ABSTRACT

Title of dissertation: FUNCTIONAL PRINCIPAL COMPONENT ANALYSIS WITH APPLICATION TO VIEWERSHIP OF MOTION PICTURES

Yue Tian, Doctor of Philosophy, 2014

Dissertation directed by: Professor Paul J. Smith
Department of Mathematics
University of Maryland, College Park

Anderson Professor Wolfgang S. Jank
Information Systems Decision Sciences Department
College Of Business, University of South Florida

Principal Component Analysis (PCA) is one widely used data processing technique in application, especially for dimensionality reduction. Functional Principal Component Analysis (fPCA) is a generalization of ordinary PCA, which focuses on a sample of functional observations and projects the original functional curves to a new space of orthogonal dimensions to capture the primary features of original functional curves. While, fPCA suffers from two potential error sources. One error source is originated from truncation when we approximate the functional subject's expansion; The other stems from estimation when we estimate the principal components from the sample. We first introduce a generalized functional linear regression model and propose it in the Quasi-likelihood setting. Asymptotic inference of the proposed functional regression model is developed.

We also utilize the proposed model to help marketing operational decision

process by analyzing viewership of motion pictures. We start with discussing customer reviews effect on movie box office sales. We use the functional regression model with function interactions to measure the effect of Word-of-Mouth on movie box office sales. One main challenge of modeling with functional interactions is the interpretation of model estimate results. We demonstrate one method to help us get important insights from model results by plotting and controlling a re-labbeld 3-D plot.

Apart from movie performance in theater, we also employ functional regression model to predict movie pre-release demand in Video-on-Demand (VOD) channel. As its growing popularity, VOD market attracts much attention in marketing research. We analyze the prediction accuracy of our proposed functional regression model with spatial components and find that our proposed model gives us the best predictive accuracy.

In summary, the dissertation develops asymptotic properties of a generalized functional linear regression model, and applies the proposed model in analyzing viewership of motion picture both in theater and Video-on-Demand channels. The proposed model not only advances our understanding of motion picture demand, but also helps optimize business decision making process.

FUNCTIONAL PRINCIPAL COMPONENT ANALYSIS WITH APPLICATION TO VIEWERSHIP OF MOTION PICTURES

by

Yue Tian

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2014

Advisory Committee:
Professor Paul J. Smith, Chair/Advisor
Professor Wolfgang S. Jank, Co-Advisor
Professor Benjamin Kedem
Professor Yuan Liao
Professor Michel Wedel

To my parents, and my family!

Acknowledgments

If someone would asked me what is my purpose of Ph.D. study, I would answer, without the slightest hesitation, to leave. The journey is about to come to an end and it is about time to leave, together with my joyful memory of all the people who motivated and helped me along the journey.

First and foremost, I would like to thank my advisors, Drs. Paul Smith and Wolfgang Jank for their guidance, support, and encouragement on my research throughout all these years. I am grateful to Dr. Paul Smith, for his patience and understanding in all aspects in my graduate study. Dr. Smith is the first person I talked with in our department, when he was the director of statistic program. I still remember my nervousness and embarrassment at that time, as I could not even say complete sentences in English. His kindness relieved my overwrought nerves, just like what he does while directing my research. Dr. Smith has given me the hope and encouragement to finish my Ph.D. thesis.

I feel so very luck to have met Dr. Jank and have the opportunity to work with him as a Ph.D. student. I started as a total rookie in applied statistical research. Dr. Jank has introduced me to this fantastic and challenging field and has offered me valuable skills in applied statistical research, competitive programming skills, and conference presentation skills, all of which make the recipe for success in my area of specialization.

I own my deepest thanks to my family. You give me the motivation to be stronger, as daughter, mother and wife, most importantly, as a person. I will con-

tinue this journal, as a powerful person.

Table of Contents

List of Figures

Chapter 1: Introduction

1.1 General Problem

The movie industry is immense business with high profile and highly variable revenue trend. The vast and growing access of Internet and social networking sites have motivated the growth in the importance of online word of mouth to movie performance. Meanwhile, Video-on-Demand (VOD) has been the subject of intense interest both in research and commercial sectors due to its convenience.

Our focus in this thesis is on measuring the impact of WOM on movie box office viewership and predicting movie demand in Video-on-Demand channel.

For the online WOM, we follow previous research and investigate three main measurements of user reviews: valence, volume and dispersion. We analyze the impact of these three WOM measurements on future box office sale of movies. In this case, we would like to study the relationship between historical information of WOM over time and movie box office sales in the near future. We use weekly movie reviews (valence, volume and dispersion) collected from Yahoo! Movies website. We also control for some other possible box office drivers and want to build a model to help us analyze whether and how customer reviews affect future box office sales.

For the VOD forecast, we concentrate on studying geographic difference of movie demand in this channel, with movie features. For this problem, we want to have an accurate forecast of movie pre-release demand trend across different geographic locations. That is to say that we need to use static movie features to predict future movie performance over time for each geographic location.

To address these two problems, we utilize functional Principal Component Analysis and functional quasi-likelihood model to help us manipulate the data and build statistical models.

1.2 Challenges of the Model

Both customer reviews (valence, volume and dispersion) and movie demand in VOD can be considered as functional observation, which change over time. Functional Data Analysis (FDA) focuses on a sample of *functional observations*, e.g. curves, and treats the observed curves as the units of observation. Functional Principal Component Analysis (fPCA) is used to find the dominant modes of variation in the data, usually after subtracting the mean from each functional observation.

We let Y denote the response and write

$$Y = g\left[\int_T \beta(t)X(t)dw(t)\right] + e.$$

Here $\beta(t)$ is an unknown parameter function and the observed covariate process is $X(t)$, where t is time and $t \in T$. We can have the functional principal component expansion and its corresponding parameter function as:

$$X(t) = \sum_{j=1}^{\infty} \epsilon_j \psi_j(t)$$

$$\beta(t) = \sum_{j=1}^{\infty} \beta_j \psi_j(t)$$

where $\psi_j(t)$'s and ϵ_j are the principal components of $X(t)$ and the corresponding principal component scores respectively.

Two potential error sources are embedded in functional principal component analysis. One error source is originated from truncation when we approximate the random processes expansion; The other error source stems from estimation when we approximate the principal components from the sample data. A generalized functional linear regression model is proposed in the quasi-likelihood setting. We develop asymptotic inference for the proposed functional regression models.

Chapter 2: Functional Quasi-likelihood Model

2.1 Functional Data Analysis

Functional Data Analysis (FDA) focuses on a sample of *functional observations*, such as online virtual stock market's history (Foutz and Jank [2010]), online auction price (Wang, Jank and Shmueli [2008]), market penetration (Sood, James and Tellis [2009]). This is in contrast to classical statistics where the focus is a set of discrete data vectors. The method of FDA was introduced by Rao [1958] for growth curves. Many theoretical properties have been developed by Ramsay and Silverman [1997] and Silverman [1996].

Recently, many classical statistical models have been generalized to the functional structure. James, Hastie and Sugar [2000] proposed functional principal components analysis for sparsely sampled curves. In practice, curves may be measured at an irregular and sparse set of time points which is even widely different across individuals. James, Hastie and Sugar [2000] used functional principal component analysis, which will be discussed in the following section, to address this issue. The case of irregular grids was also studied by Staniswalis and Lee [1998].

More recent research includes curve clustering and classification (James and Sugar [2003], Tarpey and Kinateder [2003], and James and Hastie [2001]), functional regression (Cuevas Febrero and Fraiman [2002]), functional generalized linear models (James [2002]), functional ANOVA (Guo [2002]) and time series analysis of functional data (Aguilera, Ocana and Valderrama [1999]). While this list is far from complete, it exemplifies some of the current methodological improvement in this emerging field.

Statistical models for functional data may resemble those for conventional multivariate data, for instance, principal component analysis and generalized linear models. In this paper, we use the generalized functional method, functional principal component analysis (fPCA), to model functional data. Before we move to the functional framework of PCA, we first discuss about conventional PCA first.

2.2 Principal Component Analysis

Principal Component Analysis (PCA) is an exceedingly popular technique for dimensionality reduction with minimal loss of information. The idea of PCA appeared over 100 years ago (Pearson [1901]). After its invention by Karl Pearson, this method has been used repeatedly across many different areas.

A Principal Component Analysis provides a way of characterizing covariance structure that can be more informative and efficient via linear combinations of the original variables with restrictions.

Let $X_i, i = 1, \cdots, m$, be *i.i.d.* p-dimensional vectors with $E(X_i) = \mu$ and Variance Covariance matrix Σ. The eigenvectors of Σ, $\psi_j, j = 1, \cdots, m$, are the principal components of X.

In fact, the full set of PCs, $\psi_j, j = 1, \cdots, m$ are linear transformation of the original data, and thus contain the same amount of information (variation) as the original $X_i, i = 1, \cdots, m$, but the structure of newly defined data (PCs) is important for dimension reduction. Meanwhile, the resulting uncorrelated principal components solve the potential collinearity among the original data.

2.3 Functional Principal Component Analysis

In the functional context, the counterparts of the original m-diminutional vectors $x_i = (x_{i1}, \ldots, x_{ip})^T$ are functional data $x_i(t), t \in T$. The discrete index has been replaced by continuous index t. Let $\mu(t) = E[X(t)]$.

The principal component expansion of $X(t) - \mu(t)$ can be constructed by the covariance function

$$K(s,t) = E[X(s) - \mu(s)][X(t) - \mu(t)]$$

where K is assumed to be square integrable on the space $L_2(T)$.

Let $\theta_1 \geq \theta_2 \geq \ldots \geq 0$ be the eigenvalues of K, and the corresponding orthonormal eigenfunctions $\psi_1(t), \psi_2(t), \ldots$,then

$$K(s,t) = \sum_{j=1}^{\infty} \theta_j \psi_j(s) \psi_j(t)$$

The functional principal component expansion of $X_i(t)$ (Karhunen-Loève expansion) is given by

$$X(t) - \mu(t) = \sum_{j=1}^{\infty} \epsilon_j \psi_j(t)$$

where the random variables $\epsilon_1, \epsilon_2, \ldots$, given by $\epsilon_j = \int_T X(t)\psi_j(t)dt$, are the principal component scores of $X(t)$ corresponding to the jth principal component $\psi_j(t)$. The orthogonal $\psi_j(t)$ and $\psi_k(t)$, $j \neq k$, implies $\epsilon_j's$ are uncorrelated. Meanwhile, $E(\epsilon_j) = 0$, and $\theta_j = E(\epsilon_j^2)$, $\sum_j \theta_j = \int_T E[X(t) - \mu(t)]^2 dt < \infty$.

Functional principal component analysis allows finite dimensional analysis of a problem that is intrinsically infinite dimensional, as functional objects are assumed to be smooth functions (Ramsay and Silverman [1997]). This method has been widely used in modeling continuous functional objects, for instance, price curve and derivative curves of eBay auctions (Hyde, Moore and Hodge [2004]), box office revenue in virtual stock markets (Jank and Shmueli [2007], Foutz and Jank [2010]), and functional magnetic resonance imaging (fMRI) (Viviani, Gron and Spitzer [2005]). Numerous models have been proposed by researchers for methodologically modeling functional objects. Hall and Horowitz [2007] discussed asymptotic properties of functional linear regression with functional covariates. The case where both response and predictor are functional is studied by Moyeed and Diggle [1994] and Zeger and Diggle [1994].

In traditional setting, we usually assume the sample of random functions is observed precisely, which also differentiates it from longitudinal data analysis. Hall et cl. [2006] proposed a semi-parametric method for the case when random functions are contaminated by noise, and even only a few observations are available for each function. Statistical smoothing techniques successively exploit the high-dimensional data and root-n consistent estimates are derived even in the presence of noise. fPCA is broadly utilized to model functional subjects, by approximating functional model with a series of models where the number of predictors is truncated. Muller and Stadtmuller [2005] proposed a generalized functional linear regression model considering this truncation error when applying fPCA in practical data. Asymptotic inference on the proposed model is analyzed in the thesis.

In practice, fPCA is a data-driven method. The principal components are estimated from the sample data, instead of observed, and consequently will change as sample size changes. An estimation error for the principal component is introduced in application of fPCA. In this work, we focus on incorporating functional objects in the framework of quasi-likelihood model, considering both truncation error and estimation error.

2.4 Generalized Functional Linear Regression Model and Functional Quasi-Likelihood Model

The data we observe for the ith subject are $\{X_i(t), t \in T, Y_i, i = 1, \ldots, n\}$. We assume that these data from an *i.i.d.* sample. Here the response variable Y_i is a real-valued random variable and $X_i(t), t \in T$ is a square integrable stochastic process on a real compact set T. The mean function is $\mu(t) = E[X(t)]$.

The linear predictor η given by

$$\eta = \int_T \beta(t)X(t)dt$$

If we write

$$\eta_i = \alpha + \int_T \beta(t)X_i(t)dw(t)$$
$$E[Y_i|X(t), t \in T] = \mu_i = g(\eta_i)$$

then $g(\cdot)$ will be called the link function, which is monotone, twice continuously differentiable with bounded derivatives and is thus invertible. This function links the expectation of Y_i with our random covariate function $X_i(t)$. We should note that the link function g here is the inverse link in McCullagh [1983]. We also assume $Var(Y_i|X_i(t), t \in T) = \sigma^2(\mu_i) = \tilde{\sigma}^2(\eta_i)$.

We consider generalized functional linear regression model

$$Y_i = g(\alpha + \int_T \beta(t)X_i(t)dt) + e_i, i = 1, \ldots, n,$$

where $E(e_i|X(t), t \in T) = 0$ and $Var(e_i|X(t), t \in T) = \sigma^2(\mu_i) = \tilde{\sigma}^2(\eta_i)$.

This generalized functional linear regression model is specified by a parameter function $\beta(\cdot)$, which is assumed to be square integrable on T, in addition to the link function $g(\cdot)$ and $\sigma^2(\cdot)$ ($\tilde{\sigma}^2(\cdot)$) are assumed known and satisfy previous conditions. Muller and Stadtmuller [2005] discussed the situation where the link and variance functions are unknown and are estimated nonparametrically from the data.

Because of inclusion of intercept, we have $E[X(t)] = 0$, for all t in T. Let $E(\tilde{\sigma}^2(\eta)) = \sigma^2$. Error terms e_i are *i.i.d.* with

$$\begin{aligned} E(e) &= E[E(e|X(t), t \in T)] = 0 \\ Var(e) &= Var(E(e|X(t), t \in T)) + E[Var(e|X(t), t \in T)] \end{aligned}$$

$$= \quad 0 + E[\tilde{\sigma}^2(\eta)] = \sigma^2.$$

Let $\psi_i(t), j = 1, 2, \ldots$ be the principal components of $X(t), t \in T$. Then ψ_j forms an orthonormal basis of the function space $L^2(dw)$. That is $\int_T \psi_j(t)\psi_k(t)dw(t) = \delta_{jk}$. Then the principal component expansions of the functional predictor $X(t)$ and the parameter function $\beta(t)$ are

$$\begin{aligned} X(t) &= \sum_{j=1}^{\infty} \epsilon_j \psi_j(t) \\ \beta(t) &= \sum_{j=1}^{\infty} \beta_j \psi_j(t) \end{aligned}$$

with

$$\epsilon_j = \int_T X(t)\psi_j(t)dw(t)$$

$$\beta_j = \int_T \beta(t)\psi_j(t)dw(t)$$

We note that ϵ_j's are random variables, while β_j's are coefficients. Some statistical properties of principal component ϵ_j's and coefficients β_j's are summarized in:

Fact 1. *(1)* $E(\epsilon_j) = 0$

(2) $\sum_{j=1}^{\infty} \beta_j^2 < \infty$

(3) Let $\sigma_j^2 = E(\epsilon_j^2)$, $\sum_j \sigma_j^2 = \sum_j E(\epsilon_j^2) < \infty$

Proof. (1).

$$\begin{aligned} E(\epsilon_j) &= E(\int_T X(t)\rho_j(t)dw(t)) \\ &= \int_\omega \int_T X(t)(\omega)\rho_j(t)dw(t)f_X(\omega)d\omega \end{aligned}$$

$$
\begin{aligned}
&= \int_T \rho_j(t) \int_\omega X(t)(\omega) f_X(\omega) d\omega dt \\
&= 0
\end{aligned}
$$

(2).

$$\beta(t) = \sum_j \beta_j \rho_j(t)$$

$$
\begin{aligned}
\int \beta^2(t) dw(t) &= \int [\sum_j \beta_j \rho_j(t)]^2 dw(t) \\
&= \int \sum_j \beta_j^2 \rho_j(t) + \sum_{i \neq j} 2\beta_i \rho_i(t) \beta_j \rho_j(t) dw(t) \\
&= \sum_j \int \beta_j^2 \rho_j(t) dw(t) + \sum_{i \neq j} 2\beta_i \beta_j \int \rho_i(t) \rho_j(t) dw(t) \\
&= \sum_j \beta_j^2 < \infty
\end{aligned}
$$

(3).

$$
\begin{aligned}
\int E[X^2(t)] dw(t) &= \int E[\sum_j \epsilon_j \rho_j(t)]^2 dw(t) \\
&= \int E[\sum_j \epsilon_j^2 \rho_j^2(t) + \sum_{i \neq j} \epsilon_i \epsilon_j \rho_i(t) \rho_j(t)] dw(t) \\
&= \int \sum_j \sigma_j^2 \rho_j^2(t) dw(t) \\
&= \sum_j \sigma_j^2 < \infty
\end{aligned}
$$

□

For the model error e, we have $E(e) = 0, Var(e) = \sigma^2$. Define the standardized error e' as $e = e'\sigma(\mu)$. Then,

$$
\begin{aligned}
E(e'|X) &= E(\frac{e}{\sigma(\mu)}|X) = \frac{E(e|X)}{\sigma(\mu)} = 0, \\
Var(e'|X) &= \frac{1}{\sigma^2(\mu)} Var(e|X)
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{1}{\sigma^2(\mu)}\sigma^2(\mu) = 1, \\
Var(e') &= Var[E(e'|X)] + E[Var(e'|X)] = 1.
\end{aligned}
$$

As maximum likelihood estimation is the principal method used for GLM, it is necessary to specify a probabilistic mechanism that controls the data generation process. Such a specification is unclear in many practical situation. Wedderburn [1974] proposed an important extension of likelihood function, the quasi-likelihood method. The term "quasi-likelihood" indicates, even without sufficient information about the distribution, the constructed likelihood behaves like a log-likelihood function, under mild assumptions, and thus is of statistical importance.

Suppose that the components of the response vector Y are independent with mean vector μ and covariance matrix $\sigma^2 V(\mu)$, where σ^2 is unknown but $V(\mu)$ is a matrix of known functions about μ. The mean vector μ relates to the parameter of interest β on covariates x.

The quasi-likelihood function $K(y_i, \mu_i)$ is defined as

$$
\frac{\partial K(y_i, \mu_i)}{\partial \mu_i} = \frac{y_i - \mu_i}{\sigma^2 V(\mu_i)}
$$

or equivalently,

$$
K(y_i, \mu_i) = \int^{\mu_i} \frac{y_i - s}{\sigma^2 V(s)} ds + \text{function of } y_i.
$$

As we assume the components of Y are independent, the quasi-likelihood for the complete data is

$$
K(y, \mu) = \sum_{i=1}^{n} K(y_i, \mu_i).
$$

By Wedderburn [1974], K is the log-likelihood function if y comes from a one-parameter exponential family. Meanwhile, K also possesses similar statistical properties of log-likelihood function, such as

$$\begin{aligned} E(\frac{\partial K}{\partial \mu_i}) &= 0 \\ E(\frac{\partial K}{\partial \beta_j}) &= 0 \\ E(\frac{\partial K}{\partial \mu_i})^2 &= -E(\frac{\partial^2 K}{\partial \mu_i^2}) = \frac{1}{\sigma^2 V(\mu_i)} \end{aligned} \tag{2.1}$$

The maximum quasi-likelihood estimates are obtained by setting $U(\beta)$ equal to zero, where $U(\beta)$ is the derivative of K with respect to β,

$$U(\beta) = \frac{\partial K}{\partial \beta} = \sum_{i=1}^{n} (Y_i - \mu_i) g'(\eta_i) \epsilon^{(i)} / \sigma^2(\mu_i) = 0$$

It is called the quasi-score function (score function). $U(\beta)$ has zero expectation and covariance matrix $\sigma^2 D^T V^{-1} D$, where the components of D are $D_{ij} = \partial \mu_i / \partial \beta_j$, the derivatives of $\mu(\beta)$ with respect to the parameters. McCullagh [1983] showed that quasi-likelihood estimates $\hat{\beta}$ is $\sqrt{n}$-consistent and $n^{1/2}(\hat{\beta} - \beta) \sim N_p(0, \sigma^2 (\frac{D^T V^{-1} D}{n})^{-1})$, asymptotically.

2.5 Errors of Functional Quasi-likelihood Model

2.5.1 Truncation Error

Based on the principal component expansions of functional predictor $X(t)$ and parameter function $\beta(t)$, functional Quasi-likelihood model is

$$Y_i = g(\alpha + \sum_{j=1}^{\infty} \beta_j \epsilon_j^{(i)}) + e' \tilde{\sigma}(\alpha + \sum_{j=1}^{\infty} \beta_j \epsilon_j^{(i)}), i = 1, \ldots, n, \tag{2.2}$$

where $E(e'|X(t), t \in T) = 0$, $Var(e'|X(t), t \in T) = 1$, and both $g(\cdot)$ and $\sigma^2(\cdot)(\tilde{\sigma}^2(\cdot))$ are assumed known.

In practice, we usually truncate at $p = p_n$ to approximate model (2.2) and assume the dimension $p_n \to \infty$ as $n \to \infty$.

The choice of this auxiliary parameter p is important in the procedure of modeling. Several selection methods have been analyzed, including minimization the prediction error via cross validation (Rice and Silverman [1991]) and minimization of the Akaike Information criterion (AIC) (Muller and Stadtmuller [2005]). Minka [2000] showed how to use Bayesian model selection to detect the number of components. A good overview of standard rules of thumb was given by Cangelosi and Goriely [2007].

We truncate the infinite dimensional full model at $p = p_n$ to get the p-truncated model:

$$Y_i = g(\alpha + \sum_{j=1}^{p} \beta_j \epsilon_j^{(i)})_{+} e_i' \tilde{\sigma}(\alpha + \sum_{j=1}^{p} \beta_j \epsilon_j^{(i)}), i = 1, \ldots, n, \tag{2.3}$$

Statistical properties about estimate of the $\hat{\beta}_j's$ and parameter function $\beta(\cdot)$ have been developed in Muller and Stadtmuller [2005].

2.5.2 Approximation Error

The principal component expansion of $X(t) - \mu(t), t \in T$ is constructed based on the covariance function:

$$K(s,t) = E\{[X(s) - \mu(s)][X(t) - \mu(t)]\}$$

While, in reality, only a sample of $\{X_1, X_2, \ldots, X_n\}$ of independent stochastic processes is observed, which are distributed as X. We would estimate the covariance function by the empirical approximation:

$$\tilde{K}(s,t) = \frac{1}{n}\sum_{i=1}^{n}[X_i(s) - \bar{X}(s)][X_i(t) - \bar{X}(t)]$$

where $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$.

Let $\tilde{\theta}_1 \geq \tilde{\theta}_2 \geq \ldots \geq 0$ be the eigenvalues of $\tilde{K}$ and let the corresponding orthonormal eigenfunctions be $\tilde{\psi}_1, \tilde{\psi}_2, \ldots$. Then

$$\tilde{K}(s,t) = \sum_{j=1}^{\infty} \tilde{\theta}_j \tilde{\psi}_j(s)\tilde{\psi}_j(t).$$

Consequently, the functional principal component expansion of $X_i(t)$ is

$$X_i(t) = \sum_{j=1}^{\infty} \tilde{\epsilon_j}^{(i)} \tilde{\psi}_j(t)$$

where $\tilde{\epsilon_j}^{(i)}$ is the principal component score of $X_i(t)$ corresponding to the jth principal component $\tilde{\psi}_j(t)$. We consider $\tilde{\psi}_j(t)$ as an approximation to $\psi_j(t)$, and both $\tilde{\epsilon_j}^{(i)}$ and $\tilde{\psi}_j(t)$ change with sample size n.

2.5.3 Full Model and Work Model

With the observed data $(X_i(t), t \in T, Y_i), i = 1, \ldots, n$, we assume $X_i(t), t \in T$ is a square integrable stochastic process on a real compact set T and these data form an *i.i.d.* sample. The mean function of X is $\mu(t) = E[X(t)]$. We assume there exists a known link function $g(\cdot)$, which is monotone, twice continuously differentiable with bounded derivatives, which relates the expectation of Y_i to the random function $X_i(t)$, through

$$\eta_i = \alpha + \int_t \beta(t) X_i(t) dw(t), \tag{2.4}$$

$$E(Y_i | X(t), t \in T) = \mu_i = g(\eta_i). \tag{2.5}$$

The full functional quasi-likelihood model is

$$Y_i = g(\alpha + \int \beta(t)X_i(t)dw(t)) + e_i, i = 1, \ldots, n, \tag{2.6}$$

where $E(e_i|X(t), t \in T) = 0$ and $Var(e_i|X(t), t \in T) = \sigma^2(\mu_i) = \tilde{\sigma}^2(\eta_i)$. We set $\sigma^2 = E[\tilde{\sigma}^2(\eta)]$. Both the link function $g(\cdot)$ and variance function $\sigma^2(\cdot)$ $(\tilde{\sigma}^2(\cdot))$ are assumed known.

Setting $\epsilon_j^{(i)} = \int X_i(t)\psi_j(t)dw(t)$, the full model can be rewritten as

$$Y_i = g(\alpha + \sum_{j=1}^{\infty} \beta_j \epsilon_j^{(i)}) + e_i' \tilde{\sigma}(\alpha + \sum_{j=1}^{\infty} \beta_j \epsilon_j^{(i)}), i = 1, \ldots, n, \tag{2.7}$$

where e_i' is the standardized error, $e_i = e_i'\sigma(\mu_i)$, and $E(e'|X) = 0$, $Var(e'|X) = 1$. We approximate model (2.7) with a series of models where we truncate the number of predictors at $p = p_n$ and approximate $\epsilon_j^{(i)}$ by $\tilde{\epsilon}_j^{(i)}$ corresponding to the jth principal component $\tilde{\psi}_j(t)$.

The working model becomes

$$Y_i = g(\alpha + \sum_{j=1}^{p} \beta_j \tilde{\epsilon}_j^{(i)}) + e_i' \tilde{\sigma}(\alpha + \sum_{j=1}^{p} \beta_j \tilde{\epsilon}_j^{(i)}), i = 1, \ldots, n, \tag{2.8}$$

where both $p = p_n$ and $\tilde{\epsilon}_j^{(i)} = \tilde{\epsilon}_{j,n}^{(i)}$ change with sample size n. For simplicity, we suppress the indices n.

To develop statistical inference, we also assume the dimension $p = p_n$ grows asymptotically as $n \to \infty$ and the growth rate will be specified in the next chapter.

Chapter 3: Asymptotic Properties of the Functional Quasi-likelihood Model

In Chapter 2, a functional quasi-likehood model with both truncation and approximation error is proposed. With a sample of stochastic processes and responses, we firstly evaluate the accuracy of principal component approximations $\tilde{\psi}$ and the corresponding principal component scores $\tilde{\epsilon}$ with $p_n \to \infty$ as $n \to \infty$. Asymptotic limit results for the approximated $\tilde{\epsilon}$ ($\tilde{\psi}$) and the true ϵ (ψ) are derived with increasing sample size n.

Once the full model (2.7) is truncated to a finite dimensional working model (3.7), we can apply the methodology of quasi-likelihood model, by solving a p-dimensional score equation.

With approximated $\tilde{\epsilon}$, we establish that the asymptotic properties of the resulting estimate $\tilde{\hat{\beta}}$ obtained from the score function in this functional quasi-likelihood function.

3.1 Principal Component Expansion Approximation

In Section (2.5.2) we discussed the approximation error derived from empirical approximation of covariance function $\tilde{K}(s,t)$. Here we summarize the statistical properties of the principal component expansion approximation.

Let $X(t)$ denote a random function on a real compact set T, which is square integrable, and let $\mu(t) = E[X(t)]$. The covariance function of $X(t) - \mu(t)$ is

$$K(s,t) = E\{[X(s) - \mu(s)][X(t) - \mu(t)]\}$$

where $K(\cdot,\cdot)$ is also square integrable.

The Karhunen-Loève expansion of $X(t) - \mu(t)$ is

$$X(t) - \mu(t) = \sum_{j=1}^{\infty} \epsilon_j \psi_j(t)$$

where the ϵ_j are random principal component scores of $X(t)$, and $\epsilon_j = \int_T X(t)\psi_j(t)dt$.

The empirical approximation of $K(s,t)$ is

$$\tilde{K}(s,t) = \frac{1}{n}\sum_{i=1}^{n}[X_i(s) - \bar{X}(s)][X_i(t) - \bar{X}(t)]$$

where $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$. It leads to the principal component expansion of $X(t)$ as

$$X(t) = \sum_{j=1}^{\infty} \tilde{\epsilon_j}\tilde{\psi_j}(t).$$

The set of $\tilde{\psi_1}, \tilde{\psi_2}, \ldots$ forms a complete orthonormal basis of $L_2(T)$. For any observed sample function $X_i(t)$, we have

$$X_i(t) = \sum_{j=1}^{\infty} \tilde{\epsilon_j}^{(i)}\tilde{\psi_j}(t),$$

where $\tilde{\epsilon_j}^{(i)}$ is the principal component score of $X_i(t)$ corresponding to the jth principal component $\tilde{\psi_j}(t)$. Because $\tilde{\psi_j}$ is considered as an approximation to ψ_j given a sample $\{X_1, \ldots, X_n\}$, we use $\tilde{\psi_j}$ and $\tilde{\epsilon_j}$ to distinguish from the true ψ_j and ϵ_j constructed from the true stochastic process $X(t)$. Both $\tilde{\psi_j} = \tilde{\psi}_{j,n}$ and $\tilde{\epsilon_j} = \tilde{\epsilon}_{j,n}$ change with sample size n, but here we suppress the indices n for simplicity.

As $\{\psi_1, \psi_2, \ldots\}$ forms a complete orthornaormal basis of $L_2(T)$, we may write:

$$\tilde{\psi}_j = \sum_{k=1}^{\infty} a_{jk}\psi_k. \tag{3.1}$$

According to Hall and Hosseini-Nasab [2009] the generalized Fourier coefficients a_{jk} are functions of the data, and, for each $j \neq k$, we have

$$a_{jj} = 1 - \frac{1}{2}n^{-1}\sum_{l:l\neq j}(\theta_j - \theta_l)^{-2}\left(\int Z\psi_j\psi_l\right)^2 + O_p(n^{-3/2}), \tag{3.2}$$

$$\begin{aligned} a_{jk} &= n^{-1/2}(\theta_j - \theta_k)^{-1}\int Z\psi_j\psi_k + n^{-1}\Bigg\{(\theta_j - \theta_k)^{-1}\sum_{l:l\neq k}(\theta_j - \theta_k)^{-1}\left(\int Z\psi_j\psi_l\right)\left(\int Z\psi_k\psi_l\right) \\ &\quad -(\theta_j - \theta_k)^{-2}\left(\int Z\psi_j\psi_j\right)\left(\int Z\psi_j\psi_k\right)\Bigg\} + O_p(n^{-3/2}) \end{aligned} \tag{3.3}$$

where $Z = n^{1/2}(\tilde{K} - K)$, and $\int Z\psi_j\psi_k = \int\int_{T^2} Z(s,t)\psi_j(s)\psi_k(t)dw(s)dt$. The conditions under which the infinite series in (3.2) and (3.3) converge are listed in Assumption 1. dt With functional observations, we use the $L2$ norm to analyze the accuracy of $\tilde{\psi}_j(t)$ as an estimate of $\psi_j(t)$ by the following lemma.

Lemma 3.1.1. $n\|\tilde{\psi}_j - \psi_j\|^2 = \sum_{k:k\neq j}(\theta_j - \theta_k)^{-2}(\int Z\psi_j\psi_k)^2 + O_p(n^{-1/2})$

Define $\tilde{\Delta} = (\int |\tilde{K} - K|^2)^{1/2}$,

$$\delta_j = \min_{1\leq k\leq j}(\theta_k - \theta_{k+1})$$

$$J = inf\Big[j \geq 1 : \theta_j - \theta_{j+1} \leq 2\tilde{\Delta}\Big] \tag{3.4}$$

Assumption 1. *The conditions under which the infinite series in (3.2) and (3.3) converge:*

(a) for all $C > 0$ and some $\epsilon > 0$,

$$\sup_{x\in T}\left\{E\left|X(t)^C\right|\right\} < \infty \tag{3.5}$$

$$\sup_{s,t\in T}\left(E\left[\left\{|s-t|^{-\epsilon}\left|X(s) - X(t)\right|\right\}^C\right]\right) < \infty \tag{3.6}$$

(b) for each integer $r \geq 1, \theta_j^{-r} E\Big(\int_T [X(t) - \mu(t)]\psi_j(t)dt\Big)^{2r}$ is bounded uniformly in j.

(c) The eigenvalues $\theta_j' s$ are distinct.

Based on results of Hall and Hosseini-Nasab [2009], under Assumptions 1(a) and 4, with probability 1, $\|\tilde{\epsilon}_l^{(i)} - \epsilon_l^{(i)}\| = O_p(n^{-1/2})$, for all l such that $1 \leq l \leq J-1$, where J defined as (3.4).

3.2 Principal Component Expansion Truncation

We introduced the truncation error in Section (2.5.1) while we truncated the infinite dimensional full expansion. In this section, we will talk about some useful asymptotic properties of estimates from the truncated model (2.3).

Let $\hat{\beta}_j$ denote an estimate obtained from the truncated model (2.3) and let $\chi = \{X_1, \cdots, X_n\}$.

Theorem 3.2.1. *For fixed j and each realization of χ we have*

$$E[(\hat{\beta}_j - \beta_j)^2 | \chi] = O_p(p^2/n).$$

3.3 Asymptotic Properties of Functional Quasi-likelihood Estimators

The functional quasi-likelihood model is

$$Y_i = g(\alpha + \sum_{j=1}^{p} \beta_j \tilde{\epsilon}_j^{(i)}) + e_i' \tilde{\sigma}(\alpha + \sum_{j=1}^{p} \beta_j \tilde{\epsilon}_j^{(i)}), i = 1, \ldots, n, \tag{3.7}$$

The true value β may be estimated by solving

$$U(\beta) = \sum_{i=1}^{n}(Y_i - \tilde{\mu}_i)g'(\tilde{\eta}_i)\tilde{\epsilon}^{(i)}/\sigma^2(\tilde{\mu}_i) \tag{3.8}$$

The solutions of the score equation (3.8) will be denoted by

$$\hat{\tilde{\beta}}^T = (\hat{\tilde{\beta}}_1, \cdots, \hat{\tilde{\beta}}_p) \tag{3.9}$$

The mean squared error of $\hat{\tilde{\beta}}(t) = \sum_{i=1}^{p} \hat{\tilde{\beta}}_i \tilde{\psi}_j(t)$, conditional on $\chi = \{X_1, \cdots, X_n\}$ can be written as:

$$\begin{aligned}
&\int_I E\left[\left(\hat{\tilde{\beta}}(t) - \beta_0(t)\right)^2 |\chi\right] dt \\
= \quad & \int_I E\left[\left(\sum_{i=1}^{p} \hat{\tilde{\beta}}_i \tilde{\psi}_j(t) - \sum_{i=1}^{\infty} \beta_i \psi_j(t)\right)^2 |\chi\right] dt \\
= \quad & \int_I E\left[\left(\sum_{i=1}^{p} \hat{\tilde{\beta}}_i \tilde{\psi}_j(t) - \sum_{i=1}^{\infty} \beta_i \tilde{\psi}_j(t) + \sum_{i=1}^{\infty} \beta_i \tilde{\psi}_j(t) - \sum_{i=1}^{\infty} \beta_i \psi_j(t)\right)^2 |\chi\right] dt \\
= \quad & \int_I E\left[\left(\sum_{i=1}^{p} \hat{\tilde{\beta}}_i \tilde{\psi}_j(t) - \sum_{i=1}^{\infty} \beta_i \tilde{\psi}_j(t)\right)^2 |\chi\right] dt + \int_I \left[\left(\sum_{i=1}^{\infty} \ \beta_i [\tilde{\psi}_j(t) - \psi_j(t)]\right)^2\right] dt \\
& +2\int_I \left[\sum_{j=1}^{\infty} \beta_j (\tilde{\psi}(t) - \psi_j(t))\right] E\left[\sum_{j=1}^{p} \hat{\tilde{\beta}}_j \tilde{\psi}_j(t) - \sum_{j=1}^{\infty} \beta_j \tilde{\psi}_j(t) |\chi\right] dt.
\end{aligned}$$

Theorem 3.3.1. *With probability 1, if $X_1(t), \cdots, X_n(t)$ are square integrable random functions and $\sum_{j=1}^{\infty} \delta_j^{-1} \to 0$, for all $1 \le j \le J-1$,*

$$\int_I \left[\left(\sum_{i=1}^{\infty} \ \beta_i [\tilde{\psi}_j(t) - \psi_j(t)]\right)^2\right] dt \to 0.$$

Theorem 3.3.2.

$$\int_I E\left\{\left[\sum_{j=1}^{p} \hat{\tilde{\beta}}_j \tilde{\psi}_j(t) - \sum_{j=1}^{\infty} \beta_j \tilde{\psi}_j(t)\right]^2 |\chi\right\} dt = O_p(p^{1/2} n^{-3/2}) + o_p(1).$$

From Theorem 3.3.1 and 3.3.2, we can easily get, for all $1 \le j \le J-1$

Theorem 3.3.3. *Under Assumptions 2 - 6, with probability 1, we have*

$$\int_I E\left\{\left[\hat{\tilde{\beta}}(t) - \beta_0(t)\right]^2 |\chi\right\} dt \to 0.$$

Chapter 4: Functional Shape Analysis of Movie Success in Subsequent Sales Channel

In Chapter 2, we talk about the Functional Quasi-likelihood model with both truncation and approximation errors; Chapter 3 shows us the consistency of the estimates. Now we want to utilize this method and consider functional interactions.

Suppose we have two functional subjects, $X_1(t)$ and $X_2(t)$ on T. We can write their principal component expansions as:

$$\begin{aligned} X_1(t) &= \sum_{j=1}^{\infty} \epsilon_j \psi_j(t) \\ X_2(t) &= \sum_{j=1}^{\infty} \delta_j \phi_j(t) \end{aligned}$$

and their corresponding parameter functions:

$$\begin{aligned} \beta_1(t) &= \sum_{j=1}^{\infty} \beta_j^1 \psi_j(t) \\ \beta_2(t) &= \sum_{j=1}^{\infty} \beta_j^2 \phi_j(t) \end{aligned}$$

where the ψ and ϕ are the principal components of $X_1(t)$ and $X_2(t)$.

We model the two dimensional interaction parameter function as

$$U(s,t) = \begin{bmatrix} \psi_1(s), \psi_2(s), \cdots \end{bmatrix} \begin{bmatrix} \beta_1^{12} & 0 & 0 \\ 0 & \beta_2^{12} & 0 \\ & & \ddots \end{bmatrix} \begin{bmatrix} \phi_1(t), \phi_2(t), \cdots \end{bmatrix} = \sum_{l=1}^{\infty} \beta_l^{12} \psi_l(s) \phi_l(t)$$

The linear predictor (2.4) becomes

$$\begin{aligned}\eta &= \alpha + \int \beta_1(t)X_1(t)dt + \int \beta_2(t)X_2(t)dt \\ &\quad + \int_s \int_t U(s,t)X_1(s)X_2(t)dw(s)dt.\end{aligned}$$

As

$$\begin{aligned}&\int_s \int_t U(s,t)X_1(s)X_2(t)dw(s)dt \\ &= \int_s \int_t \sum_{l=1}^{\infty} \psi_l(s)\phi_l(t)\beta_l^{12} \cdot \sum_{j=1}^{\infty} \epsilon_j \psi_j(s) \cdot \sum_{k=1}^{\infty} \delta_k \phi_k(t) \\ &= \sum_{l=1}^{\infty} \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} \beta_l^{12} \epsilon_j \delta_k \int \psi_l(s)\psi_j(s)dw(s) \int \phi_l(t)\phi_k(t)dt \\ &= \sum_{l=1}^{\infty} \beta_l^{12} \epsilon_l \delta_l \qquad (4.1)\end{aligned}$$

We use the functional model with function interaction modeled as (4.1) to analyze the effect of Word-of-Mouth on movie box office sales. In this study, the goal is to understand (and predict) the value of the sales for entertainment products, by analyzing the shapes of on-line reviews. By "shape" we mean whether the curve is trending up over time, or whether it is trending down and the rate at which the shape is changing. We model the on-line review curve using functional shape analysis. Functional shape analysis first decomposes the (infinite-dimensional) shape into a finite dimensional sum of terms and subsequently models the individual shape dimensions. In this project, we study the impact of several input shapes on box office sales. The input shapes include three different measurements of the evolution of word-of-mouth about the popularity of a movie. One challenge of our model is the incorporation of an interaction term between the word-of-mouth shapes. We demonstrate how to incorporate interaction terms into functional shape models and compare the resulting model to simpler approaches.

4.1 Introduction

Functional Data Analysis (FDA, Ramsay and Silverman [2005]) deals with curves, or functions in general. If longitudinal measurements on the same individuals are made on a suitably spaced grid, such data are typically termed a sample of curves or functional data. FDA approach regards the entire curve as an observational unit and consider it as being observed in continuum, even if in practice humankind can only observe discrete data. Some of the methodological study in this field involves functional ANOVA (Fan and Lin [1998]), functional principal component analysis (Hastie and Sugar [2000]), regression with functional responses (Faraway [1997]), or functional predictors (James [2002], James and Silverman [2005]), or both (Zeger and Diggle [1994], Moyeed and Diggle [1994]).

Studying functional data allows us to better understand the dynamic features of predictors over time, but it also raises another type of problems when we are dealing with infinite-dimensional functional predictors, compared with the classical situation with finite-dimensional predictors. The main idea is to employ Karhunen-Loeve or other expansion of the functional predictor, in order to reduce the dimensions by truncating at a few number of terms. Muller and Stadtmuller [2005] develop asymptotic inference for this kind of problem.

Based on its advantage of capturing the dynamics, the FDA approach has been applied for the exploration and analysis of data originating from various fields, for example, eBay online auctions (Jank and Shmueli [2008], Jank and Zhang [2011]), market penetration (Sood, James and Tellis [2009]), open source software evolution (Stewart, Darcy and Daniel [2006]) and monthly index of production (James O. Ramsay and James B. Ramsey [2002]).

When dealing with several predictors, care must be taken in interpreting how the effect of one predictor on the response variable depends on the magnitude of another predictor. Interaction terms are used extensively both in linear and non-linear models. Computing the magnitude of interaction effect in non-linear model requires special care and delicate treatment. Norton [2004] takes logit and probit models as examples, showing that computing the marginal effect of a change in two predictors is more complicated in non-linear models than in linear models. They present a correct estimator for the two-way interaction effect in non-linear model. However, Greene [2010] argues that the statistical testing about interaction terms might be uninformative and misleading in the context of the model and provides a useful two-step method to proceed in the analysis of interaction terms. For multiple regression models, interpretation of significant interaction terms still might be difficult in certain cases. Cortina [1993] argue that a seemly statistically significant interaction term may stem from an undetected nonlinear effect, rather than a linear multiplicative effect. After introducing interaction effects in the model, the statistical significance of the lower-order coefficients could be misleading for the typical purposes of hypotheses testing (Braumoeller [2004]). In the field of nonparametric models, many studies have been made on the detection of interactions. For instance, a new algorithm GUIDE (Generalized, Unbiased Interaction Detection and

Estimation) for regression tree construction has been developed to eliminate variable selection bias and test the presence of local two-way interactions (Loh [2002]); Sperlich [2002] proposes two statistics for detecting interactions and proves that the test procedure could detect an interaction term with probability 1.

Consumer reviews are product reviews created by consumers, which are quite different from professional reviews. Much work has been done on investigating professional reviews (King [2007], Reinstein and Synder [2005]). Dellarocas, Zhang and Awad [2007] have shown the low correlation between consumer and professional ratings. Meanwhile, consumer reviews are growing increasingly important in the consumer's decision making process. Before people go to the theatre, they may firstly sit in front of their computers to check what others say about that movie. Is the movie brilliant or just so-so? Is it worth the money they pay for the tickets? In this case, consumer reviews play a crucial role in decision process of potential consumers. Previous works (Dellarocas, Zhang and Awad [2007]; Duan, Gu and Whinston [2008]; Chintagunta, Gopinath and Venkataraman [2010]) have investigated the impact of on-line user reviews on box office performance of movies.

To summarize consumer reviews, three commonly used measurements are valence, volume and dispersion. The valence of consumer reviews is the mean user rating, an assessment of quality of a product. Dellarocas, Zhang and Awad [2007] find that valence is the most informative indicators of forecasting box office sales. Moreover, their analysis shows that consumer ratings are more influential in predicting future box office revenues than average professional reviews. Volume measures the total number of unique reviews. Intuitively, we can argue that more volume, which means more people have watched that movie, probably leads to more future movie viewers in theatre. It has been claimed by Duan, Gu and Whinston [2008] that volume, rather than valence, has a significant impact on movies' box office revenues. Interestingly, the effect of valence and volume on box office performance of movies may vary geographically. Chintagunta, Gopinath and Venkataraman [2010] investigate the impact of consumer reviews on local geographic box office sales of movies and find that valence seems to matter, not the volume. Then they do the same analysis with national data, and obtain a different result - the volume matters, but not the valence. Consumer reviews could vary a lot due to different reviewers' preference, especially on movies. Another metric of consumer reviews, dispersion, capture this difference. Big dispersion means big difference of opinions. Sun [2008] studies the informational role of consumer disagreement and suggests that dispersion is an effective marketing tool in influencing consumers purchase decision.

Sun [2008] tests the hypothesis about the significance of interaction terms between valence and dispersion. Sun finds the impact of dispersion changes with different level of average rating. Specifically, a higher dispersion has a positive influence on box office performance if and only if the average valence is relatively low. One can also argue impact of interaction between valence and volume. Intuitively, more consumers that think the movie being good may lead to larger impact on future box office sales. Nam et al. [2010] studies this interaction between valence and volume in the context of video-on-demand service. These works show the importance of including the interaction effects of on-line word-of-mouth(eWOM) in the model.

While interaction terms have received increasing attention recently, few work has been done on the functional interaction terms. When incorporating functional interaction terms in the model, we could have more flexibility of presenting how the effect of one shape on the response depends on the magnitude of another shape. In the review context, we represent the review processes (valence, volume, dispersion) of as functional objects. In that sense, every movie is associated with three functional objects describing the review evolution. We analyse how these functional objects interact with each other when affecting future sales performance by a functional regression model. Because of the difficulty of interpreting the functional interactions from a model with several functional predictors, we develop a re-scaled 3D plots which graphically illustrate how several functional predictors combined together affect the scalar response.

This chapter is organized as follow. In the next section, we provide some details about the data we use in the study. Section 4.3 gives some insights we gain from applying shape analysis to the eWOM, which are used in building the functional regression model. Section 4.4 analyse the estimate results we obtain in Section 4.3 and focus on interpreting the interaction terms in a traditional way. In Section 4.5 we develop an adjusted 3D plot which could help us better understand the interaction effect of eWOM. Finally, we conclude with further remarks in Section 4.6 .

4.2 Data

Our data are collected from 876 movies during Feb.1999 to Dec.2010 from Yahoo and IMDB. The sample data we use in this study include 405 movies with a complete history of user reviews from their release dates and have the complete corresponding box office revenue data as well. For each movie, we have the following information from Yahoo and IMDB: valence, volume, dispersion, box office revenue, launch date, distribution, rating, genre, award nominations and total advertising. Letter grade of each individual user review is converted into a numerical value by assigning F and D- to 0 and A and A+ to 10 with the rest letters corresponding to its numeric counterpart. Table 4.1 provides some key summary statistics.

Table 4.1: Descriptive Statistics

	Min.	Max.	Mean	Sd	Median
valence	0	10	5.25	3.33	6.14
volume	1	1618	73.09	149.96	16
dispersion	0	50	6.79	6.79	6
B.O. sale	5.61	19.29	14.32	1.99	14.6
adspent	1.44	10.88	9.34	1.61	9.78
awards nomi	0	75	9.14	12.52	4
Total number of movies			405		
Total number of reviews			395,297		

4.3 Estimating Review Shapes

Limitations in human perception and measurement capability allow us to record only discrete data. Firstly, we start by estimating the underlying smooth review curves for each movie. There are a variety of data smooothers. One flexible and computationally efficient choice is penalized smooth spline. Specifically, let $\tau_1, \ldots, \tau_L$ be a set of knots. Then, a polynomial spline of order p is given by

$$f(t) = \beta_0 + \beta_1 t + \beta_2 t^2 + \ldots + \beta_p t^p + \sum_{l=1}^{L} \beta_{pl}(t - \tau_l)_+^p,$$

where $u_+ = uI_{u \geq 0}$.

Define the roughness penalty

$$PEN_m(t) = \int \{D^m f(t)\}^2 dt$$

where $D^m f, m = 1, 2, 3, \ldots$, denotes the mth derivative of the function f. The penalized smoothing spline f minimizes the penalized squared error

$$PENSS_{\lambda,m} = \int [y(t) - f(t)]^2 dt + \lambda PEN_m(t)$$

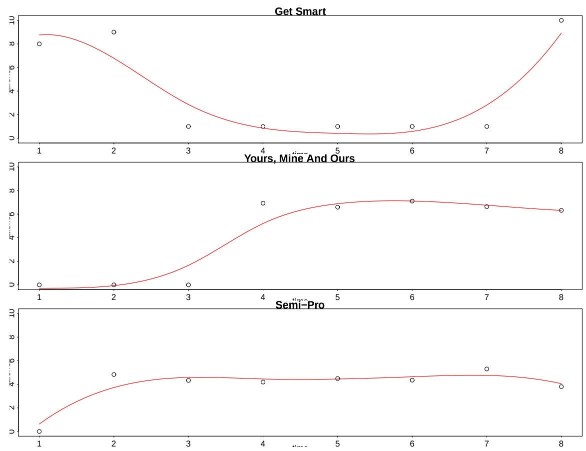
where $y(t)$ denotes the observed data at time t and the smoothing parameter λ controls the tradeoff between data fit and smoothness of the function f.

In this study, we use smoothing splines of order $p = 4$ and a smoothing parameter of $\lambda = 0.2$. The specified selection is guided by the goal of obtaining smooth functional objects that visually represent the original data well.

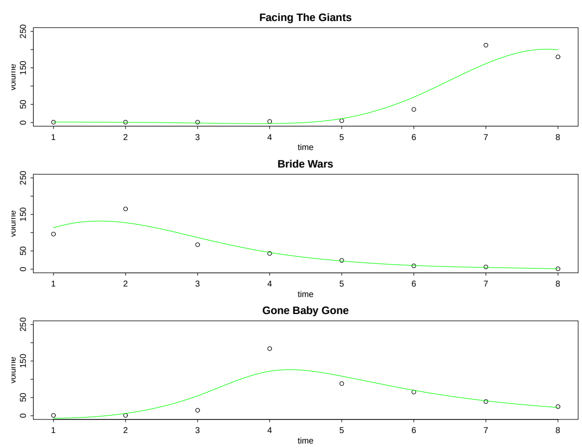
The advantages to focus on the functional aspect of consumer reviews is to capture the trend and evolution over time and investigate how different dynamics of consumer reviews affect future demanding performance. The pattern of the three metrics of consumer reviews - valence, volume and dispersion - is highly heterogeneous across different movies, which is shown in Figure 4.1.

Take a sample of consumer movie reviews as an example. Those movies share similar average valence, volume and dispersion over time, but with different evolution patterns. In the first panel, “Get Smart” has high valence at the beginning and end but considerably low in the middle; while “Yours, Mind And Ours” has low valence initially but it increases to higher level in the later weeks; the valence of “The Forbidden Kingdom” is moderate in the middle but low at two tails. Please note that these three movies share similar average valence - 4, 4.19, 3.91 respectively. In the second panel, there are three movies with similar average values, but obtaining peak values at different time. These three movies obtain their peak volume in the last two weeks, at the first two weeks and in the middle respectively. The last panel shows different evolution of dispersion for three movies with similar average values. The dispersion of “The Forbidden Kingdom” increases gradually, but faster for “Munich”; while, for “The Perfect Holiday”, dispersion rises and gets the first peak in week three, then decreases gradually, but jumps back to a higher level at the end.

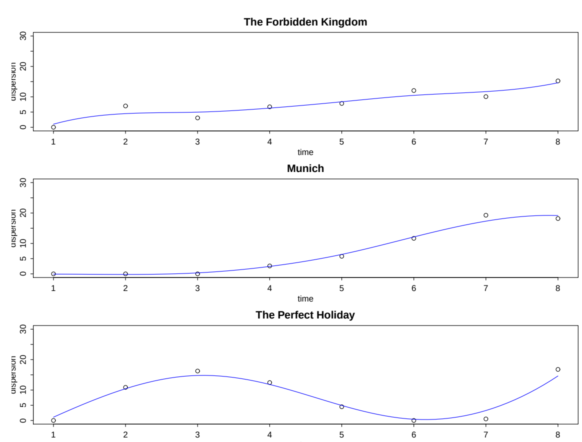
In order to capture the dynamic patterns of valence, volume and dispersion, we propose a functional model to analyse the “shapes” of consumer reviews and how they interact with each other when they influence consumers’ decision. Let $Va(t)$,



(a) Smoothed Valence Curves



(b) Smoothed Volume Curves



(c) Smoothed Dispersion Curves

Figure 4.1: Heterogeneity of Review Curves

$Vo(t)$ and $Dis(t)$ denote the function of valence, volume and dispersion over time respectively. We believe that different dynamic patterns of these three functions affect box office sale variously. Meanwhile, we analyse the two-way interaction terms: $Va(t) * Vo(t)$, $Va(t) * Dis(t)$ and $Vo(t) * Dis(t)$ in the functional regression model. These terms represent how two "shapes" interacted with each other. Functional regression model enables us to investigate the dynamic patterns of eWOM and how different dynamic patterns combined together affecting box office sales.

4.4 Shape Analysis of eWOM

In this section, we lay the foundation of our functional regression model. The functional regression model uses the shapes of previous eWOM to predict future box office sale. We separate our data into two part, the first seven weeks and week eight. We analyze the shapes of valence, volume and dispersion from week one to week seven. Then, we use the insights obtained from these shapes to predict the box office sale in week eight by building a functional regression model.

Functional Shape Analysis (FSA) focuses on a sample of *functional observations*, e.g.curves., and treat the observed curves as the units of observation. Thus our first step is to smooth the observed reviews' histories using penalized smoothing spline, a flexible and computationally efficient smoothing technique. Figure 4.1 shows the smoothed eWOM from the observed values. We can see that smoothing eliminates noise and captures the main pattern. The smoothed shapes of eWOM are considerably heterogeneous across movies. These different shapes plausibly contain

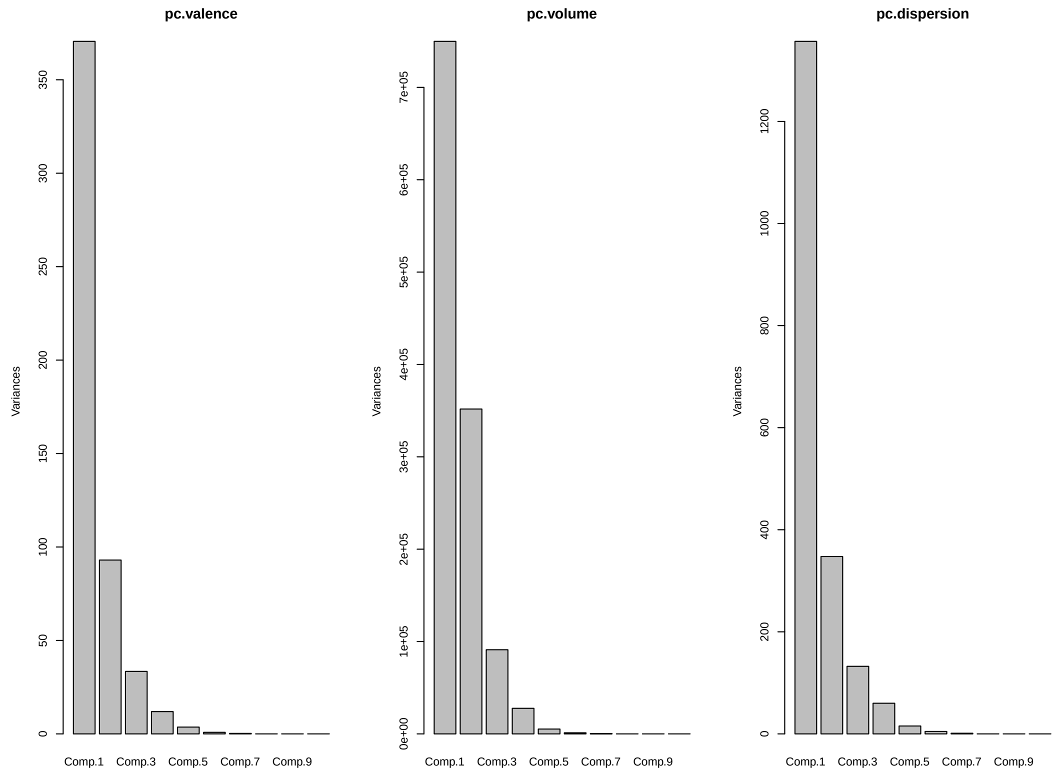


Figure 4.2: Screeplot

vital information about the potential demand, e.g. the most recent future box office sale. Thus we employ Functional Principal Component Analysis (fPCA) to extract the most indicative shapes that are common across all movies and use them to build the model.

Figure 4.2 shows the percentage of variation captured by the first few PCs of valence, volume and dispersion. It shows that the first three fPC's explain more than 95% of the total data variation. We retain the first three fPCs of valence, volume and dispersion, which are displayed in Figure 4.3.

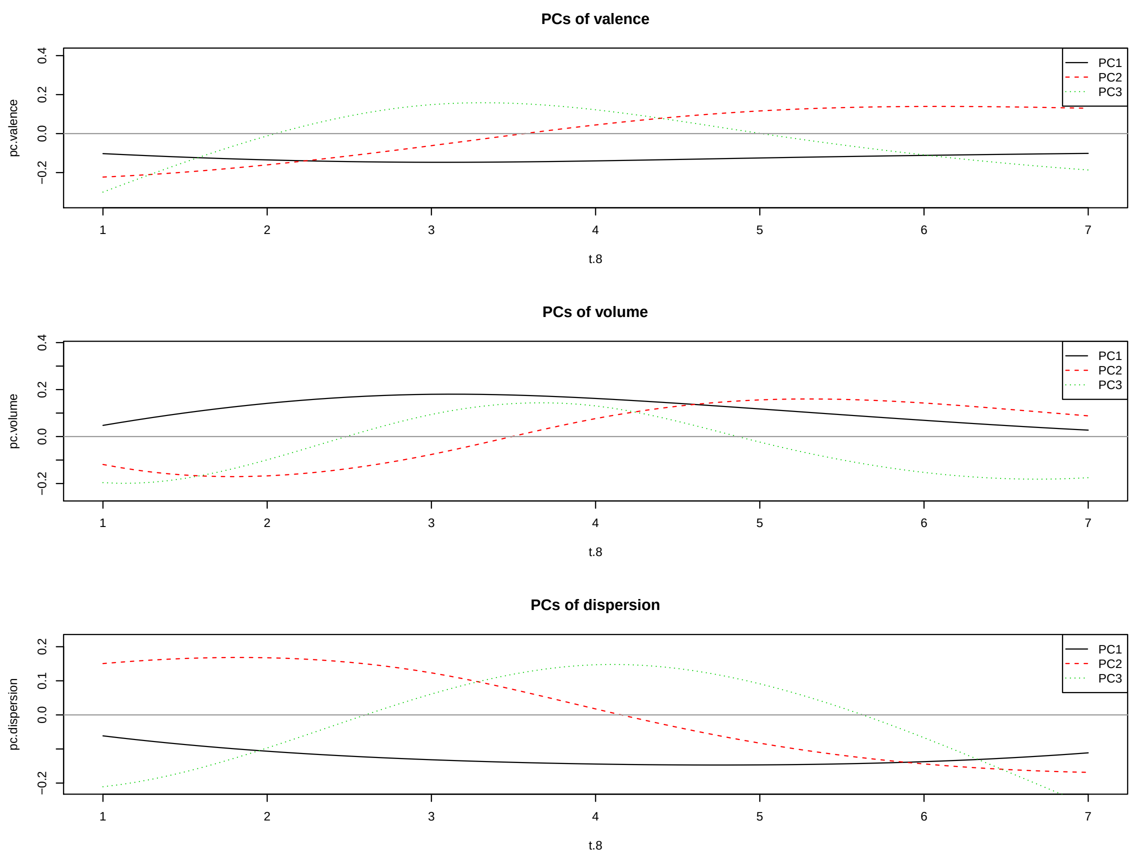


Figure 4.3: plot of PCs

As the final step, we develop a functional regression model of these nine shapes and their two-way interactions. Since while the PCs are common across all movies, the PC-scores are movie-specific, we could characterize each movie by its PC-scores of each PC. We link the box office revenue in week eight to these nine shapes, more precisely, their corresponding movie-specific PC-scores and all of their two-way interaction terms by linear regression. Some controlled variable that don't change over time are also being considered in the model, including number of awards nomination, total advertising expenditure and average box office sale. Then we utilize step-wise regression to retain four key shapes-$Va_{PC1}(t)$ $Va_{PC3}(t)$ $Dis_{PC1}(t)$ $Dis_{PC3}(t)$-two interaction terms-$Va_{PC1}(t) * Dis_{PC3}(t)$,$Va_{PC3}(t) * Dis_{PC3}(t)$-and two control variables - average Box Office sales and number of awards nomination. For illustrative purposes, we will restrict to the shapes, and not report the estimates of these two control variables.

4.5 Interpreting Interaction Terms in a Traditional Way

The parameter results are reported in Table **??**. The adjusted R^2 is 0.71, which is not bad in practice. Meanwhile, all these four shapes and their interactions are statistically significant, based on the small p-values. The $PC1$ amd $PC3$ of Valence and dispersion are the single shapes that is statistically significant.

Table 4.2: Parameter Estimation

	Estimate	Std Error	t.value	P-value
valence.Comp.1	-0.0018	0.0025	0.7127	0.4764
valence.Comp.3	-0.0111	0.0081	-1.3743	0.1701
dispersion.Comp.1	0.0068	0.0013	5.2655	0.0000
dispersion.Comp.3	-0.0078	0.0040	-1.9319	0.0541
valence.Comp.1:dispersion.Comp.3	-0.0009	0.0003	3.2095	0.0014
valence.Comp.3:dispersion.Comp.3	-0.0030	0.0008	-4.0123	0.0001
adj. R sq.		0.7072		

Firstly let us analyse the effect of single shapes of valence and dispersion in influencing the following demand. Val_{PC1} is negative together with negative coefficient, which means that high average valence would stimulate box office sale, since the movie with high rating would have negative PC-score. Meanwhile, the negative coefficient of Val_{PC3} results in "bad" box office performance for movies with positive scores of Val_{PC3} While, that kind of movies tends to have a concave-down shape of valence. Thus,even if generally high valence is good for the box office sale, the movies with high valence in week three and four and low valence at tails tend to have bad box office performance. The Dis_{PC1} and Dis_{PC3} have similar patterns with $Va_{PC1}(t)$ and $Va_{PC3}(t)$, and suggests that dispersion is especially hurtful 2-4 weeks prior to the movies demanding(week 8), while in week 1-2 or week 6-7, dispersion is not obviously deleterious.

There are two functional interaction terms in our functional regression model: $Val_{PC1} * Dis_{PC3}$ and $Val_{PC3} * Dis_{PC3}$. It shows that the interaction between valence and dispersion is important in forecasting the future demand, while volume does not interact with other measurements. While interpreting "traditional" interaction terms is relatively easy, it need to be more careful when dealing with *functional* interaction terms. Now let us discuss the effects of interaction of valence and dispersion in the same analytical procedure as previously. The coefficient of $Val_{PC1} * Dis_{PC3}$ is positive, then if a movie has positive Dis_{PC3}-score (concave-down shape of dispersion), that type of dispersion is more harmful to box office sale, when combined with high average valence which leads to negative Val_{PC1}-score. Meanwhile, the coefficient of $Val_{PC3} * Dis_{PC3}$ is negative. Then if a movie with concave-down shape of dispersion (positive Dis_{PC3}-score) would have worse box office performance when its valence is also at its peak in the middle (positive Val_{PC3}-score), because($+Val_{PC3}$-score) * ($+Dis_{PC3}$-score) * ($-$coefficient) = negative effect to box office sale. Similarly, we would expect a box office success when the movie has a concave-up shape of valence (negative Val_{PC3}-score and concave-down shape of dispersion(positive Dis_{PC3}-score).

As you may have noticed, the shortcoming of interpreting *functional* interaction terms using the same method as doing this with *traditional* interaction terms is that we can't *see* the pictures of the changes. When we do the same thing to *traditional* interaction terms, we could just show that change of the slope. In that case, from the deeper or flatter lines, we capture the idea of which typical kind of one predictor combined with which typical kind of the other predictors could give us the best (largest, or smallest) result. However, when encountering the *functional* interaction terms, we can only speculate a certain *pattern* of the shape with a certain value of principal component's score, since only one principal component score can't tell us the whole story of what the original shape would look like. In our case, the predictors are smoothed shapes and we separate one shape into several components. These components combined together build the original shape. Thus analysing one single principal component can only give us a pattern of what the original shape would look like, rather than really *see* the shapes. Based on this inconvenience, we develop a 3D plots to illustrate how these shapes interacted with each other in influencing future box office sale.

4.6 Interpreting Interaction Terms with Modified 3D Plot

Firstly, let us focus on the first interaction term - $Val_{PC1} * Dis_{PC3}$. After plotting a 3D graph with three axis's representing Val_{PC1}-score,Dis_{PC3}-score and $log(BORev)$ respectively. This plot only contains the information of scores, and how different scores of PC affect the box office revenue. We are still uncertain about the original shapes. Then we relabel the valence and dispersion axes by the corresponding shapes. For the valence axis, each point now represents one typical shapes interested in. In this case, the scores of Val_{PC1},Val_{PC2} and Val_{PC3} may all be different. For dispersion axis, each point represents a different score of Dis_{PC3}, but scores of Dis_{PC1} and Dis_{PC2} remain the same. While re-labelling the dispersion axis, we at first choose a specified shape of dispersion we're most interested in, extract its scores of all PCs. The next step is to retain the scores of PC1 and 2, change the score of PC3, and rebuild the shapes of dispersion with different Dis_{PC3}-score and the same Dis_{PC1}-score and Dis_{PC2}-score. Figure 4.4 shows how the shape of dispersion changes with Dis_{PC3}-scores, with the other components' scores unchanged. The blue curves are the *original* shapes of dispersion, corresponding with the Dis_{PC3}-score on the left axis. When Dis_{PC3}-score is positive (the top four shapes), dispersion hits peak in the middle and are relatively low at the tails, which is corresponding to concave-down curves. While, when Dis_{PC3}-score is negative (the bottom four shapes), the shape of dispersion does not change to a concave-up pattern immediately, but waits until the score being even more less (the changing point here is -60). We select three most representative shapes, with Dis_{PC3}-score 80,0 and -80, and attach them to the dispersion axis. Finally, we get Figure 4.5, the modified 3D plot.

In Figure 4.5, we attached three typical shapes on the axis of dispersion, who share the same Dis_{PC1}-score and Dis_{PC2}-score, but have different Dis_{PC3}-score. On the axis of valence, we select five shapes of valence, all of that have the same average rating, but different pattern. These five valence shapes may have different scores of all the components. Now this graph really could give us a straightforward insight of how valence and dispersion affect the future box office sale.

These three shapes of dispersion have decreasing values in week 7. In the first case with mostly concave-down shape of dispersion and high value at the end, box

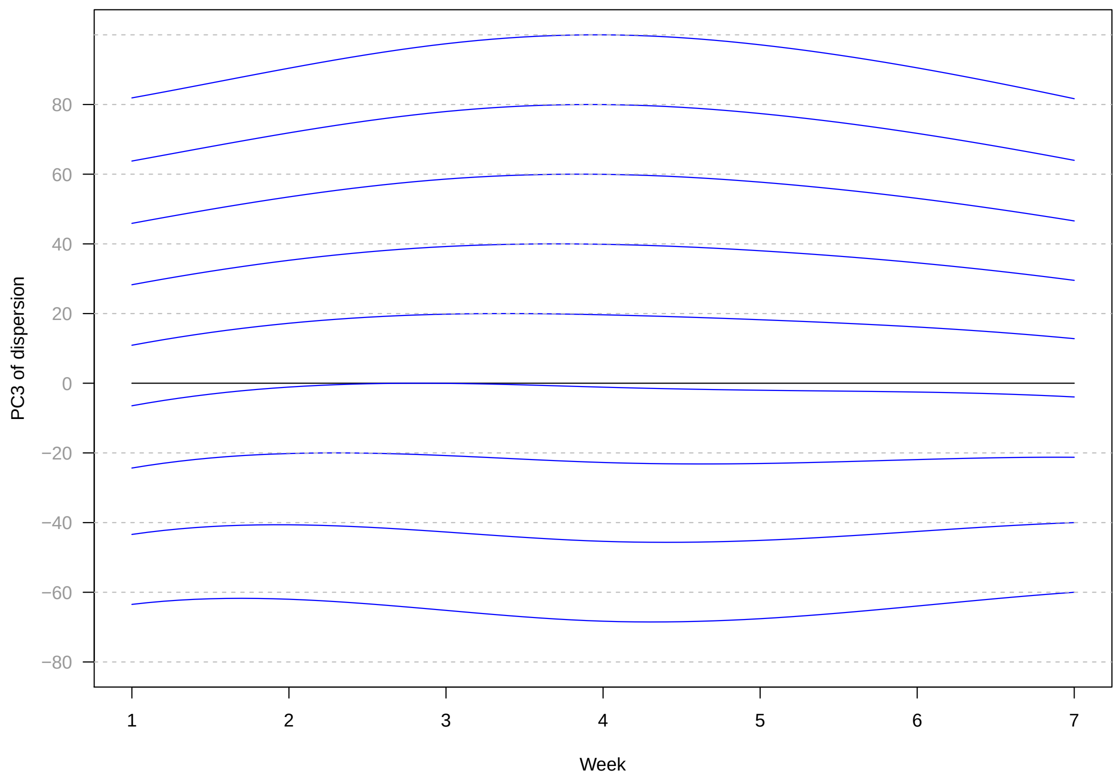


Figure 4.4: Pictures of Dispersion with Different PC3

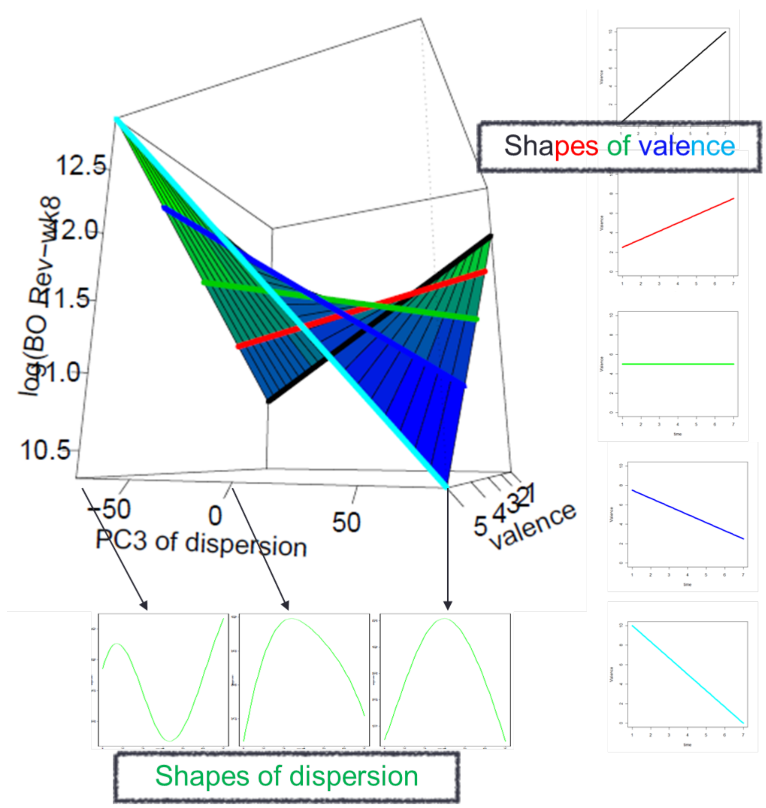


Figure 4.5: re-labelled 3D plot

office sale obtains its maximum when together with decreasing valence, where the valence is smallest at the end. One explanation is that decreasing valence means the decreasing expectation. Consumers' disagreement with the rating rises, and meanwhile, the total expectation for that movie drops. Then people may become more curious about that movie: is the movie really good or just so-so? This curiosity may stimulate recent box office sale. When we do the same analysis to the concave-down shape of dispersion (the most right panel of dispersion), we observe that low-end dispersion with high-end valence leads to most recent future box office success. This result is plausible: even if previously the opinions of the movie vary, finally most people think that movie is great, and this final revolution drives future box office performance jump. Meanwhile, if we focus on the line in the middle, the regression line when the valence is flat all the time. The slope of that regression line is small and we could say that for stable valence, the impact of dispersion on BO revenue is not obvious.

All in all, the main driver of box office performance is valence and dispersion and these two metrics of reviews interacted together influence future BO revenue. High-end dispersion is good when it's combined with low-end valence; high-end valence is good when it's combined with low-end dispersion; but for stable valence, the impact of dispersion diminishes. These results are shown in Figure 6. By choosing different shapes of valence and Dis_{PC3}-score, we could visually identify how different patterns of valence and dispersion affecting box office sale. In Figure 6, the average of valence is all the same, which emphasis the importance of dynamics of valence on affecting box office sale, which might be easily ignored by non-functional model.

4.7 Concluding Remarks

The objective of this paper is to investigate the role of on-line word of mouth of movies in affecting the future box office sales. We propose a functional regression model to measure the impact and interactions of shapes of valence, volume and dispersion of on-line consumer ratings on box office performance. We identify that the main drivers are valence and dispersion. By adjusting a traditional plot used to analysing interaction terms, we are capable of telling how valence and dispersion dynamically interact with each other influencing future off-line sales.

We acknowledge that this study is only the first step in introducing and interpreting functional interaction terms. The shortcoming of this method is that we don't have the total control of dispersion, since the scores of the first two components remain the same. Thus we can't carry out our analysis to any dispersion shapes directly. When we are interested in comparing the affect of two specified dispersion shapes, since their scores of PC1 and 2 are probably different, these two shapes can't be attached to the axis of dispersion simultaneously. To fix this disadvantage, we have to loose some accuracy: we first extract all the scores of one particular shape; by changing the Dis_{PC3}-score, we could mimic the other shape but retaining the scores of PC1 and 2.

Please note that even if we could fully control the valence shapes,the evolution of valence is not *continuous*. We have no idea about the valence shapes between point

4 and 5 on axis of valence. It means that we only have the control on discrete points of valence axis. However, we are able to rebuild of the original shapes of dispersion for any points on the axis. Thus there's some compromise between control and accuracy.

Chapter 5: Forecasting VOD Demand Curves: a Dynamic Functional Spatial-Temporal Approach

5.1 introduction

Streaming Video-on-Demand (VOD) systems offer consumers the ability to browse, select, watch and scan media content from a large various library, including pay-per-view and free contents, all from the comfort of their homes. There is extreme interest in streaming VOD because it simplifies the process of getting video to consumers, by allowing them to just point, click and watch instantly.

To be able to provide high quality services to consumers, it is crucial for the service providers to be able to understand the effect of on-demand content requests on several aspects - shape of the demand function for VOD, along with its spatial characteristics.

Introductions of new release movies in VOD channel are confronted with comparatively short lifecycle and extremely uncertain demand. Forecasting the demand shape of new release movies is not only challenging, but also important, for providing accurate and early guidance to cable operators and pay-tv service providers with regard to strategic decisions such as distribution and promotion.

Due to the large geographical coverage of VOD service providers, contents are stored at several large servers located in metropolitan areas. These servers are linked by a backbone network, thus contents can be transmitted among servers. Each customer is assigned to one particular server but s/he still can call videos from any other servers, with extra transmission cost. With the rapid increase of video quality and size, e.g. 3-D movies, it is becoming infeasible and wasteful to replicate the entire library at each server. Thus the ability to forecast the on-demand movies and their shapes at various locations is critical for providers to improve their decision making process.

Many researchers have studied optimal content allocation problems in Video-on-Demand systems. Yu et al. [2006] study the aggregate video access pattern of Video-on-Demand on the Internet. Moreover, they have shown that the popularity of individual movie changes much over time. It follows that an accurate estimate of individual video's demand is becoming imperative for optimal context placement problem. Usually optimal context placement problem assumes the popularity of individual content is known or estimated in advance of the algorithm. Applegate et al. [2010] propose a mixed integer program to optimize content distribution using the recent history (the past 7 days) as an estimate of demand for existing contents, and previous demand pattern of similar content for new-released contents. Instead, Zhou and Xu [2002] assume popularity of videos following a Zipf-like distribution and all videos share same peak period. Zipf-like distribution gives the probability of the i-th video is chosen, where video indices are ordered by popularity. This assumption works with simulation but in real-world data, it's impossible to know the order of video index in advance. Instead, Bisdikian and Patel [1995] assume a totally symmetric system: a movie is equally likely to be requested by all locations. However, this is not true for our data. We find that the demand patterns of a movie are heterogeneous among different locations, and thus spatial component should be considered in content allocation problem.

In our study, we propose a dynamic spatial-temporal model to estimate demand curve of each individual movie, rather than the aggregate demand, at different locations. An accurate pre-release demand estimate of each movie at each location helps providers optimize allocation of contents, reduce transmission cost and improve operation efficiency. After its release, the initial stage of a new movie is crucial for promotion. Typically, the providers promote a new movie for the first two weeks after its release and then move on to new titles. With an accurate demand curve model evolving over time, providers would know how long and how much to promote the movie until there is a negative opportunity cost. We utilize functional Principle Component Analysis (fPCA) to extract the main features of each movie's demand curves, which are incorporated with spatial information of each location, and construct a spatial-temporal model. It provides a powerful methodological tool to not only forecast pre-release demand shape of each individual movie at different locations, but also analyze how it evolves as time passes and valuable information is gathered.

The rest of this chapter is organized as follows. In the next section, we describe the data used in this study. In Section 5.3, we identify the geographical effect on demand shape of our data. We then describe functional shape analysis (FSA) in

Section ??. In Section ??, we discuss the results of our estimation and present the model comparison results in Section˜refModelCompVod. Then the spatial implication of movie demand is addressed in Section ??. Finally, we conclude with open questions and future research avenues in Section ??.

5.2 Data

We obtained data from a global provider of multi-platform video services, working with both content and service providers. The data used in our study are historical records of Video-on-Demand system in United States. As video contents are stored at several servers located at different cities and customers are divided into regional networks, each served by one server, the request pattern of each movie from each server represents demand trend of customers in proximity. We use the historical data ranging from October 2010 to June 2011. The data includes a complete record of the user's VCR operation (e.g. stop, pause, rewind, fast-forward, etc.). It also contains the information of the user (e.g.top-box IP), request (e.g. time and location) and content (e.g. complete titles, ID number in library, starting time, end time, etc.). We focus our analysis on each individual movie, and aggregate weekly requests of each movie at different locations throughout the whole available time window. During the record period of 273 days, 3,473,841video requests are posted over 22,017 unique video files. Even if the majority of contents in the library are television shows, which are usually free to view, we focus on movies that customers need to pay to view, which are more profitable.

Our sample consists of all movies that (a) were requested between October 2010 to June 2011; (b) were not free to watch; (c) survived at least 8 weeks. This leads to a total of 115 movies. Our dataset includes weekly requests for all 115 movies for all 11locations. In addition, we use data on a wide range of other characteristics, including production budget, star power, director power, rating, awards, cumulative Box Office sale and the gap between theatrical and Video-on-Demand release. They are obtained from such sources as Internet Movie Database, Box Office Mojo and The Official Academy Awards Database. Details about these variables, their operationalizations, and sources are in Table 5.1. Table 5.2 gives the key summary statistics of movie features we consider in this study.

Table 5.1: Movie Characteristics

Variables	Description	Measure	Source
BO	Cumulative Box Office Sale	Cumulative Box Office Sale in millions	http://www.boxofficemojo.com/
Gap	Time Gap in Weeks between Theatrical and Video-on-Demand Releases	Time Gap in Weeks	http://www.boxofficemojo.com/
Reviews	Consumer Reviews	Movies are rated on a 1-10 scale	http://www.imdb.com/
NumReviews	Total Number of Users Who Have Given a review	Total Number of Users Who Have Given a review	http://www.imdb.com/
Budget	Production Budget	Production Budget in Millions	http://www.boxofficemojo.com/
DirectorGross	Director Lifetime Gross Total	Director Lifetime Gross Total in Millions	http://www.boxofficemojo.com/
DicetorNumMvs	Total Number of Movies Director has Directed	Total Number of Movies Director has Directed	http://www.boxofficemojo.com/
StarGross	Lifetime Gross Total of the Highest Rated Star in the Movie	Lifetime Gross Total of the Highest Rated Star in the Movie in Millions	http://www.boxofficemojo.com/
StarNumMvs	Total Number of Movies of the Star	Total Number of Movies of the Star	http://www.boxofficemojo.com/
AwardsNomi	Number of Awards Nomination	Number of Awards Nomination	Academy Awards Database
AwardsWin	Number of Awards Winning	Number of Awards Winning	Academy Awards Database

Our choice of independent variables is based on previous studies conducted in analysis and prediction of movie's box office sales. As little research has been done in the field of VOD movie demand forecast and VOD is a sequential marketing channel after theaters, we believe studying these movie features will help us find whether the determinants of box office performance will play a role in the sequential release channel.

5.3 Identification of Geographical Effect on Demand Curves

The critical role of spatial factor in influencing customer demand has been documented in many marketing and economic studies. Areas close to one another share climate, history, sociodemographic and economic conditions. Therefore, customer culture, values and taste in proximity tend to be spatially associated (Parker and Tavassoli [2000]). Many articles have illustrated the heterogeneity of customer demand across geographical markets by including spatial factor into statistical models of customer choice (Jank and P.K. [2005]), customer satisfaction(Mittal [Kamakura and Govind]), online auction (Jank and Shmueli [2007]), and cumulative demand (Bronnenberg and Sismeiro [2002]).

The motion picture industry is an emerging field of increased interest to marketing scholars and researchers. A stream of research, addressing spatial movie demand difference, has begun to emerge in the marketing literature, which focus on forecasting at an aggregate level (Krider et al [2005],Eliashberg and Shugan [1997]). Given the short life span of movies, the heterogeneity of demand over time, and the

Table 5.2: Descriptive Statistics

Variables	Min	Max	Mean	Std. Dev.
logBO	0.00	19.59	12.29	7.59
Gap	-21.83	1111.45	61.28	157.26
Reviews	1.30	8.90	5.79	1.41
NumReviews	23	660700	52240	84159
Budget	0.00	160.00	43.73	44.75
DirectorGross	11.44	21.34	18.39	2.15
DirectorNumMvs	1.00	20.00	5.14	4.36
StarGross	11.44	21.99	19.99	1.92
StarNumMvs	1.00	82.00	19.99	1.06
AwardsNomi	0.00	1.00	0.14	0.35
AwardsWin	0.00	1.00	0.06	0.24
Number of Movies		115		
Number of Systems		11		

Note: logBO denotes the natural logarithmic transformation of Box Office Sale high risk of uncertain performance, the managers constantly face the challenge of adopting appropriate strategies for new released movies individually. In this study, we analyze the importance of spatial structure in affecting movie demand in VOD channel by incorporating spatial difference of movie demand in the model.

In our data of study, VOD movies are stored at 11 different systems, located at 11 cities. Each system serves nearby customers. A new movie is available to access simultaneously at all 11 systems once it is released. Thus by separating out demand by locations, we get demand curves of each movie in each system. Figure 5.1 shows the first 8 weeks demand of the movie *Takers* at 11 systems, which are denoted by capital letters. Natural logarithmic transformation is taken to demand for convenience sake and sensitive detailed location information is hidden. As shown in the figure, the demand is location specific and does not share one common pattern across all systems. Systems in top rows obtain relatively high requests of this movie once it is released, which vanish over time. Whereas, for the systems of from "F" to "M", demand does not wake up until after over 6 weeks. A potential manifestation of the above heterogeneous demand over systems could be spatial correlation in demand pattern: customers living close to one system are spatially correlated of the presence of some unobserved common geographic characteristics that affect their movie taste, besides observable factors of movies such as genre, rating, etc. If that would be the case, then we would see significant spatial covariates in later models.

Meanwhile, the various patterns of demand over systems remind us its potential application in content allocation problem. For example, in previous example, it is wise to store movie *Takers* in system A and C all the time, but in System B, D

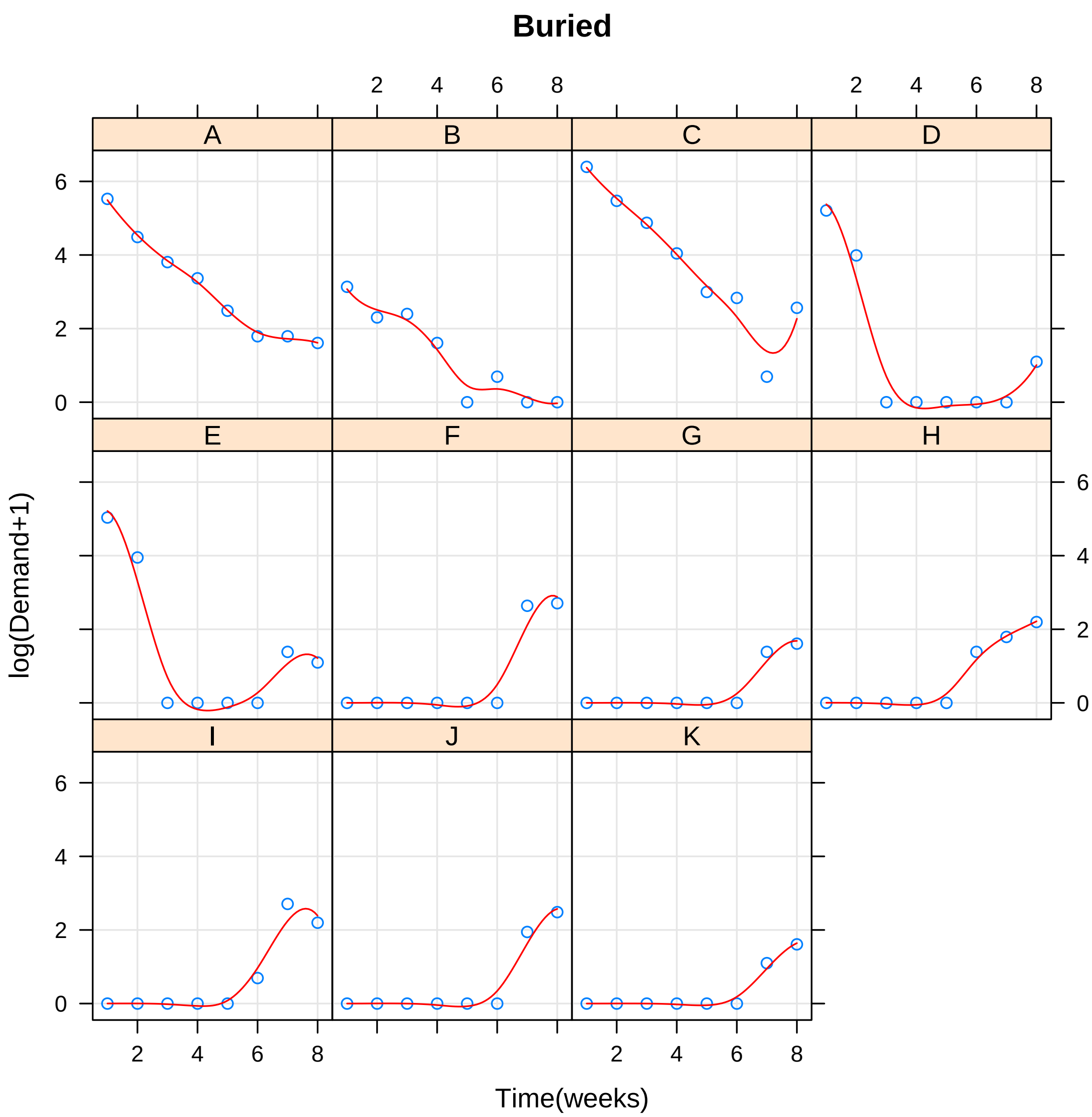


Figure 5.1: Demand Plot of Movie “Takers” of 11 locations

and E only at the beginning and “ending” period (here “ending” refers to the end of time window we look at, not the drop-off date). While for system F to M, it is better to save storage space at the first 6 weeks for other more popular movies and upload *Takers* after week 6 to get better profit with lower storage cost.

To check if demand pattern can be explained by geographic effect and movie's essential features, we estimate a functional spatial-temporal model with geographic variables to predict the demand curves. As shown in Figure 5.1, the demand curves of even one movie vary dramatically from systems. In some systems, demand decay gradually, while in others, demand picks up at the end our time window. Our goal is to characterize these patterns across movies and systems.

In the next section, we will introduce functional shape analysis as a method to extract main pattern shapes underlying demand curves across movies and systems.

5.4 Functional Data Analysis with Spatial Component

Functional Data Analysis (FDA) focuses on a sample of *functional observations*, e.g. curves, and treats the observed curves as the units of observation, such as online virtual stock market's history (Foutz and Jank [2010]), online auction price (Jank and Shmueli [2008]), market penetration (James and Tellis [2009]). This is in contrast to classical statistics where the focus is a set of discrete data vectors. The method of FDA was introduced by Rao [1958] for growth curves. Many theoretical properties have been developed by Ramsay and Silverman [1997] and Silverman [1996]. Recently, substantial classical statistical models have been generalized to the functional structure. Hastie and Sugar [2000] proposed functional principal components analysis for sparsely sampled curves. The case of irregular grids was also studied by Staniswalis and Lee [1998]. More recent research includes curve clustering and classification (James and Sugar [2003], Tarpey and Kinateder [2003], and James and Hastie [2001]), functional regression (Cuevas Febrero and Fraiman [2002]), functional generalized linear model (James [2002]), functional ANOVA (Guo [2002]) and time series analysis of functional data (Ocana and Valderrama [1999]). While this list is far from complete, it exemplifies some of the current methodological improvement in this merging field.

One rather under-explored area of functional models is functional data analysis with spatial components. As discussed before, many studies have shown the critical role of geographic effect on demand heterogeneity. However, none of these studies use functional method by considering subject demand as one single continuous curve and then to check if and how these set of curves differ with spatial components. In this study, we focus on exploring how these shapes are related to geographical effect, e.g. system.

Due to limitations in human perceptions and measurement capabilities, we can record only discrete observations of these curves. Thus the first step is to recover the underlying continuous functional objects by smoothing techniques. Smoothing can eliminate noise from observed raw data and allow important patterns to stand out. In this study, we smooth demand curves using penalized smoothing splines.

The red curves in Figure 5.1 are the smoothed demand from the observed demand denoted by dots. We can see that smoothed demand curves capture the main evolving patterns while smoothing out unusual and noisy demand spikes. Figure 5.2 shows the smoothed demand curves of four movies in all the 11 systems. The demand curves are considerably heterogeneous across movies and systems. Even for one movie, the demand curves are various across systems. These different shapes may contain important information about demand of movie at each system and our goal is to extract the most common demand shapes across all movies and systems and then use them for analyzing spatial relationship of movie demand and forecasting. We accomplish this via functional principal component analysis (fPCA). fPCA is a generalization of ordinary PCA, which projects the original curves to a new space of orthogonal dimensions to capture the primary features of original curves.

We apply fPCA to the smoothed demand curves displayed in Figure 5.2. The scree plot of Figure 5.3 shows the percentage of variance explained by each functional principal component (fPC). It shows that the first three fPC's explain more than 95.6% of the total variation in the data. We thus retain the first three fPC's for further analysis.

Figure 5.4 displays the first three fPC's of demand. We can see that each fPC captures different aspect of a movie's demand curve. To illustrate these fPC's, we introduce the *principal component scores* (fPCS) computed simultaneously with the corresponding fPC's. *Principal component score* is the inner product of its demand pattern and the corresponding fPC. For example, movie i's first fPCS, $fPCS1_i$ of demand curve is the inner product of its demand curve, $y_i = [y_{i1}, \cdots, y_{ip}]$, and the first fPC of demand curve, $fPC1 = [p_{11}, \cdots, p_{1p}]$; i.e., $fPCS1_i = y_{i1}p_{11} + \cdots + y_{ip}p_{1p}$.

The first fPC (solid black line in Figure **??**) is slightly declining and positive

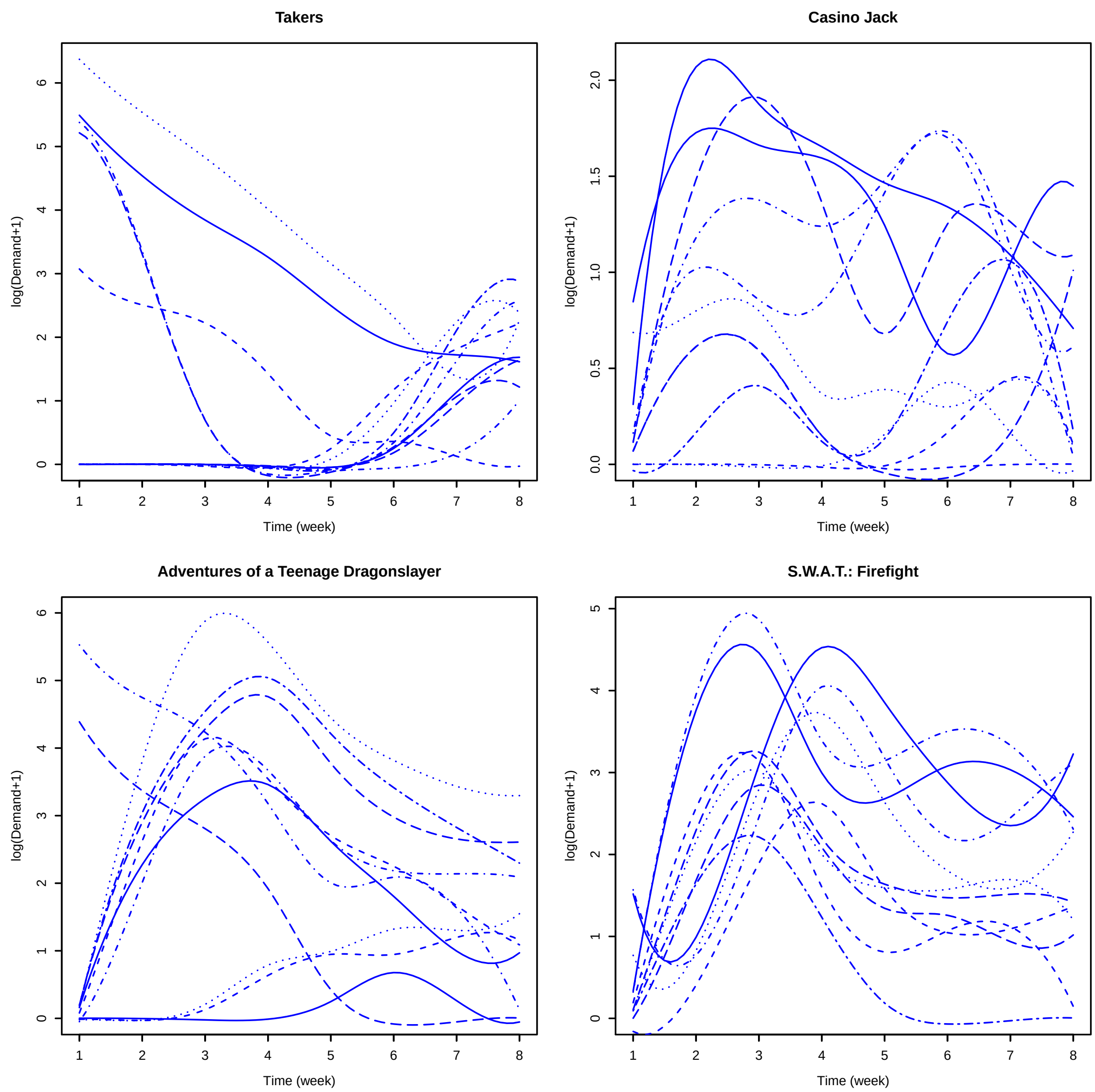


Figure 5.2: Smoothed Movie Demand Curves of Four Movies at 11 Systems

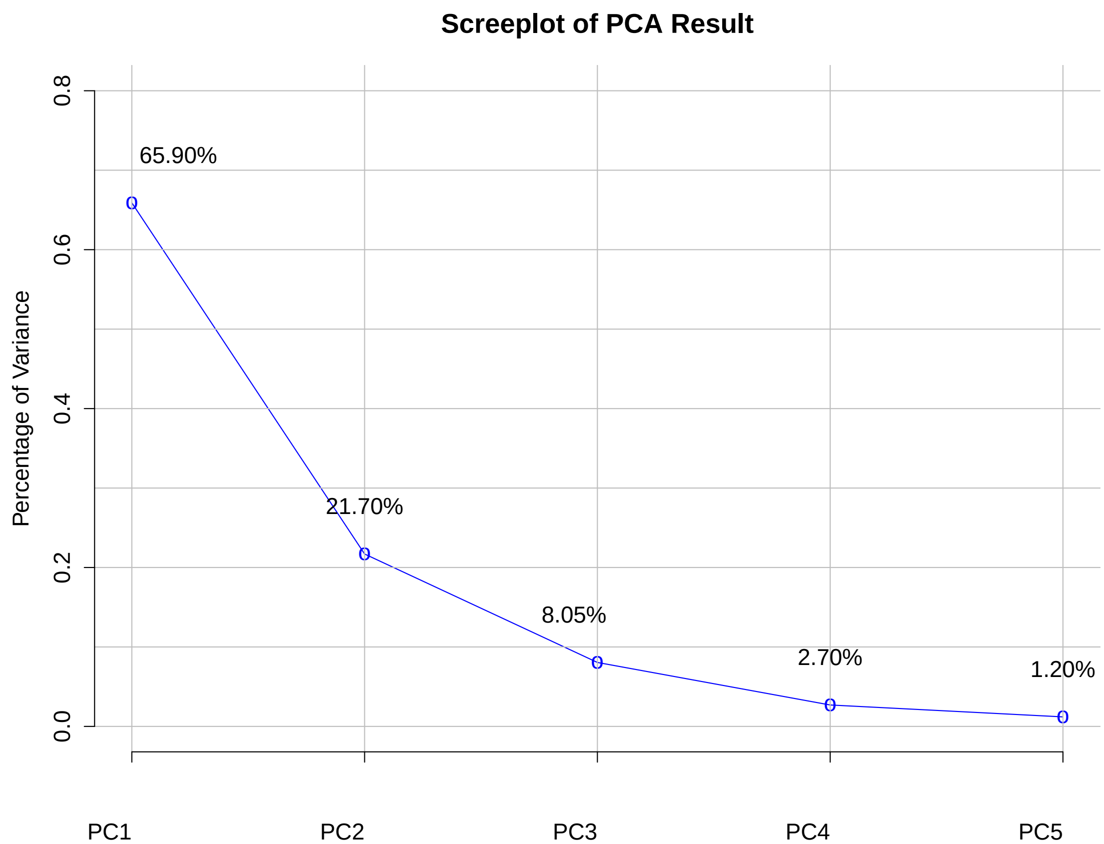


Figure 5.3: Scree Plot of PCA

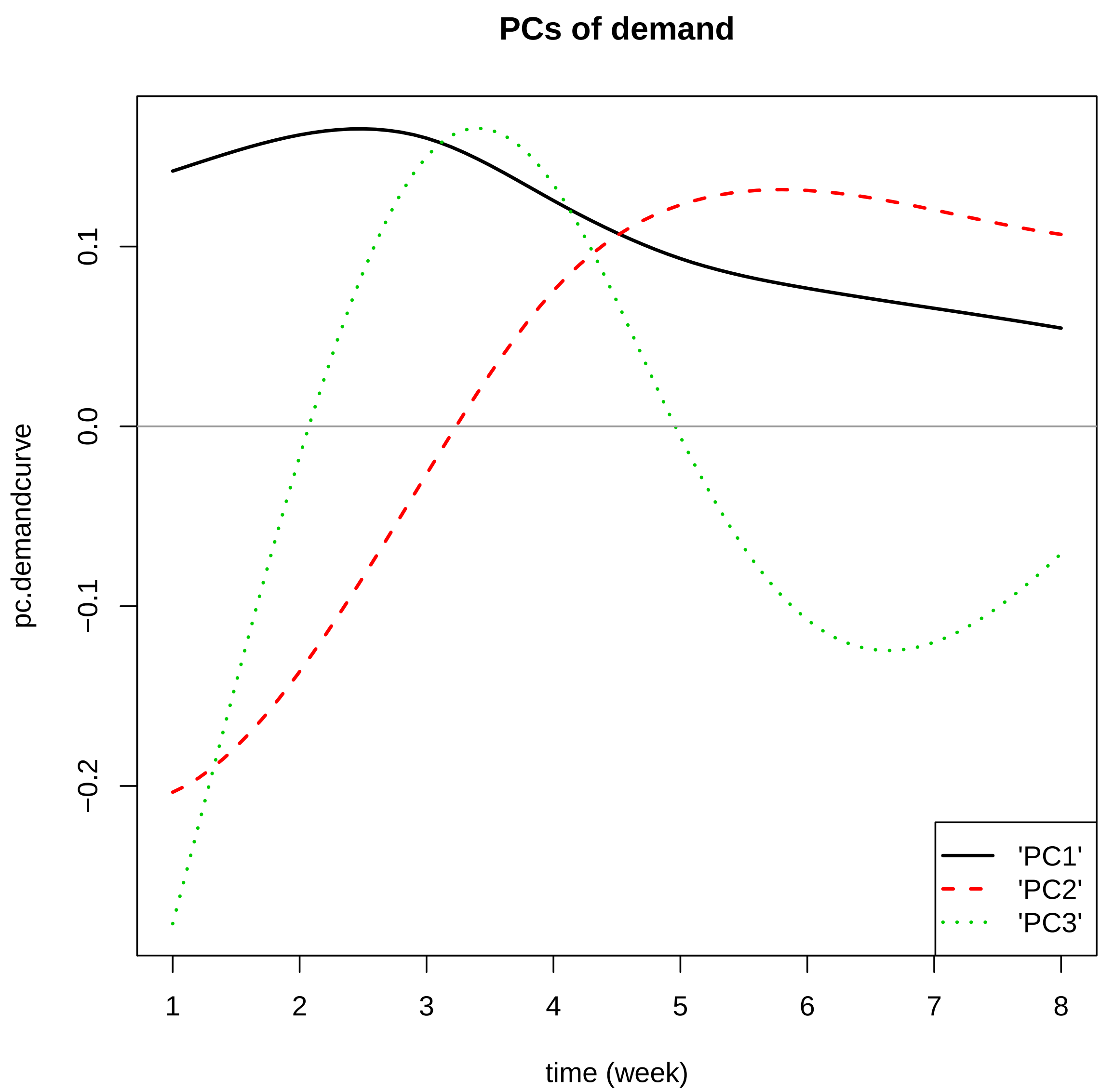


Figure 5.4: First Three Functional Principal Components of Demand

over the entire time period. In that sense, the first fPC measures a movie's weighted demand-average over the first 8 release weeks, putting more weights on movies that perform well once released. That is, a movie with a relatively high (low) 8-week average has a positive (negative) fPCS1. Take Figure 5.5 for illustration. The top panel in Figure 5.5 are demand curves of three movies at one system. The bottom panel shows the corresponding functional principal component *score* (fPCS). fPCS1 for *Little Fockers* is much larger compared with to the other two movies. By looking at the top panel, we can find an explanation: the average (log-) demand for *Little Fockers* is apparently much larger compared to the others. In fact, the 8-week average (log-) demand is 5.75 for *Little Fockers*, 2.39 for *The Social Network*, 2.27 for *3 Backyards* in System-Des Moines, and 1.32 across all movies and systems. The 8-week average (log-) demand of these three movies are both greater than the overall average, thus all fPCS1s in the bottom panel are positive.

Analogously, we can interpret fPC2 as characterizing movies that develop demand only late. An example of such a "sleeper" is *The Social Network*, its second fPCS (bottom middle panel in Figure 5.5) is highest among all three movies. fPC3 captures movies that have relatively high demand at early stage which drops down at the end. An example of such an "early bird" is *3 Backyards* in the bottom right panel of Figure 5.5.

It should reinforce the point that, fPC's are common across all demand curves, while fPCS is specific for each demand curve with its corresponding fPC, which will be used later in our forecasting model. We also want to emphasize that fPC's are not the "real" demand curves; in fact, they reflect three basic demand shapes: fPC1

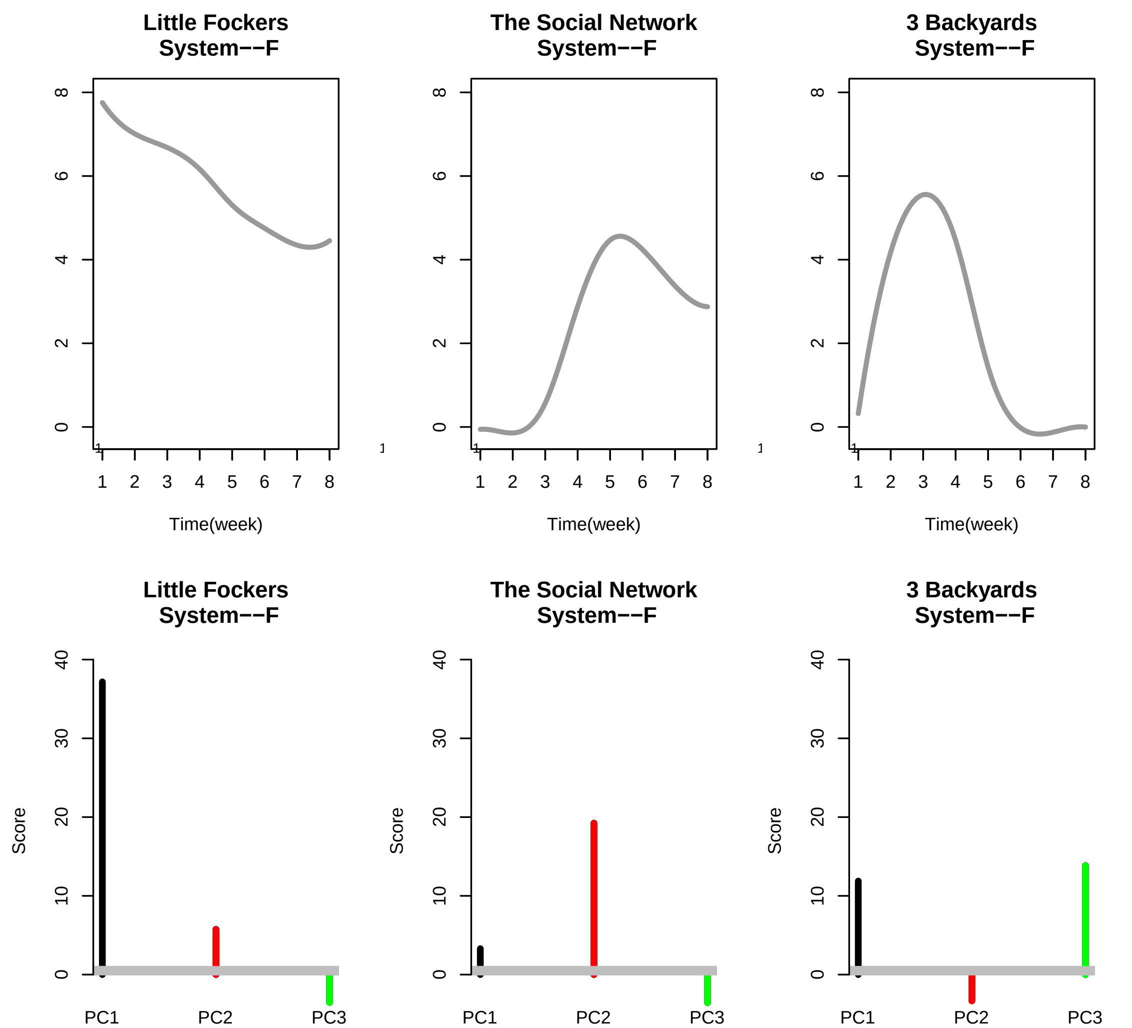


Figure 5.5: Illustration of Some Demand Curves

denotes movies with high average demand over the entire 8-week period, with special emphasis on movies that develop demand once released ("hits'); fPC2 denotes movies that develop demand only late ("sleepers"); fPC3 denotes movies with high demand at early stage but low demand at the very beginning and end ("early bird"). In summary, we can consider fPC1-3 as three orthogonal "basis" for a wide variety of demand curves, while fPCS are the coefficients of each demand curve corresponding to basis.

5.5 Parameter Estimates

In Table 5.3 and 5.4, we present the estimates of the Spat-Tem Model which is also the first step of the proposed Dynamic Spat-Tem Model. We see that the effect of movie features is heterogeneous. For instance, "Documentary" has a large positive influence on fPCS2, but negative on fPCS1 & 3.

To have a better understanding how these factors variously affect movie's performance, we propose plausible explanations. Recall that, fPC1 represents overall average demand of a movie, compared with other movies. It slightly emphasizes on early information. The significant negative effect of "Documentary" indicates a poor early performance of movie in such category of genre. fPC2 denote movies of "sleepers". A positive fPCS2 estimate of "Documentary" is a strong indicator of potential success. fPC3 points to 'early bird' movies. The negative effect of "Documentary" on demand shows a late spurt trend of such movies, which is consistent with the trend discovered by previous two fPCs.

Table 5.3: Summary of Estimation Results of Spat-Temporal Model

	Coefficient (Standard Error)		
Parameters	PC1	PC2	PC3
Action	-5.03 (0.68)***	-1.79 (5.39)***	————
Adventure	-0.16 (0.80)	-0.72 (0.55)	-0.42 (0.35)
Animation	-0.94 (1.07)	2.16 (0.74) ***	-0.85 (0.50)
Comedy	0.92 (0.63)	-0.55 (0.56)	-0.77 (0.30)***
Documentary	-17.24 (3.94)***	6.40 (2.38)***	-9.20 (1.49)***
Drama	————	————	-1.42 (0.26)***
Family	————	4.56 (0.73)***	0.72 (0.46)
Fiction	-5.90 (1.32)***	-2.75 (0.92)***	-0.57 (0.60)
Foreign	2.70 (2.20)	2.32 (1.52)	————
Horror	-2.97 (0.97)***	4.49 (0.74)***	-1.39 (0.41)***
Suspense	————	3.88 (1.06)***	-3.30 (0.66)***
Thriller	————	1.46 (0.51)***	————
Romance	-4.70 (0.96)***	————	1.53 (0.41)***
Western	————	-4.65 (1.10)***	————
Rating(PG)	-9.83 (2.38)***	-0.49 (1.48)	-1.95 (1.07)
Rating(PG-13)	-6.66 (2.50)***	2.12 (1.62)	-1.63 (1.21)
Rating(R)	-7.51 (2.48)***	3.84 (1.65)**	————
Reviews	-1.93 (0.37)***	0.73 (0.18)***	————
NumReviews	1.11 (0.33)***	————	————
Budget	1.20 (0.36)***	0.024 (0.006)***	————
BO	————	————	-0.46 (0.06)***
DirectorGross	0.54 (0.16)***	2.460e-09 (5.058e-10)***	2.156e-09 (4.898e-10)***
DirectorNumMvs	————	————	-0.10 (0.04)***
StarNumMvs	0.51 (0.27)	-0.029 (0.02)	0.03 (0.01)**
StarGross	————	-6.268e-10 (3.474e-10)	————
AwardsNomi	5.39 (0.86)***	————	0.83 (0.03)***
AwardsWin	2.79 (1.20)***	————	————
Gap	-0.03 (0.00)***	————	————

Note: System Dummies used in the model are reported separately. *** p ¡ .01, ** p ¡ .05

Table 5.4: Spatial Estimates by PCs of Spat-Temporal Model

	Coefficient		
Parameters	PC1	PC2	PC3
System A	-5.83***	-2.13***	0.55
System B	——	-2.13***	——
System C	7.00***	——	-1.10***
System D	-2.12***	-0.47	——
System E	-1.48	——	——
System F	8.84***	2.80***	-1.94***
System G	-1.84**	——	——
System H	3.09***	1.31**	-0.73**
System I	4.02***	1.33**	-0.90**
System J	3.95***	1.59***	-0.67
System K	-3.85***	——	0.50

Note: *** p¡.01, ** p¡.05

The MPAA rating G (General Audience) is the base category of MPAA rating in the model. All other three ratings (PG,PG-13, and R) are assessed against rating G. The small p-value of all three ratings in the PC1 column means that the MPAA rating influences the overall performance of movies mostly (fPCS1) and movies with MPAA rating of “G”, those with no restrictions on admittance,tend to get more demand relative to other ratings. That may stem from the fact that general audience movies are more likely to appeal to a broader audience and thereby get higher demand. This finding is in line with previous research, which has consistently shown the positive effect of MPAA rating “G” on Box Office Sales of movies (Prag and Casavant [1994],Simonoff and Sparrow [2000] and Medved [1992]). Therefore, this effect of MPAA rating on movie’s performance is inherited from the first distribution channel (theaters) to the sequential channel (Video-on-Demand).

Consumer review encompasses valuable interpersonal communications between consumers about movie. Volume of Reviews represents the awareness of the movie and has a significant influence on its performance in theater (Liu [2006]). Extant research have shown that the volume of consumer reviews positively affect Box Office Sale of movies (Gu and Whinston [2008],Zhang and Awad [2007] and Liu [2006]). Our result suggests the importance of awareness effect to movie’s performance in the sequential VOD channel. Volume of Reviews helps in the early stage (positive effect of “Number of Reviews” on fPC1), but the the stimulating effect of review incubates until late half period (estimate of “Reviews” is positive for fPC2 and negative for fPC1). This agrees with the analysis of Liu [2006].

The impact of budget, star power and director power on Box Office Sales of movies is inherited to the movie's demand in VOD channel. It shows that higher production budget movies have relative higher average demand (fPC1). In addition, our study also finds strong evidence that both star and director power has a significant effect on early releasing demand. It supports the findings of previous studies on box office sales (Foutz and Jank [2010], Elberse [2007], Chang and Ki [2005] and Desai and Basuroy [2005]).

Presence of an award nominee or award winner would reflect the quality and generate larger revenues for the movies. However, previous studies find that this factor has no relevance to a movie's performance in theater (Simonoff and Sparrow [2000]; Basuroy et al. [2003]). Besides, by the time awards are announced, most movies have finished playing in theaters. Because of the release delay in VOD channel, awards have usually been announced before a movie is available in VOD channel, and would be an efficient covariate in explaining movie demand. The positive estimates of "Awards" covariates for fPC1 (the general average performance) prove that quality reflected by winning (being nominated for) awards is significantly related to the movie's demand as long as the movie has already been in release in the channel when awards results are announced.

Based on an important characteristic of sequentially distributed products-demand in one channel of distribution system may provide a good indication of demand in the next channel of the sequential distribution channel, Box Office Sales of movies would be appropriate indicator of movie demand in VOD market. The negative estimate of "Box Office Sale" for fPC3, together with the high start value (first week when it's released in VOD market) of negative of fPC3 shape, shows us that movies with good performance in theater tend to continue to be a blockbuster in the VOD channel.

We also see that movies released in VOD channel shortly after in theaters display larger demand (negative "gap" estimate of fPC1). Such effect may be explained by a recency effect(Ross and Simonson [1991]). Movies shown in public(theaters, advertisement, newspaper, etc.) most recently are most salient and attract most attention. However, when sequential movie release is delayed substantially from the release in theater, advertising and publicity effects created in the first channel will have largely dissipated.

5.6 Model Comparison

As we've shown in the previous section, these three PCs capture the most significant variations in all the demand patterns. We then link movies characteristics, along with spatial information, to these three key shapes, or more precisely, their corresponding movie-specific fPCSs.

We compare model fit across proposed model and a set of alternate models, based on the Bayesian Information Criteria (BIC) (Table 5.5(a)). Movie Model is a basic regression model with "raw" demand, considering demand curves as discrete points and disregarding the spatial effect. Current weeks' demands are regressed on previous weeks' demands together with movie features: genre, MPAA ratings, and other features described in Table 5.1. We also include the total box office sale and the gap between the movie's close date in theater and released date in VOD channel. Spatial Model includes both movie features and spatial information about which system the purchase is made in. Similarly, spatial model focuses on the discrete demand data, and get future demand forecast based on previous demand history. Both of these two Non-Functional models yield the lowest model fits among all models.

Movie-Temporal Model links the key shapes of eight-week demand curves and movie features, and predicts eight weeks' demand shape at the very beginning of the released data, when no demand information is available. Augmenting the movie features with location information (Spat-Temporal Model) improves the model fit. In light of optimal resource allocation and efficient pricing strategy, such improvement is substantial and managerially practical for cable companies in prerelease marketing planning (e.g., media purchase, package contracts, and promotion).

As movie is released in VOD channel,sales data is readily available to us. Dynamic Spatial Temporal Model (Dynamic Spat-Tem Model) takes full advantage of all available demand information by updating the model with key shapes of previous demand curves. For instance, a movie is newly released on VOD channel four weeks ago and the key shapes of these four weeks demand are ready to be extracted from prevous sales record. Dynamic Spatial Temporal Model considers these shapes as covariates and, along with movie features and location inforamtion, forecasts future weeks' demand curve. Fully usage of resource leads to the best model fit. Actually, Spat-Temporal Model is included within Dyanamic Spat-Tem Model. Before release, we use Spat-Temporal Model to get the forecast of all eight weeks' demand simutaneously; once sales data is available, we update the model with the key shapes of those sales and get more accurate prediction of future demand. Updating model with all previous available demand data allows managers to identify indicators of a potentially successful movie early and dynamically.

Table 5.5: Comparison of Models

(a) Bayesian Information Criteria

Specification	Description	BIC
Movie Model	Only Movie Features,Non-Functional	9887.42
Spatial Model	Both Movie and Location Info., Non-Functional	9876.77
Movie-Temporal Model	Only Movie Features, Functional	6876.41
Spat-Temporal Model	Both Movie and Location Info., Functional	6752.69
Dynamic Spat-Tem Model	Dynamic Spat-Tem Model	**5111.64**

(b) Predictive Performance

Specification	Description	MAD	MSE
Movie Model	Only Movie Features,Non-Functional	110.93	5746.35
Spatial Model	Both Movie and Location Info., Non-Functiona	23.59	70.18
Movie-Temporal Model	Only Movie Features, Functional	24.12	197.80
Spat-Temporal Model	Both Movie and Location Info., Functional	1.55	2.03
Dynamic Spat-Tem Model	Dynamic Spat-Tem Model	**0.96**	**1.25**

We also compare the predictive performance of both proposed and alternate models and compute the mean absolute deviation (MAD) and mean square error (MSE) between predictions of demand from the model and the actual demand (Table 5.5(b)). The MAD and MSE values in both sections are the lowest for the

(c) Model Forecast Performance		
Specification	Description	MAPE
Movie Model	Only Movie Features,Non-Functional	7.42
Spatial Model	Both Movie and Location Info., Non-Functiona	6.84
Movie-Temporal Model	Only Movie Features, Functional	2.08
Spat-Temporal Model	Both Movie and Location Info., Functional	1.83
Dynamic Spat-Tem Model	Dynamic Spat-Tem Model	**1.11**

Dynamic Spat-Tem Model. Not surprisingly, managers can derive increasingly accurate forecasts as more information becomes available over time.

We also provide out-of-sample forecast performance based on mean absolute percentage error (MAPE) for all model forecasts, using movie-by-movie, full cross-validation (Table 5.5 (c)). That is, we hold out one movie-location demand curve at a time, estimate the model on the remaining demand curves, and then use the estimated parameters to forecast the held-out movie's demand curve. The evidence clearly supports the predictive accuracy of our proposed Dynamic Spat-Tem Model.

Table 5.6 compares the performance of dynamic spat-tem models in different stages. We calculate the PC-Scores from historical demand shapes and include them into the model as functional predictors. For instance, "Dynamic Spat-Tem Model for Wk5-8" includes all movie features described in table 5.2, spatial variables and PC-Scores of demand shape from week 1 to week 4 to predict PC-Scores of demand shape from week 5 to week 8. Overall, we find that there is a substantial improvement in both model fit and predictive accuracy when history movie demand is considered, consistent with the view that the movie demand shape is dynamic. Meanwhile, it

Models	BIC	MSE	MAD	MAPE
Dynamic Spat-Tem Model for Wk1-8	6773.43	3.44	1.56	0.64
Dynamic Spat-Tem Model for Wk3-8	6412.65	2.40	1.30	0.51
Dynamic Spat-Tem Model for Wk5-8	4888.11	2.04	1.21	0.36
Dynamic Spat-Tem Model for Wk7-8	**2428.13**	**1.68**	**1.01**	**0.27**

Table 5.6: Comparison of Dynamic Model Performance

Note: Dynamic Spat-Tem Model for Wk1-8 is identical with Spat-Temporal Model; Dynamic Spat-Tem Model for Wk3-8 uses movie demand history (from week 1 to week 2) to predict movie demand shape of week 3 to week 8 and so on.

shows that the more abundance in history demand information incorporated, the better model performs. Therefore, history demand evolution is an important factor in effecting future demand performance.

To illustrate spatial impact on model performance, we compare forecast performance of proposed dynamic model in table 5.7, which provides information in the forecast differences across systems and stages. The heterogeneity of dynamic model performance emphasizes the significance of including spatial components when analyzing movie demand in VOD channel.

5.7 Spatial Implication of Movie Demand

To further add to the understanding of the geographic effect on movie demand in terms of demand shapes' Principle Component Scores (PC-Scores) (Section 5.3

Table 5.7: Comparison of Dynamic Model Across Locations

	MAPE of Dynamic Model			
System	Wk1-8	Wk3-8	Wk5-8	Wk7-8
A	**0.56**	**0.43**	**0.40**	**0.32**
B	0.77	0.63	0.53	0.48
C	1.06	0.94	0.70	0.59
D	0.77	0.64	0.50	0.49
E	0.78	0.63	0.49	0.49
F	0.79	0.67	0.52	0.50
G	1.05	0.91	0.70	0.59
H	1.06	0.90	0.71	0.63
I	0.69	0.58	0.45	0.46
J	0.84	0.71	0.51	0.53
K	0.92	0.76	0.58	0.54

Note: Values in bold highlight system with the best forecast accuracy.

and 5.4), we present the locations of the thirteen systems in Figure 5.6, Figure 5.7 and Figure 5.8. The bars with different heights indicate average PC-Score with respect to each system. Higher bar indicates a larger absolute average PC-Scores for each location; furthermore, dark (light) gray color indicates positive (negative) value. For instance, a dark high bar represents large positive average PC-Score for that location; while a light high bar represents the opposite (very small negative average).

As can be seen from the figure, PC-Scores are highly correlated with geographical locations. Such a geographic demand map can be helpful to managers in identifying demand pattern for each system. For northern area, we find the highest movie demand, indicated by the high dark bars; while, on the western and eastern coasts, the movie demand is uniformly low. In general, movie performance is better in the northern part of United States. Furthermore, northern area also have the highest PC2-Score (Figure 5.7) and lowest PC3-Score(Figure 5.8), which indicates a potential increasing trend in the evolving movie demand.

Given our estimates in Table 5.3, we can reconsider the resource allocation decision across the 11systems. For example, we estimate that over the 8-week period of our sample, Genre "Horror" has negative fPC3 ("early bird") estimate, but positive fPC2("sleeper") estimate; In other words, "Horror" movies tend to be potential successful movies. On the other hand, "Northern" Systems have the same sign of estimates for fPC2 & 3 as "Horror" , which amplifies the late spurt trend. Therefore, a "Horror" movie in northern area is a strong indication of potential successful movies. These indicators may be used as rules of thumb for managerial

Figure 5.6: Average PC1-Scores for Each System

Figure 5.7: Average PC2-Scores for Each System

Figure 5.8: Average PC3-Scores for Each System

decision, especially pre-release marketing planning.

After the first weeks of release in VOD channel, movie sales curve reveals initial demand and can be used as significant explanatory variable. Table 1 shows us that incorporating with key shapes of previous demand remarkably improves predictive accuracy of spatial temporal model, which makes sense as we include more information in the model. An accurate and dynamic forecast of future weeks demand at each system may help managers determine the optimal drop-off/promotion date of movies at certain system and save valuable resources for other more profitable movies.

5.8 Conclusion

We have developed a dynamic spatial temporal model that considers the spatial effects of movie demand in VOD channel. The model considers the evolution of movie demand over time and for different locations, as well as movie features (genre, MPAA rating, reviews, star and director power, box office sale, etc.). We believe that this is a first attempt to help managers of cable companies allocate their resources across different systems, save transmission cost and determine drop-off and promotion timing even before the contents are released.

The results show that modeling different demand curves at different locations yields better insights and greatly improves forecast accuracy, relative to models that aggregates the demand at all locations. More importantly, our method produces early pre-release forecasts that are most valuable to managerial decision making (e.g, optimal context allocation problem and marketing).

Meanwhile, our model can be dynamically updated with early demand information after content release and further improved forecast accuracy helps managers get a clearer idea about the evolution of demand and refresh the decision making process. The main value of our model comes from the fact that we have associated demand with spatial effect and so have provided a way for managers to use these estimates and insights in making allocation and marketing decisions.

There are a few limitations of our study. Our research is highly exploratory and the principle aim is developing a more accurate forecast models for better managerial decision making. It primarily concentrates on finding the spatial effect on movie demand in VOD channel, and its application in marketing decision making process. Explanation of this relationship among location, movie preference and demand is not within the scope of this study. Further research will help establish theoretical support of spatial relevance to movie demand.

In this study, we apply functional data analysis to demand of the first 8 weeks. Truncating of the demand histories makes our results conservative in demonstrating the actual predictive accuracy of our model as potentially valuable information has been neglected by the right censoring process. Our model can be extended to account for unbalanced sample. This is a challenging problem since the exact available time period of each movie is various and one needs to find a way to incorporate variousness of movie's available time period. Other potential direction for future researches are to consider the interaction between movie features and different locations and to extend our model to develop an optimal content allocation model. These are some important issues and our study presents a possible starting point to address them.

Chapter 6: Proof of Lemma, Proposition, Theorem and auxiliary results

6.1 Proof of Lemma 3.1.1

Lemma 3.3.1 $n\|\tilde{\psi}_j - \psi_j\|^2 = \sum_{k:k\neq j}(\theta_j - \theta_k)^{-2}(\int Z\psi_j\psi_k)^2 + O_p(n^{-1/2})$

Proof. As in (3.1), we can write: $\hat{\psi}_j - \psi_j = \sum_{k-1}^{\infty}(a_{jk} - \delta_{jk})\psi_k$,where δ is the Kronecker δ. Then,

$$
\begin{aligned}
\|\hat{\psi}_j - \psi_j\|^2 &= \int(\hat{\psi}_j - \psi_j)^2 dw(t) \\
&= \int\left[\sum_{k=1}^{\infty}(a_{jk} - \delta_{jk})\psi_k\right]^2 dw(t) \\
&= \int\left[\sum_{k=1}^{\infty}(a_{jk} - \delta_{jk})^2\psi_k^2\right. \\
&\quad \left. + \sum_{k_1\neq k_2} 2(a_{jk_1} - \delta_{jk_1})(a_{jk_2} - \delta_{jk_2})\psi_{k_1}\psi_{k_2}\right] dw(t) \\
&= \sum_{k=1}^{\infty}(a_{jk} - \delta_{jk})^2 \int \psi_k^2(t)dw(t) \\
&\quad +2\sum_{k_1\neq k_2}(a_{jk_1} - \delta_{jk_1})(a_{jk_2} - \delta_{jk_2})\int \psi_{k_1}\psi_{k_2}dw(t)
\end{aligned}
$$

Because $\int \psi_k^2 dw(t) = 1$ and $\int \psi_{k_1}\psi_{k_2}dw(t) = 0$ when $k_1 \neq k_2$,

$$
\begin{aligned}
n\|\hat{\psi}_j - \psi_j\|^2 &= n\sum_{k=1}^{\infty}(a_{jk} - \delta_{jk})^2 \\
&= n\sum_{k:k\neq j} a_{jk}^2 + n(a_{jj} - 1)^2.
\end{aligned}
$$

We have from (3.2)

$$
\begin{aligned}
n(a_{jj}-1)^2 &= n\Big[-\frac{1}{2}n^{-1}\sum_{k:k\neq j}(\theta_j-\theta_k)^{-2}\Big(\int Z\psi_j\psi_k\Big)^2+O_p(n^{-3/2})\Big]^2 \\
&= n\Big\{\frac{1}{4}n^{-2}\Big[\sum_{k:k\neq j}(\theta_j-\theta_k)^{-2}\Big(\int Z\psi_j\psi_k\Big)^2\Big]^2+\Big[O_p(n^{-3/2})\Big]^2 \\
&\quad -n^{-1}\sum_{k:k\neq j}(\theta_j-\theta_k)^{-2}\Big(\int Z\psi_j\psi_k\Big)^2 O_p(n^{-3/2})\Big\} \\
&= \frac{1}{4}n^{-1}\Big[\sum_{k:k\neq j}(\theta_j-\theta_k)^{-2}\Big(\int Z\psi_j\psi_k\Big)^2\Big]^2+O_p(n^{-2}) \\
&\quad -\sum_{k:k\neq j}(\theta_j-\theta_k)^{-2}(\int Z\psi_j\psi_k\Big)^2 O_p(n^{-3/2}) \\
&= O_p(n^{-1}).
\end{aligned}
$$

Furthermore, from (3.3)

$$
\begin{aligned}
n\sum_{k:k\neq j}(a_{jk})^2 &= n\sum_{k:k\neq j}\Big\{n^{-1/2}(\theta_j-\theta_k)^{-1}\int Z\psi_j\psi_k \\
&\quad +n^{-1}\Big[(\theta_j-\theta_k)^{-1}\sum_{l:l\neq j}(\theta_j-\theta_l)^{-1}\Big(\int Z\psi_j\psi_l\Big)\Big(\int Z\psi_k\psi_l\Big) \\
&\quad -(\theta_j-\theta_k)^{-2}\Big(\int Z\psi_j\psi_j\Big)\Big(\int Z\psi_j\psi_k\Big)\Big]+O_p(n^{-3/2})\Big\}^2 \\
&= n\sum_{k:k\neq j}\Big\{n^{-1}(\theta_j-\theta_k)^{-2}(\int Z\psi_j\psi_k)^2 \\
&\quad +n^{-2}\Big[(\theta_j-\theta_k)^{-1}\sum_{l:l\neq j}(\theta_j-\theta_l)^{-1}\Big(\int Z\psi_j\psi_l\Big)\Big(\int Z\psi_k\psi_l\Big) \\
&\quad -(\theta_j-\theta_k)^{-2}\Big(\int Z\psi_j\psi_j\Big)\Big(\int Z\psi_j\psi_k\Big)\Big]^2+O_p(n^{-3}) \\
&\quad +2n^{-3/2}(\theta_j-\theta_k)^{-1}\int Z\psi_j\psi_k\Big[(\theta_j-\theta_k)^{-1}\sum_{l:l\neq j}(\theta_j-\theta_l)^{-1} \\
&\quad \Big(\int Z\psi_j\psi_l\Big)\Big(\int Z\psi_k\psi_l\Big)-(\theta_j-\theta_k)^{-2}\Big(\int Z\psi_j\psi_j\Big)\Big(\int Z\psi_j\psi_k\Big)\Big] \\
&\quad +2n^{-1/2}(\theta_j-\theta_k)^{-1}\int Z\psi_j\psi_k O_p(n^{-3/2})+ \\
&\quad +2n^{-1}\Big[(\theta_j-\theta_k)^{-1}\sum_{l:l\neq j}(\theta_j-\theta_l)^{-1}\Big(\int Z\psi_j\psi_l\Big)\Big(\int Z\psi_k\psi_l\Big)
\end{aligned}
$$

$$-(\theta_j-\theta_k)^{-2}\Big(\int Z\psi_j\psi_j\Big)\Big(\int Z\psi_j\psi_k\Big)\Big]O_p(n^{-3/2})\Big\}$$

$$= \sum_{k:k\neq j}(\theta_j-\theta_k)^{-2}(\int Z\psi_j\psi_k)^2+O_p(n^{-1})+O_p(n^{-1/2})_+O_p(n^{-3/2})$$

Therefore we have

$$n\|\hat{\psi}_j-\psi_j\|^2=\sum_{k:k\neq j}(\theta_j-\theta_k)^{-2}(\int Z\psi_j\psi_k)^2+O_p(n^{-1/2})$$

□

6.2 Proof of Theorem 3.2.1

Theorem 3.2.1 For fixed j and each realization of χ we have

$$E[(\hat{\beta}_j-\beta_j)^2|\chi]=O_p(p^2/n).$$

Proof. From Muller and Stadtmuller [2005] , we have

$$\|\sqrt{n}(\hat{\beta}-\beta)-(\frac{D^TD}{n})^{-1}\frac{U(\beta)}{\sqrt{n}}\|_2^2\to 0$$

where "$S_n\sim T_n$" means that the ratio of the random variables S_n and T_n converges to 1 as $n\to\infty$. Here $U(\beta)$ is the score function of the truncated model and D is the $n\times p$ matrix:

$$D=D_{n\times p}=\Big(g'(\eta_i)\epsilon_k^{(i)}/\sigma(\mu_i)\Big)_{1\leq i\leq n,1\leq k\leq p}.$$

Let ξ_{jk} be the jk entry of matrix $[(1/n)E(D^TD)]^{-1}$. Then,

$$(\hat{\beta}_j-\beta_j)\sim\frac{1}{n}\sum_{i=1}^n\sum_{k=1}^p\frac{(y_i-g(\eta_i))g'^2(\eta_i)}{\sigma^2(\mu_i)}\epsilon_k^{(i)}\xi_{jk}.$$

We have

$$
\begin{aligned}
E[(\hat{\beta}_j-\beta_j)^2|\chi] &\sim \frac{1}{n^2}\sum_{i_1=1}^{n}\sum_{i_2=1}^{n}\sum_{k_1=1}^{p}\sum_{k_1=1}^{p} \\
&\quad E\Bigg[\frac{(y_{i_1}-g(\eta_{i_1}))g'^2(\eta_{i_1})}{\sigma^2(\mu_{i_1})}\frac{(y_{i_2}-g(\eta_{i_2}))g'^2(\eta_{i_2})}{\sigma^2(\mu_{i_2})}\epsilon_{k_1}^{i_1}\epsilon_{k_1}^{i_1}\xi_{jk_1}\xi_{jk_2}|\chi\Bigg] \\
&= \frac{1}{n^2}\sum_{i_1=1}^{n}\sum_{i_2=1}^{n}\sum_{k_1=1}^{p}\sum_{k_1=1}^{p}\frac{g'^2(\eta_{i_1})}{\sigma^2(\mu_{i_1})}\frac{g'^2(\eta_{i_2})}{\sigma^2(\mu_{i_2})}\epsilon_{k_1}^{i_1}\epsilon_{k_1}^{i_1}\xi_{jk_1}\xi_{jk_2} \\
&\quad \times E\Bigg[(y_{i_1}-g(\eta_{i_1}))(y_{i_2}-g(\eta_{i_2}))|\chi\Bigg] \\
&= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{k_1=1}^{p}\sum_{k_1=1}^{p}\frac{g'^4(\eta_i)}{\sigma^4(\mu_i)}\sigma^2\xi_{jk_1}\xi_{jk_2}\epsilon_{k_1}^{(i)}\epsilon_{k_1}^{(i)} \\
&= O_p(\frac{p^2}{n}).
\end{aligned}
$$

□

6.3 Proof of Theorem 3.3.1

Theorem 3.3.1 With probability 1, if $X_1(t),\cdots,X_n(t)$ are square integrable random functions and $\sum_{j=1}^{\infty}\delta_j^{-1}\to 0$, for all $1\le j\le J-1$,

$$\int_I\Bigg[\Big(\sum_{i=1}^{\infty}\ \beta_i[\tilde{\psi}_j(t)-\psi_j(t)]\Big)^2\Bigg]dt\to 0.$$

Proof. If both series converge,

$$\Bigg[\sum_{j=1}^{\infty}\beta_j[\tilde{\psi}_j(t)-\psi_j(t)]\Bigg]^2\le\sum_{j=1}\beta_j^2\sum_{j=1}^{\infty}[\tilde{\psi}_j(t)-\psi_j(t)]^2$$

It follows from results of Hall et cl. [2006] that

$$
\begin{aligned}
\int_I\Bigg[\sum_{j=1}^{\infty}\beta_j[\tilde{\psi}_j(t)-\psi_j(t)]\Bigg]^2dt &\le \sum_{j=1}\beta_j^2\int_I\sum_{j=1}^{\infty}[\tilde{\psi}_j(t)-\psi_j(t)]^2dt \\
&= \sum_{j=1}\beta_j^2\sum_{j=1}^{\infty}\|\tilde{\psi}_j(t)-\psi_j(t)\|^2 \\
&\le 8\sum_{j=1}\beta_j^2\tilde{\Delta}\sum_{j=1}^{\infty}\delta_j^{-1} \qquad (6.1)
\end{aligned}
$$

□

6.4 Proof of Theorem 3.3.2

Theorem 3.3.2

$$\int_I E\Big\{\big[\sum_{j=1}^{p}\hat{\hat{\beta}}_j\tilde{\psi}_j(t)-\sum_{j=1}^{\infty}\beta_j\tilde{\psi}_j(t)\big]^2|\chi\Big\}dt = O_p(p^{1/2}n^{-3/2})+o_p(1).$$

Proof.

$$\begin{aligned}
&\int_I E\Big\{\big[\sum_{j=1}^{p}\hat{\hat{\beta}}_j\tilde{\psi}_j(t)-\sum_{j=1}^{\infty}\beta_j\tilde{\psi}_j(t)\big]^2|\chi\Big\}dt \\
=&\int_I E\Big\{\big[\sum_{j=1}^{p}(\hat{\hat{\beta}}_j-\beta_j)\tilde{\psi}_j(t)-\sum_{j=p+1}^{\infty}\beta_j\tilde{\psi}_j(t)\big]^2|\chi\Big\}dt \\
=&\int_I E\Big\{\big[\sum_{j=1}^{p}(\hat{\hat{\beta}}_j-\beta_j)\tilde{\psi}_j(t)\big]^2|\chi\Big\}dt+\int_I E\Big\{\big[\sum_{j=p+1}^{\infty}\beta_j\tilde{\psi}_j(t)\big]^2|\chi\Big\}dt \\
&-2\int_I\sum_{i=1}^{p}\sum_{j=p+1}^{\infty}E\Big\{(\hat{\hat{\beta}}_j-\beta_j)\tilde{\psi}_j(t)\beta_i\tilde{\psi}_i(t)|\chi\Big\}dt
\end{aligned}\tag{6.2}$$

Assumption 2. *The number of principal components $p=p_n$ in the sequence of estimating p_n-truncated score function (3.8) satisfies $p=p_n\to\infty$ and $p_n n^{-1/4}\to 0$ as $n\to\infty$.*

Lemma 6.4.1. *Under Assumption 2, with probability 1,*

$$\int_I E\Big\{\big[\sum_{j=p+1}^{\infty}\beta_j\tilde{\psi}_j(t)\big]^2|\chi\Big\}dt = o_p(1)\tag{6.3}$$

Assumption 3. *Under these conditions, Lemma (6.4.2) holds.*

(a) $F(\beta,\epsilon)$ is differentiable around (β,ϵ)

Lemma 6.4.2. *Under Assumption 3, with probability 1,*

$$\int_I E\left\{\left[\sum_{j=1}^{p}(\hat{\hat{\beta}}_j-\beta_j)\tilde{\psi}_j(t)\right]^2|\chi\right\}dt = O_p(p^{1/2}n^{-3/2})+O_p(pn^{-3})\tag{6.4}$$

From Lemma (6.4.1) and (6.4.2),

$$\int_I E\Big\{\big[\sum_{j=1}^{p}\hat{\tilde{\beta}}_j\tilde{\psi}_j(t)-\sum_{j=1}^{\infty}\beta_j\tilde{\psi}_j(t)\big]^2|\chi\Big\}dt = O_p(p^{1/2}n^{-3/2})+o_p(1)$$

□

6.5 Proof of Lemma 6.4.1

Lemma 6.4.1 Under Assumption 2, with probability 1,

$$\int_I E\Big\{\big[\sum_{j=p+1}^{\infty}\beta_j\tilde{\psi}_j(t)\big]^2|\chi\Big\}dt = o_p(1)$$

Proof.

$$\begin{aligned}\int_I E\Big\{\big[\sum_{j=p+1}^{\infty}\beta_j\tilde{\psi}_j(t)\big]^2|\chi\Big\}dt &= \int_I E\Big\{\sum_{j=p+1}^{\infty}\beta_j^2\tilde{\psi}_j^2(t)+\sum_{i=1}^{p}\sum_{j\neq i}\beta_j\tilde{\psi}_j(t)\beta_i\tilde{\psi}_i(t)|\chi\Big\}dt \\ &= \int_I\sum_{j=p+1}^{\infty}\beta_j^2\tilde{\psi}_j^2(t)dt \\ &= \sum_{j=p+1}^{\infty}\beta_j^2\int_I\tilde{\psi}_j^2(t)dt \\ &= \sum_{j=p+1}^{\infty}\beta_j^2\to 0 \qquad (6.5)\end{aligned}$$

□

6.6 Proof of Lemma 6.4.2

Lemma 6.4.2 Under Assumption 3, with probability 1,

$$\int_I E\left\{\left[\sum_{j=1}^{p}(\hat{\tilde{\beta}}_j-\beta_j)\tilde{\psi}_j(t)\right]^2|\chi\right\}dt = O_p(p^{1/2}n^{-3/2})+O_p(pn^{-3})$$

Proof.

$$\begin{aligned}
& \int_I E\Big\{(\tilde{\hat{\beta}}_j-\beta_j)^2|\chi\Big\}dt \\
= & \int_I E\Big\{(\tilde{\hat{\beta}}_j-\hat{\beta}_j+\hat{\beta}_j-\beta_j)^2|\chi\Big\}dt \\
= & \int_I E\Big\{(\tilde{\hat{\beta}}_j-\hat{\beta}_j)^2|\chi\Big\}dt+\int_I E\Big\{(\hat{\beta}_j-\beta_j)^2|\chi\Big\}dt+2\int_I E\Big\{(\tilde{\hat{\beta}}_j-\hat{\beta}_j)(\hat{\beta}_j-\beta_j)|\chi\Big\}dt
\end{aligned}$$

As $\tilde{\hat{\beta}}$, a p by 1 vector, is the solution of score function $U(\tilde{\beta},\tilde{\epsilon})=\sum_{i=1}^n(Y_i-\tilde{\mu}_i)\frac{g'(\tilde{\eta}_i)}{\sigma^2(\tilde{\eta}_i)}\tilde{\epsilon}^{(i)}=0$ and $\hat{\beta}$ is the soluiton of $U(\beta,\epsilon)=\sum_{i=1}^n(Y_i-\mu_i)\frac{g'(\eta_i)}{\sigma^2(\eta_i)}\epsilon^{(i)}=0$, we have

$$\begin{aligned}
&\sum_{i=1}^n(Y_i-\tilde{\hat{\mu}}_i)\frac{g'(\tilde{\hat{\eta}}_i)}{\sigma^2(\tilde{\hat{\eta}}_i)}\tilde{\epsilon}^{(i)}=0 \\
&\sum_{i=1}^n(Y_i-\hat{\mu}_i)\frac{g'(\hat{\eta}_i)}{\sigma^2(\hat{\eta}_i)}\epsilon^{(i)}=0
\end{aligned} \tag{6.6}$$

where $\tilde{\hat{\eta}}_i=\sum_{j=1}^p\tilde{\hat{\beta}}_j\tilde{\epsilon}_j$ and $\hat{\eta}_i=\sum_{j=1}^p\hat{\beta}_j\epsilon_j$.

Now we consider $\sum_{i=1}^n(Y_i-\tilde{\hat{\mu}}_i)\frac{g'(\tilde{\hat{\eta}}_i)}{\sigma^2(\tilde{\hat{\eta}}_i)}\tilde{\epsilon}^{(i)}$ as function of $(\tilde{\hat{\beta}},\tilde{\epsilon})$ and $\sum_{i=1}^n(Y_i-\hat{\mu}_i)\frac{g'(\hat{\eta}_i)}{\sigma^2(\hat{\eta}_i)}\epsilon^{(i)}$ as function of $(\hat{\beta},\epsilon)$. We do taylor expansion to $F(\tilde{\hat{\beta}},\tilde{\epsilon})$ around $(\hat{\beta},\epsilon)$,

$$F(\tilde{\hat{\beta}},\tilde{\epsilon})=F(\hat{\beta},\epsilon)+\begin{bmatrix}\frac{\partial F}{\partial\hat{\beta}} & \frac{\partial F}{\partial\epsilon}\end{bmatrix}\begin{bmatrix}\tilde{\hat{\beta}}-\hat{\beta}\\ \tilde{\epsilon}-\epsilon\end{bmatrix}+R\tilde{\epsilon} \tag{6.7}$$

where $\|R\|\le C\left\|\begin{pmatrix}\tilde{\hat{\beta}}-\hat{\beta}\\ \tilde{\epsilon}-\epsilon\end{pmatrix}\right\|^2$ and C is a constant. $\frac{\partial F}{\partial\hat{\beta}}$ is the Jacobian of F with respect to $\hat{\beta}$ and the same as $\frac{\partial F}{\partial\epsilon}$

Then we have

$$0=\frac{\partial F}{\partial\hat{\beta}}(\hat{\beta},\epsilon)(\tilde{\hat{\beta}}-\hat{\beta})+\frac{\partial F}{\partial\epsilon}(\hat{\beta},\epsilon)(\tilde{\epsilon}-\epsilon)+R$$

$$
\begin{aligned}
&\left\| \frac{\partial F}{\partial \epsilon}(\hat{\beta}, \epsilon)(\tilde{\epsilon} - \epsilon) \right\|^2 \\
= &\sum_{k=1}^{p} \left[\sum_{l=1}^{p} \sum_{i=1}^{n} \frac{g'^2(\eta_i)}{\sigma^2(\eta_i)} \hat{\beta}_l \epsilon_k^{(i)} (\tilde{\epsilon}_l^{(i)}) + \sum_{t=1}^{p} \sum_{i=1}^{n} (y_i - g(\eta_i)) \epsilon_k^{(i)} \hat{\beta}_l \left[\frac{g''(\eta_i)}{\sigma^2(\eta_i)} - \frac{g'(\eta_i)\tilde{\sigma}^{2'}}{\sigma^4(\eta_i)} \right] (\tilde{\epsilon}_t^{(l)} - \epsilon_l^{(i)}) \right]^2 \\
= &\sum_{k=1}^{p} \left[\sum_{l=1}^{p} \sum_{i=1}^{n} \left(\frac{g'^2(\eta_i)}{\sigma^2(\eta_i)} + (y_i - g(\eta_i)) \left[\frac{g''(\eta_i)}{\sigma^2(\eta_i)} - \frac{g'(\eta_i)\tilde{\sigma}^{2'}}{\sigma^4(\eta_i)} \right] \right) \hat{\beta}_l \epsilon_k^{(i)} \left(\tilde{\epsilon}_l^{(i)} - \epsilon_i^{(i)} \right) \right]^2 \\
= &\sum_{k=1}^{p} \sum_{l_1=1}^{p} \sum_{l_2=1}^{p} \sum_{i_1=1}^{n} \sum_{i_2=1}^{n} \left[\frac{g'^2(\eta_{i_1})}{\sigma^2(\eta_{i_1})} + (y_{i_1} - g(\eta_{i_1}) \left[\frac{g''(\eta_{i_1})}{\sigma^2(\eta_{i_1})} - \frac{g'(\eta_{i_1})\tilde{\sigma}^{2'}}{\sigma^4(\eta_{i_1})} \right] \right] \\
&\left[\frac{g'^2(\eta_{i_2})}{\sigma^2(\eta_{i_2})} + (y_{i_2} - g(\eta_{i_2}) \left[\frac{g''(\eta_{i_2})}{\sigma^2(\eta_{i_2})} - \frac{g'(\eta_{i_2})\tilde{\sigma}^{2'}}{\sigma^4(\eta_{i_2})} \right] \right] \hat{\beta}_{l_1} \hat{\beta}_{l_2} \epsilon_k^{(i_1)} \epsilon_k^{(i_2)} \left(\tilde{\epsilon}_{l_1}^{(i_1)} - \epsilon_{l_1}^{(i_1)} \right) \left(\tilde{\epsilon}_{l_2}^{(i_2)} - \epsilon_{l_2}^{(i_2)} \right)
\end{aligned}
$$

Assumption 4. *Assume that with probability 1, X is left-continuous at each point (or right-continuous at each point).*

Based on results of Hall and Hosseini-Nasab [2009], under Assumption 4, with probability 1, $\|\tilde{\epsilon}_l^{(i)} - \epsilon_l^{(i)}\| = O_p(n^{-1/2})$, for all l such that $1 \le l \le J-1$, where J defined as (3.4).

Assumption 5. *The link function g is monotone, invertible and has two continuous bounded derivatives with $\|g'(\cdot)\| \le c$, $\|g''(\cdot)\| \le c$ for a constant $c \le 0$. The variance function $\sigma^2(\cdot)$ has a continuous bounded derivative and there exists a $\delta > 0$ such that $\sigma(\cdot) \le \delta$.*

Assumption 6. *For each component of estimate of truncated model (2.3), $\hat{\beta}_k$,*

$$E(\hat{\beta}_k^2|\chi) \le C$$

and

$$\max\{\theta_1, \cdots, \theta_p, \cdots\} \le M'$$

where C and M' are constants and $k = 1, \cdots, p$.

Then, with Assumption 5 and 6,

$$
\begin{aligned}
E\left\{\left\|\frac{\partial F}{\partial \epsilon}(\hat{\beta},\epsilon)(\tilde{\epsilon}-\epsilon)\right\|^2|\chi\right\} &\leq M\sum_{k=1}^{p}\sum_{l_1=1}^{p}\sum_{l_2=1}^{p}\sum_{i_1=1}^{n}\sum_{i_2=1}^{n} \\
& \quad E\left\{\hat{\beta}_{l_1}\hat{\beta}_{l_2}\epsilon_k^{(i_1)}\epsilon_k^{(i_2)}\left(\tilde{\epsilon}_{l_1}^{(i_1)}-\epsilon_{l_1}^{(i_1)}\right)\left(\tilde{\epsilon}_{l_2}^{(i_2)}-\epsilon_{l_2}^{(i_2)}\right)|\chi\right\} \\
&= M\sum_{k=1}^{p}\sum_{i=1}^{n}E\left\{\hat{\beta}_k^2\epsilon_k^{(i)^2}(\tilde{\epsilon}_k^{(i)}-\epsilon_k^{(i)})^2|\chi\right\} \\
&= M\sum_{k=1}^{p}\sum_{i=1}^{n}\left[\epsilon_k^{(i)^2}(\tilde{\epsilon}_k^{(i)}-\epsilon_k^{(i)})^2\right]E(\hat{\beta}_k^2|\chi)=O_p(p/n)
\end{aligned}
$$

where M is a constant.

For the remainder term R, we have

$$
\begin{aligned}
\|R\| &\leq C\left\|\begin{pmatrix}\hat{\hat{\beta}}-\hat{\beta} \\ \\ \tilde{\epsilon}-\epsilon\end{pmatrix}\right\|^2 \\
&= C\left[\sum_{i=1}^{p}(\hat{\hat{\beta}}_j-\hat{\beta}_j)^2+\sum_{j=1}^{p}(\tilde{\epsilon}_j-\epsilon)^2\right] \\
&= C\sum_{i=1}^{p}(\hat{\hat{\beta}}_j-\hat{\beta}_j)^2+O_p(p/n^2)
\end{aligned}
$$

$$
\begin{aligned}
0 &= \frac{\partial F}{\partial\hat{\beta}}(\hat{\beta},\epsilon)(\hat{\hat{\beta}}-\hat{\beta})+\frac{\partial F}{\partial\epsilon}(\tilde{\epsilon}-\epsilon)+R \\
0 &\leq -\|\frac{\partial F}{\partial\hat{\beta}}(\hat{\beta},\epsilon)(\hat{\hat{\beta}}-\hat{\beta})\|+\|\frac{\partial F}{\partial\epsilon}(\tilde{\epsilon}-\epsilon)+R\|
\end{aligned}
$$

$$
\begin{aligned}
\|\frac{\partial F}{\partial\hat{\beta}}(\hat{\beta},\epsilon)(\hat{\hat{\beta}}-\hat{\beta})\| &\leq \|\frac{\partial F}{\partial\epsilon}(\hat{\epsilon}-\epsilon)\|+\|R\| \\
&\leq O_p(\sqrt{p/n})+O_p(p/n^2)+C\sum_{j=1}^{p}(\hat{\hat{\beta}}_j-\hat{\beta}_j)^2 \qquad (6.8)
\end{aligned}
$$

$$
\|\frac{\partial F}{\partial\hat{\beta}}(\hat{\beta},\epsilon)(\hat{\hat{\beta}}-\hat{\beta})\|^2
$$

$$= \sum_{k=1}^{p} \left\{ \sum_{l=1}^{p} \sum_{i=1}^{n} \left[\frac{g'^2(\eta_i)}{\sigma^2(\eta_i)} + (y_i - g(\eta_i)) \left[\frac{g''(\eta_i)}{\sigma^2(\eta_i)} - \frac{g'(\eta_i)\tilde{\sigma}^{2'}}{\sigma^4(\eta_i)} \right] \right] \epsilon_k^{(i)} \epsilon_l^{(i)} (\hat{\tilde{\beta}}_l - \hat{\beta}_l) \right\}^2$$

$$= \sum_{k=1}^{p} \sum_{l_1,l_2=1}^{p} \sum_{i_1,i_2=1}^{n} \left\{ \frac{g'^2(\eta_{i_1})}{\sigma^2(\eta_{i_1})} + (y_{i_1} - g(\eta_{i_1})) \left[\frac{g''(\eta_{i_1})}{\sigma^2(\eta_{i_1})} - \frac{g'(\eta_{i_1})\tilde{\sigma}^{2'}}{\sigma^4(\eta_{i_1})} \right] \right\}$$
$$\left\{ \frac{g'^2(\eta_{i_2})}{\sigma^2(\eta_{i_2})} + (y_{i_2} - g(\eta_{i_2})) \left[\frac{g''(\eta_{i_2})}{\sigma^2(\eta_{i_2})} - \frac{g'(\eta_{i_2})\tilde{\sigma}^{2'}}{\sigma^4(\eta_{i_2})} \right] \right\} \epsilon_k^{(i_1)} \epsilon_k^{(i_2)} \epsilon_{l_1}^{(i_1)} \epsilon_{l_2}^{(i_2)} (\hat{\tilde{\beta}}_{l_1} - \hat{\beta}_{l_1})(\hat{\tilde{\beta}}_{l_2} - \hat{\beta}_{l_2})$$

$$\begin{aligned} E\left\{ \| \frac{\partial F}{\partial \hat{\beta}}(\hat{\beta}, \epsilon)(\hat{\tilde{\beta}} - \hat{\beta}) \|^2 | \chi \right\} &= O_p(1) \sum_{k=1}^{p} \sum_{l=1}^{p} \sum_{i=1}^{n} \epsilon_k^{(i)^2} \epsilon_l^{(i)^2} E\left\{ (\hat{\tilde{\beta}} - \hat{\beta})^2 | \chi \right\} \\ &= \sum_{k=1}^{p} \sum_{i=1}^{n} \epsilon_k^{(i)^4} E\left\{ (\hat{\tilde{\beta}} - \hat{\beta})^2 | \chi \right\} \\ &= O_p(np) E\left\{ (\hat{\tilde{\beta}}_l - \hat{\beta}_l)^2 | \chi \right\} \end{aligned}$$

From Inequation (6.8),

$$O_p(np)E\left\{ (\hat{\tilde{\beta}}_l - \hat{\beta}_l)^2 | \chi \right\} - O_p(p) E\left\{ \hat{\tilde{\beta}}_l - \hat{\beta}_l)^2 | \chi \right\} \le O_p(\sqrt{p/n}) + O_p(p/n^2)$$
$$E\left\{ (\hat{\tilde{\beta}}_l - \hat{\beta}_l)^2 | \chi \right\} \le O_p(p^{-1/2} n^{-3/2}) + O_p(n^{-3})$$

$$\begin{aligned} & \int_I E\left\{ \left[\sum_{j=1}^{p} (\hat{\tilde{\beta}}_j - \beta_j) \tilde{\psi}_j(t) \right]^2 | \chi \right\} dt \\ = & \int_I E\left\{ \sum_{j=1}^{p} (\hat{\tilde{\beta}}_j - \beta_j)^2 \tilde{\psi}_j^2(t) + \sum_{i=1}^{p} \sum_{j \ne i} (\hat{\tilde{\beta}}_i - \beta_i) \tilde{\psi}_i(t) (\hat{\tilde{\beta}}_j - \beta_j) \tilde{\psi}_j(t) | \chi \right\} dt \\ = & \sum_{j=1}^{p} E\left\{ (\hat{\tilde{\beta}}_j - \beta_j)^2 | \chi \right\} \\ = & O_p(p^{1/2} n^{-3/2}) + O_p(pn^{-3}) \end{aligned} \quad (6.9)$$

□

Bibliography

Aguilera, A. M., F. A. Ocana and M. J. Valderrama (1999), “Forecasting Time Series by Functional PCA: Discussion of Several Weighted Approaches,” Computational Statistics, 14, 443-467.

Anderson, Eugene and Salisbury Linda (2003), “The Formation of Market-Level Expectations and Its Covariates,” Journal of Consumer Research, 30 (June), 115-24.

Applegate, D., A. Archer, V. Gopalakrishnan, S. Lee, and K. K. Ramakrishnan (2010),“Optimal Content Placement for a Large-scale VoD System,” in ACM CoNEXT.

Basuroy, S.,S. Chatterjee, and A. Ravid (2003), “How critical are critical reviews? The box office effects of film critics, star power, and budget,”Journal of Marketing, 67(4), 103-117.

Bisdikian, C.C., Patel, B.V. (1995), “Issues on movie allocation in distributed video-on-demand systems ,” Communications, ICC ’95 Seattle, ’Gateway to Globalization’, 1995 IEEE International Conference, 1, 250-255.

Bear F. Braumoeller (2004). Hypothesis Testing and Multiplicative Interaction Terms. International Organization, 58 , pp 807-820.

Bronnenberg, Bart J. and Catarina Sismeiro (2002), “Using Multimarket Data to Predict Brand Performance in Markets for Which No or Poor Data Exist,” Journal of Marketing Research, 39, 1-17.

Richard Cangelosi, and Alain Goriely (2007). “Component retention in principal component analysis with application to cDNA microarray data.” Biology Direct. 2 (1).

Chang, Byeng-Hee, and Eyun-Jung Ki (2005), “Devising a practical model for predicting theatrical movie success: Focusing on the experience good property,” Journal of Media Economics, 18.4, 247-269.

Pradeep K. Chintagunta , Shyam Gopinath , Sriram Venkataraman, The Effects of Online User Reviews on Movie Box Office Performance: Accounting for Sequential Rollout and Aggregation Across Local Markets, Marketing Science, v.29 n.5, p.944-957, September 2010.

Jose M. Cortina, Interaction, nonlinearity, and multicollinearity: Implications for multiple regression, Journal of Management, Volume 19, Issue 4, Winter 1993, Pages 915-922.

Cuevas, A., M. Febrero and R. Fraiman (2002), “Linear Functional Regression: The Case of Fixed Design and Functional Response,” Canadian Journal of Statistics, 30, 285-300.

Chrysanthos Dellarocas, Xiaoquan (Michael) Zhang, and Neveen F. Awad (2007), “Exploring the value of online product reviews in forecasting sales: The case of motion pictures,” Journal of Interactive Marketing, 21(4), 23-45.

Desai K, Basuroy S. (2005), “Interactive Influence of Genre Familiarity, Star Power, and Critics’ Reviews in the Cultural Goods Industry: The Case of Motion Picture,” Psychology & Marketing [serial online], 22(3): 203-223.

Wenjing Duan, Bin Gu, Andrew B. Whinston (2008), “The dynamics of online word-of-mouth and product sales: An empirical investigation of the movie industry,” Journal of Retailing,84(2), 233-242.

Elberse, Anita (2007), “The Power of Stars: Do Star Actors Drive the Success of Movies?,” Journal of Marketing, 71, 102-120.

Eliashberg, Jehoshua, and Steven M. Shugan (1997), “Film Critics: Influences or Predictors,” Journal of Marketing, 61(2), 68-78.

Fan,Jianqing, and Sheng-Kuei Lin (1998). ’Test of Significance When Data Are Curves.” Journal of the American Statistical Association 93 (443), 1007-21.

Faraway, J.J. (1997). Regression analysis for a functional response. Technometrics 39, 254-262.

Foutz, N. Z., and W. Jank (2010), " Prerelease demand forecasting for motion pictures: Using functional shape analysis of virtual stock markets," Marketing Science, 29(3), 568-379.

Foutz, N. Z. and W. Jank (2010), " Research Note-Prerelease Demand Forecasting for Motion Pictures Using Functional Shape Analysis of Virtual Stock Markets," Marketing Science, 29, 3, 568-579.

William Greene, Testing hypotheses about interaction terms in nonlinear models, Economics Letters, Volume 107, Issue 2, May 2010, Pages 291-296.

Guo, W. (2002), "Inference in Smoothing Spline Analysis of Variance," Journal of the Royal Statistical Society: Series B (Statistical Methodology), 64, 887-898.

Hall, Peter, Hans-Georg Muller, and Jane-Ling Wang (2006). "Principal Component Methodology - Properties of principal component methods for functional and longitudinal data analysis". The Annals of Statistics. 34 (3): 1493.

Hall, Peter, and Hosseini-Nasab M (2009). "Theory for high-order bounds in functional principal components analysis." Mathematical Proceedings of the Cambridge Philosophical Society, 146 (1): 225-256.

Hall, Peter, and Joel L. Horowitz (2007), "Methodology and convergence rates for functional linear regression," Annals of Statistics, 35, 1, 70-91.

Hyde, V., E. Moore and A. Hodge (2004), "Functional PCA for exploring bidding activity times for online auctions," Term paper, Research Interaction Team, VIGRE program, Univ. Maryland.

James, G. M. (2002), "Generalized Linear Models with Functional Predictors," Journal of the Royal Statistical Society: Series B (Statistical Methodology), 64, 411-432.

James, G. M. and C. A. Sugar (2003), "Clustering for Sparsely Sampled Functional Data," Journal of the American Statistical Association, 98, 397-408.

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

James, G. M. and T. J. Hastie (2001), “Functional Linear Discriminant Analysis for Irregularly Sampled Curves,” Journal of the Royal Statistical Society: Series B (Statistical Methodology), 63, 533-550.

James, G., and Silverman, B. (2005) ”Functional Adaptive Model Estimation”, Journal of the American Statistical Association 100, 565-576.

Jank, Wolfgang and Natasha Foutz (2007),“ Using virtual stock exchanges to forecast box-office revenue via functional shape analysis,” In The Prediction Markets Workshop at Electronic Commerce.

Jank, W., and Shmueli, G. (2007), “Modeling concurrency of events in online auctions via spatio-temporal semiparametric models,” Journal of the Royal Statistical Society, 56(1), 1-27.

Jank, Wolfgang and P.K. Kannan (2005), “Understanding Geographical Markets of Online Firms Using Spatial Models of Customer Choice,” Marketing Science, 24(4), 623-34.

Jank, W., G. Shmueli. 2008. Studying heterogeneity of price evolution in eBay auctions via functional clustering. Business Computing Handbooks in Information Systems Series, Vol.3. Emerald Group Publishing, Bingley, UK, 237-263.

Jank, Wolfgang, and Zhang Shu. ”An Automated And Data-Driven Bidding Strategy For Online Auctions.” INFORMS Journal On Computing 23.2 (2011): 238-253.

Timothy King (2007) , Does film criticism affect box office earnings? Evidence from movies released in the U.S. in 2003, Journal of Cultural Economics, Volume 31, Number 3, 171-186.

Krider, R. E., Li, T., Liu, Y, and Weinberg, C. B. (2005), “The Lead-Lag Puzzle of Demand and Distribution: A Graphical Method Applied to Movies,” Marketing Science, 24(4), 635-645.

Liu, Yong (2006), “Word-of-Mouth for Movies: Its Dynamics and Impact on Box Office Revenue,”Journal of Marketing, 70(3), 74-89.

Wei-Yin Loh, Regression Trees With Unbiased Variable Selection and Interaction Detection, Statistica Sinica (2002) vol. 12, pp. 361–386.

McCullagh, P. (1983), "Quasi-likelihood Functions." Annals of Statistics, 11, 59-67.

Medved, M. (1992), "Hollywood vs. America," New York: Harper Collins.

Minka, Thomas P. (2000). "Automatic choice of dimensionality for PCA. "In NIPS, 13, 598-604.

Mittal, Vikas, Wagner A. Kamakura, and Rahul Govind (2004), "Geographic Patterns in Customer Service and Satisfaction: An Empirical Investigation," Journal of Marketing: 68(3), 48-62.

Moyeed, R. A. and Diggle, P. J. (1994), "Rates of convergence in semi-parametric modeling of longitudinal data," Australian and New Zealand Journal of Statistics, 36, 75-93.

Muller, H.-G., and U. Stadtmuller (2005). "Generalized functional linear models." ANNALS OF STATISTICS. 33 (2): 774-805.

Sungjoon Nam , Puneet Manchanda , Pradeep K. Chintagunta, The Effect of Signal Quality and Contiguous Word of Mouth on Customer Acquisition for a Video-on-Demand Service, Marketing Science, v.29 n.4, p.690-700, July 2010.

Nelder, J. A. and Wedderburn, R. W. M. (1972), "Generalized Linear Models." J. Royal Statistics Society, A, 135, 370-384.

Edward C. Norton Hua Wang Chunrong Ai, 2004. "Computing interaction effects and standard errors in logit and probit models", Stata Journal, StataCorp LP, vol. 4(2), pages 154-167, June.

Parker, Philip M. and Tavassoli, T. Nader (2000), "Homeostasis and Consumer Behavior Across Cultures," International Journal of Research in Marketing, 17(1), 33-53.

Pearson, K. *On Lines and Planes of Closest Fit to Systems of Points in Space*, Philosophical Magazine 2(6): 559-572.

J. Prag, J. Casavant (1994), "An empirical study of the determinants of revenues and marketing expenditures in the motion picture industry," Journal of Cultural Economics,18, 217-235.

James O. Ramsay, James B. Ramsey, 2002, Functional data analysis of the dynamics of the monthly index of nondurable goods production, Journal of Econometrics, Volume 107, Issues 1-2:327-344.

Silverman, B. (1996), "Smoothed Functional Principal Components Analysis by Choice of Norm," The Annals of Statistics, 68, 45-54.

Simonoff, Jeffrey S. and Sparrow, Ilana R.(2000), "Predicting movie grosses: Winners and losers, blockbusters and sleepers," Chance, 13(3), 15-24.

Sood, Ashish, G. James and Gerard J.Tellis (2009), "Functional Regression: A New Model and Approach for Predicting Market Penetration of New Products," Marketing Science, 28(1), 36-51.

Stefan Sperlich, Dag Tjøstheim and Lijian Yang (2002). NONPARAMETRIC ESTIMATION AND TESTING OF INTERACTION IN ADDITIVE MODELS. Econometric Theory, 18 , pp 197-251.

Staniswalis, J. G. and J. J. Lee (1998), "Nonparametric Regression Analysis of Longitudinal Data," Journal of the American Statistical Association, 93, 1403-1418.

Stewart, K.J., D.P. Darcy, S. Daniel. 2006. Opportunities and challenges applying functional data analysis to the study of open source software evolution. Statistical Science 21(2) 167-178.

Sun, M. J. 2008. The informational role of consumer disagreement. Working paper, Department of Economics, Boston University, Boston.

Tarpey, T. and K. K. J. Kinateder (2003), "Clustering Functional Data," Journal of Classification, 20, 93-114.

Ramsay, J.O. and Silverman, B.W. (1997), "Functional Data Analysis," New York, Springer.

Ramsay, J. O. and Silverman, B.W. (2005), "Functional data analysis, 2nd ed., New York : Springer."

Rao, C.R. (1958), "Some Statistical Methods for Comparison of Growth Curves," Biometrics, 14, 1-17.

Reinstein, D. A. and Snyder, C. M. (2005), THE INFLUENCE OF EXPERT REVIEWS ON CONSUMER DEMAND FOR EXPERIENCE GOODS: A CASE STUDY OF MOVIE CRITICS. The Journal of Industrial Economics, 53: 27–51.

Rice, John A., and B. W. Silverman (1991).“Estimating the Mean and Covariance Structure Nonparametrically When the Data are Curves.” Journal of the Royal Statistical Society. Series B (Methodological). 53 (1): 233-243.

Ross, William T., Jr., and Itamar Simonson (1991), “Evaluations of Paris of Experiences: A Preference for Happy Endings,”Journal of Behavioral Decision Making, 4(4), 273-82.

Viviani, Roberto, George Grő n, and Manfred Spitzer (2005), “Functional principal component analysis of fMRI data,” Human Brain Mapping, 24, 2, 109-129.

Wang, S., W Jank and G. Shmueli (2008), “Explaining and Forecasting Online Auction Prices and their Dynamics using Functional Data Analysis,” Journal of Business and Economic Statistics, 26 (2), p. 144-160.

Wedderburn, R. W. M. (1974), “Quasi-likelihood Functions, Generalized Linear Models and Gauss-Newton Method.” Biometrika, 61, 439-447.

Yu, Hongliang, Dongdong Zheng, Ben Y.Zhao and Weimin Zheng (2006), “Understanding User Behavior in Large-Scale Video-on-Demand systems,” EuroSys, 18-21.

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.

Zhou, Xiaobo, Cheng-Zhong Xu (2002), “Optimal video replication and placement on a cluster of video-on-demand servers,” Proceedings. International Conference, 547- 555.