A simultaneous confidence corridor for varying coefficient regression with sparse functional data

Lijie Gu, Soochow University

Lily Wang, Iowa State University

Wolfgang K. Härdle, Humboldt-Universität zu Berlin

Lijian Yang, Soochow University

Published in TEST 23 (4), 806-843, 2014

Outline

- Motivation
- Assumptions
- Main theoretical results
- Implementation details
- Simulation
- Real data analysis
- Conclusions

Motivation

- Linear regression model: $Y = \beta_1 T_1 + \beta_2 T_2 + \cdots + \beta_d T_d + \varepsilon$

Motivation

- Linear regression model: $Y = \beta_1 T_1 + \beta_2 T_2 + \cdots + \beta_d T_d + \varepsilon$
- Varying coefficient model (VCM): Hastie and Tibshirani (1993)

Motivation

- Linear regression model: $Y = \beta_1 T_1 + \beta_2 T_2 + \cdots + \beta_d T_d + \varepsilon$

- Varying coefficient model (VCM): Hastie and Tibshirani (1993)

- VCM: $Y = \beta_1 (X_1) T_1 + \beta_2 (X_2) T_2 + \cdots + \beta_d (X_d) T_d + \varepsilon$

Motivation

- Linear regression model: $Y = \beta_1 T_1 + \beta_2 T_2 + \cdots + \beta_d T_d + \varepsilon$

- Varying coefficient model (VCM): Hastie and Tibshirani (1993)

- VCM: $Y = \beta_1 (X_1) T_1 + \beta_2 (X_2) T_2 + \cdots + \beta_d (X_d) T_d + \varepsilon$

- Strikes a delicate balance of simplicity and flexibility.

Motivation

- Linear regression model: $Y = \beta_1 T_1 + \beta_2 T_2 + \cdots + \beta_d T_d + \varepsilon$
- Varying coefficient model (VCM): Hastie and Tibshirani (1993)
- VCM: $Y = \beta_1(X_1) T_1 + \beta_2(X_2) T_2 + \cdots + \beta_d(X_d) T_d + \varepsilon$
- Strikes a delicate balance of simplicity and flexibility.
- Cobb-Douglas model for GDP growth in Liu and Yang (2010). Longitudinal model for CD4 percentages in Wu and Chiang (2000), Fan and Zhang (2000), Wang, Li and Huang (2008).

Motivation

1817 observations of CD4 cell percentages on 283 patients:

- $X_{i0} \equiv 1$: baseline.
- X_{i1}: smoking status; (smoking, $X_{i1} = 1$; nonsmoking, $X_{i1} = 0$);
- X_{i2}: centered pre-infection CD4 percentage;
- X_{i3}: centered age at the time of HIV infection;
- T_{ij}: the time (in years) of the j-th measurement on the i-th patient;
- Y_{ij}: the measurement of CD4 cell percentage at time T_{ij};
- N_i: the number of measurements for the i-th patient, $1 \leq N_i \leq 14$;
- $N_T = \sum_{i=1}^{n} N_i$: the total sample size;

Motivation

1817 observations of CD4 cell percentages on 283 patients:

- $X_{i0} \equiv 1$: baseline.
- X_{i1}: smoking status; (smoking, $X_{i1} = 1$; nonsmoking, $X_{i1} = 0$);
- X_{i2}: centered pre-infection CD4 percentage;
- X_{i3}: centered age at the time of HIV infection;
- T_{ij}: the time (in years) of the j-th measurement on the i-th patient;
- Y_{ij}: the measurement of CD4 cell percentage at time T_{ij};
- N_i: the number of measurements for the i-th patient, $1 \leq N_i \leq 14$;
- $N_T = \sum_{i=1}^{n} N_i$: the total sample size;

Idea: Modelling Y_{ij} as linearly dependent on $X_{il}, l = 0, 1, 2, 3$ with coefficients as functions of T_{ij}.

Motivation

- Observed for i-th subject $(1 \le i \le n)$: $\{\mathbf{X}_i, T_{ij}, Y_{ij}\}, 1 \le j \le N_i$, where $\mathbf{X}_i = (X_{i1}, \ldots, X_{id})$.

Motivation

- Observed for i-th subject $(1 \leq i \leq n)$: $\{\mathbf{X}_i, T_{ij}, Y_{ij}\}, 1 \leq j \leq N_i$, where $\mathbf{X}_i = (X_{i1}, \ldots, X_{id})$.
- A feasible VCM for sparse functional data:

$$Y_{ij} = \sum_{l=1}^{d} \eta_{il}(T_{ij}) X_{il} + \sigma(T_{ij}) \varepsilon_{ij}, \quad 1 \leq i \leq n, \ 1 \leq j \leq N_i$$

Motivation

- Observed for i-th subject ($1 \leq i \leq n$): $\{\mathbf{X}_i, T_{ij}, Y_{ij}\}, 1 \leq j \leq N_i$, where $\mathbf{X}_i = (X_{i1}, \ldots, X_{id})$.
- A feasible VCM for sparse functional data:

$$Y_{ij} = \sum_{l=1}^{d} \eta_{il}(T_{ij}) X_{il} + \sigma(T_{ij}) \varepsilon_{ij}, \quad 1 \leq i \leq n, \; 1 \leq j \leq N_i$$

- Karhunen-Loève L^2 representation:

$$\eta_{il}(t) = m_l(t) + \sum_{k=1}^{\infty} \xi_{ik,l} \phi_{k,l}(t), \quad t \in \mathcal{T}, \quad l = 1, \ldots, d$$

Motivation

- Observed for i-th subject ($1 \leq i \leq n$): $\{\mathbf{X}_i, T_{ij}, Y_{ij}\}, 1 \leq j \leq N_i$, where $\mathbf{X}_i = (X_{i1}, \ldots, X_{id})$.
- A feasible VCM for sparse functional data:

$$Y_{ij} = \sum_{l=1}^{d} \eta_{il}(T_{ij}) X_{il} + \sigma(T_{ij}) \varepsilon_{ij}, \quad 1 \leq i \leq n, \ 1 \leq j \leq N_i$$

- Karhunen-Loève L^2 representation:

$$\eta_{il}(t) = m_l(t) + \sum_{k=1}^{\infty} \xi_{ik,l} \phi_{k,l}(t), \quad t \in \mathcal{T}, \quad l = 1, \ldots, d$$

- Unknowns: $m_l(t) = E\{\eta_l(t)\}$, $l = 1, \ldots, d$,

$$G_l(s,t) = \mathsf{cov}\{\eta_l(s), \eta_l(t)\} = \sum_{k=1}^{\infty} \phi_{k,l}(s)\phi_{k,l}(t)$$

Motivation

- Model:

$$Y_{ij} = \sum_{l=1}^{d} m_l \left(T_{ij}\right) X_{il} + \sum_{l=1}^{d} \sum_{k=1}^{\infty} \xi_{ik,l} \phi_{k,l} \left(T_{ij}\right) X_{il} + \sigma \left(T_{ij}\right) \varepsilon_{ij}$$

Motivation

- Model:

$$Y_{ij} = \sum_{l=1}^{d} m_l \left(T_{ij}\right) X_{il} + \sum_{l=1}^{d} \sum_{k=1}^{\infty} \xi_{ik,l} \phi_{k,l} \left(T_{ij}\right) X_{il} + \sigma \left(T_{ij}\right) \varepsilon_{ij}$$

- Goals:
  - estimate the unknown **coefficient functions** $m_l(t), l = 1, \ldots, d$;
  - focus on constructing **simultaneous confidence corridors (SCCs)** for $m_l(t), l = 1, \ldots, d$ with asymptotically correct confidence level.

Motivation

- Model:

$$Y_{ij} = \sum_{l=1}^{d} m_l (T_{ij}) X_{il} + \sum_{l=1}^{d} \sum_{k=1}^{\infty} \xi_{ik,l} \phi_{k,l} (T_{ij}) X_{il} + \sigma (T_{ij}) \varepsilon_{ij}$$

- Goals:
  - estimate the unknown **coefficient functions** $m_l(t), l = 1, \ldots, d$;
  - focus on constructing **simultaneous confidence corridors (SCCs)** for $m_l(t), l = 1, \ldots, d$ with asymptotically correct confidence level.
- SCC: A sliding confidence interval for $m_l(t)$ over $t \in \mathcal{T}$

$$\lim_{n \to \infty} P\{m_l(t) \in I_t, \text{for all } \ t \in \mathcal{T}\} = 1 - \alpha$$

Motivation

- Model:

$$Y_{ij} = \sum_{l=1}^{d} m_l\left(T_{ij}\right) X_{il} + \sum_{l=1}^{d} \sum_{k=1}^{\infty} \xi_{ik,l} \phi_{k,l}\left(T_{ij}\right) X_{il} + \sigma\left(T_{ij}\right) \varepsilon_{ij}$$

- Goals:
  - estimate the unknown **coefficient functions** $m_l(t), l = 1, \ldots, d$;
  - focus on constructing **simultaneous confidence corridors (SCCs)** for $m_l(t), l = 1, \ldots, d$ with asymptotically correct confidence level.
- SCC: A sliding confidence interval for $m_l\left(t\right)$ over $t \in \mathcal{T}$

$$\lim_{n \to \infty} P\{m_l(t) \in I_t, \text{for all } \; t \in \mathcal{T}\} = 1 - \alpha$$

- More appropriate inference on the whole curve $m_l\left(t\right)$

Motivation

- Model:

$$Y_{ij} = \sum_{l=1}^{d} m_l\left(T_{ij}\right) X_{il} + \sum_{l=1}^{d} \sum_{k=1}^{\infty} \xi_{ik,l} \phi_{k,l}\left(T_{ij}\right) X_{il} + \sigma\left(T_{ij}\right) \varepsilon_{ij}$$

- Goals:
  - estimate the unknown **coefficient functions** $m_l(t), l = 1, \ldots, d$;
  - focus on constructing **simultaneous confidence corridors (SCCs)** for $m_l(t), l = 1, \ldots, d$ with asymptotically correct confidence level.

- SCC: A sliding confidence interval for $m_l\left(t\right)$ over $t \in \mathcal{T}$

$$\lim_{n \to \infty} P\{m_l(t) \in I_t, \text{for all } \ t \in \mathcal{T}\} = 1 - \alpha$$

- More appropriate inference on the whole curve $m_l\left(t\right)$
- In the following, take $\mathcal{T} = [0, 1]$

- A sequence of interior knots $\{\gamma_J\}_{J=1}^{N_{\mathrm{s}}}$ with $h_{\mathrm{s}} = 1/\left(N_{\mathrm{s}}+1\right)$

$$\gamma_0 = 0 < \gamma_1 < \cdots < \gamma_{N_{\mathrm{s}}} < 1 = \gamma_{N_{\mathrm{s}}+1}, \gamma_J = Jh_{\mathrm{s}},\ J = 0, ..., N_{\mathrm{s}}+1,$$

- Subintervals: $\chi_J = \left[\gamma_J, \gamma_{J+1}\right), J = 0, ..., N_{\mathrm{s}}-1, \chi_{N_{\mathrm{s}}} = \left[\gamma_{N_{\mathrm{s}}}, 1\right]$

- A sequence of interior knots $\{\gamma_J\}_{J=1}^{N_\mathrm{s}}$ with $h_\mathrm{s} = 1/\left(N_\mathrm{s}+1\right)$

$$\gamma_0 = 0 < \gamma_1 < \cdots < \gamma_{N_\mathrm{s}} < 1 = \gamma_{N_\mathrm{s}+1}, \gamma_J = Jh_\mathrm{s},\ J = 0, ..., N_\mathrm{s}+1,$$

- Subintervals: $\chi_J = \left[\gamma_J, \gamma_{J+1}\right), J = 0, ..., N_\mathrm{s}-1, \chi_{N_\mathrm{s}} = \left[\gamma_{N_\mathrm{s}}, 1\right]$
- The space of functions that are constant on each χ_J as

$$G^{(-1)} = \left\{ \sum\nolimits_{J=0}^{N_s} \lambda_J b_J(t) \middle| \lambda_J \in \mathbb{R},\ b_J\left(t\right) = I_{\chi_J}\left(t\right), J = 0, \ldots, N_s \right\}$$

- A sequence of interior knots $\{\gamma_J\}_{J=1}^{N_{\mathrm{s}}}$ with $h_{\mathrm{s}} = 1/(N_{\mathrm{s}}+1)$

$$\gamma_0 = 0 < \gamma_1 < \cdots < \gamma_{N_{\mathrm{s}}} < 1 = \gamma_{N_{\mathrm{s}}+1}, \gamma_J = Jh_{\mathrm{s}},\ J = 0, ..., N_{\mathrm{s}}+1,$$

- Subintervals: $\chi_J = [\gamma_J, \gamma_{J+1}), J = 0, ..., N_{\mathrm{s}}-1, \chi_{N_{\mathrm{s}}} = [\gamma_{N_{\mathrm{s}}}, 1]$
- The space of functions that are constant on each χ_J as

$$G^{(-1)} = \left\{ \sum_{J=0}^{N_s} \lambda_J b_J(t) \middle| \lambda_J \in \mathbb{R},\ b_J(t) = I_{\chi_J}(t), J = 0, \ldots, N_s \right\}$$

- The space of spline coefficient functions on $\mathcal{T} \times \mathbb{R}^d$ as

$$\mathcal{M} = \left\{ g(t, \mathbf{x}) = \sum_{l=1}^{d} g_l(t) x_l : g_l(t) \in G^{(-1)}, t \in \mathcal{T}, \mathbf{x} = (x_1, \ldots, x_d)^{\mathsf{T}} \in \mathbb{R}^d \right\}$$

- A sequence of interior knots $\{\gamma_J\}_{J=1}^{N_{\mathrm{s}}}$ with $h_{\mathrm{s}} = 1/(N_{\mathrm{s}}+1)$

$$\gamma_0 = 0 < \gamma_1 < \cdots < \gamma_{N_{\mathrm{s}}} < 1 = \gamma_{N_{\mathrm{s}}+1}, \gamma_J = Jh_{\mathrm{s}},\ J = 0, ..., N_{\mathrm{s}}+1,$$

- Subintervals: $\chi_J = [\gamma_J, \gamma_{J+1})\,, J = 0, ..., N_{\mathrm{s}} - 1, \chi_{N_{\mathrm{s}}} = [\gamma_{N_{\mathrm{s}}}, 1]$
- The space of functions that are constant on each χ_J as

$$G^{(-1)} = \left\{ \sum\nolimits_{J=0}^{N_s} \lambda_J b_J(t) \middle| \lambda_J \in \mathbb{R},\ b_J(t) = I_{\chi_J}(t)\,, J = 0, \ldots, N_s \right\}$$

- The space of spline coefficient functions on $\mathcal{T} \times \mathbb{R}^d$ as

$$\mathcal{M} = \left\{ g(t, \mathbf{x}) = \sum_{l=1}^{d} g_l(t) x_l : g_l(t) \in G^{(-1)}, t \in \mathcal{T}, \mathbf{x} = (x_1, \ldots, x_d)^{\mathsf{T}} \in \mathbb{R}^d \right\}$$

- The collection of order β Hõlder continuous function on $[0, 1]$ as

$$C^{0,r}[0,1] = \left\{ \phi : \|\phi\|_{0,r} = \sup\nolimits_{t \neq t', t, t' \in [0,1]} \frac{|\phi(t) - \phi(t')|}{|t - t'|^r} < +\infty \right\}$$

Assumptions

(A1) *The regression functions* $m_l(t) \in C^{0,1}[0,1]$, $l = 1, \ldots, d$.

(A2) *The set of random variables* $\left(T_{ij}, \varepsilon_{ij}, N_i, \xi_{ik,l}, X_{il}\right)_{i=1,j=1,k=1,l=1}^{n,N_i,\infty,d}$ *is a subset of variables* $\left(T_{ij}, \varepsilon_{ij}, N_i, \xi_{ik,l}, X_{il}\right)_{i=1,j=1,k=1,l=1}^{\infty,\infty,\infty,d}$ *consisting of independent random variables, in which all* T_{ij}*'s i.i.d with* $T_{ij} \sim T$*, where* T *is a random variable with probability density function* $f(t)$*;* X_{il}*'s i.i.d for each* $l = 1, \ldots, d$*;* N_i*'s i.i.d with* $N_i \sim N$*, where* $N > 0$ *is a positive integer-valued random variable with* $\mathsf{E}\{N^{2r}\} \leq r! c_N^r$, $r = 2, 3, \ldots$, *for some constant* $c_N > 0$*. Variables* $\left(\xi_{ik,l}\right)_{i=1,k=1,l=1}^{\infty,\infty,d}$ *and* $\left(\varepsilon_{ij}\right)_{i=1,j=1}^{\infty,\infty}$ *are i.i.d* $N(0,1)$.

(A3) *The functions* $f(t)$, $\sigma(t)$ *and* $\phi_{k,l}(t) \in C^{0,r}[0,1]$ *for some* $r \in (0,1]$ *with* $f(t) \in [c_f, C_f]$, $\sigma(t) \in [c_\sigma, C_\sigma]$, $t \in [0,1]$, *for constants* $0 < c_f \leq C_f < +\infty$, $0 < c_\sigma \leq C_\sigma < +\infty$.

Assumptions

(A4) *For* $l = 1, \ldots, d$, $\sum_{k=1}^{\infty} \left\|\phi_{k,l}\right\|_{\infty} < +\infty$, *and* $G_l(t,t) \in [c_{G,l}, C_{G,l}]$, $t \in [0,1]$, *for constants* $0 < c_{G,l} \leq C_{G,l} < +\infty$.

(A5) *There exist constants* $0 < c_{\mathbf{H}} \leq C_{\mathbf{H}} < +\infty$ *and* $0 < c_\eta \leq C_\eta < +\infty$, *such that* $c_{\mathbf{H}} I_{d\times d} \leq \mathbf{H} = \{H_{ll'}\}_{l,l'=1}^{d} = \mathsf{E}\left(\mathbf{X}\mathbf{X}^{\mathsf{T}}\right) \leq C_{\mathbf{H}} I_{d\times d}$. *For some* $\eta > 4$, $l = 1, \ldots, d$, $c_\eta \leq \mathsf{E}\left|X_l\right|^{8+\eta} \leq C_\eta$.

(A6) *As* $n \to \infty$, *the number of interior knots* $N_{\mathrm{s}} = \mathcal{o}\left(n^{\vartheta}\right)$ *for some* $\vartheta \in (1/3, 1/2)$ *while* $N_{\mathrm{s}}^{-1} = \mathcal{o}\left\{n^{-1/3}\left(\log(n)\right)^{-1/3}\right\}$. *The subinterval length* $h_{\mathrm{s}} \sim N_{\mathrm{s}}^{-1}$.

- Review the model:

$$Y_{ij} = \sum_{l=1}^{d} m_l\left(T_{ij}\right) X_{il} + \sum_{l=1}^{d} \sum_{k=1}^{\infty} \xi_{ik,l} \phi_{k,l}\left(T_{ij}\right) X_{il} + \sigma\left(T_{ij}\right) \varepsilon_{ij}$$

- Review the model:

$$Y_{ij} = \sum_{l=1}^{d} m_l \left(T_{ij}\right) X_{il} + \sum_{l=1}^{d} \sum_{k=1}^{\infty} \xi_{ik,l} \phi_{k,l} \left(T_{ij}\right) X_{il} + \sigma \left(T_{ij}\right) \varepsilon_{ij}$$

- $\hat{m}\left(t, \mathbf{x}\right) = \sum_{l=1}^{d} \hat{m}_l(t) x_l = \operatorname{argmin}_{g \in \mathcal{M}} \sum_{i=1}^{n} \sum_{j=1}^{N_i} \left\{Y_{ij} - g\left(T_{ij}, \mathbf{X}_i\right)\right\}^2$

- Review the model:

$$Y_{ij} = \sum_{l=1}^{d} m_l\left(T_{ij}\right) X_{il} + \sum_{l=1}^{d} \sum_{k=1}^{\infty} \xi_{ik,l} \phi_{k,l}\left(T_{ij}\right) X_{il} + \sigma\left(T_{ij}\right) \varepsilon_{ij}$$

- $\hat{m}\left(t, \mathbf{x}\right) = \sum_{l=1}^{d} \hat{m}_l(t) x_l = \underset{g \in \mathcal{M}}{\operatorname{argmin}} \sum_{i=1}^{n} \sum_{j=1}^{N_i} \left\{Y_{ij} - g\left(T_{ij}, \mathbf{X}_i\right)\right\}^2$

- $\sum_{l=1}^{d} \hat{m}_l(t) x_l \equiv \sum_{l=1}^{d} \sum_{J=0}^{N_s} \hat{\gamma}_{J,l} b_J(t) x_l$, where

$$\hat{\boldsymbol{\gamma}} = \underset{\boldsymbol{\gamma}=\left(\gamma_{0,1},\ldots,\gamma_{N_s,d}\right)^{\mathsf{T}} \in \mathbb{R}^{d(N_s+1)}}{\operatorname{argmin}} \sum_{i=1}^{n} \sum_{j=1}^{N_i} \left\{Y_{ij} - \sum_{l=1}^{d} \sum_{J=0}^{N_s} \gamma_{J,l} b_J\left(T_{ij}\right) X_{il}\right\}^2 .$$

- Review the model:

$$Y_{ij} = \sum_{l=1}^{d} m_l\left(T_{ij}\right) X_{il} + \sum_{l=1}^{d} \sum_{k=1}^{\infty} \xi_{ik,l} \phi_{k,l}\left(T_{ij}\right) X_{il} + \sigma\left(T_{ij}\right) \varepsilon_{ij}$$

- $\hat{m}\left(t, \mathbf{x}\right) = \sum_{l=1}^{d} \hat{m}_l(t) x_l = \operatorname{argmin}_{g \in \mathcal{M}} \sum_{i=1}^{n} \sum_{j=1}^{N_i} \left\{Y_{ij} - g\left(T_{ij}, \mathbf{X}_i\right)\right\}^2$

- $\sum_{l=1}^{d} \hat{m}_l(t) x_l \equiv \sum_{l=1}^{d} \sum_{J=0}^{N_s} \hat{\gamma}_{J,l} b_J(t) x_l$, where

$$\hat{\boldsymbol{\gamma}} = \operatorname*{argmin}_{\boldsymbol{\gamma} = \left(\gamma_{0,1}, \ldots, \gamma_{N_s,d}\right)^{\mathsf{T}} \in \mathbb{R}^{d(N_s+1)}} \sum_{i=1}^{n} \sum_{j=1}^{N_i} \left\{Y_{ij} - \sum_{l=1}^{d} \sum_{J=0}^{N_s} \gamma_{J,l} b_J\left(T_{ij}\right) X_{il}\right\}^2 .$$

- Under Assumption (A5), $\hat{m}_l(t) = \sum_{J=0}^{N_s} \hat{\gamma}_{J,l} b_J(t), \quad l = 1, \ldots, d$

- $\hat{\mathbf{m}}(t) = \{\hat{m}_l(t)\}_{l=1}^d$ has an asymptotic covariance matrix

$$\mathbf{\Sigma}_n(t) = \mathbf{H}^{-1}\mathbf{\Gamma}_n(t)\mathbf{H}^{-1} = \left\{\sigma_{n,ll'}^2(t)\right\}_{l,l'=1}^d,$$

where $\mathbf{H} = E\left(\mathbf{X}\mathbf{X}^{\mathsf{T}}\right)$,

$$\mathbf{\Gamma}_n(t) = c_{J(t),n}^{-2}\{nE(N_1)\}^{-1}E\mathbf{X}\mathbf{X}^{\mathsf{T}}\left[\int_{\chi_{J(t)}} \sigma_Y^2(u,\mathbf{X})\, f(u)\, du\right.$$

$$\left. + \frac{E\{N_1(N_1-1)\}}{EN_1}\sum_{l=1}^d X_l^2 \int_{\chi_{J(t)}\times\chi_{J(t)}} G_l(u,v)\, f(u)\, f(v)\, dudv\right],$$

$\sigma_Y^2(t,\mathbf{x}) = \mathsf{Var}(Y\,|T=t,\mathbf{X}=\mathbf{x}\,) = \sum_{l=1}^d G_l(t,t)\, x_l^2 + \sigma^2(t)$.

- $\hat{\mathbf{m}}(t) = \{\hat{m}_l(t)\}_{l=1}^d$ has an asymptotic covariance matrix

$$\mathbf{\Sigma}_n(t) = \mathbf{H}^{-1}\mathbf{\Gamma}_n(t)\mathbf{H}^{-1} = \left\{\sigma_{n,ll'}^2(t)\right\}_{l,l'=1}^d,$$

where $\mathbf{H} = E\left(\mathbf{X}\mathbf{X}^{\mathsf{T}}\right)$,

$$\mathbf{\Gamma}_n(t) = c_{J(t),n}^{-2}\{nE(N_1)\}^{-1}E\mathbf{X}\mathbf{X}^{\mathsf{T}}\left[\int_{\chi_{J(t)}} \sigma_Y^2\left(u, \mathbf{X}\right) f\left(u\right) du\right.$$

$$\left.+\frac{E\left\{N_1(N_1-1)\right\}}{EN_1}\sum_{l=1}^d X_l^2 \int_{\chi_{J(t)}\times\chi_{J(t)}} G_l\left(u,v\right) f\left(u\right) f\left(v\right) dudv\right],$$

$\sigma_Y^2(t, \mathbf{x}) = \mathsf{Var}(Y\,|T=t, \mathbf{X}=\mathbf{x}\,) = \sum_{l=1}^d G_l\left(t,t\right) x_l^2 + \sigma^2(t)$.

- $\sigma_{n,ll'}^2(t)$ is the asymptotic covariance between $\hat{m}_l(t)$ and $\hat{m}_{l'}(t)$
- $\sigma_{n,ll}^2(t)$ is the asymptotic variance of $\hat{m}_l(t)$, $l = 1, \ldots, d$

Main theoretical results

- Under Assumptions (A1)–(A6), for $\forall t \in [0,1]$,as $n \to \infty$,

$$\mathbf{\Sigma}_n^{-1/2}(t)\left\{\hat{\mathbf{m}}(t) - \mathbf{m}(t)\right\} \xrightarrow{\mathcal{L}} N\left(\mathbf{0}, \mathbf{I}_{d\times d}\right)$$

- Under Assumptions (A1)–(A6), for $\forall t \in [0,1]$, $l = 1, \ldots, d$ and any $\alpha \in (0,1)$,

$$\lim_{n\to\infty} P\left\{\sigma_{n,ll}^{-1}(t)\left|\hat{m}_l(t) - m_l(t)\right| \leq Z_{1-\alpha/2}\right\} = 1 - \alpha.$$

  while $Z_{1-\alpha/2}$ is the $100\,(1-\alpha/2)^{th}$ percentile of the standard normal distribution.

- The asymptotic $100\,(1-\alpha)\,\%$ pointwise confidence intervals (CIs) for $m_l(t)$, $t \in [0,1]$, $l = 1, \ldots, d$, are

$$\hat{m}_l(t) \pm \sigma_{n,ll}(t) Z_{1-\alpha/2}.$$

Main theoretical results

- Under Assumptions (A1)-(A6), for $l = 1, \ldots, d$ and any $\alpha \in (0,1)$,

$$\lim_{n\to\infty} P\left\{\sup_{t\in[0,1]} \sigma_{n,ll}^{-1}(t)\left|\hat{m}_l(t) - m_l(t)\right| \leq Q_{N_s+1}(\alpha)\right\} = 1-\alpha,$$

in which $Q_{N_s+1}(\alpha) = b_{N_s+1} - a_{N_s+1}^{-1}\log\left\{-\frac{1}{2}\log(1-\alpha)\right\}$,

$a_{N_\mathrm{s}+1} = \left\{2\log\left(N_\mathrm{s}+1\right)\right\}^{1/2}, b_{N_\mathrm{s}+1} = a_{N_\mathrm{s}+1} - \left(2a_{N_\mathrm{s}+1}\right)^{-1}\log\left(2\pi a_{N_\mathrm{s}+1}^2\right)$.

Main theoretical results

- Under Assumptions (A1)-(A6), for $l = 1, \ldots, d$ and any $\alpha \in (0, 1)$,

$$\lim_{n\to\infty} P\left\{\sup_{t\in[0,1]} \sigma_{n,ll}^{-1}(t)\left|\hat{m}_l(t) - m_l(t)\right| \leq Q_{N_s+1}(\alpha)\right\} = 1-\alpha,$$

in which $Q_{N_s+1}(\alpha) = b_{N_s+1} - a_{N_s+1}^{-1}\log\left\{-\frac{1}{2}\log(1-\alpha)\right\}$,

$a_{N_\mathrm{s}+1} = \{2\log(N_\mathrm{s}+1)\}^{1/2}, b_{N_\mathrm{s}+1} = a_{N_\mathrm{s}+1} - (2a_{N_\mathrm{s}+1})^{-1}\log\left(2\pi a_{N_\mathrm{s}+1}^2\right).$

- Under Assumptions (A1)-(A6), for $l = 1, \ldots, d$ and any $\alpha \in (0, 1)$,

$$\lim_{n\to\infty} P\left\{m_l(t) \in \hat{m}_l(t) \pm \sigma_{n,ll}(t)\, Q_{N_\mathrm{s}+1}(\alpha), \forall t \in [0,1]\right\} = 1-\alpha,$$

- The asymptotic $100(1-\alpha)\,\%$ SCCs for $m_l(t), l = 1, \ldots, d$ are

$$\hat{m}_l(t) \pm \sigma_{n,ll}(t) Q_{N_s+1}(\alpha), \quad t \in [0,1]$$

Implementation details

- $N_{\mathrm{s}} = \left[cN_{\mathrm{T}}^{1/3} (\log n) \right]$, constant spline estimators $\hat{m}_l(t), l = 1, \ldots, d$
- Estimate density $f(t)$ by histogram estimator $\hat{f}(t)$
- Spline estimators $\hat{\sigma}_Y^2(t, \mathbf{x})$ and $\hat{G}_l(t,t)$
- $\hat{\mathbf{\Gamma}}_n(t) \equiv \left[n^{-1} \sum_{i=1}^n X_{il} X_{il'} \hat{\sigma}_Y^2(t, \mathbf{X}_i) \left\{ \hat{f}(t) h_s N_T \right\}^{-1} \right.$ $\left. \times \left\{ 1 + \left(\frac{\sum_{i=1}^n N_i^2}{N_T} - 1 \right) \frac{\sum_{l=1}^d \hat{G}_l(t,t) X_{il}^2}{\hat{\sigma}_Y^2(t, \mathbf{X}_i)} \hat{f}(t) h_s \right\} \right]_{l,l'=1}^d$
- $\hat{\mathbf{H}} = \left\{ n^{-1} \sum_{i=1}^n X_{il} X_{il'} \right\}_{l,l'=1}^d$,
- $\hat{\mathbf{\Sigma}}_n(t) = \left\{ \hat{\sigma}_{n,ll'}^2(t) \right\}_{l,l'=1}^d = \hat{\mathbf{H}}^{-1} \hat{\mathbf{\Gamma}}_n(t) \hat{\mathbf{H}}^{-1}$;

Implementation details

- $N_{\mathrm{s}} = \left[c N_{\mathrm{T}}^{1/3} (\log n) \right]$, constant spline estimators $\hat{m}_l(t), l = 1, \ldots, d$
- Estimate density $f(t)$ by histogram estimator $\hat{f}(t)$
- Spline estimators $\hat{\sigma}_Y^2(t, \mathbf{x})$ and $\hat{G}_l(t, t)$
- $\hat{\mathbf{\Gamma}}_n(t) \equiv \left[n^{-1} \sum_{i=1}^n X_{il} X_{il'} \hat{\sigma}_Y^2(t, \mathbf{X}_i) \left\{ \hat{f}(t) h_s N_T \right\}^{-1} \right.$

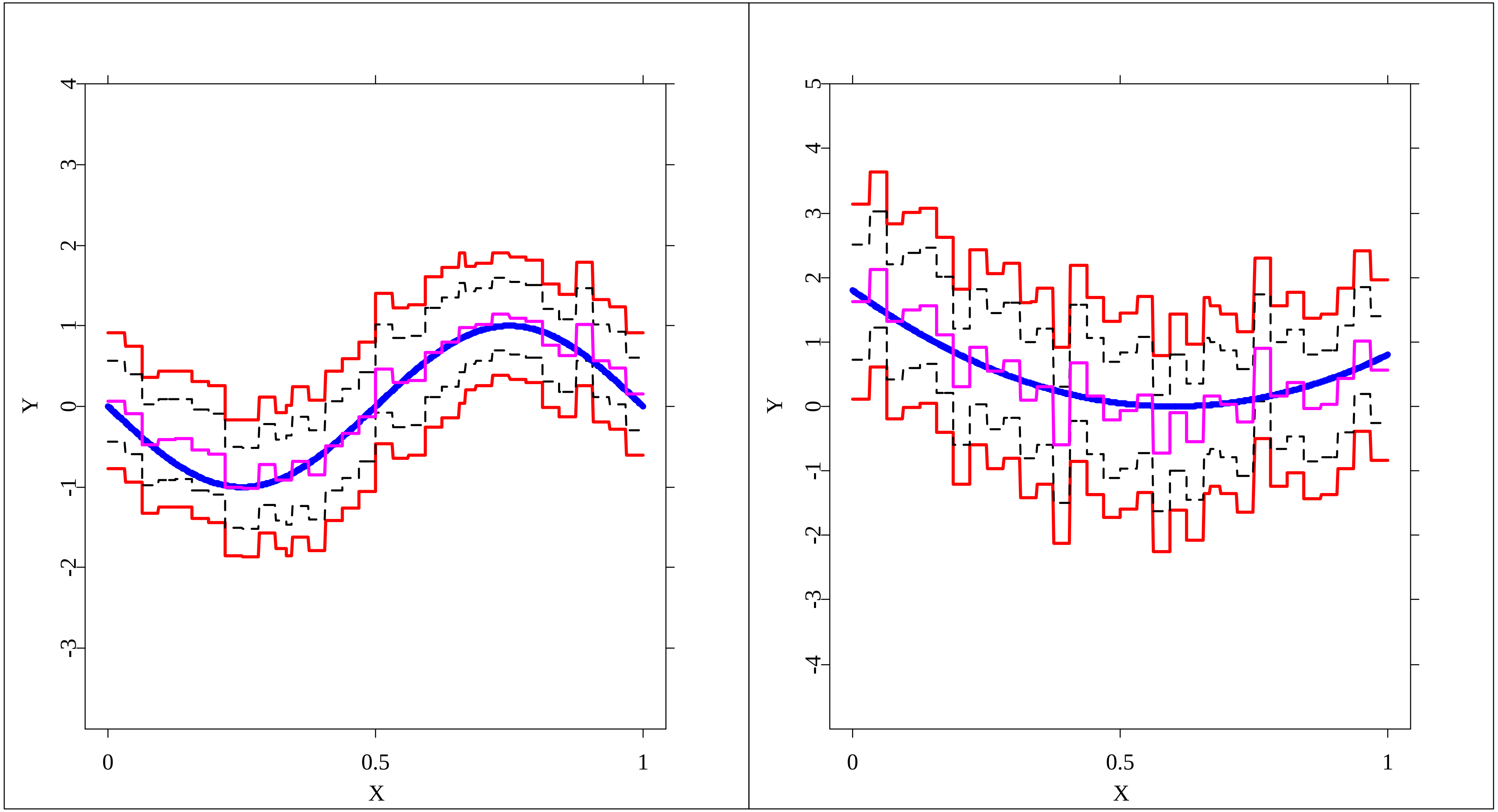
  $\left. \times \left\{ 1 + \left(\frac{\sum_{i=1}^n N_i^2}{N_T} - 1 \right) \frac{\sum_{l=1}^d \hat{G}_l(t,t) X_{il}^2}{\hat{\sigma}_Y^2(t, \mathbf{X}_i)} \hat{f}(t) h_s \right\} \right]_{l,l'=1}^d$
- $\hat{\mathbf{H}} = \left\{ n^{-1} \sum_{i=1}^n X_{il} X_{il'} \right\}_{l,l'=1}^d$,
- $\hat{\mathbf{\Sigma}}_n(t) = \left\{ \hat{\sigma}_{n,ll'}^2(t) \right\}_{l,l'=1}^d = \hat{\mathbf{H}}^{-1} \hat{\mathbf{\Gamma}}_n(t) \hat{\mathbf{H}}^{-1}$;
- SCCs for $m_l(t), l = 1, \ldots, d$ are $\hat{m}_l(t) \pm \hat{\sigma}_{n,ll}(t) Q_{N_s+1}(\alpha), l = 1, \ldots, d$

Simulation

- $Y_{ij} = \left\{ m_1 (T_{ij}) + \sum_{k=1}^{2} \xi_{ik,1} \phi_{k,1} (T_{ij}) \right\} X_{i1} + \left\{ m_2 (T_{ij}) + \sum_{k=1}^{3} \xi_{ik,2} \phi_{k,2} (T_{ij}) \right\} X_{i2} + \sigma (T_{ij}) \varepsilon_{ij}, 1 \le i \le n, 1 \le j \le N_i$
- $T \sim U[0,1]$, $X_1 \sim N(0,1)$, $X_2 \sim$ Binomial$[1, 0.5]$, $\xi_{k,1} \sim N(0,1)$, $k = 1, 2$, $\xi_{k,2} \sim N(0,1)$, $k = 1, 2, 3$, $\varepsilon \sim N(0,1)$, N_i having a discrete uniform distribution from $2, \cdots, 14$, for $1 \le i \le n$.
- $m_1(t) = \sin \{2\pi (t - 1/2)\}$, $\phi_{1,1}(t) = -2 \cos \{\pi (t - 1/2)\} / \sqrt{5}$, $\phi_{2,1}(t) = \sin \{\pi (t - 1/2)\} / \sqrt{5}$;
- $m_2(t) = 5 (t - 0.6)^2$, $\phi_{1,2}(t) = 1$, $\phi_{2,2}(t) = \sqrt{2} \sin (2\pi t)$, $\phi_{3,2}(t) = \sqrt{2} \cos (2\pi t)$;
- $N_s = \left[c N_T^{1/3} (\log n) \right]$, $c = 0.3, 0.5, 0.8, 1$. The noise level $\sigma = 0.5, 1$.

Table 1: Coverage frequencies of the 95% and 99% SCCs for functions m_1 (left) and m_2 (right), based on 500 replications.

σ	n	$1-\alpha$	$c=0.3$	$c=0.5$	$c=0.8$	$c=1$
1.0	200	0.950	0.950, 0.952	0.944, 0.948	0.920, 0.904	0.886, 0.884
		0.990	0.990, 0.998	0.990, 0.990	0.976, 0.984	0.968, 0.974
	400	0.950	0.944, 0.948	0.950, 0.930	0.922, 0.912	0.908, 0.904
		0.990	0.996, 0.984	0.990, 0.988	0.984, 0.988	0.974, 0.966
	600	0.950	0.934, 0.962	0.954, 0.946	0.930, 0.952	0.930, 0.924
		0.990	0.992, 0.996	0.992, 0.986	0.988, 0.990	0.984, 0.990
	800	0.950	0.936, 0.934	0.960, 0.966	0.950, 0.964	0.956, 0.934
		0.990	0.998, 0.996	0.994, 0.994	0.986, 0.992	0.988, 0.988
0.5	200	0.950	0.936, 0.948	0.952, 0.942	0.916, 0.900	0.912, 0.890
		0.990	0.988, 0.994	0.992, 0.990	0.972, 0.974	0.972, 0.972
	400	0.950	0.916, 0.930	0.936, 0.932	0.928, 0.916	0.904, 0.898
		0.990	0.994, 0.984	0.992, 0.988	0.996, 0.988	0.978, 0.976
	600	0.950	0.924, 0.948	0.952, 0.954	0.926, 0.958	0.936, 0.938
		0.990	0.996, 0.994	0.994, 0.986	0.984, 0.990	0.990, 0.994
	800	0.950	0.942, 0.900	0.950, 0.960	0.942, 0.962	0.960, 0.938
		0.990	0.996, 0.998	0.996, 0.994	0.990, 0.996	0.992, 0.988



Plots of 95% confidence corridors/intervals for m_1(left), m_2(right) at $\sigma = 0.5$, $n = 200$.

Real data analysis

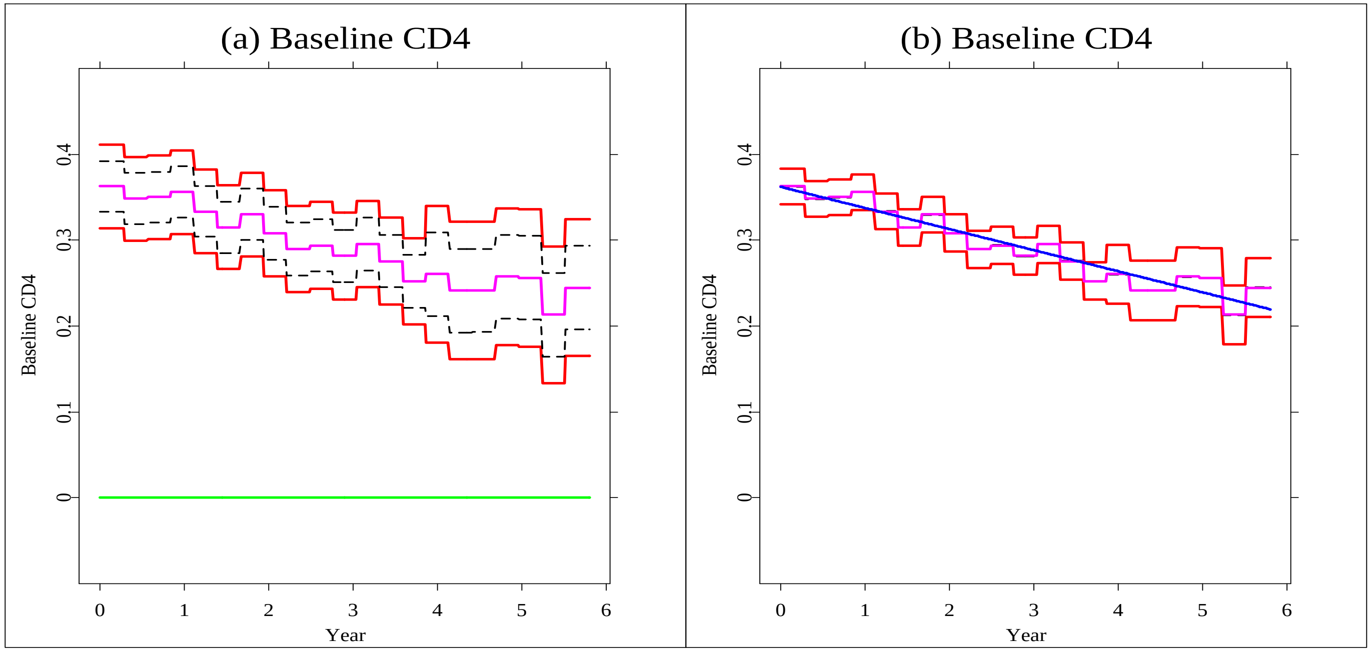
1817 observations of CD4 cell percentages on 283 patients

- Model:

$$Y_{ij} = \sum_{l=0}^{3} m_l\left(T_{ij}\right) X_{il} + \sum_{l=0}^{3} \sum_{k=1}^{\infty} \xi_{ik,l} \phi_{k,l}\left(T_{ij}\right) X_{il} + \sigma\left(T_{ij}\right) \varepsilon_{ij}$$

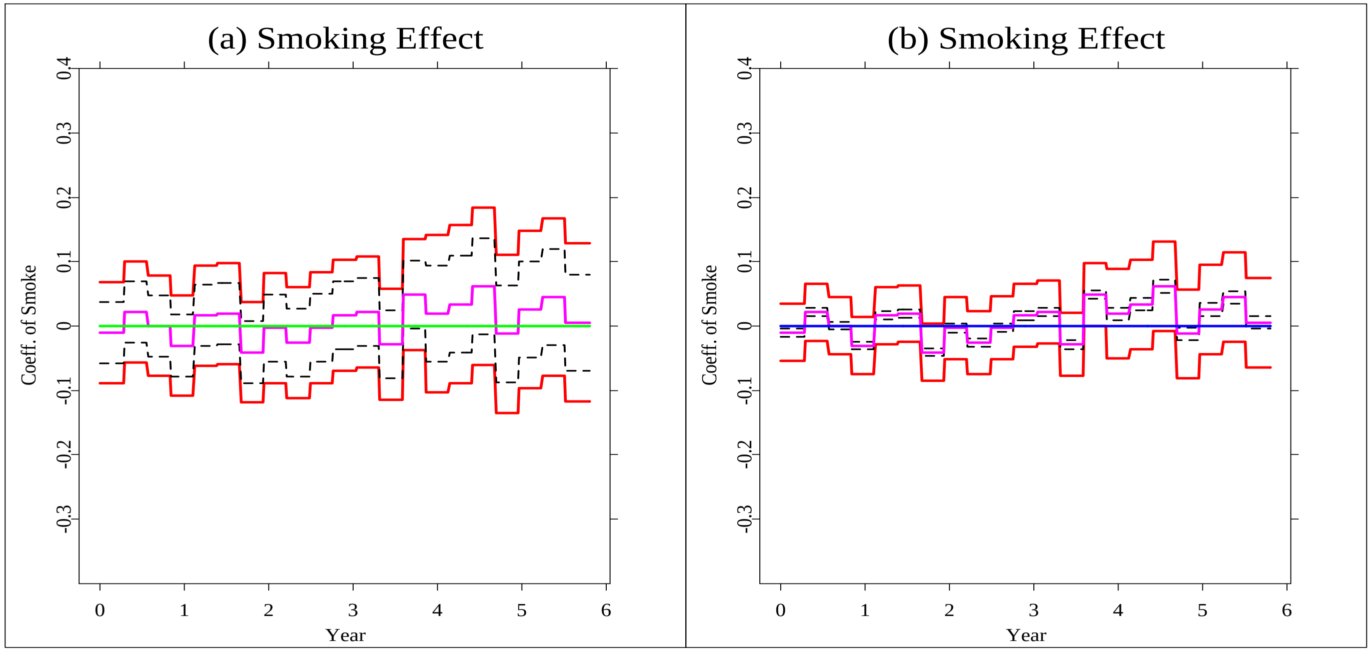
- $m_0(t)$: the coefficient function for baseline CD4 percentage;
- $m_1(t)$: the coefficient function for smoking status;
- $m_2(t)$: the coefficient function for centered pre-infection CD4 percentage;
- $m_3(t)$: the coefficient function for centered age.

- $H_{00} : m_0(t) \equiv a + bt,\ \exists a, b \in \mathbb{R}$ v.s. $H_{10} : m_0(t) \neq a + bt$, $\forall a, b \in \mathbb{R}$;



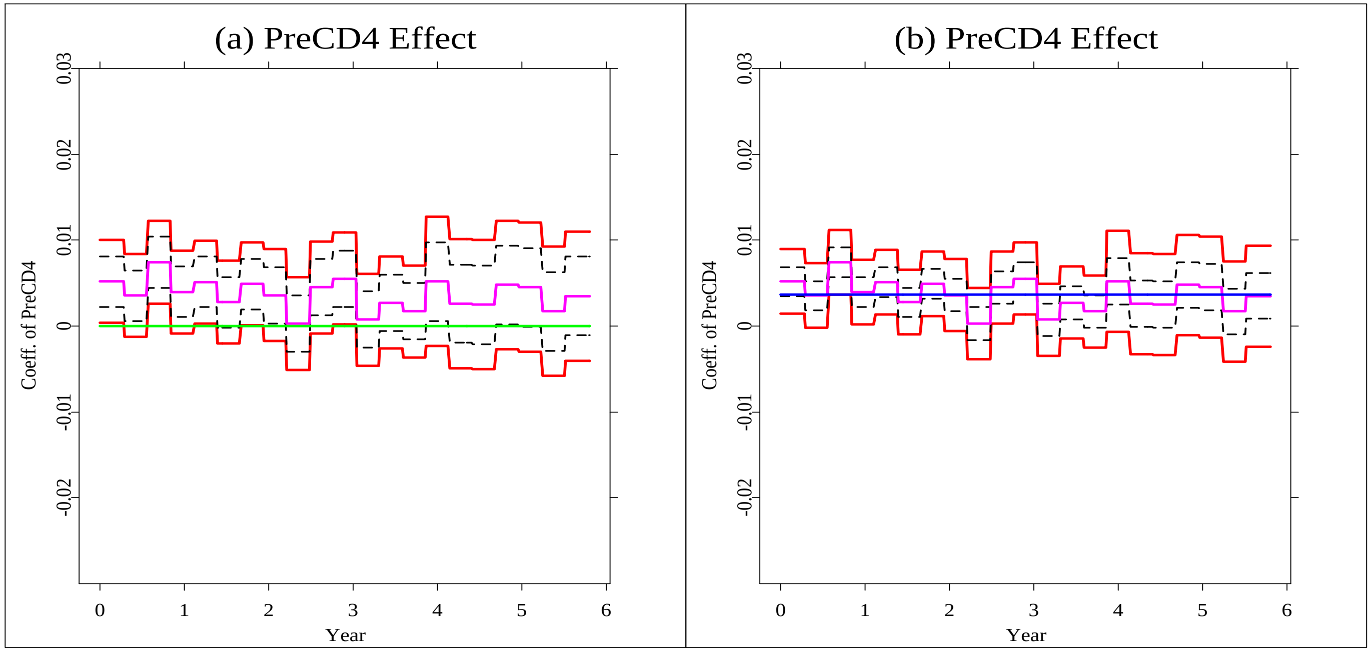
  - (a) $\hat{m}_0$ (middle solid), 95% SCC (solid) and pointwise CIs (dashed);
  - (b) the same except with confidence level $1 - \hat{\alpha}_0$ and the estimated m_0 under H_{00} (solid linear), $\hat{\alpha}_0 = 0.99072$

- $H_{01} : m_1(t) \equiv 0$ v.s. $H_{11} : m_1(t) \neq 0$, for some $t \in [0, 6]$;



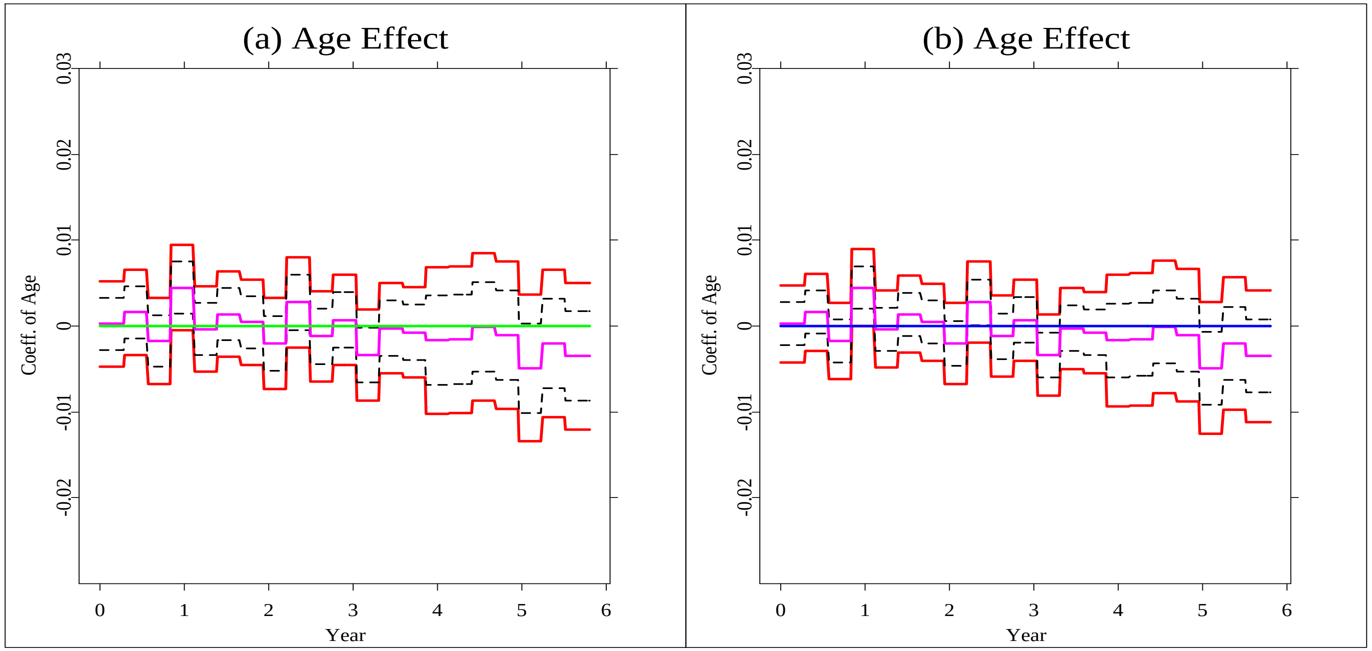
  – (a) $\hat{m}_1$ (middle solid), 95% SCC (solid) and pointwise CIs (dashed);
  – (b) the same except with confidence level $1 - \hat{\alpha}_1$ and the estimated m_1 under H_{01} (solid linear), $\hat{\alpha}_1 = 0.79723$

- $H_{02} : m_2(t) \equiv c$, for some $c > 0$ v.s. $H_{12} : m_2(t) \neq c$, for any $c > 0$;



  - (a) $\hat{m}_2$ (middle solid), 95% SCC (solid) and pointwise CIs (dashed);
  - (b) the same except with confidence level $1 - \hat{\alpha}_2$ and the estimated m_2 under H_{02} (solid linear), $\hat{\alpha}_2 = 0.25404$

- $H_{03} : m_3(t) \equiv 0$ v.s. $H_{13} : m_3(t) \neq 0$, for some $t \in [0, 6]$.



  - (a) $\hat{m}_3$ (middle solid), 95% SCC (solid) and pointwise CIs (dashed);
  - (b) the same except with confidence level $1 - \hat{\alpha}_3$ and the estimated m_3 under H_{03} (solid linear), $\hat{\alpha}_3 = 0.10775$

Conclusions

- A varying coefficient regression model for sparse functional data:

$$Y_{ij} = \sum_{l=1}^{d} m_l\left(T_{ij}\right) X_{il} + \sum_{l=1}^{d} \sum_{k=1}^{\infty} \xi_{ik,l} \phi_{k,l}\left(T_{ij}\right) X_{il} + \sigma\left(T_{ij}\right) \varepsilon_{ij}$$

- Based on spline smoothing, SCCs for $m_l\left(t\right), l = 1, \ldots, d$ are

$$\hat{m}_l(t) \pm \sigma_{n,ll}(t) Q_{N_s+1}\left(\alpha\right), l = 1, \ldots, d$$

- CD4/HIV study as an example is used to illustrate how inference is made using SCCs.

Thank you for your attention!

References

[1] Bosq, D (1998) Nonparametric statistics for stochastic processes. Springer-Verlag, New York

[2] Brumback B, Rice JA (1998) Smoothing spline models for the analysis of nested and crossed samples of curves (with Discussion). J Am Stat Assoc 93:961–994

[3] Cao G, Yang L, Todem D (2012) Simultaneous inference for the mean function based on dense functional data. J Nonparametr Stat 24:359–377

[4] Cao G, Wang J, Wang L, Todem D (2012) Spline confidence bands for functional derivatives. J Stat Plan Inference 142:1557–1570

[5] Chiang CT, Rice JA, Wu CO (2001) Smoothing spline estimation for varying coefficient models with repeatedly measured dependent variables. J Am Stat Assoc 96:605–619

[6] Claeskens G, Van Keilegom I (2003) Bootstrap confidence bands for regression curves and their derivatives. Ann Stat 31:1852–1884

[7] de Boor C (2001) A practical guide to splines. Springer-Verlag, New York

[8] Fan J, Zhang JT (2000a) Two-step estimation of functional linear models with application to longitudinal data. J R Stat Soc Ser B 62:303–322

[9] Fan J, Zhang WY (2000b) Simultaneous confidence bands and hypothesis testing in varying-coefficient models. Scand J Stat 27:715–731

[10] Fan J, Zhang WY (2008) Statistical methods with varying coefficient models. Stat Interface 1:179–195

[11] Ferraty F, Vieu P (2006) Nonparametric functional data analysis: theory and practice. Springer-Verlag, New York

[12] Gabrys R, Horváth L, Kokoszka, P (2010) Tests for error correlation in the functional linear model. J Am Stat Assoc 105:1113–1125

[13] Hall P, Müller, HG, Wang, JL (2006) Properties of principal component methods for functional and longitudinal data analysis. Ann Stat 34:1493–1517

[14] Hall P, Titterington DM (1988) On confidence bands in nonparametric density estimation and regression. J Mult Anal 27:228–254

[15] Härdle W, Luckhaus S (1984) Uniform consistency of a class of regression function estimators. Ann Stat 12:612–623

[16] Hastie T, Tibshirani R (1993) Varying-coefficient models. J R Stat Soc Ser B 55:757–796

[17] Hoover DR, Rice JA, Wu CO, Yang LP (1998) Nonparametric smoothing estimates of time-varying coefficient models with longitudinal data. Biometrika 85:809–822

[18] Horváth L, Kokoszka P (2012) Inference for functional data with applications. Springer-Verlag, New York

[19] Huang JZ, Wu CO, Zhou L (2002) Varying-coefficient models and basis function approximations for the analysis of repeated measurements. Biometrika 89:111–128

[20] Huang JZ, Wu CO, Zhou L (2004) Polynomial spline estimation and inference for varying coefficient models with longitudinal data. Stat Sin 14:763–788

[21] James GM, Hastie T, Sugar C (2000) Principal component models for sparse functional data. Biometrika 87:587–602

[22] James GM, Sugar CA (2003) Clustering for sparsely sampled functional data. J Am Stat Assoc 98:397–408

[23] Leadbetter MR, Lindgren G, Rootzén H (1983) Extremes and related properties of random sequences and processes. Springer-Verlag, New York

[24] Liu R, Yang L (2010) Spline-backfitted kernel smoothing of additive coefficient model. Econ Theory 26:29–59

[25] Ma S, Yang L, Carroll RJ (2012) A simultaneous confidence band for sparse longitudinal regression. Stat Sin 22:95–122

[26] Manteiga W, Vieu P (2007) Statistics for functional data. Comput Stat Data Anal 51:4788–4792

[27] Ramsay JO, Silverman BW (2005) Functional data analysis. Springer, New York

[28] Wang L, Li H, Huang JZ (2008) Variable selection in nonparametric varying-coefficient models for analysis of repeated measurements. J Am Stat Assoc 103:1556–1569

[29] Wang L, Yang L (2009) Polynomial spline confidence bands for regression curves. Stat Sin 19:325–342

[30] Wu CO, Chiang CT (2000) Kernel smoothing on varying coefficient models with longitudinal dependent variable. Stat Sin 10:433–456

[31] Wu CO, Chiang CT, Hoover DR (1998) Asymptotic confidence regions for kernel smoothing of a varying-coefficient model with longitudinal data. J Am Stat Assoc 93:1388–1402

[32] Wu Y, Fan J, Müller HG (2010) Varying-coefficient functional linear regression. Bernoulli 16:730–758

[33] Xue L, Yang L (2006) Additive coefficient modelling via polynomial spline. Stat Sin 16:1423–1446

[34] Xue L, Zhu L (2007) Empirical likelihood for a varying coefficient model with longitudinal data. J Am Stat Assoc 102:642–654

[35] Yao W, Li R (2013) New local estimation procedure for a non-parametric regression function for longitudinal data. J R Stat Soc Ser B 75:123–138

[36] Yao F, Müller HG, Wang JL (2005a) Functional linear regression analysis for longitudinal data. Ann Stat 33:2873–2903

[37] Yao F, Müller HG, Wang JL (2005b) Functional data analysis for sparse longitudinal data. J Am Stat Assoc 100:577–590

[38] Zhou L, Huang J, Carroll RJ (2008) Joint modelling of paired sparse functional data using principal components. Biometrika 95:601–619

[39] Zhu H, Li R, Kong L (2012) Multivariate varying coefficient model for functional responses. Ann Stat 40:2634–2666