International Journal of Hybrid Information Technology
Vol. 9, No.12 (2016), pp. 187-200
http://dx.doi.org/10.14257/ijhit.2016.9.12.17

Parametric Estimation for Semi-varying Coefficient Model via Penalized Spline

Guo Haibing[1], Zhao MingtaoZhao[2] and Li Lianqing[3]

[1]. Department of Statistics, Huaihai Institute of Technology, Lianyungang, 222005,China
[2]. School of Statistics and Applied Mathematics, Anhui University of Finance and Economics, Bangbu 233030, China
[3]. School of Science, Huaihai Institute of Technology, Lianyungang, 222005, China
ghb142@163.com

Abstract

A semi-varying coefficient model, a type of semi-parametric model between the linear model and the varying coefficient model offers a good flexibility and has become an important tool for exploring the dynamic patterns in many areas of research, including economics, finance and biomedical research. This paper proposes a new estimation method for semi-varying coefficient models based on the roughness penalized spline approach. Asymptotic properties of their estimators are also obtained. The penalized spline estimation method is applied to an economic example using a semi-varying coefficient model.

***Keywords**: Semi-varying coefficient model; Spline; Average mean square error (AMSE); Confidence Band*

1. Introduction

In recent years, driven by the spread of computer technology and its many sophisticated applications, many useful data-analytic modelling techniques have been proposed to relax traditional parametric models. For the classic linear model

$$Y=\sum_{j=1}^{s} X_j\alpha_j+\varepsilon \tag{1}$$

adequate theories have been derived and widely applied. However such linear models always have a series of strict assumptions, which sometimes cannot be completely satisfied in practice. Under this circumstance, nonparametric models offer a suitable alternative because they require relatively fewer model assumptions and offer greater estimating efficiency. However, the `curse of dimensionality' problem is a hindrance to multiple nonparametric models and can sharply increase estimation variance. To circumvent this problem, statisticians have imposed an additional structure in regression function. Examples include additive models, partially linear models, and varying coefficient models.

Among the above nonparametric or semi-parametric models, the varying coefficient model is encountered in many context and has been successfully applied to multi-dimensional nonparametric regression, and generalized linear models. The varying coefficient model is defined as follows(Hastie and Tibshriani (1993))

$$Y=\sum_{j=1}^{s} X_j\alpha_j(U)+\varepsilon \tag{2}$$

ISSN: 1738-9968 IJHIT

where Y is a response variable, X and U are covariates that depend on U . Because of the generality of function $\alpha(u)$, the model bias can be reduced significantly. Thus, the model has attracted considerable attention (Fan and Zhang (1999); Chiang *et al*. (2001); Zhang (2010)).

In practice, a question often arises as to whether a covariate affects a response or wether a coefficient is varying (Fan and Zhang (2000)). For example, in the frame work of a cross-section production function, c_l and c_f denote a firm's liquid capital and fixed capital, respectively, and u represents the firm's research and development (R\&D) expenditure. It is well known that the level of u will affect the marginal productivity of the fixed capital, but not that of the liquid capital. Hence, c_f may have a varying coefficient $\alpha_f(u)$, but c_l will not. This gives rise to a semi-varying coefficient model, which is defined by the linear model as follows:

$$Y = \sum_{j=1}^{p} Z_j \alpha_j(u) + \sum_{k=1}^{q} X_k \beta_k + \varepsilon \tag{3}$$

This model consists of two parts: a linear part, which involves the constant coefficient $\beta_k(k=1,\cdots,q)$, and a nonparametric part, which involves the varying coefficient $\alpha_j(u)(j=1,\cdots,p)$. The former part can be sufficiently incorporated with prior information, while the latter can be used to relax the model's assumptions. Hence, the semi-varying coefficient model is more flexible than both the parametric linear model and varying coefficient model. Moreover, the semi-varying coefficient model can efficiently avoid the 'curse of dimensionality' problem.

There are a sequential of literatures on the estimation procedure for the semi-varying coefficient model. Fan and Huang (2002) suggested using the kernel-based profile likelihood approach to estimate a partially varying coefficient model. Ahamad, Leelahanon and Li (2005) proposed a general series estimation method, and their results can be extended to a conditional heteroskedasticity error case in a straightforward manner, but with no asymptotic normality result. Motivated by two practical problems: the relationship of human health to air pollution and weather condition, and the transmission mechanism of epidemics, Xia, Zhang and Tong (2004) proposed a new estimating procedure for (3) based on a local linear model, which can be used to simultaneously estimate the parametric and nonparametric functions of the same model at their optimal consistency rates.

In addition to above local linear approximate method, the spline approach offers an alternative estimating method for varying coefficient functions in model (3). Hastie and Tibshirani(1993) outline a spline estimating method for the varying coefficients model and provided an improved algorithm for deriving estimators. For multivariable survival models with varying coefficients, Kauermann (2005) considered penalized spline smoothing for hazard regression. Chiang Rice and Wu(2001) applied a kind of smoothing spline estimating method to a special varying coefficients model, with a repeatedly measured dependent variable. There are many studies on the use of the spline method to estimate varying coefficient models, but the application of the spline of to semi-varying coefficient models has receive less attention, especially for asymptotic properties. Li, Xue and Lian(2011) used a B-spline to derive estimators for the semi-varying coefficient model with a diverging number of components.

In fact, compared with the local linear method, spline method place more emphasis on 'globe approximation' because there are more overlaps among basis functions as their degree increases. The present paper proposes a new estimation procedure for semi-varying coefficient models based on the roughness penalized spline approach proposed by Green and Sileverman (1994), in which varying coefficient functions are estimated in piecewise, and both the varying and fixed coefficients are estimated simultaneously in a single step. Moreover, asymptotic properties of their estimators are obtained.

The rest of paper is organized as follows. In Section 2, we present the semi-varying coefficient model and provide some notations to simplify the model in a matrix form. In Section 3, the estimation of parameters by the roughness penalized spline approach as well as their numeric characters and asymptotic properties is presented. In Section 4, we present an application of our proposed approach to economic analysis. The sketched proofs of theorems in Section 3 are given in the Appendix.

2. The Semi-varying Coefficient Model

Suppose that $(y_i, z_{i1}, \cdots, z_{ip}, x_{i1}, \cdots, x_{iq}, u_i), (i=1,\cdots,n)$ denote a random sample from model (3), where y_i is response variable, and $(z_{i1}, \cdots, z_{ip}, x_{i1}, \cdots, x_{iq}, u_i)$ are covariate variables. Then, the semi-varying coefficient model (3) can be rewritten as

$$y_i = \sum_{j=1}^{p} z_{ij}\alpha_j(u_i) + \sum_{k=1}^{q} x_{ik}\beta_k + \varepsilon_i (i=1,\cdots,n) \tag{4}$$

Let $x_k = (x_{1k}, \cdots, x_{nk})^T (k=1,\cdots,q)$, $z_j = diag(z_{1j}, \cdots, z_{nj})(j=1,\cdots,p)$, $Z=(z_1,\cdots,z_p)$, $X=(x_1,\cdots,x_p)$ and the varying coefficients $\alpha_j(U) = (\alpha_j(u_1),\cdots,\alpha_j(u_n))^T$ and $\alpha(U) = (\alpha_1(U)^T, \cdots, \alpha_p(U)^T)^T$. Correspondingly, $\beta = (\beta_1, \cdots \beta_q)^T$ denotes the fixed coefficients and $\varepsilon = (\varepsilon_1, \cdots \varepsilon_n)^T$ denotes the error vector. Hence, model (4) can be rewritten in the matrix form as follows:

$$Y = Z\alpha(U) + X\beta + \varepsilon \tag{5}$$

Without losing generality, we can suppose that the variables $\{u_i\}$ are distinct from each other at an interval [0,1]. Let $U^0 = (u_1^0, \cdots u_n^0)^T$ be the vector of ascending ordered values of $\{u_i\}$, and correspondingly, $\alpha(U^0) = (\alpha_1(U^0)^T, \cdots, \alpha_p(U^0)^T)^T$. Let N be the incidence matrix mapping U to U^0, then we have $\alpha(U) = N\alpha(U^0)$. When $Z = ZN$, model (5) can be rewritten as

$$Y = Z\alpha(U^0) + X\beta + \varepsilon \tag{6}$$

In this paper, we make following assumptions:

(A1) The varying coefficient function $\alpha_j(\cdot)(j=1,\cdots,p)$ are defined in the function space $W(a,b) = \left\{ f(x) : f^{'} \quad absolutely \quad continuous; f^{(2)} \in L^2(a,b) \right\}$.

(A2) Let h_i denoting the distance between two knots u_i^0 and u_{i+1}^0, *i.e.* $h_i = u_{i+1}^0 - u_i^0 (i=1,\cdots,n-1)$ and $h = \max\{h_1,\cdots,h_{n-1}\}$, $h_0 = \min\{h_1,\cdots,h_{n-1}\}$, there exist a constant $M > 0$, such that $h/h_0 \le M$.

(A3) The deterministic design points $u_i \in [0,1], i=1,\cdots,n$ assume that there exist a distribution function Q with corresponding positive continuous design density function $q(x)$ such that, with Q_n the empirical distribution of $u_1,\cdots,u_n$,

$$\sup_{u\in[0,1]} |Q_n(u) - Q(u)| = o(\frac{1}{n})$$

(A4) The error vector ε is white noise

$$\varepsilon \sim (0, \sigma^2 I) \tag{7}$$

i.e. $E(\varepsilon) = 0$ and $Cov(\varepsilon) = \sigma^2 I$, where 0 is a $n \times 1$ vector and I is a $n \times n$ identity matrix.

3. Estimation of Parameters

For given sample value $u_1,\cdots,u_n$, suppose that the varying coefficients function $\alpha_j(\cdot)(l=1,\cdots,p)$ is a piecewise cubic spline on each of the intervals $(0,u_1],(u_1,u_2],\cdots,(u_n,1]$. We can then plot the entire function through points $(u_1,\alpha_j(u_1)),\cdots,(u_n,\alpha_j(u_n))$ (Green and Silverman(1994)). However, in model (4), the real value of $\alpha_j(u_1),\cdots,\alpha_j(u_n)$ are unknown and we must first estimated these values. Following Ni, Zhang and Zhang (2010), we consider the penalized likelihood function as follows

$$L(\alpha(U^0),\beta;Y)=(Y-Z\alpha(U^0)-X\beta)^T(Y-Z\alpha(U^0)-X\beta)+\xi\sum_{j=1}^{p}\int\alpha_j''(s)^2ds \qquad (8)$$

Note that (8) has two parts: the first term on the right hand side is used to control the degree of model fitting and the second term is used to control the degree of smoothness for the nonparametric function $\alpha_j(\cdot)$, where ξ is a non-negative smoothing parameter. For the cubic spline function under the assumption (A1), we can rewrite the integral term of (8) in the matrix form as follows (from equation (2.3) in Green and Silverman (1994))

$$\int_0^1\alpha_j''(s)^2ds=\alpha_j(U^0)^TK_0\alpha_j(U^0),\ j=1,\cdots,p \qquad (9)$$

where K_0 is the smoothing matrix with respect to variable $U^0=(u_1^0,\cdots,u_n^0)$, which will be demonstrated in Lemma 1. Denoting $K=diag(K_1,\cdots,K_p)$, where $K_1\equiv K_2\equiv\cdots\equiv K_p\equiv K_0$, model (8) has the matrix form as follows:

$$L(\alpha(U^0),\beta;Y)=(Y-Z\alpha(U^0)-X\beta)^T(Y-Z\alpha(U^0)-X\beta)+\xi\alpha(U^0)^TK\alpha(U^0) \qquad (10)$$

Minimizing $L(\alpha(U^0),\beta;Y)$, we can establish the following theorems (their proof is sketched in the Appendix).

Theorem 1. *The penalized likelihood estimator of coefficients for model (6) is*

$$\begin{cases}\hat\alpha_j(U^0)=W_j(Z^TQ_XZ+\xi K)^{-1}Z^TQ_XY,(j=1,\cdots,p)\\ \beta=(X^TX)^{-1}X^T(I-SQ_X)Y\end{cases} \qquad (11)$$

where $W_j=(0_{n\times(j-1)n},I_n,0_{n\times(p-j)n})$ *and* $0_{n\times(j-1)n}$ *denotes the* $n\times(j-1)n$ *zero matrix,* $S=Z(Z^TQ_xZ+\xi K)^{-1}Z^T$, $Q_x=I-P_x=I-X(X^TX)^{-1}X^T$.

Remark: Under the assumption (A1--A4) and $nh^3\to 0$, it follows from (A.39) that the estimator β in (11) is consistent with a parametric estimator, *i.e.*

$$\beta\to(X^TX)^{-1}X^TY.$$

This means that the asymptotic estimation of β only depends on fixed coefficients X regardless whether X is correlated to Z.

Now that we have obtained the estimated points $(u_1,\alpha_j(u_1)),\cdots,(u_n,\alpha_j(u_n))$, we can use them to estimate the entire function $\alpha_j(\cdot)$, Green and Siverman(1994) report that for any $u\in[0,1]$

$$\alpha_j(u)=M(u)\alpha_j(u^0) \qquad (12)$$

where $M(u)=N(u)+LJ(u)$ and $L=(0_{n\times1},Q^TR^{-1},0_{n\times1})^T$, R and Q are defined in Lemma1. The vector $N(u)=(N_1(u),\cdots,N_n(u))$ satisfies

$$N_k(u)=\begin{cases}\frac{u-u_1}{h_1}I_{\{u_1<u<u_k\}}, & k=1,\\ \frac{u_{k+1}-u}{h_k}I_{\{u_k<u<u_{k+1}\}}+\frac{u-u_k}{h_{k-1}}I_{\{u_{k-1}<u<u_k\}}, & 2\le k\le n-1,\\ \frac{u_n-u}{h_{n-1}}I_{\{u_{n-1}<u<u_n\}}, & k=n,\end{cases}$$

and in the same manner, the vector $J(u)=(J_1(u),\cdots,J_n(u))$ satisfies

$$J_k(u)=\begin{cases}(u-u_1)(u_2-u)(1+\frac{u-u_1}{h_1})I_{\{u_1<u<u_2\}}, & k=1,\\ (u-u_k)(u_{k+1}-u)(1+\frac{u-u_k}{h_k})I_{\{u_k<u<u_{k+1}\}}-(u-u_{k-1})(1+\frac{u-u_{k-1}}{h_{k-1}})I_{\{u_{k-1}<u<u_k\}}, & 2\le k\le n-1,\\ (u-u_{n-1})(u_n-u)(1+\frac{u_n-u}{h_{n-1}})I_{\{u_{n-1}<u<u_n\}}, & k=n.\end{cases}$$

The local asymptotic property of $\alpha_j(\cdot)$ can be derived from the theorems as follows.

Theorem 2. *Under the assumptions (A1--A4) and* $n\to\infty$, $nh^3\to 0$, *for any fixed* $u\in(u_i,u_{i+1}]$, $i=1,\cdots,n-1$, *we* have

$$E(\alpha_j(u)-\alpha_j(u))=b_j(u)+o(n^2ph^6) \tag{13}$$

where $b_j(u)=-\frac{\alpha_i^{(3)}(u_i)h_i^3}{3!}B_3(\frac{u-u_i}{h_i})$ *and* $B_3(\cdot)$ *is the 3rd Bernoulli polynomial, which is inductively defined as* follows

$$B_0(x)=1, B_i(x)=\int_0^x iB_{i-1}(z)dz+b_i \tag{14}$$

where $b_i=-i\int_0^1\int_0^x B_{i-1}(z)dzdx$ *is the ith Bernoulli number (see Barrow and Smith(1978)).*

Theorem 3. *Under the same assumptions in Theorem2, for any* $u\in[0,1]$

$$\frac{\alpha_j(u)-(\alpha_j(u)+b_j(u))}{\sqrt{Var(\alpha_j(u))}}\to_d N(0,1) \tag{15}$$

where $b_j(u)$ *is defined as in Theorem 2.*

We now apply Theorems 2 and 3 to construct pointwise confidence intervals for the varying coefficient function $\alpha_j(u)$.

Theorem 4. Under the assumptions (A1-A4), for any $u\in[0,1]$, the $100(1-\alpha)\%$ asymptotic confidence interval for $\alpha_j(u)$ is

$$[\alpha_j(u)-\sum_{k=2}^{n-1}\frac{c_{jk}h_k^3}{3!}B_3(\frac{u-u_k}{h_k})I_{\{u_k<u\le u_{k+1}\}}]\pm z_{\alpha/2}Var(\alpha_j(u)) \tag{16}$$

where $z_{\alpha/2}$ is the $(1-\alpha/2)$ th normal percentile and function $B_3(\cdot)$ is defined in Theorem 2 , $(c_{j2},\cdots,c_{j(n-1)})=W\alpha_j(U^0)$, where W is derived from (A.33)

Theorem 5. *Under the assumption (A1--A4) and when* $h\to 0$ *and* $n\to\infty$, $\alpha_1(U^0),\cdots,\alpha_p(U^0)$ *defined in (12) satisfy*

$$\sum_{j=1}^{p} AMSE(\alpha_j(U^0)) = o(\frac{1}{n}\sum_{k=1}^{n-2}(\frac{k^4}{\xi k^4 + \frac{1}{c_0}nh^3})^2) + o(\frac{\sigma^2}{n}[l_0 + \sum_{k=1}^{n-2}\frac{n^2h^6}{(\xi c_0 k^4 + nh^3)^2}]) \quad (17)$$

where $\alpha_1(U^0),\cdots,\alpha_p(U^0)$ *are defined in (14), and* l_0 *and* c_0 *are constants that depends on the distribution of* U^0.

Theorem 6. *The fixed coefficient estimators in (10) have the expectation and covariance matrix as follows:*

$$\begin{cases} E(\beta) = \beta + (X^T X)^{-1} X^T (I - SQ_X) Z\alpha(U^0) \\ Cov(\beta) = \sigma^2 (X^T X)^{-1} X^T (I - SQ_X)(I - Q_X S) X (X^T X)^{-1} \end{cases} \quad (18)$$

where S , W_j *and* Q_X *have the same representation as in Theorem 1.*

4. Application

To introduce the semi-varying coefficient model, in Section 1, we have presented an example of the cross-section production function. We now return to this example and use it to illustrate an application of our proposed estimation approach. In this section, we consider the estimation of a production function for China's manufacturing industry in a semi-varying coefficient model. The data was drawn from the Second Economic Census conducted by the National Statistic Bureau of China in 2008. To conduct a macro-analysis of the production function, we used indices from every province for different industries as samples. To avoid heterogeneity across different industries and to achieve sufficiently large sample sizes, we include industry sectors as follows: food processing from agriculture, foods manufacturing, beverage manufacturing and tobacco manufacturing. After removing some missing values, the sample size is 113.

First, we considered the benchmark parametric linear model as follows:

$$\ln Y = \beta_0 + \gamma \ln W + \beta_l \ln L + \beta_k \ln K + \beta_z \ln Z + \varepsilon \quad (19)$$

where Y is the industrial output value, W is the liquid capital, K is the fixed capital, Z is the R\&D input, L is the labour input, and ε is error term. All monetary measures are in hundred million Renminbi, and labour is measured in ten thousands.

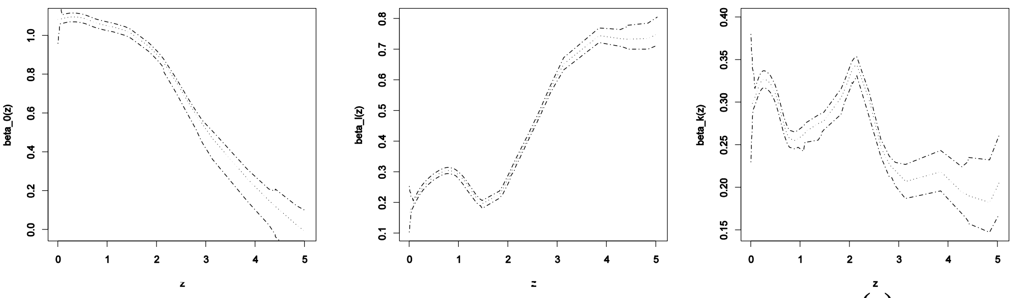
Through the least-square theory, the model (18) has the estimated form as follows:

$$\begin{array}{llllll} y = & 1.17766 & +0.61931\ln W & +0.25188\ln L & +0.32413\ln K & -0.05681\ln Z \\ & (6.61\times10^{-6}) & (3.9\times10^{-4}) & (0.000148) & (0.007608) & (0.093459) \end{array} \quad (20)$$

where the numbers in parenthesis are the p- values of the respective coefficients. Clearly, the variable Z has the maximum p-value and the null hypothesis $(\beta_z = 0)$ cannot be rejected at a significance level of 0.05. Hence, the linear model (19) is a poor fit with the the industry production function. It is likely that the level of R&D expenditure affects the marginal productivity of the fixed capital and labour inputs, but not that of the liquid capital. A semi-varying coefficients industry production model can be derived as follows (Ahamad, Leelahanon and Li(2005)):

$$\ln Y = \gamma \ln W + \alpha_0(z) + \alpha_l(z)\ln L + \alpha_k(z)\ln K + \varepsilon \quad (21)$$

and these parameters can be estimated using the roughness penalized spline approach. The estimated value of γ is 0.6651809 and the varying coefficients $\alpha_0(z)$, $\alpha_l(z)$ and $\alpha_k(z)$ and their point-wise confidence intervals are plotted in the graphs presented below:



(a) (b) (c)

Figure 1. (a) $\beta_0(z)$, (b) $\beta_l(z)$, (c) $\beta_k(z)$

Figure 1(a) shows the plot for $\alpha_0(z)$; the figure indicates that $\alpha_0(z)$ is nonlinear function of z and that $\alpha_0(z)$ has decreases as z increases. Figure 1(b) shows that the marginal productivity of labour $\alpha_l(z)$ is also a nonlinear function of z. Although $\alpha_l(z)$ locally varies with z, it has a general upward trend, which means that firms with large R\&D expenditure have higher marginal labour productivity. Figure 1(c) shows that the marginal productivity of the (fixed) capital fluctuates with z and reaches a maximum value at $z = 2.1$. These results are not surprising given the impact of the subprime crisis and the subsequent Chinese government bailout. Because of the unbalance distribution of bailout capital accross different regions and industries, some of the companies sampled have sufficient capital but others experienced capital scarcities. This somewhat affects the R&D input, and most of the R&D expense are used to improve equipment performance, rather than to train labor force. This in turn may account for the varying marginal capital productivity with respect to z. Nevertheless, the marginal capital productivity $\beta_k(z)$ exhibits a general downward trend. Overall, we can conclude that a higher R&D input can increase marginal labour productivity, and decrease marginal capital productivity. Our research also shows that the industrial output value (y) has a positive correlation with the R\&D input (z).

We next applied the local linear method to this data set. The estimated value of γ was 0.6646276 and the estimated varying coefficients were similar to those obtained by using the roughness method. To compare the efficiency of the two estimation methods, we considered the goodness-of-fit:

$$R^2 = \frac{SSR}{SST} = \frac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \quad .$$

The results show that the roughness method $(R^2 = 0.952385)$ is marginally superior to the local linear method $(R^2 = 0.950503)$.

5. Appendix

Lemma 1. Let A is a positive matrix $(A_{n\times n} \geq 0)$ and K_0 is the smoothing matrix defined in (9), under the assumption (A2) and $h \to 0$, the matrix $(A+\xi K_0)^{-1}K_0 (\xi > 0)$ has the eigenvalues as follows:

$$\lambda_i((A+\xi K_0)^{-1}K_0) = \frac{c_i}{\xi c_i + o(h^3)}, i = 1, \cdots, n,$$

where $c_1 = c_2 = 0$, $c_i = \frac{(i-2)^4}{n} c_0(1+o(1))$, c_0 is a constant depend on the distribution of U_0 and $\lambda_i(A)$ denotes the ith eigenvalue of the matrix A which is in ascending order.

Proof: It follows from Green and Silverman (1994) that the smoothing matrix in (9) satisfy $K_0 = QR^{-1}Q^T$, where

$$R = \begin{pmatrix} \frac{h_1+h_2}{3} & \frac{h_2}{6} & & \\ \frac{h_2}{6} & \frac{h_2+h_3}{3} & \ddots & \\ & \ddots & \ddots & \frac{h_{n-2}}{6} \\ & & \frac{h_{n-2}}{6} & \frac{h_{n-2}+h_{n-1}}{3} \end{pmatrix},$$

and

$$Q = \begin{pmatrix} h_1^{-1} & & \\ -(h_1^{-1}+h_2^{-1}) & \ddots & \\ h_2^{-1} & \ddots & h_{n-2}^{-1} \\ & \ddots & -(h_{n-2}^{-1}+h_{n-1}^{-1}) \\ & & h_{n-1}^{-1} \end{pmatrix}.$$

Denote $R = hR_0$ and $Q = h^{-1}Q_0$, we have

$$K_0 = h^{-3}K_0^{'} := h^{-3}Q_0R_0^{-1}Q_0^T \tag{A.1}$$

It follows from the assumption (A2) that the value of elements both in Q_0 and R_0 are independent each other when $h \to 0$.

Let c_i is a eigenvalue of $Q_0R_0^{-1}Q_0^T$ and u_i is the corresponding eigenvector, *i.e.*

$$Q_0R_0^{-1}Q_0^T u_i = c_i u_i, i = 1, \cdots, n \tag{A.2}$$

and $0 = c_1 = c_2 < c_3 \le \cdots \le c_n$. Demmler and Reinch(1975) indicate that the whole components of eigenvector becomes oscillating as their corresponding eigenvalue increases. Further, Speckman(1985) give the approximate eigenvalues in (A.2) as

$$c_1 = c_2 = 0, c_i = \frac{(i-2)^4}{n} c_0(1+o(1)), i = 3, \cdots, n \tag{A.3}$$

where c_0 is a constant that only depends on the distribution of U_0.

Let $L = A + \xi K_0$ is a invertible matrix, then $L > 0$. Extracting the factor h^{-3} from L yields

$$L = h^{-3}L_0 := h^{-3}(h^3 A + \xi K_0^{'}) .$$

Given $A \ge 0, K_0^{'} \ge 0$ and the Weyl's inequality, we get

$$h^3\lambda_n(A) \ge \lambda_i(L_0) - \xi\lambda_i(K_0^{'}) \ge h^3\lambda_1(N), i = 1, \cdots, n \tag{A.4}$$

Thus, when $h \to 0$,

$$\lambda_i(L_0) = \xi\lambda_i(K_0^{'}) + o(h^3) .$$

These means that there exist an orthogonal matrix P, which satisfies

$$P^T L_0 P = P^T(o(h^3)A)P + \xi P^T K_0^{'} P = diag(o(h^3), \cdots, o(h^3)) + \xi diag(c_1, \cdots, c_n) .$$

Use $(A+\xi K_0)^{-1}K_0 = L_0^{-1}K_0^{'}$, we have

$$P^T L_0^{-1} K_0 P = (P^T L_0 P)^{-1} P^T K_0 P = diag(\frac{c_1}{\xi c + o(h^3)}, \cdots, \frac{c_n}{\xi c_n + o(h^3)}).$$

The proof of Lemma 1 is completed.

Proof of Theorem 1. The differentiation of (10) with respect to $\alpha(U^0)$ and β are

$$\begin{cases} \dfrac{\partial L(\alpha(U^0), \beta; Y)}{\partial \alpha(U^0)} = -2Z^T(Y - Z\alpha(U^0) - X\beta) + 2\xi K\alpha(U^0), \\ \dfrac{\partial L(\alpha(U^0), \beta; Y)}{\partial \beta} = -2X^T(Y - Z\alpha(U^0) - X\beta). \end{cases}$$

Thus, the equation system is set as follows:

$$\begin{cases} Z^T Z\alpha(U^0) + Z^T X\beta + \xi K\alpha(U^0) = Z^T Y, \\ X^T Z\alpha(U^0) + X^T X\beta = X^T Y. \end{cases}$$

Then, we yield

$$\begin{pmatrix} Z^T Z + \xi K & Z^T X \\ X^T Z & X^T X \end{pmatrix} \begin{pmatrix} \alpha(U^0) \\ \beta \end{pmatrix} = \begin{pmatrix} Z^T Y \\ X^T Y \end{pmatrix}.$$

Solving this equation, we obtain

$$\begin{cases} \alpha(U^0) = (Z^T Q_X Z + \xi K)^{-1} Z^T Q_X Y, \\ \beta = (X^T X)^{-1} X^T (I - SQ_X)Y, \end{cases}$$

The proof of Theorem 1 is completed.[]

Now we introduce Barrow and Smith (1978) research results about the approximation error of spline function on $S([0,1])$:

$$\inf_{s(u)\in([0,1])} \| f(u) + b(u) - s(u) \|_{L_\infty} = o(h^3) \quad \text{(A.5)}$$

where

$$b(u) = -\frac{f^{(3)}(u_i) h_i^3}{3!} B_m(\frac{u - u_i}{h_i}).$$

It follows from (A.5) that there exists a $s_{\alpha_j}(u) \in S([0,1])$ such that

$$\alpha_j(u) + b_j(u) - s_{\alpha_j}(u) = o(h^3) \quad \text{(A.6)}$$

Proof of Theorem 2. From (12) and (A.6), we have

$$\begin{aligned} E(\alpha_j(u)) - \alpha_j(u) &= [E(\alpha_j(u)) - s_{\alpha_j}(u)] + [s_{\alpha_j}(u) - \alpha_j(u)] \\ &= [E(\alpha_j(u)) - s_{\alpha_j}(u)] + b_j(u) + o(h^3) \end{aligned} \quad \text{(A.7)}$$

To obtain (13), it is suffice to show that

$$\begin{aligned} E(\alpha_j(u)) - \alpha_j(u) &= [E(\alpha_j(u)) - s_{\alpha_j}(u)] + [s_{\alpha_j}(u) - \alpha_j(u)] \\ &= [E(\alpha_j(u)) - s_{\alpha_j}(u)] + b_j(u) + o(h^3) \end{aligned} \quad \text{(A.8)}$$

Because $s_{\alpha_j}(u) \in S([0,1])$, it follows from (11) and (10) that

$$E(\alpha_j(u)) - s_{\alpha_j}(u) = M(u) W_j [(Z^T Q_X Z + \xi K)^{-1} Z^T Q_X Z](\alpha(U^0) - s_\alpha(U^0)) \quad \text{(A.9)}$$

where $\alpha(U^0)$ is defined in (6), and $s_\alpha(U^0) = (s_\alpha(U^0)^T, \cdots, s_{\alpha_p}(U^0)^T)^T$, where $s_{\alpha_j}(U^0) = (s_{\alpha_j}(U_1^0), \cdots, s_{\alpha_j}(U_n^0))^T, j = 1, \cdots, p$ .Note that $(A + hB) = A^{-1} + hB^{-1}AB^{-1} + o(h^2)$ and the definition of matrix K in (A.1), we have

$$(Z^T Q_X Z + \xi K)^{-1} Z^T Q_X Z = h^3 \xi^{-1} K_0^{'-1} Z^T Q_X Z + o(h^6) \quad \text{(A.10)}$$

Thus, from (A.1),(A.9) and (A.10), we obtain

$$E(\alpha_j(u)) - s_{\alpha_j}(u) = h^3\xi^{-1}M(u)\sum_{k=1}^{p} K_0^{-1} Z_j Q_X Z_j \gamma_k + o(h^6) \qquad (A.11)$$

where $\gamma_k = \alpha(U^0) - s_{\alpha_j}(U^0)$ and K_0 is the diagonal elements for block diagonal matrix $K_0^{'}$. Suppose that k_{ij}^* and Q_{ij} are the $(i,j)th$ elements of K_0^{-1} and Q_X respectively, and $\gamma_k = (\gamma_{1n}, \cdots, \gamma_{kn})$, we have

$$E(\alpha_j(u)) - s_{\alpha_j}(u) = o(h^3\xi^{-1}\sum_{l=1}^{n}\sum_{k=1}^{n}\sum_{t=1}^{n} k_{lt}^* z_{jt} Q_{ts}\gamma_{ks} M_l(u)) \qquad (A.12)$$

It follows from Zhou and Shen(1978)(see proof of Theorem 2.1) that $\gamma_{ks} = o(h^3)$. Substituting γ_k and $M(u)$ into (A.12), there is

$$E(\alpha_j(u)) - s_{\alpha_j}(u) = o(n^2 ph^6) \qquad (A.13)$$

The theorem proof is completed from (A.7) and (A.13).[]

Proof of Theorem 3. It follows from (12) that

$$\begin{aligned} Var(\alpha_j(u)) &= \sigma^2 M(u)(Z^T Q_X Z + \xi K)^{-1} Z^T Q_X Z (Z^T Q_X Z + \xi K)^{-1} M^T(u) \\ &:= \sigma^2 M(u) G M^T(u) \end{aligned} \qquad (A.14)$$

From(A.10), it is easily proven that

$$G = o(h^6 W_j K_0^{'-1} Z K_0^{'-1} W_j^T) := o(h^6 G^0) \qquad (A.15)$$

Let G_{st}^0 denote the (s,t) elements of G^0, then we have

$$G_{st}^0 = \sum_{k=1}^{n}\sum_{l=1}^{n} k_{sl}^* z_{jt} Q_{tk} z_{jk} k_{kt}^* = o(n^2) \qquad (A.16)$$

Substituting (A.16) and (A.15) into (A.14), we obtain

$$Var(\alpha_j(u)) = o(n^2 h^6) \qquad (A.17)$$

By Theorem 2 and (A.17), we have

$$\frac{E(\alpha_j(u))}{\sqrt{Var(\alpha_j(u))}} - \frac{b_j(u) + \alpha_j(u)}{\sqrt{Var(\alpha_j(u))}} = \frac{o(n^2 ph^6)}{o(nh^3)} = o(nph^3) = o(1) \qquad (A.18)$$

Thus, (15) follows when

$$\frac{\alpha_j(u) - E(\alpha_j(u))}{\sqrt{Var(\alpha_j(u))}} \to_d N(0,1) \qquad (A.19)$$

From (6) and (12), we have

$$\alpha_j(u) - E(\alpha_j(u)) = \sum_{i=1}^{n} a_i \varepsilon_i \qquad (A.20)$$

where $(a_1, \cdots, a_n)^T = M(u) W_j (Z^T Q_X Z + \xi K)^{-1} Z^T Q_X$. Obviously,

$$\sum_{i=1}^{k} a_i^2 = Var(\alpha_j(u)) \qquad (A.21)$$

To check the requirement of Linderberg-Feller center theorem, it is suffice to verify

$$\max\left\{a_i^2\right\} = o(Var(\alpha_j(u))) \qquad (A.22)$$

Note that $(a_1, \cdots, a_n) = \mathrm{O}(h^3\xi^{-1}M(u)K_0^{-1}Z_j Q_X)$, then

$$a_t = o(h^3\xi^{-1}M(u)K_0^{-1}Z_jQ_t), 1 \le t \le n \tag{A.23}$$

where Q_t is the t th column of Q_X. Thus,

$$a_t^2 = o(h^6\xi^{-2}M(u)K_0^{-1}Z_jQ_tQ_t^TZ_jK_0^{-1}M^T(u)) \tag{A.24}$$

Because

$$M(u)K_0^{-1}Z_jQ_tQ_t^TZ_jK_0^{-1}M^T(u) \le \lambda_{\max}(M^T(u)M(u))\lambda_{\max}(Z_jK_0^{-2}Z_j)tr(Q_tQ_t^T) \tag{A.25}$$

where $\lambda_{\max}(A)$ denote the max eigenvalue of matrix A, and

$$tr(Q_tQ_t^T) = tr(Q_t^TQ_t) = Q_{tt} \le 1, \lambda_{\max}(M^T(u)M(u)) = \lambda_{\max}\left\{M_i^2(u)\right\} \le 1 \tag{A.26}$$

where Q_{tt} denote the (t,t) elements of Q_X, we have

$$M(u)K_0^{-1}Z_jQ_tQ_t^TZ_jK_0^{-1}M^T(u) \le \lambda_{\max}(Z_jK_0^{-2}Z_j)$$

Considering

$$\lambda_{\max}(Z_jK_0^{-2}Z_j) \le \lambda_{\max}(Z_jZ_j)\lambda_{\max}K_0^{-2}) = \max\left\{z_{ji}^2\right\}\lambda_{\max}K_0^{-2}),$$

and (A.3), we have

$$\lambda_{\max}(K_0^{-2}) = O(n^2).$$

Further,

$$\max\left\{a_i^2\right\} = o(n^2h^6) = o(Var(\alpha_j(u))) \tag{A.27}$$

Hence (15) follows from (A.27) and Lindeberg-Feller center theorem.

Proof of Theorem 4. It follows from Theorem 3 that the $100(1-\alpha)\%$ asymptotic confidence for $\alpha_j(u)+b_j(u)$ is

$$\alpha_j(u) \pm z_{\alpha/2}.$$

Noting that

$$b_j(u) = \frac{\alpha_j^{(3)}(u_i)h_i^3}{3!}B_3(\frac{u-u_i}{h_i}), u \in (u_i, u_{i+1}] \tag{A.28}$$

and $\alpha_j^{(3)}$ is a unknown value, it is critical to get the estimator of $\alpha_j^{(3)}$.Obviously, when $h \to 0$,

$$\alpha_j^{(3)}(u_i) = \lim_{u \to u_i}\frac{\alpha_j^{(2)}(u)-\alpha_j^{(2)}(u_i)}{u-u_i} \approx \frac{\alpha_j^{(2)}(u_i)-\alpha_j^{(2)}(u_{i-1})}{h_i} \tag{A.29}$$

Thus, from (A.29) we can use $\alpha_j^{(2)}(u_i)$ and $\alpha_j^{(2)}(u_{i-1})$ to estimate $\alpha_j^{(3)}(u_i)$ as

$$\alpha_j^{(3)}(u_i) = \frac{\alpha_j^{(2)}(u_i)-\alpha_j^{(2)}(u_{i-1})}{h_i}, i = 2,\cdots,n-1 \tag{A.30}$$

and define $\alpha_j^{(3)}(u_1) = \alpha_j^{(3)}(u_n) = 0$. In fact, we can use $\alpha_j(U^0)$ to estimate $(\alpha_j^{(2)}(u_2),\cdots\alpha_j^{(2)}(u_{n-1}))^T$ from Green and Silverman (1994) (formula (2.4) in this book) as

$$(\alpha_j^{(2)}(u_2),\cdots,\alpha_j^{(2)}(u_{n-1}))^T = R^{-1}Q\alpha_j(U^0) \tag{A.31}$$

Denote

$$B=\begin{pmatrix} \frac{1}{h_2} & \frac{-1}{h_2} & & \\ & \ddots & \ddots & \\ & & \frac{1}{h_{n-1}} & \frac{-1}{h_{n-1}} \end{pmatrix} \tag{A.32}$$

and $C=(0_{n\times 1},Q^T R^{-1})^T$, it follows from (A.30) and (A.31) that

$$(\overset{(3)}{\alpha_j}(u_2),\cdots,\overset{(3)}{\alpha_j}(u_{n-1}))^T = BC\alpha_j(U^0) := W\alpha_j(U^0) \tag{A.33}$$

We complete the proof.[]

Proof of Theorem 5. Considering the definition of $\alpha(U^0)=(\alpha(U^0)^T,\cdots,\alpha_p(U^0)^T)^T$ in Section 2, we have $\sum_{j=1}^{p} AMSE(\alpha_j(U^0)) = AMSE(\alpha(U^0))$ .Let $N=Z^T Q_X Z$, then the expectation of $\alpha(U^0)$ is

$$E(\alpha(U^0))=(Z^T Q_X Z+\lambda K)^{-1} Z^T Q_X E(Y)=(N+\xi K)^{-1} N\alpha(U^0)\,.$$

Further,

$$Bias(\alpha(U^0))=(I-(N+\xi K)^{-1}N)\alpha(U^0)=\xi(N+\xi K)^{-1}K\alpha(U^0):=\xi M\alpha(U^0)$$

and the covariance matrix of $\alpha(U^0)$ is

$$Cov(\alpha(U^0))=\sigma^2(N+\xi K)^{-1}N(N+\xi K)^{-1}.$$

Hence, the average mean square error for $\alpha(U^0)$ satisfy

$$AMSE(\alpha(U^0))=\frac{1}{np}\alpha(U^0)^T M^T M\alpha(U^0)+\frac{\sigma^2}{np}tr((N+\xi K)^{-1}N(N+\xi K)^{-1}) \tag{A.34}$$

Notes that K is a block diagonal matrix with K_0 , *i.e.* $K=diag(K_0,\cdots K_0)$, the eigenvalues of K satisfy

$$\lambda_{i\cdot p+k}(K)=\lambda_{i+1}(K_0), i=0,\cdots,(n-1), k=1,\cdots,p \tag{A.36}$$

By Lemma 1 and $h\to 0$, we have

$$\lambda_{i\cdot p+k}(M)=\frac{c_{i+1}}{\xi c_{i+1}+o(h^3)}, i=0,\cdots,(n-1), k=1,\cdots,p\,.$$

Supposing that $f_1,\cdots,f_{np}$ are a series of orthogonalized eigenvectors of M and satisfy

$$Mf_j=\lambda_i(M)f_j, j=1,\cdots,np\,.$$

Then, there exist a vector $\gamma=(\gamma_1,\cdots,\gamma_{np})$, which satisfy

$$\alpha(U^0)=\sum_{j=1}^{np}\gamma_j f_j$$

Hence,

$$\alpha(U^0)^T M^T M\alpha(U^0)=\sum_{j=1}^{np}\gamma_j^2\lambda_j^2(M)\le c\sum_{j=1}^{np}\lambda_j^2(M)=c\cdot p\sum_{i=1}^{n}\left(\frac{c_i}{\xi c_i+o(h^3)}\right)^2 \tag{A.36}$$

where $c=\max(\gamma_1^2,\cdots,\gamma_{np}^2)$. Substituting (A.3) into (A.36), we have

$$\alpha(U^0)^T M^T M\alpha(U^0) \le c\cdot p\sum_{k=1}^{n-2}\left(\frac{k^4}{\xi k^4+\frac{1}{c_0}no(h^3)}\right)^2 \tag{A.37}$$

Let $H=(N+\xi K)^{-2}$ and it can be gotten from (A.4) and (A.35) that

$$\lambda_{i\cdot p+k}(H)=\frac{o(h^6)}{(\xi c_{i+1}+o(h^3))^2}, i=0,\cdots,(n-1), k=1,\cdots,p \tag{A.38}$$

when $h\to 0$. For $H>0$ and $N\ge 0$, there are some inequalities as follows

$$\lambda_{np}(N)\lambda_j(H)\ge\lambda_j(NH)\ge\lambda_1(N)\lambda_j(H), j=1,\cdots,np .$$

Thus,

$$0\le tr((N+\xi K)^{-1}N(N+\xi K)^{-1})=tr(NH)\le\lambda_{np}(N)\sum_{j=1}^{np}\lambda_j(H) \tag{A.39}$$

Substituting (A.4) and (A.38) into (A.39), there are

$$0\le tr((N+\xi K)^{-1}N(N+\xi K)^{-1})\le p[l_0+\sum_{k=1}^{n-2}\frac{n^2o(h^6)}{(\xi c_0k^4+no(h^3))^2}] \tag{A.40}$$

where l_0 is a constant. Using (A.37) and (A.39) into (A.40), we obtain

$$AMSE(\alpha(U^0))\le\frac{1}{n}c\sum_{k=1}^{n-2}\left(\frac{k^4}{\xi k^4+\frac{1}{c_0}no(h^3)}\right)^2+\frac{\sigma^2}{n}[l_0+\sum_{k=1}^{n-2}\frac{n^2o(h^6)}{(\xi c_0k^4+no(h^3))^2}]$$

The proof of Theorem 5 is completed.

Proof of Theorem 6 It follows from results in Theorem 1 that

$$E(\beta)=(X^TX)^{-1}X^T(I-SQ_X)E(Y)=\beta+(X^TX)^{-1}X^T(I-SQ_X)Z\alpha(U^0).$$

and

$$\begin{aligned}Cov(\beta)&=(X^TX)^{-1}X^T(I-SQ_X)Cov(Y,Y)(I-Q_XS)X(X^TX)^{-1}\\&=\sigma^2(X^TX)^{-1}X^T(I-SQ_X)(I-Q_XS)X(X^TX)^{-1}\end{aligned}$$

The proof is completed.

Acknowledgments

Guo's research was supported by the Science Foundation Of Huaihai Institute of Technology (Z2014008) and the Teaching Reform Project of Jiangsu Province(2015JSJG517). Zhao's research was supported by the National Social Science Fund (15CTJ008).

References

[1] I. Ahmad, S. Leelahanon and Q. Li, "Efficient estimation of a semiparameteric partially linear varying coefficient model", Ann. Statist., vol. 33, **(2005)**, pp. 258-283.
[2] T. Bi, B. Zhang and R. Xu, "Dynamics of intraday serial correlation in China's stock market", Communication in Statistics - Simulation and Computation, vol. 40, **(2011)**, pp. 1637-1650.
[3] G. G. Barrow, P. W. Smith, "Asymptotic properties of the best L2 [0; 1] approximation by spline with variable knots", Quart. Appl. Math, vol. 36, **(1980)**, pp. 293-304.
[4] C.T. Chiang, J.A. Rice and C.O. Wu, "Smoothing spline estimation for varying coefficient models with repeatedly measured dependent variables", J. Amer. Statist. Assoc., vol.96, **(2001)**, pp. 605-619.
[6] A. Demmler and C. Reinch, "Oscilliation matrices with spline smoothing", Numer. Math., vol. 24, **(1975)**,pp.375-382.
[7] J. Fan and T. Huang, "Profile likelihood inferences on parametric regression models", Bernoulli, vol. 11, **(2005)**, pp. 1031-1057.
[8] J. Fan, W. Zhang, "Statistical Estimation In Varying Coefficient Models", Ann. Statist., vol. 33, **(1999)**, pp. 258-7283.

[9] J. Fan and W. Zhang, "Simultaneous confidence bands and hypothesis testing in varying-coefficient models", Scand. J. Statist., vol. 27, **(2000)**, pp. 715-731.
[10] P.J. Green and B.W. Silverman, "Nonparametric regression generalized linear model", Chapman and Hall, London, **(1994)**.
[11] T.J. Hastie and R.J. Tibshirani, "Generalized additive models", Chapman and Hall, London, **(1990)**.
[12] T.J. Hastie and R.J. Tibshriani, "Varying-Coefficient Model", J. Roy. Statist. Soc. B, vol. 55, **(1993)**, pp. 757-796.
[13] W. Härdle, H. Liang and J. Guo, "Partially linear models", Physica-Verlag, Heidelberg, **(2004)**.
[14] C.M. Hurvich, J .S. Simonoff and C.L. Tsai, "Smoothing parameter selection in nonparametric regression using an improved Akaike information criterion", J. Roy. Statist. Soc. B., vol.60, **(1998)**, pp. 271-293.
[15] K. Gäran, "Penalized Spline Smoothing in Multivariable Survival Modelswith Varying Coefficients", Computational Statistics & Data Analysis, vol. 49, **(2005)**, pp. 169-86.
[16] R. Kohn, C.F. Ansley and D. Tharm, "The performance of cross-validation and maximum likelihood estimators of spline smoothing parameters", J. Amer. Statist. Assoc., vol. 86, **(1991)**, pp. 1042-1050.
[17] X. Ni, D.Zhang and H. Zhang, "Variable selection for semi-parametric mixed models in longitudinal studies", Biometrics, vol. 66, **(2010)**, pp. 79-88.
[18] P. Speckman, "Spline smoothing and optimal rates of convergence in nonparametric models", Ann. Statist., vol. 13, **(1985)**, pp. 970-983.
[19] P. Speckman and D. Sun, "Asymptotic properties of smoothing parameter selection in spline smoothing", Manuscript, University of Missouri, Columbia, **(2000)**.
[20] Y. Xia, W. Zhang and H. Tong, "Efficient estimation for semivarying coefficient models", Biometrika, vol. 91, **(2004)**, pp. 661-681.

Authors

Haibing Guo, he received his Ph.D degree in statistics and is currently a lecturer in Huaihai Institute of Technology.. His research interest is mainly in the area of semi-parametric model and data mining. In the past years, his research focus on model.

