal of Cultural Economics, 30*(3), 201–218.

Roodman, D. (2009). A note on the theme of too many instruments. *Oxford Bulletin of Economics and Statistics*, *71*(1), 135–158.

Rossi, P. E. (2014). Invited paper—Even the rich can make themselves poor: A critical examination of IV methods in marketing applications. *Marketing Science, 33*(5), 655–672.

Sass, E. (2013). *Most marketers will spend more on social media in 2014*. Available at, http://www.mediapost.com/publications/article/213850/most-marketers-will-spend-more-on-social-media-in.html. Accessed September 4, 2015.

Scharfstein, D. S., & Stein, J. C. (1990). Herd behavior and investment. *American Economic Review, 80*(3), 465–479.

Urbany, J. E., Dickson, P. R., & Wilkie, W. L. (1989). Buyer uncertainty and information search. *Journal of Consumer Research, 16*(2), 208–215.

Vogel, H. L. (2007). *Entertainment industry economics* (7th ed.). Cambridge: Cambridge University Press.

You, Ya., Vadakkepatt, G. G., & Joshi, A. M. (2015). A meta-analysis of electronic word-of-mouth elasticity. *Journal of Marketing, 79*(2), 19–39.

Zhang, X., & Dellarocas, C. (2006). The lord of the ratings: Is a movie's fate influenced by reviews? In *ICIS 2006 proceedings*, 1959–1978.
Zwiebel, J. (1995). Corporate conservatism and relative compensation. *Journal of Political Economy, 103*(1), 1–25.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.