Estimation and Variable Selection for GAPLM

Author:Li Wang, Xiang Liu, Hua Liang and Raymond J. Carroll
Speaker: Jiyanglin Li

School of Statistics and Management
Shanghai University of Finance and Economics

Outline

1. Introduction
2. Methods
   - Estimation Method
   - Variable Selection
3. Numerical Studies
   - Simulation
   - Real Example
4. Conclusion

Models

Consider $\mathbf{X} : p \times 1$, $\mathbf{Z} : q \times 1$,
$\mu(\mathbf{X}) = E(Y|\mathbf{X})$, $\mu(\mathbf{X}, \mathbf{Z})) = E(Y|\mathbf{X}, \mathbf{Z}))$.

- Generalized Linear Model (GLM):

$$g(\mu(\mathbf{X})) = \beta_0 + \mathbf{X}^T \boldsymbol{\beta}, \mu(\mathbf{X}) = E(Y|\mathbf{X}).$$

- Generalized Additive Model (GAM):

$$\mu(\mathbf{X}) = \alpha + f_1(X_1) + \cdots + + f_p(X_p).$$

- Generalized Additive Partial Linear Model (GAPLM):

$$g(\mu(\mathbf{X}, \mathbf{Z})) = \sum_{k=1}^{p} \eta_k(X_k) + \mathbf{Z}^T \boldsymbol{\beta}.$$

$\mathbf{X} = (X_1, \cdots, X_{d_1})^T, \mathbf{Z} = (Z_1, \cdots, Z_{d_2})^T$: covariates,
Y: response variable, $\mathcal{B}, \mathcal{C}$: known functions, ξ natural parameters.

- The conditional density of Y given $(\mathbf{X}, \mathbf{Z}) = (\mathbf{x}, \mathbf{z})$ has the form

$$f_{Y|\mathbf{X},\mathbf{Z}}(y|\mathbf{x}, \mathbf{z}) = \exp\{y\xi(\mathbf{x}, \mathbf{z}) - \mathcal{B}(\xi(\mathbf{x}, \mathbf{z})) + \mathcal{C}(y)\}.$$

- The conditional mean and the conditional variance are

$$\mu(\mathbf{x}, \mathbf{z}) = E(Y|\mathbf{X} = \mathbf{x}, \mathbf{Z} = \mathbf{z}) = \mathcal{B}'(\xi(\mathbf{x}, \mathbf{z}))$$
$$V(\mathbf{x}, \mathbf{z}) = Var(Y|\mathbf{X} = \mathbf{x}, \mathbf{Z} = \mathbf{z}) = \mathcal{B}''(\xi(\mathbf{x}, \mathbf{z})).$$

Generalized Additive Partial Linear Model (GAPLM)

Define the Generalized Additive Partial Linear Model (GAPLM):

$$g(\mu(\mathbf{X}, \mathbf{Z})) = \sum_{k=1}^{d_1} \eta_k(X_k) + \mathbf{Z}^T \boldsymbol{\beta}.$$

- $\boldsymbol{\beta}$: d_1-dimensional parameter.
- $\{\eta_k\}_{k=1}^{d_1}$: unknown smooth functions which satisfy $E(\eta(X_k)) = 0$.

Main jobs

- Use polynomial spline for estimation of nonparametric functions.
- Derive quasi-likelihood based estimators for the linear parameters.
- Describe the rates of convergence of the nonparametric parts.
- Establish asymptotic normality for the estimators of the parametric components.
- Develop variable selection procedures for the parametric components.

1 Introduction

2 Methods
- Estimation Method
- Variable Selection

3 Numerical Studies

4 Conclusion

Quasi-likelihood function

If $Var(Y|\mathbf{X}=\mathbf{x},\mathbf{Z}=\mathbf{z})=\sigma^2 V(\mu(\mathbf{x},\mathbf{z}))$, replace the conditional loglikelihood function $\log(f_{Y|\mathbf{X},\mathbf{Z}}(y|\mathbf{x},\mathbf{z}))$ by a quasi-likelihood function $Q(m,y)$ satisfying

$$\frac{\partial}{\partial m}Q(m,y)=\frac{y-m}{V(m)}.$$

Note that

$$\begin{aligned}\frac{\partial}{\partial \xi}\log(f_{Y|\mathbf{X},\mathbf{Z}}(y|\mathbf{x},\mathbf{z})) &= y-\mathcal{B}'(\xi(\mathbf{x},\mathbf{z})) \\ &= y-\mu(\mathbf{x},\mathbf{z}).\end{aligned}$$

Maximum quasi-likelihood

- Given Sample:$(Y_i, \mathbf{X}_i, \mathbf{Z}_i), i = 1, \cdots, n$.
- To avoid confusion, we use some notation
  - True additive function:$\eta_0 = \sum_{k=1}^{d_1} \eta_{0k}(x_k)$.
  - True parameter values: $\boldsymbol{\beta}_0$.
- Let $\mathcal{G}_n$ be the collection of functions with the additive form $\eta(\mathbf{x}) = \sum_{k=1}^{d_1} \eta_k(x_k)$.

Goal: seek a function $\eta \in \mathcal{G}_n$ and a value of β that maximize the quasi-likelihood function

$$L(\eta, \boldsymbol{\beta}) = \frac{1}{n} \sum_{i=1}^{n} Q(g^{-1}(\eta(\mathbf{X}_i) + \mathbf{Z}_i^T \boldsymbol{\beta}), Y_i). \tag{1}$$

Maximum quasi-likelihood

Using B-spline basis functions to approximate η, i.e.,

$$\eta(\mathbf{x}) \approx \boldsymbol{\gamma}^T \mathbf{b}(\mathbf{x}),$$

thus maximization problem in (1) is equivalent to finding $\boldsymbol{\gamma}$ and $\boldsymbol{\beta}$ to maximize

$$\ell(\boldsymbol{\gamma}, \boldsymbol{\beta}) = \frac{1}{n} \sum_{i=1}^{n} Q(g^{-1}(\boldsymbol{\gamma}^T \mathbf{b}(\mathbf{X}_i) + \mathbf{Z}_i^T \boldsymbol{\beta}), Y_i). \tag{2}$$

Maximum quasi-likelihood

Denote the vector of normalized B-spline basis functions by $\mathbf{B}(\mathbf{x})$, the maximization problem in (2) is equivalent to finding $(\boldsymbol{\gamma}, \boldsymbol{\beta})$ to maximize

$$\ell(\boldsymbol{\gamma}, \boldsymbol{\beta}) = \frac{1}{n} \sum_{i=1}^{n} Q(g^{-1}(\mathbf{B_i}^T \boldsymbol{\gamma} + \mathbf{Z}_i^T \boldsymbol{\beta}), Y_i). \tag{3}$$

The spline estimator of η_0 is

$$\hat{\eta}(\mathbf{x}) = \hat{\boldsymbol{\gamma}}^T \mathbf{B}(\mathbf{x}).$$

Notations

ν: positive integer, $\alpha \in (0,1], p = \alpha + \nu > 2$. Define

$$\mathcal{H}(p) = \{g(\cdot) : \forall 0 \le m^*, m \le 1, |g^{(\nu)}(m^*) - g^{(\nu)}(m)| \le C|m^* - m|^\alpha\}.$$

$$\rho_\ell(m) = \left(\frac{dg^{-1}(m)}{dm}\right)^\ell \cdot \frac{1}{V(g^{-1}(m))},$$

$$q_\ell(m,y) = \frac{\partial^\ell}{\partial m^\ell} Q(g^{-1}(m), y).$$

It's easy to verify that

$$\begin{aligned} q_1(m,y) &= (y - g^{-1}(m))\rho_1(m), \\ q_2(m,y) &= (y - g^{-1}(m))\rho_1'(m) - \rho_2(m). \end{aligned}$$

Write $\mathbf{T} = (\mathbf{X}, \mathbf{Z}), \mathbf{A}^{\otimes 2} = \mathbf{A}\mathbf{A}^T$.

Notations

Define

$$m_0(\mathbf{T}) = \eta_0(\mathbf{X}) + \mathbf{Z}^T \boldsymbol{\beta}_0,$$

$$\Gamma(\mathbf{x}) = \frac{E(\mathbf{Z}\rho_2(m_0(\mathbf{T}))|\mathbf{X} = \mathbf{x})}{E(\rho_2(m_0(\mathbf{T}))|\mathbf{X} = \mathbf{x})}, \tilde{\mathbf{Z}} = \mathbf{Z} - \Gamma^{\mathsf{add}}(\mathbf{X}),$$

where

$$\Gamma^{\mathsf{add}}(\mathbf{x}) = \sum_{k=1}^{d_1} \Gamma_k(x_k)$$

is the projection of Γ onto the Hilbert space of theoretically centered additive functions with a norm

$$\|f\|_{\rho_2, m_0} = E(f(\mathbf{X})^2 \rho_2(m_0(\mathbf{T}))).$$

Notations

For measurable functions $\varphi_1, \varphi_2, \varphi$ on $[0,1]^d$, define

$$\langle \varphi_1, \varphi_2 \rangle_n = \frac{1}{n}\sum_{i=1}^{n}\Big(\varphi_1(\mathbf{X}_i)\varphi_2(\mathbf{X}_i)\Big), \quad \|\varphi\|_n^2 = \frac{1}{n}\sum_{i=1}^{n}\varphi^2(\mathbf{X}_i).$$

$$\langle \varphi_1, \varphi_2 \rangle = E\Big(\varphi_1(\mathbf{X})\varphi_2(\mathbf{X})\Big), \quad \|\varphi\|_2^2 = E\varphi^2(\mathbf{X}).$$

For measurable functions ψ on $[0,1]$, define

$$\|\psi\|_{nk}^2 = \frac{1}{n}\sum_{i=1}^{n}\psi^2(X_{ik}), \quad \|\psi\|_{2k}^2 = E\psi^2(X_k).$$

Assumptions

1. $\eta_0''(\cdot)$ is continuous and $\eta_{0k}(\cdot) \in \mathcal{H}(p), k = 1, \cdots, d_1$.
2. $q_2(m, y) < 0$ and $c_q < |q_2^\nu(m, y)| < C_q(\nu = 0, 1)$.
3. The distribution of $\mathbf{X}$ is absolutely continuous, and its density $0 < f(\mathbf{x}) < +\infty, \forall \mathbf{x} \in [0, 1]^{d_1}$.
4. $\forall \boldsymbol{\omega} \in \mathbb{R}^{d_2}, \|\boldsymbol{\omega}\| = 1$,
$$c \leq \boldsymbol{\omega}^T E(\mathbf{Z}^{\otimes 2} | \mathbf{X} = \mathbf{x}) \boldsymbol{\omega} \leq C.$$
5. The number of knots $n^{\frac{1}{2p}} \ll N_n \ll n^{\frac{1}{4}}$.

Assumptions

6. $\Gamma_k(\cdot) \in \mathcal{H}(p), k = 1, \cdots, d_1.$
7. $|\rho_\ell(m_0)| \leq C_\rho$, and $\forall |m - m_0| \leq C_m$,

$$|\rho_\ell(m) - \rho_\ell(m_0)| \leq C_\rho^* |m - m_0|, \ell = 1, 2.$$

8. $\exists C_0, \mathrm{s.\,t.}\, E((Y - g^{-1}(m_0(\mathbf{T})))^2 | \mathbf{T}) \leq C_0, a.s.$

Asymptotic properties

Theorem 1

Under conditions 1-5, for $k = 1, \cdots, d_1$,

$\|\hat{\eta} - \eta_0\|_2 = O_p(N_n^{\frac{1}{2}-p} + (\frac{N_n}{n})^{\frac{1}{2}})$;

$\|\hat{\eta} - \eta_0\|_n = O_p(N_n^{\frac{1}{2}-p} + (\frac{N_n}{n})^{\frac{1}{2}})$;

$\|\hat{\eta}_k - \eta_{0k}\|_{2k} = O_p(N_n^{\frac{1}{2}-p} + (\frac{N_n}{n})^{\frac{1}{2}})$;

$\|\hat{\eta}_k - \eta_{0k}\|_{nk} = O_p(N_n^{\frac{1}{2}-p} + (\frac{N_n}{n})^{\frac{1}{2}})$.

Theorem 2

Under conditions 1-8, $\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \to \mathcal{N}(0, \boldsymbol{\Omega}^{-1})$, where $\boldsymbol{\Omega} = E(\rho_2(m_0(\mathbf{T}))\tilde{\mathbf{Z}}^{\otimes 2})$.

1 Introduction

2 Methods
- Estimation Method
- Variable Selection

3 Numerical Studies

4 Conclusion

Penalized quasi-likelihood

The penalized quasi-likelihood is defined as

$$\mathcal{L}(\eta, \boldsymbol{\beta}) = \frac{1}{n}\sum_{i=1}^{n} Q(g^{-1}(\eta(\mathbf{X}_i) + \mathbf{Z}_i^T\boldsymbol{\beta}), Y_i) - n\sum_{j=1}^{d_2} p_{\lambda_j}(|\beta_j|),$$

Substitute η by its estimate to obtain

$$\mathcal{L}_P(\boldsymbol{\beta}) = \frac{1}{n}\sum_{i=1}^{n} Q(g^{-1}(\mathbf{B}_i^T\hat{\boldsymbol{\gamma}} + \mathbf{Z}_i^T\boldsymbol{\beta}), Y_i) - n\sum_{j=1}^{d_2} p_{\lambda_j}(|\beta_j|),$$

$p_{\lambda_j}(|\beta_j|)$: penalty function, i.e.,

$p_{\lambda_j}(\lvert\beta_j\rvert) = 0.5\lambda_j^2 I(\beta_j \neq 0)$	Subset selection
$p_{\lambda_j}(\lvert\beta_j\rvert) = \lambda_j\lvert\beta_j\rvert$	LASSO
$p_{\lambda_j}(\lvert\beta_j\rvert) = \lambda_j\lvert\beta_j\rvert^2$	Ridge

SCAD

Fan and Li (2001) proposed the Smoothly Clipped Absolute Deviation (SCAD) penalty, defined by

$$p'_{\lambda_j}(\beta) = \lambda_j \left(I(\beta \leq \lambda_j) + \frac{(a\lambda_j - \beta)_+}{(a-1)\lambda_j} I(\beta > \lambda_j) \right)$$

where $a > 2, \beta > 0$.

SCAD results in an estimator with oracle properties:

- Unbiasdness
- Sparsity
- Continuity

Maximum penalized likelihood estimator

Define the maximum penalized likelihood estimator

$$\hat{\boldsymbol{\beta}}^{\mathrm{MPL}} = \arg\max_{\boldsymbol{\beta}} \mathcal{L}_P(\boldsymbol{\beta})$$

$\hat{\boldsymbol{\beta}}^{\mathrm{MPL}}$ performs asymptotically as well as an oracle estimator.

Notations

Let $\boldsymbol{\beta}_0 = (\beta_{10}, \cdots, \beta_{d_2 0}) = (\boldsymbol{\beta}_{10}^T, \boldsymbol{\beta}_{20}^T)^T$.

- $\boldsymbol{\beta}_{10}$: nonzero components.
- $\boldsymbol{\beta}_{20} = \mathbf{0}$.

Write $\mathbf{Z} = (\mathbf{Z}_1^T, \mathbf{Z}_2^T)^T$ and denote

$$a_n = \max_{1 \le j \le d_2} \{|p'_{\lambda_j}(|\beta_{j0}|)| : \beta_{j0} \neq 0\},$$
$$w_n = \max_{1 \le j \le d_2} \{|p''_{\lambda_j}(|\beta_{j0}|)| : \beta_{j0} \neq 0\},$$

Notations

s: the number of nonzero components of $\boldsymbol{\beta}_0$,

$$\boldsymbol{\xi}_n = (p'_{\lambda_1}(|\beta_{10}|)\operatorname{sgn}(\beta_{10}), \cdots, p'_{\lambda_s}(|\beta_{s0}|)\operatorname{sgn}(\beta_{s0}))^T,$$
$$\boldsymbol{\Sigma}_\lambda = \operatorname{diag}(p''_{\lambda_1}(|\beta_{10}|), \cdots, p''_{\lambda_s}(|\beta_{s0}|))^T,$$

Define $\mathbf{T_1} = (\mathbf{X}, \mathbf{Z_1})^T$, $m_0(\mathbf{T}_1) = \eta_0(\mathbf{X}) + \mathbf{Z_1}^T\boldsymbol{\beta}_{10}$,

$$\Gamma_1(\mathbf{x}) = \frac{E(\mathbf{Z}_1\rho_2(m_0(\mathbf{T}_1))|\mathbf{X}=\mathbf{x})}{E(\rho_2(m_0(\mathbf{T}_1))|\mathbf{X}=\mathbf{x})}, \tilde{\mathbf{Z}}_1 = \mathbf{Z}_1 - \Gamma_1^{\text{add}}(\mathbf{X}),$$

where Γ_1^{add} is the projection of Γ_1 onto the Hilbert space of theoretically centered additive functions with the norm $\|f\|_{\rho_2,m_0}$.

Oracle properties

Theorem 3

Under conditions 1-8, and if $a_n \to 0$ and $w_n \to 0$ as $n \to \infty$, then there exists a local maximizer $\hat{\boldsymbol{\beta}}^{\mathrm{MPL}}$ of $\mathcal{L}_P(\boldsymbol{\beta})$ such that its rate of convergence is $O_p(n^{-\frac{1}{2}} + a_n)$.

Theorem 4

Suppose that conditions 1-8 hold and that $\liminf_{n\to\infty} \liminf_{\beta_j \to 0^+} \lambda_{jn}^{-1} p'_{\lambda_{jn}}(|\beta_j|) > 0$. If $\sqrt{n}\lambda_{jn} \to \infty$ as $n \to \infty$, then the $\sqrt{n}$-consistent estimator $\hat{\boldsymbol{\beta}}_1^{\mathrm{MPL}}$ in Theorem 3 satisfies $\hat{\boldsymbol{\beta}}_2^{\mathrm{MPL}} = \mathbf{0}$, and

$$\sqrt{n}(\boldsymbol{\Omega}_s + \boldsymbol{\Sigma}_\lambda)(\hat{\boldsymbol{\beta}}^{\mathrm{MPL}} - \boldsymbol{\beta}_{10} + (\boldsymbol{\Omega}_s + \boldsymbol{\Sigma}_\lambda)^{-1}\boldsymbol{\xi}_n) \xrightarrow{\mathcal{D}} \mathcal{N}(\mathbf{0}, \boldsymbol{\Omega}_s)$$

where $\boldsymbol{\Omega}_s = E(\rho_2(m_0(\mathbf{T}_1))\tilde{\mathbf{Z}}_1^{\otimes 2})$.

1 Introduction

2 Methods

3 Numerical Studies
   Simulation
   Real Example

4 Conclusion

100 data sets consisting of $n = 100, 200$ and 400 obs, respectively, from the GAPLM:

$$\text{logit}\Big(Pr(Y = 1)\Big) = \eta_1(X_1) + \eta_2(X_2) + \mathbf{Z}^T\boldsymbol{\beta}, \qquad (4)$$

where

$$\eta_1(x) = sin(4\pi x),$$
$$\eta_1(x) = 10\Big(\exp(-3.25x) - 4\exp(-6.5x) + 3\exp(-9.75x)\Big).$$

True parameters $\boldsymbol{\beta} = (3, 1.5, 0, 0, 0, 0, 2, 0)^T$.

Example 1

Data description

The authors describe that the inputs should be generated as follows:

- $X_1, X_2 \overset{\text{i.i.d}}{\sim} U[0,1]$, $Z_1, Z_2 \sim \mathcal{N}(0.5, 0.09)$, $(Z_1, \cdots, Z_6, X_1, X_2)^T$ has an autoregressive structure with correlation coefficient $\rho = 0.5$.

However, I think that their description is wrong and that the inputs should be generated as follows:

- $X_1, X_2 \overset{\text{i.i.d}}{\sim} U[0,1]$, $Z_1, \cdots, Z_8 \sim \mathcal{N}(0.5, 0.09)$, $(Z_1, \cdots, Z_8)^T$ has an autoregressive structure with correlation coefficient $\rho = 0.5$.

Evidence: Fan and Li (2001).

Example 1

Data description

The authors describe that the inputs should be generated as follows:

- $X_1, X_2 \overset{\text{i.i.d}}{\sim} U[0,1]$, $Z_1, Z_2 \sim \mathcal{N}(0.5, 0.09)$, $(Z_1, \cdots, Z_6, X_1, X_2)^T$ has an autoregressive structure with correlation coefficient $\rho = 0.5$.

However, I think that their description is wrong and that the inputs should be generated as follows:

- $X_1, X_2 \overset{\text{i.i.d}}{\sim} U[0,1]$, $Z_1, \cdots, Z_8 \sim \mathcal{N}(0.5, 0.09)$, $(Z_1, \cdots, Z_8)^T$ has an autoregressive structure with correlation coefficient $\rho = 0.5$.

Evidence: Fan and Li (2001).

Example 1

Data description

The authors describe that the inputs should be generated as follows:

- $X_1, X_2 \overset{\text{i.i.d}}{\sim} U[0,1]$, $Z_1, Z_2 \sim \mathcal{N}(0.5, 0.09)$, $(Z_1, \cdots, Z_6, X_1, X_2)^T$ has an autoregressive structure with correlation coefficient $\rho = 0.5$.

However, I think that their description is wrong and that the inputs should be generated as follows:

- $X_1, X_2 \overset{\text{i.i.d}}{\sim} U[0,1]$, $Z_1, \cdots, Z_8 \sim \mathcal{N}(0.5, 0.09)$, $(Z_1, \cdots, Z_8)^T$ has an autoregressive structure with correlation coefficient $\rho = 0.5$.

Evidence: Fan and Li (2001).

Example 1

Data description

The authors describe that the inputs should be generated as follows:

- $X_1, X_2 \overset{\text{i.i.d}}{\sim} U[0,1]$, $Z_1, Z_2 \sim \mathcal{N}(0.5, 0.09)$, $(Z_1, \cdots, Z_6, X_1, X_2)^T$ has an autoregressive structure with correlation coefficient $\rho = 0.5$.

However, I think that their description is wrong and that the inputs should be generated as follows:

- $X_1, X_2 \overset{\text{i.i.d}}{\sim} U[0,1]$, $Z_1, \cdots, Z_8 \sim \mathcal{N}(0.5, 0.09)$, $(Z_1, \cdots, Z_8)^T$ has an autoregressive structure with correlation coefficient $\rho = 0.5$.

Evidence: Fan and Li (2001).

Table: **C**, **I** and **MRME** stand for the average number of the five zero coefficients correctly set to 0, the average number of the three nonzero coefficients incorrectly set to 0, and the median of the relative model errors.

n	Method	*C*	*I*	MRME
100	ORACLE	5	0	0.27
	SCAD	4.29	0.93	0.60
	Lasso	3.83	0.67	0.51
	BIC	4.53	0.95	0.54
400	ORACLE	5	0	0.33
	SCAD	4.81	0.27	0.49
	Lasso	3.89	0.10	0.67
	BIC	4.90	0.35	0.46

Still generate Y from model (4) but with $\boldsymbol{\beta} = (3, 1.5, 2)$.

- $Z_1, Z_2, X_1, X_2 \sim \mathcal{N}(0.5, 0.09), Z_3 \sim$ Bernoulli(0.5), $(Z_1, Z_2, X_1, X_2)^T$ has an autoregressive structure with correlation coefficient ρ.
- Two scenarios: (i) $\rho = 0.5$, (ii) $\rho = 0.7$.

However, as in example 1, I still think that their description is wrong and that the inputs should be generated as follows:

- $Z_1, Z_2 \sim \mathcal{N}(0.5, 0.09), X_1, X_2 \sim U[0, 1], Z_3 \sim$ Bernoulli(0.5), $(Z_1, Z_2, X_1, X_2)^T$ has an autoregressive structure with correlation coefficient ρ.
- Two scenarios: (i) $\rho = 0.5$, (ii) $\rho = 0.7$.

Example 2

Data description

Still generate Y from model (4) but with $\boldsymbol{\beta} = (3, 1.5, 2)$.

- $Z_1, Z_2, X_1, X_2 \sim \mathcal{N}(0.5, 0.09), Z_3 \sim$ Bernoulli(0.5), $(Z_1, Z_2, X_1, X_2)^T$ has an autoregressive structure with correlation coefficient ρ.
- Two scenarios: (i) $\rho = 0.5$, (ii) $\rho = 0.7$.

However, as in example 1, I still think that their description is wrong and that the inputs should be generated as follows:

- $Z_1, Z_2 \sim \mathcal{N}(0.5, 0.09), X_1, X_2 \sim U[0, 1], Z_3 \sim$ Bernoulli(0.5), $(Z_1, Z_2, X_1, X_2)^T$ has an autoregressive structure with correlation coefficient ρ.
- Two scenarios: (i) $\rho = 0.5$, (ii) $\rho = 0.7$.

Example 2

Data description

Still generate Y from model (4) but with $\boldsymbol{\beta} = (3, 1.5, 2)$.

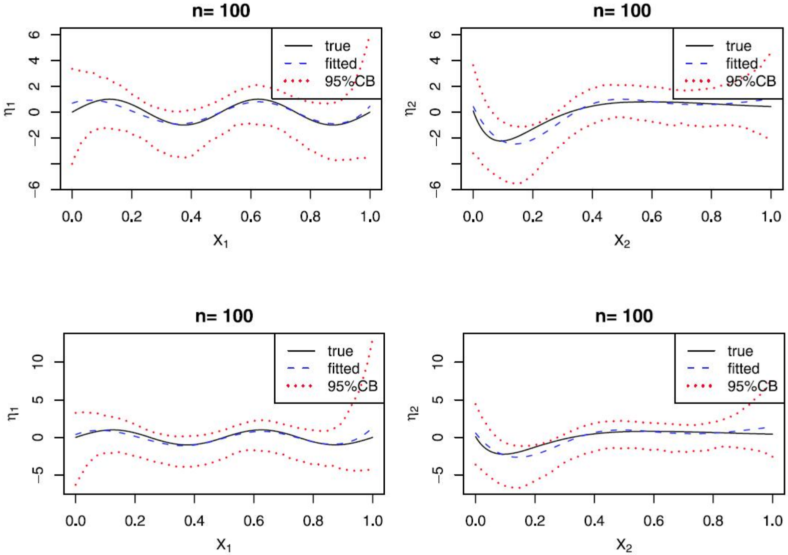
- $Z_1, Z_2, X_1, X_2 \sim \mathcal{N}(0.5, 0.09), Z_3 \sim$ Bernoulli(0.5), $(Z_1, Z_2, X_1, X_2)^T$ has an autoregressive structure with correlation coefficient ρ.
- Two scenarios: (i) $\rho = 0.5$, (ii) $\rho = 0.7$.

However, as in example 1, I still think that their description is wrong and that the inputs should be generated as follows:

- $Z_1, Z_2 \sim \mathcal{N}(0.5, 0.09), X_1, X_2 \sim U[0, 1], Z_3 \sim$ Bernoulli(0.5), $(Z_1, Z_2, X_1, X_2)^T$ has an autoregressive structure with correlation coefficient ρ.
- Two scenarios: (i) $\rho = 0.5$, (ii) $\rho = 0.7$.

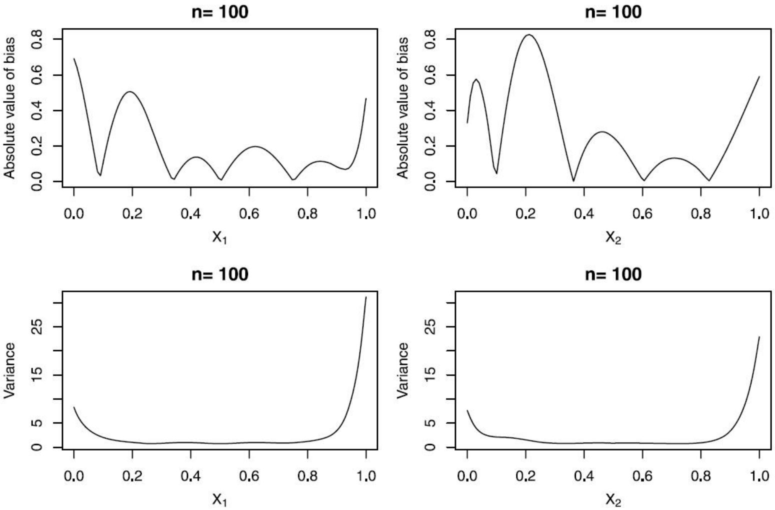
Example 2

Results



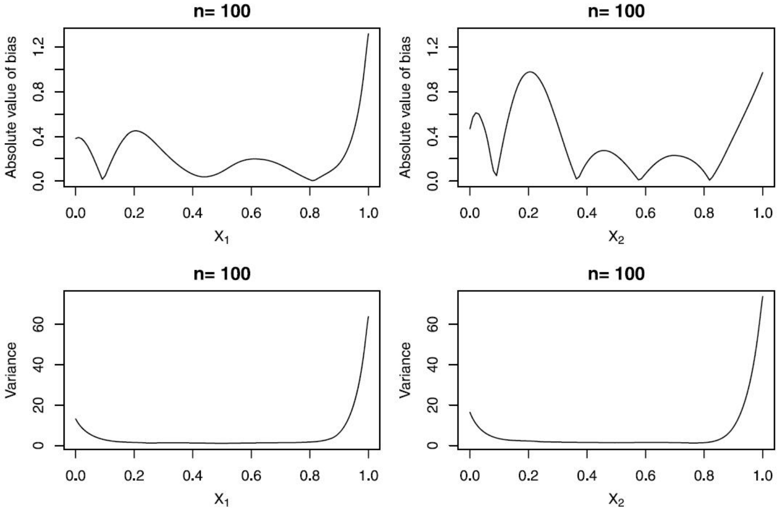
Example 2

Results



Example 2

Results



1 Introduction

2 Methods

3 Numerical Studies
   Simulation
   Real Example

4 Conclusion

Pima Indian diabetes

Data set

- Response: Y, the indicator of a positive test for diabetes.
- Input
  - $NumPreg$, the number of pregnancies.
  - DBP, diastolic blood pressure.
  - DPF, diabetes pedigree function.
  - PGC, the plasma glucose concentration after two hours in an oral glucose tolerance test.
  - BMI, body mass index.
  - AGE (years).

There are 724 complete observations in this data set.

- GLM:

$$\begin{aligned}\text{logit}\Big(Pr(Y=1)\Big) =& \beta_0 + \beta_1 NumPreg + \beta_2 DBP + \beta_3 DPF \\ & + \beta_4 PGC + \beta_5 BMI + \beta_6 AGE.\end{aligned}$$

- GAPLM:

$$\begin{aligned}\text{logit}\Big(Pr(Y=1)\Big) =& \beta_0 + \beta_1 NumPreg + \beta_2 DBP + \beta_3 DPF \\ & + \beta_4 PGC + \eta_1(BMI) + \eta_2(AGE).\end{aligned}$$

Table: Left panel: estimated values, associated standard errors and P-values by using GLM. Right panel: Estimates, associated standard errors using the GAPLM with the proposed variable selection procedures

	GLM				GAPLM		
	Est.	**s.e.**	z **value**	**Pr(>|z|)**	**SCAD (s.e.)**	**LASSO (s.e.)**	**BIC (s.e.)**
NumPreg	0.118	0.033	3.527	0	0 (0)	0.021 (0.019)	0 (0)
DBP	−0.009	0.009	−1.035	0.301	0 (0)	−0.006 (0.005)	0 (0)
DPF	0.961	0.306	3.135	0.002	0.958 (0.312)	0.813 (0.262)	0.958 (0.312)
PGC	0.035	0.004	9.763	0	0.036 (0.004)	0.034 (0.003)	0.036 (0.004)
BMI	0.091	0.016	5.777	0			
AGE	0.017	0.01	1.723	0.085			

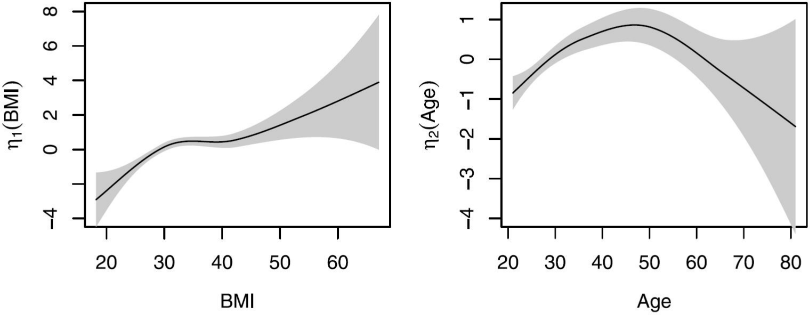


Figure: The patterns of the nonparametric functions of BMI and Age (solid lines) with se. (shaded areas)

Conclusions

- Proposed an effective polynomial spline technique for the GAPLM.
- Proved some asymptotic properties of the estimator under some regularity conditions.
- Developed variable selection procedures for the parametric components.

Contributions

- The procedures are computationally efficient, theoretically reliable, and intuitively appealing.
- The estimators of the linear components, which are often of primary interest, are asymptotically normal.
- the variable selection procedure for the linear components has an asymptotic oracle property.

Thank you!