FUNCTIONAL RESPONSE ADDITIVE MODEL ESTIMATION WITH ONLINE VIRTUAL STOCK MARKETS

BY YINGYING FAN[*,1], NATASHA FOUTZ[†],
GARETH M. JAMES[*] AND WOLFGANG JANK[‡]

*University of Southern California[*], University of Virginia[†]
and University of South Florida[‡]*

While functional regression models have received increasing attention recently, most existing approaches assume both a linear relationship and a scalar response variable. We suggest a new method, "Functional Response Additive Model Estimation" (FRAME), which extends the usual linear regression model to situations involving both functional predictors, $X_j(t)$, scalar predictors, Z_k, and functional responses, $Y(s)$. Our approach uses a penalized least squares optimization criterion to automatically perform variable selection in situations involving multiple functional and scalar predictors. In addition, our method uses an efficient coordinate descent algorithm to fit general nonlinear additive relationships between the predictors and response.

We develop our model for novel forecasting challenges in the entertainment industry. In particular, we set out to model the decay rate of demand for Hollywood movies using the predictive power of online virtual stock markets (VSMs). VSMs are online communities that, in a market-like fashion, gather the crowds' prediction about demand for a particular product. Our fully functional model captures the pattern of pre-release VSM trading prices and provides superior predictive accuracy of a movie's post-release demand in comparison to traditional methods. In addition, we propose graphical tools which give a glimpse into the causal relationship between market behavior and box office revenue patterns, and hence provide valuable insight to movie decision makers.

1. Introduction. Functional data analysis (FDA) has become an important topic of study in recent years, in part because of its ability to capture patterns and shapes in a parsimonious and automated fashion [Ramsay and Silverman (2005)]. Recent methodological advances in FDA include functional principal components analysis [James, Hastie and Sugar (2000), Rice and Wu (2001)], regression with functional responses [Zeger and Diggle (1994)] or functional predictors [Ferraty and Vieu (2002), James and Silverman (2005)], functional linear discriminant analysis [Ferraty and Vieu (2003), James and Hastie (2001)], functional clustering [Bar-Joseph et al. (2003), James and Sugar (2003)] or functional forecasting [Zhang, Jank and Shmueli (2010)].

In this paper we are interested in the regression situation involving p different functional predictors, $X_1(t), \ldots, X_p(t)$. Most existing functional regression models assume a *linear* relationship between the response and predictors [Yao, Müller and Wang (2005a)], which is often an overly restrictive assumption. Recently, several papers have suggested approaches for performing nonlinear functional regressions [Chen, Hall and Müller (2011), Fan, James and Radchenko (2014), James and Silverman (2005)] of the form

$$(1) \qquad Y_i = \sum_{j=1}^{p} f_j(X_{ij}) + \varepsilon_i, \qquad i = 1, \ldots, n,$$

where the f_j's are general nonlinear functions of $X_{ij}(t)$ and Y_i is a centered response. Generally speaking, these approaches operationalize estimation of equation (1) by using functional index models. While all of these approaches provide a very flexible extension of the linear functional model, they are designed for scalar responses only. In this paper, we generalize this framework to functional responses. That is, we consider both functional predictors $X_{ij}(t)$ and functional responses $Y_i(s)$ and allow them to be related in a nonlinear way.

We refer to our proposed nonlinear functional regression method as "Functional Response Additive Model Estimation" (FRAME), which models both multiple functional predictors as well as functional responses. Beyond the extension to functional responses, FRAME makes two additional important contributions to the existing literature. First, it uses a penalized least squares approach to efficiently fit high-dimensional functional models while simultaneously performing variable selection to identify the relevant predictors, an area that has received very little attention in the functional domain. FRAME is computationally tractable because we use a highly efficient coordinate descent algorithm to optimize our criterion. Second, FRAME extends beyond the standard linear regression setting to fit general nonlinear additive models. Since the predictors, $X_{ij}(t)$, are infinite dimensional, any functional regression model must perform some kind of dimension reduction. FRAME achieves this goal by modeling the response as a nonlinear function of a one-dimensional linear projection of $X_{ij}(t)$, a functional version of the *single index model* approach. Our method uses a supervised fit to automatically project the functional predictors into the best one-dimensional space. We believe this is an important distinction because projecting into the unsupervised PCA space is currently the dominant approach in functional regressions, even though it is well known that this space need not be optimal for predicting the response. Our nonlinear approach allows us to model much more subtle relationships and we show that, on our data, FRAME produces clear improvements in terms of prediction accuracy.

We develop our model for novel forecasting challenges in the motion picture industry. Providing accurate forecasts for the success of new products is crucial for the 500 billion dollar entertainment industries (such as motion picture, music, TV, gaming and publishing). These industries are confronted with enormous

investments, short product life-cycles, and highly uncertain and rapidly decaying demand. For instance, decision makers in the movie industry are keenly interested in accurately forecasting a product's *demand pattern* [Bass et al. (2001), Sawhney and Eliashberg (1996)] in order to allocate, for example, weekly advertising budgets according to the predicted *rate of demand decay*, that is, according to whether a film is expected to open big and then decay fast, or whether it opens only moderately but decays very slowly.

However, forecasting demand patterns is challenging since it is highly heterogeneous across different products. Take, for instance, the sample of movie demand patterns in Figure 1. Here we have plotted the log weekly box office revenues for the first ten weeks from the release date for a number of different movies. While revenues for some movies (e.g., *13 GOING ON 30* and *50 FIRST DATES*) decay exponentially over time, revenues for others (e.g., *BEING JULIA*) increase first before decreasing later. Even for movies with similar demand patterns (e.g., those on the second row of Figure 1), the speed of decay varies greatly.

In this article we develop FRAME to forecast the demand patterns of box office revenues using a number of functional predictors which capture various sources of information about movies, such as consumers' word of mouth, via a novel data source, online virtual stock markets (VSMs). In a VSM, participants trade virtual
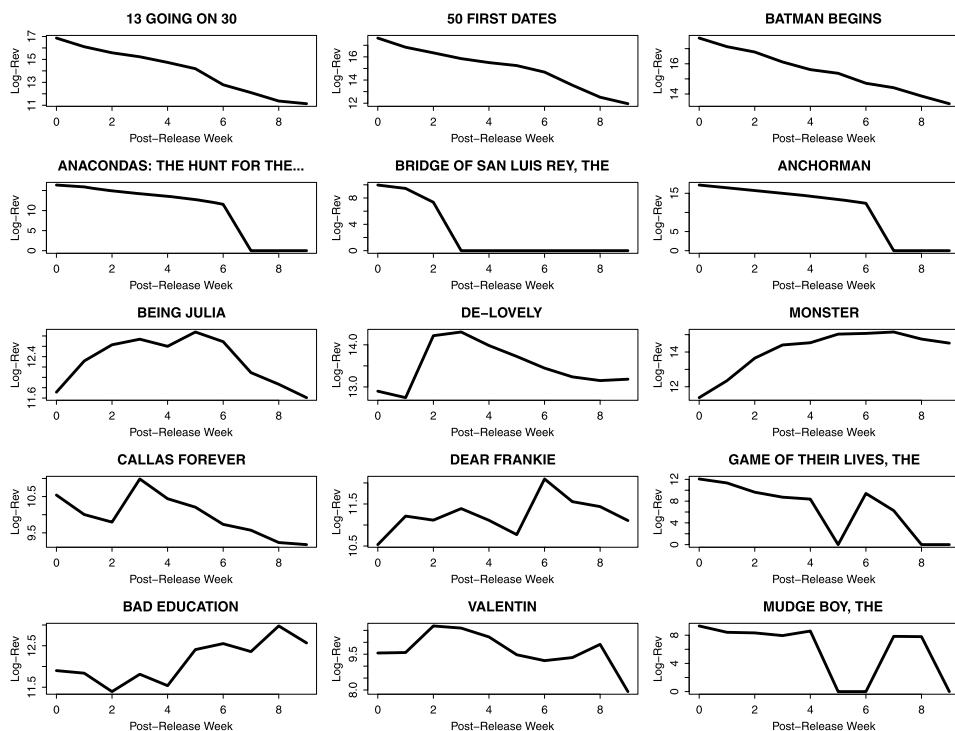

FIG. 1. *Movie demand decay rates for a sample of movies.*

stocks according to their predictions of the outcome of the event represented by the stock (e.g., the demand for an upcoming movie). As a result, VSM trading prices can provide early and reliable demand forecasts [Foutz and Jank (2010), Spann and Skiera (2003)]. VSMs are especially intriguing from a statistical point of view since the *shape* of the trading prices may reveal additional information, such as the speed of information diffusion which, in turn, can proxy for consumer sentiment and word of mouth about a new product [Foutz and Jank (2010)]. For instance, a last-moment price spurt may reveal a strengthening hype for a product and may thus be essential in forecasting its demand.

This paper is organized as follows. In the next section we provide further background on virtual stock markets in general and our data in particular. In Section 3 we present the FRAME model and its optimization criterion. We also discuss an efficient coordinate descent algorithm for fitting FRAME. In Section 4 an extensive simulation study is used to demonstrate the superior performance of FRAME, in comparison to a number of competitors. Section 5 discusses the results from applying FRAME to our movie data. In that section, we also address the challenge of interpreting the results from a model involving both functional predictors and functional responses using "dependence plots." Dependence plots graphically illustrate, for typical shapes of the predictors, the corresponding predicted response pattern. These dependence plots allow for a glimpse into the relationship between response and predictors and provide actionable insight for decision makers. We conclude with further remarks in Section 6.

2. Data. We have two different sources of data. Our input data (i.e., functional predictors) come from the weekly trading histories of an online virtual stock market for movies before their releases; our output data (i.e., functional responses) pertain to the post-release weekly demand of those movies. We have data on a total of 262 movies. The data sources are described below.

2.1. *Online virtual stock markets.* Online virtual stock markets (VSMs) operate in ways very similar to real life stock markets except that they are not necessarily based on real currency (i.e., participants often use virtual currency to make trades), and that each stock corresponds to discrete outcomes or continuous parameters of an event (rather than a company's value). For instance, a value of $54 for the movie stock *50 FIRST DATES* is interpreted as the traders' collective belief that the movie will accrue $54 million in the box office during its first four weeks of theatrical exhibition. If the movie eventually earns $64 million, then traders holding the stock will liquidate (or "cash-in") at $64 per share.

The source of our data is the *Hollywood Stock Exchange* (HSX), one of the best known online VSMs. HSX was established in 1996 and aims at predicting a movie's revenues over its first four weeks of theatrical exhibition. HSX has had well over 2 million active participants worldwide and each trader is initially endowed with $2 million virtual currency and can increase his or her net worth by

strategically selecting and trading movie stocks (such as by buying low and selling high). Traders are further motivated by opportunities to appear on the daily *Leader Board* that features the most successful traders.

For each movie we collect four functional predictors between 52 and 10 weeks prior to the movie's release date. They are the following: the intra-day average price (i.e., the average of the highest and lowest trading prices of the day, as recorded by HSX) on each Friday (which is the most active trading day of the week), each Friday's number of accounts shorting the stock, number of shares sold, and number of shares held short. Figure 2 shows the curves for one of these predictors, *average price*, for the movie demand patterns from Figure 1. Note that since our goal is to accomplish *early forecasts*, we only consider information between 52 and 10 weeks prior to a movie's release (i.e., up to week −10 in Figure 2). We form predictions of movie decay ten weeks prior to release because this provides a realistic time frame for managers to make informed decisions about marketing mix allocations and other strategic decisions. Of course our analysis could also be performed using data closer to the release date.

Our FRAME method captures differences in shapes of VSM trading histories (such as price or volume), for example, trending up or down, concavity vs. convexity or last-moment spurts. The empirical results in Section 5 show that these
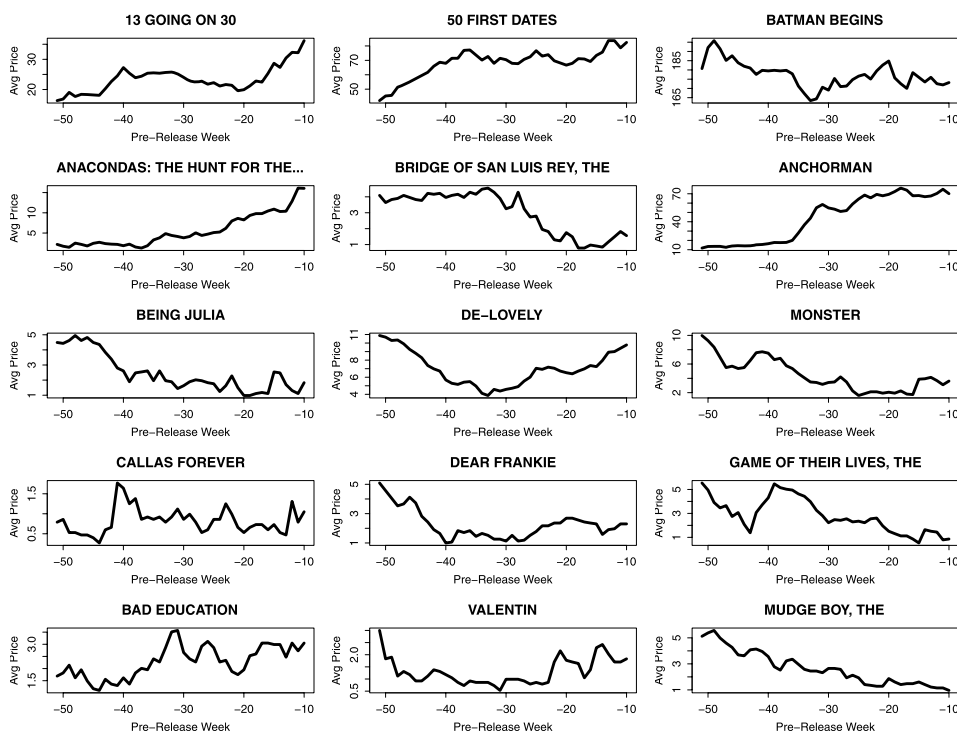

FIG. 2. *HSX trading histories for the sample of movies from Figure* 1.

shapes are predictive of the demand pattern over a product's life cycle. For example, a rapid increase in early VSM trading prices may suggest a rapid diffusion of awareness among potential adopters and a strong interest in a product. Thus, it can suggest a strong *initial* demand immediately after a new product's introduction to the market place, for example, a strong opening weekend box office for a movie. Similarly, a new product whose trading prices increase very sharply over the pre-release period may be experiencing strong positive word of mouth, which may lead to both a strong opening weekend and a reduced decay rate in demand for the movie, that is, increased longevity.

2.2. *Weekly movie demand patterns.* Our goal is to predict a movie's demand (i.e., its box office revenue). Specifically, we want to predict a movie's demand not only for a given week (e.g., at week 1 or week 5), but over its entire theatrical life cycle of about 10 weeks (i.e., from its opening week 1 to week 10). Figure 3 shows weekly demand for all 262 movies in our data (on the log-scale). The left panel plots the distribution across all movies and weeks; we can see that (log) demand is rather symmetric and appears to be bi-modal. We can also see that a portion of the data equals zero; these correspond to movies with zero demand, particularly in later weeks (the constant 1 was added to all revenues before taking the log transformation). During weeks 1 and 2, every movie has positive revenue. In week 3, only 4 movies have zero revenue; this number increases to 67 movies
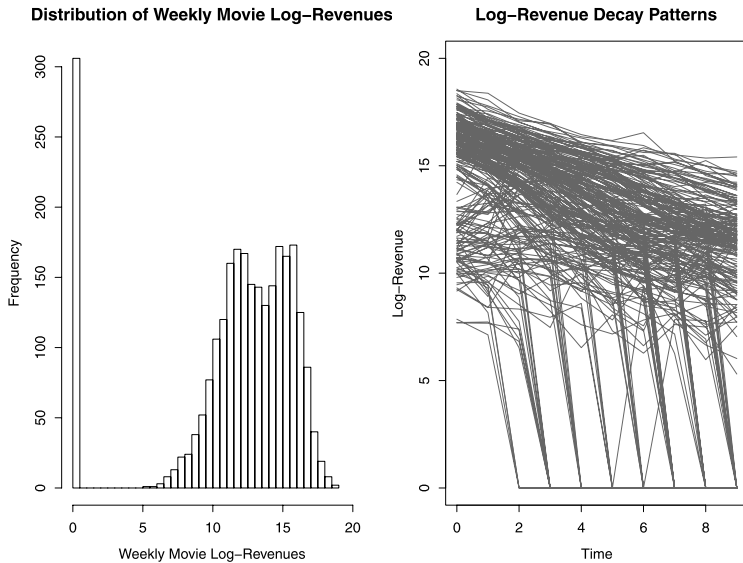

FIG. 3. *Distribution of movies' weekly demand and demand decay patterns. The right panel shows* 10-*week decay patterns* (*from the release week until* 9 *weeks after release*) *for the* 262 *movies in our sample*; *the left panel shows the distribution of the corresponding* 10 × 262 = 2620 *weekly log-revenues.*

by week 10. The right panel shows, for each individual movie, the rate at which demand decays over the 10-week period. We can see that whereas some movies decay gradually, a number have sudden drops, while others initially increase after the release week. Our goal is to characterize different demand decay *shapes* and to use the information from the VSM to forecast these shapes.

3. Functional Response Additive Model Estimation. In this section we develop our *Functional Response Additive Model Estimation* (FRAME) approach for relating a functional response, $Y_i(s)$, to a set of p functional predictors, $X_{i1}(t), \ldots, X_{ip}(t)$, and q univariate predictors, $Z_{i1}, \ldots, Z_{iq}$, where $i = 1, \ldots, n$.

3.1. *FRAME model.* The classical functional linear regression model is given by

$$(2) \qquad Y_i(s) = \int \beta(s, t) X_i(t)\, dt + \varepsilon_i(s),$$

where $\beta(s, t)$ is a smooth two-dimensional coefficient function to be estimated as part of the fitting process. Note we assume throughout that the predictors and responses have been centered so that the intercept term can be ignored. We also assume that the response curves $Y_i(s)$ are independent, given $X_i(t)$; for work on correlated response curves, see, for example, Di et al. (2009) or Crainiceanu, Caffo and Morris (2011).

The model given by (2) has been applied in many settings. However, it has two obvious deficiencies for use with our data. First, it assumes a single functional predictor, whereas our data contains $p = 4$ functional predictors and a number of univariate predictors. Second, the integral in (2) is a natural analogue of the summation term in the linear regression model. Hence, (2) assumes a linear relationship between the predictor and the response. In many situations this assumption is too restrictive, so we wish to allow for a nonlinear relationship.

In this paper we model the relationship between the response function and the predictors using the following nonlinear additive model:

$$(3) \qquad Y_i(s) = \sum_{j=1}^{p} f_j(s, X_{ij}) + \sum_{k=1}^{q} \phi_k(s, Z_{ij}) + \varepsilon_i(s),$$

where $f_j(s, x)$ and $\phi_k(s, z)$ are general nonlinear functions to be estimated. Model (3) has the advantage that it is able to incorporate all $p + q$ predictors using a natural additive model. It is also flexible enough to model nonlinear relationships. However, fitting (3) poses some significant difficulties. First, if p or q are large relative to n, we end up in a high-dimensional situation where many different nonlinear functions must be estimated. We address this issue by fitting (3) using a penalized least squares criterion. Our penalized approach has the effect of automatically performing variable selection on the predictors, in a similar fashion

to the lasso [Tibshirani (1996)] or group lasso [Yuan and Lin (2006)] methods. Hence, we can very effectively deal with a large number of predictors. Second, even for a low value of p, estimating a completely general $f_j(s, x)$ is infeasible because $X_{ij}(t)$ is itself an infinite-dimensional function. Instead we model $f_j(s, x)$ using a *functional single index model*:

$$f_j(s, X_{ij}) = g_j \left(\int \beta_j(s, t) X_{ij}(t) \, dt \right),$$

where $\beta_j(s, t)$ is a two-dimensional index function which projects $X_{ij}(t)$ into a single direction and $g_j(x)$ is a one-dimensional function representing the nonlinear impact of the projection on $Y_i(s)$. In this way the task of estimating $f_j(s, x)$ is reduced to the simpler problem of estimating $\beta_j(s, t)$ and $g_j(x)$. Note that our primary interest in this paper is in forming accurate predictions for the response, $Y_i(s)$. Hence, we are generally not concerned with identifiability of $g_j(x)$ and $\beta_j(s, t)$, which would be more important in an inference setting. Nevertheless, empirically we have found that $g_j(x)$ and $\beta_j(s, t)$ can often be well estimated.

Using this functional index model (3) reduces to

$$(4) \qquad Y_i(s) = \sum_{j=1}^{p} g_j \left(\int \beta_j(s, t) X_{ij}(t) \, dt \right) + \sum_{k=1}^{q} \phi_k(s, Z_{ij}) + \varepsilon_i(s).$$

We then model $\beta_j(s, t) = \mathbf{b}(s, t)^T \boldsymbol{\eta}_j$ and $X_{ij}(t) = \tilde{\mathbf{b}}(t)^T \boldsymbol{\theta}_{ij}$, where $\mathbf{b}(s, t)$ and $\tilde{\mathbf{b}}(t)$ are appropriately chosen basis functions. In implementation, to ensure that $\beta_j(s, t)$ and $g_j(x)$ are identifiable, we constrain $\|\boldsymbol{\eta}_j\| = 1$ for all j. Using this representation,

$$(5) \qquad \int \beta_j(s, t) X_{ij}(t) \, dt = \boldsymbol{\theta}_{ij}^T \left[\int \tilde{\mathbf{b}}(t) \mathbf{b}(s, t)^T \, dt \right] \boldsymbol{\eta}_j = \tilde{\boldsymbol{\theta}}_{ij}(s)^T \boldsymbol{\eta}_j,$$

where $\tilde{\boldsymbol{\theta}}_{ij}(s) = [\int \mathbf{b}(s, t) \tilde{\mathbf{b}}(t)^T \, dt] \boldsymbol{\theta}_{ij}$. Note that $\boldsymbol{\eta}_j$ must be estimated as part of the fitting process, but $\tilde{\boldsymbol{\theta}}_{ij}(s)$ can be assumed known for all s because $\mathbf{b}(s, t)$ and $\tilde{\mathbf{b}}(t)$ are given, so the integral can be directly computed. In addition, $\boldsymbol{\theta}_{ij}$ can be easily computed since $X_{ij}(t)$ is directly observed.

Using this basis representation, (4) becomes

$$(6) \qquad Y_i(s) = \sum_{j=1}^{p} g_j (\tilde{\boldsymbol{\theta}}_{ij}(s)^T \boldsymbol{\eta}_j) + \sum_{k=1}^{q} \phi_k(s, Z_{ik}) + \varepsilon_i(s).$$

In practice, the response function, $Y_i(s)$, will generally be observed at a finite set of time points, $s_{i1}, \ldots, s_{in_i}$. For example, for the box office data the revenues are observed at each of the first ten weeks. In this situation (6) can be represented as

$$Y_{il} = \sum_{j=1}^{p} g_j (\tilde{\boldsymbol{\theta}}_{ijl}^T \boldsymbol{\eta}_j) + \sum_{k=1}^{q} \phi_k(s_l, Z_{ik}) + \varepsilon_{il},$$

$$(7)$$

$$i = 1, \ldots, n, l = 1, \ldots, n_i,$$

where $Y_{il} = Y_i(s_{il})$, $\tilde{\boldsymbol{\theta}}_{ijl} = \tilde{\boldsymbol{\theta}}_{ij}(s_l)$ and ε_{il} are assumed to be independent for all i and l [conditional on $X_{ij}(t)$ and Z_{ik}].

3.2. *FRAME optimization criterion.* Fitting FRAME requires estimating the unobserved parameters, $g_j(x)$, $\boldsymbol{\eta}_j$ and $\phi_k(s, z)$, which we achieve using a supervised least squares penalization approach. In particular, the FRAME fit is produced by minimizing the following criterion over a grid of possible values for the tuning parameter $\lambda \geq 0$:

(8)
$$\frac{1}{2} \sum_{i=1}^n \int \left\{ Y_i(s) - \sum_{j=1}^p g_j(\tilde{\boldsymbol{\theta}}_{ij}(s)^T \boldsymbol{\eta}_j) - \sum_{k=1}^q \phi_k(s, Z_{ik}) \right\}^2 ds$$
$$+ \lambda \left(\sum_{j=1}^p \rho(\|f_j\|) + \sum_{k=1}^q \rho(\|\phi_k\|) \right),$$

where $\|f_j\|^2 = \sum_{i=1}^n \int f_j(s, X_{ij})^2 \, ds$ with $f_j(s, X_{ij}) = g_j(\tilde{\boldsymbol{\theta}}_{ij}(s)^T \boldsymbol{\eta}_j)$, $\|\phi_k\|^2 = \sum_{i=1}^n \int \phi_k(s, Z_{ik})^2 \, ds$ and $\rho(\cdot)$ is a penalty function.

The first term in (8) corresponds to the squared error between $Y_i(s)$ and the FRAME prediction, integrated over s, and ensures an accurate fit to the data. The second term places a penalty on the ℓ_2 norms of the $f_j(x)$'s and $\phi_k(s, z)$'s. Note that penalizing the squared ℓ_2 norms, $\|f_j\|^2$ and $\|\phi_k\|^2$, would be analogous to performing ridge regression. However, we are penalizing the square root of this quantity, which has the effect of shrinking some of the functions exactly to zero and hence performing variable selection in a similar fashion to the group lasso [Simon et al. (2013), Yuan and Lin (2006)].

For a response sampled at a finite set of evenly spaced time points, $s_1, s_2, \ldots, s_L$, we approximate (8) by

(9)
$$\frac{1}{2L} \sum_{i=1}^n \sum_{l=1}^L \left\{ Y_{il} - \sum_{j=1}^p g_j(\tilde{\boldsymbol{\theta}}_{ijl}^T \boldsymbol{\eta}_j) - \sum_{k=1}^q \phi(s_l, Z_{ik}) \right\}^2$$
$$+ \lambda \left(\sum_{j=1}^p \rho(\|\mathbf{f}_j\|) + \sum_{k=1}^q \rho(\|\boldsymbol{\phi}_k\|) \right),$$

where $Y_{il} = Y_i(s_l)$, $\tilde{\boldsymbol{\theta}}_{ijl} = \tilde{\boldsymbol{\theta}}_{ij}(s_l)$, $\|\mathbf{f}_j\|^2 = \sum_{i=1}^n \sum_{l=1}^L g_j(\tilde{\boldsymbol{\theta}}_{ijl}^T \boldsymbol{\eta}_j)^2$ and $\|\boldsymbol{\phi}_k\|^2 = \sum_{i=1}^n \sum_{l=1}^L \phi(s_l, Z_{ik})^2$. Note that in using (9) we are implicitly assuming that the response has been sampled at a dense enough set of points that the integral is well approximated by the summation term. This approximation worked well for our data, but for sparsely sampled responses one would need to first fit a smooth approximation of the response and sample the fitted curve over a dense set of time points.

We further assume that $g_j(x)$ and $\phi_k(s, z)$ can, respectively, be well approximated by basis functions $\mathbf{h}(x)$ and $\boldsymbol{\omega}(s, z)$ such that $g_j(x) \approx \mathbf{h}(x)^T \boldsymbol{\xi}_j$ and $\phi_k(s, z) \approx \boldsymbol{\omega}(s, z)^T \boldsymbol{\alpha}_k$. At each response time point s_l, let $\mathbf{h}_{ijl} = \mathbf{h}(\tilde{\boldsymbol{\theta}}_{ijl}^T \boldsymbol{\eta}_j)$ and $\boldsymbol{\omega}_{ikl} = \boldsymbol{\omega}(s_l, z_{ik})$ with $\tilde{\boldsymbol{\theta}}_{ijl}$ defined in (9). Then using this basis representation, (9) can be expressed as

$$
\begin{aligned}
(10) \quad & \frac{1}{2L} \sum_{i=1}^{n} \sum_{l=1}^{L} \left\{ Y_{il} - \sum_{j=1}^{p} \mathbf{h}_{ijl}^T \boldsymbol{\xi}_j - \sum_{k=1}^{q} \boldsymbol{\omega}_{ikl}^T \boldsymbol{\alpha}_k \right\}^2 \\
& + \lambda \left(\sum_{j=1}^{p} \rho \left(\sqrt{\boldsymbol{\xi}_j^T H_j^T H_j \boldsymbol{\xi}_j} \right) + \sum_{k=1}^{q} \rho \left(\sqrt{\boldsymbol{\alpha}_k^T \Omega_k^T \Omega_k \boldsymbol{\alpha}_k} \right) \right),
\end{aligned}
$$

where H_j is a matrix with rows $\mathbf{h}_{1j1}, \mathbf{h}_{1j2}, \ldots, \mathbf{h}_{1jL}, \mathbf{h}_{2j1}, \ldots, \mathbf{h}_{njL}$ and Ω_k is defined similarly using $\boldsymbol{\omega}_{ikl}$. The FRAME fit is then produced by minimizing (10) over $\boldsymbol{\eta}_j, \boldsymbol{\xi}_j$ and $\boldsymbol{\alpha}_k$.

3.3. *FRAME optimization algorithm.* For a given value of λ, we break the problem of minimizing (10) into two iterative steps, where we first estimate $\boldsymbol{\xi}_j$ and $\boldsymbol{\alpha}_k$ given $\boldsymbol{\eta}_j$, and second estimate $\boldsymbol{\eta}_j$ given $\boldsymbol{\xi}_j$ and $\boldsymbol{\alpha}_k$. One advantage of this approach is that the minimization of (10) in the first step can be achieved using an efficient coordinate descent algorithm which we summarize in Algorithm 1.

Our approach has the same general form as similar algorithms used in other settings. In particular, arguments similar to those in Ravikumar et al. (2009) and Fan, James and Radchenko (2014) prove that Algorithm 1 will minimize a penalized

---

Algorithm 1 Step 1 of FRAME algorithm
---

    0. Initialize $S_j^H = (H_j^T H_j)^{-1} H_j^T$ and $S_k^\Omega = (\Omega_k^T \Omega_k)^{-1} \Omega_k^T$ for $j = 1, \ldots, p$ and $k = 1, \ldots, q$, where the matrices H_j and Ω_k are defined in (10).

For each $j \in \{1, \ldots, p\}$ and $k \in \{1, \ldots, q\}$:

    1. Fix all $\hat{\boldsymbol{\xi}}_{j'}$ for $j' \neq j$. Compute the residual vector $\mathbf{R}_j = \mathbf{Y} - \sum_{j' \neq j} H_{j'} \hat{\boldsymbol{\xi}}_{j'} - \sum_{k=1}^{q} \Omega_k \hat{\boldsymbol{\alpha}}_k$.

    2. Let $\hat{\boldsymbol{\xi}}_j = c_j S_j^H \mathbf{R}_j$ where $c_j = (1 - \lambda / \|H_j S_j^H \mathbf{R}_j\|)_+$ is a shrinkage parameter.

    3. Center $\hat{\mathbf{f}}_j \leftarrow \hat{\mathbf{f}}_j - \text{mean}(\hat{\mathbf{f}}_j)$.

    4. Fix all $\hat{\boldsymbol{\alpha}}_{k'}$ for $k' \neq k$. Compute the residual vector $\mathbf{R}_k = \mathbf{Y} - \sum_{j=1}^{p} H_j \hat{\boldsymbol{\xi}}_j - \sum_{k' \neq k} \Omega_{k'} \hat{\boldsymbol{\alpha}}_{k'}$.

    5. Let $\hat{\boldsymbol{\alpha}}_k = c_k S_k^\Omega \mathbf{R}_k$ where $c_k = (1 - \lambda / \|\Omega_k S_k^\Omega \mathbf{R}_k\|)_+$ is a shrinkage parameter.

    6. Center $\hat{\boldsymbol{\phi}}_k \leftarrow \hat{\boldsymbol{\phi}}_k - \text{mean}(\hat{\boldsymbol{\phi}}_k)$.

Repeat 1 through 6 and iterate until convergence.

Algorithm 2 FRAME algorithm

0. Choose initial values for $\hat{\boldsymbol{\eta}}_j$ for $j \in \{1, \ldots, p\}$.

1. Compute $\mathbf{h}_{ijl}$ using the current estimates for $\boldsymbol{\eta}_j$. Estimate $\boldsymbol{\xi}_j$ and $\boldsymbol{\alpha}_k$ using Algorithm 1.

2. Conditional on the $\boldsymbol{\xi}_j$'s and $\boldsymbol{\alpha}_k$'s from step 1, estimate the $\boldsymbol{\eta}_j$'s by minimizing (11).

3. Repeat steps 1 and 2 and iterate until convergence.

---

criterion of the form given by (10) provided $\rho(t) = t$. We discuss the extension to a general penalty function in the Appendix. Note that the S_j^H and S_k^Ω matrices defined in Algorithm 1 only need to be computed once so the calculations in 1 through 6 of Algorithm 1 can all be performed efficiently.

In the second step we estimate $\boldsymbol{\eta}_j$, given current estimates for the $\boldsymbol{\xi}_j$'s and $\boldsymbol{\alpha}_k$'s, by minimizing the sum of squares term

$$(11) \qquad \sum_{i=1}^{n} \sum_{l=1}^{L} \left\{ Y_{ij} - \sum_{j=1}^{p} \mathbf{h}(\tilde{\boldsymbol{\theta}}_{ijl}^T \boldsymbol{\eta}_j)^T \boldsymbol{\xi}_j - \sum_{k=1}^{q} \boldsymbol{\omega}_{ikl}^T \boldsymbol{\alpha}_k \right\}^2$$

over $\boldsymbol{\eta}_j$. Note that we do not include the penalty when estimating $\boldsymbol{\eta}_j$ because the $\boldsymbol{\eta}_j$'s are providing a direction in which to project $X_{ij}(t)$ and are thus constrained to be norm one. Hence, applying a shrinkage term would be inappropriate. Minimization of (11) can be approximately achieved using a first order Taylor series approximation of $g_j(x)$. We provide the details on this minimization in the Appendix.

Formally, the FRAME algorithm is summarized in Algorithm 2.

3.4. *Tuning parameters.* Fitting FRAME requires selecting the regularization parameter λ and the basis functions $\tilde{\mathbf{b}}(t)$, $\mathbf{b}(s, t)$, $\mathbf{h}(x)$ and $\boldsymbol{\omega}(s, t)$ defined in (5) and (10). For our simulations and the HSX data we used cubic splines to model $\mathbf{h}(x)$, $\tilde{\mathbf{b}}(t)$ and $\mathbf{b}(s, t)$, and a simple linear representation for $\boldsymbol{\omega}(s, z)$ so $\phi_k(s, z_k) = z_k \alpha_k$. We selected the dimensions of these bases simultaneously using 10-fold cross-validation (CV) based on prediction error. More specifically, we chose a grid of values for the dimension of each basis and randomly partitioned the original sample into 10 subsamples of equal size. For each $k = 1, \ldots, 10$, we used 9 subsamples to fit the model with dimensions of these bases fixed at a given combination of the grid values, and used the remaining subsample to calculate the prediction error. The cross-validated prediction error is then calculated as the average prediction error over the 10 validation subsamples. Thus, for every combination of basis dimensions, we obtained one cross-validated prediction error. The final selected dimensions for these basis functions are the ones which minimize the 10-fold cross-validated prediction error. Since the FRAME algorithm is very efficient, this approach worked well on our data.

To compute λ, one could potentially add a grid of values for λ to the above 10-fold CV, fit FRAME over all possible combinations of the tuning parameter values, and select the "best" value. However, a more efficient approach is to compute initial estimates for $\boldsymbol{\eta}_j$, minimize (10) over $\boldsymbol{\xi}_j$ and $\boldsymbol{\alpha}_k$ for each possible value of λ, choose the $\boldsymbol{\xi}_j$'s and $\boldsymbol{\alpha}_k$'s corresponding to the value of λ with the lowest 10-fold CV, estimate the $\boldsymbol{\eta}_j$'s for only this one set of parameters, and iterate. This approach means that, for each iteration, the minimization of (11) only needs to be performed for a single value of λ. We found this approach worked well for choosing the tuning parameters in both our simulated and real data analyses.

4. Simulations. In this section we conduct a simulation study to compare the performance of FRAME to several alternative functional approaches. We first generated $p = 6$ functional predictors using $X_{ij}(t) = \mathbf{F}(t)\boldsymbol{\theta}_{ij} + \varepsilon_{ij}(t)$, where $\mathbf{F}(t)$ was a 3-dimensional Fourier basis, $\boldsymbol{\theta}_{ij}$ was simulated from a $N(\mathbf{0}, \mathbf{I}_3)$ distribution, and the $\varepsilon_{ij}(t)$'s were independent over i, j and t with a $N(0, 0.1^2)$ distribution. Each predictor was sampled at 150 equally spaced time points over the interval $t \in [0, 1]$. In addition, $q = 8$ scalar predictors, Z_{ik}, were simulated from a standard normal distribution. Next, we generated $\beta_j(s, t) = \beta_{j1}(s) + \beta_{j2}(t) + 0.1\beta_{j1}(s)\beta_{j2}(t)$, where $\beta_{j1}(s) = \mathbf{b}(s)^T\boldsymbol{\eta}_{j1}$, $\beta_{j2}(t) = \mathbf{b}(t)^T\boldsymbol{\eta}_{j2}$, $\mathbf{b}(\cdot)$ was a 5-dimensional cubic spline basis, and $\boldsymbol{\eta}_{j1}$ and $\boldsymbol{\eta}_{j2}$ were independent $N(\mathbf{0}, \mathbf{I}_5)$ vectors.

The responses were generated from the model

$$Y_i(s_\ell) = \sum_{j=1}^{p} g_j\left(\int \beta_j(s_\ell, t)X_{ij}(t)\,dt\right) + \sum_{k=1}^{q} \gamma_k Z_{ik} + \varepsilon_i(s_\ell),$$

(12)
$$i = 1, \ldots, n,$$

where $\varepsilon_i(s_\ell) \sim N(0, 0.1^2)$ and $Y_i(s_\ell)$ was sampled at 20 equally spaced time points $s_1, \ldots, s_L$ over the interval $s \in [0, 1]$. We set $g_1(x) = \sin(x)$, $g_2(x) = \cos(x)$ and $g_j(x) = 0$ for $j = 3, \ldots, 6$. Thus, only the first two functional predictors were signal variables, with the remainder representing noise. Similarly, we set $\gamma_1 = 1$ and $\gamma_k = 0$ for $k = 2, \ldots, 8$ so the last seven scalar predictors were noise variables. All training data sets were generated using $n = 200$ observations.

We compared *FRAME* to six possible competitors. The simplest, *Mean*, ignored the predictors and used the average of the training response, at each time point s, to predict the responses on the test data. This method serves as a benchmark to illustrate the improvement in prediction accuracy that can be achieved using the predictors. The next method was the *Classical Functional Linear Regression* model given by (2). We fit (2) by computing the first G functional principal components (FPC) for the response function, and also the first K FPCs for each predictor function. We then used the 8 scalar predictors and the $6K$ FPC scores from the 6 functional predictors to fit separate linear regressions to each of the first G FPC scores on the response. To form a final prediction for the response function, we multiplied the estimated FPC scores by the first G principal component functions. The value

of G, between 1 and 4, and K, between 1 and 3, were both chosen using 10-fold cross-validation. The classical functional approach does not automatically perform variable selection, so we also fit a variant (*PCA-L*). The only difference between *Classical* and *PCA-L* is that the latter method used the group Lasso to compute the linear regressions between the response and predictor principal component scores and hence selected a subset of the predictors.

The fourth method, *PCA-NL*, was identical to *PCA-L* except that a nonlinear generalized additive model (GAM) was used to regress the response principal component scores on the predictor scores. Standard GAM does not automatically perform variable selection, so we fit *PCA-NL* using a variant of SPAM [Ravikumar et al. (2009)], which implements a penalized nonlinear additive model procedure and hence selects a subset of the predictors. We used the Lasso penalty function with the tuning parameter, λ, chosen over a grid of 20 values via 10-fold CV. Similarly, the dimension of the nonlinear functions used in SPAM were chosen, between 4 and 6, using 10-fold CV.

The next method, *Last Observation*, took as inputs $Z_{i1}, \ldots, Z_{i8}$ plus the last observed values of $X_{ij}(t)$, that is, $X_{i1}(t_{150}), \ldots, X_{i6}(t_{150})$. We then used the resulting 14 scalar predictors to estimate separate GAM regressions for the response at each observed point, $Y(s_1), \ldots, Y(s_{20})$, a total of 20 different regressions. As with *PCA-NL*, we used a variant of SPAM to perform variable selection. While using only the last observed time point may appear to be a naive approach, these methods are common in situations like the HSX data, where it is often assumed that all the information is captured at the latest time point. Hence, we implemented this approach to illustrate the potential advantage from incorporating the entire functional predictor.

The final comparison method, *FPCA-FAR*, combined the FPCA approach with the FAR method proposed in Fan, James and Radchenko (2014). FAR does not directly correspond to our setting because it is designed for problems involving functional predictors but only a scalar response. FPCA-FAR addresses this limitation by producing G separate FAR fits, one for each of the first G FPC scores. The FAR method has similar tuning parameters to SPAM, which were again chosen using 10-fold CV.

In fitting FRAME we set $\beta_j(s, t) = \beta_{j1}(s) + \beta_{j2}(t)$, where $\beta_{j1}(s)$, $\beta_{j2}(t)$ and $g_j(x)$ were approximated using cubic splines. The dimension of the basis for both $\beta_{j2}(t)$ and $\tilde{\beta}(t)$ was selected as the value among 4, 5, 6, which gave the lowest prediction error to $X_{ij}(t)$ on the held-out time points. In particular, for each possible dimension we held out every 5th observed time point for each $X_{ij}(t)$, produced a least squares fit using the remaining observations, and then calculated the squared error between the observed and predicted values of $X_{ij}(t)$ at the held-out time points. The value of λ and the dimensions of $\beta_{j1}(s)$ and $g_j(x)$ were all chosen using 10-fold CV in a similar fashion to the other comparison methods. We set ρ equal to the identity function, which corresponds to a group lasso type penalty function.

In order to match a real-life setting, we deliberately generated the data from a model that does not match the FRAME fit. In particular, the true $\beta_j(s, t)$ function included an interaction term, while the FRAME estimate was restricted to be additive, the predictors were generated from a Fourier basis but approximated using a spline basis, and the nonlinear functions, $g_1(x)$ and $g_2(x)$, were generated according to sin and cos functions, respectively, but approximated using a spline basis. In addition, all the various FRAME tuning parameters were automatically selected using CV, as part of the fitting process, so the true dimension of the basis functions was not assumed to be known.

We generated 100 different training data sets and fit each of the seven methods to all 100 data sets. False negative rates (FN), the fraction of signal variables incorrectly excluded, and false positive rates (FP), the fraction of noise variables incorrectly included, were computed. The prediction error, PE $= \frac{1}{20N} \sum_{i=1}^{N} \sum_{l=1}^{20} (Y_i(s_l) - \widehat{Y}_i(s_l))^2$, was also calculated on a large test data set with $N = 1000$ observations. The results, averaged over the 100 simulations, are displayed in Table 1, with standard errors shown in parentheses. Since the Last Observation method contains separate fits for each time point, its FN and FP rates are averaged over the twenty different fits. Figure 4 plots the prediction errors over s.

All methods show significant improvement over the Mean approach, indicating that the scalar and functional variables have real predictive ability. FRAME had perfect variable selection results on the scalar predictors, with false positive and false negative rates both being zero. All methods had zero false negative rates on the scalar predictors. However, PCA-NL and Last Observation both had high false positive rates. FRAME also did a much better job than all its competitors in

TABLE 1
False positive (FP) rates, false negative (FN) rates and their prediction errors (PE) for the seven comparison methods, averaged over the 100 simulation runs. The top rows relate to the functional predictors, $X_j(t)$, and the lower rows to the scalar predictors, Z_k. Standard errors are provided in parentheses

|		Mean	Classical	PCA-L	PCA-NL	Last Obs.	FPCA-FAR	FRAME
Functional	FP	–	–	0.0600	0.4200	0.2395	0.0375	0.0000
		–	–	(0.0182)	(0.0333)	(0.0101)	(0.0114)	(0.0000)
	FN	–	–	0.4400	0.0200	0.3002	0.1300	0.0600
		–	–	(0.0163)	(0.0098)	(0.0102)	(0.0220)	(0.0163)
Scalar	FP	–	–	0.0971	0.3671	0.2419	0.0400	0.0000
		–	–	(0.0175)	(0.0247)	(0.0089)	(0.0117)	(0.0000)
	FN	–	–	0.0000	0.0000	0.0000	0.0000	0.0000
		–	–	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
	PE	1.1983	0.1108	0.1040	0.1284	0.2727	0.0680	0.0651
		(0.0035)	(0.0030)	(0.0029)	(0.0024)	(0.0035)	(0.0019)	(0.0020)
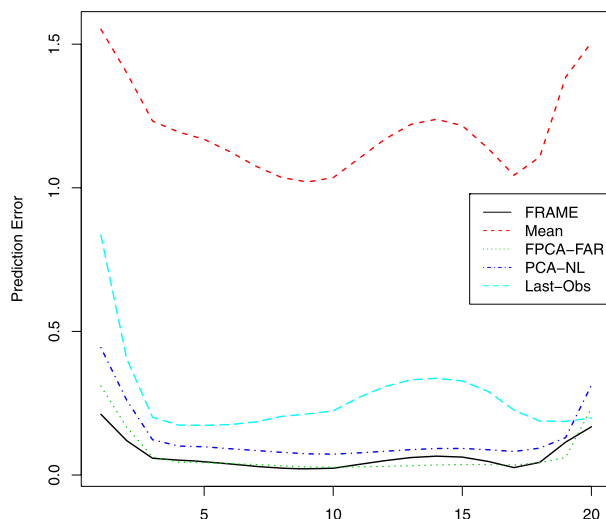
FIG. 4. *Mean prediction errors for five of the comparison methods at each of the* 20 *time points that the response function was observed over. The Classical and PCA-L curves were not plotted to make the figure easier to read.*

identifying the functional predictors. PCA-NL and Last Observation had high false positive rates for the functional predictors, and the PCA-L and Last Observation methods had high false negative rates. In terms of prediction error, FRAME is considerably superior to all methods except for FPCA-FAR. In comparing FRAME to FPCA-FAR, we note that while FRAME only results in a small improvement in terms of prediction error, it does a far better job in selecting the correct variables.

5. Forecasting demand decay rates. In this section we provide results from applying our FRAME approach to the HSX data. In doing so, we assume that the revenue curves of any two movies are independent, given the predictors. This assumption is not unreasonable because managers use strategic scheduling [Einav (2010)] to minimize the risk of two movies simultaneously competing for the same audience. More importantly, the HSX data (i.e., our predictors) have incorporated relevant information about the movies [Foutz and Jank (2010)]. Hence, one might expect much lower correlations among movies after conditioning on the predictors.

Figure 5 illustrates the modeling setup. Recall that for each movie we collect four functional predictors: the intra-day average price, the number of accounts shorting the stock, the number of shares sold and the number of shares held short. These curves capture related yet distinct aspects of consumer sentiment and word of mouth about a movie. The four functional predictors (represented using the green curve before the movie release in Figure 5) are observed from 52 up to 10 weeks prior to the movie's release. We then use FRAME to form predictions of $Y_i(s) = \log$(cumulative revenue for movie i at week s) (blue line after the movie release).
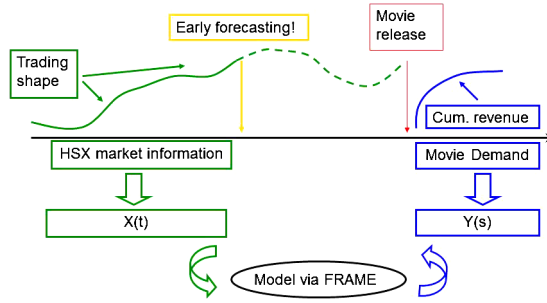
FIG. 5.   *Illustration of our model.*

In Section 5.1 we test the predictive accuracy of FRAME on the HSX data in relation to that of several competing methods. Then in Section 5.2 we discuss a graphical approach to obtain new insight into the relationship between VSMs and movies' success.

5.1. *Prediction accuracy.*   We compare a number of functional and nonfunctional methods to predict the box office cumulative revenue pattern for our 262 movies. Table 2 provides weekly mean absolute errors (MAE) between the predicted and actual cumulative box office revenue (on the log scale) for FRAME as well as six comparison methods. Specifically, we randomly divide the movies into training and test data (180 and 82 movies, resp.), fit the various methods using the training data and then compute MAE for week s on the test data:

$$(13) \qquad \text{MAE}(s) = \frac{1}{|\mathcal{T}|} \sum_{i \in \mathcal{T}} |Y_i(s) - \widehat{Y}_i(s)|,$$

TABLE 2

Mean absolute errors (*MAEs*) *on test data for FRAME and six competing methods averaged over twenty random partitions of the movies*

	Mean	Classical	PCA-L	PCA-NL	Last Obs.	FPCA-FAR	FRAME
Week 1	2.1898	1.5365	1.5856	1.1793	1.1534	1.2011	1.0952
Week 2	2.0490	1.4214	1.4582	1.0951	1.0683	1.1165	1.0116
Week 3	1.9057	1.3107	1.3372	1.0157	1.0335	1.0323	0.9482
Week 4	1.8335	1.2694	1.2900	0.9915	0.9970	1.0106	0.9364
Week 5	1.7907	1.2490	1.2666	0.9815	0.9923	1.0002	0.9305
Week 6	1.7610	1.2385	1.2527	0.9785	0.9944	0.9960	0.9324
Week 7	1.7418	1.2329	1.2431	0.9759	0.9868	0.9952	0.9371
Week 8	1.7294	1.2301	1.2379	0.9749	1.0132	0.9947	0.9397
Week 9	1.7199	1.2269	1.2337	0.9759	0.9938	0.9952	0.9432
Week 10	1.7144	1.2261	1.2322	0.9772	1.0051	0.9962	0.9460

where $\mathcal{T}$ represents the test data and $\widehat{Y}_i(s)$ the prediction for week s using a given method. We repeat this process over 20 random partitions of the movies and average the resulting MAE's. All seven methods are implemented in the same fashion as was used in the simulation analysis.

A few trends are clear from Table 2. First, all methods dominate Mean, indicating that the HSX curves contain useful predictive information. Second, the errors tend to decline over time, suggesting that there is more variability in the early weeks, but, to some extent, this averages out over time. Third, PCA-NL, FPCA-FAR and Last Observation give similar results and dominate Classical and PCA-L. Thus, there is clear evidence of a nonlinear relationship. Finally, FRAME provides superior results in comparison to the other six approaches for each of the ten weeks. The relative advantage of FRAME is highest in the first couple of weeks where predictions appear to be the most difficult.

Table 3 records the number of times each of the four predictors were selected, averaged over the 20 different training data sets. The intra-day average price variable appears to be the most important, with all methods selecting it on every run. FRAME also selected the variable of accounts trading short but ignored the remaining two predictors. By comparison, Last Observation chose the largest models, often including all four predictors. This may have been to compensate for the fact that the method only observed the final time point for each curve.

To further benchmark FRAME against alternative methods that are commonly used in the literature on movie demand forecasting [Sawhney and Eliashberg (1996)], Table 4 provides error rates for seven additional models. For each of these models, we estimate ten separate weekly linear regressions, one for each of the ten revenue weeks. We fit each regression to the training data, using the same 20 random partitions as in Table 2, and report the average MAE's on the test data. The first six models are based on individual movie features, respectively, genre (e.g., drama or comedy), sequel (yes/no), production budget (in dollars), MPAA rating, run time (in minutes) and studios (e.g., Universal or 20th Century Fox). The seventh model is based on a combination of all six features. The best individual predictor appears to be genre, but combining all six predictors gives the best results. However, the MAE's from the combined model are still significantly higher than

TABLE 3
Average number of times each of the four predictors were selected for each method

	Price	Account short	Shares sold	Shares short
FRAME	1.00	1.00	0.00	0.00
FPCA-FAR	1.00	0.30	0.00	0.00
PCA-L	1.00	0.80	0.00	0.00
PCA-NL	1.00	0.05	0.30	0.65
Last. Obs.	1.00	0.58	0.62	1.00

TABLE 4
Mean absolute errors on test data using various characteristics of the movies. Errors are averaged over twenty random partitions

	Genre	Sequel	Budget	Rating	Run time	Studio	All
Week 1	1.632	2.136	1.899	1.850	2.209	2.040	1.445
Week 2	1.589	2.003	1.762	1.749	2.064	1.915	1.395
Week 3	1.510	1.858	1.620	1.634	1.905	1.770	1.312
Week 4	1.487	1.792	1.564	1.604	1.829	1.714	1.304
Week 5	1.472	1.753	1.535	1.587	1.784	1.685	1.296
Week 6	1.463	1.728	1.516	1.578	1.755	1.668	1.291
Week 7	1.458	1.713	1.501	1.569	1.735	1.656	1.287
Week 8	1.457	1.703	1.492	1.563	1.723	1.648	1.286
Week 9	1.457	1.695	1.487	1.561	1.714	1.642	1.287
Week 10	1.458	1.691	1.484	1.559	1.709	1.639	1.287

for the best methods in Table 2, suggesting that the HSX curves provide additional information beyond that of the movie features.

5.1.1. *Why does FRAME predict so well?* We now offer a closer look into when (and potentially why) the prediction accuracy of FRAME is superior to that of the alternative methods in Tables 2 and 4. To that end, we investigate the relationship between FRAME's mean absolute percentage error (MAPE) in cumulative revenues over the first ten weeks since release and film characteristics, such as budget, genre, MPAA rating, and the volume and valence of critics' reviews. Similarly, we examine how the relative performance of FRAME (i.e., the difference between FRAME's MAPE and the lowest MAPE of either PCA-NL or FPCA-FAR) is associated with film characteristics. Tables 5 and 6 show the linear regression results.

Table 5 shows that FRAME performs well (i.e., has a low prediction error) for movies that are sequels, rated below *R*, have a shorter runtime, are released by a major studio such as Paramount, Warner Brothers, Universal or Twentieth Century Fox, and reviewed by a larger number of critics. Intuitively, these results suggest that FRAME performs especially well for movies that enjoy a greater capability for creating pre-release buzz. For instance, sequels build upon the success of their predecessors; films released by major studios benefit from significant advertising and publicity before opening; those with lower MPAA ratings, for example, G and PG, appeal to wider audiences; and greater attention from the critics, due to, for instance, a film's quality or controversies, could further fuel the public's fascination. Such firm- or consumer-generated buzz provides rich information to the HSX traders, who rapidly integrate the information into the stock trading. FRAME seems to be capable of capturing the dynamics of such buzz and translating it into accurate predictions.

TABLE 5
Linear regression of FRAME's prediction error on film characteristics

Name	Coefficient	Std err.	t	p-value
Intercept	0.098	0.068	1.439	0.151
Sequel	−0.033	0.014	−2.314	0.022
Budget	0.000	0.000	0.587	0.558
Action	−0.015	0.050	−0.296	0.768
Animated	0.016	0.054	0.306	0.760
Comedy	−0.009	0.050	−0.185	0.853
Drama	−0.011	0.050	−0.216	0.829
Horror	0.004	0.050	0.086	0.931
Other genres	0.066	0.060	1.098	0.273
Rating below R	−0.026	0.011	−2.417	0.016
Runtime	0.001	0.000	2.516	0.013
Major studio	−0.039	0.010	−3.744	0.000
Oscar	0.030	0.028	1.062	0.289
Critics volume	−0.001	0.000	−7.600	0.000
Critics valence	0.006	0.005	1.322	0.188
Consumer WOM volume	0.000	0.000	1.943	0.053
Consumer WOM valence	0.004	0.006	0.654	0.514

TABLE 6
Linear regression of the difference between FRAME's prediction error and the lowest error of either PCA-NL or FPCA-FAR on film characteristics

| Name	Coefficient	Std err.	t	p-value
Intercept	0.011	0.024	0.465	0.642
Sequel	0.000	0.005	0.036	0.971
Budget	0.000	0.000	−0.307	0.759
Action	−0.013	0.018	−0.746	0.456
Animated	−0.019	0.019	−1.023	0.308
Comedy	−0.018	0.017	−1.010	0.314
Drama	−0.023	0.017	−1.326	0.186
Horror	−0.012	0.018	−0.685	0.494
Other genres	0.015	0.021	0.721	0.471
Rating below R	−0.001	0.004	−0.302	0.763
Runtime	0.000	0.000	−1.101	0.272
Major studio	−0.006	0.004	−1.731	0.085
Oscar	0.017	0.010	1.764	0.079
Critics volume	0.000	0.000	3.198	0.002
Critics valence	−0.001	0.002	−0.533	0.595
Consumer WOM volume	−0.000	0.000	−3.901	0.000
Consumer WOM valence	0.004	0.002	1.936	0.054
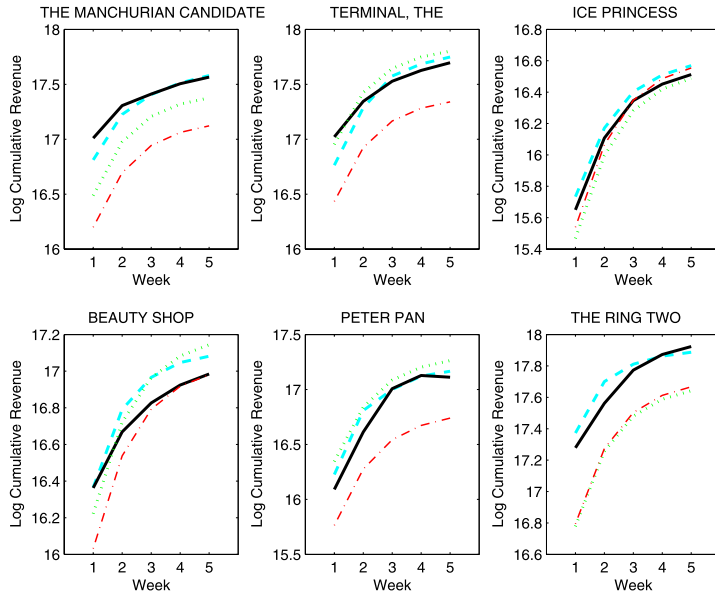
FIG. 6. *Top* 6 *movies with the smallest FRAME prediction error: the solid lines correspond to FRAME's prediction*; *the dashed lines show the corresponding true values. The two closest competitors are given by the dotted lines* (PCA-NL) *and the dash-dotted lines* (FPCA-FAR), *respectively.*

Figure 6 shows the six movies for which FRAME predicts the best in terms of MAPE. Two-thirds of these six movies were released by major studios with the exception of *THE RING TWO* and *THE TERMINAL*. Moreover, all of them were rated below R except for *THE MANCHURIAN CANDIDATE*. And all attracted more than a hundred critics' reviews. A third of them are sequels, specifically *PETER PAN* and *THE RING TWO*, as compared to 11% in the sample. Moreover, sequels are not far down the list. For example, FRAME also provides excellent predictions for sequels like *MISS CONGENIALITY 2* and *OCEAN'S TWELVE*. By contrast, FRAME predicts the least accurately for the following movies: *KAENA*: *THE PROPHECY*, *THE INTENDED* and *EULOGY*. None of these movies was a sequel or produced by a major studio. Only *KAENA*: *THE PROPHECY* had a below-R rating; and the volumes of critics' reviews for all three movies were below 35.

It is possible that movies with some of the above identified characteristics—sequels, low MPAA ratings, major studio releases and more critics' reviews—are easier to predict in general by any method, not only by FRAME. Indeed, Table 6 shows that FRAME does not have a statistically significant advantage (despite directionally so) over PCA-NL or FPCA-FAR in predicting demand for films of the above characteristics. Nonetheless, FRAME continues to outperform the alternative methods for films generating more viewer ratings online, suggesting its distinct ability to incorporate information potentially not captured by alternative methods,

such as potential viewers' interest that is not widely available ten weeks prior to a film's release.

5.2. *Model insight.* The previous section has shown that using a fully functional regression method such as FRAME can be beneficial for forecasting demand decay patterns. However, while nonlinear functional regression methods can result in good predictions, one downside is that because both model-input (HSX trading paths) as well as model-output (cumulative box office demand) arrive in the form of functions, it is hard to understand the relationship between the response and the predictors.

A useful graphical method to address this shortcoming is to visualize the relationship by generating candidate predictor curves, using the fitted FRAME model to predict corresponding responses and then plotting $X(t)$ and $Y(s)$ together. The idea is similar to the "partial dependence plots" described in Hastie, Tibshirani and Friedman (2001); however, in contrast to their approach, our plots take into account the joint effect of all predictors (and are hence not "partial"); we thus call our graphs "dependence plots."

Figure 7 displays several possible dependence plots with idealized input curves in the left panel and corresponding output curves from FRAME in the right panel. Note that since in our empirical analysis the intra-day average price was by far the most important predictor, we use that variable as $X(t)$ and fit FRAME with this single functional predictor. We study a total of four different scenarios. The top row corresponds to a situation where all input curves start and end at the same values (0 and 100, resp.); their only difference is how they get from the start to the end: the middle curve (solid line) grows at a linear rate; the upper and lower curves (dotted and dashed lines) grow at logarithmic and exponential rates, respectively. In that sense, the three curves represent movies whose HSX prices either grow at a constant (linear) rate, or grow fast early but then slow down (logarithmic) or grow slowly early only to increase toward release (exponential).

The top right panel shows the result: the logarithmic HSX price curve (dotted line) results in the largest cumulative revenue. In particular, its cumulative revenue is larger compared to the linear price curve (solid line), and both logarithmic and linear price curves beat the cumulative revenue generated by the exponential price curve (dashed line). In fact, the logarithmic price curve results in cumulative revenue that continues to grow significantly, especially in later weeks. This is in contrast to the cumulative revenue generated by the exponential price curve which becomes almost constant after week two or three.

What do these findings imply? Recall that all three HSX price curves start and end at the same value (0 and 100, resp.), so all observed differences are only with respect to their shape. This suggests that shapes matter enormously in VSMs. It also suggests that more buzz early on (i.e., the logarithmic shape) has much more impact on the overall revenue compared to a last moment hype closer to release time (i.e., the exponential shape).
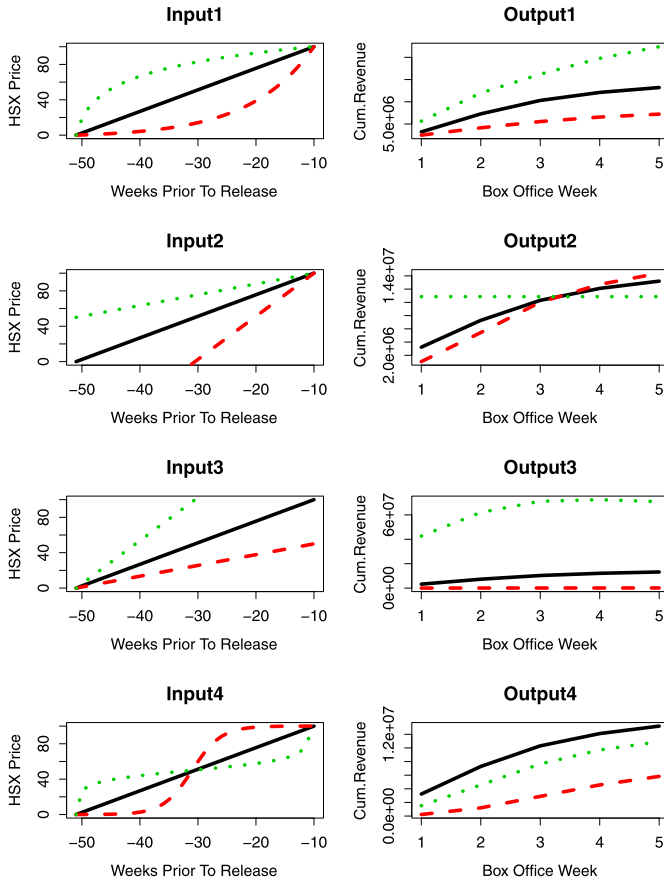
FIG. 7. *Dependence plots for different input shapes. The left panels contain various idealized input curves of HSX prices over time. Each figure plots three possible shapes for the observed HSX trading history of a movie. The right panels plot the corresponding predicted cumulative revenues using FRAME. For example, in the top row we observe that an HSX trading curve which increases rapidly and then levels off (dotted line) corresponds to a higher predicted revenue than either a linear pattern (solid line) or slow start with a large increase at the end (dashed line).*

The next two rows of Figure 7 show additional shape scenarios with both rows displaying input curves with a common linear shape. In the second row the curves are converging toward a common HSX value, while the input curves in the third row are diverging. The case of diverging curves suggests that the larger the most recent HSX value, the larger is the corresponding cumulative box office revenue. The converging case emphasizes the effect of recency of information: like in panel 1, all HSX price curves end at the same value; however, unlike in panel 1, they all have the same shape. We can see that the corresponding cumulative box office revenue also almost converges in week 5. This suggests that the difference in shape (e.g., linear vs. logarithmic vs. exponential) carries important information about

the *change* in the dynamics of word of mouth or consumer-generated buzz which translates into significant revenue differences.

The last row in Figure 7 shows yet another scenario of HSX price curves: an S-shape (dashed line) and an inverse-S shape (dotted line). Notice that the inverse-S shape features spurts of extreme growth both at the very beginning and at the very end, almost like a combination of logarithmic and exponential growth from panel 1. However, while the spurts resemble the logarithmic and exponential shapes, their overall magnitude is smaller compared to that in panel 1. As a result, the cumulative revenue is smaller compared to that of the linear growth. This suggests that while the dynamics of HSX price curves matter, their magnitude and timing matters even more, as the linear HSX price curve features a much more steady and sustained overall change in HSX prices compared to the inverse-S shape (which is constant most of the time with two small spurts at the beginning and the end). More evidence for this can be seen in the S-shaped HSX price curve (dashed line): while it does feature some change, most of the change happens in the middle of the curve which leads to the lowest of the three cumulative revenue curves.

6. Conclusion. This paper makes three significant contributions. First, we develop a new nonlinear regression approach, FRAME, which is capable of forming predictions on a functional response given multiple functional predictors and simultaneously conducting variable selection. Our results on both the HSX and simulated data demonstrate that FRAME is capable of providing a considerable improvement in prediction and variable selection accuracy relative to a host of competing methods. Second, we introduce a new and promising data source to the statistics community. Online virtual stock markets (VSMs) are market-driven mechanisms to capture opinions and valuations of large crowds in a single number. Our work shows that the information captured in VSMs is rich but requires appropriate and creative statistical methods to extract all available knowledge [Jank and Shmueli (2006)]. Finally, we make our approach practical for inference purposes by developing dependence plots to illustrate the relationship between input and output curves.

FRAME overcomes some of the technical difficulties encountered in other functional models. For instance, FRAME does not require the calculation of eigenfunctions, as is the case with our benchmark method, FPCA, in, for example, Tables 1 or 3. In FPCA, we first compute the principal components of the response curves, and then apply standard modeling techniques to the principal component scores. However, since the response curves are observed with random error, so are the corresponding eigenfunctions. While approaches for removing this random variation from the eigenfunctions exist [Yao, Müller and Wang (2005b)], FRAME does not rely on a principal component decomposition and thus does not encounter this type of challenge.

Our results have important implications for managerial practice. Equipped with the early forecasts of demand decay patterns, studio executives can make educated

decisions regarding weekly advertising allocations (both before and after the opening weekend), selection of the optimal release date to minimize competition with films from other studios and cannibalization of films from the same studio [Einav (2007)], and negotiation of the weekly revenue sharing percentages with the theater owners. Studios may be able to better manage distributional intensity and consumer word of mouth. For instance, for a movie predicted to have a strong opening weekend but fast decay afterward, the studio may consider nationwide release, as opposed to limited or platform release strategies (i.e., from initial limited release to nationwide release later on), at the same time strategically managing potentially negative word of mouth. The predicted demand decay of a film will also shed crucial light on a studio's sequential distributional strategies. For example, a studio may consider delaying (or shortening) a movie's video release or international release timing if the movie is predicted to have longevity (or faster decay) in theaters. Given that many academics have called for serious research on the optimal release timing in the subsequent distributional channels, such as home videos and international theatrical markets [Eliashberg, Elberse and Leenders (2006)], and that these channels represent five times more revenues than the domestic theatrical box office [MPAA (2007)], our results bear further crucial implications to the profitability of the motion picture industry.

A potential limitation of our approach is that it may only add value in inefficient markets where valuable information, above and beyond the information contained in the final trading price, is captured by the shape of the trading histories, such as prices, accounts and shares. However, as outlined earlier, previous research suggests that VSMs are not fully efficient. Furthermore, the strong predictive accuracy of our functional approach provides further empirical validation for this finding. In addition, the FRAME methodology is applicable beyond just VSM data. In general, it can be used on any regression problem involving functional predictors and responses.

We believe there are many other interesting applications of VSM's to different domains, such as music, TV shows and video games which all share similar characteristics to movies, such as frequent introductions of new, unique and experiential products, pop culture appeal and strong influence of hype on demand. Such research would be made possible by the increasing availability of data from VSMs for, for example, books (MediaPredict), music (HSX), TV shows (Inkling) and video games (SimExchange).

APPENDIX: ALGORITHM DETAILS

For a general penalty function, $\rho(t)$, we use the local linear approximation method proposed in Zou and Li (2008) to solve (10). The penalty function can be approximated as $\rho(\|\mathbf{f}\|) \approx \rho'(\|\mathbf{f}^*\|)\|\mathbf{f}\| + C$, where $\mathbf{f}^*$ is some vector that is close to $\mathbf{f}$ and C is a constant. Hence, the only required change to the FRAME algorithm for optimizing over general penalty functions is to replace λ by $\lambda^* = \lambda\rho'(\|\hat{\mathbf{f}}_j\|)$ in the calculation of c_j in 2, and replace λ by $\lambda^* = \lambda\rho'(\|\hat{\boldsymbol{\phi}}_k\|)$ in the calculation of c_k

in 5, where $\hat{\mathbf{f}}_j$ and $\hat{\boldsymbol{\phi}}_k$ represent the most recent estimates for $\mathbf{f}_j$ and $\boldsymbol{\phi}_k$. The initial estimates of $\hat{\mathbf{f}}_j$ and $\boldsymbol{\phi}_k$ can be obtained by using the Lasso penalty. This simple approximation allows the FRAME algorithm to be easily applied to a wide range of penalty functions.

To implement the second step of the FRAME algorithm, we minimize (11) with respect to the $\boldsymbol{\eta}_j$'s. Directly minimizing (11) is difficult due to the nonlinearity of the functions $g_j(x) \approx \mathbf{h}(x)^T \boldsymbol{\xi}_j$. To overcome this difficulty, we observe that, with the estimates $\hat{\boldsymbol{\xi}}_j$ and $\hat{\boldsymbol{\alpha}}_k$ from Algorithm 1 and the current value, $\boldsymbol{\eta}_{j,\text{old}}$, of $\boldsymbol{\eta}_j$, the first order approximation of $g(\tilde{\boldsymbol{\theta}}_{ijl}^T \boldsymbol{\eta}_j) \approx \mathbf{h}(\tilde{\boldsymbol{\theta}}_{ijl}^T \boldsymbol{\eta}_j)^T \hat{\boldsymbol{\xi}}_j$ is

$$\mathbf{h}(\tilde{\boldsymbol{\theta}}_{ijl}^T \boldsymbol{\eta}_j)^T \hat{\boldsymbol{\xi}}_j \approx \mathbf{h}(\tilde{\boldsymbol{\theta}}_{ijl}^T \boldsymbol{\eta}_{j,\text{old}})^T \hat{\boldsymbol{\xi}}_j + \mathbf{h}'(\tilde{\boldsymbol{\theta}}_{ijl}^T \boldsymbol{\eta}_{j,\text{old}})^T \hat{\boldsymbol{\xi}}_j \cdot \tilde{\boldsymbol{\theta}}_{ijl}^T (\boldsymbol{\eta}_j - \boldsymbol{\eta}_{j,\text{old}}).$$

Thus, we can approximate (11) by

$$(14) \qquad \sum_{i=1}^{n} \sum_{l=1}^{n_i} \left(R_{il} - \sum_{j=1}^{p} \mathbf{h}'(\tilde{\boldsymbol{\theta}}_{ijl}^T \boldsymbol{\eta}_{j,\text{old}})^T \hat{\boldsymbol{\xi}}_j \cdot \tilde{\boldsymbol{\theta}}_{ijl}^T (\boldsymbol{\eta}_j - \boldsymbol{\eta}_{j,\text{old}}) \right)^2,$$

where $R_{il} = Y_{il} - \sum_{j=1}^{p} \mathbf{h}(\tilde{\boldsymbol{\theta}}_{ijl}^T \boldsymbol{\eta}_{j,\text{old}})^T \hat{\boldsymbol{\xi}}_j - \sum_{k=1}^{q} \boldsymbol{\omega}_{ikl}^T \hat{\boldsymbol{\alpha}}_k$. The above approximation (14) is a quadratic function of $\boldsymbol{\eta}_j$ and can be minimized easily. Hence, the new value of $\boldsymbol{\eta}_j$ is updated as the minimizer of (14). We also note that if the estimate $\hat{\boldsymbol{\xi}}_j$ from Algorithm 1 is $\mathbf{0}$, then the corresponding value of $\boldsymbol{\eta}_j$ will not be updated.

REFERENCES

BAR-JOSEPH, Z., GERBER, G. K., GIFFORD, D. K., JAAKKOLA, T. S. and SIMON, I. (2003). Continuous representations of time-series gene expression data. *J. Comput. Biol.* **10** 341–356.

BASS, F. M., GORDON, K., FERGUSON, T. L. and GITH, M. L. (2001). DIRECTV: Forecasting diffusion of a new technology prior to product launch. *Interfaces* **31** S82–S93.

CHEN, D., HALL, P. and MÜLLER, H.-G. (2011). Single and multiple index functional regression models with nonparametric link. *Ann. Statist.* **39** 1720–1747. MR2850218

CRAINICEANU, C., CAFFO, B. and MORRIS, J. (2011). Multilevel functional data analysis. In *The SAGE Handbook of Multilevel Modeling*. SAGE, Thousand Oaks, CA.

DI, C.-Z., CRAINICEANU, C. M., CAFFO, B. S. and PUNJABI, N. M. (2009). Multilevel functional principal component analysis. *Ann. Appl. Stat.* **3** 458–488. MR2668715

EINAV, L. (2007). Seasonality in the U.S. motion picture industry. *Rand J. Econ.* **38** 127–145.

EINAV, L. (2010). Not all rivals look alike: Estimating an equilibrium model of the release date timing game. *Econ. Inq.* **48** 369–390.

ELIASHBERG, J., ELBERSE, A. and LEENDERS, M. (2006). The motion picture industry: Critical issues in practice, current research, and new research directions. *Mark. Sci.* **25** 638–661.

FAN, Y., JAMES, G. and RADCHENKO, P. (2014). Functional additive regression. Under review.

FERRATY, F. and VIEU, P. (2002). The functional nonparametric model and application to spectrometric data. *Comput. Statist.* **17** 545–564. MR1952697

FERRATY, F. and VIEU, P. (2003). Curves discrimination: A nonparametric functional approach. *Comput. Statist. Data Anal.* **44** 161–173. MR2020144

FOUTZ, N. and JANK, W. (2010). Pre-release demand forecasting for motion pictures using functional shape analysis of virtual stock markets. *Mark. Sci.* **29** 568–579.

HASTIE, T., TIBSHIRANI, R. and FRIEDMAN, J. (2001). *The Elements of Statistical Learning*. Springer, New York. MR1851606

JAMES, G. M. and HASTIE, T. J. (2001). Functional linear discriminant analysis for irregularly sampled curves. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **63** 533–550. MR1858401

JAMES, G. M., HASTIE, T. J. and SUGAR, C. A. (2000). Principal component models for sparse functional data. *Biometrika* **87** 587–602. MR1789811

JAMES, G. M. and SILVERMAN, B. W. (2005). Functional adaptive model estimation. *J. Amer. Statist. Assoc.* **100** 565–576. MR2160560

JAMES, G. M. and SUGAR, C. A. (2003). Clustering for sparsely sampled functional data. *J. Amer. Statist. Assoc.* **98** 397–408. MR1995716

JANK, W. and SHMUELI, G. (2006). Functional data analysis in electronic commerce research. *Statist. Sci.* **21** 155–166. MR2324075

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

RAVIKUMAR, P., LAFFERTY, J., LIU, H. and WASSERMAN, L. (2009). Sparse additive models. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **71** 1009–1030. MR2750255

RICE, J. A. and WU, C. O. (2001). Nonparametric mixed effects models for unequally sampled noisy curves. *Biometrics* **57** 253–259. MR1833314

SAWHNEY, M. S. and ELIASHBERG, J. (1996). A parsimonious model for forecasting gross box office revenues of motion pictures. *Mark. Sci.* **15** 113–131.

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

SPANN, M. and SKIERA, B. (2003). Internet-based virtual stock markets for business forecasting. *Manage. Sci.* **49** 1310–1326.

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

YAO, F., MÜLLER, H.-G. and WANG, J.-L. (2005a). Functional linear regression analysis for longitudinal data. *Ann. Statist.* **33** 2873–2903. MR2253106

YAO, F., MÜLLER, H.-G. and WANG, J.-L. (2005b). Functional data analysis for sparse longitudinal data. *J. Amer. Statist. Assoc.* **100** 577–590. MR2160561

YUAN, M. and LIN, Y. (2006). Model selection and estimation in regression with grouped variables. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **68** 49–67. MR2212574

ZEGER, S. L. and DIGGLE, P. J. (1994). Semiparametric models for longitudinal data with applications to CD4 cell numbers in HIV seroconverters. *Biometrics* **50** 689–699.

ZHANG, S., JANK, W. and SHMUELI, G. (2010). Real-time forecasting of online auctions via functional k-nearest neighbors. *Int. J. Forecast.* **26** 666–683.

ZOU, H. and LI, R. (2008). One-step sparse estimates in nonconcave penalized likelihood models. *Ann. Statist.* **36** 1509–1533. MR2435443

Y. FAN
G. M. JAMES
UNIVERSITY OF SOUTHERN CALIFORNIA
BRIDGE HALL 401
LOS ANGELES, CALIFORNIA 90089-0809
USA

N. FOUTZ
MCINTIRE SCHOOL OF COMMERCE
UNIVERSITY OF VIRGINIA
340 ROUSS AND ROBERTSON HALL
CHARLOTTESVILLE, VIRGINIA 22904
USA

W. JANK
INFORMATION SYSTEMS DECISION SCIENCES DEPARTMENT
UNIVERSITY OF SOUTH FLORIDA
4202 E. FOWLER AVENUE
BSN 3403
TAMPA, FLORIDA 33620
USA
E-MAIL: wjank@usf.edu