The Impact of Social Media Activities on Theater Demand

Baldin, Andrea; Bille, Trine; Mukkamala, Raghava Rao; Vatrapu, Ravi

Document Version
Accepted author manuscript

Published in:
Journal of Cultural Economics

DOI:
10.1007/s10824-023-09480-z

Publication date:
2024

License
Unspecified

Citation for published version (APA):
Baldin, A., Bille, T., Mukkamala, R. R., & Vatrapu, R. (2024). The Impact of Social Media Activities on Theater Demand. *Journal of Cultural Economics*, *48*(2), 199-220. https://doi.org/10.1007/s10824-023-09480-z

The Impact of Social Media Activities on Theater Demand

Andrea Baldin*[1,2], Trine Bille†[2], Raghava Rao Mukkamala‡[3], and Ravi Vatrapu§[4]

[1]*Department of Economics, Ca' Foscari University of Venice, Cannaregio 873, 30121 Venice, Italy*
[2]*Department of Business Humanities and Law, Copenhagen Business School, Porcelænshaven 18B, DK-2000 Frederiksberg, Denmark*
[3]*Department of Digitalization, Copenhagen Business School, Howitzvej 60, DK-2000 Frederiksberg, Denmark*
[4]*Centre for Digital Enterprise Analytics and Leadership (DEAL), Ted Rogers School of Management, Toronto Metropolitan University, Canada.*

*andrea.baldin@unive.it (ORCID: 0000-0002-4386-2423)
†tb.bhl@cbs.dk (ORCID: 0000-0002-0139-166X)
‡rrm.digi@cbs.dk (ORCID: 0000-0001-9814-3883)
§vatrapu@ryerson.ca (ORCID: 0000-0002-9109-5281)

Abstract

A well-known factor in the consumption of cultural goods is that demand is subject to the 'nobody knows' principle and therefore difficult to predict. Other sectors have successfully analyzed social media data to predict real-world outcomes; the cultural field has applied this type of data analysis in the context of movies. This paper is the first study to consider the impact of electronic word of mouth (eWOM) generated via social media in the context of performing arts. Compared to conventional word-of-mouth mechanisms, social media sites may further reduce the uncertainty caused by the 'nobody knows' principle by propagating an enormous amount of enduring and real-time information and opinions. This paper aims to test the potentiality of social media in understanding theater demand by combining booking data for the period 2010-2016 from the sales system of the Royal Danish Theater with volumetric data extracted from the theater's official Facebook Page. In particular, we take into account the different possible relationships between the feedback provided by social media (in terms of 'likes' and comments) and the purchase of tickets by consumers: 1) eWOM influences tickets sale; 2) no causal relationship between eWOM and tickets sale as both reflect unobserved characteristics of the theater production; 3) tickets sale influence eWOM activities; 4) ticket sale influence eWOM which in turn influence ticket sale and so on. The results suggest that only the number of likes, rather than the Facebook comments, is related to the decision to purchase a ticket. In particular, there is a mutual interaction between the number of likes given to posts specifically dedicated to a given production and the number of tickets sold concerning that specific production: eWOM activity (in terms of "like") influence the tickets sale, which in turn generate eWOM activity. With this study we aim to show how social media data can constitute a new and effective tool for understanding theater demand.

Keywords: theater demand, social media, word of mouth, panel data

JEL codes: C33, C53, Z10,

1. Introduction

The purpose of this paper is to analyze the effect of social media on the demand for performing arts. This issue can be of particular relevance in analyzing the consumption of performing arts. Indeed, performing arts provide an illustrative example of how consumer behavior in the cultural sector does not match standard assumptions of neoclassical economic theory. One of the main assumptions of the neoclassical theory considers customers to be fully and costless informed about the market. However, as we know, theater performances are experience goods, the quality of which is unknown before being experienced. This leads to the uncertain nature of the demand for creative goods, the 'nobody knows' principle (Richard Caves, 2000). From the supply perspective, it is difficult to understand why people choose to consume what they do, and hence to predict attendance of any given show. Similarly, it is difficult for customers to assess the quality and the value of a cultural product before committing to consume it; all the information acquired about the quality of a theatrical production can thus have a crucial role in the purchase-decision context, determining box-office success or failure. Consumers may therefore use different quality indicators in order to evaluate the quality beforehand. These quality indicators can be objective facts (such as the title, the author, the director, or the actors, etc.) or subjective quality indicators such as professional reviews (e.g. in newspapers), or evaluations made by friends and relatives (word-of-mouth mechanisms).

Because of the rapid development of the social media, the electronic word of mouth (henceforth: eWOM) based on social media and online community has attracted increased attention from researchers who analyze word-of-mouth (henceforth: WOM) mechanisms. Social media can be construed as a form of collective wisdom (Asur and Huberman, 2010) since they propagate an enormous amount of highly variable and real-time information through large user communities; they allow users to share information about feelings, emotional moods, and opinions relating to any experience whatsoever, including that of attending a theatrical performance.

Given that social media are sources of big data, several researchers have explored the possibility of exploiting the information inferred from Twitter, Facebook, YouTube, etc., and web search exploration tools such as Google Trends and Yahoo! search query logs, to predict real-world outcomes. Such studies include the prediction of iPhone sales (Lassen et al., 2014), car sales (Voortman, 2015), and political election results (Franch, 2013). For a systematic literature review on predictive analytics with big social data we refer to Lassen et al. (2017). The cultural field has applied this data analysis in the context of movies, investigating the effects of online reviews collected via YouTube, Yahoo! Movies, etc., on movie revenue and audience (Duan et al., 2008; Chu et. al, 2016; Oh et al., 2017; Basuroy et al., 2020), and examining the predictive power of social media such as Twitter (Asur and Huberman, 2010).

We can presume that the application of these tools might be more effective in areas characterized by products with a (almost) global market, such as iPhones, automobiles, and movies. This is not the case for performing arts: each production refers to the institution that supplies it. A *Traviata* staged at a theater in one country might be unlike the *Traviata* staged at a theater in another country, where the potential audience might also have a different profile. Information regarding the perceived quality of a theatrical production can be gathered from reviews (reflecting the experts' opinions) or audience surveys; the former is found to have no significant effect on the audience (see Urrutiaguer, 2002), whereas the latter is a time-consuming activity for the theater, and not readily available to researchers. Despite limitations in the performing arts context, social media may still represent an additional instrument for the straightforward collection of a larger amount of information concerning audience impressions of a production: this information is exchanged in the users' community, leading us to expect that the diffusion of eWOM is faster and more effective than the traditional WOM mechanism. Moreover, given the ample audience of online communities, social media may have a role in strengthening the so-called 'bandwagon effect', this being a typical characteristic of cultural consumption that essentially indicates how consumers, in order to reduce search and information costs, follow the crowd and make choices based on the choices of the majority.

This paper aims to test whether social media data related to theater productions have an impact on the number of tickets sold, or vice versa. We do this by combining booking data from the Royal Danish Theater with information from the theater's official Facebook Page for the 2010/2011 season to the 2015/2016 season.[1] In particular, we investigate the nature of the relationship between the feedback provided by social media (in terms of likes and comments), and the theater audience. Four possibilities are considered: first, the relationship

[1] Generally, the season starts in September and ends the following June.

between eWOM and audience suggests causality: the information that eWOM provides to the customers affects their decisions to attend or not attend the production. Second, a correlation between eWOM and attendance reflects unobserved production-specific characteristics (e.g. the quality of the performance, stage design etc.) meaning that these characteristics simultaneously determine the audience's quality perception and demand for the given production, thus such correlation does not imply causation. Third, the purchase of tickets generates eWOM activities, thus the causality relationship is reversed compared to the first hypothesis. Fourth, there exists an interdependent relationship between eWOM and total audience: past and current audience generate eWOM which in turn influences the current and future audience, who will affect the current and future eWOM, and so on: in this case both the first and the third hypothesis hold.

These different perspectives are analyzed respectively by means of three modelling approaches: panel data, a simultaneous equation system, and a vector autoregressive (VAR) model followed by a Granger causality test. The structure of the paper can be outlined as follows: Section 2 reviews the literature on the adoption of quality indicators in estimating theater demand; Section 3 provides the context of our study; Section 4 presents the dataset and the variables considered; Section 5 illustrates the methodology used together with a discussion of the results obtained; Section 6 concludes the paper with some economic and managerial implications.

2. Literature Review

The demand for performing arts is a widely analyzed topic in cultural economics. One of the main difficulties in the empirical studies concerns the inclusion of qualitative production characteristics that, in the case of 'higher arts', seem to be decisive for the audience. In the literature on the subject, different quality indicators have been used, as summarized in Table 1.

Some authors (Jenkins and Austen-Smith, 1987; Krebs and Pommerehne, 1995) have focused on the concept of 'highbrow' and 'lowbrow' productions as a quality indicator, providing evidence of the positive impact of 'lowbrowness' in season programs (measured as the ratio of works with more than 75 performances to all works played in a season) on theater demand.

Concerning the inclusion of subjective quality indicators, Throsby's pioneering works (1980 with Nielsen; 1990) incorporate reviewers' opinions, developing four variables that assess the press review opinions related to various 'technical' dimensions (source material, production, acting, and design) and showing their impact on demand.

Table 1: Quality indicators for theater performances in the literature. Adapted from Grisolia and Willis (2012)

Variable	**Description**	**Authors**
Quality review	Press review using a scale from 1 (very poor) to 5 (excellent)	Throsby (1990); Abbé-Decarroux (1994); Colbert *et al.*(1998); Corning and Levy (2002); Baldin and Bille (2023)
In-house production	Whether it is an in-house production	Abbé-Decarroux (1994); Colbert *et al.*(1998)
Reputation of the author, play, producer, and cast	Known/unknown	Abbé-Decarroux (1994); Akdede and King (2006); Willis and Snowball (2009)
Reputation of the theater	Perception of the quality of the theater (through loyalty of the attendance)	Urrutiaguer (2002)
Reputation of the director	Whether the director manages a theatrical institution or not	Urrutiaguer (2002); Willis and Snowball (2009)
Quality of the stage design and costumes	Expenses for stage set and costumes per production	Zieba (2009)
Repertoire classification	Classic, Modern, Contemporary, or Atypical play	Abbé-Decarroux (1994); Corning and Levy (2002); Urrutiaguer (2002)
Popularity	Number of times the production is performed worldwide	Baldin *et al.* (2018)

Tobias (2004) proposes a new framework for aggregating ordinal expert judgments into a single quality judgment, using it in a regression to find links between the quality measurement and economic variables. The author finds that production costs are positively related to this new quality indicator for opera and ballet, giving evidence of positive but decreasing marginal returns; for drama, however, production costs are not a factor in explaining experts' perceptions of quality.

Urrutiaguer (2002) proposes that opposing opinions on quality are the reason why regression models often reveal a low statistical significance for the quality variables used to explain theatrical demand. A regression equation is constructed in order to explain demand, with continuous variables for price and volume, and with dummy variables for drama critics, directors, growth in funding by public authorities, and repertoire classification. Detailed data on the demand for French 'theatrical institutions' in 1995 and 1996 are used to test the model. To some extent, the results support the hypothesis that the media reputation of shows, as expressed in the form of reviews, and the artistic reputation of directors, as listed in the program, have opposite effects on attendance. However, the most reliable sign of quality remains the reputation of the theatrical institution.

Finally, Baldin and Bille (2023) assess the impact of professional reviewers' evaluations (on a scale from 1-5, from reviews in newspapers) and audiences' evaluations (from surveys of attendees) by considering the heterogeneity among consumers as well as experts. They show that there are consistent and significant patterns among audiences' and critics' evaluations within the different segments.

Following this line of research, the current research paper aims to clarify the role of social media data with respect to their function in predicting and/or influencing theater demand, taking the Royal Danish Theater as a case study. To the best of our knowledge, this is the first study that considers the role of social media in a performing arts context. Remaining in the broad cultural sector, the effect of eWOM and social media has been analyzed with regard to the movie industry: Asur and Huberman (2010) predict with high accuracy the box-office revenues of movies in advance of their release, using the rate of chatter on Twitter and sentiment analysis of the tweets. Duan et al. (2008) develops a simultaneous equation system to capture the interrelationship between eWOM, collected through online reviews, and movie sales, demonstrating that eWOM is both precursor and outcome of movie sales. Chu et al. (2016) discusses the role of audience movie reviews, collected through online fora, on movie attendance, distinguishing between the influencer and predictor roles. The authors conclude that positive online reviews have both a positive prediction and a positive influencing effect; the negative reviews have a negative prediction effect and a positive influencing effect; the neutral reviews have both negative prediction and negative influencing effects.

3. The Royal Danish Theater

The Royal Danish Theater was founded in 1748 and is the Danish national theater. It has three main venues in Copenhagen: The Old Stage (Gamle Scene, built in 1874), the Royal Opera House (Operaen, opened in 2005), and the Royal Playhouse (Skuespilhuset, opened in 2008). Both the Opera House and the Playhouse have a main stage and smaller stages for experimental productions. The Royal Danish Theater is one of the few national theaters in the world offering a program of opera, ballet, and drama performances, as well as classical concerts.

The government agreement with the Royal Danish Theater states that it is the national theater for the whole country and the entire population; the theater has an obligation to produce a broad repertoire of high artistic quality within ballet, opera, and plays; it is obligated to continue the classical traditions as well as developing the performing arts in new and contemporary ways; there is special focus on productions of Danish origin.

The Royal Danish Theater receives funding from the state budget, under the Ministry of Culture, and has a number of more specific obligations in agreement with the current Minister of Culture. Included in these obligations are general cultural policy goals: performing special productions for children and young people, for example, and keeping ticket prices at a level that ensures the theater is accessible to all socio-economic groups.

In 2017, the theater had a total budget of DKK 689.2 million (€ 93 million): 75% in the form of government subsidy; DKK 171.4 million (€ 23 million) in own earnings, of which 74% (€ 17 million) was from ticket sales; the rest was income from sponsors, etc.

A total of 488,668 tickets were sold by the theater in 2017. The principal draw for attendees was plays, which accounted for 41% of the total number of tickets; opera accounted for 25%, while ballet accounted for 23%, and the final 11% of tickets sold were for classical concerts.

Due to its obligations as a national theater, the Royal Danish Theater repertoire has to be based on a number of parameters: quality and variety, understood as a fairly large number of productions from the classical repertoire as well as new productions, developing the performing arts, and Danish as well as international productions from the world repertoire.

Furthermore, the theatre has to decide the number of performances of each production during the season, as well as how they are scheduled in respect of weekdays and weekends. There will be a loss in earnings if a given production is scheduled for fewer performances than demanded by the audience, or for more performances than demanded by the audience (empty seats). Putting a new production of stage involves high fixed costs (due to rehearsal time, designing and building the stage set, etc.), but the costs involved in extending a production run with extra performances are small, and the marginal costs are lower than the marginal revenue (Bille Hansen, 1991). Finally, the Royal Danish Theater has to decide its pricing policy, including price differentiation based on diverse audience groups (young people and seniors, for example), as well as the time of performance, seating categories, the genre of performance, production costs, etc.

4. Dataset and variables

We have built a panel dataset where the ticket sale and Facebook (FB) activities of the Royal Danish Theater productions are observed on a weekly basis. We have considered the 128 productions from season 2010/2011 to season 2015/2016 that have been mentioned in at least one FB post. As shown in Table 2, we notice that the number of productions with at least one FB post increases over the years, especially in the last two seasons under consideration. This is more evident when looking at the percentage of productions among all the productions that took place in a given season with at least one FB post, going from 26% in the 2010/2011 season to 67% in the 2015/2016 season. This indicates the growing importance and awareness of social media as a marketing tool.

Table 2: Productions with at least one FB post

Season	No. of productions	Productions with FB post	Percentage
2010-2011	30	8	26%
2011-2012	48	18	38%
2012-2013	41	14	34%
2013-2014	55	20	36%
2014-2015	54	27	50%
2015-2016	61	41	67%

Our dataset combines two different sources: booking data from the sales system of the Royal Danish Theater and volumetric eWOM data extracted from the official Facebook Page of the Royal Danish Theater which, at the time of analysis, had approximately 64,000 followers. The former provides us information on the price and number of tickets sold for each day of the sale period and for each of the productions staged during the period considered. Furthermore, we disaggregate the total number of tickets sold by considering the discounted ticket for students/under 25 years old people, allowing us to verify whether the social media effect differs for this particular segment of the audience.

The Social Data were collected using our custom-built, in-house tool: Social Data Analytics Tool (SODATO) (Hussain and Vatrapu, 2014a, 2014b; Vatrapu et al., 2016). The raw dataset consisted of around 311,220 data points. Each row is equivalent to an action on the Royal Theater's Facebook Page. The data are categorized into dimensions, which include both ordered and categorical data, and contain information about user ID, timestamp, event name, username, type of post (whether it is a post, comment, like), and, if relevant, links and text value of post and comment. The dataset contains approximately 4,000 posts, 35,000 comments, 10,000 comment replies, likes on comment, and likes on comment replies. The rest of the data items are likes on the posts, approximately 260,000 likes were the most dominant social media action performed by the users of the Royal Danish Theater Facebook Page. The total downloaded dataset contains information from 2008 until 2017; however, we have only used the relevant social data regarding the performances and productions that were considered for this paper.

We chose FB rather than other social media (such as Twitter and Instagram) because it is the primary social media platform used by the Royal Danish Theater to spread awareness about its events. For example, their Facebook page is being followed by 100,231 people (as of March 2023), almost twice the number of followers of the Instagram page (52,600) while their Twitter profile is followed by 4,660 followers. Moreover, engagement on Facebook is also quite high (4000 posts during the analysis period of 2010-2016) compared to their activity on Instagram and Twitter, respectively 2058 post and 2,625 tweets in the whole period 2008-2023.Therefore, we have chosen to focus on the Facebook data for the analysis.

The main variables derived from the merged dataset are: the number of tickets sold in the week for the given production, and the social media activities related to a given production in the week. Among the latter we consider the number of post published, likes and comments. In addition, we have carried out a sentiment analysis on the text of the comments using a domain-specific custom classifier built using supervised machine-learning algorithms. The comments are classified as positive (for example "*Fantastic opera and nice nice production. Enjoyed the singers.....suggested to all opera lovers*), neutral (for example "*What days do you play the Barselsstuen?*) and negative (for example, "*I'm disappointed with the composition of our mixed subscription, so I'm afraid it won't be repeated this year*"). Out of a total of 11,835 comments, most of them are neutral (77. 7%), 22.1% are positive while there are very few negative comments (0.2%). As the number of likes depends on the number of posts (a positive relationship is confirmed by a coefficient correlation r = 0.59), we consider also the average number of likes per post in the week. We don't find the same relationship between number of post and number of comments (r = - 0.02). A possible explanation is that putting a like is a more automatic action undertaken when the content of the post is appreciated by the users, while writing a comment is a more reasoned action undertaken when the user has something to say, regardless of the number of post dedicated to a given topic.

Intuitively the number of likes and comment is expected to increase together with the increase of FB followers (the potential audience that can see the post and possibly give a like or write a comment). Thus, we include also the variable *FB WEEK* indicating the number of weeks passed since the theater opened the FB Page, as the number of followers normally increases over time. We also consider the different types of post, distinguishing 5 types: link to a website, status, photo, video, and events.

As additional time-variant variables we have included: a dummy *PRESALE* equal to 1 if the tickets are sold during the presale period; *WEEKSALE*, a variable that considers the number of weeks passed since the beginning of the sales period, including the presale (it is the time trend variable); $PERF_t$, denoting the number of performances of a given production that take place in the week considered, and $PERF_{t+1}$ considering the number of performances scheduled in the following week: indeed we have verified that, for a single performance, there is an increase in the number of tickets sold over the last two weeks, thus a larger number of performances of a given production in week $t+1$ would increase the number of tickets sold in week t for that production. Finally, the *REMAINING* variable denotes the number of performances remaining of a given production, while $CUMTICKET_t$ indicates the total number of tickets sold up to that given week. This latter variable is useful in indicating the total number of people that potentially have a bigger incentive to provide positive feedback on FB.

Table 3 provides descriptive statistics of the variables considered.

Table 3: Descriptive statistics. 3045 observations

Variable	Mean	SD	Min	Max
Tickets sold	322.69	441.85	0	5170
Young tickets	50.93	99.85	0	1421
Likes	17.42	91.86	0	2004
Likes per post	10.83	48.92	0	1002
Positive comments	0.687	10.31	0	553
Negative comments	0.005	0.094	0	3
Neutral comments	2.452	31.33	0	1041
Number of post	0.213	0.622	0	6
FB WEEK	281.66	88.04	99.54	415.55
PRESALE	0.053	0.225	0	1
WEEKSALE	17.75	11.54	1	57
$PERF_t$	0.469	1.166	0	10
$PERF_{t+1}$	0.469	1.166	0	10
REMAINING	12.52	10.05	0	47
CUM TICKET	3477.56	4396.58	5	43555

We undertook two separate analyses: in the first we follow the approach adopted by Chu et al.

(2016), to distinguish the predictor and influencer effect by estimating both a fixed-effect and random-effect panel data model. In the second analysis we estimate a simultaneous equation system in a similar way as Duan et al. (2008), which takes into account the fact that eWOM, inferred from FB data, both influences and is influenced by the number of tickets sold. The variables of the two models will be presented separately.

5. Methodology and results

In this section, we investigate the role of FB data on explaining the theater demand. A possible relationship between FB data and attendance could suggest causality or simply reflect the impact of unobservable characteristics of the product (i.e. theater performance) that are not measurable, such as the good/bad quality of the production, the skills of the performers etc. In the former case, FB data have an influencer role: providing consumers with information that affects purchasing decisions. For example, many positive likes on FB about a specific production may encourage potential customers to attend a performance of that production. In this case there would be a causal relationship from eWOM to the number of tickets sold. Alternatively, FB data could have a predictor role: a correlation between positive opinions and high demand reflects positive performance-specific characteristics (e.g. director, newness of the production). According to this perspective, such performance characteristics simultaneously cause FB activity and high demand, but there is not causal relationship between them: the change in FB data and the change in demand over time do not have a systematic relationship. Instead, it could be that the production characteristics influence its success on the box-office (as we expect) with a follow up effect on the eWOM. In such a case, there would be a causal relationship from the number of tickets sold to eWOM. Lastly there could be a mutual relationship between FB data and theater demand.
In order to distinguish between these possible effects, we estimate different models: a fixed-effect panel data model, a simultaneous equation model and a vector autoregressive model (VAR) with a Granger causality test.

5.1 Panel data

If FB data have an influencer effect, then the change over time of demand is related to a change over time of FB data, regardless of the time-invariant production characteristics. We explore this possibility estimating a fixed effect panel data model, considering as dependent variable the weekly number of tickets sold for a given production. For some productions, the box office opens for pre-sale in May, whereas for most productions the tickets are put on sale in August. We consider the whole box-office period, since the FB post related to a production can be published at any time during this period. Figure 1 shows the trend of tickets sold for the ballet *Sylfiden* during the 30-week sales period. During this period, various FB posts were published, both before the first performance and between one performance and another. In this example, the first post was published in week 19 of the box-office period, whereas the first performance took place in week 23. Clearly, the box-office period ends with the last performance of that specific production schedule.

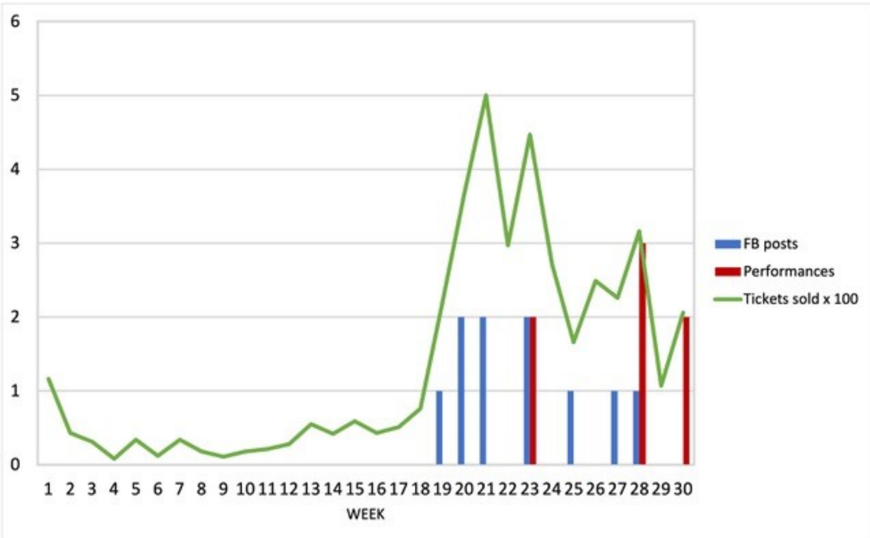

Figure 1: Tickets sold, performances, and FB posts for the production of *Sylfiden*

We choose a fixed effect model as it implies that any changes in the dependent variable must be due to influences other than these fixed characteristics. So, the parameters estimated for the FB data (which vary over time) are interpreted as influencer effect because the performance-specific characteristics cancel out (being invariant over time).

In this case, we aim to estimate the following equation:

$$(y_{it} - \bar{y}_i) = \beta' \left(A'_{SMit} - \overline{A_{SMi}}\right) + \gamma'(X'_{it} - \bar{X}_i) + (\varepsilon_{it} - \bar{\varepsilon}_i) \tag{1}$$

where y_{it} is the number of attendees for the performance i run the t^{th} time, A'_{SM} is a vector denoting the social media data, X'_{it} is the vector of control variables that are time-variant, ε_{it} is the error term, and the overlines represent the average values.

In the random effect model, on the contrary, it is possible to include time invariant variables, however it assumes that the unobservable individual production-specific characteristics are uncorrelated with the explanatory variables thus it excludes a predictor effect, i.e the possibility that social media data reflect the impact of unobservable characteristics of the production.

As social media activities we consider:

- Total *number of likes* on Facebook post related to the performance *i* in week *t*
- Total *number of comments* on Facebook post related to the performance *i* in week *t*, distinguished in positive, negative and neutral comments
- Total *number of posts* related to the performance *i* in week *t*

As mentioned in the previous section, the total number of likes depends on the number of posts, hence, we consider the ratio number of likes / number of posts, denoting the number of likes per post. However, the FB post can be seen as a 'marketing' action undertaken by the theater, which can affect the number of tickets sold. Thus, besides the main effect, in a second model we also consider the interaction terms between number of likes and number of posts. In this way, we clarify how the effect of FB likes and comments differs according to the number of posts, and vice-versa.

Table 4 shows the estimations of the panel data models. The fixed effect model reports a positive correlation (according to the model, *r* ranges between 0.61 and 0.67) between the unobserved production-specific characteristics and the explanatory variables (social media data). The Hausman test further confirms the preference for a fixed-effect model (*p-value* < 0.0001).

Table 4: Estimation results of panel data

	Total attendance		Young customers	
	Model 1	**Model 2**	**Model 1**	**Model 2**
Likes		0.351**		0.125****
		(0.146)		(0.033)
Posts		96.83****		10.142****
		(12.60)		(2.846)
Positive comments	0.456	0.177	0.049	0.009
	(0.566)	(0.558)	(0.126)	(0.126)
Negative comments	4.434	- 27.84	- 6.714	- 11.28
	(65.48)	(64.59)	(14.62)	(14.59)
Neutral comments	- 0.270	- 0.266	- 0.036	- 0.035
	(0.200)	(0.196)	(0.446)	(0.044)
Likes x Posts		- 0.011		- 0.016
		(0.047)		(0.010)
Likes per post	1.093****		0.189****	
	(0.129)		(0.008)	
PRESALE	- 228.83****	- 229.28****	- 45.63****	- 45.66****
	(33.07)	(32.56)	(7.386)	(7.35)
WEEKSALE	5.884****	5.098****	1.419****	1.330****
	(0.690)	(0.685)	(0.154)	(0.155)
$PERF_t$	141.56****	122.87****	24.51****	22.40****
	(7.541)	(7.711)	(1.684)	(1.742)
$PERF_{t+1}$	94.73****	85.66****	12.97****	12.01****
	(6.102)	(6.098)	(1.363)	(1.377)
REMAINING	35.13****	31.74****	5.660****	5.277****
	(1.665)	(1.681)	(0.372)	(0.380)
Intercept	-331408****	-276.95****	-62.18****	- 56.02****
	(29.52)	(29.65)	(6.593)	(6.70)
R^2 within	0.353	0.373	0.247	0.254
R^2 between	0.242	0.254	0.086	0.091
R^2 overall	0.246	0.273	0.137	0.147
Observations	3045	3045	3024	3045
Groups	99	99	99	99

****$p < 0.001$; ***p<0.010; **p<0.050; ***p<0.100; Robust standard errors in brackets under the estimated coefficients

Results show some expected findings: the number of tickets sold is less during the presale period and increases over time during the sales period (coefficient of the variable *WEEKSALE* is positive and statistically significant). As expected, the number of performances scheduled in the current and in the following week have a positive impact on the number of tickets sold, as well as the number of performances remaining. What interests us more, for the purpose of this paper, is the coefficient of the social media activities. We notice that the number of likes has a positive and highly significant effect on the number of tickets sold, also for the young segment. However, for the young segment is the effect smaller. In particular, the effect of the number of likes per post (Model 1) is positive, as well as the main effect of the number of likes (Model 2). The effect of the total number of likes decreases as the number of posts increase: it would be reasonable to suppose that the effect of the many likes achievable with few posts is bigger than the same quantity of likes achievable with many posts. However, this effect is not significant. Concerning the effect of FB comments, they don't seem to influence the number of tickets sold, regardless of their sentiment.

Overall, the model shows that only the number of likes will influence the number of tickets sold: a change in their number is associated with a change in the amount of ticket sold, regardless the number of post published (we rule out the effect of the increase in likes due to an increase in posts published). We can only suspect that the number of likes reflects unobserved characteristics of the production (i.e. a predictor role). This is verified with the simultaneous equation model whose results are shown in Table 5.

5.2 Simultaneous Equation System

The simultaneous equation system is used to consider a possible interdependence between FB activities and tickets sold. This approach is adapted from the conceptual framework proposed by Duan et al. (2008) in the movie context. We assume that eWOM (in the forms of FB likes and comments) influences the current number of tickets sold. However, people who have bought a ticket for a given performance may also contribute to positive feedback through social media activities, which in turn leads to an increase in the number of tickets sold. First, in a single-fixed effect equation that relates the number of tickets sold to social media effect, we determine whether there is a simultaneity issue, i.e. whether the variables *n° of likes* and *n° of comments* are endogenous. The Durbin-Watson test shows that only *n° of likes* is endogenous (*p-value* < 0.001), whereas *n° of comments* is not (*p-value* = 0.653).

Hence, we model this framework through two interdependent equations, where in one equation the dependent variable is the amount of ticket sold in a given week and in the other the number of likes per post is the dependent variable.

$$Tickets\ sold_{it} = \alpha_1 + \beta_1\, n°\ of\ likes\ per\ post_{it} + \sum_{j=1}^{2} \beta_2\, n°\ of\ likes\ per\ post_{it-j} + \beta_3\, n°\ of\ comments_{it} + \sum_{j=1}^{2} \beta_4\, n°\ of\ likes_{it-j} + \gamma' X_{it} + \mu_i + \varepsilon_{it} \quad (3)$$

$$Likes\ per\ post_{it} = \alpha_2 + \eta' POST\ TYPE'_{it} + \theta_1\, Tickets\ sold_{it} + \sum_{j=1}^{2} \theta_2\, Tickets\ sold_{it-j} + \delta_1\, FB\ Week + \delta_2\, CUM.TICKET_{it} + \lambda_i + \upsilon_{it} \quad (4)$$

In equation (3), *X* is the same vector of time-variant characteristics as in (1). The eWOM activities are the number of likes and the comments, where the latter are distinguished in positive, negative and neutral. We include also multi-lag terms (up to 2) assuming that eWOM spreads and it might take time to convince costumers to buy a ticket. The theater productions' factors, which are time-invariant, are controlled by the fixed effect, so production-specific dummy variables are included in both equations.

In equation (4), we relate the n° of likes per post in time *t* to the number of tickets sold in the same week, the type of posts published (video, photo, link etc.), the cumulative number of tickets of that production already sold (*CUMTICKET*), the n° of performance in the current week and the number of weeks passed since the theater opened the FB Page. In this equation we also include multi-lag terms (up to 2) for the variable related to the ticket sale: as most of the tickets are sold in the two weeks before the performance, we consider two weeks as a possible time span between the ticket purchase and the eWOM action. The equation also includes the different type of post (status, video, photo etc.) that could attract the FB users' attention in a different way. The three-stage least-square (3SLS) procedure is used to simultaneously estimate the system of two equations, considering both the total amount of tickets and the tickets sold to young/student customers. The results are shown in Table 5[2].

Table 5: Estimation results of the 3SLS model

	Total attendance	**Young costumers**
Ticket sold equation		
Likes per post	5.512****	0.717****
	(0.403)	(0.083)
Likes per post *t-1*	0.461***	0.096***
	(0.117)	(0.028)
Likes per post *t-2*	0.469***	0.143****
	(0.107)	(0.026)
Positive comments	0.271	0.147
	(0.512)	(0.120)
Positive comments *t-1*	- 0.048	- 0.053
	(0.445)	(0.108)
Positive comments *t-2*	- 0.194	- 0.018
	(0.444)	(0.107)
Negative comments	- 5.877	- 1.572
	(53.45)	(12.82)
Negative comments *t-1*	42.17	9.195

[2] The 3SLS model is estimated assuming homoskedasticity, as the option for robust standard error is not available for the reg3 command in STATA

	(51.61)	(12.48)
Negative comments *t-2*	39.74	28.69
	(54.21)	(13.10)
Neutral comments	- 0.101	0.044
	(0.185)	(0.042)
Neutral comments *t-1*	- 0.149	- 0.007
	(0.158)	(0.038)
Neutral comments *t-2*	- 0.125	- 0.060
	(0.157)	(0.038)
PRESALE	- 182.70****	- 39.16****
	(27.57)	(6.589)
WEEKSALE	6.101****	1.291****
	(0.856)	(0.180)
$PERF_t$	80.96****	16.14****
	(6.671)	(1.576)
$PERF_{t+1}$	60.26****	7.889****
	(5.343)	(1.199)
REMAINING	27.13****	4.572****
	(1.912)	(0.403)
Intercept	- 88.53	- 37.69*
	(93.86)	(19.54)
R^2	0.336	0.490
Likes equation		
Tickets sold	- 0.002	0.063
	(0.010)	(0.065)
Tickets sold *t-1*	0.060****	0.173****
	(0.008)	(0.049)
Tickets sold *t-2*	- 0.008***	0.030**
	(0.003)	(0.015)
Events	36.67*	68.04**
	(20.02)	(28.09)
Link	17.67****	14.93***
	(4.293)	(4.956)
Status	- 1.896	- 7.127***
	(2.161)	(2.478)
Photo	31.57****	33.46****
	(3.026)	(2.822)
Video	20.71****	26.13****
	(3.022)	(3.425)
CUM. TICKET	0.001***	0.001*
	(0.0003)	(0.0003)
FB Week	- 0.232**	- 0.159
	(0.116)	(0.123)
Intercept	85.95*	62.89
	(45.57)	(48.14)
R^2	0.143	0.169
Observations	2854	2854

****$p < 0.001$; ***$p<0.010$; **$p<0.050$; ***$p<0.100$; Robust standard errors in brackets under the estimated coefficients

In the equation for the tickets sold, we note that the number of likes is a significant predictor as well as the lagged coefficient in a similar way as in the panel data model, although the magnitude of the influence of the current number of likes is much higher. This may indicate that the eWOM influence is more effective in the concurrent week and diminishes in the following weeks. At the same time the comments, no matter their sentiment, do not influence the decision to purchase a ticket: considering this model together with the panel data models, it seems that the FB likes are far more important in understanding theater demand than are the FB comments. The sign of the other coefficients of the first equation are the same as in the panel data model: the number of tickets sold is less during the presale period and increases over time during the sales period; in addition, there is a positive relationship between the number of tickets sold and the number of performances in the current and in the following week.

Concerning the second equation, we see that the purchase of ticket has an effect on the eWOM activity, not in the concurrent week but a week later. For young theatergoers, there is a positive but weaker effect also two weeks later, while in general there is a negative (although the impact is negligible) effect two weeks after[3]. This seems to suggest that there is a span period of around one week between the ticket purchase and the eWOM activity, in which we suppose the user have attended the performance. This, together with the positive and statistically significant coefficient of the variable *CUM TICKET* (number of tickets already sold for the production) lead us to suspect that eWOM activity is not generated right after the purchase of tickets, but once having attended the performance. Only after that, they spread the eWOM in the form of likes (but not of "comments") which in turn influence the purchase of ticket (influence effect). Finally, we notice that some types of post attract a larger quantity of likes than other types: in particular, "events" and 'photos' generate more likes than "video" and "link", whereas the 'status' post is not statistically significant in the general equation and negative considering the young customers.

5.3 Panel vector autoregressive model and Granger causality test

Lastly, we try to capture the interdependencies among eWOM (number of likes per post and number of comments) and tickets sale using a panel vector autoregressive (VAR) model (Abrigo and Love, 2016), together with a Granger causality test. The test aims to verify whether the past value of the variable (for instance likes per post) are useful in determining the value of the other variable (for instance, tickets sold), conditional on the

[3] Also Duan et al. (2008) found a negative and significant coefficient of the second lag term of box-office revenue on the amount of eWOM. Their explanation is that there is a substitution effect between WOM volume between *t* and *t-2*

past values of the latter variable. The null hypothesis is that one variable does not Granger-cause the other variable. The VAR is estimated using the number of performances as exogenous variable and constant terms in each of the two equations. Table 6-7 shows the results for the tickets sale and FB likes; similar Tables 8-9 for the tickets sale and FB comments.

Table 6: Panel VAR model of tickets and FB likes per post

	Total attendance	**Young costumers**
Ticket equation		
Ticket sold		
L1	1.196****	1.190****
	(0.115)	(0.145)
L2	0.067	0.232
	(0.067)	(0.114)
Likes per post		
L1	0.963***	0.180***
	(0.348)	(0.069)
L2	0.098	0.114
	(0.384)	(0.104)
Performances	38.51**	3.378
	(19.15)	(6.673)
Likes equation		
Ticket sold		
L1	0.029***	0.181***
	(0.010)	(0.070)
L2	0.016*	0.150**
	(0.008)	(0.063)
Likes per post		
L1	0.224**	0.236**
	(0.102)	(0.102)
L2	- 0.012	- 0.008
	(0.069)	(0.073)
Performances	7.082***	6.917*
	(2.612)	(3.722)

****$p < 0.001$; ***$p<0.010$; **$p<0.050$; ***$p<0.100$; Robust standard errors in brackets under the estimated coefficients

Table 7: Granger causality test of tickets sold and FB likes per post

	Total attendance		**Young costumers**	
Causal direction	Chi-squared	p-value	Chi-squared	p-value
Tickets sold → Likes per post	7.972	0.0019	7.188	0.027
Likes per post → Tickets sold	10.805	0.005	9.754	0.008

Table 8: Panel VAR model of tickets and FB comments

	Total attendance	**Young costumers**
Ticket equation		
Ticket sold		
L1	1.233****	1.232****
	(0.118)	(0.148)
L2	0.059	0.228**
	(0.068)	(0.114)
Comments		
L1	0.183	0.053
	(0.156)	(0.062)
L2	0.114	0.009
	(0.142)	(0.044)
Performances	45.71**	6.177
	(18.91)	(7.084)
Comments equation		
Ticket sold		
L1	0.011	0.074
	(0.008)	(0.046)
L2	- 0.002	- 0.006
	(0.003)	(0.026)
Comments		
L1	0.023*	0.023*
	(0.012)	(0.012)
L2	0.002	0.001
	(0.003)	(0.004)
Performances	2.716****	3.057***
	(0.668)	(1.023)

****$p < 0.001$; ***$p<0.010$; **$p<0.050$; ***$p<0.100$; Robust standard errors in brackets under the estimated coefficients

Table 9: Granger causality test of tickets sold and FB comments

Causal direction	Chi-squared	p-value	Chi-squared	p-value
Tickets sold → Comments	1.835	0.399	0.734	0.693
Comments → Tickets sold	3.984	0.136	3.829	0.147

We find again a significant connection between the FB likes and the number of tickets sold and a not significant on between the FB comments and the purchase of tickets. We can argue for a mutual interaction between the number of likes given to posts specifically dedicated to a given production and the number of tickets sold concerning that specific production: FB likes have an influencer role as an signal of quality for consumers, leading to an increase in the amount of tickets sold. However, the volume of likes are themselves influenced by the quantity of people that have already bought a ticket for a given performance.

Discussion and conclusion

In this paper, we have shown that FB data can constitute a new effective tool for understanding theater demand. Our panel data model shows that the number of likes has a positive and highly significant effect on the number of tickets sold, also for the young segment. It is interesting to notice that for the young segment we find a smaller effect. Concerning the effect of FB comments, they do not seem influence the number of tickets sold, regardless of their sentiment. Another interesting result is that the number of likes will influence the number of tickets regardless of the number of post published. In the simultaneous equation model, the first equation shows the same results as in the panel data model, namely that the number of likes is a significant predictor, and in addition the lagged coefficient shows that the eWOM influence is more effective in the concurrent week and diminished in the following weeks. The second equation shows that the purchase of tickets has an effect on eWOM activity not in the concurrent week, but a week later. This leads us to conclude that eWOM activity in the form of likes, is not generated right after the purchase of tickets, but after the attendance of the performance. This in turn influence the purchase of tickets (influence effect). Furthermore, we notice, that “events” and “photos” generate more likes that “videos” and “links”. Finally, using a panel vector autoregressive (VAR) model we can conclude that FB likes have an influencer role as a signal of quality for consumers, leading to an increase in amounts of ticket sold. However, the volume of likes are themselves influenced by the quantity of people that have already bought a ticket for a given performance.

The results contain crucial information for theater managers. Firstly, the results show that likes given to the FB posts have both a prediction and an influence role. It means that FB likes can be an effective instrument in order to generate ticket sales as well a predictor of number of tickets sold.

Secondly, as we find that only likes and not comments, eighter positive or negative, have an effect. This can come as a surprise, but perhaps it shows that people don’t spend time reading the comments, and perhaps they don’t expect to agree with them. However, the size of likes signals a large interest for the production, and for the consumers the number of likes can be a way to reduce search coast by signaling quality of a given production. Furthermore, likes can generate a snowball effect, which is more effective that traditional WOM, because it is faster and more reachable by a large audience. Furthermore, the process can be driven by the institutions (theatre) through their social media marketing plan. This highlights the importance for a theatre to generate likes, and here “events” and “photos” seem to have larger impact the other kind of posts.

Thirdly, the general effects are also found among the younger audience, but the effects are much smaller, leading to the conclusion that the theatre should generally post on FB with the mature audience as the target – not the younger segment.

In this way the analysis of FB posts and likes (and similar studies of other social media platforms) can constitute a means by which to assist theater management in taking decisions concerning, for example, pricing and organization of the theater season program.

The adoption of social media data in understanding theater demand is not as effective as it can be in other areas of application characterized by the supply of products with a (almost) global market and audience, such as iPhones, automobiles, and movies. However, it represents an additional instrument for the straightforward collection of a larger amount of information concerning audience impressions of a theater production. Given the growing usage of social media by consumers, we believe that this stream of research will benefit from further development. In future research, it would be interesting to investigate the impact of social media for other types of cultural institutions. Furthermore, other social media platforms that FB should be included, and it would be interesting to study how the effects differ for different audience segments.

References

- Abbé-Decarroux, F. (1994). The perception of quality and the demand for services: Empirical application to the performing arts. *Journal of Economic Behavior & Organization*, *23*(1), 99-107.

- Abrigo, M. R., & Love, I. (2016). Estimation of panel vector autoregression in Stata. *The Stata Journal*, *16*(3), 778-804.

- Akdede, S. H., & King, J. T. (2006). Demand for and productivity analysis of Turkish public theater. *Journal of Cultural Economics*, *30*(3), 219-231.

- Asur, S., & Huberman, B. A. (2010, August). Predicting the future with social media. In *2010 IEEE/WIC/ACM international conference on web intelligence and intelligent agent technology* (Vol. 1, pp. 492-499). IEEE.

- Baldin, A., & Bille, T. (2023). T*he Avant-garde Consumers: A new perspective on quality evaluations of performing arts,* Poetics (forthcoming)

- Baldin, A., Bille, T., Ellero, A., & Favaretto, D. (2018). Revenue and attendance simultaneous optimization in performing arts organizations. *Journal of cultural economics*, *42*(4), 677-700.

- Basuroy, S., Abraham Ravid, S., Gretz, R. T., & Allen, B. J. (2020). Is everybody an expert? An investigation into the impact of professional versus user reviews on movie revenues. *Journal of Cultural Economics*, *44*(1), 57-96.

- Bille Hansen, T. (1991). *Det Kgl. Teater: et kulturøkonomisk studie*. AKF forlaget, Copenhagen.

- Caves, R. E. (2000). *Creative industries: Contracts between art and commerce* (No. 20). Harvard University Press.

- Chu, P. C., Liu, Y. H., & Sher, C. Y. (2016). *The impact of amateur film reviews on movie theater attendance: predictors versus influencers*. In 19th International Conference on Cultural Economics, Valladolid, Spain.

- Colbert, F., Beauregard, C., & Vallée, L. (1998). The importance of ticket prices for theatre patrons. *International Journal of Arts Management*, 8-15.

- Corning, J., & Levy, A. (2002). Demand for live theater with market segmentation and seasonality. *Journal of Cultural Economics*, *26*(3), 217-235.

- Duan, W., Gu, B., & Whinston, A. B. (2008). Do online reviews matter?-An empirical investigation of panel data. *Decision support systems*, *45*(4), 1007-1016.

- Franch, F. (2013). (Wisdom of the Crowds) 2: 2010 UK election prediction with social media. *Journal of Information Technology & Politics*, *10*(1), 57-71.

- Grisolía, J. M., & Willis, K. G. (2012). A latent class model of theatre demand. *Journal of Cultural Economics*, *36*(2), 113-139.

- Hussain, A., & Vatrapu, R. (2014, September). Social data analytics tool: Design, development, and demonstrative case studies. In *2014 IEEE 18th International Enterprise Distributed Object Computing Conference Workshops and Demonstrations* (pp. 414-417). IEEE.

- Hussain, A., & Vatrapu, R. (2014, May). Social data analytics tool (sodato). In *International Conference on Design Science Research in Information Systems* (pp. 368-372). Springer, Cham.

- Jenkins, S., & Austen-Smith, D. (1987). Interdependent decision-making in non-profit industries: A simultaneous equation analysis of English provincial theatre. *International Journal of Industrial Organization*, *5*(2), 149-174.

- Krebs, S., & Pommerehne, W. W. (1995). Politico-economic interactions of German public performing arts institutions. *Journal of Cultural Economics*, *19*(1), 17-32.

- Lassen, N. B., Madsen, R., & Vatrapu, R. (2014, September). Predicting iphone sales from iphone tweets. In *2014 IEEE 18th International Enterprise Distributed Object Computing Conference* (pp. 81-90). IEEE.

- Lassen, N. B., la Cour, L., & Vatrapu, R. (2017). Predictive analytics with social media data. *The SAGE handbook of social media research methods*, 328-340.

- Oh, C., Roumani, Y., Nwankpa, J. K., & Hu, H. F. (2017). Beyond likes and tweets: Consumer engagement behavior and movie box office in social media. *Information & Management*, *54*(1), 25-37.

- Seaman, B. A. (2006). Empirical studies of demand for the performing arts. *Handbook of the economics of art and culture*, *1*, 415-472.

- Throsby, C. D., & Nielsen, E. (1980). *Product Quality Decisions in Non Profit Performing Arts Firms*. Macquarie University, School of Economic and Financial Studies.

- Throsby, C. D. (1990). Perception of Quality in Demand for the Theatre. *Journal of cultural economics*, *14*(1), 65-82.

- Tobias, S. (2004). Quality in the performing arts: Aggregating and rationalizing expert opinion. *Journal of Cultural Economics*, *28*(2), 109-124.

- Urrutiaguer, D. (2002). Quality judgements and demand for French public theatre. *Journal of Cultural Economics*, *26*(3), 185-202.

- Vatrapu, R., Mukkamala, R. R., Hussain, A., & Flesch, B. (2016). Social Set Analysis: A Set Theoretical Approach to Big Data Analytics. *IEEE Access*, *4*, 2542-2571.

- Voortman, M. C. (2015). *Validity and reliability of web search based predictions for car sales* (Master's thesis, University of Twente).

- Willis, K. G., & Snowball, J. D. (2009). Investigating how the attributes of live theatre productions influence consumption choices using conjoint analysis: the example of the National Arts Festival, South Africa. *Journal of Cultural Economics*, *33*(3), 167-183.

- Zieba, M. (2009). Full-income and price elasticities of demand for German public theatre. *Journal of Cultural Economics*, *33*(2), 85-108.