



Research article

B-spline estimation in varying coefficient models with correlated errors

Yanping Liu and Juliang Yin*

School of Economics and Statistics, Guangzhou University, Guangzhou 510006, China

* **Correspondence:** Email: yin_juliang@hotmail.com.

Abstract: The varying coefficient model assumes that the regression function depends linearly on some regressors, and that the regression coefficients are smooth functions of other predictor variables. It provides an appreciable flexibility in capturing the underlying dynamics in data and avoids the so-called curse of dimensionality in analyzing complex and multivariate nonlinear structures. Existing estimation methods usually assume that the errors for the model are independent; however, they may not be satisfied in practice. In this study, we investigated the estimation for the varying coefficient model with correlated errors via B-spline. The B-spline approach, as a global smoothing method, is computationally efficient. Under suitable conditions, the convergence rates of the proposed estimators were obtained. Furthermore, two simulation examples were employed to demonstrate the performance of the proposed approach and the necessity of considering correlated errors.

Keywords: varying coefficient models; nonlinear time series; B-spline estimation; consistency; generalized least squares method

Mathematics Subject Classification: 62G08, 62M10, 65D07

1. Introduction

Nonlinear phenomena, namely nonnormality, asymmetric cycles, and nonlinear relationships between lagged variables, have been well observed in some classical data sets, such as the sunspot, Canadian lynx, and Australian blowfly data. However, linear ARMA models can not adequately approximate these nonlinear phenomena [1]. Due to this, nonparametric regression models have found important applications in modeling nonlinear time series [2, 3]. Yet, in the multivariate setting with more than two variables, it is difficult to estimate regression function with reasonable accuracy due to the “curse of dimensionality”. To solve this problem, various semiparametric models have been studied. For example, Stone [4] proposed the additive model, and Gao [5] investigated a class of partially linear models. In addition, one of the most popular semiparametric models is the varying coefficient model [6], whose regression function depends linearly on some regressors and regression

coefficients vary with some threshold variables.

In this study, we consider the varying coefficient model of the following form:

$$Y_t = \mathbf{X}_t^\top \boldsymbol{\beta}(U_t) + \epsilon_t \quad (1.1)$$

where $\mathbf{X}_t = (X_{t1}, \dots, X_{tp})^\top$ and $\boldsymbol{\beta}(U_t) = (\beta_1(U_t), \dots, \beta_p(U_t))^\top$. The functions $\beta_j(\cdot)$, $j = 1, \dots, p$ are assumed to be unknown but smooth. \top denotes the transpose of a vector or matrix. U_t is a univariate random variable, which is called the threshold variable. Both \mathbf{X}_t and U_t can consist of either exogenous variables or lagged variables of Y_t . ϵ_t is the error term that satisfies $\mathbb{E}(\epsilon_t | \mathbf{X}_t, U_t) = 0$, and $\boldsymbol{\epsilon} = (\epsilon_1, \dots, \epsilon_n)^\top$. As a generalization of the linear model, the Model (1.1) has attracted a great deal of attention over the past two decades. When \mathbf{X}_t and U_t are some lagged variables of Y_t , Chen and Tsay [7] proposed an arranged local regression procedure for the specification of the Model (1.1), and the consistency result and a recursive algorithm were given. Cai et al. [1] applied a local linear technique to estimate the time-varying coefficients. The asymptotic properties of the kernel estimators were investigated under the α -mixing condition. Nevertheless, as pointed out in [8, 9], the local smoothing method is computationally expensive because it requires re-fitting at every point where the fitted function needs to be evaluated. In contrast, the advantage of the B-spline estimation method is its computational efficiency. However, it is difficult to establish the asymptotic normality of the spline estimators [10, 11]. For varying coefficient models, Lai et al. [12] considered the B-spline to estimate the time-varying coefficients.

The aforementioned work depends on the assumption of independent errors. However, model misspecification, such as omitting relevant variables and wrong function form, may result in correlated errors, as mentioned in [13]. From this perspective, the assumption of independent errors is inappropriate. Many authors have studied the topic of non-/semi-parametric regression with correlated errors. Under the assumption that model errors follow an invertible linear process, Xiao et al. [14] proposed a modification of the local polynomial estimation for nonparametric regression. Su and Ullah [13] considered nonparametric regression with an error process in a nonparametric autocorrelated form. Lei et al. [15] investigated a semiparametric autoregressive model with ARMA errors. In this study, we propose a global smoothing method based on B-spline for the estimation of the time-varying coefficients. Similar to [16] and [17], we have relaxed the assumption of independence to model errors, assuming that they can be correlated. That is, when the model errors are independent, the covariance matrix of the errors is $\mathbb{E}(\boldsymbol{\epsilon}\boldsymbol{\epsilon}^\top) = \sigma^2 \mathbf{I}$. Now, we assume that $\mathbb{E}(\boldsymbol{\epsilon}\boldsymbol{\epsilon}^\top | \mathbf{X}_t, U_t) = \mathbf{V}$. Specially, if model errors follow an AR(1) process, $\epsilon_t = \rho\epsilon_{t-1} + e_t$, $|\rho| < 1$, $e_t \stackrel{\text{i.i.d.}}{\sim} (0, \sigma^2)$,

$$\mathbf{V} = \frac{\sigma^2}{1 - \rho^2} \begin{pmatrix} 1 & \rho & \dots & \rho^{n-2} & \rho^{n-1} \\ \rho & 1 & \dots & \rho^{n-3} & \rho^{n-2} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \rho^{n-2} & \rho^{n-3} & \dots & 1 & \rho \\ \rho^{n-1} & \rho^{n-2} & \dots & \rho & 1 \end{pmatrix}.$$

Our study assumes that the covariance matrix of model errors is positive definite, without requiring a specific form of autoregression, in which case, the scope of application can be further expanded. Certainly, we extend Theorem 1 of [12] to the case of correlated errors.

The remainder of this paper has been presented as follows. In Section 2, the estimation procedure for the varying coefficient model with correlated errors has been introduced. Section 3 presents the consistency and convergence rates of the spline estimators and provides an estimation algorithm. In Section 4, we present numerical examples to illustrate the performance of the proposed estimation method. In Section 5, we compare the results of the proposed spline method with the local linear method proposed by Cai et al. [1]. The conclusions are presented in Section 6 and the proofs of the main results in the Appendix.

2. Estimation method

Let $a = \xi_0 < \xi_1 < \dots < \xi_{M+1} = b$ partition the interval $[a, b]$ into subintervals $[\xi_k, \xi_{k+1}), k = 0, \dots, M$ with M internal knots. A polynomial B spline of order r is a function whose restriction to each subinterval is a polynomial of degree $r - 1$ and globally $r - 2$ times continuously differentiable. A linear space $S_{r,M}$ with a fixed sequence of knots has a normalized B-spline basis $\{B_1(u), \dots, B_K(u)\}$ with $K = M + r$. As in [18], the basis satisfies (i) $B_k(u) \geq 0, k = 1, \dots, K$, (ii) $\sum_{s=1}^K B_k(u) \equiv 1$, and (iii) $B_j(u)$ is supported inside an interval of length r/K , and at most r of the basis functions are nonzero at any given u .

Suppose that each coefficient function $\beta_j(u)$ in the model (1.1) is smooth; then, it can be well approximated by a B-spline function $\beta_j^*(u) \in S_{r,M}$ [19]. Thus, there is a set of constants $b_{js}^*, s = 1, \dots, K$, such that $\beta_j(u) \approx \beta_j^*(u) = \sum_{s=1}^K b_{js}^* B_s(u)$. Different coefficients might be approximated by B-spline with a different number of knots in principle, but for simplicity, we have assumed the same basis for the different coefficients. Let $\mathbf{b} = (\mathbf{b}_1^\top, \dots, \mathbf{b}_p^\top)^\top = (b_{11}, \dots, b_{1K}, \dots, b_{p1}, \dots, b_{pK})^\top$, $\mathbf{Z}_t = (X_{t1}B_1(U_t), \dots, X_{t1}B_K(U_t), \dots, X_{tp}B_K(U_t))^\top$, $\mathbf{Z} = (\mathbf{Z}_1, \dots, \mathbf{Z}_n)^\top$, $\mathbf{Y} = (Y_1, \dots, Y_n)^\top$. Let $\mathbb{E}(\boldsymbol{\epsilon}\boldsymbol{\epsilon}^\top | \mathbf{X}_t, U_t) = \mathbf{V}$, and it is initially known. Subsequently, we can write the criterion as

$$Q(\mathbf{b}) = \frac{1}{n}(\mathbf{Y} - \mathbf{Z}\mathbf{b})^\top \mathbf{V}^{-1}(\mathbf{Y} - \mathbf{Z}\mathbf{b}). \quad (2.1)$$

Denoting the minimizer by $\hat{\mathbf{b}}$, we estimate $\beta_j(u)$ by $\hat{\beta}_j(u) = \sum_{k=1}^K \hat{b}_{jk} B_k(u)$.

3. Asymptotic property

We have imposed the following conditions. (C1) and (C2) are mild regularity conditions. (C3) is necessary for the identification of the coefficient functions in varying coefficient models, as mentioned in Huang and Shen [20]. (C4), (C6), and (C7) are the same as those used in [1] for stationary mixing data. (C5) imposes some smoothness conditions on the coefficient functions.

3.1. Assumptions

(C1) The eigenvalues of \mathbf{V} are bounded away from zero and infinity.

(C2) The smoothing variable U_t has a bounded density supported on $[a, b]$.

(C3) The eigenvalues of the matrix $\mathbb{E}(\mathbf{X}_t \mathbf{X}_t^\top | U_t = u)$ are uniformly bounded away from zero and infinity for all $u \in [a, b]$.

- (C4) The conditional density of (U_1, U_{l+1}) given $(\mathbf{X}_1, \mathbf{X}_{l+1})$ is uniformly bounded on the support of $(\mathbf{X}_1, \mathbf{X}_{l+1})$. The conditional density of U_1 given \mathbf{X}_1 is uniformly bounded on the support of \mathbf{X}_1 .
- (C5) For $g = \beta_j$, $1 \leq j \leq p$, g satisfies a Lipschitz condition of order $d > 1/2$: $|g^{(l,d)}(t) - g^{(l,d)}(s)| \leq C|s - t|^{d-[d]}$, where $[d]$ is the biggest integer strictly smaller than d and $g^{(l,d)}$ is the $[d]$ -th derivative of g . The order of the B-spline used satisfies $q \geq d + 1/2$.
- (C6) The process $\{Y_t, \mathbf{X}_t, U_t\}_{t \in \mathbb{Z}}$ is jointly strictly stationary with U_t taking values in \mathbb{R} and \mathbf{X}_t taking values in \mathbb{R}^p . The α -mixing coefficient $\alpha(l)$ of $\{Y_t, \mathbf{X}_t, Z_t\}_{t \in \mathbb{Z}}$ satisfies $\sum_{l=1}^{\infty} l^c \alpha(l)^{1-2/\delta} < \infty$ for some $\delta > 2$ and $c > 1 - 2/\delta$.
- (C7) $\mathbb{E}(|X_{tj}|^{2\delta}) < \infty$, $j = 1, \dots, d$, where δ is given in condition (C6).

3.2. Theoretical result

Theorem 3.1: Assume (C1)–(C7) and that $K \rightarrow \infty$, $K^3/n \rightarrow 0$, then we have

$$\|\hat{\beta}_j - \beta_j\|_2^2 = O_p(K/n + 1/K^{2d}), \quad 1 \leq j \leq p.$$

As noted in Lai et al. [12], we know that by choosing $K \asymp n^{1/(2d+1)}$, the well-known optimal convergence rate $O_p(n^{-2d/(2d+1)})$ is achieved.

If \mathbf{V} is unknown and $\hat{\mathbf{V}}$ is an estimator of \mathbf{V} , then an application of the substitution principle leads to

$$Q(\mathbf{b}) = \frac{1}{n}(\mathbf{Y} - \mathbf{Z}\mathbf{b})^\top \hat{\mathbf{V}}^{-1}(\mathbf{Y} - \mathbf{Z}\mathbf{b}). \quad (3.1)$$

A two-stage estimator $\bar{\mathbf{b}}$ can be obtained by minimizing (3.1); then, $\bar{\beta}_j(u) = \sum_{k=1}^K \bar{b}_{jk} B_k(u)$.

Theorem 3.2: Assume (C1)–(C7), $K \rightarrow \infty$, $K^3/n \rightarrow 0$ and that $\hat{\mathbf{V}}$ is consistent in probability estimating \mathbf{V} , then we have

$$\|\bar{\beta}_j - \beta_j\|_2^2 = O_p(K/n + 1/K^{2d}), \quad 1 \leq j \leq p.$$

Remark 1. In practice, \mathbf{V} is usually unknown. It is customary to replace \mathbf{V} in (2.1) with its consistent estimator $\hat{\mathbf{V}}$ [17, 21]. Then, $\bar{\mathbf{b}}$ can be derived through a two-stage estimation. If the sample size is sufficiently large, we can split the data set into the training and test parts, and provide a consistent estimator of \mathbf{V} from the residuals of ordinary least squares (OLS) for fitting (1.1) by grouping the training set.

3.3. Computational aspects

However, the method in Remark 1 is not feasible when the sample size is small. It should be noted that Montoril et al. [17] have presented an iterative procedure for estimating \mathbf{V} , but the iterative procedure is computationally expensive. Hence, they have provided a more efficient method. Generally, if the errors of the model follow an autoregressive process, we can estimate the $\beta_j(\cdot)$ by adopting some ideas of [17]. The estimation algorithm is as follows.

Step 1. Estimate the coefficient vector \mathbf{b} by OLS, and denote it by $\mathbf{b}^{(0)}$.

Step 2. Compute the residuals via $\epsilon_t^{(0)} = Y_t - \mathbf{Z}_t \mathbf{b}^{(0)}$, and fit an autoregressive model to the residuals, i.e.,

$$\epsilon_t^{(0)} = \varphi_1^{(0)} \epsilon_{t-1}^{(0)} + \cdots + \varphi_p^{(0)} \epsilon_{t-p}^{(0)} + e_t.$$

Step 3. Letting $\mathbf{b}^{(0)}$ and $(\varphi_1^{(0)} \cdots \varphi_p^{(0)})$ in step 1 and step 2 as initial values, estimate \mathbf{b} by minimizing numerically

$$\ell(\mathbf{b}) = \sum_{t=1}^n \{\varphi_p(L)(Y_t - \mathbf{Z}_t \mathbf{b})\}^2,$$

where $\varphi_p(L) = 1 - \varphi_1 L - \cdots - \varphi_p L^p$, with the backshift operator satisfying $L^k V_t = V_{t-k}$, $k > 0$.

4. Numerical examples

In this section, two simulated examples are considered: the threshold variable U_t is an exogenous variable and the lagged variable of Y_t . For a given data set, we used equally spaced knots. The values used as candidates for the number of internal knots varied from one to five. The optimal number of internal knots is selected using the Bayesian information criterion (BIC) (Schwarz [22]). The BIC criterion function is defined as

$$\text{BIC} = \log(\text{RMS}) + \log(n) \times \frac{p}{n},$$

where n denotes the sample size, RMS denotes the residual mean square, p is equal to the sum of number of autoregressive coefficients assumed for the errors and number of B-spline basis. The B-spline basis $\{B_1(U_t), \cdots, B_K(U_t)\}$ can be obtained from function **bs** using the package **splines** in R language [3]. We consider time series data with lengths $n = 200, 400$ and 600 , and replicate the simulation 200 times in each case. For each replication, a total of $1000+n$ observations were generated, and only the last n observations were used to ensure approximate stationarity. The performance of estimators $\{\hat{\beta}_j(\cdot)\}$ can be demonstrated by the square root of average squared errors (RASE):

$$\text{RASE}^2 = \sum_{k=1}^p \text{RASE}_j^2,$$

with

$$\text{RASE}_j = \{n_{\text{grid}}^{-1} \sum_{k=1}^{n_{\text{grid}}} [\hat{\beta}_j(u_k) - \beta_j(u_k)]^2\}^{1/2},$$

and $\{u_k, k = 1, \cdots, n_{\text{grid}}\}$ are grid points on an interval over which the functions are evaluated. Because the range of the time series data varies from simulation to simulation, we need to select a common interval to compare the RASE values. The intervals selected for Example 4.1 and Example 4.2 are $[-0.45, 0.45]$ and $[-2, 2]$, respectively.

Example 4.1. Consider the following data generating process

$$Y_t = \beta_1(U_t)Y_{t-1} + \beta_2(U_t)Y_{t-2} + \epsilon_t,$$

with $\beta_1(u) = 0.9 \sin(\pi u)$ and $\beta_2(u) = 0.85 \cos(\pi u)$. We study two autoregressive errors, i.e. AR(1): $\epsilon_t = 0.8\epsilon_{t-1} + e_t$ and AR(2): $\epsilon_t = 0.5\epsilon_{t-1} + 0.45\epsilon_{t-2} + e_t$, where $e_t \stackrel{\text{i.i.d.}}{\sim} N(0, 0.2^2)$ and $U_t \stackrel{\text{i.i.d.}}{\sim} U[-0.5, 0.5]$.

Example 4.2. We now discuss an exponential autoregressive (EXPAR) model

$$Y_t = \beta_1(Y_{t-1})Y_{t-1} + \beta_2(Y_{t-1})Y_{t-2} + \epsilon_t,$$

with $\beta_1(u) = 0.6 + 0.9e^{(-u^2)}$ and $\beta_2(u) = -0.4 - 1.2e^{(-u^2)}$. In this case, the autoregressive errors we study are AR(1): $\epsilon_t = 0.9\epsilon_{t-1} + e_t$ and AR(2): $\epsilon_t = 0.6\epsilon_{t-1} + 0.35\epsilon_{t-2} + e_t$, where $e_t \stackrel{\text{i.i.d.}}{\sim} N(0, 0.2^2)$.

5. Numerical results and discussion

In this section, the numerical results for Examples 4.1 and 4.2 are presented. We compare the performance of local linear (Local) estimators proposed by Cai et al. [1], the spline estimators under the assumption of independent errors (Spl.ind) and spline estimators (Spl.cor) proposed by us. Tables 1 and 2 show the mean and standard deviation (in parentheses) of the RASEs for $\hat{\beta}_1(\cdot)$ and $\hat{\beta}_2(\cdot)$ with linear ($k = 1$) and cubic ($k = 3$) splines under different AR(p) errors. It is apparent that the standard deviation of the RASEs in columns Spl.cor decrease with the increase in the sample size n . In addition, the results of linear splines are similar to those of the cubic splines. Moreover, it can be seen that the proposed approach performs better than the method that ignores the correlated errors. Based on cubic splines, Tables 3 and 4 provide the resulting estimates while considering different AR (1) errors: $\epsilon_t = \theta\epsilon_{t-1} + e_t$ with $\theta = 0.3, 0.6, 0.9, 0.95$ when $n = 600$. We can see that with the increase in the correlation levels θ , the performance of Local and Spl.ind worsens, but Spl.cor remains relatively stable. Furthermore, Spl.cor is always better than Spl.ind and Local, which emphasizes the importance of considering the correlated errors. The computational times of Example 4.2 are reported in Table 5. The local linear method is computationally expensive when the sample size is large. Moreover, when considering the correlated errors, the algorithm needs to search for the optimal solution iteratively. Therefore, the calculation time of Spl.cor is longer than that of Spl.ind.

Table 1. Simulation results of Example 4.1 for difference sample sizes.

degree	AR(p)	sample size	Spl.ind		Spl.cor	
			$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_1$	$\hat{\beta}_2$
$k = 1$	1	$n = 200$	0.1554(0.0366)	0.1028(0.0297)	0.0694(0.0320)	0.0635(0.0269)
		$n = 400$	0.1380(0.0286)	0.0937(0.0239)	0.0667(0.0311)	0.0606(0.0265)
		$n = 600$	0.1335(0.0240)	0.0885(0.0193)	0.0606(0.0292)	0.0563(0.0245)
	2	$n = 200$	0.1484(0.0382)	0.1366(0.0355)	0.0624(0.0315)	0.0633(0.0318)
		$n = 400$	0.1340(0.0261)	0.1258(0.0270)	0.0626(0.0313)	0.0630(0.0312)
		$n = 600$	0.1289(0.0224)	0.1236(0.0226)	0.0624(0.0310)	0.0629(0.0309)
$k = 3$	1	$n = 200$	0.1513(0.0372)	0.1015(0.0296)	0.0664(0.0330)	0.0595(0.0279)
		$n = 400$	0.1340(0.0293)	0.0903(0.0249)	0.0592(0.0277)	0.0516(0.0254)
		$n = 600$	0.1301(0.0240)	0.0857(0.0211)	0.0530(0.0270)	0.0479(0.0225)
	2	$n = 200$	0.1448(0.0385)	0.1366(0.0365)	0.0724(0.0342)	0.0718(0.0331)
		$n = 400$	0.1309(0.0264)	0.1245(0.0277)	0.0654(0.0321)	0.0638(0.0325)
		$n = 600$	0.1264(0.0229)	0.1217(0.0236)	0.0631(0.0270)	0.0619(0.0282)

Table 2. Simulation results of Example 4.2 for difference sample sizes.

degree	AR(p)	sample size	Spl.ind		Spl.cor	
			$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_1$	$\hat{\beta}_2$
$k = 1$	1	$n = 200$	0.1452(0.0683)	0.1095(0.0299)	0.0700(0.0307)	0.0658(0.0199)
		$n = 400$	0.1049(0.0422)	0.1022(0.0248)	0.0615(0.0239)	0.0634(0.0192)
		$n = 600$	0.0894(0.0337)	0.0972(0.0209)	0.0553(0.0208)	0.0599(0.0183)
	2	$n = 200$	0.2242(0.1250)	0.1307(0.0413)	0.0743(0.0391)	0.0648(0.0227)
		$n = 400$	0.1610(0.0807)	0.1305(0.0375)	0.0618(0.0284)	0.0591(0.0180)
		$n = 600$	0.1397(0.0691)	0.1274(0.0330)	0.0608(0.0282)	0.0587(0.0166)
$k = 3$	1	$n = 200$	0.1420(0.0528)	0.1132(0.0290)	0.0647(0.0252)	0.0503(0.0165)
		$n = 400$	0.1108(0.0411)	0.1021(0.0264)	0.0560(0.0220)	0.0431(0.0152)
		$n = 600$	0.0883(0.0309)	0.0964(0.0196)	0.0501(0.0198)	0.0371(0.0146)
	2	$n = 200$	0.2020(0.0988)	0.1349(0.0437)	0.0763(0.0282)	0.0531(0.0196)
		$n = 400$	0.1584(0.0660)	0.1347(0.0373)	0.0585(0.0239)	0.0430(0.0147)
		$n = 600$	0.1402(0.0567)	0.1291(0.0311)	0.0538(0.0213)	0.0389(0.0136)

Table 3. Simulation results of Example 4.1 for difference AR(1) error structures.

θ	Local		Spl.ind		Spl.cor	
	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_1$	$\hat{\beta}_2$
0.3	0.0799(0.0207)	0.0447(0.0156)	0.0760(0.0206)	0.0400(0.0145)	0.0443(0.0155)	0.0400(0.0147)
0.6	0.1221(0.0229)	0.0474(0.0171)	0.1187(0.0226)	0.0566(0.0188)	0.0442(0.0214)	0.0392(0.0178)
0.9	0.1353(0.0226)	0.0752(0.0228)	0.1316(0.0238)	0.1064(0.0220)	0.0641(0.0297)	0.0585(0.0267)
0.95	0.1387(0.0229)	0.0971(0.0314)	0.1317(0.0234)	0.1181(0.0233)	0.0739(0.0306)	0.0681(0.0282)

Table 4. Simulation results of Example 4.2 for difference AR(1) error structures.

θ	Local		Spl.ind		Spl.cor	
	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_1$	$\hat{\beta}_2$
0.3	0.0679(0.0266)	0.0373(0.0168)	0.0376(0.0148)	0.0309(0.0106)	0.0359(0.0139)	0.0287(0.0104)
0.6	0.0691(0.0281)	0.0469(0.0132)	0.0455(0.0161)	0.0434(0.0129)	0.0393(0.0154)	0.0325(0.0114)
0.9	0.1427(0.0678)	0.0888(0.0207)	0.0883(0.0309)	0.0964(0.0196)	0.0501(0.0198)	0.0371(0.0146)
0.95	0.2008(0.0977)	0.1504(0.0416)	0.1435(0.0510)	0.1508(0.0318)	0.0563(0.0221)	0.0456(0.0153)

Table 5. Computation time for the spline estimators and the local linear estimators in seconds based on Example 4.2.

Sample size	Spl.ind	Spl.cor	Local
$n = 200$	4.4	85.7	57.3
$n = 400$	5.0	94.5	120.1
$n = 600$	5.5	102.4	208.6

The estimated $\beta_1(\cdot)$ and $\beta_2(\cdot)$ from a typical sample using cubic splines for Examples 4.1 and 4.2 with AR(1) error structures are plotted in Figures 1 to 4. The solid curve is the true curve, and the dotted curve represents the typical estimated curve. The typical sample is chosen in such a way that its RASE value is equal to the median in the 200 simulations [23]. Clearly, the proposed estimators capture the main features of the true functions well. Although it is not shown here, the examples with AR(2) error structures show similar results.

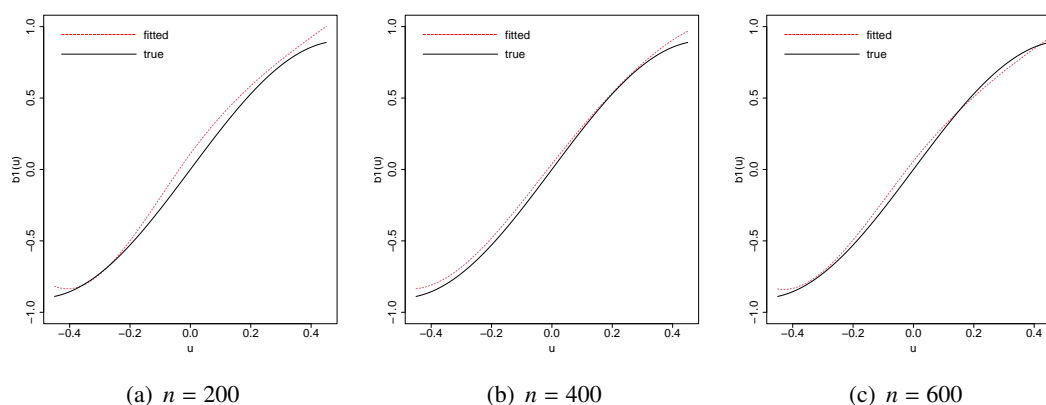


Figure 1. Typical estimated curves of β_1 in Example 4.1.

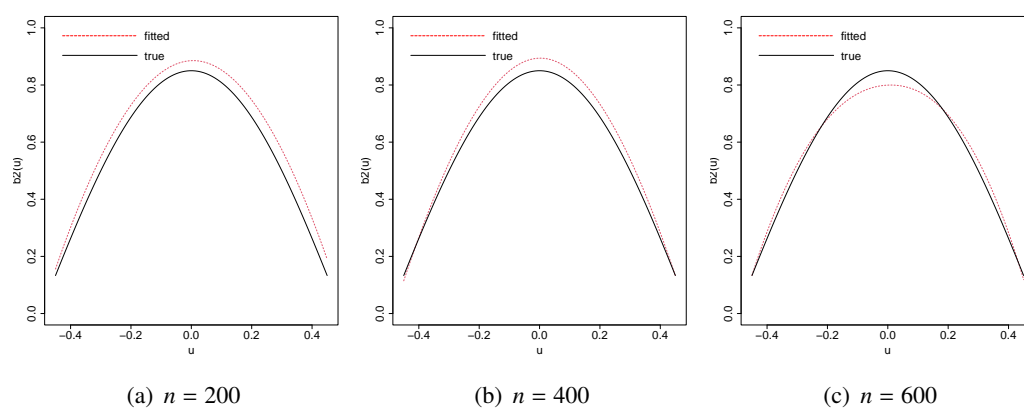


Figure 2. Typical estimated curves of β_2 in Example 4.1.

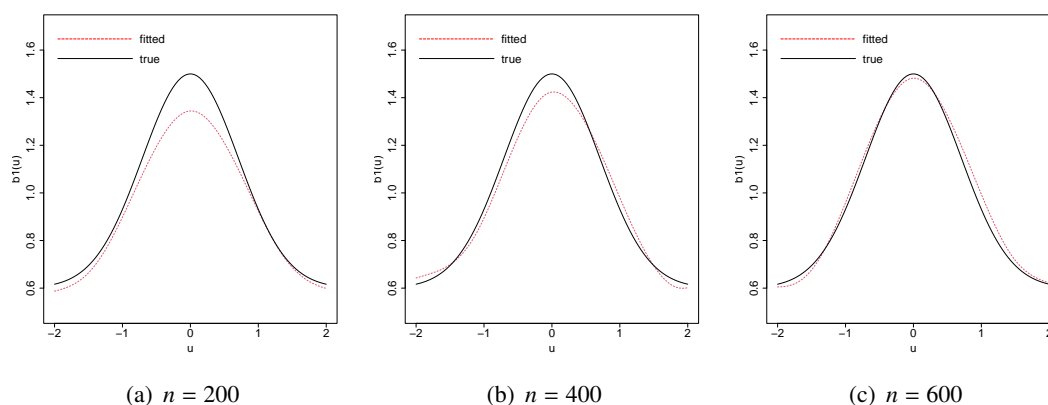


Figure 3. Typical estimated curves of β_1 in Example 4.2.

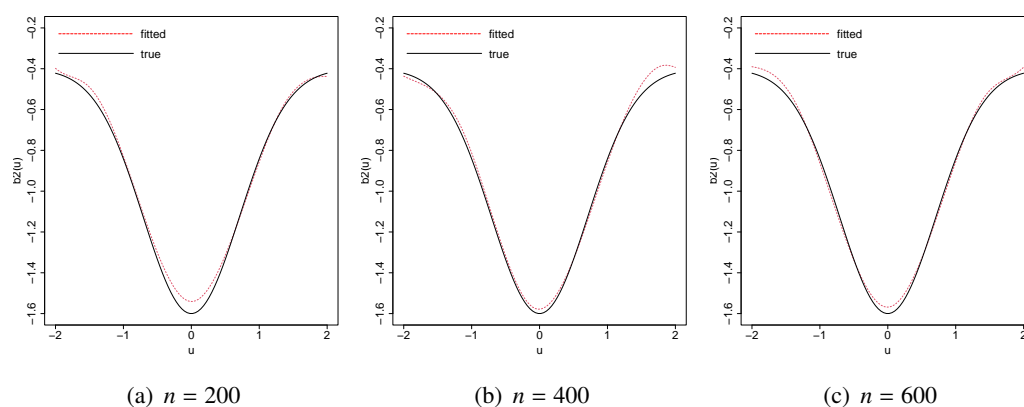


Figure 4. Typical estimated curves of β_2 in Example 4.2.

6. Conclusions

This study considered the B-spline estimation for varying coefficient models with correlated errors using a weighted least squares method. In terms of the theoretical results, convergence rates were derived under the α -mixing condition. Simulation studies were conducted to illustrate the performance of the proposed estimation method which showed better performance while considering correlated errors in fitting varying coefficient models.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under Grant No. 61973096 and by GDUPS (2019).

Conflict of interest

The authors declare that they have no any competing interests.

Appendix

A. Proof of the main results

Notation: For two positive sequences a_n and b_n , $a_n \lesssim b_n$ means that a_n/b_n is uniformly bounded, $a_n \asymp b_n$ if $a_n \lesssim b_n$ and $b_n \lesssim a_n$. Let $\|g\|_2 = \{\int_{[a,b]} g^2(x)dx\}^{1/2}$ be the L_2 -norm of a square integrable function $g(\cdot)$ on $[a, b]$. \mathbb{R} denotes the set of real numbers, and \mathbb{Z} denotes the set of integers. $|\cdot|$ denotes the Euclidean norm of a vector. \mathbf{I} is a identity matrix. \mathbf{e}_j denotes a unit vector whose j -th entry is 1 and all other entries of which are 0. We denote by $\lambda_{\min}(\cdot)$ ($\lambda_{\max}(\cdot)$) the smallest (the largest) eigenvalue of a matrix. Let C denote a generic constant that might assume different values at different places. Given random variables $V_n, n \geq 1$, let $V_n = O_p(b_n)$ means that the random variables $V_n/b_n, n \geq 1$ are bounded in probability, that is,

$$\lim_{M \rightarrow \infty} \lim_{n \rightarrow \infty} \mathbb{P}(|V_n| > Mb_n) = 0.$$

And $V_n = o_p(b_n)$ means that the random variables $V_n/b_n, n \geq 1$ converge to zero in probability, namely

$$\lim_{n \rightarrow \infty} \mathbb{P}(|V_n| > \epsilon b_n) = 0, \text{ for } \forall \epsilon > 0.$$

In order to proof the theorems, we need the following Lemmas.

Lemma 1. (Lemma 2, Lai et al. [12]) Under assumptions (C2) and (C3), there are constants $0 < b_1 < b_2 < \infty$ such that the eigenvalues of $\mathbb{E}\mathbf{Z}_1\mathbf{Z}_1^\top$ fall in $[\frac{b_1}{K}, \frac{b_2}{K}]$, and under the additional conditions (C4)–(C7), there are constants $0 < b_3 < b_4 < \infty$ such that the eigenvalues of $\sum_{t=1}^n \mathbf{Z}_t\mathbf{Z}_t^\top/n$ fall in $[\frac{b_3}{K}, \frac{b_4}{K}]$ with probability approaching 1 as $n \rightarrow \infty$.

Lemma 2. Under the conditions (C1)–(C7), we have $\boldsymbol{\epsilon}^\top \mathbf{V}^{-1} \mathbf{Z} (\mathbf{Z}^\top \mathbf{V}^{-1} \mathbf{Z})^{-1} \mathbf{Z}^\top \mathbf{V}^{-1} \boldsymbol{\epsilon} = O_p(K)$.

Proof. Note that $\mathbb{E}(\mathbf{Z}^\top \mathbf{Z}) = \mathbb{E}(\sum_{t=1}^n \mathbf{Z}_t\mathbf{Z}_t^\top) = n\mathbb{E}(\mathbf{Z}_1\mathbf{Z}_1^\top)$. Hence, by Lemma 1 we obtain

$$\begin{aligned} \mathbb{E}(\boldsymbol{\epsilon}^\top \mathbf{V}^{-1} \mathbf{Z} \mathbf{Z}^\top \mathbf{V}^{-1} \boldsymbol{\epsilon}) &= \mathbb{E}\text{tr}(\mathbf{Z}^\top \mathbf{V}^{-1} \boldsymbol{\epsilon} \boldsymbol{\epsilon}^\top \mathbf{V}^{-1} \mathbf{Z}) \\ &= \text{tr}\mathbb{E}(\mathbf{Z}^\top \mathbf{V}^{-1} \mathbb{E}(\boldsymbol{\epsilon} \boldsymbol{\epsilon}^\top | \mathbf{Z}) \mathbf{V}^{-1} \mathbf{Z}) \\ &= \text{tr}\mathbb{E}(\mathbf{Z}^\top \mathbf{V}^{-1} \mathbf{Z}) \\ &\lesssim \text{tr}\mathbb{E}(\mathbf{Z}^\top \mathbf{Z}) \\ &\lesssim n. \end{aligned} \tag{A.1}$$

Using the Markov inequality, we can derive

$$\|\boldsymbol{\epsilon}^\top \mathbf{V}^{-1} \mathbf{Z}\|^2 = O_p(n). \tag{A.2}$$

On the other hand, by Lemma 1 and (C1),

$$\begin{aligned} &\lim_{M \rightarrow \infty} \lim_{n \rightarrow \infty} \mathbb{P}(\lambda_{\max}(\mathbf{Z}^\top \mathbf{V}^{-1} \mathbf{Z})^{-1} < M \frac{K}{n}) \\ &= \lim_{M \rightarrow \infty} \lim_{n \rightarrow \infty} \mathbb{P}\left(\frac{1}{\lambda_{\min}(\mathbf{Z}^\top \mathbf{V}^{-1} \mathbf{Z})} < M \frac{K}{n}\right) \end{aligned}$$

$$\begin{aligned}
&\geq \lim_{M \rightarrow \infty} \lim_{n \rightarrow \infty} \mathbb{P}\left(\frac{1}{\lambda_{\min}(\mathbf{V}^{-1})\lambda_{\min}(\mathbf{Z}^{\top}\mathbf{Z})} < M\frac{K}{n}\right) \\
&\geq \lim_{M \rightarrow \infty} \lim_{n \rightarrow \infty} \mathbb{P}\left(\frac{1}{\lambda_{\min}(\mathbf{Z}^{\top}\mathbf{Z})} < CM\frac{K}{n}\right) \\
&\geq \lim_{M \rightarrow \infty} \lim_{n \rightarrow \infty} \mathbb{P}\left(\lambda_{\min}\left(\frac{\mathbf{Z}^{\top}\mathbf{Z}}{n}\right) > \frac{C}{MK}\right) \\
&\geq \lim_{n \rightarrow \infty} \mathbb{P}\left(\lambda_{\min}\left(\frac{\mathbf{Z}^{\top}\mathbf{Z}}{n}\right) > \frac{b_3}{K}\right) \\
&= 1.
\end{aligned}$$

Consequently,

$$\lambda_{\max}(\mathbf{Z}^{\top}\mathbf{V}^{-1}\mathbf{Z})^{-1} = O_p\left(\frac{K}{n}\right). \quad (\text{A.3})$$

Then, it follows from (A.2) and (A.3) that

$$\begin{aligned}
&\lim_{M \rightarrow \infty} \lim_{n \rightarrow \infty} \mathbb{P}(\boldsymbol{\epsilon}^{\top}\mathbf{V}^{-1}\mathbf{Z}(\mathbf{Z}^{\top}\mathbf{V}^{-1}\mathbf{Z})^{-1}\mathbf{Z}^{\top}\mathbf{V}^{-1}\boldsymbol{\epsilon} \leq KM) \\
&\geq \lim_{M \rightarrow \infty} \lim_{n \rightarrow \infty} \mathbb{P}(\lambda_{\max}(\mathbf{Z}^{\top}\mathbf{V}^{-1}\mathbf{Z})^{-1}\|\boldsymbol{\epsilon}^{\top}\mathbf{V}^{-1}\mathbf{Z}\|^2 \leq KM) \\
&= \lim_{M \rightarrow \infty} \lim_{n \rightarrow \infty} \mathbb{P}(\lambda_{\max}(\mathbf{Z}^{\top}\mathbf{V}^{-1}\mathbf{Z})^{-1}\|\boldsymbol{\epsilon}^{\top}\mathbf{V}^{-1}\mathbf{Z}\|^2 \leq KM, \|\boldsymbol{\epsilon}^{\top}\mathbf{V}^{-1}\mathbf{Z}\|^2 \leq \sqrt{Mn}) \\
&\geq \lim_{M \rightarrow \infty} \lim_{n \rightarrow \infty} \mathbb{P}(\lambda_{\max}(\mathbf{Z}^{\top}\mathbf{V}^{-1}\mathbf{Z})^{-1}\sqrt{Mn} \leq KM, \|\boldsymbol{\epsilon}^{\top}\mathbf{V}^{-1}\mathbf{Z}\|^2 \leq \sqrt{Mn}) \\
&= \lim_{M \rightarrow \infty} \lim_{n \rightarrow \infty} \mathbb{P}(\lambda_{\max}(\mathbf{Z}^{\top}\mathbf{V}^{-1}\mathbf{Z})^{-1} \leq \frac{K}{n}\sqrt{M}, \|\boldsymbol{\epsilon}^{\top}\mathbf{V}^{-1}\mathbf{Z}\|^2 \leq \sqrt{Mn}) \\
&= \lim_{M \rightarrow \infty} \lim_{n \rightarrow \infty} \mathbb{P}(\lambda_{\max}(\mathbf{Z}^{\top}\mathbf{V}^{-1}\mathbf{Z})^{-1} \leq \frac{K}{n}\sqrt{M}) \\
&= 1.
\end{aligned}$$

This completes the proof of Lemma 2. \square

Lemma 3. Suppose $\beta_{0j}(u) = \sum_{k=1}^K b_{0jk}B_k(u)$ is the best approximating B-spline for $\beta_j(u)$ with $\|\beta_{0j} - \beta_j\|_{\infty} = O(K^{-d})$ (such approximation property is well-known for B-spline under smoothness assumptions (C5), see for example [19]). Let $\mathbf{b}_0 = (b_{011}, \dots, b_{01K}, \dots, b_{0p1}, \dots, b_{0pK})^{\top}$, $\boldsymbol{\eta} = \mathbf{P}_{\mathbf{V}^{-\frac{1}{2}}\mathbf{Z}}\mathbf{V}^{-\frac{1}{2}}(\mathbf{Y} - \mathbf{Z}\mathbf{b}_0)$, where $\mathbf{P}_{\mathbf{V}^{-\frac{1}{2}}\mathbf{Z}} = \mathbf{V}^{-\frac{1}{2}}\mathbf{Z}(\mathbf{Z}^{\top}\mathbf{V}^{-1}\mathbf{Z})^{-1}\mathbf{Z}^{\top}\mathbf{V}^{-\frac{1}{2}}$ is a projection matrix. Then under (C1)–(C7) we have

$$\|\boldsymbol{\eta}\|^2 = O_p\left(K + \frac{n}{K^{2d}}\right).$$

Proof. Denote $r_t = \sum_{j=1}^p X_{tj}\beta_j(U_t)$ and $\mathbf{r} = (r_1, \dots, r_n)^{\top}$, thus $\mathbf{Y} = \mathbf{r} + \boldsymbol{\epsilon}$. We have

$$\begin{aligned}
\|\boldsymbol{\eta}\|^2 &= \|\mathbf{P}_{\mathbf{V}^{-\frac{1}{2}}\mathbf{Z}}\mathbf{V}^{-\frac{1}{2}}(\mathbf{Y} - \mathbf{Z}\mathbf{b}_0)\|^2 \\
&= \|\mathbf{P}_{\mathbf{V}^{-\frac{1}{2}}\mathbf{Z}}\mathbf{V}^{-\frac{1}{2}}\boldsymbol{\epsilon} + \mathbf{P}_{\mathbf{V}^{-\frac{1}{2}}\mathbf{Z}}\mathbf{V}^{-\frac{1}{2}}(\mathbf{r} - \mathbf{Z}\mathbf{b}_0)\|^2 \\
&\leq 2\|\mathbf{P}_{\mathbf{V}^{-\frac{1}{2}}\mathbf{Z}}\mathbf{V}^{-\frac{1}{2}}\boldsymbol{\epsilon}\|^2 + 2\|\mathbf{P}_{\mathbf{V}^{-\frac{1}{2}}\mathbf{Z}}\mathbf{V}^{-\frac{1}{2}}(\mathbf{r} - \mathbf{Z}\mathbf{b}_0)\|^2
\end{aligned}$$

$$\triangleq 2S_1 + 2S_2.$$

Noting that $\mathbf{P}_{\mathbf{V}^{-\frac{1}{2}}\mathbf{Z}}$ is an idempotent matrix, by Lemma 2 we obtain that

$$\begin{aligned} S_1 &= \boldsymbol{\epsilon}^\top \mathbf{V}^{-\frac{1}{2}} \mathbf{P}_{\mathbf{V}^{-\frac{1}{2}}\mathbf{Z}} \mathbf{V}^{-\frac{1}{2}} \boldsymbol{\epsilon} \\ &= \boldsymbol{\epsilon}^\top \mathbf{V}^{-1} \mathbf{Z} (\mathbf{Z}^\top \mathbf{V}^{-1} \mathbf{Z})^{-1} \mathbf{Z}^\top \mathbf{V}^{-1} \boldsymbol{\epsilon} \\ &= O_p(K). \end{aligned}$$

By the approximation property of B-spline we have $\|\mathbf{r} - \mathbf{Z}\mathbf{b}_0\|^2 = O_p(\frac{n}{K^{2d}})$, hence

$$\begin{aligned} &\lim_{M \rightarrow \infty} \lim_{n \rightarrow \infty} \mathbb{P}(\|\mathbf{P}_{\mathbf{V}^{-\frac{1}{2}}\mathbf{Z}} \mathbf{V}^{-\frac{1}{2}} (\mathbf{r} - \mathbf{Z}\mathbf{b}_0)\|^2 \leq M \frac{n}{K^{2d}}) \\ &= \lim_{M \rightarrow \infty} \lim_{n \rightarrow \infty} \mathbb{P}((\mathbf{r} - \mathbf{Z}\mathbf{b}_0)^\top \mathbf{V}^{-\frac{1}{2}} \mathbf{P}_{\mathbf{V}^{-\frac{1}{2}}\mathbf{Z}} \mathbf{V}^{-\frac{1}{2}} (\mathbf{r} - \mathbf{Z}\mathbf{b}_0) \leq M \frac{n}{K^{2d}}) \\ &\geq \lim_{M \rightarrow \infty} \lim_{n \rightarrow \infty} \mathbb{P}(\lambda_{\max}(\mathbf{V}^{-\frac{1}{2}} \mathbf{P}_{\mathbf{V}^{-\frac{1}{2}}\mathbf{Z}} \mathbf{V}^{-\frac{1}{2}}) \|\mathbf{r} - \mathbf{Z}\mathbf{b}_0\|^2 \leq M \frac{n}{K^{2d}}) \\ &\geq \lim_{M \rightarrow \infty} \lim_{n \rightarrow \infty} \mathbb{P}(\|\mathbf{r} - \mathbf{Z}\mathbf{b}_0\|^2 \leq CM \frac{n}{K^{2d}}) \\ &= 1. \end{aligned}$$

Thus, $S_2 = O_p(\frac{n}{K^{2d}})$, the proof is complete. □

Proof of Theorem 3.1.

Proof. By the definition of $\hat{\mathbf{b}}$, we get

$$\begin{aligned} 0 &\geq nQ(\hat{\mathbf{b}}) - nQ(\mathbf{b}_0) \\ &= (\mathbf{Y} - \mathbf{Z}\hat{\mathbf{b}})^\top \mathbf{V}^{-1} (\mathbf{Y} - \mathbf{Z}\hat{\mathbf{b}}) - (\mathbf{Y} - \mathbf{Z}\mathbf{b}_0)^\top \mathbf{V}^{-1} (\mathbf{Y} - \mathbf{Z}\mathbf{b}_0) \\ &= 2(\mathbf{Y} - \mathbf{Z}\mathbf{b}_0)^\top \mathbf{V}^{-1} \mathbf{Z}(\mathbf{b}_0 - \hat{\mathbf{b}}) + [\mathbf{Z}(\mathbf{b}_0 - \hat{\mathbf{b}})]^\top \mathbf{V}^{-1} \mathbf{Z}(\mathbf{b}_0 - \hat{\mathbf{b}}) \\ &= 2[\mathbf{P}_{\mathbf{V}^{-\frac{1}{2}}\mathbf{Z}} \mathbf{V}^{-\frac{1}{2}} (\mathbf{Y} - \mathbf{Z}\mathbf{b}_0)]^\top \mathbf{V}^{-\frac{1}{2}} \mathbf{Z}(\mathbf{b}_0 - \hat{\mathbf{b}}) + \|\mathbf{V}^{-\frac{1}{2}} \mathbf{Z}(\mathbf{b}_0 - \hat{\mathbf{b}})\|^2. \end{aligned} \tag{A.4}$$

Applying the Cauchy-Schwartz inequality, the equation (A.4) can be continued as

$$0 \geq -4\|\mathbf{P}_{\mathbf{V}^{-\frac{1}{2}}\mathbf{Z}} \mathbf{V}^{-\frac{1}{2}} (\mathbf{Y} - \mathbf{Z}\mathbf{b}_0)\|^2 + \|\mathbf{V}^{-\frac{1}{2}} \mathbf{Z}(\mathbf{b}_0 - \hat{\mathbf{b}})\|^2.$$

Hence, by Lemma 3 we obtain

$$\|\mathbf{V}^{-\frac{1}{2}} \mathbf{Z}(\mathbf{b}_0 - \hat{\mathbf{b}})\|^2 = O_p(K + \frac{n}{K^{2d}}).$$

Then, it follows from Lemma 1 and (C1) that

$$\begin{aligned} 0 &= \lim_{M \rightarrow \infty} \lim_{n \rightarrow \infty} \mathbb{P}(\|\mathbf{V}^{-\frac{1}{2}} \mathbf{Z}(\mathbf{b}_0 - \hat{\mathbf{b}})\|^2 > M(K + \frac{n}{K^{2d}})) \\ &\geq \lim_{M \rightarrow \infty} \lim_{n \rightarrow \infty} \mathbb{P}(\lambda_{\min}(\mathbf{Z}^\top \mathbf{V}^{-1} \mathbf{Z}) \|\mathbf{b}_0 - \hat{\mathbf{b}}\|^2 > M(K + \frac{n}{K^{2d}})) \end{aligned}$$

$$\begin{aligned}
&\geq \lim_{M \rightarrow \infty} \lim_{n \rightarrow \infty} \mathbb{P}(\lambda_{\min}(V^{-1})\lambda_{\min}(Z^T Z)\|\mathbf{b}_0 - \hat{\mathbf{b}}\|^2 > M(K + \frac{n}{K^{2d}})) \\
&\geq \lim_{M \rightarrow \infty} \lim_{n \rightarrow \infty} \mathbb{P}(\lambda_{\min}(\frac{Z^T Z}{n})\|\mathbf{b}_0 - \hat{\mathbf{b}}\|^2 > CM(\frac{K}{n} + \frac{1}{K^{2d}})) \\
&= \lim_{M \rightarrow \infty} \lim_{n \rightarrow \infty} \mathbb{P}(\lambda_{\min}(\frac{Z^T Z}{n})\|\mathbf{b}_0 - \hat{\mathbf{b}}\|^2 > CM(\frac{K}{n} + \frac{1}{K^{2d}}), \lambda_{\min}(\frac{Z^T Z}{n}) > \frac{b_3}{K}) \\
&\geq \lim_{M \rightarrow \infty} \lim_{n \rightarrow \infty} \mathbb{P}(\|\mathbf{b}_0 - \hat{\mathbf{b}}\|^2 > CM(\frac{K^2}{n} + \frac{1}{K^{2d-1}})),
\end{aligned}$$

which means that $\|\mathbf{b}_0 - \hat{\mathbf{b}}\|^2 = O_p(\frac{K^2}{n} + \frac{1}{K^{2d-1}})$. Then, by the property of B-spline (De Boor [19])

$$\frac{b_5}{K}\|\hat{\mathbf{b}}_j - \mathbf{b}_{0j}\|^2 \leq \|\hat{\beta}_j - \beta_{0j}\|_2^2 \leq \frac{b_6}{K}\|\hat{\mathbf{b}}_j - \mathbf{b}_{0j}\|^2, \quad (\text{A.5})$$

for some constants $b_5, b_6 > 0$, we can derive

$$\|\hat{\beta}_j - \beta_j\|_2^2 \lesssim \frac{1}{K}\|\hat{\mathbf{b}}_j - \mathbf{b}_{0j}\|^2 + \|\beta_j - \beta_{0j}\|_\infty^2 = O_p(\frac{K}{n} + \frac{1}{K^{2d}}).$$

□

Proof of Theorem 3.2.

Proof. The proof of this theorem is similar to that of Theorem 2.2 in [17].

Let $\mathbf{R} = \mathbf{Z}^T \mathbf{V}^{-1} \mathbf{Z}$ and $\hat{\mathbf{R}} = \mathbf{Z}^T \hat{\mathbf{V}}^{-1} \mathbf{Z}$. Note that $\hat{\mathbf{V}}$ is consistent, then by the Theorem 1.4.2 in [24] we have

$$\lambda_{\max}(\mathbf{R}\hat{\mathbf{R}}^{-1}) \leq \lambda_{\max}(\mathbf{V}^{-1}\hat{\mathbf{V}}) = 1 + o_p(1).$$

This implies that $\hat{\mathbf{R}}$ is a consistent estimator of \mathbf{R} , namely $\hat{\mathbf{R}}^{-1}\mathbf{R} - \mathbf{I} = o_p(1)$. It follows from (2.1) that

$$\hat{\mathbf{b}} = (\mathbf{Z}^T \mathbf{V}^{-1} \mathbf{Z})^{-1} \mathbf{Z}^T \mathbf{V}^{-1} \mathbf{Y} = \mathbf{R}^{-1} \mathbf{Z}^T \mathbf{V}^{-1} \mathbf{Y}.$$

Therefore, by (3.1) we can derive that

$$\begin{aligned}
\bar{\mathbf{b}} &= (\mathbf{Z}^T \hat{\mathbf{V}}^{-1} \mathbf{Z})^{-1} \mathbf{Z}^T \hat{\mathbf{V}}^{-1} \mathbf{Y} \\
&= (\hat{\mathbf{R}}^{-1} - \mathbf{R}^{-1})\mathbf{Z}^T \hat{\mathbf{V}}^{-1} \mathbf{Y} + \mathbf{R}^{-1}\mathbf{Z}^T \hat{\mathbf{V}}^{-1} \mathbf{Y} \\
&= (\hat{\mathbf{R}}^{-1}\mathbf{R} - \mathbf{I})\mathbf{R}^{-1}\mathbf{Z}^T \hat{\mathbf{V}}^{-1} \mathbf{Y} + \mathbf{R}^{-1}\mathbf{Z}^T (\hat{\mathbf{V}}^{-1}\mathbf{V} - \mathbf{I})\mathbf{V}^{-1} \mathbf{Y} + \hat{\mathbf{b}} \\
&= (\hat{\mathbf{R}}^{-1}\mathbf{R} - \mathbf{I})\mathbf{R}^{-1}\mathbf{Z}^T (\hat{\mathbf{V}}^{-1}\mathbf{V} - \mathbf{I})\mathbf{V}^{-1} \mathbf{Y} + (\hat{\mathbf{R}}^{-1}\mathbf{R} - \mathbf{I})\mathbf{R}^{-1}\mathbf{Z}^T \mathbf{V}^{-1} \mathbf{Y} \\
&\quad + \mathbf{R}^{-1}\mathbf{Z}^T (\hat{\mathbf{V}}^{-1}\mathbf{V} - \mathbf{I})\mathbf{V}^{-1} \mathbf{Y} + \hat{\mathbf{b}} \\
&= o_p(1)\hat{\mathbf{b}} + \hat{\mathbf{b}}.
\end{aligned}$$

This ensures that

$$\|\bar{\mathbf{b}} - \hat{\mathbf{b}}\|^2 = o_p(1). \quad (\text{A.6})$$

By (A.5), (A.6) and Theorem 3.1, we conclude that

$$\|\bar{\beta}_j - \beta_j\|_2^2 \leq \|\bar{\beta}_j - \hat{\beta}_j\|_2^2 + \|\hat{\beta}_j - \beta_j\|_2^2$$

$$\begin{aligned} &\lesssim \|\bar{\mathbf{b}} - \hat{\mathbf{b}}\|^2 + \|\hat{\beta}_j - \beta_j\|_2^2 \\ &= O_p\left(\frac{K}{n} + \frac{1}{K^{2d}}\right). \end{aligned}$$

This completes the proof of Theorem 3.2. □

References

1. Z. Cai, J. Fan, Q. Yao, Functional-coefficient regression models for nonlinear time series, *J. Am. Stat. Assoc.*, **95** (2000), 941–956. doi: 10.1080/01621459.2000.10474284.
2. W. Härdle, H. Liang, J. Gao, *Partially linear models*, Heidelberg: Physica-Verlag, 2000. doi: 10.1007/978-3-642-57700-0.
3. R. S. Tsay, R. Chen, *Nonlinear time series analysis*, Hoboken: John Wiley & Sons, 2018. doi: 10.1002/9781119514312.
4. C. J. Stone, Additive regression and other nonparametric models, *Ann. Stat.*, **13** (1985), 689–705. doi: 10.1214/aos/1176349548.
5. J. Gao, *Nonlinear time series: Semiparametric and nonparametric methods*, London: Chapman & Hall, 2007. doi: 10.1201/9781420011210.
6. T. Hastie, R. Tibshirani, Varying coefficient models, *J. R. Stat. Soc. B*, **55** (1993), 757–779. doi: 10.1111/j.2517-6161.1993.tb01939.x.
7. R. Chen, R. S. Tsay, Functional-coefficient autoregressive models, *J. Am. Stat. Assoc.*, **88** (1993), 298–308. doi: 10.1080/01621459.1993.10594322.
8. X. Wu, Z. Tian, H. Wang, Polynomial spline estimation for nonparametric (auto-) regressive models, *Stud. Sci. Math. Hung.*, **46** (2009), 515–538. doi: 10.1556/sscmath.2009.1105.
9. L. Xue, H. Liang, Polynomial spline estimation for a generalized additive coefficient model, *Scand. J. Stat.*, **37** (2010), 26–46. doi: 10.1111/j.1467-9469.2009.00655.x.
10. Y. Lu, R. Zhang, L. Zhu, Penalized spline estimation for varying-coefficient models, *Commun. Stat. Theor. M.*, **37** (2008), 2249–2261. doi: 10.1080/03610920801931887.
11. S. Ma, L. Yang, Spline-backfitted kernel smoothing of partially linear additive model, *J. Stat. Plan. Infer.*, **141** (2011), 204–219. doi: 10.1016/j.jspi.2010.05.028.
12. P. Lai, J. Meng, H. Lian, Polynomial spline approach for variable selection and estimation in varying coefficient models for time series data, *Stat. Probabil. Lett.*, **96** (2015), 21–27. doi: 10.1016/j.spl.2014.09.008.
13. L. Su, A. Ullah, More efficient estimation in nonparametric regression with nonparametric autocorrelated errors, *Economet. Theor.*, **22** (2006), 98–126. doi: 10.1017/S026646660606004X.
14. Z. Xiao, O. B. Linton, R. J. Carroll, E. Mammen, More efficient local polynomial estimation in nonparametric regression with autocorrelated errors, *J. Am. Stat. Assoc.*, **98** (2003), 980–992. doi: 10.1198/016214503000000936.
15. H. Lei, Y. Xia, X. Qin, Estimation of semivarying coefficient time series models with ARMA errors, *Ann. Stat.*, **44** (2016), 1618–1660. doi: 10.1214/15-AOS1430.

16. M. H. Montoril, P. A. Morettin, C. Chiann, Spline estimation of functional coefficient regression models for time series with correlated errors, *Stat. Probabil. Lett.*, **92** (2014), 226–231. doi: 10.1016/j.spl.2014.05.021.
17. M. H. Montoril, P. A. Morettin, C. Chiann, Wavelet estimation of functional coefficient regression models, *Int. J. Wavelets Multi.*, **16** (2018), 1850004. doi: 10.1142/S0219691318500042.
18. H. Lian, P. Lai, H. Liang, Partially linear structure selection in Cox models with varying coefficients, *Biometrics*, **69** (2013), 348–357. doi: 10.1111/biom.12024.
19. C. De Boor, *A practical guide to splines*, New York: Springer-Verlag, 1978. doi: 10.2307/2006241.
20. J. Z. Huang, H. Shen, Functional coefficient regression models for non-linear time series: A polynomial spline approach, *Scand. J. Stat.*, **31** (2004), 515–534. doi: 10.1111/j.1467-9469.2004.00404.x.
21. J. Shao, *Mathematical statistics*, New York: Springer-Verlag, 2003. doi: 10.1007/b97553.
22. G. Schwarz, Estimating the dimension of a model, *Ann. Stat.*, **6** (1978), 461–464. doi: 10.1214/aos/1176344136.
23. J. Fan, Q. Yao, *Nonlinear time series: Nonparametric and parametric methods*, New York: Springer-Verlag, 2003. doi: 10.1007/978-0-387-69395-8.
24. P. K. Sen, J. M. Singer, *Large sample methods in statistics: An introduction with applications*, New York: Chapman & Hall, 1994. doi: 10.1201/9780203711606.



©2022 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>)