Smoothing Spline Varying Coefficient Models

Chenlei Leng

July 3, 2003

Abstract

This paper considers model selection in varying-coefficient models.

1 Introduction

Varying-coefficient models (Hastie and Tibshirani (1993)) are a powerful family of nonparametric models. By allowing the linear model coefficients vary with some covariates, the models circumvent the curse of dimensionality to some extent. Quantification of forms of coefficients is usually done in a hypothesis testing frame, see for example Hastie and Tibshirani (1993), Cai, Fan, and Li (2000), Huang, Wu, and Zhou (2002). This paper will tackle this problem from a model (component) selection perspective. By imposing a norm other than the usual square norm on the component to be penalized of each coefficient, an efficient and automatic quantification is obtained.

Assume we have observed $(u_i, x_{ij}, y_i),\ i = 1, ..., n, j = 0, ..., p$ from a process $Y_i = \sum_{j=0}^{p} X_{ij}\eta(U_i)+\epsilon_{ij}$ where ϵ_{ij} is iid normal with unknown variance. The intercept is incorporated into $X_{i0} = (1, ..., 1)^T$. Without loss of generality, we assume $U \in R$, extension to $U \in R^d$ is straightforward.

A penalized least square estimate using smoothing splines has the form

$$A = \frac{1}{n}\sum_{i=1}^{n}[y_i - \sum_{j=0}^{p} x_{ij}\eta_j(u_i)]^T[y_i - \sum_{j=0}^{p} x_{ij}\eta_j(u_i)] + \lambda J(\eta),\ \text{where } J(\eta) = \sum_{j=0}^{p} J(\eta_j). \tag{1.1}$$

2 Model specification

In the smoothing spline analysis of variance models over $[0, 1]^d$, it is assumed η_i's lie in a reproducing kernel Hilbert space. See Wahba (1990). Let $W_2^{(j)}[0, 1] = \{f : f, f' \text{ are absolutely continuous and } f'' \in L_2[0, 1]\}, j = 1, ..., d$ be the second order Sobolev Hilbert space corresponding to the function space over $X^{(j)}$ on $[0, 1]$. When endowed with the inner product

$$(f, g)_{W_2^{(j)}[0,1]} = \{\int_0^1 f(t)dt\}\{\int_0^1 g(t)dt\} + \{\int_0^1 f'(t)dt\}\{\int_0^1 g'(t)dt\} + \int_0^1 f''(t)g''(t)dt,$$

The reproducing kernel corresponds to $W_2^{(j)}$ is $1 + k_1(s)k_1(t) + k_2(s)k_2(t) - k_4(|s - t|)$, with $k_1(s) = s - 0.5, k_2(s) = (k_1^2(s) - 1/12)/2, k_4(s) = (k_1^4(s) - k_1^2(s)/2 + 7/240)/24$.

Now we can use different kernels depending on the goals of the analysis:

(a) if the goal is to select important variables from $X_j,\ j = 1, ..., p$, we can use the reproducing kernel

$$k(s, t) = 1 + k_1(s)k_1(t) + k_2(s)k_2(t) - k_4(|s - t|),$$

under this setting, $\eta_i = Kc_i$ where $K = \{k(u_j, u_k)\}$ and $\eta_i = \{\eta_i(u_j)\}$. We denote $W_2^{(j)}[0,1] = W^{(j)}$. If the model matrix

$$\begin{pmatrix} x_{10} & ... & x_{1p} \\ \vdots & ... & \vdots \\ x_{n0} & ... & x_{np} \end{pmatrix}$$

is of full rank, a unique solution exists;

(b) if the goal is to see whether some of the coefficients are invariant, we can use the reproducing kernel

$$k(s,t) = k_1(s)k_1(t) + k_2(s)k_2(t) - k_4(|s-t|),$$

under this setting, $\eta_i = c_{i0} + Kc_i$ where $K = \{k(u_j, u_k)\}$ and $\eta_i = \{\eta_i(u_j)\}$. $W_2^{(j)}[0,1] = \{1\} \oplus W^{(j)}$ where $W^{(j)}$ has reproducing kernel $k(s,t)$;

(c) if the goal is to see whether the interactions between U and X's can be modeled by simple linear interaction XU, we can use the reproducing kernel

$$k(s,t) = k_2(s)k_2(t) - k_4(|s-t|),$$

under this setting, $\eta_i = c_{i0} + c_{i1}(u - 1/2) + Kc_i$ where $K = \{k(u_j, u_k)\}, u = (u_1, ..., u_n)^T, \eta_i = \{\eta_i(u_j)\}$. $W_2^{(j)}[0,1] = \{1\} \oplus \{k_1\} \oplus W^{(j)}$ where $W^{(j)}$ has reproducing kernel $k(s,t)$.

The COSSO type estimate (Lin and Zhang (2002)) penalizes the L_1 norm of $J(\eta)$ instead of the usual L_2 norms. Namely,

$$A = \frac{1}{n}\sum_{i=1}^{n}[y_i - \sum_{j=0}^{p} x_{ij}\eta_j(u_i)]^T[y_i - \sum_{j=0}^{p} x_{ij}\eta(u_i)] + \lambda J(\eta), \text{ where } J(\eta) = \sum_{j=0}^{p} ||P^j \eta_j||, \tag{2.1}$$

where P^j is the projection of η_j onto $W^{(j)}$.

3 Mathematical details

We will derive the formulation in the setting of (b) in the previous section. This setting can be used to detect varying or constant coefficients. Derivations for the other settings are straightforward.

(2.1) is equivalent to

$$A = \frac{1}{n}\sum_{i=1}^{n}[y_i - \sum_{j=0}^{p} x_{ij}\eta_j(u_i)]^T[y_i - \sum_{j=0}^{p} x_{ij}\eta_j(u_i)] + \lambda\sum_{j=0}^{p}\theta_j^{-1}||P^j\eta_j||^2, \text{ s.t.} \sum_{j=1}^{p}\theta_j \leq M. \tag{3.1}$$

In theory, the smoothing spline estimates are cubic splines with knots placed at each distinct value of U. In this case, we have to solve a $(n+1)(p+1)$ by $(n+1)(p+1)$ linear system which obviously is time consuming and inversion of a large matrix can cause numerical problems. Instead we randomly select N samples $u_{(1)}, ..., u_{(N)}$ from $u_1, ..., u_n$ and use them as knots. A justification of this may be found in Gu and Kim (2002).

Some notations follow. Let $K_{n\times N} = \{k(u_i, u_{(j)})\}$ and $Q_{N\times N} = \{k(u_{(i)}, u_{(j)})\}$ where $k(s,t) = k_1(s)k_1(t) + k_2(s)k_2(t) - k_4(|s-t|)$. The solution can be written as $\eta_j = ec_{j0} + \theta_j K c_j$, where $\eta_j = (\eta_j(u_1), ..., \eta_j(u_n))^T, e = (1, ..., 1)^T, c_{j0} \in R$ and $c_j = (c_{j1}, ..., c_{jN})^T$. θ_j is the smoothing parameter associated with η_j. Denote $D_j = \mathrm{diag}(x_{1j}, x_{2j}, ..., x_{nj}), j = 0, ..., p$ and $Y = (y_1, ..., y_n)^T$. (3.1) can be written as

$$A = \frac{1}{n}||Y - \sum_{j=0}^{p} D_j(ec_{j0} + \theta_j K c_j)||^2 + \lambda \sum_{j=1}^{p} \theta_j c_j Q c_j. \tag{3.2}$$

Taking derivative w.r.t. c_{j0} and c_j, the estimation equation is

$$\begin{pmatrix} e^T D_j^2 e & \theta_j e^T D_j^2 K \\ \theta_j K^T D_j^2 e & \theta_j^2 K^T D_j^2 K + n\lambda\theta_j Q \end{pmatrix} \begin{pmatrix} c_{j0} \\ c_j \end{pmatrix} = \begin{pmatrix} e^T D_j \\ \theta_j K^T D_j \end{pmatrix} [Y - \sum_{i \neq j} D_i(ec_{i0} + \theta_i K c_i)].$$

Rearraging terms, one gets

$$(U^T U + \lambda R_\theta)C = U^T Y,$$

where $U = (D_0 e, \theta_0 D_0 K, D_1 e, \theta_1 D_1 K, ..., D_p e, \theta_p D_p K)$, $C = (c_{00}, c_0^T, ..., c_{p0}, c_p^T)^T$ and

$$R_\theta = \begin{pmatrix} 0 & & & & & \\ & \theta_0 Q & & & & \\ & & 0 & & & \\ & & & \theta_1 Q & & \\ & & & & \cdots & \\ & & & & & \theta_p Q. \end{pmatrix}$$

So $\hat{Y} = HY$ where $H = U(U^T U + n\lambda R_\theta)^{-1} U^T$. A GCV type estimate is then

$$GCV = \frac{nY^T(I - H)^2 Y}{(\mathrm{tr}(I - H))^2}. \tag{3.3}$$

For fixed C, let $Z = Y - \sum_{j=0}^{p} D_j e c_{j0}, G = [D_0 K c_0, ..., D_p K c_p]$ and $S = [c_0^T Q c_0, ..., c_p^T Q c_p]^T$,

$$A = \frac{1}{n}||Z - G\theta||^2 + \lambda S^T \theta, \text{ s.t. } \sum_{j=1}^{p} \theta_j \leq M. \tag{3.4}$$

The algorithm is the following

1 Fix $(\theta_0, ..., \theta_p) = (1, ..., 1)$, find best $\lambda = \lambda_0$ to minimize GCV in (3.3). Assume the coefficients $C = C_\lambda$;

2 For M on a reasonable range, for $C = C_\lambda$, use quadratic programming in (3.4) to solve for θ and then C. Find the $M = \hat{M}$ which gives minimum GCV.

3 The solution corresponding to $\hat{M}$ is the final solution.

Remark 3.1 (A further refinement). Notice the naive smoothing spline estimate (3.1) penalizes $\eta_j, j = 0, ..., p$. Being aware of the parametric (in X) and nonparametric (in U) interactions, it is more natural to put penalties on $X\eta$. The reproducing kernel for such penalties are $k_X K_\eta$, where $k_X(s, t) = k1(s)k1(t)$ and K_η is the reproducing kernel according to the specification in section 2. This might lead to improvement in component selection.

Remark 3.2 (Different starting θ_j's). In the cases all the components have similiar smoothness, the starting value $\theta = (1, .., 1)^T$ works well. In real application especially in biomedical research, the effects of risk factors usually have nice smooth properties. However, if component adopt highly heteorgenous smoothness, a more sensible starting value for θ would be the optimal penalties in the usual smoothing spline cases. By doing this, different degrees of smoothness can be incorporated in this modelling framework.

Remark 3.3 (Backfitting). For large p, a similiar backfitting algorithm can be derived. This avoids the need to invert a large matrix.

4 Simulation

We simulate data $(u, x_1, ..., x_6)$ from $[0, 1]^7$ with

$$y = 6(2u - 1)^2 x_1 + \frac{7\sin(2\pi u)}{(2 - \sin(2\pi u))} x_2 + 4x_3 + \epsilon,$$

with $\text{var}(\epsilon) = 0.2\text{var}(m(u, X_1, X_2, X_3))$ where $m(u, X_1, X_2, X_3) = E(Y|U, X_1, X_2, X_3)$. The signal to noise ration in this case is $5:1$. U is uniform from $[0, 1]$ and $x_1, ..., x_6$ are uniformly on $[0, 1]^6$ with pairwise correlation 0.5. We randomly choose 20 basis functions. From Table 4.1, our estimates have very good component slection ability.

	x_0	x_1	x_2	x_3	x_4	x_5	x_6
(a)	14	100	100	100	8	16	11
(b)	28	100	100	15	15	12	17

Table 4.1: Frequency of models: (a) whether this term is needed; (b) whether this term should be modeled nonparametric.

5 Extension to longitudinal data

Suppose we have repeated measurement $(u_{ij}, x_{ij}, y_{ij}, i = 1, ..., I, j = 1, ..., n_i$, where I is the number of subjects and n_i is the number of measurements on ith subject.

$$A = \frac{1}{n}\sum_{i=1}^{I}\sum_{j=1}^{n_i}[y_{ij} - \sum_{l=0}^{p} x_{ijl}\eta_l(u_{ij})]^T[y_{ij} - \sum_{l=0}^{p} x_{ijl}\eta_l(u_{ij})] + \lambda J(\eta), \text{ where } J(\eta) = \sum_{l=0}^{p} J(\eta_l), \quad (5.1)$$

where $n = \sum_{i=1}^{I} n_i$.

We derive the formulation for the situation (b) in section (2). Extensions to other settings are straightforward.

Some notation follow. Again we use a subset of $u_{(i)}, i = 1, .., N$ as the basis function. Denote $Y_i = (y_{i1}, ..., y_{in_i})^T$, $X_{i\cdot l} = \text{diag}(x_{i1l}, ..., x_{in_il})$,$K_i = \{k(u_{ij}, u_{(k)})\}$ and $Q_{N\times N} = \{k(u_{(i)}, u_{(j)})\}$ where $k(s, t) = k_1(s)k_1(t) + k_2(s)k_2(t) - k_4(|s - t|)$. The solution can be written as $\eta_l = ec_{l0} + \theta_l K_l c_l$, where $\eta_l = (\eta_j(u_{11}), ..., \eta_j(u_{In_I}))^T, e = (1, ..., 1)^T, c_{j0} \in R$ and $c_j = (c_{j1}, ..., c_{jN})^T$. θ_l is the smoothing parameter associated with η_l.

Now, (5.1) can be written as

$$\frac{1}{n}\sum_{i=1}^{I}\{Y_i-\sum_{l=0}^{P}X_{i\cdot l}(c_{l0}+\theta_l K_i c_l)\}^T\{Y_i-\sum_{l=0}^{P}X_{i\cdot l}(c_{l0}+\theta_l K_i c_l)\}+\lambda\sum_{l=0}^{p}\theta_l c_l^T Q c_l \text{ s.t. } \sum_{l=1}^{p}\theta_l\le M. \quad (5.2)$$

Taking derivative with respect to c_{l0} and c_l, the estimation equation is

$$\begin{pmatrix}\sum_{i=1}^{I}e^T X_{i\cdot j}^2 e & \sum_{i=1}^{I}\theta_j e^T X_{i\cdot j}^2 K_i\\ \sum_{i=1}^{I}\theta_j K_i^T X_{i\cdot j}^2 e & \sum_{i=1}^{I}\theta_j^2 K_i^T X_{i\cdot j}^2 K_i+n\lambda\theta_j Q\end{pmatrix}\begin{pmatrix}c_{j0}\\ c_j\end{pmatrix}=\sum_{i=1}^{I}\begin{pmatrix}e^T X_{i\cdot j}\\ \theta_j K_i^T X_{i\cdot j}\end{pmatrix}[Y_i-\sum_{k\ne j}X_{i\cdot k}(ec_{k0}+\theta_k K_i c_k)].$$

Rearraging terms, one gets

$$(\sum_{i=1}^{I}U_i^T U_i+n\lambda R_\theta)C=\sum_{i=1}^{I}U_i^T Y_i,$$

where $U_i=(X_{i\cdot 0}e,\theta_0 X_{i\cdot 0}K_i,X_{i\cdot 1}e,\theta_1 X_{i\cdot 1}K_i,...,X_{i\cdot p}e,\theta_p X_{i\cdot p}K_i)$, $C=(c_{00},c_0^T,...,c_{p0},c_p^T)^T$ and

$$R_\theta=\begin{pmatrix}0 & & & & & \\ & \theta_0 Q & & & & \\ & & 0 & & & \\ & & & \theta_1 Q & & \\ & & & & \cdots & \\ & & & & & \theta_p Q.\end{pmatrix}$$

So $\hat\eta=HY$ where

$$H=\begin{pmatrix}U_1\\ \vdots\\ U_I\end{pmatrix}(\sum_{i=1}^{I}U_i^T U_i+n\lambda R_\theta)^{-1}\begin{pmatrix}U_1\\ \vdots\\ U_I\end{pmatrix}^T,$$

with $H_{jk}=U_j(\sum_{i=1}^{I}U_i^T U_i+n\lambda R_\theta)^{-1}U_k^T$.

For fixed C, let $Z_i=Y_i-\sum_{l=0}^{p}X_{i\cdot l}ec_{l0}$, $G_i=[X_{i\cdot 0}K_i c_0,...,X_{i\cdot p}K_i c_p]$ and $S=[c_0^T Q c_0,...,c_p^T Q c_p]^T$,

$$A=\frac{1}{n}\sum_{i=1}^{I}||Z_i-G_i\theta||^2+\lambda S^T\theta,\text{ s.t. }\sum_{j=1}^{p}\theta_j\le M. \quad (5.3)$$

Following Rice and Silverman (1991) and Hoover, Rice, Wu, and Yang (1998), a leave-one-subject-out cross validation score is used. Denote $\hat\eta^{(-i)}$ as the estimate by leaving out the ith subject. While evaluated at the covariates of ith subject, the estimate is $\hat\eta^{(-i)}(t_i)$, where $t_i=(u_{i1},...,u_{in_i})$,

$$CV=\frac{1}{n}\sum_{i=1}^{I}\{Y_i-\hat\eta^{(-i)}(t_i)\}^T\{Y_i-\hat\eta^{(-i)}(t_i)\}. \quad (5.4)$$

From (5), denote $U = (U_1^T, ..., U_I^T)^T$ and $Y = (Y_1^T, ..., Y_I^T)^T$

$$\begin{aligned}\hat{\eta}^{(-i)}(t_i) =& U_i(\sum_{j\neq i} U_j^T U_j + \lambda R_\theta)^{-1}(U_1^T, ..., U_{(i-1)}^T, U_{(i+1)}^T, ..., U_I^T)\begin{pmatrix} Y_1 \\ \vdots \\ Y_{i-1} \\ Y_{i+1} \\ \vdots \\ Y_I \end{pmatrix} \\ \approx& U_i(\sum_{j=1}^{p} U_j^T U_j + \lambda R_\theta)^{-1}(U_1^T, ..., U_{(i-1)}^T, U_{(i+1)}^T, ..., U_I^T)\begin{pmatrix} Y_1 \\ \vdots \\ Y_{i-1} \\ Y_{i+1} \\ \vdots \\ Y_I \end{pmatrix} \\ =& U_i(\sum_{j=1}^{p} U_j^T U_j + \lambda R_\theta)^{-1}(U^T Y - U_i^T Y_i),\end{aligned} \tag{5.5}$$

in second equation, we used $(\sum_{j\neq i} U_j^T U_j + \lambda R_\theta)^{-1} \approx (\sum_{j=1}^{p} U_j^T U_j + \lambda R_\theta)^{-1}$, i=1,...,I. Let $\Omega = (\sum_{j=1}^{p} U_j^T U_j + \lambda R_\theta)$,

$$CV = \frac{1}{n}\sum_{i=1}^{I}(Y_i - \hat{\eta}_i + U_i\Omega^{-1}U_i^T Y_i)^T(Y_i - \hat{\eta}_i + U_i\Omega^{-1}U_i^T Y_i) \tag{5.6}$$

Remark 5.1 (Connection with Leave-one-out lemma). It's obvious

$$\begin{aligned}\hat{\eta}^{(-i)}(t_i) =& U_i(\sum_{j\neq i} U_j^T U_j + \lambda R_\theta)^{-1}(U_1^T, ..., U_I^T)\begin{pmatrix} Y_1 \\ \vdots \\ Y_{i-1} \\ \hat{\eta}^{(-i)}(t_i) \\ Y_{i+1} \\ \vdots \\ Y_I \end{pmatrix} \\ \approx& U_i(\sum_{j=1}^{p} U_j^T U_j + \lambda R_\theta)^{-1}U^T\begin{pmatrix} Y_1 \\ \vdots \\ Y_{i-1} \\ \hat{\eta}^{(-i)}(t_i) \\ Y_{i+1} \\ \vdots \\ Y_I \end{pmatrix} \\ =& U_i\Omega^{-1}(U^T Y - U_i^T Y_i + U_i^T\hat{\eta}^{(-i)}(t_i)).\end{aligned}$$

So

$$\hat{\eta}^{(-i)}(t_i) = \frac{U_i\Omega^{-1}U^T Y - H_{ii}Y_i}{I - H_{ii}},$$

where I is the identity matrix. Now we get

$$CV \approx \frac{1}{n}\sum_{i=1}^{I} \frac{(Y_i - \hat{\eta}_i)^T (Y_i - \hat{\eta}_i)}{(I - H_{ii})^2}. \tag{5.7}$$

Notice if H is a diagonal matrix, we have exactly the same form while we assume there is no intra-subject correlation. When intra-subject correlation is negligible, replacing $(I - H_{ii})^2$ by $(1 - \frac{1}{n}\mathrm{tr}(H))^2$ might yield significant saving in computational time.

Remark 5.2 (Weighted penalized least square). We can easily modify (5.1) to incorporate weights for each subject,

$$A = \frac{1}{n}\sum_{i=1}^{I} w_i \sum_{j=1}^{n_i} [y_{ij} - \sum_{l=0}^{p} x_{ijl}\eta_l(u_{ij})]^T [y_{ij} - \sum_{l=0}^{p} x_{ijl}\eta_l(u_{ij})] + \lambda J(\eta), \text{ where } J(\eta) = \sum_{l=0}^{p} J(\eta_l), \tag{5.8}$$

where w_i is a nonnegative weight associated with ith subject. Popular choices are $w_i = 1$, in which each observation has equal weight, and $w_i = 1/n_i$, in which each subject has equal weight.

6 Conditional inference

Since the estimated components are linear smoothers conditional on estimated θ, techniques developed in section 3.8 of Hastie and Tibshirani (1990) can be applied. The inference is conditional on the fact we know which component is missing. More spefically, we estimate the variance of Y by

$$\hat{\sigma}^2 = \frac{(Y - \hat{\eta})^T (Y - \hat{\eta})}{\mathrm{tr}(I - H)}$$

in the varying coefficient cases and

$$\hat{\sigma}^2 = \hat{CV}$$

in the longitudinal cases. The pointwise confidence interval for estimated component j can be derived from the covariance matrix V of C,

$$V = \mathrm{var}(C) = \mathrm{var}(\Omega^{-1}U^T Y) = \hat{\sigma}^2 \Omega^{-1} U^T U \Omega^{-1},$$

$$\mathrm{var}(\hat{\eta}_j) = \mathrm{var}([e, \theta_j K]C_j) = [e, \theta_j K]\mathrm{var}(C_j)[e, \theta_j K]^T = [e, \theta_j K]V_{jj}[e, \theta_j K]^T; \tag{6.1}$$

where V_{jj} is a submatrix of V corresponding to C_j. Pointwise standard error bands can be computed using the diagonal of $\mathrm{var}(\hat{\eta}_j)$.

7 Simulation

In the following simulation, a simple random sample of 100 subjects is generated from the model

$$Y_{ij} = g_0(t_{ij}) + \sum_{k=1}^{3} X_{ijk} g_k(t_{ij}) + \epsilon_{t_{ij}}, \; i = 1, ..., 50; j = 1, ..., n_i; \tag{7.1}$$

where

$$g_0(t) = 6(2t-1)^2, \quad g_1(t) = \frac{7\sin(2\pi)}{(2 - \sin(2\pi t))};$$

$$g_2(t) = 4, \quad g_3(t) = 3t.$$

The covariates are generated using the following scheme: $X_{\cdot\cdot 1}(t)$ is uniformly on $[t, t+2]$; $X_{\cdot\cdot 2}(t)$ is normally distributed with zero mean and variance $\{1 + X_{\cdot\cdot 1}(t)\}/\{2 + X_{\cdot\cdot 1}(t)\}$; $X_{\cdot\cdot 3}(t)$, independent of $X_{\cdot\cdot 1}(t)$ and $X_{\cdot\cdot 2}(t)$, is a binary random variable with probability of success 0.6. $X_{\cdot\cdot 4}(t), X_{\cdot\cdot 5}(t), X_{\cdot\cdot 6}(t)$ are nomrally distributed with variance 2, independt of the rest covariates.

To generate the observation time, each individual has a scheduled time points $\{0, 1, ..., 30\}$ and each scheduled time has a probabilty of 60% being skipped. Random perturbations uniform on $[-0.5, 0.5]$ are added to the non-skipped time leading to unequal number of observation times n_i for each subject. The atcual time is rescaled onto $[0, 1]$ and we get t_{ij}. The random errors $\epsilon_i(t_{ij}) = a(Z_i(t_{ij}) + E_{ij})$, where E_{ij} follows $N(0, 4)$ distribution and the $Z_i(t_{ij})$ are generated from a stationary Gaussian preocess with zero mean and decayed exponential covariance function

$$\mathrm{cov}(Z_{i_1}(t_{i_1 j_1}), Z_{i_2}(t_{i_2 j_1})) = \begin{cases} 4\exp(-|t_{i_1 j_1} - t_{i_2 j_1}|) & \text{if } i_1 = i_2, \\ 0 & \text{if } i_1 \neq i_2. \end{cases}$$

and a is a constant which makes the empirical signal to noise ration $5 : 1$. The randomly choose 20 basis functions. A typical fit is shown in Figure 7.1 and 7.2. Both estimated components are very close to the true components. Using the kernel (a) in section yields typical estimates in Figure 7.3 and 7.4.

References

Cai, Z., J. Fan, and R. Li (2000). Efficient estimation and inferences for varying-coefficient models. *Journal of the American Statistical Association 95*(451), 888–902.

Gu, C. and Y.-J. Kim (2002). Penalized likelihood regression: general formulation and efficient approximation. *The Canadian Journal of Statistics 30*(4), 619–628.

Hastie, T. and R. Tibshirani (1990). *Generalized additive models*. Chapman & Hall Ltd.

Hastie, T. and R. Tibshirani (1993). Varying-coefficient models. *Journal of the Royal Statistical Society, B 55*, 757–796.

Hoover, D. R., J. A. Rice, C. O. Wu, and L.-P. Yang (1998). Nonparametric smoothing estimates of time-varying coefficient models with longitudinal data. *Biometrika 85*, 809–822.

Huang, J. Z., C. O. Wu, and L. Zhou (2002). Varying-coefficient models and basis function approximations for the analysis of repeated measurements. *Biometrika 89*(1), 111–128.

Lin, Y. and H. Zhang (2002). Component selection and smoothing in smoothing spline analysis of variance model. Technical Report 1072, University of Wisconsin, Madison.

Rice, J. 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*, 233–243.

Wahba, G. (1990). *Spline Models for Observational Data*. Society for Industrial and Applied Mathematics.

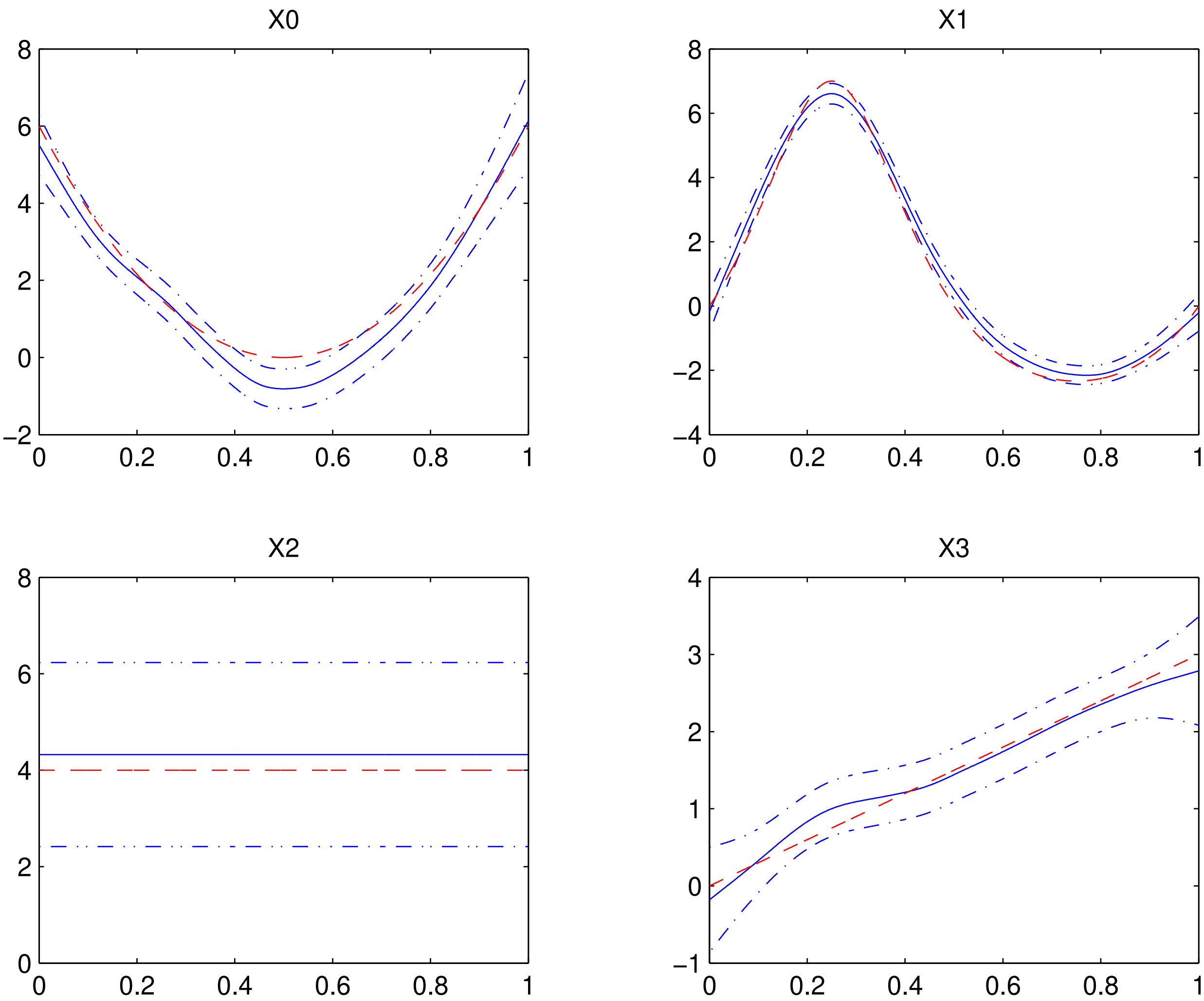

Figure 7.1: Typical fitted curves when $I = 100$ for x_0, x_1, x_2, x_3. This uses kernel in **(b)** of section 2. The blue solid lines indicate the fitted curves. The bluse dash dot line is the ± 2 pointwise error bounds. Red dashed lines are the true components.

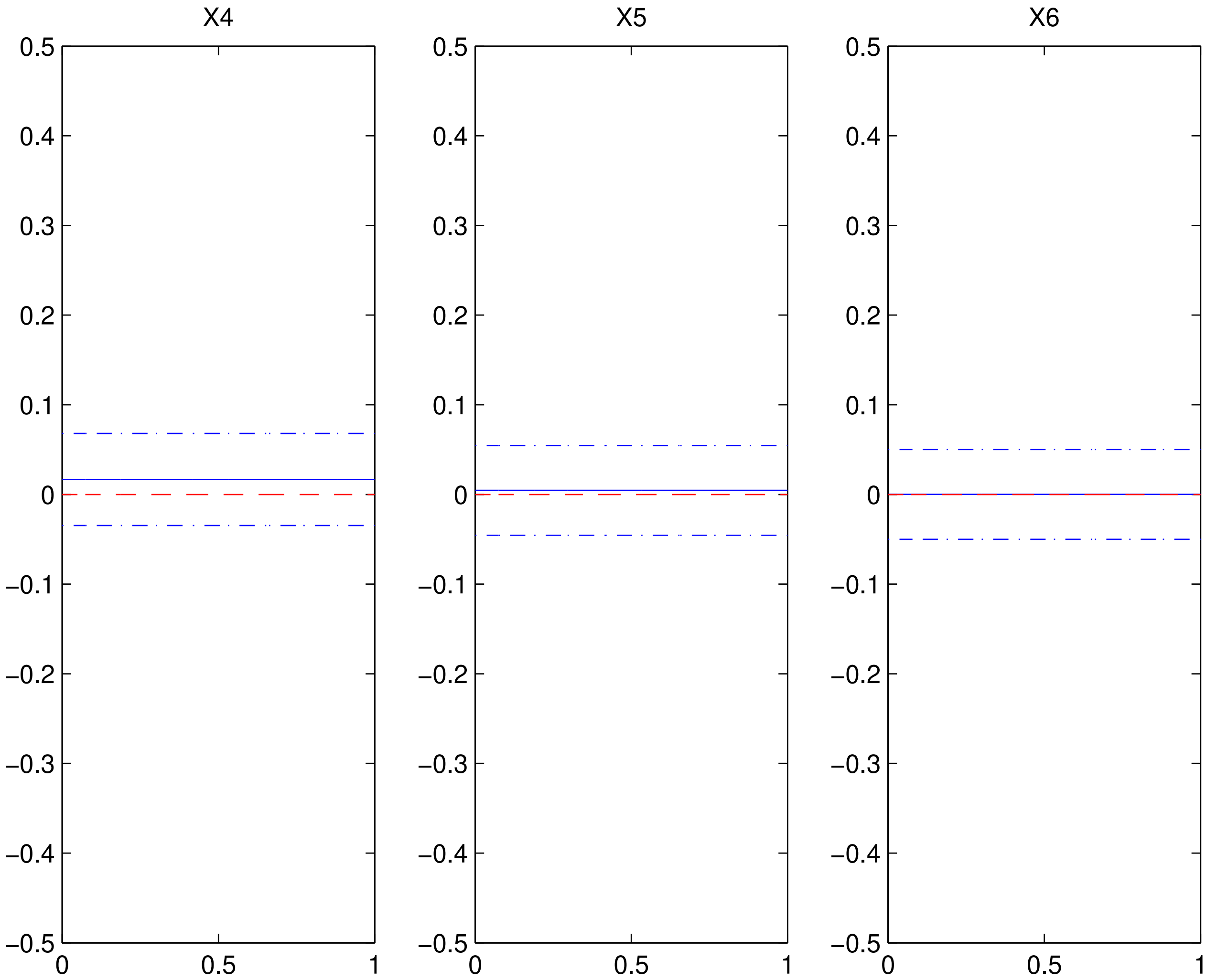

Figure 7.2: Typical fitted curves when $I = 100$ for x_4, x_5, x_6. This uses kernel in **(b)** of section 2. The blue solid lines indicate the fitted curves. The bluse dash dot line is the ± 2 pointwise error bounds. Red dashed lines are the true components.

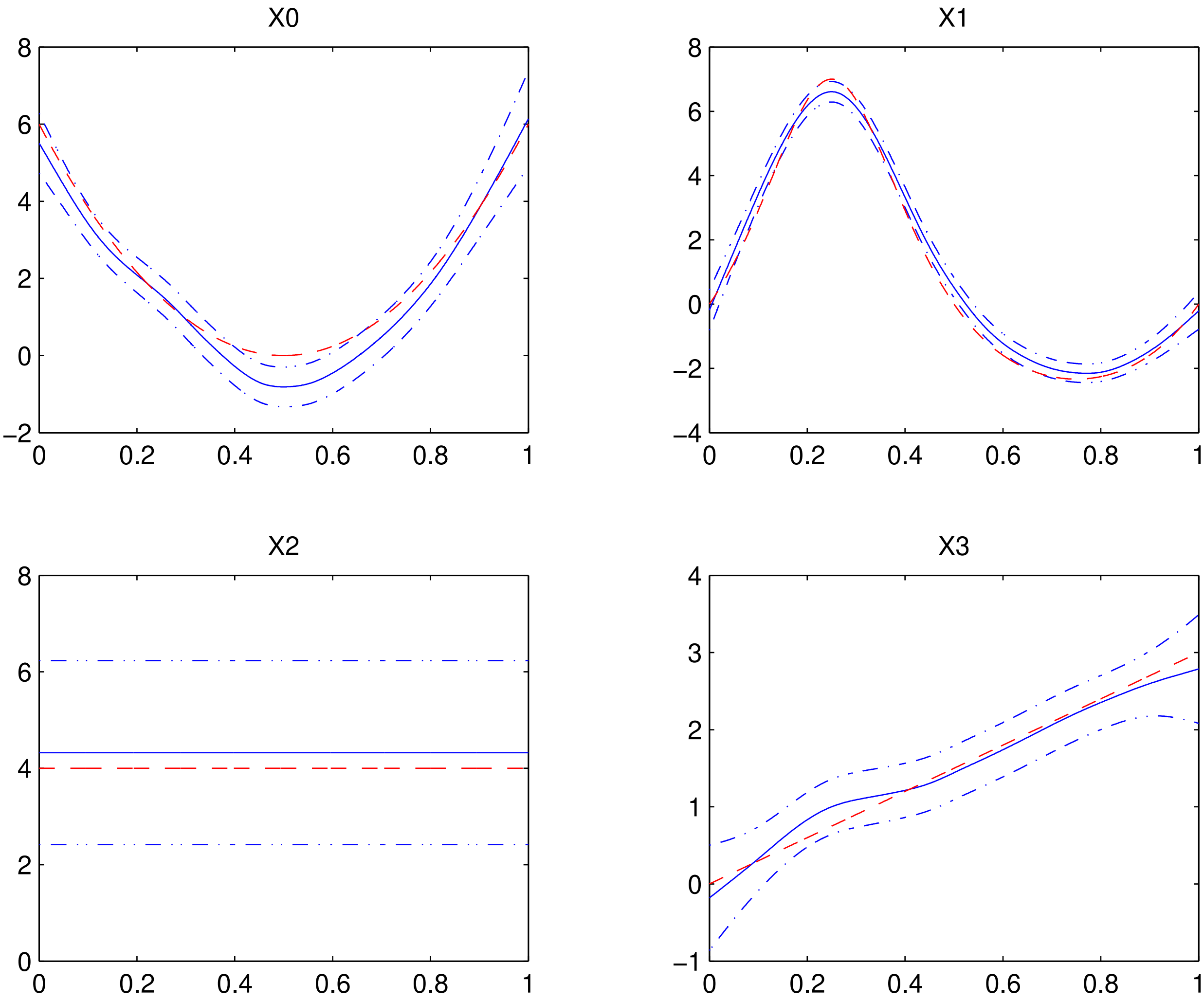

Figure 7.3: Typical fitted curves when $I = 100$ for x_0, x_1, x_2, x_3. This uses kernel in **(a)** of section 2. The blue solid lines indicate the fitted curves. The bluse dash dot line is the ± 2 pointwise error bounds. Red dashed lines are the true components.

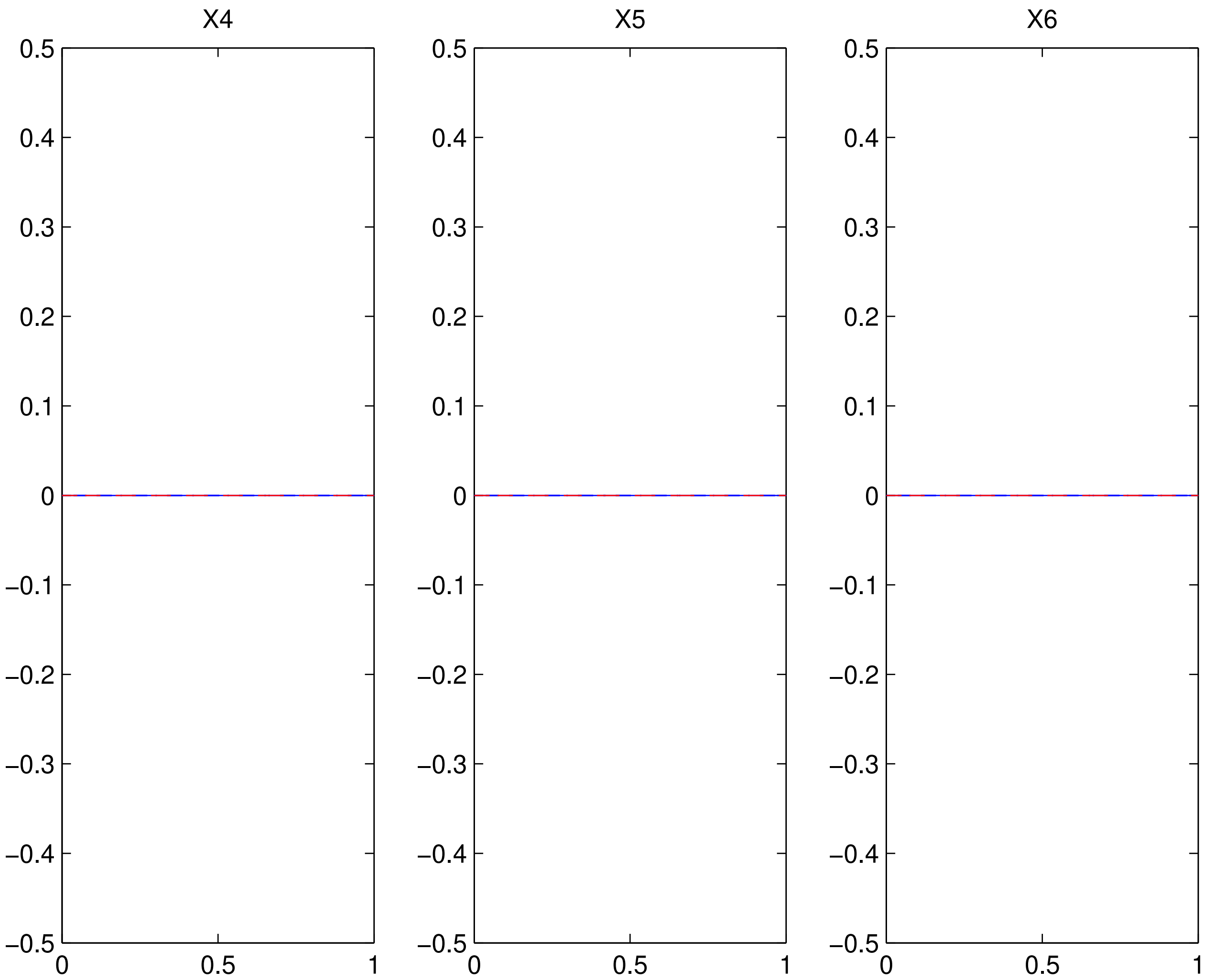

Figure 7.4: Typical fitted curves when $I = 100$ for x_4, x_5, x_6. This uses kernel in **(a)** of section 2. The blue solid lines indicate the fitted curves. The bluse dash dot line is the ± 2 pointwise error bounds. Red dashed lines are the true components.