The Annals of Applied Statistics
2024, Vol. 18, No. 2, 1686–1708
https://doi.org/10.1214/23-AOAS1854

HOW ARE PRELAUNCH ONLINE MOVIE REVIEWS RELATED TO BOX OFFICE REVENUES?

BY TIANYU GUAN[1,a], JASON HO[2,b], ROBERT KRIDER[2,c], JIGUO CAO[3,d] AND ANDREW FOGG[4,e]

[1]*Department of Mathematics and Statistics, York University,* [a]*tianyug1988@gmail.com*
[2]*Beedie School of Business, Simon Fraser University,* [b]*jason_ho_3@sfu.ca,* [c]*robert_krider@sfu.ca*
[3]*Department of Statistics and Actuarial Science, Simon Fraser University,* [d]*jiguo_cao@sfu.ca*
[4]*Roku, Inc.,* [e]*afogg@roku.com*

This paper studies the dynamic patterns of the prelaunch online movie reviews, or movie electronic word-of-mouth (eWOM), over time and investigates their relations to the subsequent box office revenues. The volume and valence of prelaunch eWOM have been shown to be early indicators of strong or weak box office. The time patterns of prelaunch eWOM evolution, which are essentially functional data, on the other hand, tend to be overlooked. We apply the functional principal component analysis, a dimension reduction technique in functional data analysis, to analyze the dynamic patterns of various quantile trajectories of the movie eWOM, instead of directly studying the whole eWOM functional data. The functional principal component (FPC) scores of quantile trajectories at various quantile levels are used to predict the box office revenues. We use the sparse group lasso method to select the quantile levels and individual FPC scores that make significant contributions to the prediction of box office revenues. The results show that compared with other measures, such as valence and variance, the top-end quantiles would be a better measure in capturing the relations between the prelaunch product ratings time pattern and launch sales.

1. Introduction. Online product reviews, commonly conceptualized as electronic word of mouth (eWOM), are one of the most active research areas in a variety of disciplines such as marketing, management information system, statistics, and data science (Babić Rosario, De Valck and Sotgiu (2020), Verma and Yadav (2021), Qahri-Saremi and Montazemi (2019)). Many of the eWOM studies focus on the effects of eWOM on consumer purchase behaviors, which in turn drive key business outcomes like purchase intention or sales. Such eWOM effects have been established and replicated across these studies, as shown in the meta-analyses by Purnawirawan et al. (2015), You, Vadakkepatt and Joshi (2015) and Babić Rosario et al. (2016).

In this research we consider quantile trajectories at various quantile levels, which are obtained from the 60-hour prelaunch users' ratings for movies widely released from October 2017 to March 2020 on imdb.com. We explore the evolution patterns of the quantile trajectories using functional principal component analysis (FPCA), a dimension reduction technique in functional data analysis (FDA), and we investigate how the resulting low-rank FPC scores help explain the box office revenues.

1.1. *Research questions.* There are two general research approaches in studying eWOM. The first approach uses individual consumer data, typically in experimental settings, to understand the influences of eWOM at a disaggregate level. For example, the meta-analysis of

Received June 2022; revised November 2023.
Key words and phrases. Electronic word of mouth, functional data analysis, functional principal component analysis, quantile functions, variable selection.

Purnawirawan et al. (2015) shows there is a robust positive effect of positive eWOM on individual consumers' attitudes toward the focal products. The second approach, on the other hand, uses field data, studying the relations of eWOM and subsequent sales at an aggregate level, like the whole U.S. market. Meta-analyses of these aggregate level by You, Vadakkepatt and Joshi (2015) and Babić Rosario et al. (2016) found that eWOM was positively correlated with aggregated sales, but the magnitude of the correlation would depend on what summary statistics or metrics were used in the specific studies.

We intend to contribute to the aggregate eWOM literature by answering two specific research questions:

1. Would the evolution patterns of product review ratings before product launch be associated with the launch sales?
2. Would quantiles be better summary statistics than arithmetic averages and standard deviations in capturing the relations between the prelaunch product ratings time pattern and launch sales?

To answer these two research questions, we collected users' ratings on a 1–10 scale for movies released from October 2017 to March 2020 on imdb.com (a popular movie rating website in the U.S.) and measured at hourly intervals during the 60 hours prior to the movies' releases in the U.S. The movies' box office for the opening week and the subsequent week (week two) were collected from Box Office Mojo by IMDbPro (boxofficemojo.com), which is the U.S. online box office reporting and analysis service that tracks box office revenues both domestically and internationally. It is important to note that online product reviews in general and movie reviews in particular are not naturally a single variable occurring only once but a data stream over time. Figure 1 shows 3D plots of the users' ratings for four selected movies—*Bohemian Rhapsody*, *What Men Want*, *The Predator*, and *Jurrasic World: Fallen Kingdom*, on imdb.com. As we can see, online product reviews are essentially a data stream along two dimensions, namely, the time dimension and the rating dimension. We observe that the online reviews for different movies exhibit different patterns over the time and rating dimensions, suggesting an opportunity to characterize each movie's data stream by a few parameters, and relate these parameters to the subsequent sales. As such, we propose to summarize the rating dimension by quantiles and then capture the timing dimension of the quantiles by FPCA. The resultant FPCA scores can then be used as predictors for the opening week (and second week) box office sales, essentially answering research question 1. To assess the usefulness of quantiles to summarize the rating dimension, we compare our approach to several benchmark models, like arithmetic averages and lagged variable coefficients, essentially addressing question 2.

While we will stay away from making any causation claim just like most field data studies in the literature, we intend to preserve the time order of occurrence between eWOM and sales, thus restricting the eWOM time series to the prelaunch period, that is, the 60 hours prior to the U.S. theatrical releases. In fact, our focus on prelaunch eWOM is particularly relevant to business decision makers in the movie industry. For example, if prelaunch eWOM can be used as an early indicator of a strong or weak opening week box office, movie theaters could adjust staffing and screen scheduling accordingly to improve their profitability (Eliashberg et al. (2009)). On the other hand, if prelaunch eWOM shows a weaker than expected opening week box office, the movie studio could consider to increase their online advertising buys to boost up the ticket sales.

1.2. *Literature review.* We discuss our intended contributions by reviewing the related literature. Houston et al. (2018) argued that prelaunch buzz was different from postlaunch eWOM, which has received much research attention in a variety of areas, like management

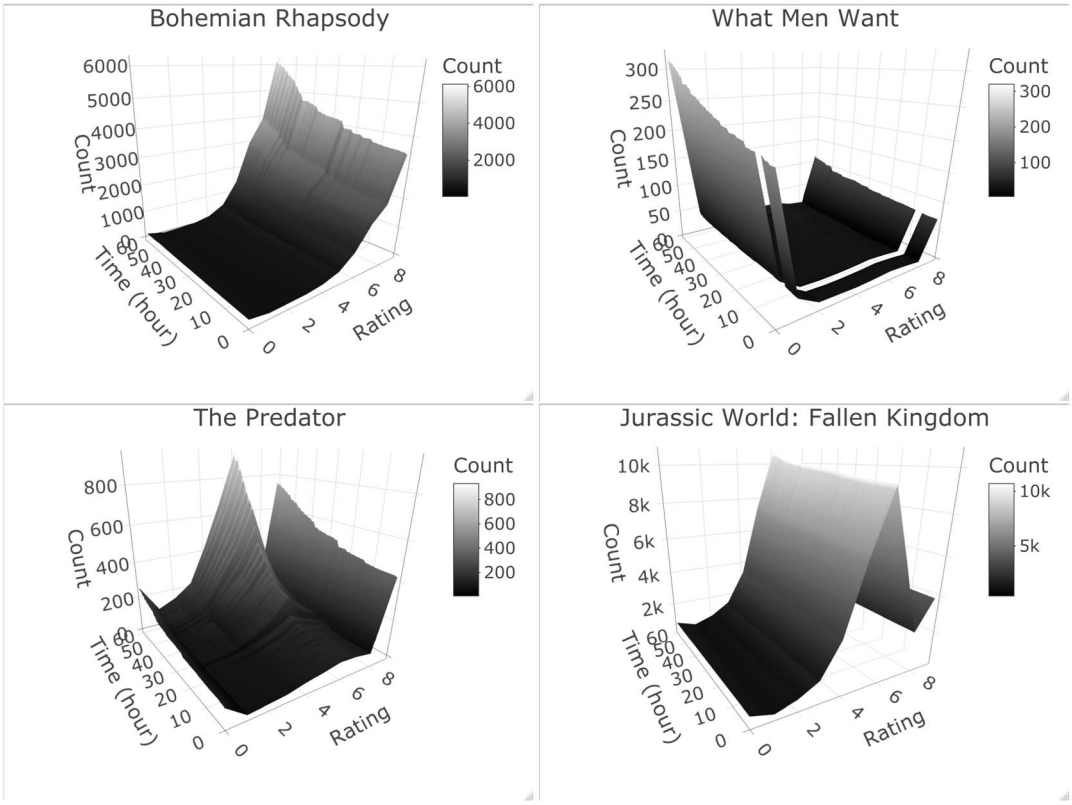

FIG. 1. *Counts of users of four selected movies across rating levels* $1-10$ *and across* 60 *hours before the theatrical release Hour* 60. *For* What Men Want, *the users' ratings are missing for hour* 8.

information systems (e.g., Clemons, Gao and Hitt (2006)). It is important to note that our prelaunch product ratings are different from their notion of prelaunch buzz (or interests) as they would be cumulated with postlaunch product ratings on the same platform such as imdb.com. Specifically, we believe there are four possible groups of IMDb users providing such prelaunch product reviews, namely, the keeners (who post reviews before watching the movies, acting more in line with buzz), the preview moviegoers, overseas moviegoers, and promotional reviews. Being theory-agnostic, one of our focuses is to study if this aggregation of product ratings would be helpful to managers.

Previous studies tend to ignore the time dimension of the prelaunch eWOM and capture the prelaunch eWOM as some cumulative measures at the point of product launch (e.g., Chintagunta, Gopinath and Venkataraman (2010), Dhar and Chang (2009), Liu (2006)). On the other hand, by using prelaunch buzz (not prelaunch product ratings) a few studies focus on identifying common shapes of the time evolution of prelaunch eWOM. Xiong and Bharadwaj (2014) found some common shapes in the detailed time structure of blogs and forum postings about upcoming video game releases could add predictive power beyond aggregate buzz, advertising spending, and other covariates to the first week video game sales. Gelper, Peres and Eliashberg (2018) identified spikes in the volume of prerelease social media mentioning that contribute to the predictability of product sales. Using the time series of a virtual stock market by amateur box office forecasters, Hollywood Stock Exchange, Foutz and Jank (2010) identified a small number of distinguishing shapes in the stock prices (e.g., the last-moment velocity spurt), which can help predict the subsequent movie box office performance.

While these "shape" studies are not exactly tracking prelaunch product ratings, like our study (e.g., Foutz and Jank (2010) were tracking amateur box office forecasters, not nec-

essarily overlapped with the actual moviegoers), the "interests" tracked can be considered consumer behaviors in an early stage of a consumer decision journey, suggesting the potential of the "shape" of the prelaunch product ratings in predicting the subsequent launch sales. The first research question is thus about whether the evolution patterns of product review ratings before product launch would be associated with the launch sales. To answer this research question, we use the FPCA method to recognize the time patterns of users' ratings, reducing the dimensionality of each movie's whole data stream to several principal components. Foutz and Jank (2010) and Xiong and Bharadwaj (2014) also used FDA in their studies of prelaunch interests and found FDA to be helpful in forecasting. Compared to their works, our paper studies the 60-hour quantile trajectories of the prelaunch movie ratings, and we apply a special version of FPCA, the PACE method to account for irregular missing data. FDA is a vital statistical tool specifically developed to analyze random trajectories, surfaces, or any multidimensional functions. There is an extensive literature on FDA (e.g., Ramsay, Hooker and Graves (2009), Ramsay and Silverman (2005), Wang, Chiou and Müller (2016)). FPCA has been widely used as an FDA dimension reduction method, which reduces the random trajectory to a set of random low-dimensional vectors called FPC scores. A general introduction to FPCA is provided in Chapters 8 and 9 in Ramsay and Silverman (2005). Asymptotic properties of FPCA have been studied by Dauxois, Pousse and Romain (1982), Zhang and Chen (2007), and Benko, Härdle and Kneip (2009). Rice and Silverman (1991), Cardot (2000), and Hall and Hosseini-Nasab (2006) have studied FPCA for fully or densely observed functional data. For a more difficult case where the data are sparsely and irregularly sampled, James, Hastie and Sugar (2000) handled it using a reduced rank mixed effects framework, whereas Yao, Müller and Wang (2005) proposed the principal components analysis through conditional expectation (PACE). A useful application of PACE is to impute the missing values for some subjects from the predicted trajectories.

In contexts where the time series of both eWOM and sales are available, lagged variables are the logical choice to study the dynamic effects of eWOM (e.g., vector autoregressive regression (VAR), as in Pauwels, Aksehirli and Lackman (2016)). In contrast, in the prelaunch period study like ours, where product ratings are the only time series and sales is only observed at one specific time point (e.g., opening week box office), the lagged variable method would cause interpretation and estimation issues. For example, in our 60-hour movie rating time series (see Figure 1), there is no clear way to determine the required number of lags in capturing the variation in the patterns across different movies. Even if we have enough movies (which is the unit of analysis in such a cross-sectional analysis) to estimate all 60 lags, we would have difficulties interpreting the signs and magnitudes of all 60 lagged variables. One of the few ways to apply conventional time series models, like the autoregressive (AR) model to this context, is probably to first fit the 60-hour time series to a single-variate AR and then use the estimated coefficients as predictors for the opening week box office. This is indeed one of the benchmark models we used in this study and will be discussed in more details later on.

Volume, valence, and to a lesser extent, variance of ratings are the common metrics to measure eWOM at the aggregate level (You, Vadakkepatt and Joshi (2015), Babić Rosario et al. (2016)). Their wide adoption in the literature can be attributed to two reasons. First, they correspond to the summary statistics we routinely compute and report for any samples. Specifically, the volume of eWOM is essentially the sample size while the valence is the sample's arithmetic average of the product review distribution, say on a $1-10$ review scale. Second, they allow intuitive interpretation for business decision makers: volume captures the "popularity" of the new product; valence reflects the opinion of an "average" consumer, and variance shows the diversity of opinions.

While volume and/or valence are robustly shown to be associated with sales, the relative effects of the two metrics seem to vary from study to study. For example, in the same context

of U.S. movie market, Liu (2006) found volume, not valence, to be the significant predictor of sales, whereas Chintagunta, Gopinath and Venkataraman (2010) obtained opposite results; that is, valence, not volume, would be the main driver to predict the box office sales. As shown in the meta-study by You, Vadakkepatt and Joshi (2015), there are several factors that may explain such mixed results, but the present study intends to examine the appropriateness of using arithmetic average in representing the positivity and negativity of eWOM. In other words, does the opinion of an "average" consumer a useful metric in predicting subsequent sales?

As shown in Figure 1, eWOM does not necessarily follow a unimodal distribution. In fact, most online product review ratings exhibit a positivity bias and extremity. Hu, Pavlou and Zhang (2009) termed the nonunimodal distribution J-shaped distribution. It is unclear to what extent the arithmetic average of a J-shaped distribution would be the informative metric. Instead of using the arithmetic average, we use quantiles as alternative metrics in comparison to the arithmetic average and answer the research question of whether quantiles would be better summary statistics than arithmetic averages in capturing the relations between the prelaunch product ratings time pattern and launch sales. To answer this research question, we compare the prediction power of FPC scores of various quantile trajectories to the arithmetic average and standard deviation trajectories of users' ratings while controlling for volume in a regression model. We apply variable selection methods, that is, lasso (Tibshirani (1996)) and sparse group lasso (Simon et al. (2013)), to select the most significant quantile level and individual FPC scores. There are multiple advantages of using quantile trajectories, which are discussed in detail in Section 2.2.

The present study applies a theory-agnostic variable reduction technique, FPCA, a core technique in FDA to the data stream of online product reviews so as to find the most efficient way to summarize online product reviews in both the time and rating dimensions in predicting the subsequent sales. By comparing with other summary statistics, such as the average and standard deviation, we find that the variations of the time trends of the 0.8th quantile trajectory across movies are most associated with the movies' opening week and week two box offices. Intuitively, as opposed to tracking an "average" moviegoer, movie industry practitioners would be better off tracking the eWOM of the top 20 percent fans in the days leading to official movie releases. We believe such a data-driven approach would complement the more theory-driven approach in the literature, moving us closer to the full understanding of this important phenomenon.

We structure the rest of the paper as follows. In Section 2 we begin with a discussion of the eWOM data and how to obtain the quantile trajectories from the eWOM data. In Section 3 we introduce how to apply the PACE method to explore the dynamic patterns of the quantile trajectories. We outline the regression models that use the FPC scores of various quantile trajectories to explain the box office revenues. In addition, we demonstrate that our models are reliable. In Section 4 we present the results and their interpretation. Section 5 discusses the managerial and theoretical implications. Conclusion is given in Section 6. The paper focuses on the prediction of the opening week's box office. Additional results for predicting week two box office revenues are provided in the Supplementary Material (Guan et al. (2024)).

2. Data description.

2.1. *eWOM data*. Users' ratings of movies with scales $1, \ldots, 10$ were collected hourly from IMDb (imdb.com) between October 2017 and March 2020 for all 851 movies released in this time period. IMDb is a subsidiary of Amazon, which is a worldwide popular movie database that provides extensive information of movie, TV show, cast information, and entertainment programs. Specifically, we extracted the users' ratings for the 60 hours preceding

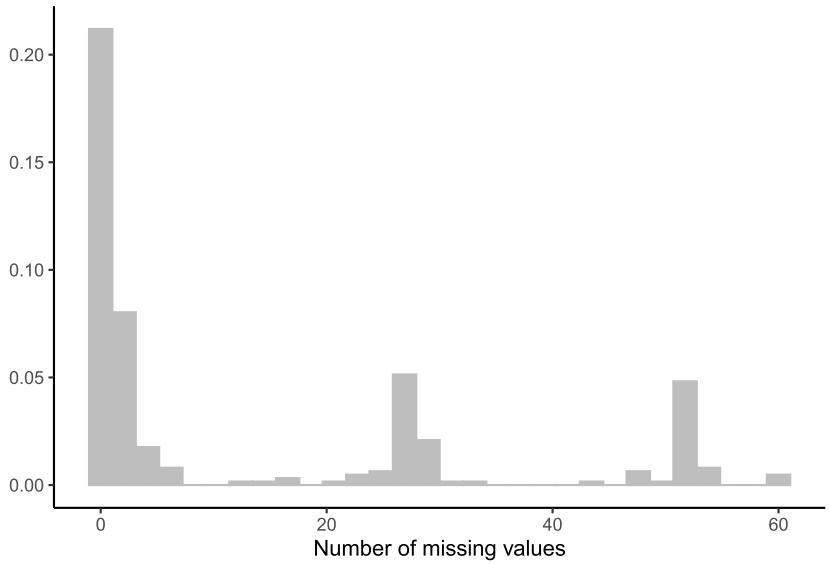

FIG. 2. *Distribution of the number of missing values for the* 301 *movies that opened in more than* 1000 *theatres*.

the theatrical weekend release, starting at 9 p.m. Eastern Standard Time (EST) on Tuesday and ending at 9 a.m. Friday EST. IMDb does not have a consistent policy of the time when consumer reviews first appear on the site, and often reviews are available several days earlier, but the 60-hour window captures most movies. The movies' box office for the opening week and the subsequent week (week two) were collected from boxofficemojo.com, another Amazon subsidiary.

We restrict the data to wide release movies, defined as those that opened in more than 1000 theatres. This filters out independent or platform release movies, which have time patterns of users' ratings that likely behave quite differently from the wide release movies. For example, independent movies would be exhibited in film festivals before the major release at theaters, resulting in movie reviews available much earlier. This restriction shrinks our dataset to 301 movies. We occasionally encountered technical problems extracting user ratings for some movies in some hours, creating some missing values in the ratings. The distribution of the number of missing values of the 301 wide released movies is displayed in Figure 2. We can observe a spike when the number of missing values is over 40. Therefore, we further removed the movies that have more than 40 missing ratings over 60 hours. This results in 257 movies in our final analysis dataset. Among the 257 movies, 150 movies have at least *one* missing value and 107 movies have no missing values. The missing data can be imputed by the PACE method, which was originally proposed by Yao, Müller and Wang (2005) to perform functional principal component analysis for sparse longitudinal data. We conduct simulation studies to verify the validity of applying the PACE method to impute the missing values; see the Supplementary Material for details.

The final analysis dataset has 60 hours of the cumulative counts of ratings from 1 to 10, for 257 movies, comprising over one million counts of ratings. In Figure 3 we plot the cumulative number of ratings pooled across all movies at time 60 (9 a.m. on Friday). We observe that the distribution is similar to the usual J-shaped distribution of WOM rating counts (Hu, Pavlou and Zhang (2009)). Interestingly, it varies from the common J-shape in a peak at rating 8 followed by a drop to rating 9, like a "W" superimposed on the "J."

2.2. *Quantile trajectories*. The sample quantile trajectories can be obtained from the cumulative users' ratings at each hour for each movie. The sample quantiles retain the detailed

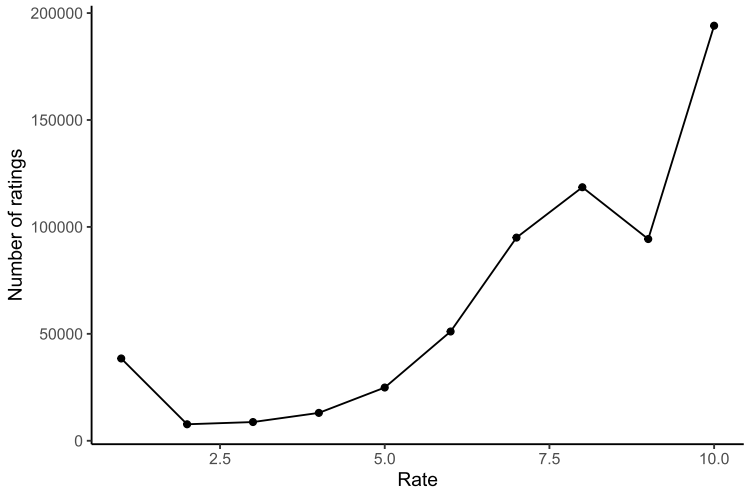

FIG. 3. *Cumulative rating counts distribution of all movies accumulated at* 9 *a.m. on Friday of the North American opening weekend.*

shape of the rating counts across time and ratings. There is an extensive literature on quantile functions (Gilchrist (2000), Sang, Begen and Cao (2021)) in statistical modelling. The advantages of using quantile functions are multiple. First, the quantile functions are defined on a fixed domain $[0, 1]$, which simplifies the process of analysis. Second, they enjoy many theoretical properties, which make them useful for modelling distributions. Details about the mathematical properties of quantile functions can be found in Ghosal et al. (2023). Third, as pointed out by Yang et al. (2020), quantile functions are easy to estimate by order statistics.

Let Q_{ij}^{α} denote the αth sample quantile of movie i at time t_j, where $\alpha \in [0, 1]$, $i = 1, \ldots, 257$ and $t_j = j$, $j = 1, \ldots, 60$. The contour plots of the sample quantiles allow easy visualization of distribution changes over time. Figure 4 shows the contour plots of the sample quantiles for four selected movies: *Pacific Rim Uprising*, *Fantastic Beasts: The Crimes of Grindelwald*, *The Nutcracker and the Four Realms*, and *Alita: Battle Angel*. The four contour plots exhibit different patterns across time. For example, for *Pacific Rim Uprising* and *Fantastic Beasts: The Crimes of Grindelwald*, the upward curvature of the color boundaries reflects the shift of the distribution to lower ratings. Specifically, the rating level of the 0.8th quantile of *Pacific Rim Uprising* goes from 10 to 7 as the time goes from *one* to 60 hours. The 0.8th quantile remains at 10 throughout the entire 60-hour time period of *Fantastic Beasts: The Crimes of Grindelwald*, while the 0.2, 0.4, and 0.6 quantile ratings decline over time. *The Nutcracker and the Four Realms* and *Alita: Battle Angel* are examples that exhibit different patterns. A pattern where the ratings are less polarized over time is observed for *The Nutcracker and the Four Realms*, whereas *Alita: Battle Angel* shows a pattern where the distribution is very stable over time.

3. Data analysis. We first apply the FPCA method (Lin, Wang and Cao (2016), Sang, Wang and Cao (2017), Nie et al. (2018), Nie and Cao (2020), Shi et al. (2021), Shi et al. (2022), Nie et al. (2022)) to discover the major sources of variations in the αth quantile trajectories of the users' ratings over the 60 hours prior to the releases of movies. The functional principal components (FPCs) are found by maximizing the variations of the quantile trajectories. The first few leading FPC scores of quantile trajectories with various α are used to predict the box office of the opening week and the subsequent week. We use variable selection methods (sparse group lasso and lasso) to select the α and individual FPC scores that make the significant contribution to predicting the box office.

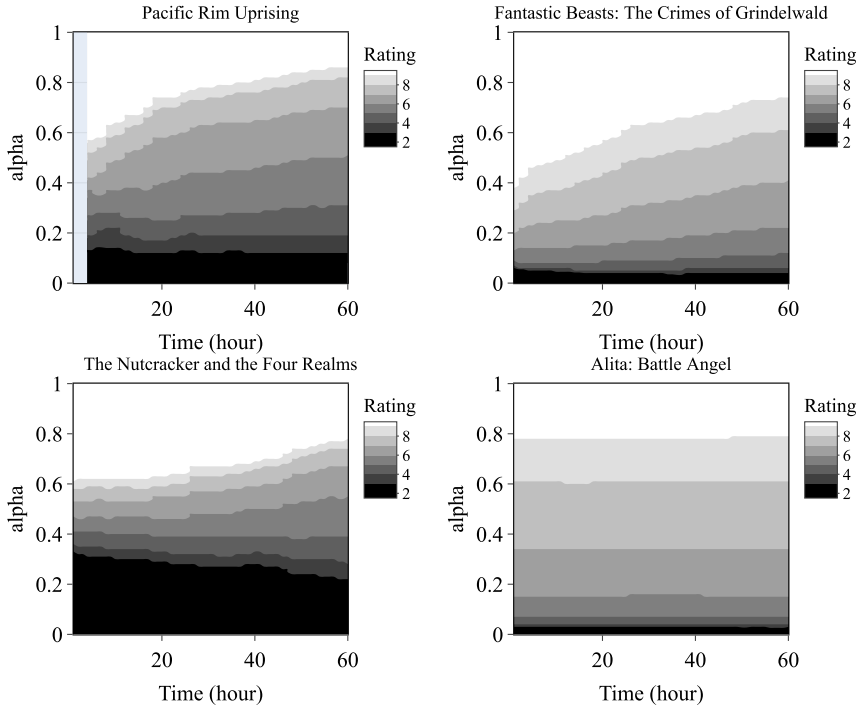

FIG. 4. *Contour plots of the sample quantile functions for four selected movies. For* Pacific Rim Uprising, *the users' ratings are missing for the first three hours.*

3.1. *Functional principal component analysis.* For fixed α, let $Q_1^\alpha(t), \ldots, Q_n^\alpha(t)$ be n independent realizations of a smooth random quantile function $Q^\alpha(t)$ with unknown mean function $\mu^\alpha(t) = E(Q^\alpha(t))$ and covariance function $G^\alpha(s,t) = \mathrm{Cov}(Q^\alpha(s), Q^\alpha(t))$, where $s, t \in \mathcal{T} = [0, 60]$. Based on Mercer's Theorem, $G^\alpha(s,t)$ has an orthogonal expansion in $L^2(\mathcal{T})$,

$$G^\alpha(s,t) = \sum_{k=1}^{\infty} \lambda_k^\alpha \psi_k^\alpha(s) \psi_k^\alpha(t),$$

where λ_k^α are nonincreasing eigenvalues and $\psi_k^\alpha(t)$ are orthonormal eigenfunctions of the covariance function $G^\alpha(s,t)$. We assume that each individual's underlying random function $Q_i^\alpha(t)$ can be expressed in terms of a Karhunen–Loève expansion as

$$Q_i^\alpha(t) = \mu^\alpha(t) + \sum_{k=1}^{\infty} \xi_{ik}^\alpha \psi_k^\alpha(t),$$

where ξ_{ik}^α are uncorrelated random variables with mean 0 and variance λ_k^α. We call $\psi_k^\alpha(t)$ functional principal components (FPCs) and ξ_{ik}^α the corresponding FPC scores. Let Q_{ij}^α be observed αth sample quantile of movie i at time t_j, $t_j = 1, \ldots, 60$, and assume Q_{ij}^α are not missing at N_i ($N_i \le 60$) time points $t_{i_1}, \ldots, t_{i_{N_i}}$. $Q_{ii_l}^\alpha$ is modeled as

$$\begin{aligned} Q_{ii_l}^\alpha &= Q_i^\alpha(t_{i_l}) + \epsilon_{ii_l}^\alpha \\ &= \mu^\alpha(t_{i_l}) + \sum_{k=1}^{\infty} \xi_{ik}^\alpha \psi_k^\alpha(t_{i_l}) + \epsilon_{ii_l}^\alpha, \end{aligned} \tag{1}$$

where $l = 1, \ldots, N_i$ and ϵ_{ii_l} are random errors following the normal distribution with mean 0 and variance $(\sigma^\alpha)^2$.

In (1) we first estimate the mean function $\mu^\alpha(t)$ by applying local linear smoothers based on the pooled data from all the quantile trajectories. The estimated covariance function $\hat{G}^\alpha(s,t)$ is obtained by local linear surface smoother. The eigenvalues λ_k^α and eigenfunctions $\psi_k^\alpha(t)$ are obtained by solving the eigenequations $\int_{\mathcal{T}} \hat{G}^\alpha(s,t)\psi_k^\alpha(s)\,\mathrm{d}s = \lambda_k^\alpha \psi_k^\alpha(t)$, subject to $\int_{\mathcal{T}}(\hat{\psi}_k^\alpha(t))^2\,\mathrm{d}t = 1$ and $\int_{\mathcal{T}} \hat{\psi}_k^\alpha(t)\hat{\psi}_j^\alpha(t)\,\mathrm{d}t = 0$ for $k \neq j$. The estimation of σ^α is based on a local quadratic surface smoother along the direction perpendicular to the diagonal and a local linear surface smoother in the direction of the diagonal. Details of the above estimations are given in the Appendix of Yao, Müller and Wang (2005).

As discussed in Section 2, there exist irregular missing values in the eWOM data, due to some unexpected data collection issues like down servers. We applied the PACE (Yao, Müller and Wang (2005)) method to impute the missing values and obtain the FPC scores by the conditional expectations. They assume that, in (1), ξ_{ik}^α and ϵ_{iil}^α are jointly Gaussian distributed. For simplicity of notation, we ignore the α superscript in the following definition and formula. Let $\tilde{\boldsymbol{Q}}_i = (Q_{ii_1}^\alpha, \ldots, Q_{ii_{N_i}}^\alpha)^{\mathrm{T}}$ be the vector with N_i observations that are not missing for movie i at times $t_{i_1}, \ldots, t_{i_{N_i}}$. We estimate the ξ_{ik} by

$$\hat{\xi}_{ik} = \hat{E}(\xi_{ik} \mid \tilde{\boldsymbol{Q}}_i) = \hat{\lambda}_k \hat{\boldsymbol{\psi}}_{ik}^{\mathrm{T}} \hat{\boldsymbol{\Sigma}}_{\tilde{\boldsymbol{Q}}_i}^{-1}(\tilde{\boldsymbol{Q}}_i - \hat{\boldsymbol{\mu}}_i),$$

where $\hat{\boldsymbol{\psi}}_{ik} = (\hat{\psi}_k(t_{i_1}), \ldots, \hat{\psi}_k(t_{i_{N_i}}))^{\mathrm{T}}$, $\hat{\boldsymbol{\mu}}_i = (\hat{\mu}(t_{i_1}), \ldots, \hat{\mu}(t_{i_{N_i}}))^{\mathrm{T}}$, and $\hat{\boldsymbol{\Sigma}}_{\tilde{\boldsymbol{Q}}_i}$ is an $N_i \times N_i$ matrix with elements $(\hat{\boldsymbol{\Sigma}}_{\tilde{\boldsymbol{Q}}_i})_{(j,l)} = \hat{G}(t_{i_j}, t_{i_l}) + \hat{\sigma}^2\delta_{jl}$ with δ_{jl} being 1 if $j = l$ and 0 otherwise. The estimated $\hat{\lambda}_k$, $\hat{\boldsymbol{\psi}}_{ik}$, $\hat{\boldsymbol{\mu}}_i$, $\hat{G}(t_{i_j}, t_{i_l})$, and $\hat{\sigma}$ are obtained based on the procedures introduced in the previous paragraph.

In practice, we usually select the top K FPCs to represent the random quantile trajectory. Then $Q_i(t) = \mu(t) + \sum_{k=1}^K \xi_{ik}\psi_k(t)$. The value of K is chosen such that the cumulative percentage of variance explained by the top K estimated FPCs exceeds a desired level. In this paper, we set the desired level at 99% which may results in a large value for K. Now, suppose that the observed sample quantile trajectory is missing at t_{i_0}; the missing value can be imputed as

$$\hat{Q}_{ii_0} = \hat{\mu}(t_{i_0}) + \sum_{k=1}^K \hat{\xi}_{ik}\hat{\psi}_k(t_{i_0}). \tag{2}$$

3.2. *Box office prediction.* We use the estimated FPC scores of the quantile trajectories to predict the box office revenues in the opening week and the subsequent week. p quantile levels $\alpha_1, \ldots, \alpha_p$ are considered. We expect to select a set of FPC scores that are predictive of box office and, at the same time, interpretable. To achieve these objectives, we treat each quantile level as a group and use the sparse group lasso technique (Simon et al. (2013)) to select the FPC scores at both the group and within-group levels.

Let Y_i be the logarithm of the box office revenue with a unit of a million dollars for the opening week (or subsequent week), and let $\zeta_{ik}^{(l)} = \hat{\xi}_{ik}^{\alpha_l}/\sqrt{\hat{\lambda}_k^{\alpha_l}}$ denote the standardized kth FPC score of the α_lth quantile trajectory for movie i. In addition, we include the logarithm of the total number of ratings, that is, the logarithm of the cumulative number of ratings at time 60 for movie i, to account for the rating volume effect. Let X_i denote the standardized logarithm of the total number of ratings at hour 60 to have mean zero and unit norm. We use the linear regression model for the data,

$$\boldsymbol{Y} = b_0 + \beta_0 \boldsymbol{X} + \sum_{l=1}^p \boldsymbol{\zeta}^{(l)}\boldsymbol{\beta}^{(l)} + \boldsymbol{\epsilon}, \tag{3}$$

where b_0 is the intercept, $\boldsymbol{Y} = (Y_1, \ldots, Y_n)^{\mathrm{T}}$, $\boldsymbol{X} = (X_1, \ldots, X_n)^{\mathrm{T}}$, $\boldsymbol{\zeta}^{(l)}$ is an $n \times K_l$ matrix with elements $\boldsymbol{\zeta}^{(l)}_{(i,k)} = \zeta^{(l)}_{ik}$, $\boldsymbol{\beta}_l = (\beta_{l1}, \ldots, \beta_{lK_l})^{\mathrm{T}}$, and $\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_n)^{\mathrm{T}}$ are random errors. The sparse group lasso estimators of b_0, β_0 and $\boldsymbol{\beta}^{(l)}$, $l = 1, \ldots, l$ are obtained by minimizing

$$\frac{1}{2}\left\|\boldsymbol{Y} - b_0 - \beta_0 \boldsymbol{X} - \sum_{l=1}^{p} \boldsymbol{\zeta}^{(l)}\boldsymbol{\beta}^{(l)}\right\|_2^2 + (1-\gamma)\lambda \sum_{l=1}^{p} \sqrt{K_l}\left\|\boldsymbol{\beta}^{(l)}\right\|_2 + \gamma\lambda\|\boldsymbol{\beta}\|_1, \tag{4}$$

where $\gamma \in [0, 1]$ and $\boldsymbol{\beta} = (\boldsymbol{\beta}^{(1)\mathrm{T}}, \ldots, \boldsymbol{\beta}^{(p)\mathrm{T}})^{\mathrm{T}}$. As Simon et al. (2013) pointed out, $\gamma \in [0, 1]$ leads to a convex combination of the lasso and group lasso penalties. When $\gamma = 0$, the sparse group lasso penalty reduces to the group lasso penalty which is designed for group selection. When $\gamma = 1$, the sparse group lasso penalty is the lasso penalty and can only do the individual variable selection. When $0 < \gamma < 1$, the sparse group lasso penalty can be used for variable selection at the group and within-group individual levels simultaneously.

We compare the predictive accuracy of model (3) with several other models. We first compare model (3) to a baseline model with only one covariate, the rating volume effect. The model has the form $\boldsymbol{Y} = b_0 + \beta_0 \boldsymbol{X} + \boldsymbol{\epsilon}$. We further compare (3) with a set of p models, one for each of the p quantile trajectories. All p models account for the rating volume effect, and each model uses the FPC scores obtained from one quantile trajectory with a specific quantile level l. The p models are $\boldsymbol{Y} = \beta_0 \boldsymbol{X} + \boldsymbol{\zeta}^{(l)}\boldsymbol{\beta}^{(l)} + \boldsymbol{\epsilon}$, $l = 1, \ldots, p$. For each model, we use the lasso method to select the relevant FPC scores at each α_lth quantile level. We estimate $\boldsymbol{\beta}^{(l)}$ by minimizing

$$\frac{1}{2}\left\|\boldsymbol{Y} - b_0 - \beta_0 \boldsymbol{X} - \boldsymbol{\zeta}^{(l)}\boldsymbol{\beta}^{(l)}\right\|_2^2 + \kappa_l\left\|\boldsymbol{\beta}^{(l)}\right\|_1, \tag{5}$$

where κ_l are nonnegative tuning parameters.

3.3. *Model validation.* We estimated the FPC scores for each quantile trajectory based on all 257 movies. After we obtained the estimated FPC scores, we fit the linear regression models introduced in the previous section. In the fitting procedure, we used the cross-validation (CV) method to select the optimal values for the sparse group lasso tuning parameter λ in (4) and the lasso tuning parameters κ_l in (5). The model prediction performance is evaluated based on a randomly selected test dataset. Specifically, 47 of the 257 movies are randomly selected and held out as a test dataset, and the remaining 210 movies are used in a seven-fold CV to select the sparse group lasso and lasso tuning parameters. The sparse group lasso and lasso tuning procedures allow us to select the FPC scores to be entered into or dropped from the regression model. With this setting we used $47/257 = 18\%$ randomly selected data as the test dataset and the remaining 82% data as the training and validation datasets. Our choice is close to the commonly used split percentages in the literature: 80% of the data for the training and validation datasets and 20% for the test dataset. For the remaining 210 movies, $180/210 = 86\%$ of the data are used as a training dataset to fit the model, and 14% of them are used as a validation dataset to select the tuning parameters. We repeated this process 1000 times, with a different randomly selected set of 47 movies held out each time to evaluate the prediction performance. We repeated the process a sufficiently large number of times such that every movie is selected into the test dataset multiple times. With 1000 times of random partitions, each movie is expected to appear in the test dataset for $1000 \times 47/257 = 183$ times. One commonly used metric for evaluating the prediction performance is the prediction mean squared error (PMSE). However, since the opening week box office revenues varied a lot for the 257 movies, the PMSE tend to be dominated by the movies with large week one box office revenues. For example, the opening week box office revenues of *Playmobil: The Movie* and *Avengers: Endgame* were \$822,723 and \$473,894,638, respectively. Therefore, we evaluate

the prediction performance by the root mean squared relative prediction error (RMS-RPE). Let S_j be the test dataset randomly selected in the jth partition of the entire dataset, and let $\hat{Y}_i^{(-j)}$, $i \in S_j$, denote the estimated response obtained on the remaining training set S_j^c. The RMS-RPE for the jth run is defined as

$$\sqrt{\frac{1}{n_{\text{test}j}} \sum_{i \in S_j} \left(\frac{\exp(Y_i) - \exp(\hat{Y}_i^{(-j)})}{\exp(Y_i)} \right)^2}, \quad j = 1, \ldots, 1000,$$

where $\exp(Y_i)$ and $\exp(\hat{Y}_i^{(-j)})$ are the observed and estimated box office revenues of the opening week, respectively, and $n_{\text{test}j}$ is the size of the jth test dataset for all 1000 data partitions. In our setting, $n_{\text{test}j} = 47$ for all j.

4. Results. We first performed the FPCA to quantile trajectories at multiple quantile levels of users' ratings over the 60 hours prior to the releases of movies for 257 movies. The FPCA provides insights for the dynamic effects of the prelaunch eWOM time series. The leading FPC scores extract major information from the quantile trajectories and are thus used to predict the opening week's and the second week's box office revenues.

4.1. *FPCA of the prelaunch eWOM quantile trajectories and missing data imputation.* We first consider five quantile levels at $\alpha_1 = 0.1$, $\alpha_2 = 0.25$, $\alpha_3 = 0.5$, $\alpha_4 = 0.75$, $\alpha_5 = 0.9$. The FPCs of the quantile trajectories at the five quantile levels are estimated with the PACE method. Although the prelaunch eWOM quantile trajectories exhibit complicated dynamic patterns, the top few FPCs not only display simple shapes but also capture over 90% of the total variations. For example, the top three FPCs of the 0.75th quantile trajectory explain 93.29% of the total variation among 257 curves. The top FPCs of the five quantile trajectories are plotted in the Supplementary Material. We will defer the discussion of the specific shapes of these top FPCs to Section 4.3, when we interpret the results of a model with the best predictive performance.

After we obtained the estimated FPCs and the FPC scores, we imputed the missing values from the predicted quantile trajectories (see (2)). For example, the movie *The Grinch* has 27 out of 60 missing data from hour 1 to hour 27. The left panel of Figure 5 displays the observed sample quantile contour for *The Grinch*, and the recovered sample quantile function is shown in the right panel.

As discussed in Section 3.1, the number of FPCs are chosen such that the cumulative percentage of variance, explained by the first K FPCs, exceeds 99%. Based on this selection

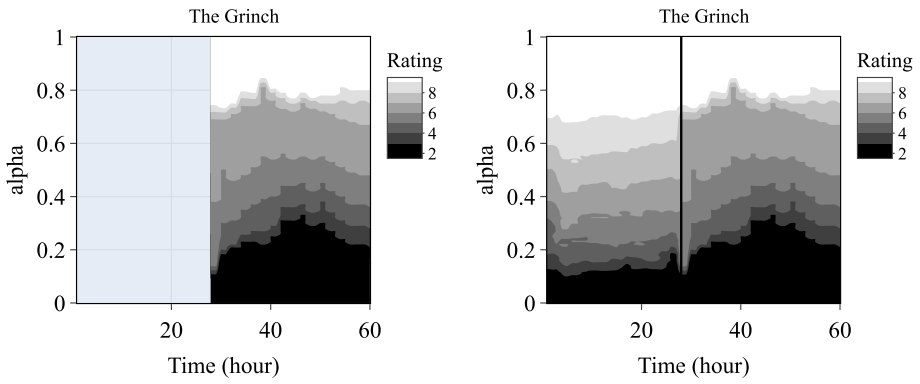

FIG. 5. *Contour plots of the observed sample quantiles for* The Grinch (*left panel*) *and of the recovered quantiles with missing data imputed by the PACE method* (*right panel*).

TABLE 1

Percentages of variance explained by the top K estimated FPCs with the cumulative percentage of variation included in the parentheses. The cumulative percentage of variance of the top K estimated FPCs exceeds a desired level of 99% of the total variations

FPC	$Q^{0.10}$	$Q^{0.25}$	$Q^{0.5}$	$Q^{0.75}$	$Q^{0.9}$
1	95.64 (95.64)	95.22 (95.22)	89.14 (89.14)	78.74 (78.74)	60.89 (60.89)
2	1.83 (97.47)	2.46 (97.68)	5.91 (95.05)	8.68 (87.42)	20.89 (81.78)
3	0.66 (98.13)	0.76 (98.44)	1.50 (96.55)	5.87 (93.29)	7.41 (89.19)
4	0.44 (98.57)	0.53 (98.97)	1.00 (97.55)	2.00 (95.29)	3.61 (92.80)
5	0.41 (98.98)	0.24 (99.21)	0.63 (98.18)	1.81 (97.10)	1.98 (94.78)
6	0.20 (99.18)		0.42 (98.60)	0.56 (97.66)	1.34 (96.12)
7			0.28 (98.88)	0.51 (98.17)	1.10 (97.22)
8			0.24 (99.12)	0.31 (98.48)	0.78 (98.00)
9				0.27 (98.75)	0.65 (98.65)
10				0.21 (98.96)	0.52 (99.17)
11				0.20 (99.16)	

rule, the number of FPCs selected for each of the 0.1, 0.25, 0.5, 0.75, and 0.9th quantile trajectories are *six*, *five*, *eight*, 11, and 10, respectively. The corresponding FPC scores of the selected FPCs are used for predicting the box office revenues. Table 1 shows the amount of total variation explained by each FPC. We can observe that, in general, the number of FPCs selected is greater for the quantile trajectory Q^{α} with a larger α. The first FPCs of $Q^{0.10}$, $Q^{0.25}$, and $Q^{0.5}$ capture over 89% of the total variance, indicating that the type of variation of the first FPC dominates all other types of variations. For $Q^{0.75}$ and $Q^{0.9}$, the type of variations of the subsequent FPCs are also important to show the full picture of the total variations.

4.2. *Relations between the quantile FPC scores and the box office revenues.* We use the selected FPC scores of the five quantile trajectories to predict the box office revenues in the opening week and the second week. We start with predicting the opening week's box office. We first consider model (3), where the response is the logarithm of the box office revenue for the opening week and the predictors are the standardized logarithm of the total number of ratings, that is, the rating volume effect, and the estimated FPC scores from the five quantile trajectories. There are a total of 257 movies. Each of the five quantile trajectories is treated as a group, that is, $p = 5$, and the selected number of estimated FPC scores for each group are $K_1 = 6$, $K_2 = 5$, $K_3 = 8$, $K_4 = 11$, and $K_5 = 10$. The sparse group lasso method (4) is used for discovering the sparsity of groups and within each group. As introduced in Section 3.3, we randomly select 47 movies from the 257 movies as a test dataset and use the remaining 210 movies to select the sparsity tuning parameter λ by a seven-fold CV. The procedure is repeated independently 1000 times. For the other parameter γ in (4), since we expect the "overall sparsity" and would like to encourage grouping, we follow Simon et al. (2013) and set $\gamma = 0.95$. Table 2 shows the percentage of times that each predictor (FPC score) is chosen in the 1000 runs. The first two FPC scores of the 0.75th quantile trajectory are selected by almost all the 1000 runs with proportions of 100% and 99.9%. Other FPC scores that are selected for more than 90% of the time include the second FPC scores of the 0.5th and 0.9th quantiles (the proportions are 99.5% and 96.1%) and the third FPC score of the 0.75th quantile with a proportion of 90.2%. In summary, the FPC scores of the 0.1th and 0.25th quantiles seem to be selected much less than the higher-end quantiles of $Q^{0.5}$, $Q^{0.75}$, and $Q^{0.9}$. The boxplot of the RMS-RPE is shown in Figure 6.

TABLE 2
The proportion (%) of the 1000 runs that each FPC score is chosen

FPC score	$Q^{0.10}$	$Q^{0.25}$	$Q^{0.5}$	$Q^{0.75}$	$Q^{0.9}$
1	71.9	19.3	48.1	100	22.4
2	31.7	71.3	99.5	99.9	96.1
3	74.2	32.6	76.2	90.2	14.2
4	44.2	38.2	37.6	57.5	41.8
5	35.3	59.8	59.5	19.1	41.9
6	22.3		28.6	34.7	49.6
7			31.1	16.4	20.0
8			14.5	33.0	30.6
9				21.1	32.5
10				24.9	30.6
11				65.4	

To investigate the effects of the FPC scores of each quantile trajectory on the box office revenues in the opening week, we fit five separate linear regression models, each of which includes the rating volume effect and only the FPC scores estimated from one quantile trajectory. The lasso method (5) is used to select the FPC scores for each model. In other words, we compare across models, each containing K_l FPC scores of quantile l, and search for the model with the lowest RMS-RPE of the opening week box office calculated on the test dataset for each run. In addition, a baseline model with the rating volume effect as the only predictor is considered. Figure 6 shows the boxplots of the RMS-RPE of box office in the opening week on the test datasets based on 1000 runs. We observe that the model with the 0.75th quantile FPC scores outperforms the other models. Models with the 0.1th and 0.25th quantile FPC scores have similar performance as the baseline model, indicating that the FPC scores from the two quantile trajectories do not increase the prediction power. Models with the 0.9th quantile and all five quantiles have the larger variances in the prediction errors.

The model using the 0.75th quantile FPC scores perform the best in terms of the prediction error. Next, we refine our search for the best quantile model to smaller intervals of 0.5 around 0.75. Specifically, we consider a set of lasso regression models, each of which uses the FPC

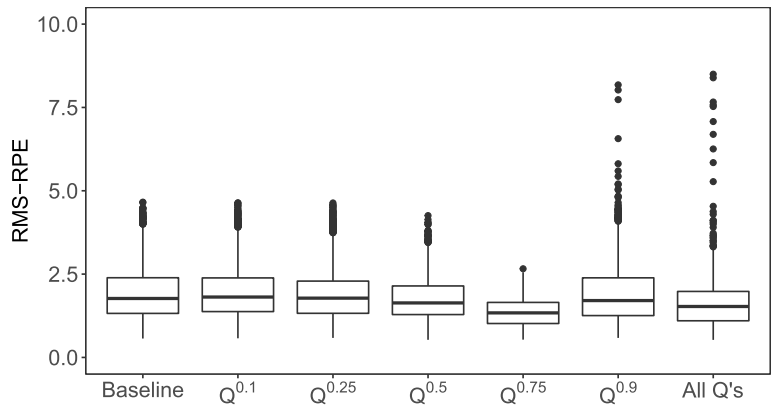

FIG. 6. *Boxplots of the root mean squared relative prediction error (RMS-RPE) of the box office in the opening week on the test datasets based on 1000 random training/test splits for various models ("Baseline:" the baseline model; "Q^p", $p = 0.1, 0.25, 0.5, 0.75,$ and 0.9: the lasso regression model with predictors being the FPC scores from one of the quantile trajectory Q^p; "All Q's": the sparse group lasso regression with predictors being the FPC scores from all five quantile trajectories).*

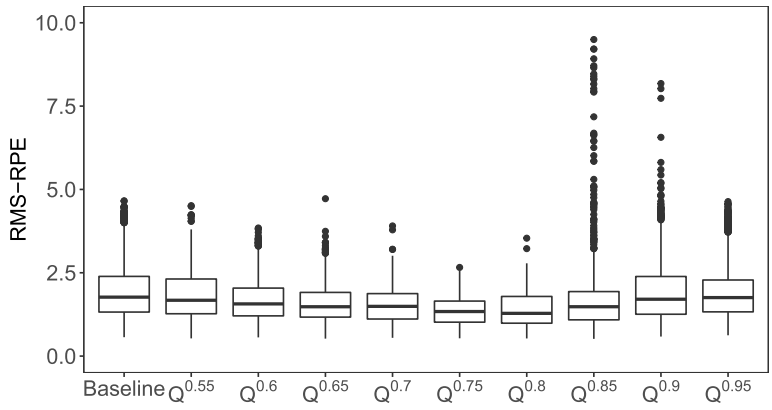

FIG. 7. *Boxplots of the root mean squared relative prediction error* (*RMS-RPE*) *of the box office in the opening week on the test datasets based on* 1000 *random training/test splits for various models* (*"Baseline:" the baseline model*; *"Q^p", $p = 0.55, 0.6, \ldots, 0.95$: the lasso regression model with predictors being the FPC scores from the quantile trajectory Q^p*.

scores of one of the quantile trajectories of $Q^{0.55}$, $Q^{0.6}$, $Q^{0.65}$, $\ldots$, $Q^{0.95}$ as predictors. The boxplots of the RMS-RPE of box office revenues for these models are displayed in Figure 7. The 0.75th and 0.8th quantile FPC scores perform the best. The prediction errors based on the 0.8th quantile FPC scores have the smallest median, whereas the 0.75th quantile achieves the smallest variance and least extreme outliers.

The above models are also used to predict the second week's box office revenues. The corresponding results regarding the prediction errors are provided in the Supplementary Material. The results for the opening week's and second week's box office revenues are similar. We conclude that the 0.75th quantile FPC scores perform the best among all five quantile FPC scores. However, by a finer grid search, the 0.8th quantile FPC scores outperform all the other quantiles, which have both the smallest median of the prediction errors.

We next compare the models with quantile FPC scores to several benchmark models. The 0.8th quantile FPC scores perform comparably to the 0.75th quantile FPC scores for predicting the opening week's box office revenues and outperform all the other quantiles for the second week's box office prediction. We, therefore, compare the benchmark models with the model based on the 0.8th quantile FPC scores. The first benchmark model uses an autoregressive (AR) model to fit the 0.8th quantile trajectories. For movie i, let $Q_{i,t}^{0.8} = Q_i^{0.8}(t)$. We use the PACE method to impute the missing values and fit an AR(2) model to the detrended quantile time series $Q_{i,t}^{0.8} = \alpha_{i1} Q_{i,t-1}^{0.8} + \alpha_{i2} Q_{i,t-2}^{0.8} + e_{i,t}$, where α_{i1} and α_{i2} are parameters and $e_{i,t}$ is the white noise error. Although the lack of box office sales in the prelaunch period makes it impossible to use dynamic models like VAR, as discussed in Section 1, we can check to what extent the two parameters in AR(2) can capture the time pattern during the prelaunch period and be related to the box office revenues in a cross-sectional manner. The rating volume effect X and estimated time series parameters $\hat{\alpha}_{i1}$ and $\hat{\alpha}_{i2}$ are used in a linear regression model to predict the box office revenues in the opening week and subsequent week. Each of the other benchmark models uses lasso method with one of the following set of predictors: (1) the rating volume effect X and the FPC scores of the average ratings over time, a valence by time measure, (2) X and the FPC scores of the standard deviation of the ratings over time, a standard deviation (variance) by time measure, (3) FPC scores of the logarithm of the counts of ratings over time, a volume by time measure, (4) FPC scores of the logarithm of the counts of ratings that are greater than 5 over time, a positive volume by time measure, and (5) FPC scores of the logarithm of the counts of ratings that are less than or equal to 5 over time, a negative volume by time measure. The RMS-RPE are shown in

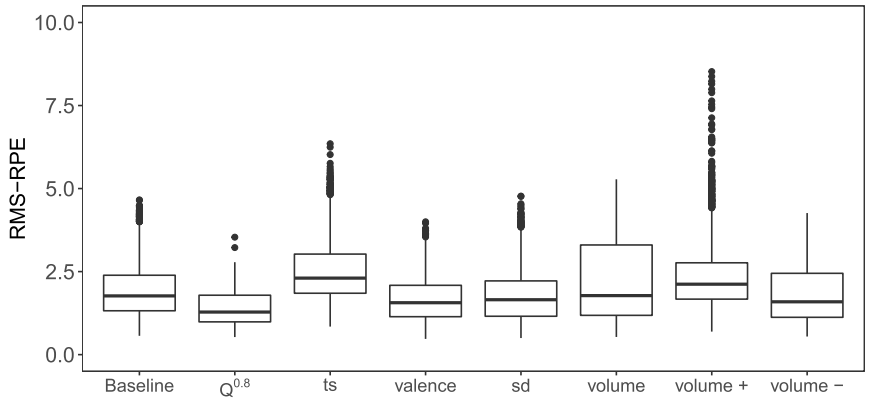

FIG. 8. *Boxplots of the root mean squared relative prediction error (RMS-RPE) of the box office in the opening week on the test datasets based on* 1000 *random training/test splits for various models (“Baseline:” the baseline model; “$Q^{0.8}$”: the lasso model with the* 0.8*th quantile FPC scores; “ts:” linear regression model with the* $AR(2)$ *parameters; “valence, sd, volume, volume +,” and “volume -” represent the lasso regression models with the valence by time measure, standard deviation by time measure, volume by time measure, positive volume by time measure, and negative volume by time measure, respectively.*

Figure 8 for the opening week's box office. Compared with the benchmark models, the 0.8th quantile FPC scores achieve the most significant prediction power. The comparisons of the benchmark models with the 0.8th quantile FPC scores for the second week's box office are provided in the Supplementary Material.

4.3. *Investigation of the prediction power of the* 0.8*th quantile FPC scores.* In Section 4.2 we used 10 FPC scores of the 0.8th quantile trajectory in the lasso regression model such that the cumulative percentage of variance explained by the first 10 FPCs exceeds 99%. Table 3 summarizes for the prediction of the opening week's box office revenues by the proportion of the 1000 runs that each of the 10 FPC scores is chosen and the proportion of variation explained by each FPC. The results for the week two box office prediction is in the Supplementary Material. We can observe that the top three FPC scores are chosen by the lasso method in almost all runs, and they capture 92.34% of the total variation.

We initially include as many FPCs as possible in our model by setting a high threshold for the total percentages of variance explained by these FPCs, because a high-order FPC could be

TABLE 3

The proportion of the 1000 *random training/test splits for the prediction of first week's box office that each* 0.8*th quantile FPC score is chosen and the proportion of variation (in percentages) explained by each FPC. The cumulative percentage of variations explained are included in the parentheses*

FPC	Chosen %	% Variation explained
1	100	77.91 (77.91)
2	99.0	8.33 (86.24)
3	99.2	6.10 (92.34)
4	90.8	2.52 (94.86)
5	87.8	2.28 (97.14)
6	93.4	0.65 (97.79)
7	88.9	0.37 (98.16)
8	92.9	0.35 (98.51)
9	91.4	0.35 (98.86)
10	89.9	0.23 (99.09)

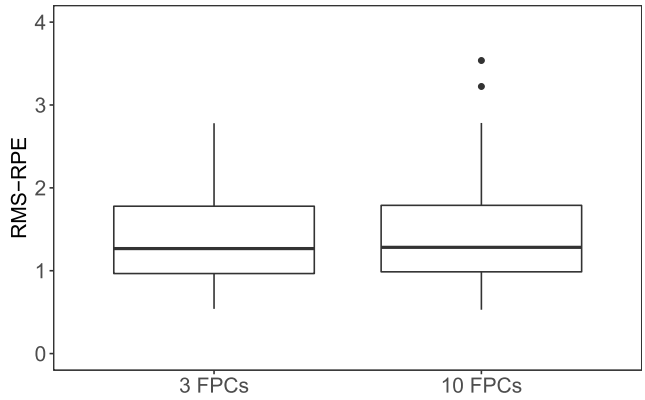

FIG. 9. *Boxplots of the root mean squared relative prediction error* (*RMS-RPE*) *of the box office in the opening week on the test datasets based on* 1000 *random training/test splits for various models* (*"*3 *FPCs:" the OLS regression with the first three* 0.8*th quantile FPC scores*; *"*10 *FPCs:" the lasso regression with the first* 10 0.8*th quantile FPC scores*.

helpful for prediction. Specifically, we include 10 FPCs by setting the 99% threshold for the total percentages of variance explained by these FPCs. We then apply the same train and test procedure, described in Section 3.3, to run an ordinary least squares (OLS) regression using only the top three FPC scores. Figure 9 shows the RMS-RPE using the top 10 FPC scores in comparison with using the top three FPC scores. We observe that dropping the high-order FPC has a negligible impact on prediction error, which greatly simplifies the interpretation. Therefore, we reduce the number of FPCs to three in our following analysis. The results for the second week's box office are provided in the Supplementary Material. On this basis we focus on the shape of the first three FPCs.

The first three FPCs provide a low-rank representation of the 0.8th quantile trajectory, that is, the quantile trajectory is decomposed into the sum of the mean function and the first three FPCs,

$$Q_i^{0.8}(t) \approx \mu^{0.8}(t) + \sum_{k=1}^{3} \xi_{ik}^{0.8} \psi_k^{0.8}(t).$$

Thus, each movie's 600 (60 hours by 10 ratings) rating counts have been reduced to three FPCs for the 0.8th quantile, a dramatic dimensional reduction. The contribution of each component $\psi_k^{0.8}(t)$ to explaining $Q_i^{0.8}(t)$ depends on the movie specific scores $\xi_{ik}^{0.8}$. The distribution of the standardized scores (the regressions were run with standardized scores) across movies is shown in Figure 10, with scores beyond two standard deviations removed. Note that the first and second scores are mostly positive and the third is mostly negative.

The medians of the three standardized scores for the week one box office prediction are 0.41, 0.24, and −0.23, respectively. Figure 11 shows the mean curve and the first three FPCs of the 0.8th quantile. The first FPC accounts for 77.91% of the total variation. We can observe that the first FPC is positive throughout the 60 hours, and the weight placed on the 30–50 hours is larger than that placed on the other time intervals. In addition, the percentage 77.91% indicates that the main source of variation among the movies comes from the weighted average of their quantile trajectories. Movies with high FPC 1 scores are the ones that have much higher than average 0.8th quantiles during 30–50 hours and with high 0.8th quantiles in other time intervals. For example, the movie *Green Book* is one of the the movies with highest FPC 1 scores because its 0.8th quantiles over hours 1–60 are all 10 s. The second FPC explains 8.33% of the total variation and consists of a negative contribution for the time interval from hour 12 to hour 43, which can be interpreted as the difference in the quantile

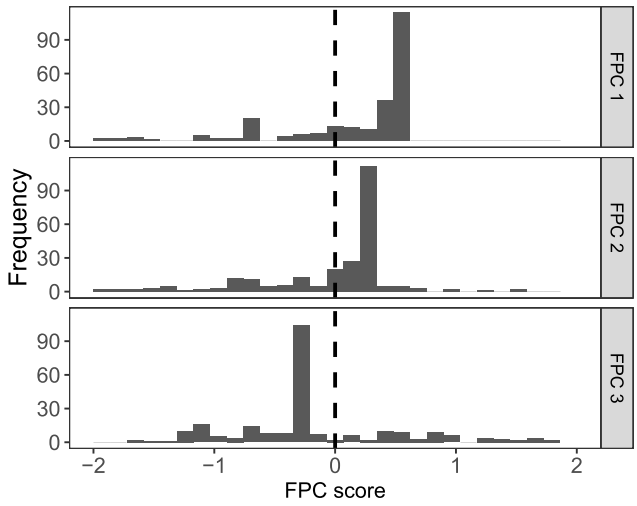

FIG. 10. *Histograms of the first three standardized FPC scores of the* 0.8*th quantile trajectory. The vertical dashed lines indicate the zero values of the FPC scores.*

trajectory between the early and late time intervals and the time interval in the middle of [12, 43]. The third FPC accounts for 6.10% of the total variation and has positive contribution for the first 25 hours and a negative contribution for the remaining hours prior to the movie releases. Movies with large difference of 0.8th quantiles in the first 25 hours and the last 35 hours tend to have high FPC 3 scores.

To see the effect of the shape of each FPC and the relative impact across the three FPCs, we plot the quantile mean plus/minus the median FPC score times the FPC for each of the three components. Specifically, in Figure 12 we plot $\mu^{0.8}(t) \pm 0.41\psi_1^{0.8}(t)$, $\mu^{0.8}(t) \pm 0.24\psi_2^{0.8}(t)$,

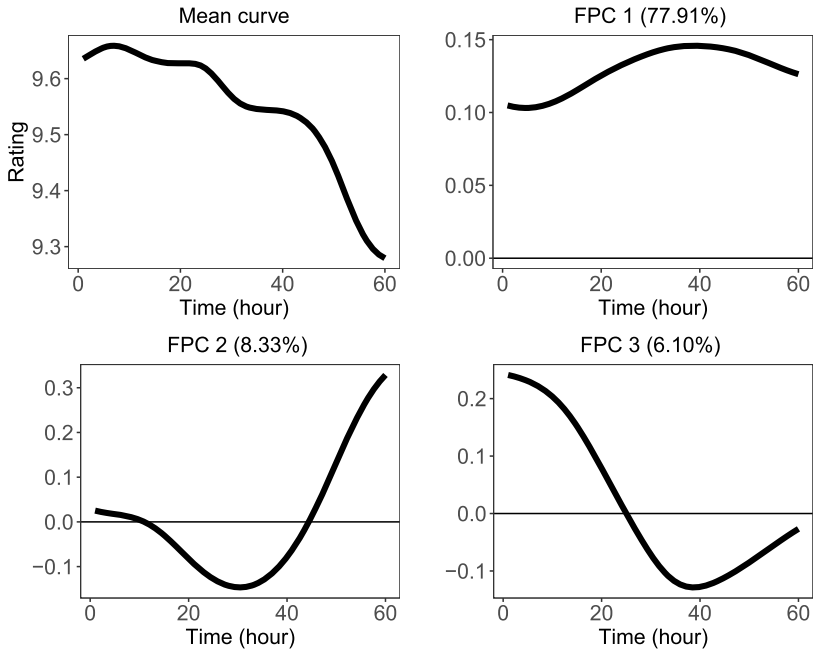

FIG. 11. *The mean curve and the first three functional principal component curves of the* 0.8*th quantile trajectory. The percentages in the parentheses indicate the amount of total variation explained by each FPC.*

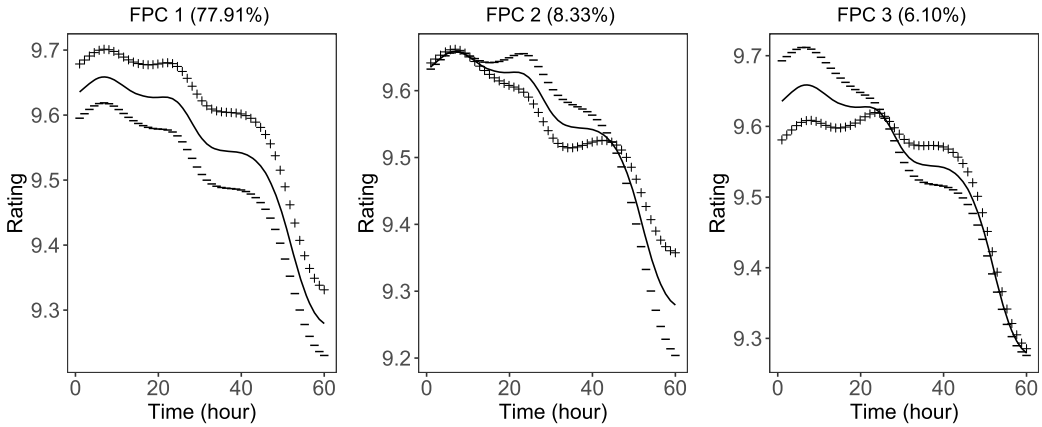

FIG. 12. *Effects of the first three principal components of the* 0.8*th quantile trajectory* (*The mean curve of the* 0.8*th quantile trajectory* (*solid black curves*), *the mean plus median score times the component* (*plus signs*), *and the mean minus the median score times the component* (*negative signs*)). *The percentages in the parentheses indicate the amount of total variation explained by each FPC.*

and $\mu^{0.8}(t) \mp 0.23\psi_3^{0.8}(t)$. The mean curve of the 0.8th quantile declines over the 60-hour time period, consistent with previous research on WOM ratings (e.g., Li and Hitt (2008)).

Next, we provide some intuition into the interpretation of the results for the prediction of the box office in the opening week. The OLS coefficients of the first three FPC scores are all positive with median values 0.35, 0.17, and 0.13, respectively (see Figure 13). We can be confident of positive effects, that is, as each of the first three FPC scores increases, the box office increases. Since the first component never changes sign over the time span (it will always be above or below the mean; see Figure 12), the greater the 0.8th quantile throughout the time span, the higher the box office revenues. This is unsurprising and consistent with the valence effect in past research (e.g., Liu (2006)). Figure 12 shows that the adding of the first FPC times its median score to the quantile mean is associated with about a 0.05 increase in the quantile rating above the mean over the entire time span. Some sense of the magnitude of the effect of such a rating increase can be gained by noting that a first FPC score of 0.41 (the median) increases the week one box office by a factor of about $e^{0.35\times 0.41} = 1.15$ (15% increase from a zero FPC 1 score).

Figure 12 shows that adding the second FPC times its median score to the quantile mean is associated with a more rapid than average drop in rating between hours 12 and 43 but a less

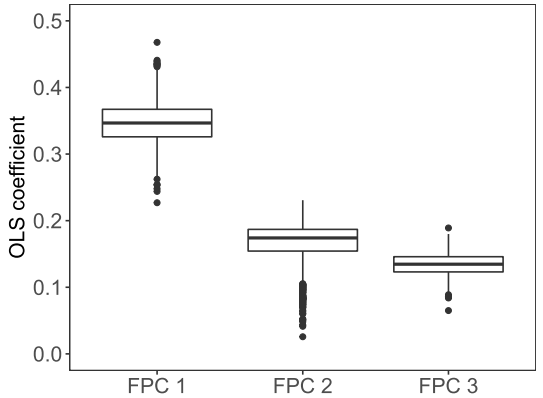

FIG. 13. *Boxplots of the coefficients of the first three FPC scores of the* 0.8*th quantile trajectory, from the cross validation runs of the OLS regression model with the first three FPC scores.*

rapid drop than the mean after hour 43. The OLS coefficients are all positive with a median of 0.17, therefore, a second FPC score of 0.24 (the median) increases the opening week's box office by a factor of about $e^{0.17\times0.24} = 1.04$ (4% increase from a zero FPC 2 score).

Adding the third FPC times its median score is associated with a lower than average 0.8th quantile trajectory before hour 25 but a higher than average 0.8th quantile trajectory afterward. The interpretation of this plot is complicated by the fact that the median score is negative (-0.23) so that the positive curve in Figure 12 is actually showing the result of subtracting an amount from the median. All the coefficients associated with the third FPC are positive, with median 0.13. As the FPC 3 score increases, the box office increases. As shown in Figure 11 and essentially the negative curve in Figure 12, a higher positive FPC 3 score is associated with a rapid decline of the 0.8th quantile from a higher than the average starting point at hour 1, all the way to hour 40 and a small recovery back to the mean afterward. In other words, if a movie has a high positive FPC 3 score, say 0.23, it will have a higher box office than a movie with zero FPC 3 by a factor of about $e^{0.13\times0.23} = 1.03$ (3% increase).

The week 2 models have coefficients that are slightly larger than the week 1 coefficients, with the same relative magnitudes across the three components. The effects of the ratings on week 2 box office are thus similar to week 1, but slightly larger.

5. Discussion. Our data driven approach to studying the prelaunch eWOM and its relation to the subsequent box office generates useful implications for the movie industry practitioners and academic, who research about adoption of movies and other entertainment products with prelaunch eWOM publicly available.

5.1. *Managerial implications*. The ability to predict the box office performance of a movie before it even hits the movie theaters is important to both movie studios and movie theaters. The valence (the arithmetic average rating) and volume (number of users, who rate the movie online) of eWOM are well-established metrics. The current study, on the other hand, demonstrates the value of comparing a movie's prelaunch eWOM to a benchmark trajectory, the 0.8th quantile of IMBb user rating in our context. Specifically, movie industry practitioners may have a hard time calculating a single metric like average rating as the eWOM evolves constantly, but it would be relatively easy for them to compare a movie's 0.8th trajectory to the benchmark/mean trajectory. Although a movie's rating would continuously decline during the 60 hours prior to its release, if its trajectory is above the benchmark (i.e., similar to the plus curve of FPC 1 in Figure 12), movie industry practitioners, like marketers and movie theaters, can then adjust their marketing and operational activities accordingly. For example, if a movie is about 0.05 higher than the mean of the 0.8th quantile, there may be 15% more box office than a movie with zero FPC 1 score, as discussed earlier. This is particularly important to movie theaters, which may not have the same resource or technology as movie studios that have their proprietary forecasting system.

As shown in Eliashberg et al. (2009), it is particularly challenging to predict the opening week and second week box office performance of a new movie. Although using some relatively crude forecasting methods, Eliashberg et al. (2009)'s forecasting and movie scheduling system was shown to be able to increase the annual revenue for one theater by U.S. $220,000. It is quite possible that a forecasting system based on tracking the 0.8th quantile trajectory would provide similar, if not more additional revenues.

In addition to scheduling high demand movies at the optimal starting time, a better forecasting system, based on tracking quantile trajectory, would be of great value to movie theaters pricing and staffing decisions. Specifically, given how crucial concession sales are to the profitability of a movie theater (Gil and Hartmann (2009)), being able to more accurately predict a high demand movie would allow a movie theater to have enough staff for the concession stores so as to ensure all orders can be taken and delivered before the starting time of

individual movie screenings. In addition, even if movie theaters are not fully adopting differential pricing, the ability to more accurately forecast demand of specific tent pole movies and thereby show them in "premium" pricing screening formats, like 3D, is expected to improve the profitability of any movie theater (Ho et al. (2018)).

5.2. *Theoretical implications*. Our theory-agnostic approach abstracts away from the individual level behaviors by the internet users posting those prelaunch movie reviews and/or the potential moviegoers considering watching specific movies. That being said, the data patterns revealed by our analysis have interesting implications for such micro-level social influence work as social impact theory (Latane and Wolf (1981)) and informational cascades (Bikhchandani, Hirshleifer and Welch (1992)). Specifically, while the positive relation of FPC 1 score with the opening week box office is consistent with the notion that earlier consumers' behaviors, like purchase, and eWOM would positively influence the subsequent consumers' decisions, the positive relation of FPC 3 score to the opening box office suggests a faster than average drop of the 0.8th quantile would actually be associated with a higher opening box office. Although FPC 3 accounts for a small percentage of variation in the 60-hour data stream, this empirical pattern cannot be explained by existing social influence work. It may be due to the possibility that the prelaunch period eWOM has not reached the equilibrium state required by theoretical work like Bikhchandani, Hirshleifer and Welch (1992) or something unique to the behaviors during the prelaunch period, suggesting interesting research direction for further theoretical work.

In addition, our finding that the 0.8th quantile trajectory is more informative or predictive than those of the arithmetic means or other summary statistics indicates the importance of the rapid slope change of the J-shaped distribution on the high end of the eWOM distribution. This empirical pattern may be related to the notion of proximity in the social impact theory (Latane and Wolf (1981))—the potential moviegoers in the opening week tend to be the ones, who want to like the movie and, thereby, be more similar or "proximate" to the top 20 percentile of the rating distribution in terms of taste and anticipation for the movie. This would be an interesting research direction to complement the more established explanations of the J-shaped eWOM distribution in the information system literature, like the self-selection biases (Hu, Pavlou and Zhang (2009), Li and Hitt (2008), Hu, Pavlou and Zhang (2017)).

6. Conclusion. This research shows that the evolution pattern of a top-end quantiles of the cumulative users' ratings' distribution (e.g., 0.75th and 0.80th quantiles) would be a good alternative to the commonly used arithmetic average (the valence) and variance as an eWOM metric. Situating the study in the theatrical market, a market where new product launches are frequent and competing products aim at maximizing product awareness and interest at product launches, we use the movie ratings from imdb.com to demonstrate how the 0.8th quantile and its time pattern before a movie's release would be more useful to managers than the arithmetic average in predicting the movie's box office in the first and second week. We use the FPCA to capture the dominant sources of variations of the time patterns of the functional quantiles of the prelaunch users' ratings. The FPCs are estimated by the PACE method, which also impute the missing data from the predicted quantile trajectories.

The first three functional principal components of the 0.8th quantile, plus the log of the counts to capture scale, provide better prediction of box office revenues than volume and valence type of summaries of the ratings. The 0.8th quantile is also the location where the intermediate peak of the WJ pattern of the average rating distribution appears. The first and second FPCs indicate that there are positive effects of eWOM counts over time, beyond what the log of the counts—the static volume—captures. The third FPC shows that the more positive the score, the higher the box office, and the more positive the score, the more rapidly declining the quantile.

There are various possible directions of future work. Our research not only is useful in the product prelaunch prediction from the eWOM data but also is applicable to other applied settings. For example, our work can be applied to early diagnosis of disease using the longitudinal disease trajectory. In fact, in any setting where observed time series of functions need to be related to a subsequent single variable metric, our FPCA and lasso method can be a good alternative to the black-box machine learning algorithms, like deep learning. Another possible direction of future work is the choice of the number of components K. In our research we select K such that the first K FPCs explain over 99% of the total variance. But there are also several other approaches we can use for the selection. For example, we can use the cross-validation method proposed by Rice and Silverman (1991). We can also use the approaches based on information criteria (e.g., Li, Wang and Carroll (2013), Yao, Müller and Wang (2005)). Last but not least, as discussed in Section 5.2, our work is expected to complement the micro-level social influence work, like social impact theory (Latane and Wolf (1981)) and informational cascades (Bikhchandani, Hirshleifer and Welch (1992)), and possibly extend to key business outcomes other than sales, for example, perceived helpfulness of product reviews, a mainstay variable in the management information system literature (Mudambi and Schuff (2010)).

Acknowledgments. The authors would like to thank the Editor, the Associate Editor and two anonymous referees for many insightful comments. These comments are very helpful for us to improve our work.

T. Guan is an Assistant Professor at the Department of Mathematics and Statistics at York University. J. Ho is an Associate Professor at Beedie School of Business, Simon Fraser University. R. Krider is a Professor Emeritus at Beedie School of Business, Simon Fraser University. J. Cao is a Professor at the Department of Statistics and Actuarial Science at Simon Fraser University. A. Fogg is an Engineering Manager in the Advanced Development group at Roku, Inc.

J. Cao is the corresponding author for this article.

Funding. This research is supported by the Discovery grants (RGPIN-2023-04057 to J. Cao and RGPIN-2022-05140 to T. Guan) from the Natural Sciences and Engineering Research Council of Canada (NSERC).

SUPPLEMENTARY MATERIAL

Supplementary document (DOI: 10.1214/23-AOAS1854SUPP; .pdf). The Supplementary Material contains the mean curves and the top three FPCs estimated from the observed quantile trajectories at quantile levels 0.1, 0.25, 0.5, and 0.9 and additional results for predicting the week two box office revenues.

REFERENCES

BABIĆ ROSARIO, A., DE VALCK, K. and SOTGIU, F. (2020). Conceptualizing the electronic word-of-mouth process: What we know and need to know about eWOM creation, exposure, and evaluation. *J. Acad. Mark. Sci.* **48** 422–448.

BABIĆ ROSARIO, A., SOTGIU, F., DE VALCK, K. and BIJMOLT, T. H. A. (2016). The effect of electronic word of mouth on sales: A meta-analytic review of platform, product, and metric factors. *J. Mark. Res.* **53** 297–318.

BENKO, M., HÄRDLE, W. and KNEIP, A. (2009). Common functional principal components. *Ann. Statist.* **37** 1–34. MR2488343 https://doi.org/10.1214/07-AOS516

BIKHCHANDANI, S., HIRSHLEIFER, D. and WELCH, I. (1992). A theory of fads, fashion, custom, and cultural change as informational cascades. *J. Polit. Econ.* **100** 992–1026.

CARDOT, H. (2000). Nonparametric estimation of smoothed principal components analysis of sampled noisy functions. *J. Nonparametr. Stat.* **12** 503–538. MR1785396 https://doi.org/10.1080/10485250008832820

CHINTAGUNTA, P. K., GOPINATH, S. and 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** 944–957.

CLEMONS, E. K., GAO, G. and HITT, L. M. (2006). When online reviews meet hyperdifferentiation: A study of the craft beer industry. *J. Manage Inf. Syst.* **23** 149–171.

DAUXOIS, J., POUSSE, A. and ROMAIN, Y. (1982). Asymptotic theory for the principal component analysis of a vector random function: Some applications to statistical inference. *J. Multivariate Anal.* **12** 136–154. MR0650934 https://doi.org/10.1016/0047-259X(82)90088-4

DHAR, V. and CHANG, E. A. (2009). Does chatter matter? The impact of user-generated content on music sales. *J. Interact. Mark.* **23** 300–307.

ELIASHBERG, J., HEGIE, Q., HO, J., HUISMAN, D., MILLER, S. J., SWAMI, S., WIERENGA, C. B. and WIERENGA, B. (2009). Demand-driven scheduling of movies in a multiplex. *Int. J. Res. Mark.* **26** 75–88.

FOUTZ, N. Z. and JANK, W. (2010). Research note—Prerelease demand forecasting for motion pictures using functional shape analysis of virtual stock markets. *Mark. Sci.* **29** 568–579.

GELPER, S., PERES, R. and ELIASHBERG, J. (2018). Talk bursts: The role of spikes in prerelease word-of-mouth dynamics. *J. Mark. Res.* **55** 801–817.

GHOSAL, R., VARMA, V. R., VOLFSON, D., HILLEL, I., URBANEK, J., HAUSDORFF, J. M., WATTS, A. and ZIPUNNIKOV, V. (2023). Distributional data analysis via quantile functions and its application to modeling digital biomarkers of gait in Alzheimer's Disease. *Biostatistics* **24** 539–561. MR4615240 https://doi.org/10.1093/biostatistics/kxab041

GIL, R. and HARTMANN, W. R. (2009). Empirical analysis of metering price discrimination: Evidence from concession sales at movie theaters. *Mark. Sci.* **28** 1046–1062.

GILCHRIST, W. (2000). *Statistical Modelling with Quantile Functions*. CRC Press/CRC, Boca Raton, FL.

GUAN, T., HO, J., KRIDER, R., CAO, J. and FOGG, A. (2024). Supplement to "How are PreLaunch online movie reviews related to box office revenues?." https://doi.org/10.1214/23-AOAS1854SUPP

HALL, P. and HOSSEINI-NASAB, M. (2006). On properties of functional principal components analysis. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **68** 109–126. MR2212577 https://doi.org/10.1111/j.1467-9868.2005.00535.x

HO, J. Y. C., LIANG, Y., WEINBERG, C. B. and YAN, J. (2018). An empirical study of uniform and differential pricing in the movie theatrical market. *J. Mark. Res.* **55** 414–431.

HOUSTON, M. B., KUPFER, A. K., HENNIG-THURAU, T. and SPANN, M. (2018). Pre-release consumer buzz. *J. Acad. Mark. Sci.* **46** 338–360.

HU, N., PAVLOU, P. A. and ZHANG, J. (2009). Overcoming the J-shaped distribution of product reviews. *Commun. ACM* **52** 144–147.

HU, N., PAVLOU, P. A. and ZHANG, J. (2017). On self-selection biases in online product reviews. *MIS Q.* **41** 449–471.

JAMES, G. M., HASTIE, T. J. and SUGAR, C. A. (2000). Principal component models for sparse functional data. *Biometrika* **87** 587–602. MR1789811 https://doi.org/10.1093/biomet/87.3.587

VERMA, S. and YADAV, N. (2021). Past, present, and future of electronic word of mouth (eWOM). *J. Interact. Mark.* **53** 111–128.

LATANE, B. and WOLF, S. (1981). The social impact of majorities and minorities. *Psychol. Rev.* **88** 438–453.

LI, X. and HITT, L. M. (2008). Self-selection and information role of online product reviews. *Inf. Syst. Res.* **19** 456–474.

LI, Y., WANG, N. and CARROLL, R. J. (2013). Selecting the number of principal components in functional data. *J. Amer. Statist. Assoc.* **108** 1284–1294. MR3174708 https://doi.org/10.1080/01621459.2013.788980

LIN, Z., WANG, L. and CAO, J. (2016). Interpretable functional principal component analysis. *Biometrics* **72** 846–854. MR3545677 https://doi.org/10.1111/biom.12457

LIU, Y. (2006). Word of mouth for movies: Its dynamics and impact on box office revenue. *J. Mark.* **70** 74–89.

MUDAMBI, S. M. and SCHUFF, D. (2010). What makes a helpful online review? A study of customer reviews on Amazon.com. *MIS Q.* **34** 185–200.

NIE, Y. and CAO, J. (2020). Sparse functional principal component analysis in a new regression framework. *Comput. Statist. Data Anal.* **152** 107016, 15. MR4114152 https://doi.org/10.1016/j.csda.2020.107016

NIE, Y., WANG, L., LIU, B. and CAO, J. (2018). Supervised functional principal component analysis. *Stat. Comput.* **28** 713–723. MR3761351 https://doi.org/10.1007/s11222-017-9758-2

NIE, Y., YANG, Y., WANG, L. and CAO, J. (2022). Recovering the underlying trajectory from sparse and irregular longitudinal data. *Canad. J. Statist.* **50** 122–141. MR4389173 https://doi.org/10.1002/cjs.11677

PAUWELS, K., AKSEHIRLI, Z. and LACKMAN, A. (2016). Like the ad or the brand? Marketing stimulates different electronic word-of-mouth content to drive online and offline performance. *Int. J. Res. Mark.* **33** 639–655.

PURNAWIRAWAN, N., EISEND, M., DE PELSMACKER, P. and DENS, N. (2015). A meta-analytic investigation of the role of valence in online reviews. *J. Interact. Mark.* **31** 17–27.

QAHRI-SAREMI, H. and MONTAZEMI, A. R. (2019). Factors affecting the adoption of an electronic word of mouth message: A meta-analysis. *J. Manage Inf. Syst.* **36** 969–1001.

RAMSAY, J. O., HOOKER, G. and GRAVES, S. (2009). *Functional Data Analysis with R and Matlab*. Springer, New York.

RAMSAY, J. O. and SILVERMAN, B. W. (2005). *Functional Data Analysis*, 2nd ed. *Springer Series in Statistics*. Springer, New York. MR2168993

RICE, J. A. and SILVERMAN, B. W. (1991). Estimating the mean and covariance structure nonparametrically when the data are curves. *J. Roy. Statist. Soc. Ser. B* **53** 233–243. MR1094283

SANG, P., BEGEN, M. A. and CAO, J. (2021). Appointment scheduling with a quantile objective. *Comput. Oper. Res.* **132** Paper No. 105295, 20. MR4255393 https://doi.org/10.1016/j.cor.2021.105295

SANG, P., WANG, L. and CAO, J. (2017). Parametric functional principal component analysis. *Biometrics* **73** 802–810. MR3713114 https://doi.org/10.1111/biom.12641

SHI, H., DONG, J., WANG, L. and CAO, J. (2021). Functional principal component analysis for longitudinal data with informative dropout. *Stat. Med.* **40** 712–724. MR4198440 https://doi.org/10.1002/sim.8798

SHI, H., YANG, Y., WANG, L., MA, D., BEG, M. F., PEI, J. and CAO, J. (2022). Two-dimensional functional principal component analysis for image feature extraction. *J. Comput. Graph. Statist.* **31** 1127–1140. MR4513375 https://doi.org/10.1080/10618600.2022.2035738

SIMON, N., FRIEDMAN, J., HASTIE, T. and TIBSHIRANI, R. (2013). A sparse-group lasso. *J. Comput. Graph. Statist.* **22** 231–245. MR3173712 https://doi.org/10.1080/10618600.2012.681250

TIBSHIRANI, R. (1996). Regression shrinkage and selection via the lasso. *J. Roy. Statist. Soc. Ser. B* **58** 267–288. MR1379242

WANG, J.-L., CHIOU, J.-M. and MÜLLER, H.-G. (2016). Review of functional data analysis. *Annu. Rev. Stat. Appl.* **3** 257–295.

XIONG, G. and BHARADWAJ, S. (2014). Prerelease buzz evolution patterns and new product performance. *Mark. Sci.* **33** 401–421.

YANG, H., BALADANDAYUTHAPANI, V., RAO, A. U. K. and MORRIS, J. S. (2020). Quantile function on scalar regression analysis for distributional data. *J. Amer. Statist. Assoc.* **115** 90–106. MR4078447 https://doi.org/10.1080/01621459.2019.1609969

YAO, F., MÜLLER, H.-G. and WANG, J.-L. (2005). Functional data analysis for sparse longitudinal data. *J. Amer. Statist. Assoc.* **100** 577–590. MR2160561 https://doi.org/10.1198/016214504000001745

YOU, Y., VADAKKEPATT, G. G. and JOSHI, A. M. (2015). A meta-analysis of electronic word-of-mouth elasticity. *J. Mark.* **79** 19–39.

ZHANG, J.-T. and CHEN, J. (2007). Statistical inferences for functional data. *Ann. Statist.* **35** 1052–1079. MR2341698 https://doi.org/10.1214/009053606000001505