DO VERIFIED CONSUMER REVIEWS AFFECT SALES? AN EMPIRICAL ANALYSIS OF MIXED REVIEW SYSTEMS IN THE FILM INDUSTRY

Research in Progress

Dongyeon Kim, Korea Advanced Institute of Science and Technology, Seoul, South Korea, dykim88@business.kaist.ac.kr

Kyuhong Park, Korea Advanced Institute of Science and Technology, Seoul, South Korea, x64576@business.kaist.ac.kr

Jae-Hyeon Ahn, Korea Advanced Institute of Science and Technology, Seoul, South Korea, jahn@business.kaist.ac.kr

Abstract

Consumers refer to reviews in online review systems to collect information for their purchasing decisions. Therefore, the introduction of 'verified purchase badges' in some review systems can make an interesting and dynamic impact on product sales. The main objective of this study is to find the differences in the effect of verified and non-verified purchase reviews on product sales from the perspective of WOM metrics (i.e., volume, valence, and helpfulness). Therefore, this study performs an empirical analysis by collecting 866,221 online movie reviews. The results show that volume and valence have a positive effect on the movie sale. However, two group analyses based on whether the 'verified purchase badge' or not show some contrasting results. Valence, helpfulness of verified reviews, and volume of non-verified reviews have positive effects on ticket sales. Furthermore, in terms of reviewers' rating patterns, as the ratio of clusters with a high ratings pattern in the verified group and the ratio of clusters with a varied ratings pattern in the mixed group increase, the effect on sales also increases. Our results break the preconception that all verified purchase reviews always affect sales.

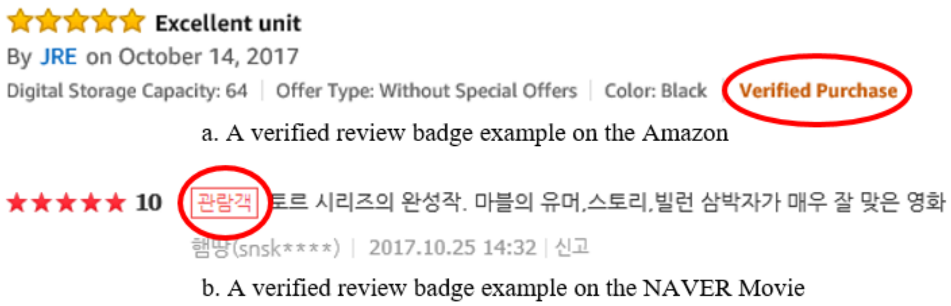
Keywords: Verified Purchase Badge, Verified Reviews, Online Review Systems, WOM, Rating Patterns, Film Industry

1 Introduction

In recent years, the growth of user-generated content (UGC) has led to a dramatic increase of online consumer reviews. This increase has been facilitated by new online review systems (or reputation systems) that provide virtual platforms that make sharing opinions easy. Given the importance and usefulness of online reviews, the credibility of online reviews has become extremely important to both consumers and companies. As such, the increasing number and salience of online review systems has attracted considerable academic and practical attention to the role of online word-of-mouth (WOM) on consumer decisions. Many previous investigations on the effect of reviews on sales or on other users have assumed that reviews have high credibility (Hung and Li, 2007; Liu, 2006; Yang et al., 2012; Ye et al., 2011). For instance, one study on movies found that moviegoers considered online reviews to be more credible and trustworthy than traditional review sources (e.g., newspaper reviews) (Bickart and Schindler, 2001). According to Nielsen's American Moviegoing Report (2014), 80 percent of the polled moviegoers consider online movie reviews a barometer for what movies to watch.

Online review systems have evolved in various ways to provide more reliable reviews to users. Typically, existing online review systems can be classified into open review systems and closed review systems (Reevoo, 2016). Open review systems (e.g., IMDb.com, Yelp.com) allow anyone to write reviews. These systems have the advantage of generating a large volume of reviews. However, open review systems potentially lower the quality and reliability of some reviews, because fake reviewers can easily write positive ratings about their own products or negative ratings about competitors' products (TIME, 2012). On the other hand, closed review systems (e.g., Airbnb.com, Booking.com) only allows users who have actually purchased a product or service to post reviews, which increases the review credibility. In summary, whereas open review systems offer accuracy, closed review systems offer breadth.

Lately, however, open review systems that have the ability to sell products or provide services to users have begun to add similar functions of closed review systems. For example, when users who buy products or services through a site write a review, a badge is displayed in the review. This type of online review system is called a mixed review system (e.g., Amazon.com, BestBuy.com). The features of this type of system include allowing anyone to post a review like open review systems while also adding a verifying badge (e.g., 'Verified Purchase' on Amazon shown in Figure 1a) to reviews written by real purchasers, like closed review systems. Best Buy also has a 'Verified Purchaser' badge and Naver Movie shows an 'Audience' badge (see Figure 1b). That is, a mixed review system is a new form that attempts to combine the advantages of open and closed review system.



a. A verified review badge example on the Amazon

b. A verified review badge example on the NAVER Movie

Figure 1. Screenshots of verified reviews. a. Amazon's 'Verified Purchase' badge (Amazon, 2017); b. Naver Movie review 'Audience' badge.

With the systematic changes made to move from the open review system to the mixed review system using verified purchase badges in some review systems, there have been few studies that examined the differences and the effects of this change. Anderson and Simester (2014) explored statistical differences between verified and non-verified purchase reviews (e.g., volume, valence, and the number of words), and Kokkodis and Lappas (2016) explored the effect of introducing verified purchase badges on the helpfulness and rating of reviews. However, the effect of differences between verified and non-verified purchase reviews on product sales has not yet been explored, as far as we understand.

To contribute to research in this area, we examine how verified reviews and non-verified reviews affect sales in the film industry. It is not easy to get sales information from most review systems that sell merchandise from different sellers, but the film industry is suitable for this study because it automatically records all ticket sales. Movies are also a good platform for verifying the effectiveness of WOM, because they are representative experience goods and a successful box office is determined in the first few weeks after a movie's initial release. As a result, we aim to explore the following research question: *Is there a difference in the effect of verified and non-verified purchase reviews on sales from the perspective of WOM (i.e., volume, valence, and helpfulness)?*

In the next section, there is related literatures. Section 3 describes the data and econometric models used for the analysis. Sections 4 and 5 describe the analysis results and conclusions, respectively.

2 Related Literature

Consumers have considered online WOM as a trustworthy source of information that can help overcome their uncertainty about a product (Gruen et al., 2006). This, in turn, can affect product sales. Thus, many people and companies are concerned about the credibility of online reviews.

A considerable body of research has studied the impact of online WOM on product sales (Chen and Xie, 2008; Gu et al., 2012; Liu, 2016; Moon et al., 2014; Senecal and Nantel, 2004; Ye et al., 2011) and sales rank (Chevalier and Mayzlin, 2006; Jabr and Zheng, 2014). These studies have found that the primary WOM characteristics that affect sales include the volume (i.e., the number of reviews) and the valence (i.e., the numerical ratings of the reviews). Considering a cognitive behavioural perspective, the volume of WOM plays a major role in promoting consumer awareness, while the valence of WOM is known to influence consumer attitudes (Liu, 2006; Mizerski, 1982).

Volume has generally been shown to have a significant impact on product sales (Duan et al., 2008; Gu et al., 2012; Liu, 2006; Moon et al., 2010; Yang et al., 2012). Research has shown that the volume of information increases consumer awareness about a product and increases sales (Godes and Mayzlin, 2004). Valence, however, has been shown in studies to have a weak or insignificant relationship on sales (Duan et al., 2008; Liu, 2006; Moon et al., 2010; Moon et al., 2014). In the Amazon.com book market, for instance, customers have found it difficult to obtain an objective valence for some books because of the possibility of manipulated reviews (Hu et al., 2012).

In addition to volume and valence, many recent studies have also attempted to investigate other variables that affect sales, such as review length and helpfulness votes (Anderson and Simester, 2014; Hu et al., 2012; Jabr and Zheng, 2014). Moon et al. (2014) examined the effect of helpfulness votes on product sales, finding that it is not significant. Nonetheless, they captured consumer heterogeneities in product evaluation by clustering the full text of the reviews.

Most online review systems do not provide monetary rewards for users who write reviews. Amazon's 'Vine program' provides pre-release items to top-ranked reviewers for free and lets them write reviews. However, the 'Verified Purchase' badge is not eligible for this monetary compensation. In the present study, differences in online status among peers due to the 'Verified Purchase' badge will be able to change the effect of the reviews (there is no effect from monetary rewards) (Goes et al., 2016; Lerner and Tirole, 2002; Roberts et al., 2006; Wasko and Faraj, 2005), and eventually it may have a different impact on product sales.

3 Data and Method

3.1 Data collection

The data for this study were collected from two sources: Naver Movie (http://movie.naver.com) and the Korean Film Council (http://www.kobis.or.kr). The Naver Movie website is the largest online movie review system in Korea. As of October of 2017, it contained about 12 million movie reviews; about 2,500 new reviews are added daily. The Korean Film Council runs an integrated movie ticket network system that stores movie sales information (e.g., daily box office receipts) from all movie theatres in Korea.

The Naver Movie site identifies verified purchase reviews posted by reviewers who actually watched the movie. This group is called 'Audience' (see Figure 1b). For this dataset, the number of verified purchase reviews was 293,098, while the number of non-verified purchase reviews was 573,123.

Our data collection for the Korean Film Council was conducted in several steps. First, a list of commercial movies released from August 2014 to July 2015 was obtained from the Korean Film Council. We excluded art movies and independent movies which is not commercial movies. Second, we narrowed this list of commercial movies to those with obtainable box office revenue records for the first two weeks following their initial release. We selected the two-week period because ticket sales during the first two weeks of a movie's release are very important to its commercial success and because the volume of reviews decreases rapidly after this time period (De Vany and Walls, 1999; Dual et al., 2008). Third, we chose movies that had at least one verified review and one non-verified review each day for two weeks. Based on these criteria, our dataset included 221 movies, 866,221 reviews, and 593,573 unique IDs.

The review data included the reviewer's unique ID, the review posting date, the rating given in the review, the full text of the review, and the helpfulness votes given by other users. In this dataset, it was able to determine how many and what reviews were submitted by each reviewer ID for the dataset period. It was also possible to calculate the daily averages for several variables, including each day's rating (i.e., valence), volume, and helpfulness votes. From the Korean Film Council, it was possible to collect information about each selected movie's title, release date, genre (SF, family, horror, drama, romance, adventure, mystery, crime, thriller, animation, action, war, comedy, and fantasy), Motion Picture Association of America (MPAA) film rating (G, PG, PG-13, and above R), distributor (top 9 companies or not), nation (America, Korea, and others), and daily box office revenue. Table 1 presents the descriptive statistics of the main variables.

Variable	Description	Mean	Min	Max
$Revenue_{it}$	The box office revenue ($) for movie *i* at day *t*	592348.14	4.43	9650028.72
$Volume_{it}$	The number of reviews for movie *i* at day *t*	277.79	2.00	5107.00
$Valence_{it}$	The average ratings of reviews for movie *i* at day *t*	7.74	1.81	10.00
$Helpfulness_{it}$	The average helpfulness ratio of reviews for movie *i* at day *t*	0.59	0.00	1.00
Note: Movie *i* is from 1 to 221 and day *t* is from 1 to 14. Helpfulness is calculated for each review by helpful votes/total votes and averaged for each movie and day.				

Table 1. Descriptive statistics.

3.2 Verified purchase econometric model

For this study, we developed panel regression equations that can predict box office revenue for a chosen movie. Previous studies have found that reviews have a significant influence on box office revenue (Duan et al., 2008; Liu, 2006; Moon et al., 2010; Moon et al., 2014; Yang et al., 2012). Liu (2006) showed that box-office revenue is affected positively by the previous week's WOM, particularly the volume of WOM. In our equation, we also included variables related to the characteristics of the reviews themselves (e.g., helpfulness ratio) (Kokkodis and Lappas, 2016; Mudambi and Schuff, 2010).

First, because the mixed review systems have a mixture of verified reviews and non-verified reviews, we considered the following model for the effect of all reviews on sales.

$$\ln(Revenue_{it}) = \alpha_0 + \alpha_1 \ln(Volume_{it-1}) + \alpha_2 Valence_{it-1} + \alpha_3 Helpfulness_{it-1} + \theta InitialScreen_i + \pi Z_i + \mu_i + \varepsilon_{it} \qquad (1)$$

where *i* indicates the movies and *t* indicates the days from 1 to 14. Investigating this early box office data can help control the effect of the distributors' ex-post supply decisions (e.g., increasing or decreasing the number of screens after a release). In addition, a daily dataset can control for more specific environmental effects (e.g., the competitive situation between films). In equation (1), the dependent variable ln($Revenue_{it}$) is the natural logarithm of the box office revenue for movie *i* at day *t*. ln($Volume_{it-1}$) is the number of reviews for movie *i* at day *t-1*. These reviews include both verified and non-verified reviews. $Valence_{it-1}$ is the average star ratings of reviews for movie *i* at day *t-1*. These ratings also did not differentiate between verified and non-verified reviews. $Helpfulness_{it-1}$ is the average helpfulness ratio of all reviews the movie *i* at day *t-1*. If we were to use actual values instead of ratios, it is possible to have trouble interpreting the results. Even if the reviews are not helpful, if the movie is popular, the actual helpfulness value will increase. For example, the movie 'Interstellar' released on November 6, 2014 and recorded 1,815 volume, 9.45 valence, and 0.70 helpfulness on the first day of release, and $2,432,967.12 USD revenue on the following day.

We also set up time-invariant variables. One of them is $InitialScreen_i$, which represents the number of screens movie i showed on during the opening day. Z_i is a movie characteristic vector that contains genre, film rate, nation, whether it is a major distributor or not, and whether the movie is a sequel or not. Lastly, μ_i captures the movie level fixed effects.

Second, from the perspective of WOM, we separated the reviews to explore the impact of verified and non-verified reviews on sales to make equation (2).

$$\begin{aligned}\ln(Revenue_{it}) = {} & \beta_0 + \beta_1 \ln(Verified\_Volume_{it-1}) + \beta_2 Verified\_Valence_{it-1} + \\ & \beta_3 Verified\_Helpfulness_{it-1} + \gamma_1 \ln(Non\_Verified\_Volume_{it-1}) + \\ & \gamma_2 Non\_Verified\_Valence_{it-1} + \gamma_3 Non\_Verified\_Helpfulness_{it-1} + \\ & \theta InitialScreen_i + \pi Z_i + \mu_i + \varepsilon_{it} \end{aligned} \tag{2}$$

The variables i, t, Z_i, $InitialScreen_i$, and $\ln(Revenue_{it})$ are the same as in equation (1), but volume, valence, and helpfulness were divided into two based on the audience badge (or verified purchase badge). The volume, valence, and helpfulness of reviews with badges are identified as $\ln(Verified\_Volume_{it-1})$, $Verified\_Valence_{it-1}$, and $Verified\_Helpfulness_{it-1}$, respectively. On the other hand, reviews without badges are represented by $\ln(Non\_Verified\_Volume_{it-1})$, $Non\_Verified\_Valence_{it-1}$, and $Non\_Verified\_Helpfulness_{it-1}$, respectively. For example, the movie 'Interstellar' recorded 141 verified volume, 1,674 non-verified volume, 9.46 verified valence, 9.45 non-verified valence, 0.65 verified helpfulness, and 0.70 non-verified helpfulness on the first day of release.

3.3 Reviewer behaviour econometric model

Since simply dividing into verified reviews and non-verified reviews does not take into account reviewer behaviour, we added users' rating patterns to equations (1) and (2). During the dataset period, reviewers were divided into three groups: people who wrote only verified reviews, people who wrote only non-verified reviews, and people who mixed the two types. We defined each group as 'verified group', 'non-verified group', and 'mixed group', respectively.

Next, a k-means cluster analysis was conducted to identify patterns for how reviewers in each group assigned their ratings. Because the total number of reviews by each reviewer was different, the number of reviews by star rating was then converted into a ratio. In the k-means cluster analysis, it is also important to determine the number of clusters k. Bapna et al. (2004) proposed a dissimilarity ratio that is calculated in a manner that divides the inter-cluster distance from the intra-cluster distance. When the dissimilarity ratio is largest, k represents the most-valuable clustering result. In this study, the largest dissimilarity ratio was derived when k was three in all groups. The clusters representing the rating patterns for each group were as follows (see Figure 2).

Using the clustering results, we developed equations (3) and (4) to explore how the change in the relative ratio of each cluster affected the sales. Equation (3) adds the result of clustering the entire reviewers based on equation (1), and equation (4) adds the result of clustering reviewers of each group based on equation (2) as independent variables.

$$\begin{aligned}\ln(Revenue_{it}) = {} & \alpha_0 + \alpha_1 \ln(Volume_{it-1}) + \alpha_2 Valence_{it-1} + \alpha_3 Helpfulness_{it-1} + \\ & \lambda_j \textstyle\sum_{j=2}^{n} Entire\_cluster\_ratio^1_{ijt-1} + \theta InitialScreen_i + \pi Z_i + \mu_i + \varepsilon_{it} \end{aligned} \tag{3}$$

$$\begin{aligned}\ln(Revenue_{it}) = {} & \beta_0 + \beta_1 \ln(Verified\_Volume_{it-1}) + \beta_2 Verified\_Valence_{it-1} + \\ & \beta_3 Verified\_Helpfulness_{it-1} + \gamma_1 \ln(Non\_Verified\_Volume_{it-1}) + \\ & \gamma_2 Non\_Verified\_Valence_{it-1} + \gamma_3 Non\_Verified\_Helpfulness_{it-1} + \\ & \delta_j \textstyle\sum_{j=2}^{o} Verified\_cluster\_ratio^1_{ijt-1} + \\ & \zeta_j \textstyle\sum_{j=2}^{p} Non\_verified\_cluster\_ratio^1_{ijt-1} + \\ & \eta_j \textstyle\sum_{j=2}^{q} Mixed\_cluster\_ratio^1_{ijt-1} + \theta InitialScreen_i + \pi Z_i + \mu_i + \varepsilon_{it} \end{aligned} \tag{4}$$

n, o, p, and q in the sigma indicate the number of clusters in each group. For example, when o is 3, $Verified\_cluster\_ratio^1_{i2t-1}$ means the ratio of cluster 2 (target cluster) to cluster 1 (base cluster), and

$Verified\_cluster\_ratio^{1}_{i3t-1}$ means the ratio of cluster 3 (target cluster) to cluster 1 (base cluster). We used relative ratios because using simple ratios for each cluster could lead to endogeneity issues.

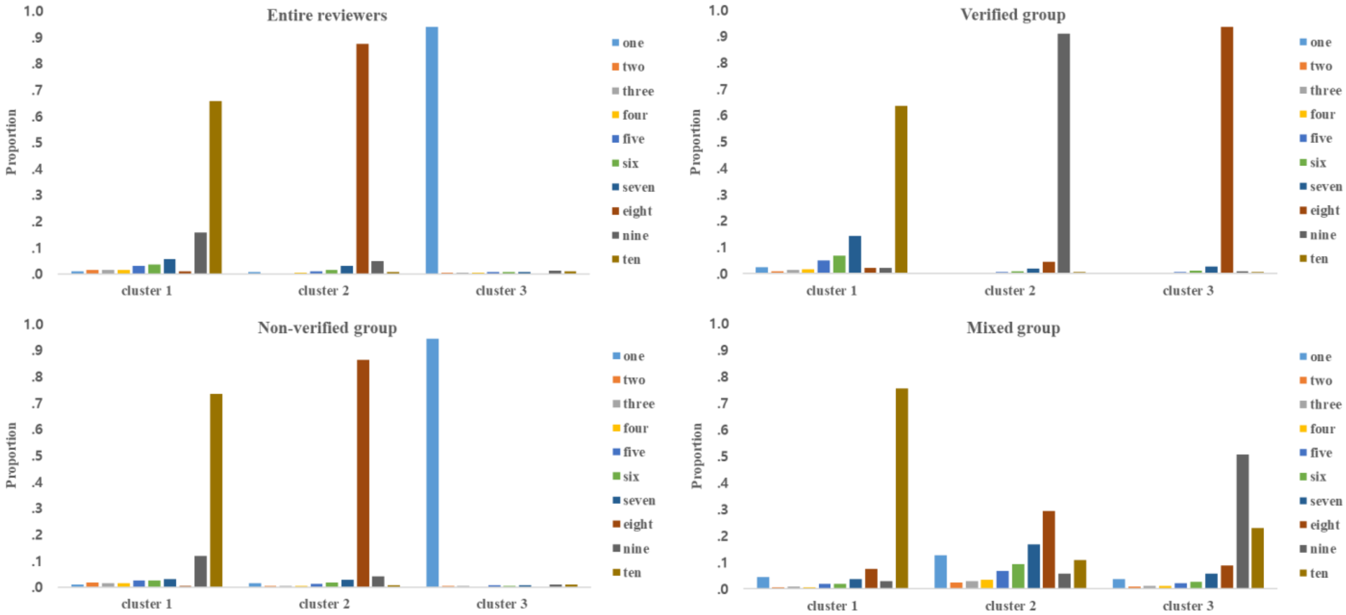


Figure 2. Rating patterns for each group.

4 Results

A general ordinary least squares fixed effects estimation was not suitable because these equations' post-diagnosis results have the following problems: heteroscedasticity, autocorrelation, and cross-sectional dependency. That is, it is a biased standard error estimation with a covariance matrix estimation. Statistics based on such a standard error have no effect. To solve this problem, Driscoll and Kraay (1998) proposed the fixed effects regression models with Driscoll-Kraay standard errors using an estimator based on the non-parametric covariance matrix. This estimator makes the equation robust in the cross-sectional dependency based on the heteroscedasticity-consistent standard errors and in the autocorrelation based on a particular lag (Hoechle, 2007).

The estimation results are presented in Table 2. From the viewpoint of all reviews in the mixed review system, volume and valence have a significant positive impact on sales, but helpfulness is not significant (see Models 1 and 3 results). However, when all reviews were divided into verified and non-verified reviews, the verified reviews' valence and helpfulness showed a significant positive impact on sales. Only the non-verified reviews' volume had a significant impact (see Models 2 and 4 results).

Verified reviews ensure that people have really watched the movie. In this case, valence and helpfulness of their reviews have a significant impact on box office. On the other hand, those who write reviews without watching movies or with purchasing a ticket on another site are classified as non-verified reviews. Because volume is one indicator of people's interest, non-verified reviews that many people can write reviews have a significant impact on box office.

From the viewpoint of the reviewers' rating patterns (when clustered for all users), it was found that the increase in the ratio of cluster 2 (which wrote most of the 9-star ratings) compared to cluster 1 (which wrote most of the 10-star ratings) had a significant negative effect on sales. Interestingly, even if the ratio of cluster 3 (which wrote most of the 1-star ratings) increased compared with cluster 1, it did not have a significant effect. This can be interpreted as that the users who used the mixed review system did not consider the low score, such as 1 point.

The results of the impact on sales when the reviewers were divided into the verified group, the non-verified group, and the mixed group (see Model 4) are as follows. First, the higher the ratio of cluster 3 (which wrote most of the 8-star ratings) compared to cluster 1 (which wrote most of the 10-star ratings) in the verified group, the more negative impact they had on sales. This looks natural because the average rating is lowered, but the users did not make difference for 10-star ratings and 9-star ratings. Second, there was no significant difference in the non-verified group. From the perspective of all reviews (see Model 1), valence did not affect sales, which is in line with the previous interpretation only having a volume effect. Lastly, the higher the ratio of cluster 3 compared to cluster 1 in the mixed group, the more positive the effect on sales. In the mixed group, the characteristics of the verified group and the non-verified group appear simultaneously because the verified reviews are mixed. It seems that the various star ratings influenced sales rather than merely having high star ratings reviews.

	Model 1	Model 2	Model 3	Model 4
Variable	Coef. (S.E.)	Coef. (S.E.)	Coef. (S.E.)	Coef. (S.E.)
$\ln(Volume_{it-1})$	1.01*** (.16)		.99*** (.16)	
$\ln(Verified\_Volume_{it-1})$		.24 (.13)		.23 (.13)
$\ln(Non\_Verified\_Volume_{it-1})$		.75*** (.13)		.74*** (.13)
$Valence_{it-1}$	.09** (.03)		.08*** (.02)	
$Verified\_Valence_{it-1}$		.09** (.03)		.10*** (.02)
$Non\_Verified\_Valence_{it-1}$		.02 (.02)		-.01 (.01)
$Helpfulness_{it-1}$	.08 (.12)		.09 (.13)	
$Verified\_Helpfulness_{it-1}$		.20* (.07)		.19* (.07)
$Non\_Verified\_Helpfulness_{it-1}$		-.16 (.11)		-.15 (.11)
$Entire\_Cluster\_Ratio_{it-1}$ (cluster 2 against base cluster 1)			-.56* (.21)	
$Entire\_Cluster\_Ratio_{it-1}$ (cluster 3 against base cluster 1)			-.03 (.18)	
$Verified\_Cluster\_Ratio_{it-1}$ (cluster 2 against base cluster 1)				.05 (.05)
$Verified\_Cluster\_Ratio_{it-1}$ (cluster 3 against base cluster 1)				-.16** (.04)
$Non\_Verified\_Cluster\_Ratio_{it-1}$ (cluster 2 against base cluster 1)				-.04 (.07)
$Non\_Verified\_Cluster\_Ratio_{it-1}$ (cluster 3 against base cluster 1)				-.16 (.09)
$Mixed\_Cluster\_Ratio_{it-1}$ (cluster 2 against base cluster 1)				.03 (.03)
$Mixed\_Cluster\_Ratio_{it-1}$ (cluster 3 against base cluster 1)				.08** (.02)
Within R-squared	0.510	0.514	0.513	0.518
Observations	3,094	3,094	3,094	3,094
Note: $^{***}p$<0.001, $^{**}p$<0.01, $^{*}p$<0.05. Movie *i* is from 1 to 221 and day *t* is from 1 to 14. Observations is derived from multiplying 221 movies by 14 days. DV is $\ln(Revenue_{it})$. S.E. is Driscoll-Kraay standard errors.				

Table 2. Fixed-effects regression results.

5 Conclusions and Future Directions

The main objective of this study was to find a difference in the effects of verified and non-verified purchase reviews on sales from the perspective of WOM (i.e., volume, valence, and helpfulness). To do that, we first compared the verified and non-verified (purchase) reviews in the mixed review system through empirical models, deriving the different effects between them. The results showed that verified reviews' valence and helpfulness and non-verified reviews' volume have a positive effect on movie sales. Furthermore, we analysed the reviewers' rating patterns of three groups: verified group, non-verified group, and mixed group. The higher the ratio of clusters having high ratings in the verified group and clusters having various ratings in the mixed group, the more positive the effect on sales. This result breaks the generalized conclusion that simply all verified purchase reviews have a positive impact on sales.

This study has several implications. First, for academic researchers, this is the first study to examine the impact of verified purchase reviews on sales. This suggests that investigations into the effect of WOM on sales should consider not only the type of reviews but also reviewers' rating patterns as important factors. Second, review system managers should try to show more useful reviews to users first. Most mixed review systems have sorted reviews by verified purchase badge and high helpfulness scores. However, they can refer to this study to sort differently according to the rating pattern of the reviewers or the purpose of the review systems' users. For example, for users who want more reliable reviews, managers can use reviews of clusters with high ratings in the verified group and reviews of the clusters with various ratings in the mixed group.

One of the limitations of this study is that we did not explore the effect of the verified purchase badge on sales in terms of system changes. That is, when the open review system is changed to a mixed review system, the effect can be measured by comparing the users who use the verified purchase badge with the users who do not use the badge. Our future research will explore this effect through the diff-in-diff methodology. Second, we analysed clusters with different patterns, but we cannot really know the specific contents each group mentioned. Our future research will make the results more robust by comparing the text for each group through text mining. We may also discover users who have intentionally breached the accuracy of sales forecasts. Lastly, we will increase the robustness of the current results by (1) comparing reviewers who frequently write reviews with reviewers who do not (i.e., heavy users vs. light users), and (2) comparing movies that had successful box office results with movies that did not.

References

Amazon (2017). *Verified Purchase Reviews*. URL: http://www.amazon.co.uk/gp/community-help/amazon-verified-purchase (visited on 11/22/2017).

Anderson, E. T. and D. I. Simester (2014). "Reviews without a Purchase: Low Ratings, Loyal Customers, and Deception." *Journal of Marketing Research* 51 (3), 249-269.

Bapna, R., P. Goes, A. Gupta, and Y. Jin (2004). "User Heterogeneity and Its Impact on Electronic Auction Market Design: An Empirical Exploration." *MIS Quarterly* 28 (1), 21-43.

Bickart, B. and R. M. Schindler (2001). "Internet Forums as Influential Sources of Consumer Information." *Journal of Interactive Marketing* 15 (3), 31-40.

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

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

De Vany, A. and W. D. Walls (1999). "Uncertainty in the Movie Industry: Does Star Power Reduce the Terror of the Box Office?" *Journal of Cultural Economics* 23 (4), 285-318.

Driscoll, J. C. and A. C. Kraay (1998). "Consistent Covariance Matrix Estimation with Spatially Dependent Panel Data." *Review of Economics and Statistics* 80 (4), 549-560.

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

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

Goes, P. B., C. Guo and M. Lin (2016). "Do Incentive Hierarchies Induce User Effort? Evidence from an Online Knowledge Exchange." *Information Systems Research* 27 (3), 497-516.

Gruen, T. W., T. Osmonbekov and A. J. Czaplewski (2006). "eWOM: The Impact of Customer-to-Customer Online Know-How Exchange on Customer Value and Loyalty." *Journal of Business Research* 59 (4), 449-456.

Gu, B., J. Park and P. Konana (2012). "The Impact of External Word-of-Mouth Sources on Retailer Sales of High-Involvement Products." *Information Systems Research* 23 (1), 182-196.

Hoechle, D. (2007). "Robust Standard Errors for Panel Regressions with Cross-Sectional Dependence." *Stata Journal* 7(3), 281-312.

Hu, N., I. Bose, N. S. Koh, and L. Liu (2012). "Manipulation of Online Reviews: An Analysis of Ratings, Readability, and Sentiments." *Decision Support Systems* 52 (3), 674-684.

Hung, K. H. and S. Y. Li (2007). "The Influence of eWOM on Virtual Consumer Communities: Social Capital, Consumer Learning, and Behavioral Outcomes." *Journal of Advertising Research* 47 (4), 485-495.

Jabr, W. and Z. Zheng (2014). "Know Yourself and Know Your Enemy: An Analysis of Firm Recommendations and Consumer Reviews in a Competitive Environment." *MIS Quarterly* 38 (3), 635-654.

Kokkodis, M. and T. Lappas (2016). "The Relationship between Disclosing Purchase Information and Reputation Systems in Electronic Markets." *Thirty Seventh International Conference on Information Systems, Dublin*.

Lerner, J. and J. Tirole (2002). "Some Simple Economics of Open Source." *The Journal of Industrial Economics* 50 (2), 197-234.

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

Mizerski, R. W. (1982). "An Attribution Explanation of the Disproportionate Influence of Unfavorable Information." *Journal of Consumer Research* 9 (3), 301-310.

Moon, S., P. K. Bergey, and D. Iacobucci (2010). "Dynamic Effects among Movie Ratings, Movie Revenue, and Viewer Satisfaction." *Journal of Marketing* 74 (1), 108-121.

Moon, S., Y. Park and Y. S. Kim (2014). "The Impact of Text Product Reviews on Sales." *European Journal of Marketing* 48 (11/12), 2176-2197.

Mudambi, S. M. and D. Schuff (2010). "What Makes a Helpful Online Review? A Study of Customer Reviews on Amazon.com." *MIS Quarterly* 34 (1), 185-200.

Nielsen (2014). *Two Thumbs Up: Moviegoing at a Theater near You*. URL: http://www.nielsen.com/om/en/insights/news/2014/twothumbsup-moviegoing-at-a-theater-near-you.html (visited on 11/22/2017).

Reevoo (2016). *6 Reasons to Opt for a Closed Review System*. URL: https://blog.reevoo.com/6-reasons-you-should-opt-for-a-closed-review-system/ (visited on 11/22/2017).

Roberts, J. A., I. H. Hann and S. A. Slaughter (2006). "Understanding the Motivations, Participation, and Performance of Open Source Software Developers: A Longitudinal Study of the Apache Projects." *Management Science* 52 (7), 984-999.

Senecal, S. and J. Nantel (2004). "The Influence of Online Product Recommendations on Consumers' Online Choices." *Journal of Retailing* 80 (2), 159-169.

TIME (2012). *9 Reasons Why You Shouldn't Trust Online Reviews*. URL: http://business.time.com/2012/02/03/9-reasons-why-you-shouldnt-trust-online-reviews/ (visited on 11/22/2017)

Wasko, M. M. and S. Faraj (2005). "Why Should I Share? Examining Social Capital and Knowledge Contribution in Electronic Networks of Practice." *MIS Quarterly* 29 (1), 35-57.

Yang, J., W. Kim, N. Amblee and J. Jeong (2012). "The Heterogeneous Effect of WOM on Product Sales: Why the Effect of WOM Valence is Mixed?" *European Journal of Marketing* 46 (11), 1523-1538.

Ye, Q., R. Law, B. Gu and W. Chen (2011). "The Influence of User-Generated Content on Traveler Behavior: An Empirical Investigation on the Effects of e-Word-of-Mouth to Hotel Online Bookings." *Computers in Human Behavior* 27 (2), 634-639.