Functional Latent Feature Regression Models For Data With Longitudinal Covariate Process

Naisyin Wang
Texas A&M University

Joint with Li, E., Li, YH, Wang, NY

University of Michigan, January 16, 2009

Outline

- Parametric Longitudinal Covariate Process
  – numerical outcomes to show issues to be considered.
- Examples: JHU Precursor study.
- Function Linear Model and Functional Latent Feature Model.
- Sufficiency score method and its properties.
- Concluding Remarks.

Latent Covariate Regression

- JHU Precursor Hypertension Study
  - Goal: Effects of life-time longitudinal processes on the latter-in-life health outcomes.
  - 782 subjects from Precursor study.
  - Covariates: baseline average SBP, mean and slope of BMI; Response, Hypertension status after age 40.
  - Logistic regression primary model (of Y_i and ξ_i) $+$ latent covariate model:

$$
\begin{aligned}
W_{ij'}^{SBP} &= \xi_i^{SBP} + U_{ij'}^{SBP}, \\
W_{ij}^{BMI} &= \xi_{1i}^{BMI} + \xi_{2i}^{BMI}\text{age}_{ij} + U_{ij}^{BMI}.
\end{aligned}
$$

- A long literature for survival data (joint modeling): reviewed by Tsiatis & Davidian (2004); Taylor and co-authors (surrogate longitudinal markers).

- Focus on GLM for Y & X: Li, et al (2004).

$$W_i = D_i \xi_i + U_i, U_i \sim \mathcal{N}(0, \Sigma_u);\ W_i = (W_{[1]i}^T, \ldots, W_{[K]i}^T)^T.$$

♠ Possible choice of Σ_u:

♠ unspecified,

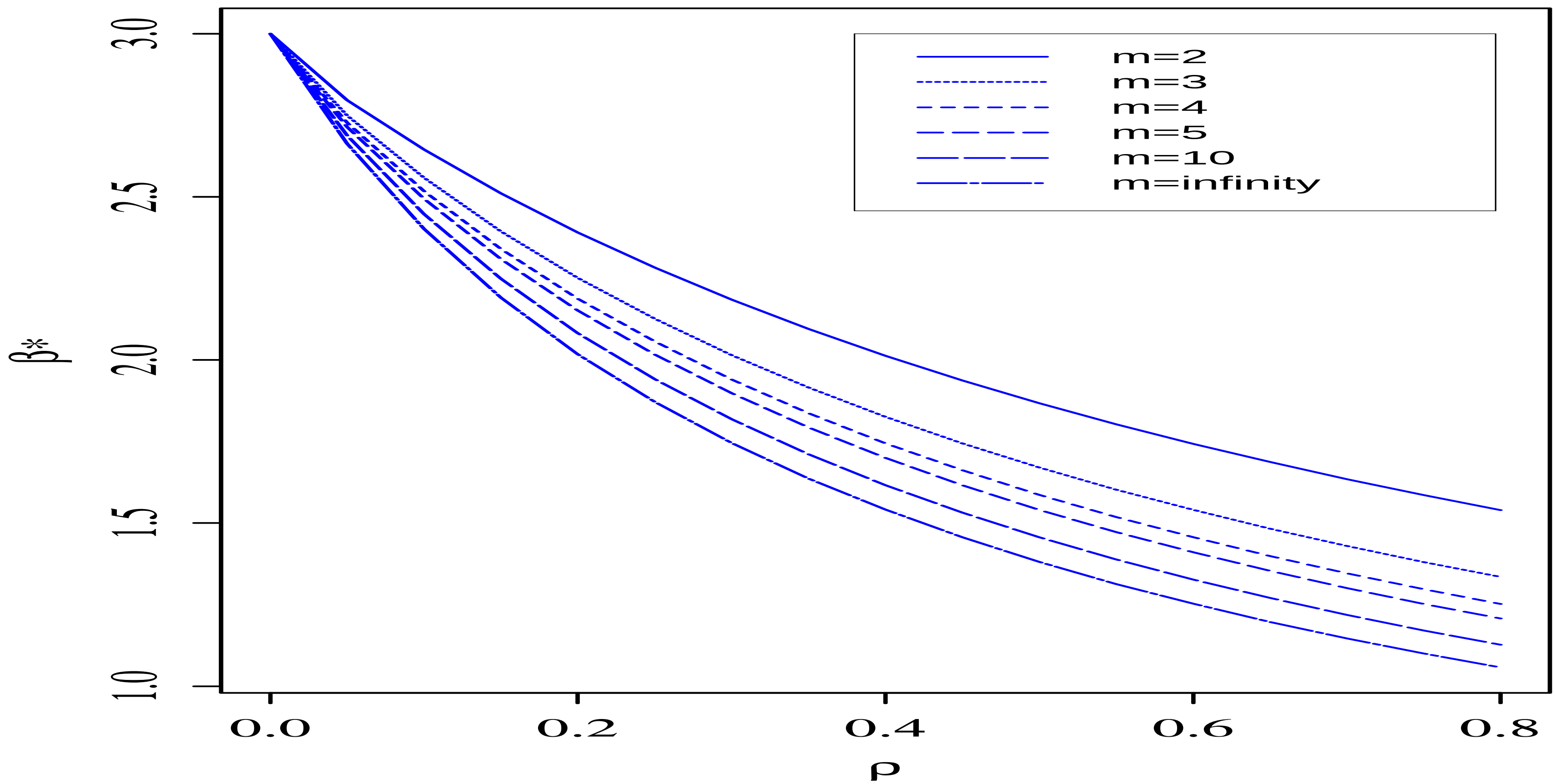
♠ block-diagonal* – specified process by process,

♠ diagonal (WI) – most common.

- Leave the distributions of ξ_i completely unspecified.

Bias Consideration in Joint Modeling

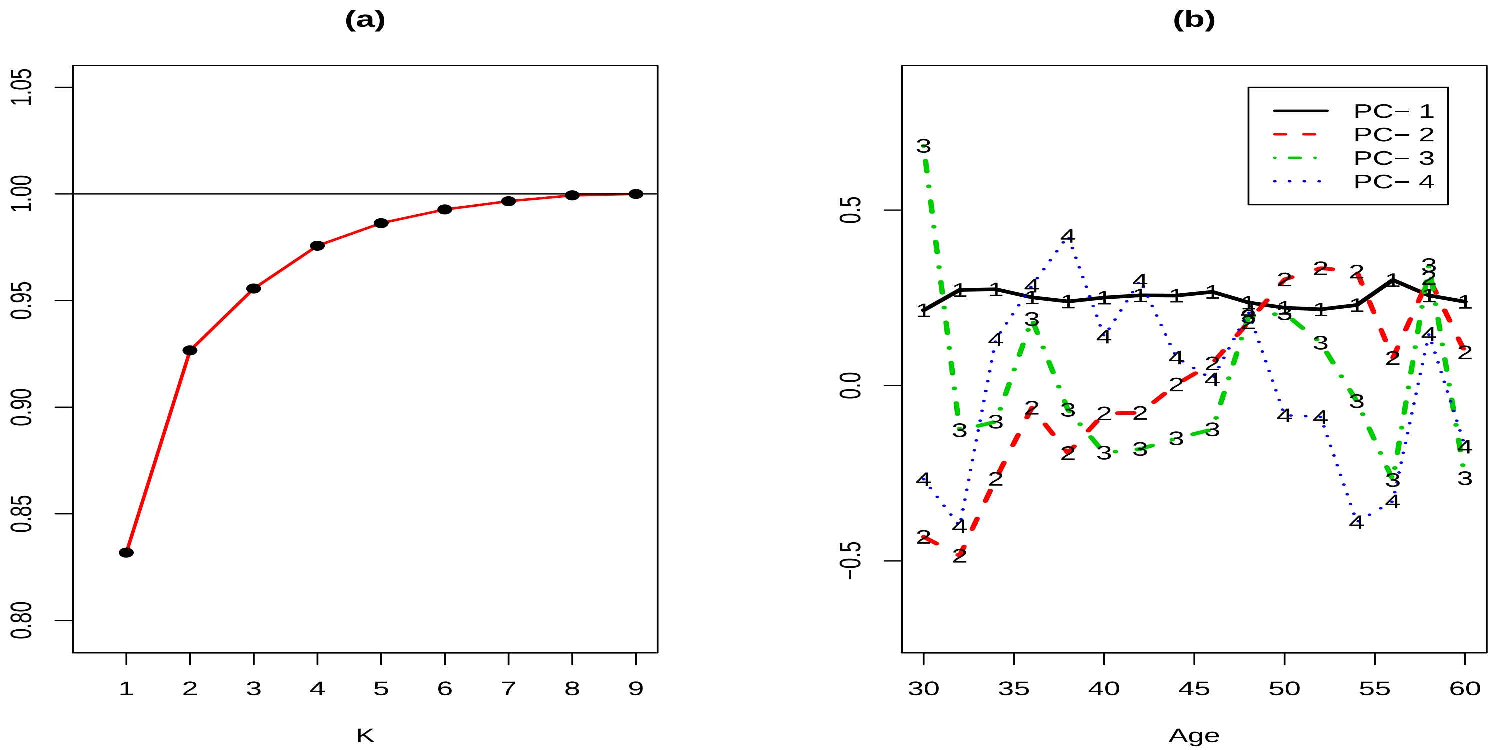
- Focus on sufficient score approach (Li, Wang & Wang, 2007);
  - Sufficient score methods: Quasilikelihood method with likelihood of Y given S, where S is the "sufficient statistic" of X given Y and W. Logistic regression: $S_i = Y_i\beta + D_i^T \Sigma_u^{-1} W_i$. Ref: Stefanski & Carroll (1987); Tsiatis & Davidian (2001).
  - We use a simple numerical study to illustrate the concerns regarding ignoring correlations.
  - Setup: binary Y from logistic regression with mean $H(\beta_0 + \beta_1 X)$; $\beta_0 = -2.5$ and $\beta_1 = 3$. $W_i = 1\xi_i + U_i$, $U_i \sim \mathcal{N}(0, \Sigma_u)$; $\Sigma_u = \sigma^2\{(1-\rho)I + \rho J\}$.
  - Results verified by simulation studies. Bias is worse when X is skew normal or t.



Bias Analysis for β^*

Method	RB (%)	SD	CP	RB (%)	SD	CP
$\rho = 0.25$	ξ_i Normal			ξ_i Bimodal mixture		
RRC	-16.8	0.27	0.52	6.7	0.43	0.97
GRRC	1.9	0.44	0.96	44.8	1.06	0.99
SS	-16.1	0.28	0.52	-7.3	0.30	0.83
GSS	3.8	0.50	0.96	3.5	0.44	0.96
GCS	3.7	0.49	0.96	3.7	0.43	0.96
	ξ_i Skew-normal			ξ_i Bivariate t_5		
RRC	-17.0	0.27	0.52	-24.0	0.25	0.22
GRRC	1.5	0.43	0.95	-8.8	0.37	0.82
SS	-16.5	0.27	0.52	-19.6	0.28	0.44
GSS	2.8	0.47	0.96	4.1	0.55	0.96
GCS	2.8	0.46	0.95	4.0	0.51	0.95

- Outcomes of parametric data analysis with block-diagonal Σ_u:
  - baseline SBP is significant at 5% level (increasing effect).
  - Conditional on that, BMI slope is significant (increasing effect), but average BMI is not.
  - The relative changes of the estimated coefficients from WI to Cov for intercept, baseline SBP, average BMI, BMI slope are, in the relative scale, 13%, 25%, 100% and 2%, respectively.
  - Testing conclusions remain the same w/ or w/o accounting for covariance but the CI's are at different places.
  - Is the use of BMI latent intercepts, slopes sufficient? Are they meaningful?



Functional PC of BMI

Eigen Values: 99.8 (83.2%), 11.4 (9.5%), 3.5 (2.9%), 2.4 (2.0%)

Functional Linear/Latent-Feature Models

- Observe Y_i, Z_i, $W_i(t_{ij})$ and consider
  $W_i(t) = \sum_{k=1}^{\infty} \xi_{ik}\, \phi_k(t) +$ white noise errors, where
  $\phi_k(t)$: functional eigen functions; $\phi_k(t)$, an orthonormal basis.
  $\lambda_k =$var(ξ_{ik}):eigen-values ($\lambda_k \geq \lambda_{k+1}$)

- Consider $\beta(t) = \sum_k b_k \phi_k(t)$ and $Y_i = \int \beta(t) X_i(t) dt +$ error.
  The primary model can be written as $Y_i = \sum_k b_k \xi_{ik} +$ error.

  - Take ξ_{ik} as latent features, the latent covariate model is
    $W_i(t) = \sum_{k=1}^{K} \xi_{ik}\, \phi_k(t) + U_i(t)$.
  - In traditional FLM, $K \to \infty$ and $\phi_k(t)$ is consistently estimated with negligible error.

- It is commonly
  - Use an approximate AIC to select K ; often $K \leq 4$.
  - Ignore the errors after truncation: use WI Σ_U.
  - Asymptotic for $\hat{\phi}_k$ requires λ_k and $\lambda_\ell - \lambda_k$ bounded away from 0 (Hall and Hosseini-Nasab, 2006).
- Like the parametric case, correlated ME might cause bias. In FPC setup, un-used ξ_{ik} for $k > K$ could cause correlations.
- Not an issue when $\lambda_k \to 0$ fastly, "zero-coefficient" for ξ_{ik} for "large" K and $\phi_k(t)$ being precisely estimated.

- In Latent Feature Regression:
  - We pick a fixed K – the first K features (ξ_{ik} uncorrelated).
    - We take into account variation due to $\phi_k(t)$ being estimated.
    - Using sufficiency score, no need to "estimate" ξ_{ik}.

- Let S_i be the sufficient statistics for $\xi_{[K],i} = (\xi_{i1}, \cdots, \xi_{iK})^T$; the unknown parameters in S_i contain $\phi_k(t)$ (D)and $R_U(s,t)$ $=$ Cov$\{U_i(s), U_i(t)\}$(Σ_u).
- The sufficiency score estimation is as in the parametric case. We use kernel method to estimate $R_W(s,t) =$ Cov$\{W_i(s), W_i(t)\}$ and consequently, $\hat{\phi}_k(t)$ and $\hat{R}_U(s,t)$.
- Estimate of $R_W(s,t)$, $R_U(s,t)$ and $\phi_k(t)$ use only W.
- Under requalrity conditions with $\delta_n = h^2 + \log(n)(nh^2)^{-1/2}$,

$$sup_{s,t}|\widehat{R}_X(s,t) - R_X(s,t)| = O_p(\delta_n),$$

$$sup_{s,t}|\widehat{R}_W(s,t) - R_W(s,t)| = O_p(\delta_n).$$

- $\star$ Let $\widetilde{K}_{k,h}(t, t_{ij'}) = \sum_{\ell \neq k}(\lambda_k - \lambda_\ell)^{-1}\phi_\ell(t)\int \phi_\ell(u)K_h(u, t_{ij'})du$, $K_h(s,t) = h^{-1}K\{(s-t)/h\}$, $W_{jj',i} = W_{ij}W_{ij'}$, and $W^*_{jj',i} = W_{jj',i} - \mathrm{E}(W_{jj',i})$, we have $\{\widehat{\phi}_k(t) - \phi_k(t)\}\{1 + o_p(n^{-1/2})\}$ $= \frac{1}{nm(m-1)}\sum_{i=1}^n \sum_{j \neq j'} f_2^{-1}(t_{ij}, t_{ij'})\phi_k(t_{ij})\widetilde{K}_{k,h}(t, t_{ij'})W^*_{jj',i}$.

- $\star$ Let $\mathbf{\Theta}$ denote the parameters of main interest, which includes b_k, $\widehat{\mathbf{\Theta}}_{[K]}$ is root-n consistent and

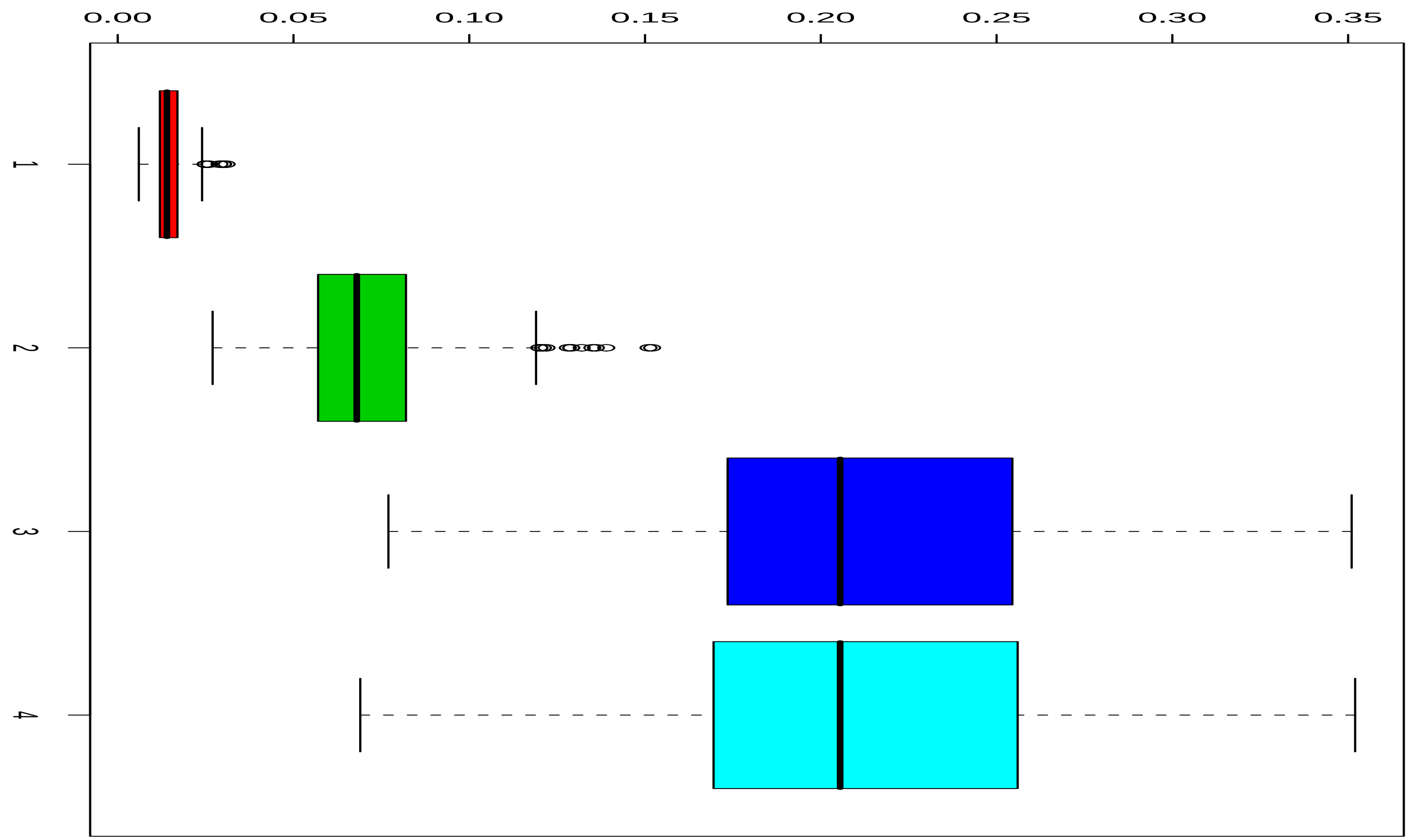
$$\sqrt{n}(\mathbf{\Theta}_{[K]} - \mathbf{\Theta}_{[K]}) \xrightarrow{d} \text{Normal}\{\mathbf{0}, \Gamma_{\mathbf{\Theta}}^{-1}(\Xi_{\mathbf{\Theta}} + \Pi_{\mathbf{\Theta}})\Gamma_{\mathbf{\Theta}}^{-1}\},$$

In comparison, if $\phi_k(t)$ and $R_U(s,t)$ were known,

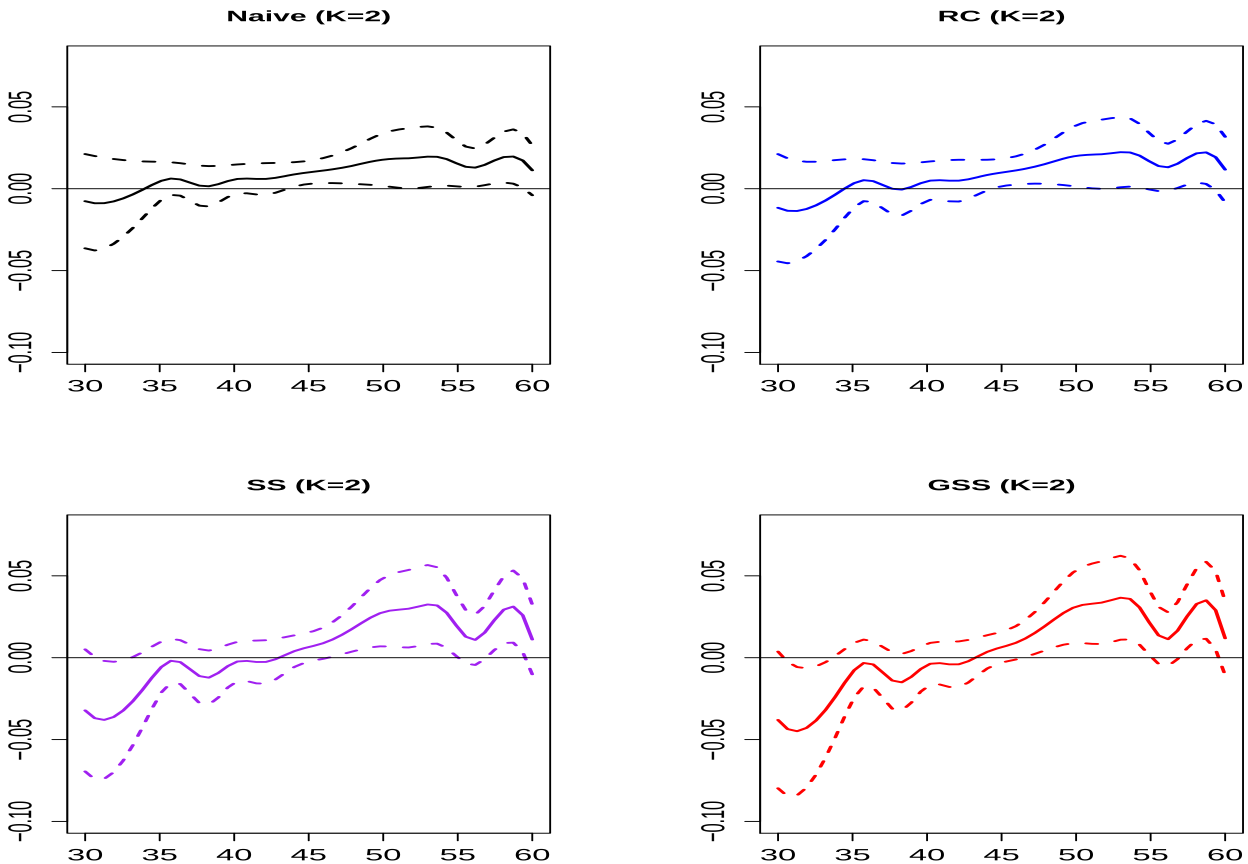
$$\sqrt{n}(\mathbf{\Theta}^*_{[K]} - \mathbf{\Theta}_{[K]}) \xrightarrow{d} \text{Normal}\{\mathbf{0}, \Gamma_{\mathbf{\Theta}}^{-1}(\Xi_{\mathbf{\Theta}})\Gamma_{\mathbf{\Theta}}^{-1}\},$$

- For the SBP-BMI data, the approximate AIC selected $K = 3$.
  - For $K = 2$, the naive estimated coefficients for SBP and BMI FPC directions are $.054^{*}$ $.034^{*}$ and $.034$ while the GSS estimates are $.129^{*}$, $.012$ and $.094^{*}$.
  - What we observed above are mainly attenuation.
  - Naive estimate would conclude the increasing trend in BMI does not contribute to hypertension.

- A Box-Plot to show estimation errors for the four FPC directions.



Root-MSE for estimated Eigen Functions
Median Correlations: .998, .962, .662, .662.



Estimated $\hat{\beta}(t)$

	RC			GSS		
	Est	s.e.	p-val	Est	s.e.	p-val
(a) $K = 3$						
Intercept	-13.769	2.687	0.00	-20.154	4.896	0.00
$\mathrm{SBP}_{\mathrm{baseline}}$	0.105	0.022	0.00	0.157	0.040	0.00
$\mathrm{BMI}_{\mathrm{PC1}}$	0.037*	0.010	0.01	0.018	0.013	0.16
$\mathrm{BMI}_{\mathrm{PC2}}$	0.021	0.036	0.55	0.078	0.049	0.11
$\mathrm{BMI}_{\mathrm{PC3}}$	-0.128	0.118	0.28	-0.143	0.171	0.65
(b) $K = 2$						
Intercept	-13.776	2.566	0.00	-16.770	4.003	0.00
$\mathrm{SBP}_{\mathrm{baseline}}$	0.105	0.021	0.00	0.129	0.033	0.00
$\mathrm{BMI}_{\mathrm{PC1}}$	0.032*	0.010	0.00	0.012	0.011	0.25
$\mathrm{BMI}_{\mathrm{PC2}}$	0.043	0.029	0.13	0.094*	0.036	0.01

Regression parameter estimates for the hypertension study.

- Consider a simple simulation study,
  - $\phi_k(t) = \sqrt{2}\sin(\pi k t)$, $\xi_{ik} \sim [0, 1/k^{1.1}]$, $U_i \sim [0, .09]$.
  - $W_i(t_{ij}) = \sum_{k=1}^{10} \xi_{ik}\, \phi_k(t_{ij}) + U_i(t_{ij})$,
  - FPC coefficients for $\beta(t)$ are $\theta_1 = 3$, $\theta_2 = 1.5$, $\theta_3 = 1.5$, $\theta_4 = 1.5$, $\theta_k = 0$ for $k \geq 5$.
  - $n = 500$; $m = 25.t_{ij} = j/m + \mathcal{N}(0, 0.0001)$.

- We consider ξ_{ik} were normal or t_3.
  - For normally distributed ξ_{ik}, the range of absolute relative biases of $(\hat\theta_1, \hat\theta_2, \hat\theta_3, \hat\theta_4)^T$ for naive, RC, SS and GSS were 24.4% to 50.8%, 11.8% to 17.6%, 13.8% to 22.7% and 4.3% to 5.8%, respectively. For t_3, RC performs much worse (17.6% to 29.3%) and GSS and SS suffer in $\hat\theta_4$.

	Naive	RC	SS	GSS
		$\xi_{ik} \sim$ normal		
IBias	4.894	2.433	3.048	1.081
IVar	0.374	0.816	2.443	1.396
IMSE	2.146	1.172	2.994	1.477
		$\xi_{ik} \sim t_3$		
IBias	6.436	4.328	3.270	2.248
IVar	0.453	1.030	3.656	2.189
IMSE	3.372	2.231	4.324	2.508

Concluding Remarks

- The basic principle behind the ME consideration in parametric and nonparametric modeling should be the same.

- Nonparametic functional linear model might require a bit more caution.

- Modeling and selection of covariance structure in non- and semiparametric longitudina/functional models deserve attention.