Journal of Economic Behavior and Organization 156 (2018) 103–113

Contents lists available at ScienceDirect

Journal of Economic Behavior and Organization

journal homepage: www.elsevier.com/locate/jebo

More buzz, more vibes: Impact of social media on concert distribution

Daegon Cho[a,*], Youngdeok Hwang[b], Jongwon Park[c]

[a] College of Business, KAIST, 85 Hoegiro, Dongdaemun-gu, Seoul 02455 Korea
[b] Department of Statistics, College of Economics, Sungkyunkwan University, 25-2, Sungkyunkwan-ro, Jongno-gu, Seoul 03063, Korea
[c] Department of Accounting, Monash Business School, Monash University, 900 Dandenong Road, Caulfield East, VIC 3145, Australia

ARTICLE INFO

Article history:
Received 17 April 2017
Revised 31 July 2018
Accepted 16 September 2018
Available online 22 September 2018

Keywords:
Concert distribution
Music industry
Zero-inflated Poisson
Generalized linear mixed model

ABSTRACT

This paper examines the impact of social media on decision-making regarding concert locations in the music industry. The use of social media allows musicians to communicate directly with their current and potential fans, which provides useful information about where concerts will be more successful. In particular, with the use of social media, musicians are likely to reach out relatively unexplored regions in choosing their concert locations. To examine the effect of social media use on the distribution of concert locations, we introduce an empirical methodology, using a zero-inflated generalized linear mixed model with a log link function. The model accounts for potential heterogeneous locational and temporal traits, allowing us to measure the impact of characteristics of the population on location selection. The parameters estimated from this model support an argument that the use of social media encourages musicians to pursue unexplored markets that they may not have considered before the use of social media.

1. Introduction

> *"I think social media and mobile technologies are clearly creating many more opportunities and so much more innovation at the show. I continue to see social media and mobile technologies to be a much bigger driver, pre-show, related to people purchasing tickets. It will be the biggest marketing vehicle for live music." (Russell Wallach, President of Live Nation Network, September 7, 2011)*

A recent study reveals that 74% of adults in the United States regularly use social media websites in 2014 (Duggan et al., 2015). Due to this widespread use, firms attempt to make use of online social media websites to communicate with existing and potential consumers. Online media help firms to increase brand recognition by enhancing consumer loyalty and advertising new products and services (Kim and Ko 2012; Bianchi and Andrews 2015; Godey et al., 2016). This phenomenon is also notable in the music industry, and musicians' social media accounts dominate the public attention. For example, more than half of Twitter users follow at least one musician's account and eight out of ten most-followed Twitter accounts is created by the musician.[1] An increasing number of musicians create their own pages on social media, such as YouTube,

* Corresponding author.
E-mail addresses: daegon.cho@kaist.ac.kr (D. Cho), yhwang@skku.edu (Y. Hwang), Jongwon.Park@monash.edu (J. Park).

[1] "Why Social Networks Need Musicians (The Wall Street Journal, May 28, 2012)" (http://blogs.wsj.com/tech-europe/2012/05/28/why-social-networks-need-musicians/).

https://doi.org/10.1016/j.jebo.2018.09.012
0167-2681/

Twitter, and Facebook, to promote their new songs and upcoming concerts through those sites. Fans also provide feedback, communicating directly with musicians and other fans.

The use of social media in the music industry has several characteristics. On the one hand, online social media allows musicians to reach thousands of people at very low cost, and musicians can obtain useful information and responses from their fans. For example, most online social media provide connection statistics regarding location and demographics through applications such as Facebook Analytics or YouTube Analytics, which may be utilized to reveal the preferences of fans. On the other hand, fans have the incentive to access their favorite musician's social media frequently, because it provides up-to-date information on upcoming events. They can also communicate with other fans outside their immediate circle in real time. These significant shifts in marketing strategy led by the rise of online social media are at the center of our enquiries.

In this study, we investigate the impact of the use of social media on the geographic distribution of concert locations. In other words, we question whether the use of social media has an effect on musicians' tour itineraries. As recording sales have declined substantially in the last decade due to widespread illegal file sharing, income sources for musicians have shifted toward revenue from concerts (Krueger 2005).[2] We conjecture that the use of social media by musicians affects this new tendency by influencing their choice of suitable concert locations. In other words, assuming that the number of concerts and distribution of concert locations is the results of profit-maximizing decision-making, the use of social media may play an essential role in the decision-making process.

To test how the use of social media affects the choice of concert locations, we identify a sample of 234 well-known musicians and collect data about their 57,411 concerts performed in the United States between 2000 and 2011. We also obtain information regarding different time points of the introduction of social media accounts by these musicians, most of whom created their own social media pages around 2008–2010. Combining these two sources of data provides us a unique opportunity to examine how the geographic distribution of concerts has changed in the era of social media.

For empirical analysis, we apply a zero-inflated Poisson (ZIP) model and isolate the effect of the use of social media from the overall temporal trend. By using the ZIP model, we can allow for overdispersion by assuming that the sample is a mixture of a Bernoulli distribution and a Poisson distribution. We find evidence that adoption and use of social media by musicians have a significant impact on the choice of concert locations. In particular, musicians tend to reach out the unexplored places for concerts where they had not given a concert before, after they create personal pages on YouTube, Twitter, and/or Facebook. For instance, prior to use of social media accounts, musicians mainly performed their concerts in the more populous places such as New York City, Los Angeles, and Chicago. However, after musicians start using social media in late 2000s, they tend to give a concert in relatively small regions, implying that the communication through social media may bring the concerts to the unexplored areas for to maximize the profit of musicians. Our findings do not indicate that artists decrease their concerts in large cities to perform concerts in smaller regions. Instead, we document that, after creating new channels on online social media, musicians tend to give a concert in unexplored regions as they perform more concerts over time. This phenomenon results in the lower concentration ratio of concerts held in populous places.

The intended contributions of this research are twofold. On one hand, we attempt to provide insight into how the use of social media has impacted the experience goods market with a special emphasis on the supply-side dynamics, which was scarcely dealt with in the prior studies. On the other hand, we use a novel methodology that can also be applied to other contexts in which a myriad of observations have zero value and the distribution is highly skewed. These types of complicated dataset occur when long-tail distributions are placed. The remainder of this paper is organized as follows. In Section 2, we review related prior literature, and we describe our data in Section 3 and present the empirical methodology and results in Section 4. The discussion and conclusions follow in Section 5.

2. Related literature

Our study relates to previous literature focusing in three different areas: i) the impact of social media in the media industry, ii) the long tail phenomenon, and iii) the effect of the Internet on the music industry. The widespread use of online social media has sparked interest in how they affect sales of media goods. On one hand, there is evidence for the influence of word-of-mouth on the sales of media goods, including books (Chevalier and Mayzlin, 2006), movies (Chintagunta et al., 2010), music (Dewan and Ramaprasad, 2014), and niche products (Dewan and Hsu, 2004; Zhu and Zhang, 2010). Focusing more on the impact of social media, several studies investigate the effects of volume and valence in social media on the sales of media goods. For example, Rui et al. (2013) examine the movie industry and Dhar and Chang (2009) examine the music industry. In the same vein, Chellappa and Chen (2009) and Chen et al. (2015) highlight that sampling on MySpace is positively related to music sales and Morales-Arroyo and Pandey (2010) and Abel et al. (2010) find a positive association between the volume of buzz on social media and music album sales. However, to the best of our knowledge, there have been few studies on the impact of the use of social media on the concert market, and our research aims to fill this gap in the literature.

Investigating the location distribution of concerts, we observe the long tail effect (Anderson, 2006), in which prevalent use of the Internet may lead to a situation where niche products account for a larger share of total sales. Many studies

[2] According to *Pollstar* magazine, the size of the concert market increased from $1.75 billion in 2001 to $4.25 billion in 2010 led by an increase in ticket prices and the number of concerts.

have examined this phenomenon: Brynjolfsson et al. (2006), Dellarocas and Narayan (2007), Brynjolfsson et al. (2010), and Brynjolfsson et al. (2011), all of whom examine the long tail effect by identifying different drivers from both the supply (e.g., lower inventory and distribution costs) and demand sides (e.g., easy search tools and useful recommender systems to access niche products) of the media industry. However, the long tail effect has been disputed in recent academic research. Some researchers have found evidence of an association between Internet use and a higher concentration of popular products in movie markets (Elberse and Oberholzer-Gee, 2007; Tan and Netessine, 2009). Zentner et al. (2013) examine changes in video rental market sales as consumers move from offline to online markets and find that consumers are more likely to rent niche titles from the online channel than from the offline channel. In this paper, we explain the long tail effect from the perspective of geography, identifying to what extent musicians choose small centers to perform their concerts in the era of social media.

Our paper is also related to the rich academic literature on the effect of the Internet on the music industry. While most empirical studies have found evidence that piracy hinders sales of recordings, a few theoretical studies identify some positive aspects of file sharing (Peitz and Waeldbroeck, 2006). In particular, Gayer and Shy (2006) identify certain conditions under which musicians may benefit from music piracy by showing that the demand for live performances increases with the increased popularity of musicians resulting from the consumption of both legal and illegal copies of their music. Mortimer et al. (2012) extend this idea by empirically showing that while file sharing has a negative impact on the sales of music recordings, it also has a positive influence on the demand for live concerts. Waldfogel (2012) also argues that the distribution of songs at zero marginal cost makes concerts an increasingly important source of revenue for the industry.

As can be observed, the music market radically changed during the last two decades led by the prevalent use of the Internet. Early studies have shown an extensive impact of Internet adoption on the recorded music market as well as the demand-side dynamics. More recently, researchers have also investigated the impact of the Internet use on the rapidly-growing concert industry. However, as most studies primarily focus on the demand-side, relatively little is known about the supply-side dynamics. In this context, our study delves into the topic by filling the gap in the literature. In fact, to generate higher profits and reduce potential risks, it is important for artists to choose appropriate places to attract more audience. We conjecture that a more significant number of interactions and volume of buzz through social media activities may provide valuable knowledge to artists and concert promoters to find better places to perform their concerts. In particular, we expect that the geographic distribution of concert locations can be different after the musicians start using social media accounts. We also believe that the notable Long Tail phenomenon can be applied in this context by which concert locations are more likely to be dispersed while the artists find unmet needs in smaller regions.

3. Data and summary statistics

3.1. Data and sample construction

We collect data on historical concert information from Songkick.com between 2000 and 2011. Songkick.com is the largest concert information service that offers rich concert information using an index of ticket vendors, venue websites, and local news media. The available data allows us to identify the main headliner of the concert, associated music genre, performance date, and geographical location (city and state). To restrict our sample to well-known musicians, we use annual charts of the top 100 North American tours provided by *Pollstar Magazine* for the corresponding years. Specifically, we account for all musicians listed on the chart at least once but excluding group tours (e.g., *American Idol Live*) and theatrical shows (e.g., *Cirque Du Soleil*). The final sample includes 57,511 concerts performed by 234 musicians from 13 genres.[3,4]

We then match each concert location to its corresponding Core Based Statistical Area (hereafter, CBSA) code. According to the U.S. Census Bureau, "Core Based Statistical Areas (CBSAs) consist of the county or counties or equivalent entities associated with at least one core (urbanized area or urban cluster) of at least 10,000 population, plus adjacent counties having a high degree of social and economic integration with the core as measured through commuting ties with the counties associated with the core." We employ CBSA coding as the geographic standard for two reasons. First, our research strategy requires us to establish appropriate geographic divisions in which concerts occur. For this purpose, zip code-level descriptions seem too specific, while state-level descriptions seem too broad. The Federal Office of Management and Budget (OMB) identifies over 900 CBSAs within the United States, 366 of which are classified as metropolitan areas with populations greater than 50,000. The population of the CBSAs included in our study varies widely from nearly 20 million in the New York City area to 55,000 in Carlson City, NV. Second, but more importantly, residents of each CBSA are socio-economically tied to the urban center by the definition provided by the OMB, indicating that people in the same CBSA share infrastructure and facilities in close proximity.[5] Therefore, it is plausible to assume that a concert held in a certain CBSA primarily targets the potential audience residing in the same area.

[3] For better understanding of the data, we present the list of top 50 musicians by total number of concerts performed between 2000 and 2011 in Appendix A.

[4] These 13 genres include: pop/rock, blues, children's, classical, comedy, country, international, Latin, new age, R&B, rap, religious, and vocal. The pop/rock genre accounts for more than 65% of concerts in our data set.

[5] For example, although the New York City CBSA includes twenty four different counties spread over three states (New York, New Jersey, and Pennsylvania), many residents' daily routines occur within New York City. In other words, residents in this particular CBSA may also be regarded as part of the potential audience of concerts held in New York City or other cities within its CBSA.

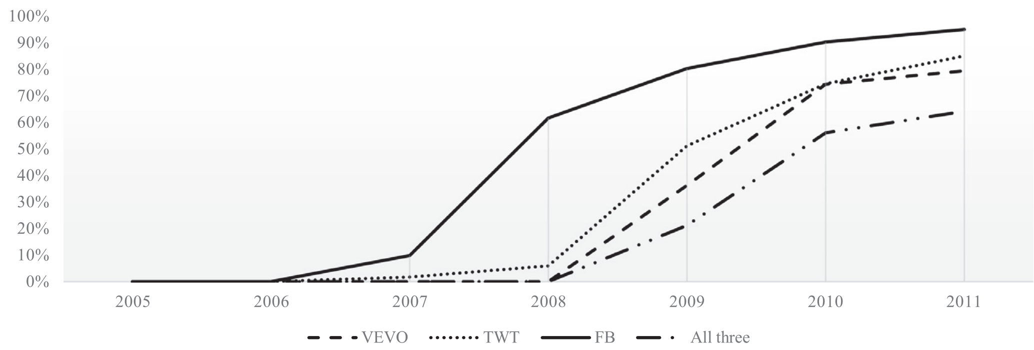

Fig. 1. Proportion of concerts held by musicians with social media accounts.
Note: VEVO, Vevo account in YouTube; TWT, Twitter account; FB, Facebook account; All three, Vevo in YouTube, Twitter and Facebook accounts.

In addition to the data from Songkick.com, we create social media activity variables for each musician by determining the date of initiation of their Vevo channels on YouTube, Twitter, and/or Facebook. Fig. 1 presents the yearly distribution of the proportion of concerts held by musicians with Vevo channels on YouTube, Twitter, Facebook, or all three social media accounts. Concerts performed by musicians with social media accounts first appeared in 2007, with Facebook in the lead at the beginning, showing rapid growth in 2008. The musicians' Twitter accounts also appeared in 2007 and significantly increase in 2009. Vevo accounts on YouTube first began to spread among musicians in 2009, which is two years later than other social media services; they reached a comparable level in 2010.[6] Among all concerts held in 2011, 79.4, 84.9, 94.9, and 64.1 percent of them were performed by musicians with Vevo channels on YouTube, Twitter, Facebook, or all three social media accounts, respectively.

3.2. Summary statistics

Table 2 presents the summary statistics of our final sample.[7] First, in *Panel A*, we report the yearly statistics of the number of musicians and concerts, and the number of CBSAs in which the concerts were held. The number of musicians increases over the sample period, because, by nature, those who appear on the *Pollstar Magazine* chart at least once are increasing. Both the number of concerts and the number of concerts per musician are increasing, although year-to-year variation is greater than the overall increasing trend. In 2009, the number of concerts substantially decreased by 18 percent (7,623 – 6,247 / 7,623) due to the global financial crisis, after which it remained at a similar level throughout the sample period. We also find an increasing trend in the number of CBSAs appearing in the data, which suggests a change in the geographical distribution of concert locations during the sample period. More importantly, during the years when social media services were increasing in popularity among musicians (2008–2011), the number of CBSAs increases while the number of concerts decreases, providing evidence that musicians' active use of social media may have influenced decision-making on the geographic and demographic attributes of concert locations.

In *Panel B*, to take a closer look at changes in musicians' choice of concert locations, we check the yearly statistics for the distribution of concert locations (CBSAs) and concert concentration ratio. For the first two columns, *AvgCBSAPop* is the average population of all CBSAs in which concerts were held and *MinCBSAPop* is the population of the least populated CBSA included in our final sample. Both *AvgCBSAPop* and *MinCBSAPop* show slightly decreasing trends and their values remain at the lowest level during the years when musicians actively began using social media services (2008–2011).

In the last three columns, we calculate three different ratios of concert concentration level, each indicating to what extent concert locations are concentrated in highly populated areas. *ConcentT10* (*ConcentT20*) is the number of concerts held in the top ten (top twenty) populated CBSAs in year t, deflated by the total number of concerts held in the same year. A higher value of *ConcentT10* (*ConcentT20*) implies that musicians have preferences for populous areas in choosing their concert locations. For both *ConcentT10* and *ConcentT20*, we find a decreasing trend and lower values in the more recent years of our sample period (2008–2011), suggesting that musicians are more likely to choose less populated CBSAs as their concert venues with the passage of time. *ConcertHerf* is the ratio computed as the Herfindahl index; it captures the level of concentration of concert locations among CBSAs.[8] Consistent with previous results, the value of *ConcertHerf* decreases throughout the sample period and reaches its lowest point in 2010 (0.0162). This also implies that, in recent years, musicians are less likely to stick to large cities and instead consider a variety of areas in selecting their concert locations. Taken together, our summary statistics reveal that during the last decade, decision-making in the music industry regarding

[6] For each social media service, the first appearance date in our sample is as follows: Facebook (Oct 1, 2007), Twitter (June 13, 2007), Vevo channel on YouTube (Jan 15, 2009).

[7] All variables reported in Table 2 are defined as in Table 1.

[8] *ConcertHerf* is measured as $\sum_{c=1}^{n} S_c{}^2$, where c is the CBSA and S_c is the number of concerts held in CBSA c deflated by the total number of concerts held in the same year.

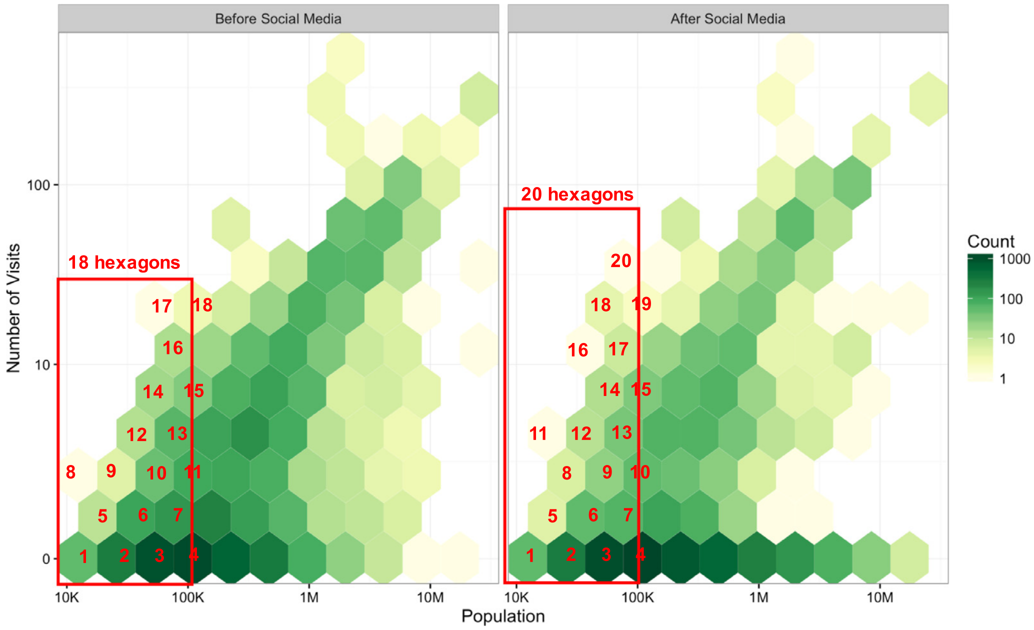

Fig. 2. Relationship between number of concerts and CBSA population.
Note: y-axis denotes the number of visits to account for areas with no concert; x-axis denotes CBSA population size.

where to locate concerts has substantially changed, and this change was more pronounced in the years when the use of social media services became widespread among the musicians.

We also present the relationship between population size and the number of concerts before and after the presence of social media accounts in Fig. 2. In fact, there is no significant difference between two charts at a glance. However, when focusing on the shaded hexagons under 100k on x-axis – the number of visits to the relatively smaller regions, there are 18 in the left, and 20 in the right chart, respectively. This difference provides evidence that there may be the greater number of visits to smaller regions after starting social media activity during the study period. We will explore further in the next section.

4. Empirical model and results

4.1. Model

In this section, we lay out the methodology and empirical results of our main analysis. We use a Poisson mixed model with an inflated zero, which is a generalized version of the usual linear regression model, to analyze the concert event data. There are therefore three major components in our model: Poisson regression, mixed effects, and inflated zeros.

First, we include the number of concert events as a response variable. Since the error structure cannot be assumed to be a normal distribution, it is natural to consider a generalized linear model (GLM). GLM is a linear regression technique that allows the observed response to be related to the linear model via a link function (e.g., Poisson regression uses a log link function). Although GLM can be used to model count response variables, our dataset contains more complications. The events are traced repeatedly over the same set of CBSAs. Additionally, there may exist some lingering effect over time. In such a case, the model must recognize the relationship between serial observations on the same unit (Laird and Ware, 1982). Hence we consider the generalized linear mixed model (GLMM), which combines the ideas of generalized linear models with random effects.[9] In GLMM, the distinction between fixed effects and random effects is important. In general, fixed effects are associated with parameters to be estimated in the model, while random effects are unobserved random variables (see Faraway, 2016 for more information). For example, in our study, the effect of CBSA population on concert occurrence is fixed, but the effect of an individual year is random. When the structure includes such complicated structure with assumptions on error structure, it is not straightforward to fit the models with methods such as maximum likelihood estimation method. We, therefore, adopted Bayesian approach which provides a very straightforward way to obtain the inferences from the model. Third, the collected concert event data has excess zeros by nature, because only data for top-ranking musicians is examined in this study. This sampling issue is taken care of by using a zero-inflated Poisson (ZIP) regression model. We

[9] Throughout this paper, we use the term "random effect" to refer the zero-inflated Poisson mixed model. Specifically, "random effect" indicates that the estimated parameters of variables are random coefficients that account for artists, genres, regions, and years.

Table 1
Variable descriptions.

Variable	Definition
Concerts	Total number of concerts held.
Log(population)	Natural logarithm of total population size.
SocialMedia	A variable that indicates the presence of the musician's social media account. This variable is set to one if the musician has his/her social media account, and zero otherwise.
Musicians	Total number of active musicians that hosted at least one concert.
CBSAs	Total number of CBSAs in which at least one concert was held.
AvgCBSAPop	Average population of all CBSAs in year t.
MinCBSAPop	Population of the least populated CBSA included in final sample.
ConcentT10	Total number of concerts held in top-ten populated CBSAs in year t.
ConcentT20	Total number of concerts held in top-twenty populated CBSAs in year t.
ConcertHerf	Herfindahl index reflecting concert concentration among CBSAs, calculated as $\sum_{c=1}^{n} S_c{}^2$, where c is the CBSA and S_c is the number of concerts held in CBSA c deflated by the total number of concerts in year t.

provide a detailed formulation of the model in the following section. Without properly taking into account of all these challenges, not only we have less accurate results, but also often fail to estimate the parameters. For example, a simple GLM without random effects cannot be estimated due to the numeric issues.

We analyze the location and time of the individual concert events. To associate them with the other information, we reorganize the data set as an aggregation count. Specifically, we count the number of concert events for each CBSA, year, genre of the music, and presence of the social media activity. To facilitate presentation of the proposed modeling approach, we introduce some useful notation. Assume that we have n CBSAs, T years, G genres in total. Hence, we have $N = n \times T \times G \times 2$ counts after aggregation, where the last multiplication by two is because we treat social media activity as binary (i.e., before and after). We have 961 CBSA units, 12 years, and 13 genres in our data set; thus, 299,832 ($961 \times 12 \times 13 \times 2$) observations are analyzed. Let y_i denote the number of events associated with i th case, where $i = 1, \cdots, 299,832$. For each y_{i}, we have associated CBSA, year, genre of the music, which are collectively denoted by Z_i. Note that Z_i is a q-dimensional incidence vector consists of 0 and 1, and the length q is determined by the sum of number of categories in CBSA, year, and the genre. For example, if y_i is the count associated with CBSA A in 2010 and genre 1, the elements in Z_i associated with these specific categories are ones, while all else being zeros.

For each concert event, the population associated with its CBSA location is available; the population, social media activity indicator and their interaction. Hence there are three elements of key variables of interest – population associated with this CBSA ($\log(population_i)$), indicator variable for social media presence ($SocialMedia_i$) and interaction of these two variables ($\log(population_{it}) \times SocialMedia_i$). We assume y_i follows the Poisson distribution with mean, μ_i. That is,

$$p(y_i = number\ of\ concerts_i |\ \mu_i) = \frac{(\mu_i)^{y_i} exp(-\mu_i)}{\mu_i!}, \tag{1a}$$

where μ_{i} is given by

$$\eta_i = \log(\mu_i) = \beta_0 + \beta_1 \log(population_i) + \beta_2 SocialMedia_i + \beta_3 \log(population_i) \times SocialMedia_i + Z_i \mathrm{b}_i + \xi^t + \varepsilon_i$$
$$where\ Z_i = \begin{pmatrix} a\ set\ of\ 961\ CBSA\ indicators\ (0\ or\ 1) \\ a\ set\ of\ 12\ year\ indicators (0\ or\ 1) \\ a\ set\ of\ 13\ genre\ indicators (0\ or\ 1) \end{pmatrix} \tag{1b}$$

and $\mathrm{b_i}|\mathrm{Q} \sim \mathrm{N}(0, \mathrm{Q}^{-1})$, where the precision matrix Q depends on parameters ϕ with $\dim(\phi) = \mathrm{V}$. The error terms are assumed to be independent of each other. Due to the presence of the random effects $\mathrm{b_i}$, the covariance matrix of η_{i}'s are not diagonal. Z_i is a q-dimensional incidence vector consists of 0 and 1, and the length q is 986 that is determined by the sum of number of categories in CBSA (961), year (12), and the genre (13). ξ^t denotes a series of year dummy to control for the average time trend that are common for all CBSAs. Among the random effects, the annual component has a temporal correlation over the years, in the autoregressive (AR) model of order 1 with the AR parameter ρ. The key variables are defined as in Table 1.

Intuitively, model (1) allows us to take into account the effect of population density as well as the CBSA-, year-, musician- and genre-specific effects when we examine the effects of social media activity. Specifically, the estimates associated with β provides the information on the impact of the population, and social media activity, and their interaction. A positive and significant coefficient on *SocialMedia* and the interaction term between log*(population)* and *SocialMedia* could imply that the use of social media account increased the number of concerts in all regions. Stated differently, this may suggest that after artists start using social media accounts, they are likely to hold more concerts not only in populous areas but also in less-populous regions. Inclusion of the random effects can help better estimation of β by properly adjusting the serial correlation of the observations from the same CBSA, year, or the same genre. For the detail of the mixed effects model, including the selection between fixed and random effects, see Bates (2010).

Table 2
Yearly trends.

Panel A. Yearly statistics for musicians, concerts, and CBSAs					
Year	Musicians	Concerts	Concerts / Musicians	CBSAs	CBSAs / Musicians
2000	117	3,143	26.86	246	2.10
2001	117	2,419	20.68	213	1.82
2002	143	3,824	26.74	283	1.98
2003	146	4,348	29.78	299	2.05
2004	149	4,210	28.26	318	2.13
2005	150	4,386	29.24	289	1.93
2006	149	3,992	26.79	276	1.85
2007	174	5,061	29.09	295	1.70
2008	195	7,623	39.09	320	1.64
2009	192	6,247	32.54	307	1.60
2010	188	6,225	33.11	336	1.79
2011	193	5,933	30.74	333	1.73
Average (total)	159.42	4,784	29.41	292.92	1.86
Panel B. Yearly statistics for CBSA populations, and concert concentration ratio					
Year	AvgCBSAPop	MinCBSAPop	ConcentT10	ConcentT20	ConcentHerf
2000	4,110,297	24,467	0.3010	0.4438	0.0196
2001	4,643,710	16,806	0.3410	0.5022	0.0244
2002	4,127,360	16,699	0.3047	0.4393	0.0194
2003	3,838,320	13,396	0.2827	0.4078	0.0188
2004	3,658,270	22,474	0.2627	0.3855	0.0181
2005	4,027,396	18,817	0.2948	0.4273	0.0222
2006	3,916,271	23,243	0.2700	0.4056	0.0238
2007	4,012,636	32,234	0.2774	0.4074	0.0215
2008	4,038,456	16,617	0.2840	0.4223	0.0181
2009	4,034,175	16,617	0.2704	0.4111	0.0197
2010	3,793,798	16,624	0.2562	0.3907	0.0162
2011	4,024,857	16,601	0.2695	0.3934	0.0182
Average (total)	4,018,795	19,550	0.2845	0.4197	0.0200

Note: This table presents the summary statistics for musicians and concert characteristics for each sample year. All variables are defined as in Table 1.

Now we briefly outline the estimation steps for our model. Let $\gamma = (\beta, \mathrm{b}) = (\gamma_1, \ldots, \gamma_G)$, where $G = \dim(\gamma)$. For the GLMM, the posterior is given by

$$\begin{aligned} \pi(\gamma, \phi|y) &\propto \pi(\gamma|\phi)\pi(\phi)\prod_{i=1}^{N} p(y_i|\gamma,\phi) \\ &\propto \pi(\phi)\pi(\beta)|Q|^{\frac{1}{2}}\exp\left\{-\frac{1}{2}b^TQb + \sum_{i=1}^{N}\log\, p(y_i|\gamma,\phi)\right\}, \end{aligned} \tag{2}$$

Define $y = (y_1, \ldots y_N)$. We are interested in the posterior marginal distributions $\pi(\gamma_g|y),\ g = 1, \ldots, G$, and $\pi(\phi_\upsilon|y),\ \upsilon = 1, \ldots, V$. It is straightforward to see that

$$\pi(\gamma_g|y) = \int \pi(\gamma_g|\phi, y)\pi(\phi|y)d\phi. \tag{3}$$

The collected concert event data has excess zeros by nature, because only those among the top-ranking musicians are examined in this study. Thus, many CBSAs have no concert events for years and most CBSAs have no visits during the study period, which causes difficulty fitting the data to the GLMM and making associated inferences. To alleviate this issue, we utilize the zero-inflated Poisson (ZIP) regression model (Lambert, 1992; Hall, 2000). In the ZIP model,

$$y_i \sim \begin{cases} 0 & \text{with } p, \\ \text{Poisson } (\eta_i) & \text{with } 1-p, \end{cases} \tag{4}$$

where η_i is from (2) and p is independently modeled from other covariates for simplicity. The mixing probability p is a hyperparameter with $p = \exp(\xi)/(1 + \exp(\xi))$ and the initial value and prior is given in terms of ξ.

4.2. Results

To examine the effect of social media on concert location, we estimate the parameters for the model (2). A common approach to evaluate the posterior marginal in Eq. (4) is the Markov chain Monte Carlo (MCMC) method, but it is well known to be prohibitively slow. Instead of using MCMC, we use integrated nested Laplace Approximations (INLAs). The INLA approach is useful to compute posterior marginals and is well known to be accurate. We use package INLA

Table 3
Results from the Main Model: Mean, median (the 50th percentile) of the posterior distribution for the fixed-effects parameters, along with 95% credible intervals (between the 2.5th and 97.5th percentiles).

	Mean	Std. dev.	2.5% Percentile	50% Percentile	97.5% Percentile
Intercept	−9.5298	0.5501	−10.6276	−9.5282	−8.4443
Log(Population)	1.5760	0.0675	1.4461	1.5752	1.7108
SocialMedia	0.1439	0.0359	0.0733	0.1439	0.2143
Log(Population) SocialMedia*	0.0395	0.0069	0.0260	0.0395	0.0530

Note: This table presents the mean, median (the 50th percentile) of the posterior distribution for the fixed effects parameters, along with 95% credible intervals (between the 2.5th and 97.5th percentiles). Dependent variable is the total number of concerts held (*Concerts*). All key variables are defined as in Table 1. In the model, parameters are assumed to be random variables and have a distribution. The credible intervals are given instead of confidence intervals; the probability that the parameters lie within the 95% credible interval is 95%.

Table 4
Results from the Main Model: Mean, median (the 50th percentile) of the posterior distribution for the model hyperparameters, along with 95% credible intervals (between the 2.5th and 97.5th percentiles).

	Mean	Std. dev.	2.5% Percentile	50% Percentile	97.5% Percentile
Zero-probability	0.5715	0.0047	0.5624	0.5715	0.5808
Precision for Year	35.1115	15.0633	12.8575	35.8753	70.9519
AR - year	0.4132	0.2140	−0.0340	0.4243	0.7851
Precision for CBSA	0.1825	0.0140	0.1543	0.1829	0.2092
Precision for Genre	0.3490	0.1418	0.1373	0.3284	0.6825

Note: This table presents the mean, median (the 50th percentile) of the posterior distribution for the model hyperparameters, along with 95% credible intervals (between the 2.5th and 97.5th percentiles). In the model, parameters are assumed to be random variables and have a distribution. The credible intervals are given instead of confidence intervals; the probability that the parameters lie within the 95% credible interval is 95%.

(Martins et al., 2013) with an R Core Team (2013) environment.[10] To assess the impact of social media activity after taking the population effects into account, we include the logarithm of population size and the use of social media, along with their interaction terms. To improve interpretability of the intercept term, we re-scale the log population by subtracting the log minimum population among the CBSAs. Then, the intercept can be interpreted as the mean rate of the event for the lowest population among the CBSAs.

To interpret the results of the estimated model, we need to see the posterior marginal of the parameter of interest. Since there is no prior information on parameters, we use vague prior distributions. Among the fixed effects, the population is a dominant component, as seen in Fig. 2. In Table 3, the posterior mode for a parameter associated with the log*(population)* is 1.576 with a 95% credible interval (1.44, 1.71), suggesting that a ten-times-more-populous city attracts approximately five times more concert events. The estimated coefficients of *SocialMedia* and the interaction term between log*(population)* and *SocialMedia* is positive and statistically significant, indicating that the presence of social media indeed shifts the curve upwardly. In other words, the number of concerts increased after opening the social media account in all regions (that is associated with the positive and significant estimated coefficient of the interaction term). At the same time, the number of concerts in less populous places also increased (that is associated with the positive and significant estimated coefficient of *SocialMedia*), because the y-intercept has a higher value when social media is in place.

In Table 4, the posterior mode for zero-probability parameter p for the ZIP model is 0.5715, or equivalently, $\hat{\xi} = 0.29$, with a 95% credible interval (0.56, 0.58). The parameter associated with social media activity is estimated to be 0.14 (0.07, 0.21), which implies that usage of social media increases the mean concert visit intensity by 15% ($\exp(0.14) = 1.15$). In addition, the slope for the log population increases by 0.04 (0.02, 0.05) after the musicians begin using social media. Combining these results suggests that social media encourages the musicians included in our sample to pursue markets with smaller populations more aggressively (because of values for both the intercept and slope increase). Thus, we may conclude that the musicians tend to visit relatively less populous places that they might have not visited without the use of social media.

Table 4 also documents the results of random effects model. The most informative information regarding a random effect is its variability. The random effects consist of annual effects, location-specific effects, and genre effects, where distinct random effect terms are assumed to be independent. Again, random effects are not parameters of interest, so we only look at parameters associated with the variance–covariance specification of the random effects. The posterior mode estimates of the precision parameters—reciprocal of the variance—for the annual, location-specific, and genre variables are 35.11 (12.86, 70.95), 0.18 (0.15, 0.21), and 0.35 (0.13, 0.68), respectively. These findings suggest that there are significant variations across years, regions, and genres. The posterior mode estimate for ρ is 0.35, but its 95% credible interval includes 0 (−0.03, 0.78), which suggests that autoregressive trend over the years is not evident.

[10] We refer readers to Rue et al. (2009) and Fong et al. (2010) for the theoretical details, and to Lindgren and Rue (2015) for the use and interface of the R package.

Table 5
Results from Fixed Effects Model: Mean, median (the 50th percentile) of the posterior distribution for the fixed-effects parameters, along with 95% credible intervals (between the 2.5th and 97.5th percentiles).

	Mean	Std. dev.	2.5% Percentile	50% Percentile	97.5% Percentile
Intercept	−9.6855	0.5389	−10.7581	−9.6849	−8.6175
Log(Population)	1.5829	0.0686	1.4506	1.5822	1.7195
SocialMedia	0.1449	0.0359	0.0743	0.1449	0.2154
*Log(Population)***SocialMedia*	0.0396	0.0069	0.0261	0.0396	0.0530
Year2001	−0.1031	0.0292	−0.1604	−0.1031	−0.0459
Year2002	0.1824	0.0260	0.1315	0.1824	0.2334
Year2003	0.3348	0.0252	0.2854	0.3348	0.3843
Year2004	0.2736	0.0254	0.2239	0.2736	0.3234
Year2005	0.3264	0.0252	0.2770	0.3264	0.3759
Year2006	0.2214	0.0257	0.1710	0.2214	0.2718
Year2007	−0.2093	0.0251	−0.2586	−0.2094	−0.1600
Year2008	0.0769	0.0239	0.0300	0.0769	0.1239
Year2009	−0.0420	0.0249	−0.0909	−0.0421	0.0069
Year2010	0.0565	0.0252	0.0070	0.0564	0.1060
Year2011	0.1930	0.0262	0.1416	0.1930	0.2445

Note: This table presents the mean, median (the 50th percentile) of the posterior distribution for the fixed effects parameters, along with 95% credible intervals (between the 2.5th and 97.5th percentiles). Dependent variable is the total number of concerts held (*Concerts*). All key variables are defined as in Table 1. In the model, parameters are assumed to be random variables and have a distribution. The credible intervals are given instead of confidence intervals; the probability that the parameters lie within the 95% credible interval is 95%.

As a robustness check, we replicate the estimation with an alternative model treating year as a fixed effect, and the results are presented in Table 5. One may easily notice that the estimated coefficients in the table are very close to corresponding outcomes in Table 4, which strengthens our main findings.

5. Discussion and conclusions

The objective of this paper is to examine how the use of social media impacts musicians' concert tours. Understanding the role of social media in concert distribution is important. Social media may be regarded as a new interactive platform through which fans and musicians communicate and the latter are able to acquire more accurate information on the needs of potential audiences. Our findings suggest that musicians utilize online social media to advantage, perhaps due to their expansive reach and leading-edge features. Musicians can make announcements and share their recent news, retain old fans and acquire new fans, promote their songs and concerts, express their emotions and so on via YouTube, Twitter, and Facebook. Indeed, musicians have great incentive to use social media to promote their concerts because of low advertising costs compared to those incurred through use of traditional media channels. In addition, social media provides key statistics on frequency and location of visitors; musicians can use this information in planning upcoming tours. In many cases, the unmet needs of audiences in unexplored areas can be realized through concerts given based on communications through social media.

The positive effect of buzz on social media is likely to increase ticket sales.[11] Social interaction, the influence of other fans, and correspondence with musicians may affect the decision-making of potential audience members to purchase concert tickets. Fans who attended previous concerts are likely to share their experiences and opinions of the concert with other fans, which may influence others. This mechanism may be particularly significant for experience goods, as discussed in previous research. Also, on social media, fans have an opportunity to listen either to the full track or part of a song for free. This is likely to affect their intention to purchase the physical or digital music, as addressed in Dewan and Ramaprasad (2014). However, attending the musician's concert physically may be a different experience from watching video clips of previous concerts. In summary, our findings suggest that adoption and use of social media by musicians and their fans may play a positive role in the concert market, increasing the likelihood that musicians will give their concerts in small, previously unexplored places.

The contribution of this study is twofold: First, while some recent literature examines the relationship between Internet use and dynamics in the concert industry, only few have addressed the effects of social media. Our study fills this void in the literature. Furthermore, we add some insights for managers about using social media as a marketing tool. Second, related prior studies are mostly based on analyses of anecdotal cases or use of simple regression models. In this study, we employ an advanced methodology, leveraging a large data set including the majority of superstars in the American music industry in the era of social media. This large-scale data analysis adds empirical evidence to the discussion on the relationship between the use of social media and the dynamics of the music market. In addition, the suggested methodology can be applied to

[11] Eventbrite, a social network platform for event organizers, quantifies the dollar value of each event's "share" made through social media such as Facebook, and Twitter. According to Eventbrite, share is most valuable for music concerts, and it is worth over $12 per share (https://www.eventbrite.com/blog/ds00-social-commerce-2/).

other contexts. In particular, the methodology may be useful in studies for which only representative data are available due to the difficulty of attaining exhaustive samples (e.g., high-ranked sites, most-elected politicians, etc.).

While our study contributes to a growing body of literature on the impact of Internet use on strategic behavior and market structure, this paper is not without limitations. First, the different time point of the introduction of a social media channel across musicians may not be an exogenous factor. To account for such information, it would be necessary to utilize either a multi-level approach to unveiling the hierarchical structure of the data or an instrumental variable. Second, we do not observe actual activity and communication between fans and musicians on the social media sites. Analyzing the interaction and dynamics of social media activities may provide additional insights regarding the impact of online social media. Third, we include only data on the number of concerts. Using ticket sales data, our study could be extended to include the economic effects of the use of social media, which in turn can measure the impact of social media marketing across different channels. Lastly, our empirical analysis is conducted at the CBSA level, while the location of concert is chosen by each musician. An analysis made at the musician level may extend this study by identifying the various traits of musicians that triangulate our main results.

Acknowledgements

This work was supported by the Ministry of Education and the Ministry of Science and ICT of the Republic of Korea and the National Research Foundation of Korea (NRF-2016S1A5A8019917, NRF-2017R1C1B5075436, and NRF-2018M3A9E8021503). Also, the author, Daegon Cho, is grateful for the research support by College of Business, KAIST (A0601003029).

Appendix A. Top 50 musicians by number of concerts performed between 2000 and 2011

Rank	Musician	Genre	# of concerts	Rank	Musician	Genre	# of concerts
1	Styx	Pop/Rock	948	26	B.B. King	Blues	471
2	Willie Nelson	Country	914	27	Def Leppard	Pop/Rock	465
3	Celine Dion	Pop/Rock	860	28	Kenny Chesney	Country	459
4	Widespread Panic	Pop/Rock	844	29	Celtic Woman	International	437
5	Indigo Girls	Pop/Rock	765	30	Sugarland	Country	434
6	311	Pop/Rock	760	31	Brooks And Dunn	Country	433
7	Barry Manilow	Pop/Rock	752	32	Harry Connick, Jr.	Vocal	424
8	Bob Dylan	Pop/Rock	739	33	Disturbed	Pop/Rock	420
9	Rascal Flatts	Country	730	34	Martina McBride	Country	414
10	Ani DiFranco	Pop/Rock	700	35	Snoop Dogg	Rap	414
11	Dave Matthews Band	Pop/Rock	681	36	Taylor Swift	Pop/Rock	412
12	Elton John	Pop/Rock	670	37	Sheryl Crow	Pop/Rock	402
13	Brad Paisley	Country	638	38	Michael W. Smith	Religious	398
14	Poison	Pop/Rock	618	39	Foo Fighters	Pop/Rock	392
15	Fall Out Boy	Pop/Rock	598	40	Santana	Pop/Rock	382
16	Korn	Pop/Rock	598	41	Bonnie Raitt	Pop/Rock	381
17	Trace Adkins	Country	571	42	Tim McGraw	Country	380
18	ZZ Top	Pop/Rock	560	43	Motley Crue	Pop/Rock	379
19	String Cheese Incident	Pop/Rock	550	44	Elvis Costello	Pop/Rock	376
20	The Moody Blues	Pop/Rock	517	45	Jimmy Buffett	Pop/Rock	373
21	Keith Urban	Country	506	46	Good Charlotte	Pop/Rock	372
22	Counting Crows	Pop/Rock	501	47	Carrie Underwood	Country	370
23	Cher	Pop/Rock	485	48	Hinder	Pop/Rock	368
24	Toby Keith	Country	483	49	John Mayer	Pop/Rock	366
25	The Black Crowes	Pop/Rock	481	50	Aerosmith	Pop/Rock	365

References

Abel, F., Diaz-Aviles, E., Henze, N., Krause, D., Siehndel, P., 2010. Analyzing the blogosphere for predicting the success of music and movie products. In: Advances in Social Networks Analysis and Mining (ASONAM), 2010 International Conference on IEEE, pp. 276–280.
Anderson, C., 2006. The Long tail: Why the Future of Business is Selling Less of More. Hyperion, New York.
Bates, D.M., 2010 lme4: Mixed-effects modeling with R. 470-474.
Bianchi, C., Andrews, L., 2015. Investigating marketing managers' perspectives on social media in Chile. J. Bus. Res. 68 (12), 2552–2559.
Brynjolfsson, E., Hu, Y.J., Smith, M.D., 2006. From niches to riches: Anatomy of the long tail. Sloan Manag. Rev. 47 (4), 67–71.
Brynjolfsson, E., Hu, Y., Smith, M.D., 2010. Research commentary-long tails vs. superstars: the effect of information technology on product variety and sales concentration patterns. Inf. Syst. Res. 21 (4), 736–747.
Brynjolfsson, E., Hu, Y., Simester, D., 2011. Goodbye Pareto principle, hello long tail: The effect of search costs on the concentration of product sales. Manag. Sci. 57 (8), 1373–1386.
Chellappa, R.K., Chen, C., 2009. MySpace killed the radio star? The impact of online sampling on song sales. In: ICIS 2009 Proceedings, p. 157.
Chen, H., De, P., Hu, Y.J., 2015. IT-enabled broadcasting in social media: an empirical study of artists' activities and music sales. Inf. Syst. Res. 26 (3), 513–531.
Chevalier, J.A., Mayzlin, D., 2006. The effect of word of mouth on sales: online book reviews. J. Mark. Res. 43 (3), 345–354.
Chintagunta, P.K., Gopinath, S., Venkataraman, S., 2010. The effects of online user reviews on movie box office performance: Accounting for sequential rollout and aggregation across local markets. Mark. Sci. 29 (5), 944–957.

Dellarocas, C., and Narayan, R., 2007. Tall heads vs. long tails: do consumer reviews increase the informational inequality between hit and niche products?. Robert H. Smith School of Business Research Paper, (06-056).
Dewan, S., Ramaprasad, J., 2014. Social media, traditional media, and music sales. MIS Q. 38 (1), 101–121.
Dewan, S., Hsu, V., 2004. Adverse selection in electronic markets: evidence from online stamp auctions. J. Ind. Econ. 52 (4), 497–516.
Dhar, V., Chang, E.A., 2009. Does chatter matter? The impact of user-generated content on music sales. J. Interact. Mark. 23 (4), 300–307.
Duggan, M., Ellison, N.B., Lampe, C., Lenhart, A., Madden, M., 2015. Social media update 2014. Pew Res. Cent. 19.
Elberse, A., Oberholzer-Gee, F., 2007. Superstars and underdogs: an examination of the long tail phenomenon in video sales. Mark. Sci. Inst. 4, 49–72.
Faraway, J.J., 2016. Extending the Linear Model With R: Generalized linear, Mixed Effects and Nonparametric Regression Models, second ed. CRC Press / Taylor & Francis, Florida.
Fong, Y., Rue, H., Wakefield, J., 2010. Bayesian inference for generalized linear mixed models. Biostatistics 11, 397–412.
Gayer, A., Shy, O., 2006. Publishers, artists, and copyright enforcement. Inf. Econ. Policy 18 (4), 374–384.
Godey, B., Manthiou, A., Pederzoli, D., Rokka, J., Aiello, G., Donvito, R., Singh, R., 2016. Social media marketing efforts of luxury brands: influence on brand equity and consumer behavior. J. Bus. Res. 69 (12), 5833–5841.
Hall, D.B., 2000. Zero-inflated Poisson and binomial regression with random effects: a case study. Biometrics 56, 1030–1039.
Kim, A.J., Ko, E., 2012. Do social media marketing activities enhance customer equity? An empirical study of luxury fashion brand. J. Bus. Res. 65 (10), 1480–1486.
Krueger, A.B., 2005. The economics of real superstars: The market for rock concerts in the material world. J. Labor Econ. 23 (1), 1–30.
Laird, N.M., Ware, J.H., 1982. Random-effects models for longitudinal data. Biometrics 963–974.
Lambert, D., 1992. Zero-inflated Poisson regression, with an application to defects in manufacturing. Technometrics 34, 1–14.
Lindgren, F., Rue, H., 2015. Bayesian Spatial Modelling with R-INLA. J. Stat. Softw. 63, 1–25.
Martins, T.G., Simpson, D., Lindgren, F., Rue, H., 2013. Bayesian computing with INLA: new features. Comput. Stat. Data Anal. 67, 68–83.
Morales-Arroyo, M., Pandey, T., 2010, June. Identification of critical eWOM dimensions for music albums. In: Management of Innovation and Technology (ICMIT), 2010 IEEE International Conference on IEEE, pp. 1230–1235.
Mortimer, J.H., Nosko, C., Sorensen, A., 2012. Supply responses to digital distribution: recorded music and live performances. Inf. Econ. Policy 24 (1), 3–14.
Peitz, M., Waelbroeck, P., 2006. Piracy of digital products: A critical review of the theoretical literature. Inf. Econ. Policy 18 (4), 449–476.
R Core Team, 2013. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
Rue, H., Martino, S., Chopin, N., 2009. Approximate Bayesian inference for latent Gaussian models by using integrated nested laplace approximations. J. R. Stat. Soc.: Ser. B (Stat. Methodol.) 71, 319–392.
Rui, H., Liu, Y., Whinston, A., 2013. Whose and what chatter matters? The effect of tweets on movie sales. Decis. Support Syst. 55 (4), 863–870.
Tan, T.F., Netessine, S., 2009. Is Tom Cruise threatened? Using Netflix Prize data to Examine the Long Tail of Electronic Commerce. Wharton Business School, University of Pennsylvania, Philadelphia.
Waldfogel, J., 2012. Music piracy and its effects on demand, supply, and welfare. In: Innovation Policy and the Economy, Vol. 12. University of Chicago Press, pp. 91–109.
Zentner, A., Smith, M., Kaya, C., 2013. How video rental patterns change as consumers move online. Manag. Sci. 59 (11), 2622–2634.
Zhu, F., Zhang, X., 2010. Impact of online consumer reviews on sales: the moderating role of product and consumer characteristics. J. Mark. 74 (2), 133–148.