



Marketing Science

Publication details, including instructions for authors and subscription information:
<http://pubsonline.informs.org>

The Effects of Online User Reviews on Movie Box Office Performance: Accounting for Sequential Rollout and Aggregation Across Local Markets

Pradeep K. Chintagunta, Shyam Gopinath, Sriram Venkataraman,

To cite this article:

Pradeep K. Chintagunta, Shyam Gopinath, Sriram Venkataraman, (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. <https://doi.org/10.1287/mksc.1100.0572>

Full terms and conditions of use: <http://pubsonline.informs.org/page/terms-and-conditions>

This article may be used only for the purposes of research, teaching, and/or private study. Commercial use or systematic downloading (by robots or other automatic processes) is prohibited without explicit Publisher approval, unless otherwise noted. For more information, contact permissions@informs.org.

The Publisher does not warrant or guarantee the article's accuracy, completeness, merchantability, fitness for a particular purpose, or non-infringement. Descriptions of, or references to, products or publications, or inclusion of an advertisement in this article, neither constitutes nor implies a guarantee, endorsement, or support of claims made of that product, publication, or service.

Copyright © 2010, INFORMS

Please scroll down for article—it is on subsequent pages



INFORMS is the largest professional society in the world for professionals in the fields of operations research, management science, and analytics.

For more information on INFORMS, its publications, membership, or meetings visit <http://www.informs.org>

The Effects of Online User Reviews on Movie Box Office Performance: Accounting for Sequential Rollout and Aggregation Across Local Markets

Pradeep K. Chintagunta

Booth School of Business, University of Chicago, Chicago, Illinois 60637,
pradeep.chintagunta@chicagobooth.edu

Shyam Gopinath

Kellogg School of Management, Northwestern University, Evanston, Illinois 60208,
s-gopinath@kellogg.northwestern.edu

Sriram Venkataraman

Goizueta Business School, Emory University, Atlanta, Georgia 30322, svenka2@emory.edu

Our objective in this paper is to measure the impact (valence, volume, and variance) of national online user reviews on designated market area (DMA)-level local geographic box office performance of movies. We account for three complications with analyses that use national-level aggregate box office data: (i) aggregation across heterogeneous markets (spatial aggregation), (ii) serial correlation as a result of sequential release of movies (endogenous rollout), and (iii) serial correlation as a result of other unobserved components that could affect inferences regarding the impact of user reviews. We use daily box office ticket sales data for 148 movies released in the United States during a 16-month period (out of the 874 movies released) along with user review data from the Yahoo! Movies website. The analysis also controls for other possible box office drivers. Our identification strategy rests on our ability to identify plausible instruments for user ratings by exploiting the sequential release of movies across markets—because user reviews can only come from markets where the movie has previously been released, exogenous variables from previous markets would be appropriate instruments in subsequent markets.

In contrast with previous studies that have found that the main driver of box office performance is the volume of reviews, we find that it is the valence that seems to matter and not the volume. Furthermore, ignoring the endogenous rollout decision does not seem to have a big impact on the results from our DMA-level analysis. When we carry out our analysis with aggregated national data, we obtain the same results as those from previous studies, i.e., that volume matters but not the valence. Using various market-level controls in the national data model, we attempt to identify the source of this difference.

By conducting our empirical analysis at the DMA level and accounting for prerelease advertising, we can classify DMAs based on their responsiveness to firm-initiated marketing effort (advertising) and consumer-generated marketing (online word of mouth). A unique feature of our study is that it allows marketing managers to assess a DMA's responsiveness along these two dimensions. The substantive insights can help studios and distributors evaluate their future product rollout strategies. Although our empirical analysis is conducted using motion picture industry data, our approach to addressing the endogeneity of reviews is generalizable to other industry settings where products are sequentially rolled out.

Key words: online word of mouth; sequential new product release; endogeneity; instrumental variables; generalized method of moments; motion pictures

History: Received: February 20, 2009; accepted: February 22, 2010; processed by Carl Mela. Published online in *Articles in Advance* May 27, 2010.

Introduction

The vast and expanding reach of the Internet and social networking sites has facilitated the growth in the importance of online word of mouth (WOM). Recent research suggests that firms should pay attention to such WOM; e.g., Godes and Mayzlin (2004) find that online postings have an impact on the ratings of TV shows, and Liu (2006) and Duan et al. (2008) find

that the volume of user reviews has a positive impact on future box office revenues of movies.

Our focus in this paper is on measuring the impact of online WOM, as reflected in online user reviews, on off-line purchase behavior, specifically the box office performance of movies in the U.S. market. We follow previous research and investigate three measures of online user reviews—the *volume* of reviews (Liu 2006,

Duan et al. 2008), the *valence* of reviews or the mean user rating (Liu 2006, Duan et al. 2008, Chevalier and Mayzlin 2006), and the *variance* in reviews (Godes and Mayzlin 2004). Two aspects of our study distinguish it from previous studies in this area. The first is that we use local geographic market box office data rather than national-level data. (Note that user review data are typically only available at the national aggregate level.¹) Second, we account for the sequential release of movies across geographic markets; i.e., even for a wide-release movie, there are at least a few markets where the movie is released after its national release date. Hence, when a movie is released in a subsequent market, the user review information can only have come from markets where the film was previously released.

Using market-level data is one way of avoiding the standard aggregation problem. However, the presence of aggregate data in conjunction with the sequential release strategy of studios creates several additional complications that can be addressed using market-level data. The first is that it could lead to situations such as the following: suppose a movie is released in market 1 in week t and is released in market 2 in week $t + 1$; it is possible that the national box office increases from t to $t + 1$. However, because most of the user reviews are coming from market 1, it is possible that the average user review valence for the movie declines from period t to $t + 1$ (users in market 1 did not seem to like the movie), whereas the aggregate box office (across markets) actually goes up because of the addition of a new market (an increase in sales from market 2 compensates for the decrease from market 1). This could bias the estimated effect of the average user rating or valence. Such an aggregation problem is not unique to the movie industry and can occur for other product markets as well—e.g., as noted in Bronnenberg and Mela (2004), packaged goods are often sequentially released across geographic markets.

The second complication from aggregation and sequential release stems from studios' strategic choice of the order of release across geographic markets. Studios have information about factors that influence demand in the various geographic markets—factors not observed to researchers interested in understanding the impact of online WOM. If the order of release reflects such knowledge, the error term in any statistical model at the geographic market level relating online user reviews to box office performance will be

correlated across geographic markets. Combined with the sequential release strategy, the studios' strategic choice of sequential release markets implies that the error term in any aggregate national-level model will exhibit serial correlation.

This implication is a problem because in models that use national aggregate time-series data, the volume of user reviews in period $t - 1$ is used as a measure of online WOM influencing box office performance in period t . Now the volume of user reviews in $t - 1$ is correlated with box office performance in $t - 1$, which contains the error term in $t - 1$; based on our aforementioned discussion, this is likely to be correlated with the error term in period t . This implies that a driver variable is correlated with the error term in any aggregate national model. Finding a valid instrument that varies over time for each movie and that is correlated with the user reviews, but not with the box office error term, could be a challenge.

Note that the mere availability of market-level data or in our case, DMA (designated market area)-level data, is not a panacea for addressing the problem of measuring the effects of user reviews. Within a DMA, we will still be unable to exploit the time-series information for a movie if we believe that there is some unobserved component of movie quality that gets revealed over time. This will manifest itself as serial correlation in the DMA-level error term, which we will not be able to accommodate for the same reason as we could not with national-level data. The reason for this serial correlation, however, is different from the reason articulated previously for national data. At the same time, such an unobserved component, if present, adds a third complication to any analysis that uses national-level data.

The presence of an unobserved component implies that our analysis has to be restricted to the opening day box office performance of a movie that is sequentially released across markets, reducing the dimension of our data to movie-market (*after* initial release) combinations. This is because, as previously noted, from the second day onwards, the user reviews will include those from the new market as well and will reflect unobserved quality factors whose presence complicates our analysis. Hence, our measures of user reviews at the time a movie is released in a market will then come from users in markets where the movie was previously released. We would still have to allow for endogeneity of these user reviews because as noted before, error terms will likely be correlated across markets as a result of strategic release. However, in this situation, we can construct plausible instruments for user reviews because any exogenous variable driving box office performance in previously

¹ This may not always be the case. Depending on the source structure (e.g., website format), type of online WOM (e.g., reviews, forums, video clips), and profile settings of the poster, it might be possible to identify a poster's location. For example, YouTube.com allows you to identify video-clip uploaders from a specific geographical region. We thank the editor for this insight.

released markets will be an appropriate instrument for user reviews in subsequent markets.²

Our objective in this paper is to measure the impact (valence, volume, and variance) of online user reviews on DMA-level local geographic box office performance of movies. We account for three complications with analyses that use national-level aggregate box office data: (i) aggregation across heterogeneous markets (spatial aggregation), (ii) serial correlation as a result of sequential release (endogenous rollout), and (iii) serial correlation as a result of other unobserved components (that also preclude the use of DMA-level time-series data).

We use daily box office ticket sales data for 148 movies released in the United States during a 16-month period (out of the 874 movies released) for which we could also collect online user review data from the Yahoo! Movies website. Our analysis controls for other possible box office drivers such as the number of theaters playing the movie, prerelease advertising levels, and the extent of intermovie competition. In contrast with previous studies that have largely found that the main driver of box office performance is the volume of reviews, we find that it is the valence that seems to matter and not the volume. Furthermore, ignoring the endogenous rollout decision does not seem to have a big impact on the results from our DMA-level analysis, which implies that once we control for the various factors as we do in our analysis, there does not appear to be much evidence for the presence of variables that are driving studios' release decisions that we as researchers do not observe. When we carry out our analysis with aggregated national data, we obtain the same results as those from previous studies, i.e., that volume matters but not the valence. Thus the difference between our results and those using national-level data likely stems from a combination of spatial aggregation and serial correlation induced by factors other than endogenous rollout. We take some initial steps to parsing out these explanations. Table 1 describes the various models and analyses we consider in this paper. We focus on cells I, III, and IV. The absence of appropriate instruments at the national level precludes us from examining cell II. Based on our empirical results, we explore some managerial implications for studios.

Data

Our data combine market-level box office information with user ratings of movie goers. We have daily box office ticket sales data on all movies released from November 2003 to February 2005. The online WOM

Table 1 Taxonomy of Models

Conditions	Without endogeneity correction	With endogeneity correction
With aggregation bias	[I] National-level models without endogeneity correction	[II] Not applicable ^a
Without aggregation bias	[III] Disaggregate market-level models without endogeneity correction	[IV] Disaggregate market-level models with endogeneity correction

^aFor reasons discussed in this paper, in the case of a product sequentially rolled out across markets, it is impossible to identify suitable instruments that allow one to correct for rollout endogeneity using national-level demand data.

data for this study were collected from the Yahoo! Movies website (see <http://movies.yahoo.com/>). Specifically, we collected the following for each user review: user ID, review date, overall rating, rating for story, rating for acting, rating for direction, rating for visuals, number of people who found the review helpful, and total number of people who read the reviews. The ratings for the reviews can range from F to A+. These ratings were translated to a numeric range from 1 to 13. We need to collect the user ratings for the entire duration of our data because we need information during the entire sequential release period. Hence our WOM data are for 148 movies in the above time period. Table 2 gives the variable definitions along with their descriptions and key summary statistics. Additional summaries between the online WOM measures are provided in Appendix B of the electronic companion, available as part of the online version that can be found at <http://mktsci.pubs.informs.org>. In the ensuing, we first discuss our online WOM data and then the box office data.

From the user rating data collected from the Yahoo! Movies website, we came up with the three measures of online WOM—valence, variance, and volume. These measures are described in Table 2. The key measure of online WOM used in our analysis is *valence* of the ratings—in our case, this corresponds to the mean rating of a movie at the time the movie is released in a local market, i.e., based on feedback from moviegoers in other markets where the movie has been released before that time. Previous research in contexts other than movies, e.g., Chevalier and Mayzlin (2006) in the case of books, underscores the importance of the valence metric.

One can also argue that the interaction between mean user rating and the volume of ratings might be important in quantifying the online WOM effect. Specifically, the more raters there are that view the movie as being good, the bigger is the impact on subsequent market box office performance. This is

²Note that in the context of high-end products where people may cross market boundaries to purchase and consume the product, the situation can be far more complex.

Table 2 Descriptive Statistics

Variable	Description	Min	Max	Mean	Std. dev.
<i>AVGRATING</i>	Mean rating of reviews until the movie is released in a new market (1–13)	5.27	12.36	9.90	1.44
<i>PRERATING</i>	Inverse of the variance in ratings until the movie is released in a new market	0.04	100.00	13.96	33.69
<i>VOLRATING</i>	Cumulative volume of reviews until the movie is released in a new market	1.00	17,519.00	474.82	1,483.57
<i>AVGADV</i>	Average per-day advertising until the movie is released in a new market (in 1,000 dollars)	72.57	33,344.70	1,783.81	3,283.57
<i>OPENINGGROSS</i>	Total opening earnings (in dollars)	5.00	201,393.60	4,066.70	12,667.45
<i>THEATERS</i>	Total number of theaters where the movie is simultaneously playing	1.00	37.00	2.25	3.54
<i>AVGSTARS COMP</i>	Average star power of competing movies in the new market (0–6)	0	5.00	2.06	0.35
<i>AVGREVIEWCOMP</i>	Average critic score of competing movies in the new market (0–100)	0	94.00	57.71	9.21
<i>DAYS</i>	Days since initial release	1.00	392.00	28.85	56.53
Number of observations		3,766			
Number of movies		148			
Number of markets (DMAs)		253			

akin to the interaction effect discussed in Nam et al. (2010) in the context of a video-on-demand service. Furthermore, Sun (2008) makes the argument for including interactions between the mean and variance of ratings.

In addition to the online WOM data, we have detailed box office data collected from multiple sources, both public and private. We have theater-level box office data for each movie title for the period from November 2003 to February 2005. The data we use for the analysis include 3,830 exhibitor locations belonging to 631 exhibitor chains (includes independent theaters) and 148 movie titles. The data are aggregated to the ACNielsen-defined DMAs. Opening day gross for a title in a local geographic market is our dependent variable. The correlation between market-level opening day gross and overall gross is 0.69 across all movies (minimum correlation of 0.60; maximum correlation of 0.89). This indicates that the opening day gross is in fact highly correlated with the overall movie revenues. We also include theater count (e.g., Neelamegham and Chintagunta 1999), prerelease national advertising (e.g., Moul 2007, Ainslie et al. 2005), days since initial release (e.g., Elberse and Eliashberg 2003), and two competition variables (Ainslie et al. 2005 emphasize the importance of competition). Table 2 gives the summary statistics and the definitions for these variables.

Methodology

The unit of analysis is a movie (*i*)–market (*j*) combination (*ij*). The revenues (*OPENINGGROSS*³) on the

opening day for movie *i* in market *j* is given by

$$\begin{aligned}
 \log(\text{OPENINGGROSS}_{ij}) &= \alpha + \gamma_1 \text{AVGRATING}_{ij} + \gamma_2 \text{VOLRATING}_{ij} \\
 &\quad + \gamma_3 \text{PRERATING}_{ij} + \gamma_4 \text{AVGRATING} \\
 &\quad \times \text{VOLRATING} + \gamma_5 \text{AVGRATING} \\
 &\quad \times \text{PRERATING} + \eta \text{DAYS}_{ij} + \lambda \text{AVGADV}_{ij} \\
 &\quad + \beta \text{THEATERS}_{ij} + \theta_1 \text{AVGSTARS COMP}_{ij} \\
 &\quad + \theta_2 \text{AVGREVIEWCOMP}_{ij} + \sigma_1 S_1 + \sigma_2 S_2 \\
 &\quad + \sigma_3 S_3 + \mu_i + \rho_j + \varepsilon_{ij}.
 \end{aligned} \tag{1}$$

On the right-hand side of Equation (1) we have the following variables. *AVGRATING*, *VOLRATING*, and *PRERATING* are the valence, volume, and precision (inverse of variance), respectively, for movie *i* at the time when the movie is released in market *j*; i.e., these are user reviews from markets where the movie was released prior to market *j*. Then we have the two interactions involving the valence—one with the volume and one with the precision. The remaining variables are largely “control” variables. The variable *DAYS* denotes the number of days since the movie’s first release at the time the movie is released in market *j*. This variable could partially account for the unobservable that might be influencing the release sequence for a movie although we are estimating a

count (*THEATERS*). We thank the area editor for this alternative specification. Here, we present the model shown in Equation (1) to be consistent with prior literature on movie box office forecasting and on online WOM and movies because comparing our results with those studies is one key objective of this research.

³ We also considered a model with opening gross per theater as the dependent variable. In this model we no longer would have theater

common effect across movies here. The model also accounts for other factors driving box office revenue—the number of theaters (*THEATERS*) in which the movie is released in market j and the average advertising expenditure until release date (*AVGADV*); both of these variables have been discussed in Table 2. In addition, we also include two variables to account for competition—the average star power of competing movies (*AVGSTARSCOMP*) and the average critic score of competing movies (*AVGREVIEWCOMP*). We include three seasonal dummies to account for seasonal variations in box office performance. In addition, we include dummy variables for holidays such as Christmas and the Fourth of July.

To account for movie and market (DMA) differences in terms of their characteristics (movie genre, rating, stars, critics evaluations, market size, population, etc.), as well as for other unobserved characteristics, we include movie (μ_i) and market (ρ_j) fixed effects. Although movie dummies have been included in past research (e.g., Duan et al. 2008, Neelamegham and Chintagunta 1999), DMA or market dummies have been difficult to control for because of the non-availability of market-level data. We also looked at the effect of including market fixed effects \times movie characteristics interactions and movie fixed effects \times market demographics interactions in the model. Including these variables did not change our results materially.

Endogeneity Concerns

As mentioned earlier, because of the sequential roll-out decisions of studios, the error term (ε_{ij}) in Equation (1) will be correlated across markets. Now because user ratings from a given market are correlated with the error term in that market, and because error terms are correlated across markets, the error term in a subsequent market will be correlated with the user ratings obtained from previously released markets. Not accounting for this correlation would lead to biased estimates for the effects of our online WOM measures. What would be valid instruments for these user ratings? Because user ratings at the time of release in market j depend on box office performance of movie i in markets prior to j , exogenous factors that drive the box office in those markets are uncorrelated with the error terms in those previously released markets and hence are uncorrelated with the error term in market j . Consequently, exogenous factors driving box office in previously released markets would be valid instruments for user ratings in market j . For the average rating and volume measures (*AVGRATING* and *VOLRATING*), we use competitive controls in earlier markets as instruments for user ratings in market j : we use the average critic score of competing movies in markets prior to market j (*COMPVIEW* _{ij}), the average star power of competing movies in those earlier markets (*COMPSTAR* _{ij}),

and the average proportion of movies of the same genre as the focal movie playing in the previous markets from which the user ratings are generated (*PROPGEN* _{ij}). Why are these appropriate instruments for the valence of the user ratings? First, note that these variables are functions of the competition variables from all markets k where the movie is released prior to its release in market j . Now the average and volume of ratings for movie i in previously released markets (i.e., prior to market j)—our endogenous variables—are likely to be higher (lower) if the quality of the competing movies in those markets (as represented by the competition variables) is lower (higher). Furthermore, because these competition variables in any given market are assumed to be uncorrelated with the unobservables in that market k (i.e., *AVGSTARSCOMP* and *AVGREVIEWCOMP* in Equation (1)) they are also uncorrelated with the unobservables in the focal market j (where release occurs *after* market k). It follows then that the instruments that are derived from these competition variables (from markets k prior to j) are also uncorrelated with the unobservables in the focal market j .

Thus the validity of our instruments hinges on the validity of our assumption that the variables *AVGSTARSCOMP* and *AVGREVIEWCOMP* are exogenous to markets k . Note that we are taking the averages in *AVGSTARSCOMP* and *AVGREVIEWCOMP* across *all* competing movies playing in markets k , not just the ones that were released at the time that the focal movie i was released (or just the subset of movies included in our sample, for that matter). Indeed, the number of movies that are actually released at the same time as the focal movie i is only a small fraction of the total number of movies we are averaging. The exogeneity of *AVGSTARSCOMP* and *AVGREVIEWCOMP* would come into question if studios are able to fully endogenize the release of all their movies, across all markets, over a very long time horizon while explicitly accounting for the strategic entry and exit decisions of each and every movie of each and every rival studio across each and every market over that same time horizon. Under these circumstances, it is reasonable to assume that *AVGSTARSCOMP* and *AVGREVIEWCOMP* are exogenous to the markets k , whence it follows that *COMPSTAR* and *COMPVIEW* are exogenous to the markets j . In our analysis, the instruments are constructed based on *all* previous time periods since initial release. Another option is to use information only from periods prior to the most recent introduction. However, we did not assess the sensitivity of our results to these alternative definitions because doing so would pose difficulties when computing instruments for, e.g., second-release markets.

As an additional instrument for *VOLRATING*, we use the number of days since the movie's release

(until the time of release in the focal market) that the rain and snow levels in the markets where the movie was released were worse than the average level of rain and snow in those markets. This instrument is likely to lower the number of moviegoers (because of worse-than-average weather conditions) and consequently the volume of user ratings from those markets. For example, *Yes Man* and *Seven Pounds* had to be content with modest opening weekend performances because of bad weather (Cieply 2008). It is less clear what an appropriate instrument would be for the variance of ratings because the potential source of correlation between the variance and the error term is not known. Although we could use instruments such as the average variance of movies of a similar genre to the focal movie that were previously released in the markets where the focal movie was previously released—and doing so does not alter our conclusions—we only report the results for the non-instrumented variance measure here and interpret its effects with some caution.

Another endogeneity concern involves the number of theaters variable in Equation (1). It is reasonable to assume that studios adjust the number of theaters or screens in each market based on knowledge regarding the unobservable in that market (Elberse and Eliashberg 2003). As an instrument for the number of theaters, we use the average number of theaters that show movies of the same genre as the focal movie in the focal market in the period prior to the movie's release. The intuition is that the number of theaters is likely to be higher for movies of a certain genre when opening in markets that might have a higher preference for that movie without being correlated with the demand shock for the specific movie under consideration. Another potentially endogenous variable is advertising. We note that (a) prerelease advertising accounts for a vast majority (e.g., Elberse and Anand 2007 find that 88% of television advertising was spent prior to initial release) of advertising spending in the movie industry, (b) prerelease advertising budgets are typically a fixed proportion of the product budget (see Vogel 2007), and (c) prerelease advertising only varies across movies and not markets, and we include movie fixed effects. Together, these factors alleviate concerns regarding the endogeneity of prerelease advertising.

Estimation

Although the availability of instruments enables us to address the correlation between user review measures and the error term, we still have the issue of the correlation in the error term for a film across the markets in which the movie is sequentially released. We use a generalized method of moments (GMM) procedure that accommodates such correlations while also allowing us to explicitly deal with the endogeneity of user ratings as previously described. For more

details, see Greene (2003). The different steps in the estimation procedure are as follows. (i) Estimate the regression model (Equation (1)) using standard instrumental variables methods. (ii) Use the residuals from the aforementioned regression to obtain the optimal GMM weighting matrix. Because we have an overidentified case (the number of instruments is greater than the number of endogenous variables), we use the inverse of the asymptotic covariance matrix (\hat{V}) as the optimal weighting matrix (W). (iii) Allow for heteroskedasticity and correlation between error terms for different movie–market combinations for every movie:

$$\hat{V} = \frac{1}{N} \left[\sum_{f=1}^F \left[\sum_{i=1}^{n_f} e_i^2 z_i z_i' + \sum_{l=1}^{n_f-1} \sum_{t=l+1}^{n_f} \left(1 - \frac{l}{(n_f+1)} \right) e_l e_{t-l} (z_t z_{t-l}' + z_{t-l} z_t') \right] \right],$$

$$W = \hat{V}^{-1},$$

where \hat{V} is the asymptotic covariance of the moment vector; W is the optimal weighting matrix; N is the number of observations; F is the number of movies; n_f is the number of observations for movie f , i.e., the number of subsequent markets in which that movie f is released; e is the vector of residuals from the instrumental variables regression; and z is the vector of instruments.

Heteroskedasticity (White 1980) is captured by the first term on the right-hand side and the second term accounts for the correlation in movie errors across markets akin to a Newey and West (1987) autocorrelation consistent covariance estimator. The double summation sign in this latter term allows for the errors for a particular movie in a given market to be correlated with the errors in all subsequent markets. The GMM estimator and its asymptotic variance are

$$\hat{\beta}_{\text{GMM}} = [(X'Z)W(Z'X)]^{-1}(X'Z)W(Z'y),$$

$$V(\hat{\beta}_{\text{GMM}}) = [(X'Z)W(Z'X)]^{-1},$$

where X is the matrix of regressors in the model in Equation (1), Z is the matrix of instruments, and y is the dependent variable, i.e., $\log(\text{OPENINGGROSS})$. In our empirical analysis, we will examine the impact of accounting only for the correlation across markets (and not for heteroskedasticity) as well as accounting for both factors.

Empirical Results

We describe our empirical results as follows. First, we discuss the results from the regression of opening day box office in each market on the three online WOM

measures (*AVGRATING*, *PRERATING*, *VOLRATING*) and their interactions ($AVGRATING \times VOLRATING$, $AVGRATING \times PRERATING$). This model falls under the class of models in cell III of Table 1. This regression controls for factors noted previously, e.g., advertising level, number of theaters, competitive factors, movie and market fixed effects, etc., and for endogeneity in number of theaters. We do not report the fixed effects in Tables 3–5 because of the large number of these parameters. Then we provide results from instrumental variables estimation where we account for the endogeneity of user ratings (*AVGRATING* and *VOLRATING*) but not for the non-IID (independent and identically distributed) nature of the error term. Finally, we discuss the results from the GMM estimation of the model parameters, i.e., cell IV of Table 1. The parameter results are reported in Table 3.

Focusing on the “No endogeneity correction for online WOM measures” column in Table 3, the results indicate that all the significant parameters are of the expected sign. Most important, we find that average user ratings (valence) have a positive and statistically significant impact on box office revenues. None of the other measures of online WOM (volume and variance) seems to influence the dependent measure of interest. Thus after controlling for advertising, distribution,⁴ competition, age, and movie and market fixed effects, only the valence of the user ratings seems to influence box office performance. This is an interesting finding in so much that most previous studies have found only the volume of posts or ratings to have a significant impact on box office revenues. What is particularly interesting about our finding, we believe, is that the significant impact is seen after controlling for both movie fixed effects as well as for market fixed effects. This means that the user ratings contain information over and above the “average quality” of the film, as reflected in the coefficient of that movie’s dummy variable.

Investigating the Extent of Endogeneity Bias. As noted previously, if the user ratings are correlated with movie–market-specific unobservables, then the estimated effects for the user ratings variable would not be valid. Thus we now discuss the impact of instrumenting for these user ratings in the aforementioned regression. The estimation results are presented in Table 3 under the column “Instrumental variables correction for online WOM measures.” We interact our instruments with the movie fixed effects to allow them to have movie-specific effects. From

the first-stage regression of user ratings on the instruments as well as on the other included exogenous factors (the various dummy variables, theaters, advertising, etc.), we find that the instruments as well as the exogenous variables explain most of the variation in the endogenous variable. This is not very surprising given the inclusion of movie and market dummy variables in our analysis. It is important to note that the competitive instruments have a significant negative impact on the average and volume of user ratings, whereas the weather variable affects the volume of the ratings. Next, we use an *F*-test to examine whether the proposed instruments jointly explain *AVGRATING* and *VOLRATING*; the results indicate that the instruments cannot be jointly excluded from the first-stage models (P -value < 0.001 and P -value < 0.001 , respectively). Furthermore, the P -values for the overidentifying restrictions test, Hansen’s *J*-test, are all greater than 0.1. This suggests that the instruments appear to be orthogonal to the error term in the box office equation.

From Table 3 we find that even after instrumenting the potentially endogenous online WOM variables, the impact of average user ratings on box office revenues is still statistically significant at the 5% level of significance. Moreover, the effects of volume and variance continue to not have an impact on box office performance. In other words, even after accounting for endogeneity, online WOM appears to be playing a role in influencing box office performance of a movie via how much users like the movie on which they are commenting. The results in Table 3 also indicate that instrumenting for user ratings has little or no impact on the effects of the other variables in the regression.

Finally, we look at the impact of allowing for non-IID error terms in the regression. As we have described previously, we accomplish this via GMM estimation with a Newey–West (see Newey and West 1987, Andrews 1991) formulation for the weight matrix. The key takeaway from Table 3 is that accounting for correlation across markets in the error term and for heteroskedastic errors does little to change our conclusions based on the analysis described above. Thus the effect of valence remains robust to these effects as well.

Although not reported here, we also examined the robustness of our results to whether a movie is considered a wide or limited release. A movie is a wide release if it is released at 600 or more theaters and a limited release if it is released at less than 600 theaters (*Box Office Mojo*). We included a dummy variable for wide release and interacted it with the valence measure. We find that the main effect of valence of user ratings is still positive and statistically significant. Although the interaction effect is negative, it is not significant underscoring the robustness of our

⁴ Distribution, i.e., *THEATERS*, is instrumented in the results reported in Table 2. However, the nature of results is similar even without instrumenting *THEATERS* (e.g., see Table 3). The *F*-test results indicate that the instruments cannot be jointly excluded from the first-stage model (P -value < 0.001).

Table 3 Impact of Measures of Online WOM on Opening Box Office Earnings

	DV: log(<i>OPENINGGROSS</i>)			
	No endogeneity correction for online WOM measures	Instrumental variables correction for online WOM measures	GMM autocorrelation	GMM heteroskedasticity + Autocorrelation
<i>AVGRATING</i>	0.1506*** (0.0470)	0.0995** (0.0504)	0.0956** (0.0470)	0.0936** (0.0434)
<i>VOLRATING</i>	−0.0002 (0.0001)	−6.7e−05 (0.0002)	−6.9e−05 (0.0002)	−0.0002 (0.0002)
<i>PRERATING</i>	−0.0348 (0.0211)	−0.0289 (0.0245)	−0.0285 (0.0232)	−0.0052 (0.0204)
<i>AVGRATING</i> × <i>VOLRATING</i>	4e−06 (1.4e−05)	−7e−06 (1.5e−05)	−6e−06 (1.4e−05)	2e−06 (1.3e−05)
<i>AVGRATING</i> × <i>PRERATING</i>	0.0028 (0.0018)	0.0024 (0.0020)	0.0024 (0.0019)	0.0004 (0.0017)
<i>AVGADV</i>	1.7e−05 (8e−06)	1.9e−05*** (7e−06)	1.9e−05*** (7e−06)	2e−05*** (5e−06)
<i>THEATERS</i>	0.1384*** (0.0542)	0.1257*** (0.0090)	0.1258*** (0.0091)	0.1301*** (0.0095)
<i>AVGSTARS COMP</i>	−0.2463*** (0.0887)	−0.2370*** (0.0814)	−0.2363*** (0.0768)	−0.2594*** (0.0587)
<i>AVGREVIEW COMP</i>	0.0023 (0.0032)	0.0019 (0.0028)	0.0019 (0.0026)	0.0032 (0.0024)
<i>DAYS</i>	−0.0250*** (0.0018)	−0.0254*** (0.0014)	−0.0255*** (0.0013)	−0.0253*** (0.0013)
Number of observations			3,766	

Notes. Coefficients of movie and market fixed effects are not reported. These findings hold even when the number of theaters is not instrumented.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

online WOM finding. The details of this analysis are provided in Appendix A of the electronic companion.

Investigating the Extent of Aggregation Bias. We now investigate the extent of aggregation bias by conducting a national-level analysis of the type used by previous researchers, i.e., cells I and II in Table 1. In the introduction, we argued that the absence of exogenous *national*-level factors driving user ratings precluded us from commenting on cell II; thus our focus is on cell I. Also, earlier in this section of the paper, we argued that one of the reasons for the findings reported in previous studies is the aggregation across markets, which masks the true marginal impact of the user ratings (aggregation bias). One way of verifying this line of reasoning is to actually aggregate the data across markets to obtain “national”-level data and then run the analysis as in previous studies. In Table 4, we provide the results of such an analysis. From our data, we first create an aggregate time series of daily box office, mean user ratings, the volume of ratings, the precision of the ratings, and other variables all at the national level.⁵ Then, we regress the aggregate box office on the measures of online WOM as well as on other control variables.

⁵ In other words, this analysis uses more information than we do in cells III and IV because there we do not use the time-series information from any given market.

Table 4 provides two sets of aggregate regressions. One set (represented by models 3 and 4) uses online WOM measures similar to those in our previous analyses (Table 3)—the mean user rating, precision, and cumulative volume of ratings across all raters until the previous time period. The other set, represented by model 5, uses the same measures as those in Liu (2006), Duan et al. (2008), and Sun (2008), which are mean user ratings, the precision of the user ratings, and the volume of the user rating from only the previous time period. The models based on aggregate data do not instrument for *THEATERS* as well because time-varying exogenous instruments for *THEATERS* are not readily available, and we cannot use a lagged box office as an instrument because of serial correlation in the aggregate error term described previously. To be consistent, models 1 and 2 also do not instrument *THEATERS*.

Table 4 shows that, consistent with the previous literature, the mean user ratings variable is not statistically significant in any of the aggregate regression models. When we use the previous period’s ratings instead of the cumulative measures, we obtain results for the volume of ratings and the precision of ratings that are consistent with the previous literature (Liu 2006, Duan et al. 2008 for the volume, Sun 2008 for the precision/variance). These results therefore reinforce the argument made earlier about the potential pitfalls

Table 4 Comparison of National-Level and Market-Level Analyses

	DV: $\log(GROSS)$				
	Market-level models ^a (OLS)		National-level models ^b (OLS)		
	Model 1	Model 2	Model 3	Model 4	Model 5
<i>AVGRATING</i>	0.1415*** (0.0413)	0.1346*** (0.0423)	0.0231 (0.0158)	0.0118 (0.0168)	
<i>VOLRATING</i>		−0.0002 (0.0002)		−0.0006*** (0.0001)	
<i>PRERATING</i>		−0.0257 (0.0179)		0.0007 (0.0033)	
<i>AVGRATING</i> × <i>VOLRATING</i>		3e−06 (1.4e−05)		0.00005*** (0.00001)	
<i>AVGRATING</i> × <i>PRERATING</i>		0.0021 (0.0015)		−0.0004 (0.0003)	
<i>AVGRATING</i> _{DAILY}					−0.0062 (0.0054)
<i>VOLRATING</i> _{DAILY}					0.0013*** (0.0004)
<i>PRERATING</i> _{DAILY}					−0.0039*** (0.0010)
<i>AVGRATING</i> _{DAILY} × <i>VOLRATING</i> _{DAILY}					−0.0001 (0.0001)
<i>AVGRATING</i> _{DAILY} × <i>PRERATING</i> _{DAILY}					0.00009 (0.0001)
<i>AVGADV</i>	2.5e−05*** (7e−06)	1.9e−05*** (7e−06)	0.0001*** (0.00001)	0.0001*** (0.00002)	0.0001*** (0.00002)
<i>THEATERS</i>	0.0936*** (0.0039)	0.0922*** (0.0039)	0.0010*** (0.00002)	0.0011*** (0.00002)	0.0010*** (0.00002)
<i>AVGSTARS</i> <i>COMP</i>	−0.1937** (0.0806)	−0.2174*** (0.0803)	0.2814 (0.1749)	0.1146 (0.1742)	0.1282 (0.1734)
<i>AVGREVIEW</i> <i>COMP</i>	0.0012 (0.0028)	0.0009 (0.0028)	0.0042 (0.0040)	0.0032 (0.0039)	0.0039 (0.0038)
<i>DAYS</i>	−0.0268*** (0.0013)	−0.0239*** (0.0013)	−0.0408*** (0.0015)	−0.0388*** (0.0016)	−0.0398*** (0.0016)
Adj. R^2	0.9892	0.9894	0.9344	0.9361	0.9376

Notes. *THEATERS* is not to be instrumented so as to be consistent with other national-level models used in prior online WOM literature on movies. OLS, ordinary least squares.

^aCoefficients of movie and market fixed effects are not reported.

^bCoefficients of movie and day-of-the-week fixed effects are not reported.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

of using national-level data for these analyses. Hence even though our data are from a different time period and represent a different set of movies, we arrive at the same conclusions as previous researchers.

From both these analyses, we find that the coefficient of average user rating, our key variable of interest, is influenced by both endogeneity and aggregation biases. However, aggregation bias seems to be more critical of the two biases for the movie industry in which a movie is sequentially rolled out into different markets. Specifically, we find that failure to account for aggregation bias results in the average user rating becoming statistically insignificant. This could be one reason why past research that uses national-level box office data (e.g., Liu 2006, Duan et al. 2008) does not find valence to be an important predictor of future box office performance. How-

ever, as noted before, the difference can also stem from the use of the time-series information in the national models. We explore this further in the following subsection.

National-Level Aggregate Models with Market-Level Controls. Here, we fit different aggregate models that take into account market-level differences to better understand the difference in results between the aggregate national-level model and the disaggregate market-level model. The final set of models is in Table 5.

• *Models 1a and 1b.* These are models 1 and 2, respectively, from Table 4, i.e., the results obtained from our market-level analysis but without any endogeneity corrections. These are provided for purposes of comparison with the other models.

Table 5 Comparison of National-Level and Market-Level Analyses

	DV: log(<i>GROSS</i>)							
	Market-level models ^a (OLS)		National-level models ^b (OLS)					
	Model 1a	Model 1b	Model 2a	Model 2b	Model 2c	Model 2d	Model 2e	Model 2f
#NEW MARKETS					0.0116*** (0.0018)	0.0117*** (0.0017)	0.0123*** (0.0019)	0.0119*** (0.0018)
#OLD MARKETS					0.0028*** (0.0004)	0.0025*** (0.0005)	0.0023*** (0.0006)	0.0028*** (0.0007)
AVGRATING	0.1415*** (0.0413)	0.1346*** (0.0423)	0.0231 (0.0158)	0.0118 (0.0168)	0.0529** (0.0245)	0.0574** (0.0244)	0.1047*** (0.0397)	0.1084*** (0.0406)
VOLRATING		−0.0002 (0.0002)		−0.0006*** (0.0001)		−0.00007 (0.00005)		−0.00006 (0.00008)
PRERATING		−0.0257 (0.0179)		0.0007 (0.0033)		0.0015 (0.0013)		0.0017 (0.0021)
AVGRATING × VOLRATING		3e−06 (1.4e−05)		0.00005*** (0.00001)		0.0001*** (0.00002)		−0.00001 (0.00001)
AVGRATING × PRERATING		0.0021 (0.0015)		−0.0004 (0.0003)		−0.0002*** (0.00005)		−0.0004 (0.0003)
AVGADV	2.5e−05*** (7e−06)	1.9e−05*** (7e−06)	0.0001*** (0.00001)	0.0001*** (0.00002)	0.0002*** (0.00003)	0.0002*** (0.00003)	0.0001** (0.00006)	0.0001** (0.00006)
THEATERS	0.0936*** (0.0039)	0.0922*** (0.0039)	0.0010*** (0.00002)	0.0011*** (0.00002)	0.0008*** (0.00004)	0.0009*** (0.00004)	0.0008*** (0.00005)	0.0008*** (0.00006)
AVGSTARS COMP	−0.1937** (0.0806)	−0.2174*** (0.0803)	0.2814 (0.1749)	0.1146 (0.1742)	0.4474 (0.2837)	0.3234 (0.2821)	0.3454 (0.3854)	0.2822 (0.3857)
AVGREVIEW COMP	0.0012 (0.0028)	0.0009 (0.0028)	0.0042 (0.0040)	0.0032 (0.0039)	−0.0075 (0.0063)	−0.005 (0.0063)	0.0035 (0.0085)	0.0033 (0.0086)
DAYS	−0.0268*** (0.0013)	−0.0239*** (0.0013)	−0.0408*** (0.0015)	−0.0388*** (0.0028)	−0.0403*** (0.0016)	−0.0352*** (0.0026)	−0.0442*** (0.0038)	−0.0387*** (0.0043)

Notes. *THEATERS* is not to be instrumented so as to be consistent with other national-level models used in prior online WOM literature on movies. We get similar results when we replace the number of new and old markets with the number of new and old theaters added, respectively.

^aCoefficients of movie and market fixed effects are not reported.

^bCoefficients of movie and day-of-the-week fixed effects are not reported.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

- *Models 2a and 2b.* These are the base models (i.e., similar to the models used in prior literature on online WOM and movies) using aggregate national data and correspond to models 3 and 4, respectively, from Table 4. These models do not have any market-level controls. The difference between 2a and 2b is the former only includes the volume and valence measures, whereas the latter also includes the precision measure and various interactions.

- *Models 2c and 2d.* These are the base models with two market-level controls—#NEW MARKETS, i.e., the number of markets in which the movie has been released on that particular day; and #OLD MARKETS, i.e., the number of existing markets in which the movie is continuing to play on that particular day (this accounts for any dropping-out of previously released markets). We feel that with some effort such data may be available for analysis because it does not require market-level box office information per se.

- *Models 2e and 2f.* These models also have the two market level controls like models 2c and 2d. However, the key difference is that the time series considered

is restricted. In other words, only days in which the movie was added to a market or the movie was removed from an market were considered. The reason for such restricted data is to reduce, to some extent, the potential confounding as a result of the inclusion of the entire time series of observations.

The results of this analysis are shown in Table 5. Interestingly, we find that the valence of user reviews (*AVGRATING*) is significant once we control for market-level differences (models 2c–2f). Furthermore, we now find a lack of significance of the volume (*VOLRATING*) and precision (*PRERATING*) variables. In addition, the coefficients for the two market-level variables are significant. Note that these effects are found after controlling for the effects of the number of screens—the typical control used in previous research. What this tells us is that it is not simply the total number of screens but whether these are “new markets” or “old markets” that are playing the movie.

What do these results imply? First, they demonstrate the importance of accounting for new markets and previous markets when studying national-level

box office performance. This appears to reverse the nonsignificant finding for the sentiment contained in user reviews when looking at only national-level data. Recall from the introduction the example where aggregate box office data across markets increased from week t to week $t + 1$ as a result of the addition of new markets, although the average user review valence for the movie declined from period t to $t + 1$ as users in the markets where the movie played at time t did not like the movie. This could bias the estimated effect of the average user rating or valence. By explicitly including the number of new markets as a control variable in the regression, any increase in box office as a result of the addition of the new markets will be accounted for via this control variable, thereby allowing the “true” effect of the user ratings to be estimated.

Second, our analysis also underscores the role of the time series in confounding the results. By focusing only on those time points when a movie is released in a new market or dropped from an existing market, the effect of the valence measure almost doubles in magnitude and is now more similar to what we find in the market-level analysis that did not instrument for user ratings (0.1346 from the market-level analysis and 0.1084 from the aggregate analysis with market controls).

At the same time, an important caveat to this analysis is that the effects are obtained after including two potentially endogenous variables into the analysis (i.e., number of new and existing markets). Again, it is not clear what the appropriate instruments for these variables might be because these are strategic variables for the studios, and so unlike our market-level analysis presented in this paper, we cannot fully account for their potential endogeneity. Nevertheless, the analysis enables us to shed some light on the mechanism that appears to be causing the bias with aggregate data.

Managerial Implications

Now that we have demonstrated the robustness of the effect of advertising and valence of user ratings on box office performance, we extend our analysis

to the market level. Because previous models have been estimated using national-level aggregate data, little is known about how consumers in different parts of the United States differ in terms of their sensitivities to online WOM effects. To advance managerial understanding of spatial variation in WOM effects, we include market dummy variable interactions with the average user rating and advertising variables in Equation (1). Then, based on a median split of the estimated effects, we are able to classify markets based on whether they are high (low) impact in terms of the effects of firm actions (advertising) and high (low) in terms of the impact of consumer actions (online WOM, or OWOM). Such an analysis may provide insights to movie distributors and studios that can help them evaluate their current and future rollout strategies. Table 6 provides examples of markets that fall into the four “quadrants.” So how can studios use the outputs of such an analysis?

The market-specific responsiveness to online WOM and advertising allows us to compare and contrast release strategies across movies, movie genres, distributors, etc. It is important to note that the market selection problem is inherently dynamic in nature. One contribution of the market-level analysis is that studios can use the information to appropriately classify markets based on their responsiveness to user reviews. Once studios have this market-level classification, they can choose the appropriate set of markets for future movies depending on the initial buzz generated for the movie. In Table 7, we report the rollout sequence of three titles in the comedy genre: *Along Came Polly*, *13 Going on 30*, and *50 First Dates*. Comparing the performance of these titles, we can offer the following insights for movie distributors.

INSIGHT 1. When the early WOM reviews for a title are unfavorable, delay its release in “consumer in control” markets.

From Table 7, it is clear that *Along Came Polly* does not follow this strategy but instead releases in all the markets in the second-release phase. This would have been a fruitful strategy had the buzz generated been positive, as was the case for *50 First Dates*. Both of

Table 6 Market-Specific Responsiveness—Definitions and Examples of Markets

Quadrant	Definitions	Proportion of markets	No. of markets	Examples
Neither in control	Low advertising coefficient + Low OWOM coefficient	0.22	46	Charlotte, NC; Hartford, CT; Norfolk, VA
Consumer in control	Low advertising coefficient + High OWOM coefficient	0.17	35	Atlanta, GA; Chicago, IL; Washington, DC
Firm in control	High advertising coefficient + Low OWOM coefficient	0.23	47	Charleston, KY; Johnstown, PA; Tulsa, OK
Both in control	High advertising coefficient + High OWOM coefficient	0.38	78	Honolulu, HI; Roanoke, VA; Springfield, MA

Table 7 Comparison of Release Strategies of Three Movies of Comedy Genre

Release phase	Responsiveness quadrant	No. of markets		
		<i>Along Came Polly</i>	<i>13 Going on 30</i>	<i>50 First Dates</i>
1	Neither in control		1	
	Consumer in control	2		
	Firm in control			
2	Both in control			1
	Neither in control	43		3
	Consumer in control	31	1	2
3	Firm in control	46		
	Both in control	78		1
	Neither in control		1	40
4	Consumer in control	1		33
	Firm in control			47
	Both in control			76
5	Neither in control	1	26	1
	Consumer in control	1	24	
	Firm in control		18	
6	Both in control		38	
	Neither in control			1
	Consumer in control			
7	Firm in control			
	Both in control		1	
	Neither in control		5	
Total	Consumer in control		1	
	Firm in control		10	
	Both in control		14	
Total	Neither in control		12	
	Consumer in control		9	
	Firm in control		19	
	Both in control		25	
Total		203	205	205

these movies are of the same genre and have similar rollout strategies in the same the set of markets. The key differentiating factor is that one movie (*50 First Dates*) received positive early WOM, whereas the other movie (*Along Came Polly*) had a negative online WOM. This could be one of the reasons why *50 First Dates* grossed more than *Along Came Polly* in the domestic market.

INSIGHT 2. When the early WOM reviews for a title are favorable, release in consumer in control markets as early as possible.

Unfortunately, *13 Going on 30* was released in a substantial number of consumer in control markets in the later phases (6 and 7). Thus the movie was unable to capitalize on the positive early online WOM, unlike *50 First Dates*. The less-than-optimal release strategy coupled with the low demand in the later-released markets as a result of the age of the movie resulted in the overall gross of *13 Going on 30* being less than half that of *50 First Dates*.

INSIGHT 3. Irrespective of the nature of online WOM, it might be helpful to release the movie in “neither in control” markets in the last phase of the movie rollout.

For example, if *13 Going on 30* had switched the 19 consumer in control markets in phases 6 and 7 with 19 neither in control markets in phase 3, the studio might have generated more revenue. Note that the aforementioned analysis does not account for various other factors that might influence studios’ decisions and should hence be interpreted in that light.

Discussion and Conclusions

The objective of this paper was to investigate the role of online word of mouth in influencing the off-line sales of sequentially released new products. We lay out potential difficulties in trying to measure this effect with national-level data and discuss how having access to local market data alleviates some issues with the measurement but raises other issues that need to be addressed as well. We then provide an empirical application for our proposed generalized method of moments-based approach to measuring the impact of the valence, volume, and variance of online user ratings from Yahoo! Movies on box office performance while accounting for both these types of biases. Our identification strategy rests on our ability to identify plausible instruments for user ratings by exploiting the sequential release of movies across markets—because user reviews can only come from markets in which the movie has previously been released, exogenous variables from previous markets would be appropriate instruments in subsequent markets.

We find that the valence of online WOM (mean user rating) has a significant and positive impact on box office earnings. The importance of the valence of online WOM has been found by past research in other contexts as well. Chevalier and Mayzlin (2006) find that negative reviews have a greater impact than positive reviews. In a related study, Shin et al. (2008) find that the level of purchase involvement acts as a moderator for the impact of both positive and negative WOM on the prices charged. Specifically, they find that negative buzz leads to price cuts for high-ticket items, whereas positive buzz enables price increases for low-priced items in websites. The adverse impact of negative online WOM has also been found in non-website-specific studies as well. For instance, Luo (2007) finds that consumer complaints in the airlines industry have a detrimental impact on firm’s future stock returns. Finally, in the context of movies, Dellarocas et al. (2007) find that volume, valence, and dispersion of user ratings in the opening weekend have a positive and significant impact on future national box office performance.⁶

⁶ The modeling framework and the role of the user ratings data in Dellarocas et al. (2007) is very different and not comparable with

This study reinforces the idea that valence of user ratings is important in the predicting movie box office performance, a finding new to the literature linking user ratings to box office performance. We find that the other online WOM measures—volume and variance—have no significant impact on future box office performance. Hence our analysis reverses previously established findings in this domain. Given our access to local market-level data, we are able to aggregate our data to the national level and replicate the findings from the previous literature—that it is volume and not valence that influences box office performance. Our results indicate that it is the aggregation across markets (along with the use of time-series data) that seems to be driving the reversal in results when going from disaggregate to aggregate data. Moreover, we are able to demonstrate that national-level models would give less biased estimates for the online WOM measures if we include appropriate market-level controls. We also look at market-level responsiveness to firm-generated media (advertising) and consumer-generated media (average user rating), and we provide implications for studios regarding their market rollout strategies.⁷

Our study also has some bearing on the aggregation and sequential product rollout literatures. As noted previously, although our box office data are at the local market level, the user review data are at the national level. In the standard aggregation literature, researchers either record the bias by comparing results from aggregate and disaggregate data (Gupta et al. 1996) or propose a "debiasing" procedure (Christen et al. 1997). What we show in this paper is that the aggregation bias problem can be resolved with only aggregated data on the predictor variable as we are able to exploit an institutional feature—the sequential rollout of movies across markets.

Previous researchers, including Bronnenberg and Mela (2004) and Van Heerde et al. (2004), have pointed out that it is important to account for the rollout decisions by firms of products across geographic markets. Both the studies examine the sequential rollout of DiGiorno pizza. In this case, it would be reasonable to assume the following: (1) Kraft adjusted its

advertising (and perhaps other marketing variables) in each of the launch markets based on some factors that are unobservable to the econometrician—the classic endogeneity problem. (2) Kraft decided the sequence of launch markets based on some knowledge of how these unobservables varied across markets and perhaps how these unobservables are correlated across markets (e.g., Chicago is a better market for pizza than Seattle because of factors not captured by the market fixed effect with a negative correlation of the unobservables across markets). Such studies that have examined product rollout decisions impose a specific structure on the nature of the rollout; i.e., they use a full-information estimation approach to account for endogeneity. Instead, we propose a limited information estimation approach, which is likely more appropriate in situations where one does not know the process by which the rollout is determined.

Our study focuses on studying the impact of user ratings. However, consumers could be influenced by other sources of consumer generated media such as blogs and online discussion boards as well. Unlike user reviews, blogs and boards may contain data generated by consumers before they consume the title.⁸ Accommodating such variables along with online reviews within a unified framework would require an identification approach different from the one outlined in this study. It would also be interesting to compare relative impacts of online WOM measures from these different sources. Future researchers could also try to identify the drivers of the different online WOM measures. Although present online WOM could be affected by prior online WOM, firm-generated media such as advertising might also influence future online WOM. Quantifying this "indirect" effect of advertising on sales through online WOM could also be of interest. Our analysis focuses on the "supply" side of online WOM; i.e., the moviegoers have supplied others with their evaluation of the product (movie). This consumer-generated content would have an impact only if there are individuals searching for this information. Thus, a fruitful research direction would be to understand search, i.e., the "demand" side of online WOM, as well. Finally, future research might be well served by focusing on the time series of box office and user reviews while dealing with the additional challenges stemming from the difficulty in obtaining feasible instruments for the online WOM measures.

the role it has in our model and other related aggregate national-level models (Liu 2006, Duan et al. 2008). Those studies examine the direct impact of online WOM measures on movie performance by exploiting the time series of user reviews. Dellarocas et al. (2007) estimate a diffusion model for each movie and estimate the diffusion parameters " p " and " q " in the first stage. In the second stage they regress the recovered first-stage qs on title-specific opening weekend covariates, including measures of user reviews. Thus, static online WOM measures indirectly affect the adoption rates via the imitation parameter.

⁷ Chevalier and Mayzlin (2006) study the impact of negative and positive reviews on sales of books.

⁸ For example, Gopinath et al. (2010) look at the impact of prerelease blogs on the box office performance of the movie on its opening day in its opening markets.

Electronic Companion

An electronic companion to this paper is available as part of the online version that can be found at <http://mktsci.pubs.informs.org/>.

Acknowledgments

The authors are indebted to Yesim Orhun for extensive discussions on identification and for help at various stages of the paper's execution. They also thank Doug Bowman, David Godes, Wes Hartmann, Sandy Jap, Dina Mayzlin, Harikesh Nair, Monic Sun, and Ting Zhu for their input. The usual disclaimer applies. The first author thanks the Kilts Center for Marketing at the University of Chicago for financial support. The authors are grateful to the editor, associate editor, and two reviewers for their valuable suggestions. This research forms a part of the second author's doctoral dissertation. The authors dedicate this paper to the memory of Fred Zufryden. The authors contributed equally, and their names are listed in alphabetical order.

References

- Ainslie A., X. Drèze, F. Zufryden. 2005. Modeling movie life cycles and market share. *Marketing Sci.* **24**(3) 508–517.
- Andrews, D. W. K. 1991. Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* **59**(3) 817–858.
- Box Office Mojo. About movie box office tracking and terms. Accessed December 10, 2009, <http://www.boxofficemojo.com/about/boxoffice.htm>.
- Bronnenberg, B. J., C. F. Mela. 2004. Market roll-out and retailer adoption for new brands. *Marketing Sci.* **23**(4) 500–518.
- Chevalier, J. A., D. Mayzlin. 2006. The effect of word of mouth on sales: Online book reviews. *J. Marketing Res.* **43**(3) 345–354.
- Christen, M., S. Gupta, J. C. Porter, R. Staelin, D. R. Wittink. 1997. Using market-level data to understand promotional effects in a nonlinear model. *J. Marketing Res.* **34**(3) 322–334.
- Cieply, M. 2008. The box office and the weather. *New York Times* (December 21), <http://artsbeat.blogs.nytimes.com/2008/12/21/the-box-office-and-the-weather/>.
- Dellarocas, C., X. M. Zhang, N. F. Awad. 2007. Exploring the value of online product reviews in forecasting sales: The case of motion pictures. *J. Interactive Marketing* **21**(4) 23–45.
- Duan, W., B. Gu, A. B. Whinston. 2008. The dynamics of online word-of-mouth and product sales—An empirical investigation of the movie industry. *J. Retailing* **84**(2) 233–242.
- Elberse, A., B. Anand. 2007. The effectiveness of pre-release advertising for motion pictures: An empirical investigation using a simulated market. *Inform. Econom. Policy* **19**(3–4) 319–343.
- Elberse, A., J. Eliashberg. 2003. Demand and supply dynamics for sequentially released products in international markets: The case of motion pictures. *Marketing Sci.* **22**(3) 329–354.
- Godes, D., D. Mayzlin. 2004. Using online conversations to study word-of-mouth communication. *Marketing Sci.* **23**(4) 545–560.
- Gopinath, S., P. K. Chintagunta, S. Venkataraman. 2010. Do pre-release blogs influence movie opening box-office performance? Working paper, University of Chicago, Booth School of Business, Chicago.
- Greene, W. H. 2003. *Econometric Analysis*, 5th ed. Prentice Hall, Upper Saddle River, NJ.
- Gupta, S., P. K. Chintagunta, A. Kaul, D. R. Wittink. 1996. Do household scanner data provide representative inferences from brand choices: A comparison with store data. *J. Marketing Res.* **33**(4) 383–398.
- Liu, Y. 2006. Word of mouth for movies: Its dynamics and impact on box office revenue. *J. Marketing* **70**(3) 74–89.
- Luo, X. 2007. Consumer negative voice and firm-idiosyncratic stock returns. *J. Marketing* **71**(3) 75–88.
- Moul, C. C. 2007. Measuring word of mouth's impact on theatrical movie admissions. *J. Econom. Management Strategy* **16**(4) 859–892.
- Nam, S., P. Manchanda, P. Chintagunta. 2010. The effect of signal quality and contiguous word of mouth on customer acquisition for a video-on-demand service. *Marketing Sci.* **29**(4) 690–700.
- Neelamegham, R., P. Chintagunta. 1999. A Bayesian model to forecast new product performance in domestic and international markets. *Marketing Sci.* **18**(2) 115–136.
- Newey, W. K., K. D. West. 1987. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* **55**(3) 703–708.
- Shin, H. S., D. M. Hanssens, B. Gajula. 2008. The impact of positive vs. negative online buzz on retail prices. Working paper, Anderson School of Management, University of California, Los Angeles, Los Angeles.
- Sun, M. J. 2008. The informational role of consumer disagreement. Working paper, Department of Economics, Boston University, Boston.
- Van Heerde, H. J., C. F. Mela, P. Manchanda. 2004. The dynamic effect of innovation on market structure. *J. Marketing Res.* **41**(2) 166–183.
- Vogel, H. L. 2007. *Entertainment Industry Economics: A Guide for Financial Analysis*, 7th ed. Cambridge University Press, Cambridge, UK.
- White, H. 1980. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica* **48**(4) 817–838.