Zwick Center for Food and Resource Policy
Working Paper No. 8

Whose and What Chatter Matters?
The Effect of Tweets on Movie Sales

Huaxia Rui, Yizao Liu, and Andrew Whinston

May 7, 2012

Department of Agricultural and Resource Economics
College of Agriculture and Natural Resources
1376 Storrs Road, Unit 4021
Storrs, CT 06269-4021
Phone: (860) 486-1927 Fax: (860) 486-2461
Contact: ZwickCenter@uconn.edu

University of Connecticut

Whose and What Chatter Matters? The Effect of Tweets on Movie Sales *

Huaxia Rui †, Yizao Liu ‡, and Andrew Whinston §

Abstract

Social broadcasting networks such as Twitter in the U.S. and "Weibo" in China are transforming the way online word of mouth (WOM) is disseminated and consumed in the digital age. In the present study, we investigated whether and how Twitter WOM affects movie sales by estimating a dynamic panel data model using publicly available data and well-known machine learning algorithms. We found that chatter on Twitter does matter; however, the magnitude and direction of the effect depends on whom the WOM is from and what the WOM is about. Measuring Twitter users' influence by how many followers they had, we found that the effect of WOM from more influential users is significantly larger than that from less influential users. In support of some recent findings about the importance of WOM valence on product sales, we also found that positive Twitter WOM increases movie sales, whereas negative WOM decreases them. Interestingly, we found that the strongest effect on movie sales comes from those tweets in which the authors expressed their intention to watch a certain movie. We attribute this finding to the dual effects of such intention tweets on movie sales: the direct effect through the WOM author's own purchase behavior, and the indirect effect through either the awareness effect or the persuasive effect of the WOM on its recipients. Our findings provide new perspectives to understand the effect of WOM on product sales and have important managerial implications. For example, our study reveals the potential values of monitoring people's intentions and sentiments on Twitter and identifying influential users for companies wishing to harness the power of social broadcasting networks.

JEL Classification: M3, C2
Keywords: Twitter, word-of-mouth, dynamic panel data

*The authors thank the NET Institute (www.NETinst.org) for financial support. We thank participants of the "Online Community and Group Collaborations I" session of ICIS 2010 and participants of the "social broadcasting" track of SXSW 2011 for useful discussions.

†University of Texas at Austin, E-mail: huaxia@utexas.edu
‡University of Connecticut, E-mail: yizao.liu@uconn.edu
§University of Texas at Austin, E-mail: abw@uts.cc.utexas.edu

1 Introduction

The rise of social broadcasting networks (SBN) such as Twitter in the U.S. and "Weibo" in China[1] is rapidly changing the landscape of online word of mouth (WOM). As a leading example, Twitter has experienced explosive growth in the last few years. As of January 2011, there were nearly 200 million registered users on Twitter who posted 110 million tweets per day.[2] Sina Weibo and Tecent Weibo, two of the leading SBN in China, both reported more than 200 million registered users this year. Although there is little doubt about the significant social and political impact of these fast-growing SBN, an interesting question for the marketing professionals is whether the huge amount of brief messages generated and consumed by the vibrant SBN community in real time has any effect on product sales. If yes, then what is the nature of such effect? The purpose of the present study was to take an initial step into answering this important question by examining the effect of Twitter messages (also known as "tweets") on movie sales.

Tweets are a relatively new type of WOM, which is often considered to be the most credible information source to consumers for the purchase of a new product or new service (Katz and Lazarsfeld 1955). While practitioners have been experimenting with strategies such as buzz management, viral marketing, and referral programs to harness the power of WOM, researchers have also been actively studying the influence and management of WOM. For example, Godes and Mayzlin (2004) used WOM conversations from Usenet to study their influence on TV ratings, and they found that the dispersion of conversations across communities has explanatory power but that the volume of WOM does not. Many researchers have used posts from Yahoo!Movies to study the effect of WOM on movie box office revenue (Liu 2006, Duan et al. 2008, Chintagunta et al. 2011, etc), and the results

[1] We call these sites *social broadcasting networks* because they each are simultaneously a social network and a broadcasting network.

[2] http://www.forbes.com/sites/oliverchiang/2011/01/19/twitter-hits-nearly-200m-users-110m-tweets-per-day-focuses-on-global-expansion/

are mixed. For example, both Liu (2006) and Duan et al. (2008) found that the volume of reviews matters, but the valence does not. This view is also partially supported by Dhar and Chang (2009), who found that future sales of a music album are positively correlated with the volume of blog posts about that album. On the other hand, Chintagunta et al. (2011) measured the effect of national online user reviews on designated market area-level local geographic box office performance of movies, and their findings suggest that it is the valence that drives box office performance, not the volume. A recent study by Sonnier et al. (2011) seems to support this view, although their results were based on the analysis of a different product category.[3] Adding to this debate is another study by Onishi and Manchanda (2010), who found mixed results for volume and valence using Japanese blog and sales data of three different products. From a different angle, Dellarocas et al. (2007) focused on the forecasting of movie sales and their results suggested the values of online product reviews.

Researchers have also examined the effect of online WOM on the sales of products other than movies or TV. For example, Chevalier and Mayzlin (2006) examined the effect of consumer reviews on relative sales of books at Amazon.com and Barnesandnoble.com and they found that an improvement in a book's reviews at one site leads to an increase in relative sales at that site. Trusov, Bucklin, and Pauwels (2009) studied the effect of WOM on member growth at an Internet social networking site and found that WOM referrals have substantially longer carryover effects than traditional marketing actions and produce substantially higher response elasticities.

Our study differs from previous WOM literature in several important ways, which are closely related to the unique features of Twitter WOM. First, as compared with online forums such as Yahoo!Movies, Twitter is a more natural environment to study the awareness effect of WOM. The awareness effect of WOM on product sales refers to its function of spreading

[3] Product names are not mentioned in the paper; however, the products are characterized as durable search goods.

basic information about the product among the population. As the name suggests, the awareness effect influences people's behavior only by informing them and thereby putting the product in their choice set. This influence is in contrast with the so-called persuasive effect which refers to WOM's function of altering people's preferences toward the product and thus influencing their purchase decisions. Because people who visit online forums such as Yahoo!Movies to find out movie review information are most likely already aware of these movies, the awareness effect of WOM on these sites is quite limited. On the other hand, WOM generated on Twitter is actually pushed to the followers of the authors. This difference between the *pull* mode on Yahoo!Movies and the *push* mode on Twitter makes Twitter a better environment for researchers to study the awareness effect of WOM.

Second, unlike many online forums in which no social structural information is available, Twitter provides an application program interface (API) structure with which we can extract the number of followers each author has. This seemingly simple information may be useful for the study of WOM. It allows us to know the number of direct recipients of each message. The number of followers a Twitter user has is like the size of her audience. The more followers she has, the more people she can reach, and the larger the effect of her WOM. Probably at a more profound level, the number of followers of a user could be a coarse proxy of the user's social influence. The very same WOM message may have quite a different impact on the recipients, depending on whom the message is from. There is little debate as to whether chatter matters to firms and earlier literature has studied extensively the question of what kind of chatter matters. But whose chatter matters? The two-step flow theory in sociology suggests that some people (opinion leaders) are more influential than others (imitators) and that information often moves first from mass media to opinion leaders and then from opinion leaders to imitators (Katz and Lazarsfeld 1955, Gladwell 2000, Slywotzky and Shapiro 1993). Applying this theory to WOM leads to the hypothesis that WOM messages from a certain group of people may have disproportionate influence on a firm's product sales. Surprisingly, there is little research in marketing science addressing this question. One notable exception is the work done by Van den Bulte and Lilien (2001) who developed a model of innovation diffusion in markets with two segments. Inspired by these works, we divided the WOM messages from Twitter into two groups, according to the number of followers each author had, and we investigated whether the proportion of each type of tweets matters in terms of their effect on movie sales. To the best of our knowledge, we are the first in the marketing literature to take into account the different degrees of influence associated with each WOM message. Intuitively, WOM from Twitter users with more followers might have larger impact on movie sales than WOM from users with fewer followers; however, this argument is not indisputable. For example, people who have fewer followers may have greater influence among the followers because they have closer relationships, and their social ties are stronger. Our empirical results suggest that the WOM effect from the group of users with more followers is significantly larger than that from the group of users with fewer followers.

Third, although most researchers in previous literature have focused on the study of review-type WOM, we have deliberately disentangled the different effects of post-consumption WOM (i.e., WOM generated by people who have consumed the product) and pre-consumption WOM (i.e., WOM generated by people who have not consumed the product).[4] Pre-consumption WOM is generally about people's intentions or plans to purchase a product,[5] whereas post-consumption WOM is usually about people's experiences and/or attitudes toward a product after consumption. Although previous literature seems to suggest that all WOM after the release of a movie is post-consumption WOM, this is not true for Twitter WOM. People on Twitter frequently talk about their plans or intentions of taking certain actions, such as

[4]Liu (2006) considered pre-release WOM which is a subset of pre-consumption WOM, and found that it has significant explanatory power for aggregate box office revenue. However, pre-consumption WOM is not limited to prerelease WOM in the case of a movie.

[5]For this reason, we use pre-consumption tweets and intention tweets interchangeably in this paper.

watching a specific movie or having a certain breakfast. The intention may be expressed directly or indirectly. For example, when people express their intentions to watch a movie, they may explicitly say so or they may complain about the difficulty of getting movie tickets. We believe the prevalence of pre-consumption WOM on Twitter poses new challenges but also offers new opportunities for managers and researchers. The main challenge comes from the automatic identification of these tweets, whereas an obvious advantage is the addition of this new dimension of WOM measurement besides valence. Intuitively, pre-consumption WOM should be treated differently when it is used to explain movie box office revenue. The authors of post-consumption WOM have already consumed the product and are less likely to purchase the product again in the near future;[6] hence, post-consumption WOM affects future product sales indirectly through its awareness effect and persuasive effect. On the other hand, because the authors of pre-consumption WOM have not consumed the product yet and are more likely than the average population to consume the product in the near future, we would expect pre-consumption WOM to have both a direct and indirect effect on future product sales. However, it is hard to predict whether pre-consumption WOM would have a larger or smaller impact on movie sales as compared with post-consumption WOM. The direct effect of pre-consumption WOM on movie sales seems to suggest a larger impact on movie sales, but on the other hand, pre-consumption WOM contains less information about product quality, which may result in a smaller impact. Our empirical results suggest that the effect of pre-consumption WOM on movie sales is actually larger than that of post-consumption WOM.

Finally, because of Twitter's simplicity and popularity, there are a huge number of tweets on a vast number of topics. For example, on March 4, 2010, one day before the release of the movie *Alice in Wonderland*, there were nearly 15,000 tweets about this movie. On February 18, 2010, two months after the release of the movie *Avatar*, there were still around 13,000

[6] This is true for many products, such as movies, and durable goods.

tweets mentioning it. In our empirical study, we used a total of more than four million tweets about 63 movies, which is significantly more than the 12,136 posts used in Liu (2006) and the 95,867 posts used in Duan et al. (2008). The large number of WOM messages means that we may have less bias in our sample than in the samples used in previous literature. However, the downside of the large data size is that we had to rely heavily on computer programs to automatically classify WOM messages into pre-consumption WOM and post-consumption WOM, and to conduct sentiment analysis on post-consumption WOM further, which, of course, is less reliable than human judgment and inevitably introduces new errors.

Our WOM data were collected from Twitter.com, and the movie sales data were collected from BoxOfficeMojo.com, both of which allow public retrieval of their data. The algorithms we used to classify tweets are also well-known machine learning algorithms. We used a dynamic panel data model to estimate the effect of WOM on movie sales in order to handle the endogeneity problem typical in the study of the effect of WOM on firm product sales. The openness of our data, algorithms, and econometric models means that our study can be easily utilized by researchers and practitioners for various purposes.

The rest of the paper is organized as follows: We first describe in detail how we collected and processed our data in Section 2. We then introduce our methodology in Section 3 and present our empirical results in Section 4. We discuss the managerial implications of this study in Section 5. Finally, we conclude our paper and point out some future research directions in Section 6.

2 Data

2.1 Data Collection

Movie sales data were collected from BoxOfficeMojo.com[7] and tweet information was collected from Twitter.[8] We collected daily box office revenues of 63 movies that were widely released between June 2009 and February 2010.[9] Although obtaining movie sales information is straightforward, collecting tweets on those movies is a little bit challenging because of the real-time nature of the data.[10] Our program queried the Twitter server for tweets mentioning the 63 movies once an hour during this period, which resulted in a total number of 4,166,623 tweets. For each tweet, we collected the content of the tweet, the time when it was posted, and the author's account name. Using each author's account name, we further retrieved the number of followers each author had.[11]

2.2 Data Processing

After the tweets were collected, we used a filtering program to filter out advertising tweets. Although we used several rules to determine whether a tweet was an advertising tweet, the most effective one was simply checking whether the tweet contained a URL. There were also some irrelevant tweets containing the search keyword. This was particularly a problem if the movie name was a single word or a commonly used phrase. We first randomly selected 200 tweets for each movie and manually checked for irrelevant tweets. For some movies like *Ninja Assassin* and *Shutter Island*, there was almost no irrelevant tweet because these two phrases

[7] http://www.boxofficemojo.com

[8] http://www.twitter.com

[9] We excluded movie titles for which it was difficult to correctly identify tweets that were related to those movies. For example, it is very hard to distinguish tweets talking about the movie *2012* from tweets talking about the year 2012.

[10] Twitter streaming API would be more suitable for this purpose but they were not available at the time we started the data collection.

[11] A user's followers are the people who subscribe to receive the user's tweets.

are rarely used on Twitter in contexts other than those movies. However, for some movies such as *Wolfman*, *The Hangover*, and *It's Complicated*, there were a considerable amount of irrelevant tweets. To reduce those irrelevant tweets, we adopted two approaches. First, we used a movie lexicon containing words or phrases such as *movie, cinema, film, theater* , and *ticket* to pick out relevant tweets containing words or phrases in the lexicon. Second, for each movie, we used a customized lexicon for those irrelevant tweets and eliminated tweets containing words or phrases in that lexicon. For example, for the movie *The Hangover*, if a tweet contained the phrase "suffering from a hangover" or the word "drunk," then that tweet was classified as an irrelevant tweet. We manually checked tweets whose relevancy could not be determined by the two procedures.

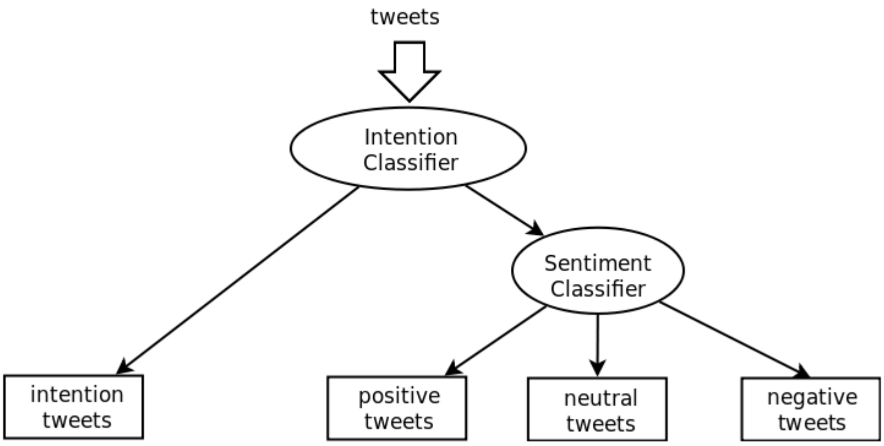

Figure 1: Tweet Classification

After filtering out the advertising tweets and irrelevant tweets, we classified a tweet into one of the four mutually exclusive categories: intention, positive, negative, and neutral. By intention tweets, we mean those tweets whose authors clearly expressed their willingness to watch the movie in the future. For example, the tweet, "Wow! I wanna see the lovely bones!!" is clearly an intention tweet. On the other hand, the tweet, "DAMN IT!!! Didn't make it... Sold out tickets for Avatar!!!" is also an intention tweet, even though it is not obvious at first glance. A positive tweet was a tweet in which the author expressed positive sentiment toward the movie. Similarly, a negative tweet was a tweet in which the author expressed negative sentiment toward the movie. Neutral tweets were all other tweets that did not belong to any of the above three categories. Figure 1 illustrates the classification scheme.

We used an intention lexicon to extract features from tweets and then used a support vector machine (Steinwart and Christmann 2008) to construct the intention classifier. The intention lexicon was built from the movie tweets in our sample. In the Appendix, we list some of the regular expressions in the intention lexicon. For the sentiment analysis of tweets, we constructed a Naive Bayesian classifier (Mitchell 1997), which drew upon a lexicon of positive words/phrases and negative words/phrases. Naive Bayesian classifiers are often used in the literature for text mining because of their simplicity. Of course, there are more sophisticated classifiers for sentiment analysis that might yield higher accuracy. An in-depth study of these methods is not the focus of this paper. Both classifiers were trained and tested on a corpus of over 3,000 tweets that were manually labeled. The precisions and recalls for the intention classifier and the sentiment classifier are reported in Table 1 and Table 2 respectively.[12]

Table 1: Precisions and Recalls of the Intention Classifier

	Precision	Recall
Intention tweets	98%	87%
Non-intention tweets	93%	99%

[12]Precision is the fraction of retrieved instances that are relevant, whereas recall is the fraction of relevant instances that are retrieved. For example, a precision of 98% and a recall of 87% for the class of intention tweets mean that 98% of the tweets classified by the program as intention tweets are indeed intention tweets based on human judgment, and 87% of the intention tweets based on human judgment are classified by the program as intention tweets.

Table 2: Precisions and Recalls of the Sentiment Classifier

	Precision	Recall
positive	75%	80%
negative	65%	71%
neutral	75%	68%

2.3 Variables

Table 3 lists the description and measurements of the key explanatory variables related to our model, and Table 4 provides the summary statistics for these variables. All variables are extracted from weekly aggregated data.[13] *Revenue* is the total movie gross revenue for a week, which is defined as the 7-day period from Friday to Thursday. On average, a movie's weekly box office revenue is around 9.5 million dollars. To capture the effect of Twitter WOM on movie box office revenues, we included the total number of tweets mentioning a movie for each week as an explanatory variable (*Total tweets*). On Twitter, a user often has a certain number of followers who subscribe to receive the author's tweets in real time. However, the number of followers for a user varies from zero to millions. We divided the tweets for each movie in each week into two groups according to the number of followers each tweet author had, which could be a proxy of their influence. Specifically, Type-1 tweets were those tweets whose authors had fewer than 400 followers and the rest of the tweets, whose authors had more than 400 followers, are Type-2 tweets. In our sample of tweets, about 85% of the tweet authors had fewer than 400 followers. On average, each movie had 5,208 Type-1 tweets and 853 Type-2 tweets each week. We computed the ratio of each type of tweets (except for neutral tweets) among all the tweets for each movie in each week.

[13] The rational for this weekly aggregation is as follows. The focus of a typical dynamic panel data model is on panels on which a large number of individuals are observed for a small number of time periods. However, in our analysis, we have a total number of 63 movies and many of them are in the theater for over 90 days. Therefore, we do the weekly aggregation to shorten the number of time periods. After the weekly aggregation, the longest time period for a movie is 15 weeks.

Table 3: Key Variables Used in Dynamic Panel Data Estimation

Revenue	Movie gross box office revenue from Friday to next Thursday
Total tweets	Total number of tweets mentioning the name of the movie i from Friday to next Thursday
Type-1 tweets	Total number of tweets with followers less than 400 (fewer audiences) from Friday to next Thursday
Type-2 tweets	Total number of tweets with followers more than 400 (more audiences) from Friday to next Thursday
Type-1 tweets ratio	Ratio of Type-1 tweets in a week
Type-2 tweets ratio	Ratio of Type-2 tweets in a week
Intention tweets ratio (%)	Total number of tweets showing intention of seeing movie i from Friday to next Thursday
Positive tweets ratio (%)	Ratio of tweets with positive comments in a week
Negative tweets ratio (%)	Ratio of tweets with negative comments in a week

Table 4: Summary Statistics of Key Variables for All Movies

Variable	Mean	Standard Deviation
Revenue	9,435,003	21,000,000
Total tweets	5,997.04	10,652.23
Type-1 tweets	5,208.16	9,246.04
Type-1 tweets ratio (%)	86.17	6.75
Type-2 tweets	853.27	2117.66
Type-2 tweets ratio (%)	13.97	6.65
Intention tweets ratio (%)	30.42	9.61
Positive tweets ratio (%)	26.69	7.13
Negative tweets ratio (%)	4.86	3.84
No. of Weekly Observation	572	

3 Model

3.1 Model Specification

To capture the dynamic nature of the data as well as the cross-sectional effects, we formulate and estimate a dynamic panel data model using the method of Arellano and Bond (1991).[14] More specifically, we write a dynamic panel data model with predetermined variables and autoregressive specification of the form:

$$y_{it} = \alpha y_{i,t-1} + \beta' x_{i,t-1} + \eta_i + \nu_{it}, \tag{1}$$

where the dependent variable y_{it} is the movie gross revenue for movie i at week t, $y_{i,t-1}$ is its own one-period lag value, and $x_{i,t-1}$ is a set of explanatory variables, including *Total tweets*, *Type-2 tweets ratio*, *Intention tweets ratio*, *Positive tweets ratio*, and *Negative tweets ratio*.[15] Hence, we have

$$\begin{aligned} Revenue_{it} &= \alpha Revenue_{i,t-1} + \beta_1 Total\ tweets_{i,t-1} \\ &+ \beta_2 Type-2\ tweets\ ratio_{i,t-1} + \beta_3 Intention\ tweets\ ratio_{i,t-1} \\ &+ \beta_4 Positive\ tweets\ ratio_{i,t-1} + \beta_5 Negative\ tweets\ ratio_{i,t-1} \\ &+ \eta_i + \nu_{it} \end{aligned} \tag{2}$$

[14]Strictly speaking, our data is a cluster sample because movies in our sample were not released on a same week. Nevertheless, we refer to it as panel data in this paper because the dynamic structure of the data is better described as a panel and also because such data was often referred as panel data in the literature. Our model structure is similar to Godes and Mayzlin (2004).

[15]De Vanny and Walls (1996) explored the path of film revenues and showed that information transmission among consumers will cause autocorrelated growth rates among movies. We did not use more lagged values for $x_{i,t}$ because Twitter is well known for its real-time feature. In fact, Twitter has on its homepage the following description of itself: *Instant updates from your friends, industry experts, favourite celebrities, and what's happening around the world.* It is thus reasonable to assume that tweets affecting next week's sales are mainly from the current week, rather than from the previous week, or even earlier weeks.

As has been acknowledged by many previous works, the relationship between movie box office revenue and WOM is intertwined. More WOM messages implies a larger awareness effect which leads to higher movie revenue. On the other hand, higher movie revenue in turn results in more people talking about the movie. Therefore, there is a problem of potential endogeneity. To account for this, we use one-period lagged values of tweet-related variables as our explanatory variables. It is reasonable to believe that the previous week's WOM would have effects on the current week's movie revenue but would not be affected by this week's movie revenue. We expect that this treatment, along with the control of the previous week's movie revenue, will alleviate the endogeneity concern typical in the study of the effect of WOM on product sales.

To estimate the parameters, note that we can not assume strict exogeneity of x_{it}. Strict exogeneity of x_{it} implies $E[x_{is}\nu_{it}] = 0$ for all s and t. However, this is not true in our model for WOM and movie box office revenue. For example, it is plausible that a shock to the movie revenue this week (captured by ν_{it}) will affect the total number of tweets mentioning the movie next week and later (captured in x_{is}, $s > t$). In such a case, even though we could assume that $E[x_{is}\nu_{it}] = 0$ for $s < t$, we have to allow $E[x_{is}\nu_{it}] \neq 0$ for $s \geq t$. Variables satisfying these conditions are commonly referred to as *predetermined variables*. When the variables are predetermined, we should not include the whole vector of first differences of observed x_{it} into the instrument matrix. Instead, we should include just the levels of x_{it} for those time periods that are assumed to be unrelated to $\triangle\nu_{it}$.

There are many other variables that will affect movie revenue, such as movie genre, director and actors, MPAA rating, and production budget. However, because these variables do not vary over time, they appear in η_i and are eliminated after first-differencing. Therefore, even though these movie-specific effects are not observed in our data, they are controlled in our model indirectly, which is an advantage of using panel data.

3.2 Inference and Estimation

Following Arellano and Bond (1991), we estimate Equation (1) using an optimal GMM method. To remove the movie-specific time-invariant factors such as movie genre and star power, we do first difference on Equation (1) to obtain

$$\bar{y}_{it} = \alpha \bar{y}_{i,t-1} + \beta' \bar{x}_{i,t-1} + \triangle \nu_{it}, \tag{3}$$

where

$$\begin{aligned} \bar{y}_{it} &= y_{it} - y_{i,t-1} \\ \bar{x}_{i,t-1} &= x_{i,t-1} - x_{i,t-2} \\ \triangle \nu_{it} &= \nu_{it} - \nu_{i,t-1}. \end{aligned}$$

We need to estimate Equation (3). However, the error terms $\bar{\nu}_{it}$ are now correlated with both $\bar{y}_{i,t-1}$ and $\bar{x}_{i,t-1}$. Therefore, we need instrument variables for them to address the endogeneity concern. To use as much information available as possible, we use all the lagged values of $\bar{y}_{i,t-1}$ and $\bar{x}_{i,t-1}$ as instruments. For example, y_{i1} and x_{i1} are the only instruments that can be used in Equation (3) for period $t = 3$, whereas we can use y_{i1}, y_{i2}, x_{i1}, x_{i2} for period $t = 4$. In general, we can use $y_{i1}, \cdots, y_{i,t-2}$, $x_{i1}, \cdots, x_{i,t-2}$ as instruments for period t.

Define X_i, Y_i, and Z_i as follows:

$$\bar{X}_i = \begin{bmatrix} \bar{y}_{i,2} & \bar{x}_{i,2} \\ \vdots & \vdots \\ \bar{y}_{i,T-1} & \bar{x}_{i,T-1} \end{bmatrix}, \qquad \bar{Y}_i = \begin{bmatrix} \bar{y}_{i,3} \\ \vdots \\ \bar{y}_{i,T} \end{bmatrix},$$

$$
Z_i = \begin{bmatrix} y_{i,1} & x_{i,1} & 0 & 0 & 0 & 0 & \dots & 0 & \dots & 0 & 0 & \dots & 0 \\ 0 & 0 & y_{i,1} & y_{i,2} & x_{i,1} & x_{1,2} & \dots & 0 & \dots & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 & \dots & y_{i,1} & \dots & y_{i,T-2} & x_{i,1} & \dots & x_{i,T-2} \end{bmatrix}.
$$

The GMM estimator $\delta_{GMM} = (\alpha, \beta')'$ minimizes the criterion[16]

$$
J = \left[\sum_{i=1}^{N} Z_i'(\bar{Y}_i - \bar{X}_i\delta)\right]' \mathbf{W} \left[\sum_{i=1}^{N} Z_i'(\bar{Y}_i - \bar{X}_i\delta)\right]
$$

where

$$
\mathbf{W} = \left[\frac{1}{N}\sum_{i=1}^{N}(Z_i'\triangle\hat{\nu}_i\triangle\hat{\nu}_i{}'Z_i)\right]^{-1}
$$

is the weighting matrix and $\triangle\hat{\nu}_i$ are consistent estimators of the first-differenced residuals obtained from a preliminary consistent estimator.[17] If we denote the total number of explanatory variables in Equation 3 by k (including $\bar{y}_{i,t-1}$), we have the following estimator:

$$
\delta_{GMM} = (\bar{X}'Z\mathbf{W}Z'\bar{X})^{-1}\bar{X}'Z\mathbf{W}Z'\bar{Y}, \tag{4}
$$

where $\bar{X} = (\bar{X}_1', \bar{X}_2', \cdots, \bar{X}_N')'$ is a $(T-2)N \times k$ matrix, $\bar{Y} = (\bar{Y}_1', \bar{Y}_2', \cdots, \bar{Y}_N')'$ is a $(T-2)N \times 1$ matrix, and $Z = (Z_1', Z_2', \cdots, Z_N')'$ is a $(T-2)N \times k(T-2)(T-1)/2$ matrix.

[16]For simplicity, we describe and formulate the estimator when the panel is balanced. But it can accommodate unbalanced panel data easily with minor changes. With unbalanced panel, the total time period for each movie is T_i. The number of columns in Z_i is $p = (\tau - 2)(\tau - 1)/2$, where τ is the total number of periods on which observations are available for some movies in the sample, and $Z_i = diag(y_{i1}, ..., y_{is}x_{i1}, ..., x_{is}), (s = 1, ..., \tau - 2)$ only if τ observations are available on movie i. For movies with $T_i < \tau$, the rows of Z_i corresponding to the missing equations are deleted and the missing values in the remaining rows are replaced by zeroes. See Page 281 of Arellano and Bond (1991) for a discussion on this.

[17]A reasonable choice for the initial consistent estimator could be obtained by using $W_{1N} = [\frac{1}{N}\sum_{i=1}^{N}(Z_i'HZ_i)]^{-1}$ where H is a $(T-2)$ square matrix with 2's on the main diagonal, -1's on the first off-diagonals and zeros elsewhere.

4 Empirical Results

Using the weekly cross-sectional data for 63 movies, we estimate the unbalanced dynamic panel data model specified in Section 3. The panel we used is unbalanced because some movies are in theaters longer than others. In addition to the advantage of more data points, the use of the unbalanced panel may lessen the effect of self-selection of movies in the sample. We first discuss the estimation results from the dynamic panel data model in Section 4.1. We then conduct a robustness check for alternative classifications for Type-1 tweets and Type-2 tweets in Section 4.2.

4.1 Results from Dynamic Panel Data Model

In Table 5, we report estimates of the dynamic panel data model (2). As is expected, the previous week's total movie revenue has positive effects on the current week's revenue, suggesting the positive autocorrelation between box office revenues of consecutive weeks.

Table 5: Estimation Results from Dynamics Panel Data for All Movies

Variable	Estimate	SD	P > $\|z\|$
Lag Revenue	0.30	0.01	0.00
Total tweets	5.34	0.36	0.00
Type-2 tweets ratio (%)	76,348.75	18,250.69	0.00
Intention tweets ratio (%)	157,905.00	38,432.42	0.00
Positive tweets ratio (%)	125,881.30	62,131.49	0.00
Negative tweets ratio (%)	-137,451.10	70,214.58	0.00
No. Weekly Observations:	568		

The total number of tweets has a positive and significant effect on movie box office revenue, which means that, holding other factors fixed, more chatter about a movie increases its sales. The positive effects of WOM volume on movie box office revenue is consistent with results from several previous studies. One plausible explanation for this result is that large WOM volume amplifies the awareness effect of WOM.

One advantage of our data is the ability to identify the different effects of tweets from users with different numbers of followers, which, to the best of our knowledge, has not been studied in the literature. We explore this effect by examining whether a bigger proportion of tweets from users with a large audience (Type-2 tweets) is associated with higher or lower movie box office revenue. If the awareness effect is dominant, then most likely we will observe a positive and significant result for the *Type-2 tweets ratio*, because a higher *Type-2 tweets ratio* means more users will be reached by WOM messages on average. On the other hand, if the persuasive effect is dominant, then the sign of *Type-2 tweets ratio* is difficult to predict because a user having more followers does not necessarily mean that he or she is more influential. The result in Table 5 suggests that a higher ratio of Type-2 tweets does have a positive and significant effect on movie box office revenue in our model. Specifically, on average in our sample, an increase of one percent of the ratio of Type-2 tweets from the previous 7 days leads to a $76,349 increase in movie revenue for the current week. There are a few caveats when we interpret this coefficient. First, it should be emphasized that we need to keep all other variables fixed in order to interpret the ceteris paribus effect of *Type-2 tweets ratio*. The following hypothetical scenario illustrates this point. Suppose there are 10,000 tweets mentioning a movie for a particular week, and 14% of these tweets are Type-2 tweets. Now, without any change to the content of each tweet, what would the movie revenue for the next week have been if the authors of 100 Type-1 tweets (i.e., 1% of the total tweets) had had a much larger audience (more than 400) than they actually did have (less than 400)? The answer to this hypothetical question is exactly the meaning of the coefficient for the variable *Type-2 tweets ratio*. Second, the coefficient value reported in Table 5 should be interpreted as the average effect of one percent increase of the *Type-2 tweets ratio* in our sample. Third, the positive and significant coefficient of *Type-2 tweets ratio* suggests that WOM messages from those with a large audience have a larger impact on movie box office revenue, but our results do not reveal the exact mechanism of why this is the case. It could be that the awareness effect is dominant. It could also be that the persuasive effect is dominant, in which case the increase of the effect of intention and positive WOM from the increase of audience size is larger than the decrease of the effect of negative WOM from the increase of audience size. Although we do not explore the detailed mechanism underlying the stronger effect of Type-2 tweets in the current model framework, it is certainly an interesting and important question to pursue in future research.

Another variable of special interests that is featured in our study is *Intention tweets ratio*, which turns out to be a significant predictor of movie revenue in the subsequent period. On average, a one percent increase in intention tweets will increase the movie box office revenues by \$157,905, which is even higher than the effects of positive tweets in terms of the size of coefficient. On the surface, this is surprising because positive WOM is like the endorsement of a product and is supposed to have the most positive effect on movie revenues. However, it is not that surprising if we take into account the dual effects of pre-consumption WOM—the direct effect on movie sales through the WOM author's own purchase behavior, and the indirect effect on movie sales through either the awareness effect or the persuasive effect of the WOM on its recipients. The significant and positive effect of the *Intention tweets ratio* on movie revenues suggests that there is significant amount of credibility in the pre-consumption WOM, which is a strong indication of the value of recognizing people' purchase intention through the analysis of Twitter data. It also suggests that there are potential opportunities of targeted advertising and marketing on Twitter.

In the literature, many studies have tried to identify the effects of WOM valence, but most of their results have led to the conclusion that the valence of WOM does not affect movie revenues. However, in the present analysis, we find that valence does have nonnegligible effects on movie box office revenues, which is consistent with the recent findings by Chintagunta et al. (2011). Specifically, WOM with positive sentiment toward the movie will increase movie sales significantly. On average, from our sample, one percent of increase in the positive tweets ratio will increase movie revenues by $125,881. On the other hand, the negative tweets from the previous week turn out to have significant and negative effects on a movie's box office revenue. On average from our sample, one percent of increase in the ratio of negative tweets will decrease the weekly gross revenues by $137,451. Note that the magnitude of this impact is larger than that of the positive tweets. Even though negative tweets account for only a very small proportion of all tweets about a movie, they might severely discourage people from watching that movie. The impact of negative WOM on movie revenue puts some doubt on an old saying that has long been circulated in marketing: Any publicity is good publicity.

4.2 Robustness Check for Different Tweets Classifications

In the benchmark dynamic panel data model discussed in Section 4.1, we classified the tweets into two types according to the number of followers of the author of each tweet. The associated cut-off number we chose was 400, which is roughly the 85th percentile of all tweets ranked according to the number of followers of the authors. This means the authors of roughly 85% of the tweet have fewer than 400 followers, and the author of roughly 15% of the tweets have more than 400 followers. In order to check whether the estimation results are robust to different cut-off points, we run the same model six times using 100, 200, 300, 500, 600, and 700 as cut-off points, respectively. For example, a cut-off point of 200 suggests that the author of each Type-1 tweet has fewer than 200 followers, and the author of each Type-2 tweet has more than 200 followers, and so on. In order to run these models, we regroup the raw tweets according to different cut-off numbers and then aggregate the data into weekly observations for dynamic panel data estimation.

In Table 6, we present estimation results for six different robust models. In general, the estimation results do not vary significantly from the results of the benchmark model in Table 5. As in the benchmark model, *Total tweets*, *Intention tweets ratio*, and *Positive tweets ratio* all have positive and significant impacts on movies' box office revenues for all six models, whereas *Negative tweets ratio* always has negative impact on movies' box office revenues.

Type-2 tweets, whose authors have a relatively larger number of followers, have a noticeably higher impact on movie revenue. It is interesting to note that the coefficient of the *Type-2 tweets ratio* roughly has an increasing trend as we increase the cut-off number, which is consistent with our intuition given that the coefficient of *Type-2 tweets ratio* is significantly positive.

5 Managerial Implications

For managers, there are at least three key take-aways from our study. First, our results regarding the different impacts of WOM from people with different degrees of influence confirms the perception that firms that want to harness the online WOM about their products should actively monitor or even seek WOM messages produced by the more influential people. An important caveat to this point is that people who are influential on some topics might not be influential on others. It is thus important for companies to identify those who are influential on topics related to their products. For the purpose of managing WOM on Twitter, using a user's number of followers as a coarse measure of influence is a good starting point, but coming up with a customized measure of influence and frequently refining and updating such a measure is very important.

Probably the most surprising result found in this study is that pre-consumption WOM has more to say about future movie sales than does post-consumption WOM. This observation has two important implications. First, it suggests that companies may carefully monitor

Table 6: Robustness Result Check from Dynamics Panel Data for All Movies

	Cut-off Point: 100 Followers			Cut-off Point: 200 Followers		
Variable	Estimate	SD	P > $\|z\|$	Estimate	SD	P > $\|z\|$
Lag Revenue	0.30	0.01	0.00	0.30	0.01	0.00
Total tweets	5.34	0.36	0.00	5.34	0.36	0.00
Type-2 tweets ratio (%)	57,075.00	13,268.20	0.00	68,758.49	15,386.40	0.00
Intention tweets ratio (%)	166,252.40	38,986.67	0.00	165,252.30	38,747.64	0.00
Positive tweets ratio (%)	154,467.20	62,958.14	0.01	147,361.70	62,773.98	0.02
Negative tweets ratio (%)	-110,023.50	66,972.74	0.10	-132,234.30	68,819.56	0.06

	Cut-off Point: 300 Followers			Cut-off Point: 500 Followers		
Variable	Estimate	SD	P > $\|z\|$	Estimate	SD	P > $\|z\|$
Lag Revenue	0.30	0.01	0.00	0.30	0.01	0.00
Total tweets	5.35	0.36	0.00	5.34	0.36	0.00
Type-2 tweets ratio (%)	73,886.63	16,886.65	0.00	78,897.60	19,074.68	0.00
Intention tweets ratio (%)	161,031.10	38,541.20	0.00	162,974.90	38,698.65	0.00
Positive tweets ratio (%)	136,802.20	62,496.97	0.03	122,939.60	61,809.90	0.05
Negative tweets ratio (%)	-127,575.40	68,600.83	0.06	-138,414.90	70,342.58	0.05

	Cut-off Point: 600 Followers			Cut-off Point: 700 Followers		
Variable	Estimate	SD	P > $\|z\|$	Estimate	SD	P > $\|z\|$
Lag Revenue	0.29	0.01	0.00	0.29	0.01	0.00
Total tweets	5.37	0.36	0.00	5.38	0.36	0.00
Type-2 tweets ratio (%)	75,899.41	19,575.80	0.00	74,796.21	19,945.52	0.00
Intention tweets ratio (%)	159,233.00	38,565.15	0.00	157,079.40	38,513.34	0.00
Positive tweets ratio (%)	119,849.70	61,625.73	0.05	113,291.10	61,612.84	0.07
Negative tweets ratio (%)	-133,041.20	70,878.51	0.06	-118,440.00	69,438.95	0.09

people's intention toward certain products on Twitter and incorporate that information to better forecast future sales.[18] Nowadays, a smart phone is often equipped with a GPS device and many Twitter users are using their smart phones to send out tweets along with their geolocations. This additional information may be useful for companies not only to forecast future sales but also to better manage their inventory. Such an idea may sound futuristic; however, considering the extent to which information technology has transformed our way of life and doing business in the last two decades, we believe it is crucial for managers to think ahead and to act swiftly in this rapidly changing business environment. Second, the dual effect of intention tweets revealed in our study suggests the possibility of targeted advertising on SBN. An important question that firms face in advertising is how to develop effective targeting strategies. The success of Google's AdWords is an example of how important targeted advertising is to firms. Our results suggest that Twitter is a natural environment where people express their intention to purchase certain products, and firms could potentially make use of this information to run targeted advertising on Twitter. For example, firms may encourage people to express their purchase intention on Twitter by providing coupons or other incentives.[19] This will not only facilitate people's purchase but will also increase the amount of buzz about their product. Of course, pre-consumption WOM generated in response to such marketing strategy would not be "organic" anymore, and the impact of "non-organic" pre-consumption WOM on product sales would be expected to be attenuated. It would be valuable for future research to study this issue, possibly through field experiments.

In summary, managers who want to leverage the power of SBN should (1)identify users who are influential on topics of their interest; (2)monitor people's intentions toward topics of their interest; (3)experiment targeted advertising based on people's expressed intentions.

[18]The idea of using Twitter data to analyze people's intention toward certain products for marketing/management purpose was first proposed in Rui et al. 2009.

[19]How to implement this in practice is an interesting question. For example, firms may do this selectively and randomly to discourage strategic behavior.

6 Conclusion

The goal of this paper was to investigate whether and how Twitter WOM—a recently popular and relatively new form of online WOM—affects movie sales. We collected Twitter WOM data using Twitter API and movie sales data from BoxOfficeMojo.com, both of which are publicly available. We then carried out the classification of tweets and conducted a sentiment analysis using well-known machine learning algorithms. Having extracted variables characterizing Twitter WOM, we used a dynamic panel data model to explore the effect of Twitter WOM on movie sales. Our study adds several important contributions to the literature. We take a first step into measuring the potentially different impacts of WOM on movie sales from people with different degrees of influence. Assuming that the number of followers a Twitter user has is a coarse proxy of his or her influence, our empirical results suggest that indeed, the effect of WOM on product sales from a more influential person is significantly larger than that from a less influential person. Our second contribution is identifying and estimating the effect of a new type of WOM on movie sales, namely, the pre-consumption WOM. The prevalence of pre-consumption WOM is most likely a result of the recent popularity of SBN, which probably explains why it was largely ignored in the earlier literature on WOM. We find that the effect of pre-consumption WOM on movie sales is larger than that of post-consumption WOM. We attribute this finding to the dual effects of pre-consumption WOM on movie sales: the direct effect on movie sales through the WOM author's own purchase behavior, and the indirect effect on movie sales through either the awareness effect or the persuasive effect of the WOM on its recipients. The third contribution of this study is to support the view that the valence of WOM does matter. Unlike most of the previous literature that uses ratings provided by users, we analyzed the sentiment of each tweet using a Naive Bayesian classifier and the estimation results of our model suggest that positive WOM increases product sales whereas negative WOM decreases them.

All data and algorithms used in the paper are readily available to marketing researchers and practitioners. Given the tremendous amount of data on Twitter, we exploited only a very small portion of it in this paper. With Twitter's easy-to-use API structure and its ever-growing popularity, we believe it could be particularly rewarding for marketing researchers and practitioners to dig into this goldmine. The following issues, which are also the limitations of this study, could be promising directions to pursue in the future.

First, the number of followers is obviously a very coarse measure of a Twitter user's personal influence. The practice of dividing tweets into two groups according to the number of followers the author of each tweet has is a compromise between considering and accurately measuring users' influence while evaluating the impact of WOM on movie sales. Future research could refine the measurement of users' influence and incorporate a better influence measurement into the econometric model, which may potentially yield interesting and useful results regarding personal influence, WOM, and firm product sales.

Second, sentiment analysis is another challenge in studying the effect of online WOM on product sales in the Web 2.0 era. On the one hand, we are happy to see large amounts of WOM data because it reduces the sample bias; on the other hand, analyzing people's attitudes becomes a challenge because manually checking each WOM message is obviously not feasible. The algorithms we used to classify tweets are very effective, but far from perfect.[20] Moreover, identifying a tweet as positive, neutral, or negative offers only one dimension of WOM sentiment, although this is probably the most important dimension. Still, there might be other dimensions of WOM sentiment that are interesting to explore, such as the intention feature used in this study. After all, human language contains far more information than valence.

[20] Currently, sentiment analysis is an active research field in computational linguistics and could be particularly useful to marketing researchers.

References

Arellano, M. and S. Bond. 1991. Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations. *Review of Economic Studies*, **58(2)**, 277-297.

Chevalier, J., and D. Mayzlin. 2006. The Effect of Word of Mouth on Sales: Online Book Reviews. *Journal of Marketing Research*, **43(3)**, 345-354.

Chintagunta, P. K., S. Gopinath, and S. Venkataraman. The Effect of Online User Reviews on Movie Box Office Performance: Accounting for Sequential Rollout and Aggregation Across Local Markets. *Marketing Science*, forthcoming

Dellarocas, C., X. Zhang, and N. F. Awad. 2007. Exploring the Value of Online Product Reviews in Forecasting Sales: The Case of Motion Pictures. *Journal of Interactive Marketing*, **21(4)**, 23-45.

De Vany, A. and W.D. Walls. 1996. Bose-Einstein Dynamics and Adaptive Contracting in the Motion Picture Industry. *Economic Journal*, **106(439)**, 1493-1514.

Dhar, V., and E. A. Chang. 2009. Does Chatter Matter? The Impact of User-Generated Content on Music Sales. *Journal of Interactive Marketing*, **23(4)**, 300-307.

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

Gladwell, M. (2000): *The Tipping Point*, New York: Little, Brown and Company.

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

Katz, E., and P. F. Lazarsfeld. 1955. *Personal Influence*, Glencoe, IL: Free Press.

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

Mitchell, T.M. *Machine Learning.* McGraw-Hill, New York, 1997.

Onishi, H., and P. Manchanda. 2010. Marketing Activity, Blogging and Sales. *Working paper.*

Pai, S., S. Siddarth. 2011. The Impact of Word-of-Mouth on Purchase Decisions: The Case of Motion Pictures. *Working paper.*

Slywotzky, A. J., B. P. Shapiro. 1993. Leveraging to Beat the Odds: The New Marketing Mind-Set. *Harvard Business Review*, **71(5)**, 97-107.

Sonnier, G. P., L. McAlister, and O. J. Rutz. 2011. A Dynamic Model of the Effect of Online Communications on Firm Sales. *Marketing Science*, **30(4)**, 702-716.

Steinwart, I, and A., Christmann. 2008. *Support Vector Machines.* New York: Springer-Verlag.

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

Van den Bulte, C., and Y. V. Joshi. 2007. New Product Diffusion with Influentials and Imitators. *Marketing Science*, **26(3)**, 400-421.

Appendix

The table below is a sample of the intention lexicon used to classify tweets as intention or non-intention tweets. The the exact name of each movie is replaced by the word "movie".

Table : Sample Regular Expressions for the Intention Lexicon

i wanna (go|see)
dying to see (the)*movie
(wait|waiting) (for|4|(to see)|(to watch))
looking forward to (.*)movie
wanna (.*) see movie
(plan|need) (to|2) (watch|see|c|catch)(the)* movie
wait (until|till) (Monday|Tuesday|Wednesday|Thursday|Friday|Saturday|Sunday)
(sold|sell) out|no ticket
(go|going) to the movies (tonight |today)*to see (the *)movie
i must (watch|see)
hope(.*)as good as