Functional Response Additive Model Estimation with Online Virtual Stock Markets

Yingying Fan‡, Natasha Foutz§, Gareth M. James‡, Wolfgang Jank†

‡Marshall School of Business, University of Southern California
§McIntire School of Commerce, University of Virginia
†College of Business, University of South Florida

February 24, 2012

Abstract

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 which extends the usual linear regression model to situations involving both functional responses, $X_j(t)$, and functional predictors, $Y(t)$. Our approach uses a penalized least squares optimization criterion to automatically perform variable selection in situations involving multiple functional predictors. In addition our method uses an efficient coordinate descent algorithm to fit general non-linear additive relationships between the predictors and response.

We apply our model to the context of forecasting product demand in the entertainment industry. In particular, we 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' opinion about a particular product. Our fully functional model captures the pattern of pre-release VSM trading values and provides superior predictive accuracy of a movie's demand distribution 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.

Key words and phrases: Functional data; non-linear regression; penalty functions; forecasting; virtual markets; movies; Hollywood.

1 Introduction

Functional data analysis (FDA) has become very popular in recent years, in part because of its ability to capture patterns and shapes in a parsimonious and automated fashion (Ramsay and Silverman, 2005). Some of the areas in which FDA has been applied include functional principal components analysis (James *et al.*, 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 (James and Hastie, 2001; Ferraty and Vieu, 2003), functional clustering (James and Sugar, 2003; Bar-Joseph *et al.*, 2003), or functional forecasting (Zhang et al., 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 et al., 2005), which is often an overly restrictive assumption. Recently Fan and James (2011) proposed an approach,"Functional Additive Regression" (FAR), for fitting a non-linear functional regression model of the form

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

where the f_j's are general non-linear functions of $X_{ij}(t)$ and Y_i is a centered response. Their method uses a penalized least squares criterion and is capable of automatically performing variable selection even for very large values of p. While the approach of Fan and James has several desirable properties it is only designed for data with scalar responses. The data that motivated our research includes not only functional predictors but also functional responses.

Hence, we propose an extension of (1), called "Functional Response Additive Model Estimation" (FRAME), to model functional responses. Our non-linear approach allows us to model much more subtle relationships and we show that, on our data, it produces clear improvements in prediction accuracy. We also address the difficulty of interpreting the results from a model involving functional predictors and responses using "dependence plots" which graphically illustrate, for typical shapes of the predictors, the corresponding predicted response pattern. These dependence plots allow for a glimpse into the causal relationship between response and predictors and provide actionable insight for decision makers.

We illustrate the predictive power of our model in the context of the entertainment industry. Providing accurate forecasts of the success of new products is crucial for the major entertainment industries (motion picture, music, TV, gaming, or publishing), which 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* (Sawhney and Eliashberg, 1996; Bass et al, 2001) in order to allocate, for example, weekly advertising budgets according to the predicted *rate of demand decay*, i.e. 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. Even for movies with similar demand patterns (e.g. those on the second row of Figure 1), the speed of decay varies greatly.

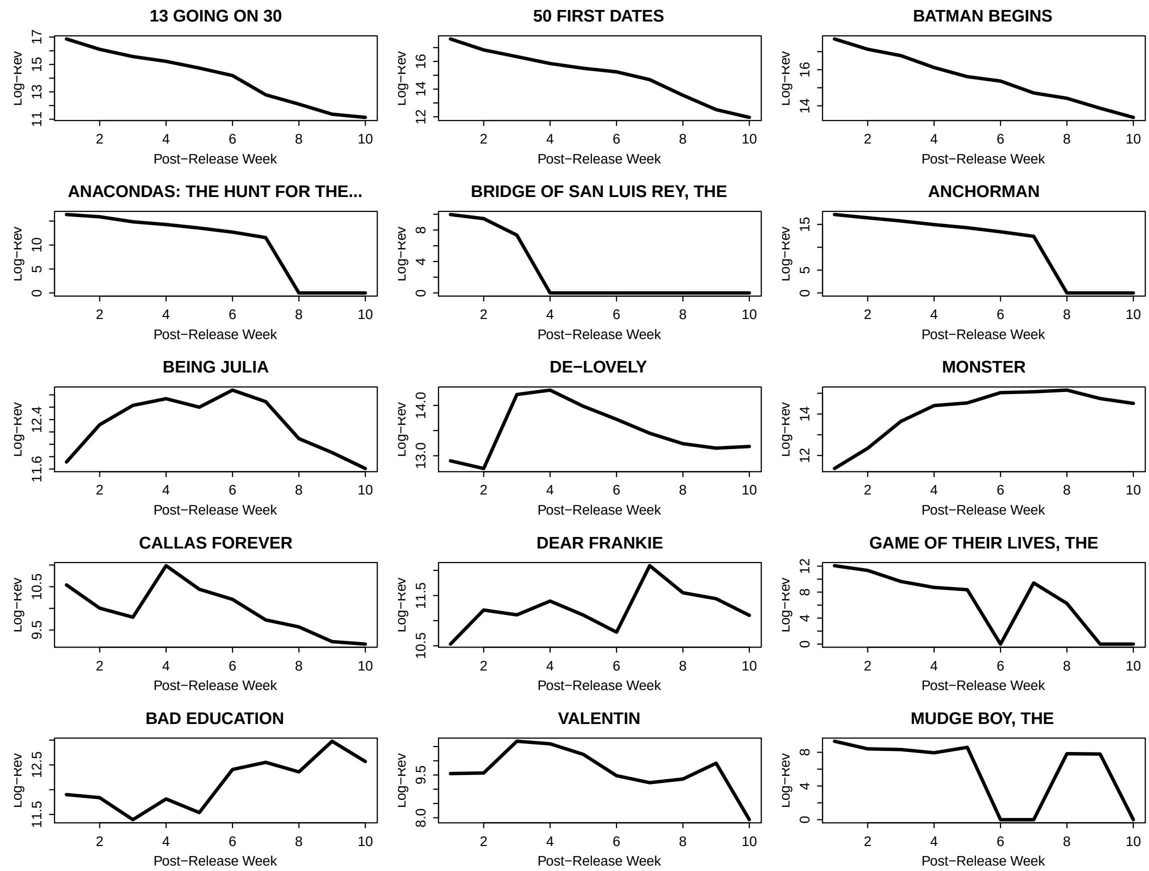

Figure 1: Movie demand decay rates for a sample of movies.

In this paper we propose to use our FRAME method to forecast the demand patterns of box office revenues. The functional predictors of our model capture consumers' word-of-mouth via a novel data source, online virtual stock markets (VSMs). In a VSM, participants trade virtual 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 may provide early and reliable demand forecasts (Spann and Skiera, 2003; Foutz and Jank, 2009). 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, 2009). 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. Section 3 briefly describes the approach of Fan and James (2011). We then develop FRAME, an extension to the functional response domain, and provide an efficient coordinate descent fitting algorithm. We demonstrate the superior performance of FRAME, in comparison to a large number of competitors, with an extensive simulation study in Section 4. Section 5 contains the results from our implementation of FRAME on the movie data. We also illustrate the insights that can be gained from our approach using dependence plots. 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 daily trading histories of an online virtual stock market for movies; our output data (i.e. functional responses) pertain to the 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 an event or a parameter (rather than a company's shares). For instance, a value of 54 cents for the stock "A democratic candidate will win the Presidential election" could be interpreted as the traders' collective belief that the democratic candidate has a 54% chance of winning. If in fact the democratic candidate wins, then traders holding the democratic candidate's stock will liquidate (or "cash-in") at $1 per share; otherwise they receive $0.

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 the first 4 weeks of a movie's revenues. HSX has nearly 2 million active participants worldwide and each trader is initially endowed with $2

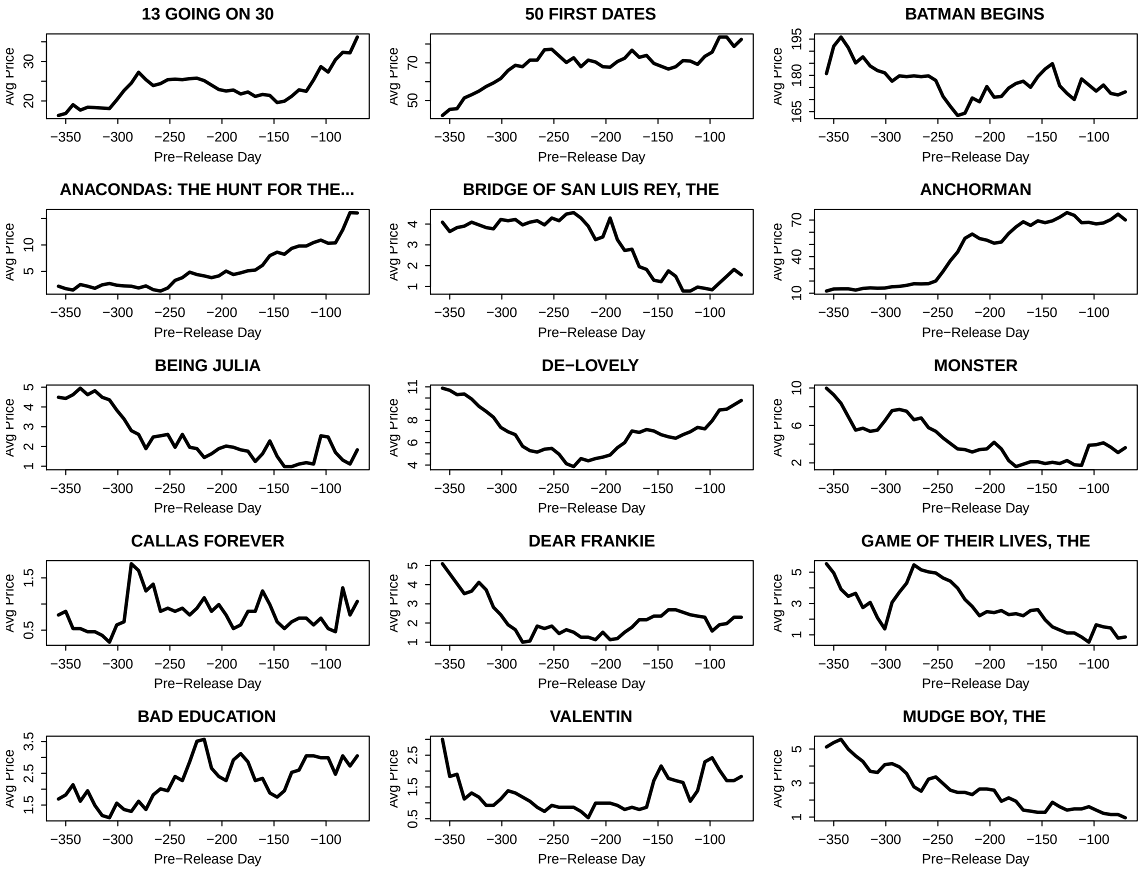

Figure 2: HSX trading histories for the sample of movies from Figure 1.

million virtual currency and can increase his or her net worth by strategically selecting and trading movie stocks (i.e. buying low and selling high). Traders are further motivated by opportunities to exchange the accrued currency for merchandize and to appear on the daily *Leader Board* that features the most successful traders. Figure 2 shows the sample of HSX trading histories corresponding to 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 day -70 in Figure 2). Predicting movie decay ten weeks prior to release gives managers ample time to make informed decisions about marketing mix allocations and other strategic decisions.

Our FRAME method captures differences in shapes of VSM price histories, e.g. trending up or down, concavity vs. convexity, or last-moment spurts. The empirical results in Section 5 show that these 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 strong interest in a product. Thus it can suggest a strong *initial* demand immediately after a new product's introduction to the market place, e.g. 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 last-moment positive word-of-mouth, which may lead to both a strong opening weekend and a reduced decay rate in demand (or increased longevity) of a new product.

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 theater 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. (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 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 while some movies decay gradually, a number have sudden drops, while other movies initially increase after the release week. Our goal is to characterize different demand decay *shapes* and to use the information from VSMs to forecast these shapes.

3 Functional Response Additive Model Estimation

In this section we first briefly summarize a method for fitting non-linear functional regressions involving a scalar response and then develop our extension to functional responses.

3.1 Functional Adaptive Regression

Fan and James (2011) propose a non-linear functional method for fitting (1) which they call Functional Additive Regression (FAR). The FAR approach models $f_j(X_{ij})$ using a single index model of the form,

$$f_j(X_{ij}) = g_j\left(\int \beta_j(t)X_{ij}(t)dt\right), \tag{2}$$

where g_j and β_j are both smooth non-parametric functions with the constraint $\|\beta_j\| = 1$. Using this non-linear representation the FAR model can be expressed as,

$$Y_i = \sum_{j=1}^{p} g_j\left(\int \beta_j(t)X_{ij}(t)dt\right) + \varepsilon_i. \tag{3}$$

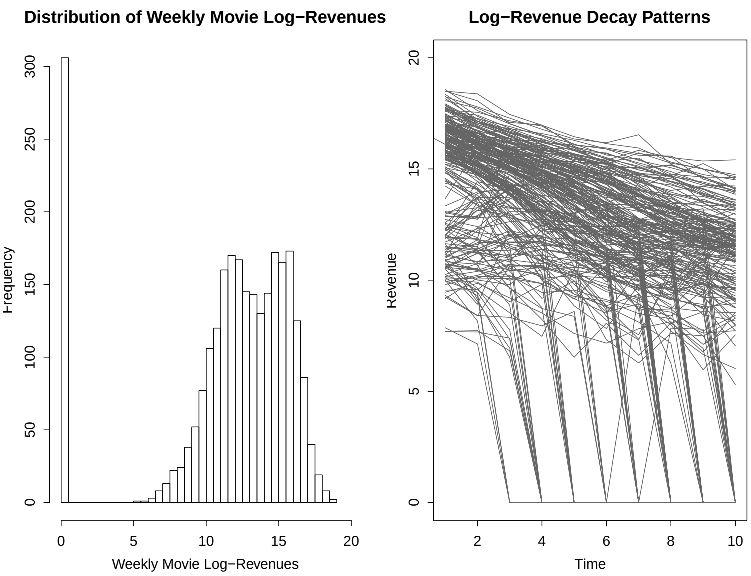

Figure 3: Distribution of movies' weekly demand and demand decay patterns. The right panel shows 10-week decay patterns (from the release week-end until 9 weeks after release) for the 262 movies in our sample; the left panel shows the distribution of the corresponding $10 \times 262 = 2,620$ weekly log-revenues.

This model is fit by minimizing a penalized least squares criterion,

$$\frac{1}{2}\left\|\mathbf{Y} - \sum_{j=1}^{p} \mathbf{f}_j\right\|^2 + \lambda \sum_{j=1}^{p} \rho(\|\mathbf{f}_j\|). \tag{4}$$

where $\mathbf{f}_j = (f_j(X_{1j}), ..., f_j(X_{nj}))^T \in R^n$ with $f_j(X_{ij})$ given by (2), $\mathbf{Y} = (Y_1, \ldots, Y_n) \in R^n$ and $\|\mathbf{f}_j\| = \sqrt{\mathbf{f}_j^T \mathbf{f}_j}$ represents the 2-norm of the vector $\mathbf{f}_j$. The penalty function, ρ, can take on different forms. When ρ is set to the identity function, (4) has a close relationship to the group lasso criterion but other functional forms, such as the SCAD penalty, can also be implemented. The penalty term has the effect of shrinking certain f_j to zero and hence performing variable selection in a similar fashion to the Lasso (Tibshirani, 1996).

The FAR approach models $\beta_j(t)$ and $X_{ij}(t)$ using an orthogonal q-dimensional basis, $\mathbf{b}(t)$, such that $\beta_j(t) = \mathbf{b}(t)^T \boldsymbol{\eta}_j$ and $X_{ij} = \mathbf{b}(t)^T \boldsymbol{\theta}_{ij}$. In addition $g_j(t)$ is approximated by a d-dimensional basis, $\mathbf{h}(t)$, such that $g_j(t) = \mathbf{h}(t)^T \boldsymbol{\xi}_j$. Using these basis representations (4) can be expressed as,

$$l_\lambda(\boldsymbol{\xi}|\boldsymbol{\eta}) = \frac{1}{2}\left\|\mathbf{Y} - \sum_{j=1}^{p} H_j \boldsymbol{\xi}_j\right\|^2 + \lambda \sum_{j=1}^{p} \rho(\|H_j \boldsymbol{\xi}_j\|), \tag{5}$$

where H_j is an n by d matrix with ith row given by $\mathbf{h}\left(\boldsymbol{\theta}_{ij}^T \boldsymbol{\eta}_j\right)^T$. The basis coefficient $\boldsymbol{\theta}_{ij}$ can be computed directly from $X_{ij}(t)$, so provided the predictors are densely sampled, $\boldsymbol{\theta}_{ij}$ is assumed known. Hence, to fit FAR one must minimize (5) over $\boldsymbol{\eta}$ and $\boldsymbol{\xi}$.

One could potentially use a coordinate descent algorithm to minimize $l_\lambda(\boldsymbol{\xi}|\boldsymbol{\eta})$ jointly over $\boldsymbol{\xi}$ and $\boldsymbol{\eta}$. However, this turns out to be a poor approach for several reasons, related to the lack of convexity of (5) and potentially unstable parameter estimates. Instead, FAR adopts a two stage algorithm, analogous to a profile likelihood approach, where the $\boldsymbol{\eta}_j$'s are first estimated in a supervised fashion, and then (5) is optimized conditional on $\boldsymbol{\eta}_j$. This approach has the advantage of providing a more accurate estimate for the $\boldsymbol{\eta}_j$'s in comparison to an unsupervised approach such as PCA, while avoiding the computational and practical difficulties of the joint optimization method. Fan and James (2011) show that the following algorithm can be successfully used to fit FAR.

FAR Algorithm

A. Given initial values for the $\boldsymbol{\xi}_j$'s, compute the $\boldsymbol{\eta}_j$'s as the values minimizing

$$Q = \sum_{i=1}^{n} \left(Y_i - \sum_{j=1}^{p} \mathbf{h}\left(\boldsymbol{\theta}_{ij}^T \boldsymbol{\eta}_j\right)^T \boldsymbol{\xi}_j \right)^2 . \tag{6}$$

B. Conditional on the estimates for $\boldsymbol{\eta}_1, \ldots, \boldsymbol{\eta}_p$ from Step A., minimize $l_\lambda(\boldsymbol{\xi}|\boldsymbol{\eta})$ over $\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_p$ using the following coordinate descent algorithm.

For each $j \in \{1, ..., p\}$,

1. Fix all $\widehat{\mathbf{f}}_k$ for $k \neq j$. Compute the residual vector $\mathbf{R}_j = \mathbf{Y} - \sum_{k \neq j} \widehat{\mathbf{f}}_k(X_k)$. Let $S_j = \widehat{H}_j(\widehat{H}_j^T \widehat{H}_j)^{-1} \widehat{H}_j^T$ where the ith row of $\widehat{H}_j$ is given by $\mathbf{h}\left(\boldsymbol{\theta}_{ij}^T \widehat{\boldsymbol{\eta}}_j\right)^T$.
2. Let $\widehat{\mathbf{P}}_j = S_j \mathbf{R}_j$ and $\phi_j = \rho'(\|\widehat{\mathbf{f}}_j\|)$ where $\widehat{\mathbf{f}}_j$ represents the most recent estimate for $\mathbf{f}_j$.
3. Let $\widehat{\mathbf{f}}_j = \alpha_j \widehat{\mathbf{P}}_j$ where $\alpha_j = \left(1 - \phi_j \lambda / \|\widehat{\mathbf{P}}_j\|\right)_+$ is a shrinkage parameter and $[x]_+$ represents the positive part of x.
4. Center $\widehat{\mathbf{f}}_j \leftarrow \widehat{\mathbf{f}}_j - \text{mean}(\widehat{\mathbf{f}}_j)$.

Repeat Steps 1. through 4. until convergence.

The coordinate descent algorithm is easy to implement because the estimate of $\mathbf{f}_j$ can be broken down into two distinct steps. First, step 2. computes the standard least squares estimate and then step 3. applies a simple shrinkage term, α_j. If $\alpha_j = 0$ then the predictor is automatically removed from the model. In practice, this algorithm is implemented over a grid of tuning parameters, $\lambda_1, \ldots, \lambda_T$. Hence, for a given λ_t, the values for the $\boldsymbol{\xi}_j$'s in Step A. are obtained as the final Step

B. estimates from the previous iteration using λ_{t-1}. Fan and James (2011) discuss approaches for selecting initial values of the $\boldsymbol{\xi}_j$'s and $\boldsymbol{\eta}_j$'s for $\lambda = \lambda_1$, and for minimizing (6).

3.2 Functional Response and Predictor

Our data contains both functional predictors, $X_{ij}(t)$ and functional responses, $Y_i(t)$. In general one can decompose the centered response as,

$$Y_i(t) = \sum_{m=1}^{M} \gamma_{im} e_m(t) + r_i(t), \tag{7}$$

where $e_1(t), \ldots, e_M(t)$ represent an approximate M-dimensional basis, $\gamma_{i1}, \ldots, \gamma_{iM}$ the corresponding basis coefficients for the ith response and $r_i(t)$ is an error term. Hence, we extend the FAR approach to the functional response domain by modeling γ_{im} as a non-linear function of $X_{ij}(t)$ i.e.

$$\gamma_{im} = \sum_{j=1}^{p} g_{jm} \left(\int \beta_{jm}(t) X_{ij}(t) dt \right). \tag{8}$$

In this formulation $e_m(t), \gamma_{im}, g_j(t)$, and $\beta_j(t)$ are all unobserved parameters. We call our method "Functional Response Additive Model Estimation" (FRAME). To ease notation we will first describe FRAME for $M = 1$.

A natural generalization of the FAR criterion is to fit FRAME by minimizing the penalized least squares criterion,

$$\frac{1}{2} \sum_{i=1}^{n} \int \left\{ Y_i(t) - \gamma_i e(t) \right\}^2 dt + \lambda \sum_{j=1}^{p} \rho(\|\mathbf{f}_j\|), \tag{9}$$

subject to

$$\int e^2(t) dt = \int \beta_j^2(t) dt = 1, \quad j = 1, \ldots, p, \tag{10}$$

and

$$\sum_{i=1}^{n} \left\{ \int Y_i(t) e(t) dt \right\}^2 = s. \tag{11}$$

The constraints given by (10) respectively ensure that γ and $e(t)$ are identifiable and that $g(t)$ and $\beta(t)$ are identifiable. The motivation for the second constraint, (11), it more subtle and we discuss this point further after the FRAME algorithm. The following Lemma is integral to optimizing (9).

Lemma 1. *Subject to the constraints (10) and (11) there is a one to one correspondence between the parameters that minimize (9) and those which minimize*

$$\frac{1}{2} \left\| \int \mathbf{Y}(t) e(t) dt - \sum_{j=1}^{p} g_j \left(\int \beta_j(t) \mathbf{X}_j(t) dt \right) \right\|^2 + \lambda \sum_{j=1}^{p} \rho(\|\mathbf{f}_j\|) \tag{12}$$

Lemma 1 can be verified by expanding and rearranging (9). This lemma suggests a two step algorithm for fitting FRAME.

FRAME Algorithm

0. Initialize $e(t)$ using the first functional principal component of $Y_i(t)$.

1. Given the current estimate for $e(t)$ fit the scalar FAR model using the pseudo response, $\tilde{Y}_i = \int Y_i(t)\hat{e}(t)dt$.

2. Given the current estimates for γ_i from (8), compute $e(t)$ as the function that minimizes

$$\left\| \widehat{\boldsymbol{\gamma}} - \int \mathbf{Y}(t)e(t)dt \right\|^2 \tag{13}$$

   subject to (10) and (11).

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

Note that, as a direct consequence of Lemma 1, at each step of this algorithm (9) is guaranteed to decline. Step 2. is implemented by modeling $Y(t)$ and $e(t)$ using a k-dimensional orthogonal basis function, $\mathbf{s}(t)$, i.e. $Y_i(t) = \mathbf{s}(t)^T\boldsymbol{\tau}_i$ and $e(t) = \mathbf{s}(t)^T\boldsymbol{\pi}$. In this case (13) reduces to,

$$\|\widehat{\boldsymbol{\gamma}} - \tau\boldsymbol{\pi}\|^2, \tag{14}$$

where τ is a matrix with ith row $\boldsymbol{\tau}_i$, which can be computed directly from the $Y_i(t)$'s. Hence, minimizing (13) becomes a constrained quadratic programming problem which can be solved using a variety of standard optimization packages.

The implementation of the constraint, (11), has two advantages. First, it guarantees a direct correspondence between (9) and (12). However, more importantly, it provides some direct control over the direction of $e(t)$. Let d_j be the jth eigenvalue of $\Sigma = \tau^T\tau$. Then it can be shown that only values of s such that $d_k \le s \le d_1$ are feasible solutions to (9) through (11). As s converges to d_1 we are forcing $e(t)$ equal to the first principal component function, an unsupervised approach often used in practice, which does not utilize the response to estimate $e(t)$. Similarly as s declines to d_k, $e(t)$ approaches the kth PC function. In general larger values of s encourage estimates for $e(t)$ that are similar to the larger PC functions while the opposite is true for smaller values of s.

For $M > 1$, once $\widehat{e}_1(t)$ and $\widehat{\boldsymbol{\gamma}}_1$ have been estimated their product can be subtracted from the response i.e. $\mathbf{Y}^*(t) = \mathbf{Y}(t) - \widehat{e}_1(t)\widehat{\boldsymbol{\gamma}}_1$. Then the FRAME algorithm is repeated using $\mathbf{Y}^*(t)$ as the response to iteratively estimate the remaining $e_m(t)$'s. In practice we implemented FRAME with $M = 1$ because we found that gave good answers on our HSX data.

4 Simulations

In this section we compare the performance of FRAME to several alternative linear and non-linear functional approaches in a series of simulation studies. We generated the responses from the model given by (7) and (8). The functional predictors, $X_{ij}(t)$, were simulated from a B-spline basis with two internal knots plus an error term,

$$X_{ij}(t_k) = \mathbf{b}(t_k)^T \boldsymbol{\theta}_{ij} + w_{ijk}, \quad w_{ijk} \sim N(0, \sigma_x^2), \quad \boldsymbol{\theta}_{ij} \sim N(0, \Theta),$$

where $\sigma_x = 0.05$ and each predictor was observed at 20 equally spaced time points, $0 = t_1, t_2, \ldots, t_{20} = 1$. For each simulation scenario we created $p = 4$ predictors. The basis coefficients, $\boldsymbol{\theta}_{ij}$, and the error terms, w_{ijk}, were all sampled independently from each other. The functional response was generated using $M = 1$ with $e_1(t) = \sin(\pi t/4)$ and $Y(t)$ sampled at 20 equally spaced points while the corresponding error term, $r_i(t_j)$, was sampled independently from a mean zero Gaussian distribution with $\sigma = 0.05$. In equation (8) the γ_i's were generated with $g_1(t)$ produced using a spline basis with random Gaussian coefficients and $g_j(t) = 0$ for $j = 2, 3, 4$, so the data contained one signal and three noise variables. The first coefficient function was generated from the same basis function used for the predictors i.e. $\beta_1(t) = \mathbf{b}(t)^T \boldsymbol{\eta}$.

Most functional regression methods utilize a functional principal components analysis (FPCA) decomposition of the predictors to form a low dimensional representation of the $X(t)$'s. The resulting PCA scores are then used as the predictors in the final regression model; the functional analogue of traditional principal components regression. In order to compare FRAME to the FPCA approach we generated a range of situations where the first principal component of $X(t)$ had varying predictive ability. In particular, let ω represent the proportion of variation in $\int X_{i1}(t)\beta_1(t)dt$ that is explained by the first principal component of $X_{i1}(t)$. Then, in order to facilitate comparisons with the FPCA approach we choose Θ and $\boldsymbol{\eta}$ in such a way that ω ranged from approximately 90% to 99%, depending on the simulation setup. In the $\omega \approx 99\%$ situation almost all the information about the response was contained in the first principal component of $X(t)$; an extremely favorable situation for the FPCA based methods. Alternatively, with $\omega \approx 90\%$ most, but not all, of the predictive information could be captured by the first principal component.

We compared *FRAME* to seven possible competitors. The first two methods, *Last Observation Linear* and *Last Observation Non-Linear*, both used just the last observed values of $X_j(t)$ as the predictor, i.e. $X_1(t_{20}), \ldots, X_4(t_{20})$. In both approaches, we estimated separate regressions for the response at each observed point, $Y(t_1), \ldots, Y(t_{20})$ using only the $X_j(t_{20})$'s as the predictors; a total of 20 separate regressions. The key difference between the two methods was that the former assumed a linear relationship while the latter modeled a non-linear fit to the response function. We implemented this comparison because for the HSX data it is common to assume an efficient market in which case one should form predictions using the last observed market value.

The next two methods, *Last Observation with PCA Linear* and *Last Observation with PCA Non-Linear* also used $X_1(t_{20}), \ldots, X_4(t_{20})$ as the predictors. However, instead of fitting separate regressions at each response time point, we computed the first functional principal component (FPC) of the $Y_i(t)$'s and formed a model to predict the corresponding FPCA scores. Hence, only one model needed to be fit. To form a prediction for the response function these methods multiple the estimated response FPCA score by the first principal component function. In contrast to the FRAME approach, these methods used the unsupervised FPCA procedure to estimate $e_1(t)$. Thus their accuracy depended on the correlation between the first FPC and $e_1(t)$ so we tested out different levels of correlation in our simulation settings.

The *FPCA Linear* and *FPCA Non-Linear* methods used a fully functional regression approach by computing the first FPCA for both the response function, and the predictor function. They then modeled either linear or non-linear relationships between the response and predictor FPCA scores.

All six of these methods used a penalized SCAD fitting procedure (Fan and Li, 2001) to automatically perform variable selection on the four predictors. In the linear settings we adopted the standard SCAD method. For the non-linear approaches we used a variant of SPAM (Ravikumar *et al.*, 2009) which implements a penalized non-linear additive procedure to perform variable selection.

The final comparison method, *FPCA FAR*, combined the FPCA approach with the FAR method of Fan and James (2011). FPCA FAR again computed the first FPCA scores for the response functions but then used these scores as the scalar response required to implement FAR. Hence, while the FPCA score for the response was still computed in an unsupervised fashion, the projection of the predictor, $X(t)$, was estimated using the supervised FAR method. As a result one might expect the performance of FPCA FAR to fall between the unsupervised FPCA Non-linear method and the fully supervised FRAME method.

We considered a total of four different simulation settings corresponding to $\omega = 90\%$ or 99% and correlations between the first FPCA of $Y(t)$ and $e_1(t)$ of approximately 0.7 or 0.9. For each setting we generated 50 different training data sets, each containing 100 observations and $p = 4$ functional predictors. We also generated a validation data set, with identical characteristics to the training data, which was used to select the tuning parameters for the various methods. Each of the eight comparison methods were fit to the data and false negative "FN" (fraction of signal variables incorrectly excluded), false positive "FP" (fraction of noise variables incorrectly included), and mean prediction errors "PE" were computed. The prediction errors corresponded to the mean squared error between the responses and predictions on a large test data set with $n = 1,000$ observations. The results are displayed in Tables 1 through 4. There are no false negative or false positive rates for the Last Observation method because different models were selected for each of the 20 regressions that were fit to the data.

	Last Obs.		Last Obs. with PCA		FPCA		FPCA	FRAME
	Lin	Non. Lin	Lin	Non Lin	Lin	Non Lin	FAR	
FN	NA	NA	0.840	0.880	0.000	0.000	0.000	0.000
SE(FN)	NA	NA	(0.052)	(0.046)	(0.000)	(0.000)	(0.000)	(0.000)
FP	NA	NA	0.073	0.187	0.253	0.667	0.120	0.167
SE(FP)	NA	NA	(0.024)	(0.046)	(0.047)	(0.062)	(0.037)	(0.064)
PE	3.901	1.671	1.363	1.360	1.086	1.217	0.645	0.672
SE(PE)	(0.003)	(0.019)	(0.011)	(0.010)	(0.044)	(0.009)	(0.003)	(0.006)

Table 1: $\omega = 99\%$, correlation between the first PCA of $Y(t)$ and the true direction $e_1(t)$ is approximately 0.9.

Table 1 corresponds to the most favorable situation for the FPCA methods where $\omega = 99\%$ and the first FPCA has a 0.9 correlation with $e_1(t)$. However, even in this setting the FPCA FAR and FRAME methods significantly outperform the other approaches, both in terms of model accuracy and prediction error. The FRAME and FPCA FAR methods give similar levels of accuracy though the latter approach has a slight advantage in this setting because there is such a high correlation between $e_1(t)$ and the first FPCA, which is used as the response for FPCA FAR.

Table 2 shows results when the first FPCA of the predictors explains a slightly lower fraction of the variation in the response. This change has almost no effect on the prediction errors for FRAME and FPCA FAR because they do not rely on a FPCA decomposition. However, most of the other methods exhibit a noticeable deterioration in performance.

	Last Obs.		Last Obs. with PCA		FPCA		FPCA	FRAME
	Lin	Non. Lin	Lin	Non Lin	Lin	Non Lin	FAR	
FN	NA	NA	0.380	0.540	0.000	0.000	0.000	0.000
SE(FN)	NA	NA	(0.069)	(0.071)	(0.000)	(0.000)	(0.000)	(0.000)
FP	NA	NA	0.100	0.173	0.233	0.753	0.253	0.113
SE(FP)	NA	NA	(0.027)	(0.043)	(0.045)	(0.054)	(0.048)	(0.061)
PE	4.000	1.777	1.483	1.486	1.077	1.353	0.653	0.677
SE(PE)	(0.005)	(0.018)	(0.011)	(0.010)	(0.033)	(0.010)	(0.005)	(0.004)

Table 2: $\omega = 90\%$, correlation between the first PCA of $Y(t)$ and the true direction $e_1(t)$ is approximately 0.9.

	Last Obs.		Last Obs. with PCA		FPCA		FPCA	FRAME
	Lin	Non. Lin	Lin	Non Lin	Lin	Non Lin	FAR	
FN	NA	NA	0.800	0.840	0.000	0.000	0.000	0.000
SE(FN)	NA	NA	(0.057)	(0.052)	(0.000)	(0.000)	(0.000)	(0.000)
FP	NA	NA	0.073	0.227	0.287	0.773	0.400	0.380
SE(FP)	NA	NA	(0.024)	(0.045)	(0.052)	(0.050)	(0.054)	(0.069)
PE	5.378	3.275	2.830	2.819	2.614	2.680	2.289	2.195
SE(PE)	(0.004)	(0.052)	(0.013)	(0.012)	(0.057)	(0.011)	(0.018)	(0.014)

Table 3: $\omega = 99\%$, correlation between the first PCA of $Y(t)$ and the true direction $e_1(t)$ is approximately 0.7.

Tables 3 and 4 show results where there is a lower correlation between $e_1(t)$ and the first FPCA of the response. In this situation we might expect the performance of the Last Observation with PCA, FPCA and FPCA FAR methods to deteriorate because their unsupervised method for estimating $e_1(t)$ is inefficient. Indeed we now see that FRAME, which uses a supervised method for estimating $e_1(t)$, is statistically superior to all other methods in terms of prediction error and has the second best model selection performance.

	Last Obs.		Last Obs. with PCA		FPCA		FPCA	FRAME
	Lin	Non. Lin	Lin	Non Lin	Lin	Non Lin	FAR	
FN	NA	NA	0.440	0.660	0.000	0.000	0.000	0.000
SE(FN)	NA	NA	(0.071)	(0.068)	(0.000)	(0.000)	(0.000)	(0.000)
FP	NA	NA	0.040	0.173	0.253	0.807	0.387	0.353
SE(FP)	NA	NA	(0.016)	(0.038)	(0.047)	(0.047)	(0.050)	(0.071)
PE	5.481	3.412	2.967	2.954	2.619	2.827	2.302	2.226
SE(PE)	(0.007)	(0.060)	(0.013)	(0.012)	(0.045)	(0.013)	(0.018)	(0.019)

Table 4: $\omega = 90\%$, correlation between the first PCA of $Y(t)$ and the true direction $e_1(t)$ is approximately 0.7.

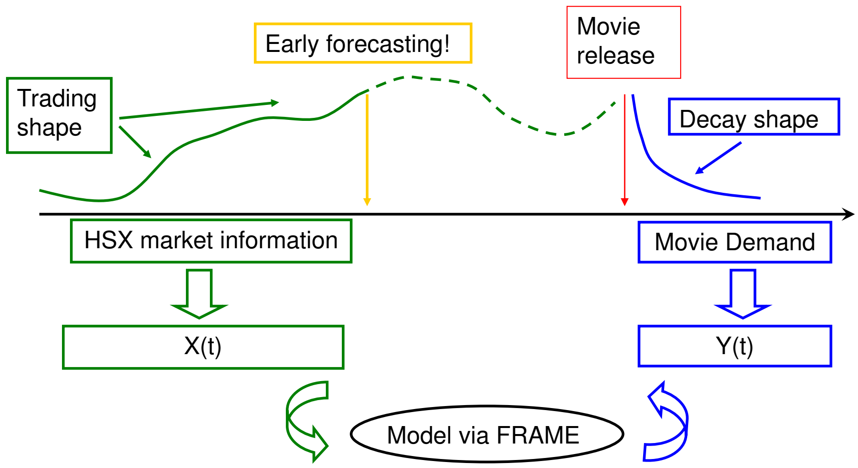

Figure 4: Illustration of our fully functional shape model.

5 Forecasting Demand Decay Rates

In this section we provide results from applying our FRAME approach to the HSX data. Figure 4 illustrates the prediction setup. For each movie we observe the HSX trading shapes (green line) between 52 and 10 weeks prior to the movie's release date. This curve forms our predictor X(t). We then use FRAME to form predictions of the box office revenue, Y(t), for the first 5 to 10 weeks after release (blue line). 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 prediction markets and movies' success.

5.1 Prediction Accuracy

We compare a number of functional and non-functional methods to predict the box office decay pattern for our 262 movies. Table 5 provides weekly mean absolute percentage errors (MAPE) between predicted and actual box office revenue for FRAME as well as six comparison methods. The MAPE scores were all calculated using ten-fold cross-validation to provide approximately unbiased estimates of the true test error rate. Since box office revenues drop to zero for several movies in later release-weeks, we computed trimmed means, averaging the middle 50% of movies. We have restricted attention to the first five weeks because more than 92% of all movie revenues occur in this time period (e.g. Einav, 2007). We fitted all models on the log revenues and then transformed the predictions back to the original scale to compute error rates since the log-scale

	Mean	Last Obs. Linear	Last Obs. Non-Linear	FPCA Linear	FPCA Non-Linear	FPCA FAR	FRAME
Week 1	1.312	1.114	0.693	1.089	0.896	0.822	0.569
Week 2	0.912	0.820	0.678	0.779	0.678	0.676	0.579
Week 3	0.872	0.805	0.658	0.775	0.685	0.681	0.584
Week 4	0.894	0.827	0.735	0.821	0.762	0.781	0.796
Week 5	0.909	0.838	0.842	0.837	0.883	0.859	0.916

Table 5: Ten-fold cross-validation trimmed mean absolute percentage error for FRAME and six alternative methods.

considerably improved prediction accuracy for all methods.

The first method in Table 5, *Mean*, corresponds to a simple estimator where we use the average of log revenue in the training set to predict (log) revenue in the holdout set. Note that we produce five such averages, one for each of the five revenue weeks (week 1 - 5). This approach serves as a baseline for the remaining methods since it does not make any use of the HSX data.

The next five methods, *Last Observation Linear, Last Observation Non-Linear, FPCA Linear, FPCA Non-Linear*, and *FPCA FAR*, all match their counterparts described in the simulation section. In particular the Last Observation methods both use, as the sole predictor, the HSX value recorded at ten weeks prior to the movie release. If the HSX market is truly efficient then all the relevant information would be captured by the last observed value, and forming predictions using the entire HSX curve would provide no further value; in which case we would expect these two approaches to generate the best results. The *FPCA Linear* and *FPCA Non-Linear* methods compute the first functional principal component (FPCA) for both the response function (movie returns), and the predictor function (HSX data). They then model the relationship between the predictor and response scores using the SCAD based penalization method. Finally, FPCA FAR again computes the first FPCA score for the response function but then uses this score as the scalar response required to implement FAR.

We compare these six approaches to our FRAME method detailed in Section 3. This approach not only produces a supervised estimate for the predictor projection, $\beta(t)$, but also a supervised estimate for the response projection, $e(t)$. The basis function used to model the response, $\mathbf{s}(t)$, is constructed using the first three FPCA of $Y_i(t)$. FRAME involves two tuning parameters, λ and s but since our data only contained a single functional predictor, so does not require any penalized variable selection, we set $\lambda = 0$. This avoids performing a detailed search over the parameters, which would likely overestimate the real world performance of FRAME. Recall that s must lie between the smallest and largest eigenvalues of Σ. Hence we set $s = d_k + a(d_1 - d_k)$ where $0 \leq a \leq 1$. Note that $a = 1$ and $a = 0$ correspond to $s = d_1$ and $s = d_k$ which respectively force the projection of the response to lie in the direction of the first or last FPCA. Hence, the FPCA FAR approach can be seen as a special case of FRAME with $a = 1$. Again, to avoid biasing our results we only test three values for a corresponding to $a = 0.05, a = 0.5$ and $a = 0.95$. Not surprisingly given our previous discussion, $a = 0.95$ gives similar results to FPCA FAR. More interestingly, we find that $a = 0.05$ gives significantly better results than either $a = 0.5$ or $a = 0.95$; suggesting that projecting the response in the first FPCA direction is suboptimal for this data. Hence, the results for FRAME in Table 5 correspond to $\lambda = 0$ and $a = 0.05$.

Table 5 shows that, not surprisingly, the error rates for all seven methods grow as we predict further away from the release date (i.e. further into the future). Only in Week 2, error rates dip slightly – which is curious – but they increase towards later weeks. In the early weeks, all six methods that incorporate HSX data improve considerably over the baseline (*Mean*), but this advantage diminishes for later weeks. During Weeks 1 through 4 the non-linear methods all outperform their linear counterparts. In particular FRAME shows large improvements over the other methods in the first 3 weeks which account for almost 80% of all movie box office revenues. From Week 5 on the signal becomes weak enough that none of the methods offer a large improvement over the *Mean* approach.

Figure 5 offers another way to view the performance of the various methods. In Figure 5a) we have plotted the cumulative percentage of movie revenue by week. As discussed earlier, most revenue occurs in the first few weeks where FRAME performs best. In Figure 5b) we have computed the cumulative revenue, by week, for each movie and tested the accuracy of our various approaches. Each curve corresponds to the cross validated absolute error between a movie's cumulative revenue, at each week, and the corresponding prediction for a given method, averaged over our 262 movies. The Last Observation Non-Linear, FPCA Non-Linear and FPCA FAR methods provide similar results and all offer significant improvements over the Mean approach. However, FRAME is clearly superior to all other methods in each of the ten weeks.

To further benchmark our functional model against alternate methods that are commonly used in the literature on movie demand forecasting (Sawhney and Eliashberg, 1996), Table 6 provides error rates for ten additional models. For each of these models, we estimate five separate weekly linear regressions, one for each of the five revenue weeks. Just as in Table 5, we again use ten-

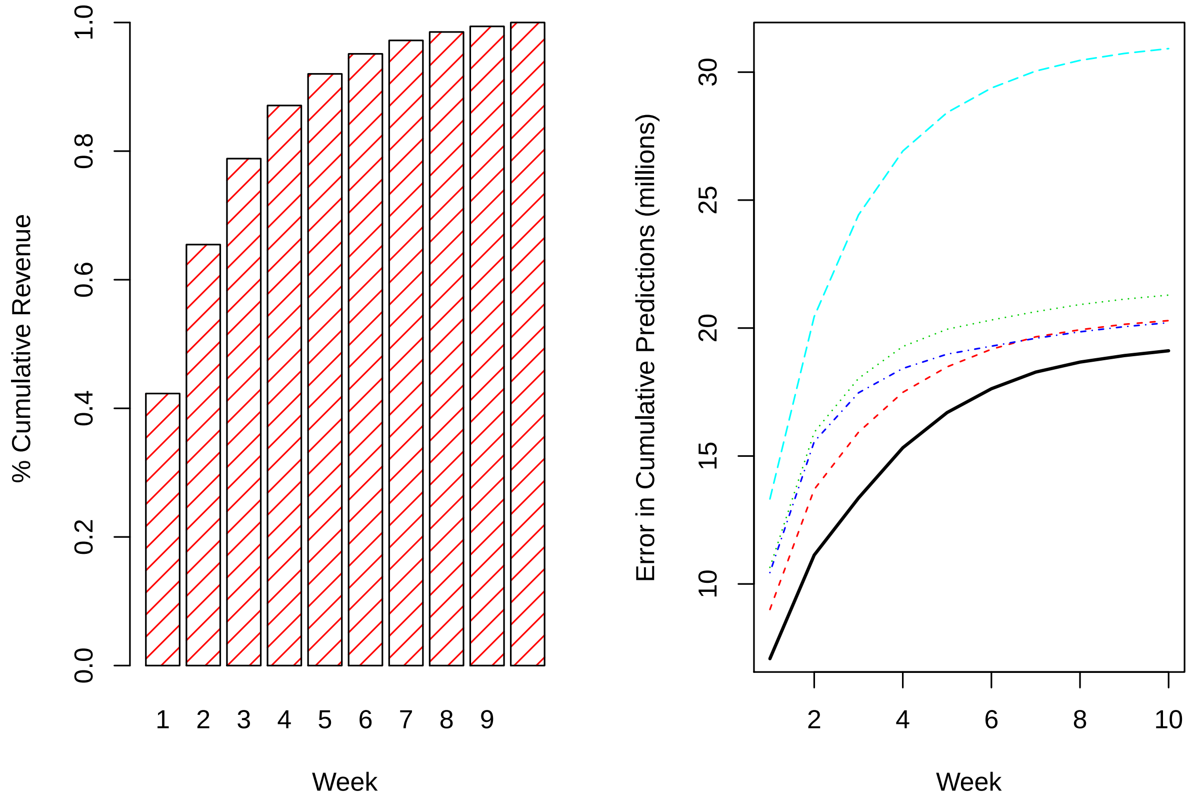

Figure 5: Left panel: Cumulative percentage of box office revenue by week. Right panel: Weekly mean absolute error in cumulative predictions for FRAME (black solid), Last Observation Non-Linear (red short dash), FPCA Non-Linear (blue dash dot), FPCA FAR (green dotted) and Mean (cyan long dash).

	Genre	Sequel	Budget	Rating	Run Time	Studio	B/G	B/Last	All	All/Last
Week 1	0.890	1.279	1.130	0.912	1.317	1.096	0.833	1.053	0.823	0.863
Week 2	0.900	0.909	0.852	0.808	0.887	0.884	0.815	0.810	0.812	0.778
Week 3	0.852	0.894	0.844	0.819	0.850	0.868	0.793	0.797	0.923	0.835
Week 4	0.898	0.902	0.850	0.882	0.886	0.940	0.851	0.814	0.926	0.865
Week 5	0.955	0.897	0.855	0.905	0.888	0.995	0.897	0.836	0.972	0.933

Table 6: Ten-fold cross-validation trimmed mean absolute percentage error for other alternative models.

fold cross-validation and report the trimmed mean absolute percentage errors between predicted and actual box office revenues. The first six models are based on movie features, respectively, genre (e.g. drama or comedy), sequel (yes/no), production budget (in dollars), MPAA rating, run time (in minutes), studios (e.g. Universal or 20th Century Fox). The seventh model is based on a combination of budget and genre (B/G), the two predictors that, overall, provide the highest prediction accuracy among the six movie features. The first seven models make no use of the HSX data at all. The eighth model includes budget and the last observed daily HSX value (ten weeks prior to release, as before). The ninth model uses all six movie features (All) and the tenth model uses the six movie features plus the last observed HSX value (All/Last).

We can see that, overall, all models have significantly higher prediction errors compared to the FRAME results from Table 5. Also note that augmenting the last observed HSX value with common movie features generally leads to a decrease in forecast accuracy; for example, compare All/Last from Table 6 with Last Observation Non-Linear from Table 5. We also investigated a model that includes both the HSX trading shapes as well as the movie features. The results (not shown here) are somewhat worse compared to the functional model without the features. This strongly suggests that HSX price paths incorporate all the information about a movie's features and characteristics, even ten weeks prior to release.

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 5 and 6. To that end, we investigate the relationship between FRAME's prediction error (i.e. the mean absolute percentage error between FRAME's predicted revenue values and their true values) 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 prediction error and the lowest error of either FPCA Non-linear or Last Observation Non-linear) is associated with film characteristics. Tables 7 and 8 show the final regression models obtained from variable selection via stepwise regression (using a combination of forward selection and backward elimination).

Table 7 shows that FRAME performs well (i.e. has a low prediction error) for movies that are sequels, rated below R, released by a major studio such as Paramount, Warner Brothers, Universal, or Twentieth Century Fox, and reviewed by a large number of critics. Intuitively, these results

Name	Coef	StErr	T-Value
Sequel	-0.033	0.014	-2.437
Horror-Scifi-Suspense	0.024	0.011	2.239
Documentary-Musical	0.091	0.033	2.737
Rating Below R	-0.017	0.010	-1.667
Runtime	0.001	0.000	4.500
Major Studio	-0.036	0.009	-4.033
Critics Volume	-0.001	0.000	-12.000
Critics Valence	0.011	0.004	3.167
User Volume	0.001	0.000	3.500

Table 7: Stepwise regression of FRAME's prediction error on film characteristics.

Name	Coef	StErr	T-Value
Animated	0.014	0.008	1.741
Horror-Scifi-Suspense	0.011	0.004	2.409
Documentary-Musical	0.069	0.014	4.971
Critics Volume	-0.001	0.000	-2.000
User Valence	0.004	0.002	2.563

Table 8: Stepwise regression of the difference between FRAME's prediction error and the lowest error of either FPCA Non-linear or Last Observation Non-linear on film characteristics.

suggest that FRAME performs especially well for movies that enjoy a greater capability for creating pre-release "buzz". Consider, sequels build upon the success of their predecessors; films released by major studios enjoy significant advertising and publicity before opening; those with lower MPAA ratings, e.g. PG and PG-13, appeal to wider audiences; and greater attention from the critics, due to, for instance, a film's quality or controversies, 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 prices. FRAME seems to be capable of capturing the dynamics of such buzz and translating it into accurate predictions.

Figure 6 shows the six movies for which FRAME predicts the best in terms of cross-validated error rate over the first five weeks. These six movies were all released by major studios with the exception of *Laws of Attraction*. Moreover, four of them, including *Catwoman*, *Laws of Attraction*, *Van Helsing*, and *King Arthur* are rated below *R*. And all attract a large number of critics' reviews, with four of them in the top 20th percentile. While none of them is a sequel, they are not far down the list. For example, FRAME provides excellent predictions for sequels like *XXX: State of the Union*, *Ocean's Twelve* or *Scooby-Doo 2*. By contrast, FRAME predicts the least accurately for the following movies: *Kaena: The Prophecy*, *It's All About Love*, and *Eulogy*. None of these movies was a sequel or produced by a major studio. Only *Kaena: The Prophecy* has a below-*R* rating; and

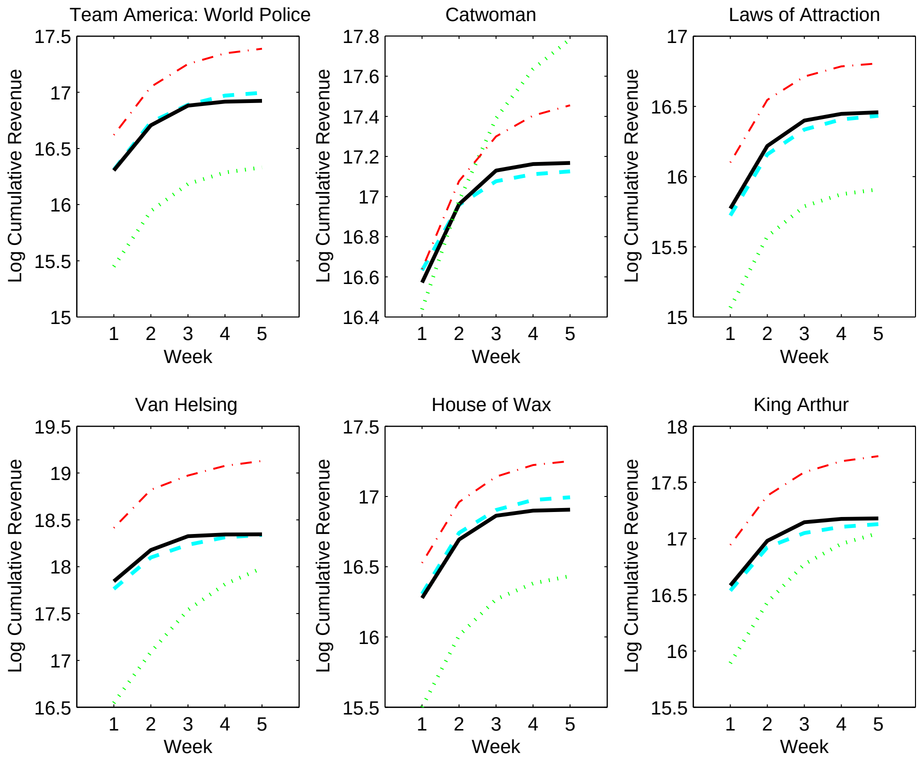

Figure 6: Top 6 movies with the smallest FRAME prediction error: The black solid lines correspond to FRAME's prediction; the blue dashed lines show the corresponding true values. The two closest competitors are given by the dotted lines (FPCA Non-linear) and the dash-dotted lines (Last Observation Non-linear), respectively.

the volumes of critics' reviews for all of three movies fall in the bottom 5th percentile.

It is possible that movies with some of the above identified characteristics – sequels, low MPAA ratings, major studio releases, and better critics' reviews – are easier to predict in general by any method, not only by FRAME. Indeed, Table 8 shows that FRAME does not have a statistically significant advantage over FPCA Non-linear or Last Observation Non-linear in predicting sequels, below-R-rated, and major studio films. Nonetheless, FRAME continues to stand out for those films catching critics' eyes, suggesting a distinct ability to incorporate information potentially not captured by alternative methods, such as the critics' 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 decay rates. However, while non-linear functional regression methods can result in good predictions, one downside is that because both model-input (HSX trading paths) as well as model-output (box office demand decay rates) arrive in the form of functions it is hard to understand the relationship between response and predictor.

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 $Y(t)$ together. The idea is similar to the "partial dependence plots" described in Hastie et al (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. 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, respectively); their only difference is how they get from the start to the end: The middle curve (solid black line) grows at a linear rate; the upper and lower curves (dotted green and dashed red 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 towards release (exponential). Movies that grow fast early may be sequels who often enjoy early awareness from their predecessors; on the other hand, slow early growth with faster increases towards the release date may be a trademark of so-called "sleeper" movies. All three curves end at exactly the same HSX market value, so any difference in estimated box office revenue is only due to their difference in shapes.

The top right panel shows the result: Both the linear (black solid line) as well as the logarithmic (green dotted line) HSX price curves result in similar decay of box office revenue, and both significantly out-perform the exponential curve (red dashed line). In fact, while both linear and logarithmic HSX price growth result in very high box office revenues during the first few weeks, exponential price growth leads to very low revenues (which stay low). Comparing linear and logarithmic price growth with one another, we notice that the logarithmic shape results in a slight revenue advantage during the first few weeks.

What do these findings imply? Recall that all three HSX price curves start and end at the same value (0 and 100, respectively), so all observed differences are only with respect to their shape. This suggests that shapes matter enormously in prediction markets. It also suggests that more "buzz"

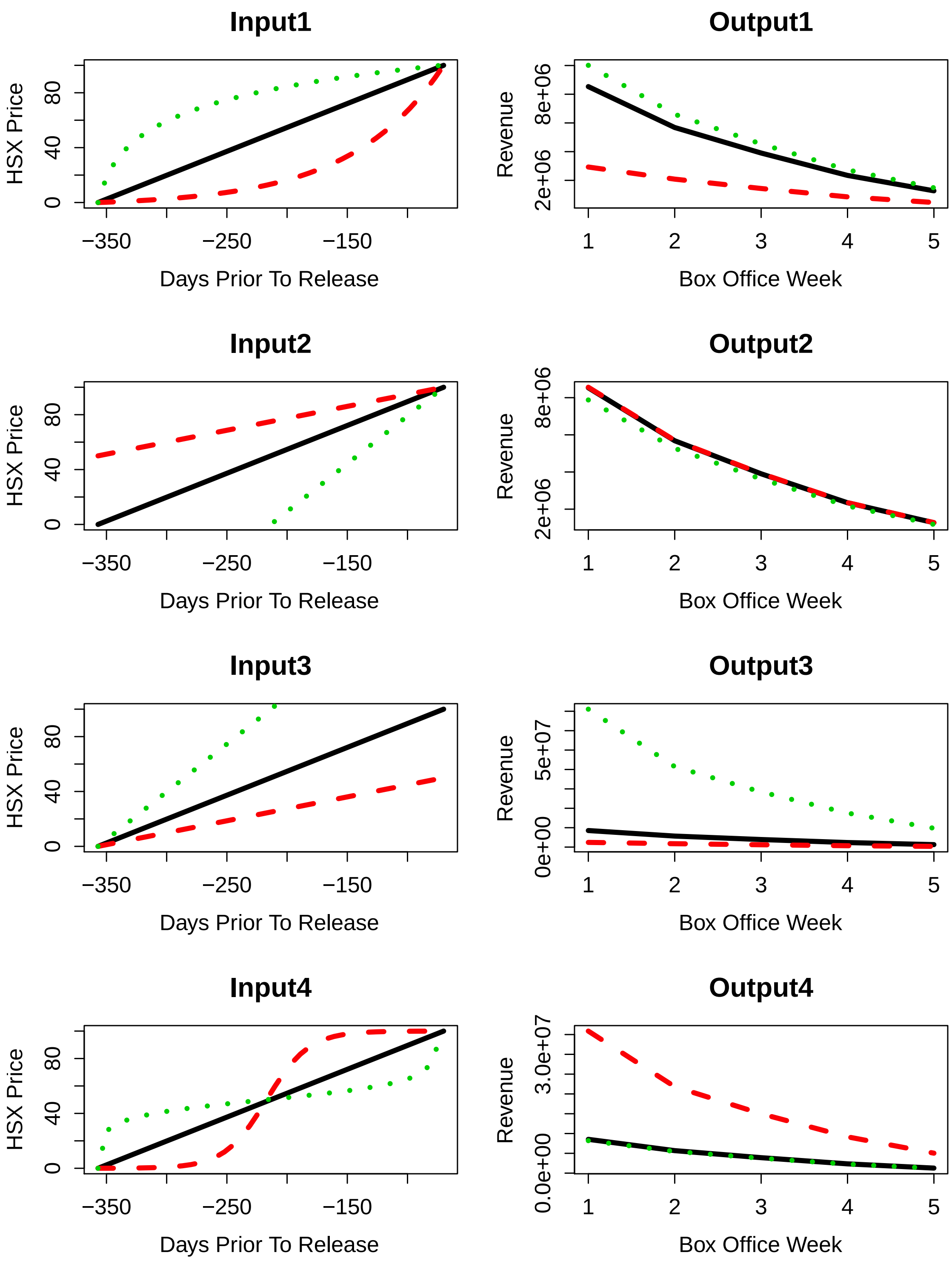

Figure 7: Dependence plots for different input shapes.

early on (i.e. the logarithmic shape) has much more impact on overall revenue compared to a "last moment hype" during the very last days (i.e. the exponential shape).

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 towards a common HSX value (100, in this case) while the input curves in the third row are diverging away from their initial HSX value. The effect on box office revenues can again be seen in the right panel: While the diverging input curves result in very different box office decay curves, HSX price curves that converge towards the same value all result in approximately the same decay curve.

The case of diverging curves suggests that "size matters": the larger the HSX value (and the faster it grows), the larger is the corresponding box office revenue (and the slower is its decay). The converging case emphasizes the effect of the shapes: Like scenario 1, all HSX price curves end at the same value; unlike scenario 1, they all have the same shape. 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 (red dashed line) and an inverse-S shape (green dotted line). Notice that while the inverse-S shape is almost constant with only very small spurts at the very beginning and at the very end, the S shape features a very prominent period of strong and steady growth, one that is similar to (but exceeds in magnitude) the logarithmic shape from the first scenario. The result supports our previous findings: While the almost-constant growth of the inverse-S shape produces very low box office revenues, the strong dynamics of the S-shape lead to revenues that exceed those of the logarithmic shape in the first scenario. All-in-all, this suggests that, in addition to the magnitude of the HSX price curve (which captures level-differences in perception about a movie), its shape is capable of capturing information about buzz. It also appears that the timing of the buzz is an important predictor of box office success.

6 Conclusion

This paper makes three significant contributions. First, 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). Second, we develop a new non-linear regression approach, FRAME, which is capable of forming predictions on a functional response given a large number of functional predictors. Our results on both the HSX data as well as extensive simulations demonstrates that FRAME is capable of providing a considerable improvement in prediction accuracy relative to a host of competing methods. Finally, we make our approach practical for inference purposes by developing dependence plots to illustrate the relationship between input and output curves.

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 afterwards, the studio may consider nationwide release, as opposed to limited or platform (i.e. from initial limited release to nationwide release later on) release strategies, 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, et al. 2006), and that these channels represent five times more revenues than 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 history. 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 market 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 cultur appeal, and strong influence of hype on demand. Such research would be made possible by the increasing availability of data from VSMs for, e.g., books (MediaPredict), music (HSX), TV shows (Inkling), and video games (SimExchange).

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. *Journal of Computational Biology* **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(3):S82–S93.

Einav, L. (2007). Seasonality in the u.s. motion picture industry. *Rand Journal of Economics*, 38(1):127-145.

Eliashberg, J., Elberse, A., and Leenders, M. (2006). The motion picture industry: Critical issues in practice, current research, and new research directions. *Marketing Science*, 25(6): 638–661.

Eliashberg, J. and Shugan, S. M. (1997). Film critics: Influencers or Predictors? *Journal of Marketing*, 61(2): 68–78.

Fan, J. and Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association* **96**, 1348–1360.

Fan, Y. and James, G. (2012). Functional Additive Regression. *Under Review* .

Ferraty, F. and Vieu, P. (2002). The functional nonparametric model and applications to spectrometric data. *Computational Statistics* **17**, 545–564.

Ferraty, F. and Vieu, P. (2003). Curves discrimination: a nonparametric functional approach. *Computational Statistics and Data Analysis* **44**, 161–173.

Foutz, N. and Jank, W. (2009). Pre-release demand forecasting for motion pictures using functional shape analysis of virtual stock markets. *Marketing Science*, Volume 29, Issue 3, Pages 568–579.

Friedman, J. and Hastie, T. and Tibshirani, R. (2000). Additive logistic regression: A statistical view of boosting. *Annals of Statistics*, 28:337-374.

Gasser, T., Mller, H. G., Khler, W., Molinari, L., and Prader, A. (1984). Nonparametric regression analysis of growth curves (Corr: V12 p1588). *The Annals of Statistics* **12**, 210–229.

Gervini, D. and Gasser, T. (2005). Nonparametric maximum likelihood estimation of the structural mean of a sample of curves. *Biometrika* **92**, 801–820.

Hastie, T. J. and Tibshirani, R. J. (1990). *Generalized Additive Models.* Chapman and Hall.

Hastie, T. J., Tibshirani, R. J. and Friedman, J. (2001). *The Elements of Statistical Learning.* Springer.

James, G. M. and Hastie, T. J. (2001). Functional linear discriminant analysis for irregularly sampled curves. *Journal of the Royal Statistical Society, Series B* **63**, 533–550.

James, G. M., Hastie, T. J., and Sugar, C. A. (2000). Principal component models for sparse functional data. *Biometrika* **87**, 587–602.

James, G. M. and Silverman, B. W. (2005). Functional adaptive model estimation. *Journal of the American Statistical Association* **100**, 565–576.

James, G. M. and Sugar, C. A. (2003). Clustering for sparsely sampled functional data. *Journal of the American Statistical Association* **98**, 397–408.

Jank, W. and Shmueli, G. (2006). Functional data analysis in electronic commerce research. *Statistical Science*, 21:155–166.

Kneip, A. and Gasser, T. (1992). Statistical tools to analyze data representing a sample of curves. *Annals of Statistics* **20**, 1266–1305.

Kneip, A., Li, X., MacGibbon, K. B., and Ramsay, J. O. (2000). Curve registration by local regression. *The Canadian Journal of Statistics* **28**, 1, 19–29.

Muller, H. and Yao, F. (2008). Functional Additive Models. *Journal of the American Statistical Association.* To appear.

Ramsay, J. O. (1998). Estimating smooth monotone functions. *Journal of the Royal Statistical Society B*, 60(2):365–375.

Ramsay, J. O. and Li, X. (1998). Curve registration. *Journal of the Royal Statistical Society, B.* **60**, 351–363.

Ramsay, J. O. and Silverman, B. W. (2005). *Functional Data Analysis (Second Ed.).* Springer-Verlag, New York.

Ravikumar, P., Lafferty, J., Liu, H., and Wasserman, L. (2009). Sparse additive models. *Journal of the Royal Statistical Society, B.* **71**, 1009–1030.

Rice, J. A. and Silverman, B. W. (1991). Estimating the mean and covariance structure non-parametrically when the data are curves. *Journal of the Royal Statistical Society, Ser. B* **53**, 233–243.

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

Rønn, B. B. (2001). Nonparametric maximum likelihood estimation for shifted curves. *Journal of the Royal Statistical Society, Series B, Methodological* **63**, 2, 243–259.

Sakoe, H. and Chiba, S. (1978). Dynamic programming algorithm optimization for spoken word recognition. *IEEE Transactions on Acoustics, Speech, and Signal Processing* **26**, 43–49.

Sawhney, M. S. and Eliashberg, J. (1996). A parsimonious model for forecasting gross box office revenues of motion pictures. *Marketing Science*, 152:113-131.

Shao, J. (2003). *Mathematical Statistics.* Springer-Verlag New York, Inc., 2nd edn.

Silverman, B. W. (1995). Incorporating parametric effects into functional principal components analysis. *Journal of the Royal Statistical Society, Sec. B* **57**, 673–689.

Spann, M. and Skiera, B. (2003). Internet-based virtual stock markets for business forecasting. *Management Science*, 49(10):1310–1326.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society, Series B* **58**, 267–288.

Tuddenham, R. D. and Snyder, M. M. (1954). Physical growth of California boys and girls from birth to eighteen years. *University of California Publications in Child Development* **1**, 183–364.

Yao, F., Muller, H. and Wang, J. (2005). Functional Linear Regression Analysis For Longitudinal Data. *Annals of Statistics*, 33: 2873-2903.

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., and Jank, W., and Shmueli, G. (2010). Real-Time Forecasting of Online Auctions via Functional K-Nearest Neighbors. *International Journal of Forecasting* **26**, 666– 683.