# Functional Latent Feature Models With A Single-Index Interaction

Naisyin Wang
Department of Statistics
University of Michigan

Yehua Li and Raymond J. Carroll
University of Georgia and Texas A&M University

PL Hsu 100 Conference, Peking University

July 4, 2010

# Our Goals

- Generalize the Standard FDA Model to allow for low-dimensional interactions
- Quantify the cost due to estimating functional principal components
- Analyze a nutrition data set

# Data Structure

- Scalar Response $Y_i$
- Longitudinal covariate $X_i(t)$, with mean $\mu_X(t)$
- Possibly observed covariate $W_i(t) = X_i(t) + U_i(t)$, with $U_i(\cdot)$ white noise.
- Fixed covariate $Z_i$, includes a 1.0 for an intercept
- Relate $Y$ to $\{X(\cdot), Z\}$

# Standard FDA Model

- The standard FDA model says that there is an unknown function $\mathfrak{A}(\cdot)$ such that

$$Y_i = \int \mathfrak{A}(t) X_i(t) dt + Z_i^{\mathrm{T}} \beta + \epsilon_i .$$

- We start with the simplest case that $X_i(\cdot)$ is entirely observed.

# Standard FDA Model

- Let $(t) = \{\psi_1(t), \psi_2(t), \cdots, \psi_p(t)\}^{\mathrm{T}}$ be $p$ orthonormal functions.
- A standard model is that

$$\mathfrak{A}(t) = {}^{\mathrm{T}}(t)\alpha.$$

- De ne for $j = 1, ..., p$,

$$\xi_{ij} = \int \psi_j(t)\{X_i(t) - \mu_X(t)\}dt.$$

- Then we have that with $\xi_i = (\xi_{i1}, ..., \xi_{ip})^{\mathrm{T}}$,

$$\int \mathfrak{A}(t)X_i(t)dt = \alpha_0 + \xi_i^{\mathrm{T}}\alpha_1.$$

# Functional Latent Features

- $\xi$ is a vector of latent variables, which we refer as the latent feature of the functional data.
- Two typical structural considerations in functional data analysis.
  - Fixed bases: $(\psi_j)$ are known basis functions, such as Fourier or wavelet basis functions.
  - Data-driven bases, e.g., the $(\psi_j)$ are the leading principal components of $X(t)$.

# Thoughts Behind the Interaction Structure

- Recall that we reduce $_1(t)\alpha_1(Z_1,\theta) + \cdots + \quad _p(t)\alpha_p(Z_1,\theta)$ to
$$^{\mathrm{T}}(t)\{\alpha_1 + \mathcal{S}(Z_1^{\mathrm{T}}\theta)\alpha_2\}.$$
- We consider only ONE $\mathcal{S}(\cdot)$ instead of $p$ of them.
- It would have been over-ambitious to estimate $p$ $\mathcal{S}(\cdot)$ due to the unstability of estimation of a higher PCA direction.
- $Z_1^{\mathrm{T}}\theta$ explains the main \$Z$-direction" on which $X(t)$ interact with $Z$.
- We use $\alpha_2$ (parametric) to accommodate di erences in interactions among di erent PCA directions.

# Allowing Interactions

- With the single index model and the known $\xi$'s, we are left with the model

$$Y = \xi^{\mathrm{T}}\{\alpha_1 + \mathcal{S}(Z_1^{\mathrm{T}}\theta)\alpha_2\} + Z^{\mathrm{T}}\beta + \epsilon$$

- This is a semiparametric single index model with an unknown function $\mathcal{S}(\cdot)$.

- There are various things needed for identi ability, we use $\|\theta\| = \|\alpha_2\| = 1$, $\mathsf{E}\{\mathcal{S}(\theta^T Z_1)\} = 0$.

# Allowing Interactions With Non-Gaussian Data

- The material is easily generalized to quasilikelihood.
- Let the mean and variance of $Y$ given $(X, Z)$ be given as

$$E(Y_i|X_i, Z_i) = \mu\left[\xi_i^{\mathrm{T}}\{\alpha_1 + \mathcal{S}(Z_{1i}^{\mathrm{T}}\theta)\alpha_2\} + Z_i^{\mathrm{T}}\beta\right];$$

$$\mathrm{var}(Y_i|X_i, Z_i) = V\left(\mu\left[\xi_i^{\mathrm{T}}\{\alpha_1 + \mathcal{S}(Z_{1i}^{\mathrm{T}}\theta)\alpha_2\} + Z_i^{\mathrm{T}}\beta\right]\right).$$

# Algorithm

- The basic mean model is that

$$E(Y_i|X_i, Z_i) = \mu\left[\xi_i^{\mathrm{T}}\{\alpha_1 + \mathcal{S}(Z_{1i}^{\mathrm{T}}\theta)\alpha_2\} + Z_i^{\mathrm{T}}\beta\right].$$

- The terms $\xi_i$ are still known to us, for now.
- There are options in tting quasilikelihood, we follow the use of the **MAVE** method of Xia and Hardle (06).
- The MAVE idea is to do two steps.
- **Step 1**: local quasilikelihood back tting with full multivariate kernel weights to get a consistent estimate of $\theta$, say $\widetilde{\theta}$.
- **Step 2**: local quasilikelihood back tting with univariate kernel weights with arguments $Z_1^{\mathrm{T}}\widetilde{\theta}$ to get e cient estimates.

# Results when the Basis Functions are Known

- We have shown that the **final estimated function** has the **usual type of bias-variance decomposition**. Under given conditions, $\|\widetilde{\ } - \ _0\| \to 0$ with probability 1.
- We have shown that the nal estimates of all the parameters has $\sqrt{n}$ **rate of convergence**, is **asymptotically normal**, and has an asymptotic covariance matrix of the form $A^{-1}BA^{-1}$, where $A^{-1}$ is a generalized inverse. Under some conditions,

$$\sqrt{n}(\widehat{\ } - \ _0) \to \text{Normal}(\mathbf{0}, \mathcal{A}^{-}\mathcal{B}\mathcal{A}^{-}),$$
$$\sqrt{nh}\{\widehat{\mathcal{S}}(u) - \mathcal{S}_0(u) - h^2\mathcal{S}^{(2)}(u)\sigma_k^2/2\} \to \text{Normal}\{0, \sigma_{\mathcal{S}}^2(u)\},$$

# What if the Basis Functions are Unknown?

- The most widely used data driven method uses **functional principal components** (FPCA).
- De ne $\mu_X(t) = E\{X(t)\}$,

$$R(s,t) = \text{cov}\{X(s), X(t)\} = \sum_{k=1}^{\infty} \omega_k \psi_k(s)\psi_k(t).$$

- The **Karhunen-Loève expansion** says that

$$X_i(t) = \mu(t) + \sum_{j=1}^{\infty} \xi_{ij}\psi_j(t),$$

where $\mathrm{E}(\xi_j) = 0$, $\text{cov}(\xi_{ij}, \xi_{ik}) = I(j = k)\omega_j$.

# Discrete functional data

- We **do not observe the process** $X_i(\cdot)$.
- Instead, **we observe discrete observations with noise**, i.e., at observation times $T_{ij}$, we observe

$$W_{ij} = X_i(T_{ij}) + U_{ij}, \quad j = 1, \cdots, m_i,$$

where $U_{ij}$ are independent zero-mean errors independent of $X_i(\cdot)$ and $Z_i$, with $\text{var}\{U_i(t)\} = \sigma_u^2$.

# Discrete functional data

- It is commonly assumed that there are in nite number non-zero eigenvalues.
- A more realistic assumption is that $Y$ only depends on a **finite number** $p$ of the leading principal components.
- To push this through, we have to estimate $\text{cov}\{X(s), X(t)\}$; we use kernel methods and $W$ to do this.
- We have $\widehat{R}(s,t) = \widehat{\sigma}_{XX}(s,t) - \widehat{\mu}_X(s)\widehat{\mu}_X(t)$, $\sigma_{XX}(s,t) = \mathrm{E}\{X(s)X(t)\}$.
- We then get $\widehat{\sigma}_u^2$ based on $\sigma_W^2(t) = \text{var}\{W(t)\} = R(t,t) + \sigma_u^2$.
- Estimate of PC-scores: numerical integration (NI) method (Müller, 2007).

# E ect of Estimating Basis Function

- Remember our basic model

$$E(Y_i|X_i, Z_i) = \mu\left[\xi_i^{\mathrm{T}}\{\alpha_1 + \mathcal{S}(Z_{1i}^{\mathrm{T}}\theta)\alpha_2\} + Z_i^{\mathrm{T}}\beta\right].$$

- We now substitute estimates $\widehat{\xi_i}$ for $\xi_i$.
- We have shown **under certain technical conditions** that estimating the PC scores **will not** impact the variability of the estimate of $\mathcal{S}(\cdot)$ but **will increase** the variability of the parameter estimates.

# Simulation

- We generate Gaussian longitudinal process $X(t)$ for $t \in [0, 1]$, with mean function $\mu_X(t) = (t - 0.6)^2 - 0.1$.
- The covariance function of the process had 2 principal components, $\psi_1(t) = 1$, $\psi_2(t) = \sqrt{2}\sin(2\pi t)$, and the eigenvalues were $\omega_1 = 1.0$ and $\omega_2 = 0.6$.
- Assume $m = 30$ discrete observations on each curve, with random observation time points being uniformly distributed on the interval $[0, 1]$.
- Discrete observations on $X$ are contaminated with zero-mean Gaussian error with variance $\sigma_u^2 = 0.1$.

# Results of the Simulation

- We found some variance in ation due to estimating the PC scores.
- There was a 50% variance in ation for estimating the parameters associated with the second principal component
- Main e ect parameters were badly biased if the interaction was ignored.

# Simulation results

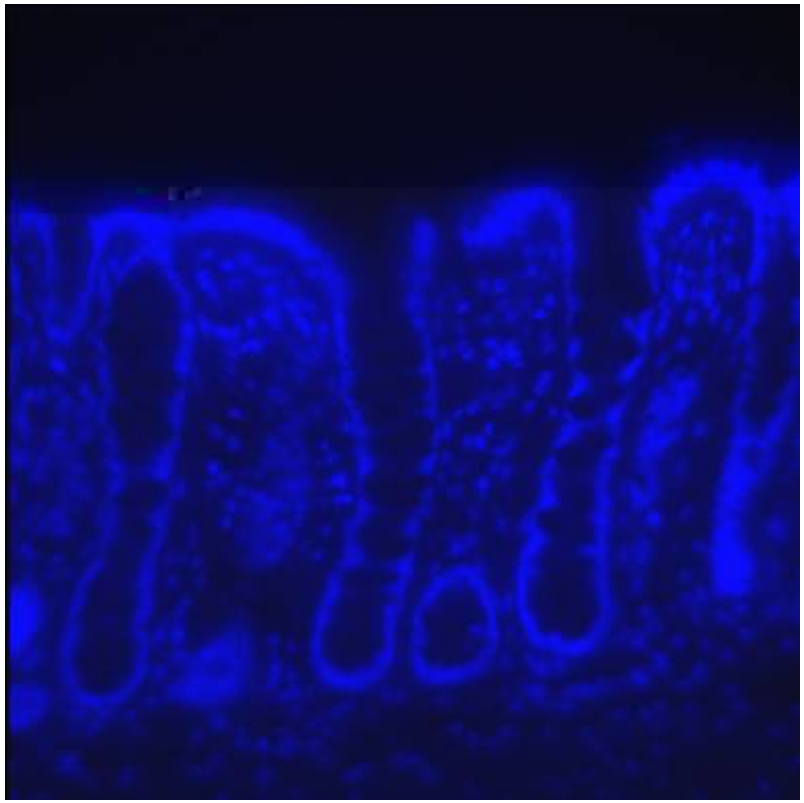
1.7071(1.497)(-07 T F225.1877(0.061) T 23.975-0.038) T F265.976ula6 0.207 -3

| | Full Model | | | | Reduced Model | | | |
|---|---|---|---|---|---|---|---|---|
| | Truth | Mean | SD | Bias | | Mean | SD | Bias |
| $_0$ | -1 | -1.038 | 0.418 | -0.038 | $_0$ | -0.939 | 0.368 | 0.061 |
| $_1$ | 2 | 2.062 | 0.494 | 0.062 | $_1$ | 1.757 | 0.429 | -0.243 |
| $_2$ | -2 | -2.021 | 0.466 | -0.021 | $_2$ | -1.741 | 0.406 | 0.259 |
| $_3$ | 2 | 2.076 | 0.299 | 0.076 | $_3$ | 1.888 | 0.270 | -0.112 |
| $^*_{11}$ | 1.7986 | 1.755 | 0.194 | -0.044 | $_1$ | 1.497 | 0.171 | -0.301 |
| $^*_{12}$ | -0.0014 | -0.096 | 0.207 | -3 (3)F12813.0.096 (3)F12813.0.096 (3)F12813.0.096 (3)F12813.0.096 ( | | | | |

# Colon Carcinogenesis Data

- Beneath the colon tissue, there are pore structures called `colonic crypts', see next slide.
- A crypt typically contains 25-30 cells. Functional covariate, $X(t)$,: *p27* measured at cell level, and $t$: relative cell location within the crypt.
- *p27* is a protein that inhibits the cell cycle.
- We sampled about 20 crypts from each of the 12 rats, with a total of $n = 249$ crypts.
- There are 2 diet groups (corn oil diet or  sh oil diet) and 2 treatment groups (with/without butyrate supplement).

# Colonic crypts



# Colon Carcinogenesis Data

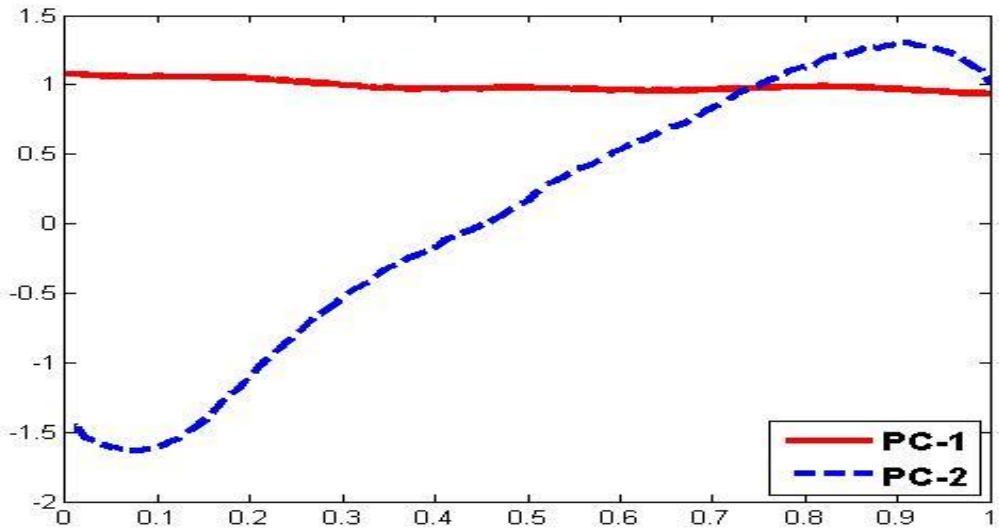
**Goal of the study:** to build a regression model between $Y$ = mean apoptotic (programmed cell death) rate within a crypt and

- $X$ = *p27* pro le curve within a crypt
- $Z$ = environment variables (diet, treatment), and mean proliferation rate in each crypt
- interaction between $X$ and $Z$. so that $Z_1$ is the same as $Z$ but without the intercept.

# PCA for *p27* data

- The first 3 eigenvalues are 0.871, 0.019 and 0.005 respectively.
- In our regression, we will only use the first 2 PC's.

# PCA for *p27* data



# The Shape of the Interaction Surface

- Our model is

$$\begin{aligned} Y_i &= \int \mathfrak{A}(t, Z_{i1}^{\mathrm{T}}\theta) X_i(t) dt + Z_i^{\mathrm{T}}\beta + \epsilon_i \\ \mathfrak{A}(t, Z_1^{\mathrm{T}}\theta) &= {}^{\mathrm{T}}(t)\{\alpha_1 + \mathcal{S}(Z_1^{\mathrm{T}}\theta)\alpha_2\} \end{aligned}$$

- The shape of $\mathfrak{A}(t, Z_1^{\mathrm{T}}\theta)$ varies dramatically based on where the cell is located, see next.

# Estimated functional coe cient function

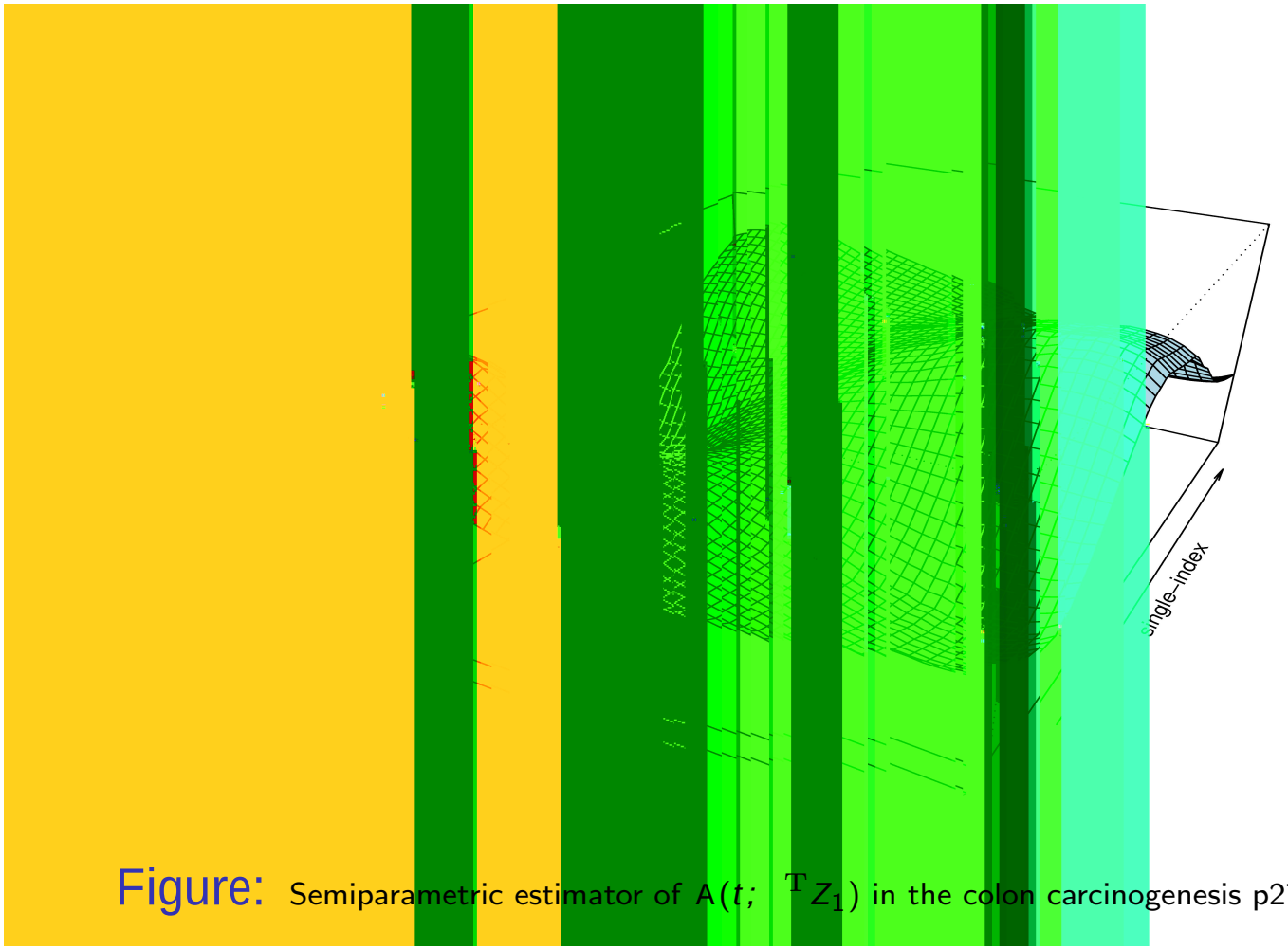



Figure: Semiparametric estimator of $A(t; {}^{T}Z_1)$ in the colon carcinogenesis p27 data.

# The Single Index

- The main e ects for $Z$ in $Z^{\mathrm{T}}\beta$ were modest but dominated by an increase in Apoptosis for the sh oil diet.
- The interaction $\theta$ was dominated by a highly signi cant e ect of butyrate exposure and the proliferative index.
- If we ignore the possible interaction, NOTHING is statistically signi cant in the resulting model, including sh oil intake.

# Interaction

- If $\mathcal{S}(\cdot)$ is constant, then there should be no relationship between it and $X$, the p27 biomarker or on $Y$
- By implication, there should thus be no relationship between $\mathcal{S}(\cdot)$ and the PC scores.
- We thus first divide the function estimates $\widehat{\mathcal{S}}(\cdot)$ into three subgroups:
  - high values of $\mathcal{S}(\cdot)$, ($S_{\text{high}}$: $\mathcal{S} > 1.5$);
  - low values of $\mathcal{S}(\cdot)$ ($S_{\text{low}}$: $\mathcal{S} < -1.5$);
  - and the ones that are in between ($S_{\text{mid}}$).

# Interaction

- We then dichotomized each of the two p27 PC scores according whether they belong to the top or bottom 50% of the scores.
- This produces 4 groups in the data: PC1-Low, PC1-High, PC2-Low and PC2-High.
- If $\mathcal{S}(\cdot)$ is constant, there there should be no systematic relationship between $\widehat{\mathcal{S}}(\cdot)$ and the average apoptotic index of these four PC groups.
- The next graph shows a clear relationship.

# Fitted model

| | 11 | 12 | 21 | 22 |
|---|---|---|---|---|
| Estimate | -0.0004 | 0.0235 | -0.0480 | 0.9988 |
| SE | 0.0108 | 0.1003 | 0.1653 | 0.0767 |
| p-value | 0.9719 | 0.8145 | 0.7714 | 0.0000 |

| | 0 | fish | buty | prolif |
|---|---|---|---|---|
| Estimate | 0.2627 | 0.0514 | 0.0223 | -0.0062 |
| SE | 0.0247 | 0.0193 | 0.0201 | 0.0135 |
| p-value | 0.0000 | 0.0078 | 0.2667 | 0.6484 |

| | fish | buty | prolif |
|---|---|---|---|
| Estimate | 0.4208 | -0.7143 | -0.5592 |
| SE | 0.2847 | 0.2419 | 0.2005 |
| p-value | 0.1394 | 0.0031 | 0.0053 |