

# Partial Time-Varying Coefficient Regression and Autoregressive Mixed Model

Hui Li<sup>1</sup>, Zhiqiang Cao<sup>2\*</sup>

<sup>1</sup>School of Statistics, Beijing Normal University, Beijing, China

<sup>2</sup>Department of Mathematics, College of Big Data and Internet, Shenzhen Technology University, Shenzhen, China

Email: \*caozhiqiang@sztu.edu.cn

**How to cite this paper:** Li, H. and Cao, Z.Q. (2023) Partial Time-Varying Coefficient Regression and Autoregressive Mixed Model. *Open Journal of Statistics*, 13, 514-533.

<https://doi.org/10.4236/ojs.2023.134026>

**Received:** July 7, 2023

**Accepted:** August 8, 2023

**Published:** August 11, 2023

Copyright © 2023 by author(s) and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## Abstract

Regression and autoregressive mixed models are classical models used to analyze the relationship between time series response variable and other covariates. The coefficients in traditional regression and autoregressive mixed models are constants. However, for complicated data, the coefficients of covariates may change with time. In this article, we propose a kind of partial time-varying coefficient regression and autoregressive mixed model and obtain the local weighted least-square estimators of coefficient functions by the local polynomial technique. The asymptotic normality properties of estimators are derived under regularity conditions, and simulation studies are conducted to empirically examine the finite-sample performances of the proposed estimators. Finally, we use real data about Lake Shasta inflow to illustrate the application of the proposed model.

## Keywords

Regression and Autoregressive, Time Series, Partial Time-Varying Coefficient, Local Polynomial

## 1. Introduction

Regression and autoregressive mixed models are widely used tools to study the correlation between random indicators of time series types and their influencing factors. The expression of the model is

$$Y_i = \alpha_0 + \alpha_1 X_{1i} + \cdots + \alpha_s X_{si} + \beta_1 Y_{i-1} + \cdots + \beta_p Y_{i-p} + \varepsilon_i, \quad (1)$$

where time series  $\{Y_i\}$ ,  $i = p+1, \dots, n$ , is the dependent variable; time series  $\{X_{ji}\}$ ,  $j = 1, \dots, s$ , represents  $s$  covariates; with  $\alpha_j$  as their regression coefficients;  $\alpha_0$  is the intercept term;  $\beta_j$ ,  $j = 1, \dots, p$ , represents  $p$  autoregressive

coefficients and satisfies stationary conditions; and  $\{\varepsilon_i\}$  is a white noise series, which is independent of both the covariates  $\{X_{ji}\}$  and response variable  $\{Y_i\}$ . Model (1) has been studied extensively in classical textbooks [1] [2]. Model (1) can also be regarded as a special case of an autoregressive model with exogenous variables or an autoregressive distributed lag model [3], which has a wide range of applications in finance, econometrics and biomedicine.

The unknown parameters in model (1) are assumed to be constant, indicating that the effects between the dependent variable and the covariates will not change with the observation time. However, in practice, complex nonlinear correlations between dependent variable and covariates are sometimes encountered. The varying-coefficient (or functional coefficient) time series model is an effective tool to handle data with such characteristics. The so-called varying-coefficient is that the parameters in the model are not constants but functions of some variables such as observation time or some delay components of time series. If the expression of the coefficient function in the model is known, it is called a parametric varying-coefficient model. Otherwise, it is a nonparametric model. It is well known that a parametric model will fit better than a nonparametric model if the coefficient function is known. However, if we make a mistake in the expression of the coefficient function, then it will cause serious deviations in the estimation and even result in misleading outcomes. In this case, the nonparametric model has more advantages. Nonparametric varying-coefficient models in time series were studied extensively, see [4]-[16]. In addition, there is a long history of models in the Bayesian realm to track the issue, where the time-variability of coefficients is (semi-) automatically shrunk to constancy in various ways [17] [18] [19].

It is noted that many scholars have carried out research on model extension for the time-varying coefficient model (*i.e.*, coefficients are functions of time) proposed by Robinson [4], see [20]-[29]. Since the model in Robinson [4] is the basis of many studies and the model proposed in this article is an extension of the time-varying coefficient model of Robinson [4] and Cai [9], we first give the model expression:

$$Y_i = \beta_0(t_i) + \beta_1(t_i)X_{1i} + \cdots + \beta_d(t_i)X_{di} + \varepsilon_i, \quad i = 1, \dots, n, \quad (2)$$

where time series  $\{Y_i\}$  is the dependent variable; time series  $\{X_{ji}\}$ ,  $j = 1, \dots, d$ , represent  $d$  covariates;  $\beta_j(\cdot)$ ,  $j = 0, \dots, d$ , are  $d + 1$  coefficient functions of observation time whose form are completely unknown and satisfy a certain degree of smoothness; and  $\{\varepsilon_i\}$  is an i.i.d (independent and identical distribution) series and is independent of  $\{X_{ji}\}$ . It should be noted that the independent variable of the coefficient function is  $t_i = i/n$  instead of  $i$ . This is a constraint that we need to satisfy when deducing large-sample properties of nonparametric smooth estimators of coefficients in a time-varying coefficient model, and the specific reasons can be seen in [4] [5] [9].

To avoid the “curse of dimensionality”, many researchers have studied nonparametric estimation methods for varying-coefficient models. For example,

Robinson [4] proposed a pseudo-local maximum likelihood estimation method to study the local estimates of functional coefficients in the time-varying coefficient time series model (2). Chen and Tsay [6] and Hastie and Tibshirani [30] proposed a kernel smooth nonparametric estimation method for functional coefficients in autoregressive models. Ramsay and Silverman [31] and Brumback and Rice [32] estimated the parameters of the varying-coefficient regression model by using the B-spline method. Cai *et al.* [8] studied local polynomial estimates for functional-coefficient regression models. Huang *et al.* [33] proposed a B-spline estimation for a varying-coefficient regression model based on repeated measures data. Aneiros-Perez and Vieu [34] used Nadaraya-Watson-type weights to estimate the semi-functional partial linear regression model (SFPLR model) and derived the corresponding asymptotic properties of the estimators. Li *et al.* [20] and Fan *et al.* [22] applied local polynomial expansion techniques to explore parameter estimation for time series models with partial time dependencies. Liu *et al.* [10] used a local linear approach to estimate the nonparametric trend and seasonal effect functions. Cai *et al.* [11] adopted a nonparametric generalized method of moments to estimate a new class of semiparametric dynamic panel data models. Chen *et al.* [23] applied a local linear method with cross-validation bandwidth selection to estimate the time-varying coefficients of the heterogeneous autoregressive model. By combining B-spline smoothing and the local linear method, Hu *et al.* [27] proposed a two-step estimation method for a time-varying additive model. Tu and Wang [14] applied adaptive kernel weighted least squares to estimate functional coefficient cointegration models. Li *et al.* [28] used classical kernel smoothing methods to estimate the coefficient functions in nonlinear cointegrating models. Karmakar *et al.* [29] applied local linear M-estimation to estimate the time-varying coefficients of a general class of nonstationary time series, among others.

For time series data with complex correlations between sample components, sometimes neither the constant coefficient regression and autoregressive mixed model (1) nor the time-varying coefficient regression model (2) can fit the data well. Motivated by this, in this article, we introduce the idea of a time-varying coefficient into a simple mixed model (1) and propose a partial time-varying coefficient regression and autoregressive mixed model. The proposed model can not only estimate the constant effects of some covariates on the dependent variable but also characterize the non-constant effects of other covariates, which greatly increases the flexibility and scope of the models (1) and (2). To the best of our knowledge, explicitly introducing part of the delay terms of the dependent variable into time-varying regression models as covariates is less actively studied. However, for time series with a strong correlation between components, sometimes introducing the delay term of the dependent variable as part of covariates can make full use of the data information and improve the model fitting degree. We will study this problem from the way of Cai [9] in this article, although the Bayesian parallel stream can be incorporated and put into perspective.

The remaining parts of this article are organized as follows. In Section 2, we introduce a partial time-varying coefficient regression and autoregressive mixed model, give the estimation method of model parameters and derive the large sample properties of the proposed estimators. We conduct a simulation study in Section 3 to examine the finite-sample performances of the proposed estimators. In Section 4, we use Shasta Lake inflow data to illustrate the application value of the model. Finally, we conclude with a discussion in Section 5.

## 2. Model and Estimation

### 2.1. Proposed Model

As mentioned in Section 1, many researchers have studied different types of varying-coefficient regression models, but research on partial time-varying coefficient regression and autoregressive mixed models is less actively studied. To fill this gap and study the correlation effects of dependent variables with covariates and delay components of partial dependent variables, we propose the following partial time-varying coefficient regression and autoregressive mixed model:

$$Y_i = \boldsymbol{\alpha}_1^T(t_i) \mathbf{X}_{i,s_1} + \boldsymbol{\alpha}_2^T \mathbf{X}_{i,s_2} + \boldsymbol{\beta}_1^T(t_i) \mathbf{Y}_{i,p_1} + \boldsymbol{\beta}_2^T \mathbf{Y}_{i,p_2} + \varepsilon_i, \quad (3)$$

where  $\boldsymbol{\alpha}_1(\cdot) = \{\alpha_{1,1}(\cdot), \dots, \alpha_{1,s_1}(\cdot)\}^T$  with the superscript T as transposition,

$$t_i = i/n \quad \text{with} \quad i = p_1 + p_2 + 1, \dots, n, \quad \boldsymbol{\alpha}_2 = (\alpha_{2,1}, \dots, \alpha_{2,s_2})^T,$$

$$\boldsymbol{\beta}_1(\cdot) = \{\beta_{1,1}(\cdot), \dots, \beta_{1,p_1}(\cdot)\}^T, \quad \boldsymbol{\beta}_2 = (\beta_{2,1}, \dots, \beta_{2,p_2})^T, \quad \mathbf{X}_{i,s_1} = (X_{i,1}, \dots, X_{i,s_1})^T,$$

$$\mathbf{X}_{i,s_2} = (X_{i,(s_1+1)}, \dots, X_{i,(s_1+s_2)})^T, \quad \mathbf{Y}_{i,p_1} = (Y_{i-i_1}, \dots, Y_{i-i_{p_1}})^T,$$

$\mathbf{Y}_{i,p_2} = (Y_{i-i_{p_1+1}}, \dots, Y_{i-i_{p_1+p_2}})^T$  with  $i_1, \dots, i_{p_1+p_2}$  as a permutation of time indicator  $1, \dots, p_1 + p_2$ . For the convenience of derivation, in this article, we set  $i_1, \dots, i_{p_1+p_2}$  to be the order from 1 to  $(p_1 + p_2)$ ;  $s = s_1 + s_2$  and  $p = p_1 + p_2$  represent the number of regression covariates and the order of autoregression, respectively; and  $\{\varepsilon_i\}$  is assumed to be a white noise series. Model (3) is an iterative formula for the sequence  $\{Y_i\}$ . To make the time series  $\{Y_i\}$  have a solution, certain constraints on the model parameters should be made, such as the autoregressive coefficient satisfies the stationary condition and  $E(\varepsilon_i Y_j) = 0$  when  $j < i$ . Since the autoregressive coefficients in model (3) are not necessarily constant, the corresponding constraints are stricter; for example, the time series in the model satisfies the  $\alpha$ -mixing condition. We will give the conditions that need to be satisfied in Section 2.3.

Model (3) enhances the generalizability of time series regression models and contains many existing statistical models as specials. For example, when  $\boldsymbol{\alpha}_1(\cdot) = \mathbf{0}$ ,  $\boldsymbol{\beta}_1(\cdot) = \mathbf{0}$ ,  $\boldsymbol{\beta}_2 = \mathbf{0}$ , model (3) is the traditional linear regression model; when  $\boldsymbol{\alpha}_2 = \mathbf{0}$ ,  $\boldsymbol{\beta}_1(\cdot) = \mathbf{0}$ ,  $\boldsymbol{\beta}_2 = \mathbf{0}$ , model (3) reduces to the time-varying coefficient linear regression model; when  $\boldsymbol{\beta}_1(\cdot) = \mathbf{0}$ ,  $\boldsymbol{\beta}_2 = \mathbf{0}$ , model (3) is a partial time-varying coefficient linear regression model; when  $\boldsymbol{\alpha}_1(\cdot) = \mathbf{0}$ ,  $\boldsymbol{\alpha}_2 = \mathbf{0}$ ,  $\boldsymbol{\beta}_1(\cdot) = \mathbf{0}$ , model (3) is the autoregressive time series model; when  $\boldsymbol{\alpha}_1(\cdot) = \mathbf{0}$ ,

$\alpha_2 = \mathbf{0}$ ,  $\beta_2 = \mathbf{0}$ , model (3) is the time-varying coefficient autoregressive time series model; when  $\alpha_1(\cdot) = \mathbf{0}$ ,  $\beta_1(\cdot) = \mathbf{0}$ , model (3) becomes the traditional constant coefficient regression and autoregressive mixed model.

Model (3) is a semi-parametric model since some of its parameters are unknown functions, and we need to apply non-parametric estimation methods, such as local polynomial expansion methods [35] and spline techniques [36]. In this article, we combine the local polynomial expansion technique with the least squares estimation method to estimate unknown model parameters.

## 2.2. Estimation Method

First, we briefly introduce the local polynomial expansion method. The main idea of this method is that for a function whose form is completely unknown but satisfies certain smoothness conditions at a fixed local point, we apply Taylor expansion to approximate the function as a polynomial about the local point. By using this method, the estimation of the function can be transformed into a parameter estimation problem of local polynomial coefficients. In this article, we elucidate the estimation method by using the first-order Taylor expansion. The higher-order Taylor expansion only increases the number of parameters to be estimated, and there is no essential difference in the algorithm. Specifically, the coefficient functions  $\alpha_1(\cdot)$  and  $\beta_1(\cdot)$  in model (3) are assumed to have a second continuous derivative, denoted as  $\alpha_1'(\cdot)$ ,  $\alpha_1''(\cdot)$ ,  $\beta_1'(\cdot)$  and  $\beta_1''(\cdot)$ , respectively.  $\forall t_0 \in (0,1)$ , by Taylor expansion, we have

$$\alpha_1(t_i) \approx \alpha_1(t_0) + \alpha_1'(t_0)(t_i - t_0), \quad \beta_1(t_i) \approx \beta_1(t_0) + \beta_1'(t_0)(t_i - t_0).$$

Then, in the local neighborhood of  $t_0$ , model (3) can be approximated as follows:

$$\begin{aligned} Y_i &\approx \mathbf{X}_{i,s_1}^T \alpha_1(t_0) + \mathbf{Y}_{i,p_1}^T \beta_1(t_0) + \mathbf{X}_{i,s_2}^T \alpha_2(t_0) + \mathbf{Y}_{i,p_2}^T \beta_2(t_0) \\ &\quad + \mathbf{X}_{i,s_1}^T \alpha_1'(t_0)(t_i - t_0) + \mathbf{Y}_{i,p_1}^T \beta_1'(t_0)(t_i - t_0) + \varepsilon_i \\ &= \boldsymbol{\gamma}^T \mathbf{Z}_i + \varepsilon_i, \quad i = 1, \dots, n, \end{aligned}$$

where

$$\boldsymbol{\gamma} = \boldsymbol{\gamma}(t_0) = \left\{ \alpha_1^T(t_0), \beta_1^T(t_0), \alpha_2^T(t_0), \beta_2^T(t_0), (\alpha_1'(t_0))^T, (\beta_1'(t_0))^T \right\}^T,$$

and

$$\mathbf{Z}_i = \mathbf{Z}_i(t_0) = \left\{ \mathbf{X}_{i,s_1}^T, \mathbf{Y}_{i,p_1}^T, \mathbf{X}_{i,s_2}^T, \mathbf{Y}_{i,p_2}^T, \mathbf{X}_{i,s_1}^T(t_i - t_0), \mathbf{Y}_{i,p_1}^T(t_i - t_0) \right\}^T.$$

We can obtain a locally weighted least squares estimate of the coefficient function in the model (3) by taking the minimum of the weighted sum of squared errors as follows:

$$\sum_{i=p+1}^n K_h(t_i - t_0) (Y_i - \boldsymbol{\gamma}^T \mathbf{Z}_i)^2,$$

where  $K(\cdot)$  is the kernel function,  $K_h(\cdot) = K(\cdot/h)/h$ , and  $h$  is the bandwidth. Using the least squares estimation theory, it is not difficult to obtain the estima-

tion of parameter  $\gamma$ , denoted as  $\hat{\gamma}$ , which has the following expression:

$$\hat{\gamma} = \hat{\gamma}(t_0) = (\mathbf{Z}^T \mathbf{K} \mathbf{Z})^{-1} \mathbf{Z}^T \mathbf{K} \mathbf{Y}, \quad (4)$$

where  $\mathbf{K} = \text{diag}\{K_h(t_{p+1} - t_0), \dots, K_h(t_n - t_0)\}$ ,  $\mathbf{Y} = (Y_{p+1}, \dots, Y_n)^T$ , and  $\mathbf{Z} = \mathbf{Z}(t_0)$  is a  $(n-p) \times (p+s+s_1+p_1)$  matrix with the  $k$ th row as  $\{\mathbf{Z}_{k+p}(t_0)\}^T, k=1, \dots, n-p$ .

For the residual term  $\varepsilon_i$  in model (3), its variance can be estimated locally as:

$$\hat{\sigma}^2 = \hat{\sigma}^2(t_0) = \frac{1}{n-p} \sum_{i=p+1}^n (Y_i - \hat{\gamma}^T \mathbf{Z}_i)^2.$$

### 2.3. Asymptotic Normality

For the convenience of description, we introduce the following definitions and notations. Let  $\mathbf{H}$  be a diagonal matrix of order  $(p+s+p_1+s_1)$ ; the first  $p+s$  elements on the main diagonal are 1, and the other  $p_1+s_1$  elements of the main diagonal are  $h$ . Let  $\mu_j = \int u^j K(u) du$ ,  $\nu_j = \int u^j K^2(u) du$  for

$j=0,1,2,3$ . Denote  $\mathbf{G}_i = (\mathbf{X}_{i,s_1}^T, \mathbf{Y}_{i,p_1}^T, \mathbf{X}_{i,s_2}^T, \mathbf{Y}_{i,p_2}^T)^T$ ,  $\mathbf{V}_i = (\mathbf{X}_{i,s_1}^T, \mathbf{Y}_{i,p_1}^T)^T$ ,

$i=p+1, \dots, n$  and  $\mathbf{\Omega}_1 = \mathbf{\Omega}_1(t_0) = E\{\mathbf{G}_i \mathbf{G}_i^T | t=t_0\}$ ,

$\mathbf{\Omega}_2 = \mathbf{\Omega}_2(t_0) = E\{\mathbf{G}_i \mathbf{V}_i^T | t=t_0\}$ ,  $\mathbf{\Omega}_3 = \mathbf{\Omega}_3(t_0) = E\{\mathbf{V}_i \mathbf{V}_i^T | t=t_0\}$ . Define

$$\mathbf{D} = \mathbf{D}(t_0) = \begin{pmatrix} \mathbf{D}_1 & \mathbf{D}_2 \\ \mathbf{D}_2^T & \mathbf{D}_3 \end{pmatrix}, \quad \mathbf{\Sigma} = \mathbf{\Sigma}(t_0) = \begin{pmatrix} \mathbf{\Sigma}_1 & \mathbf{\Sigma}_2 \\ \mathbf{\Sigma}_2^T & \mathbf{\Sigma}_3 \end{pmatrix},$$

where  $\mathbf{D}_i = \mathbf{D}_i(t_0) = \mu_{i-1} \mathbf{\Omega}_i(t_0), i=1,2,3$ , and the definition of  $\mathbf{\Sigma}_i = \mathbf{\Sigma}_i(t_0)$  is provided in the Appendix.

Throughout the derivation, we impose the following conditions:

**C1** The measurable functions  $\alpha_{1,i}(\cdot) (i=1, \dots, s_1)$  and  $\beta_{1,j}(\cdot) (j=1, \dots, p_1)$  have second continuous derivatives in the interval  $[0,1]$ .

**C2** The kernel function  $K(\cdot)$  is symmetrically bounded and has a compact support on the interval  $[-1,1]$ .

**C3**  $(Y_i, \mathbf{X}_{i,s_1}, \mathbf{X}_{i,s_2}, i=1, \dots, n)$  is a strictly stationary  $\alpha$ -mixing process, there exists  $\delta > 0$  such that  $E|Y_i|^{2(2+\delta)} < \infty$ , and the mixing coefficient  $\alpha(i) = O(i^{-\tau})$ , where  $\tau = (2+\delta)(1+\delta)/\delta$ .

**C4** The bandwidth satisfies  $nh^{1+4/\delta} \rightarrow \infty$ .

**C5** For  $t_0 \in (0,1)$ , the matrix  $\mathbf{D}(t_0)$  is full rank.

**C6** For  $t_0 \in (0,1)$ , the matrix  $\mathbf{\Sigma}(t_0)$  is positively definite.

Note that **C1**, **C2** and **C4** are the conditions that the local polynomial expansion method needs to satisfy, **C3** is the condition required for the moment of random variables, and **C5** and **C6** are requirements to be satisfied by the large sample properties of the estimators. Conditions **C1-C4** are similar to those in Robinson [4] and Cai [9].

**Theorem 1.** Under Conditions **C1-C6**, when  $h \rightarrow 0$  and  $nh \rightarrow \infty$ , we have

$$\sqrt{nh}\mathbf{H}\left(\hat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}_0-\frac{h^2}{2}\mathbf{b}\right)\xrightarrow{\mathcal{D}}N(\mathbf{0},\boldsymbol{\Sigma}_{\boldsymbol{\gamma}}),$$

where  $\mathbf{b}=\mathbf{b}(t_0)=\begin{pmatrix} \boldsymbol{\Omega}_1^{-1}\boldsymbol{\Omega}_2\boldsymbol{\mu}_2\boldsymbol{\eta}'(t_0) \\ \mathbf{0} \end{pmatrix}$  and  $\boldsymbol{\eta}'(t_0)=\begin{pmatrix} \boldsymbol{\alpha}_1''(t_0) \\ \boldsymbol{\beta}_1''(t_0) \end{pmatrix}$ ,

$$\boldsymbol{\Sigma}_{\boldsymbol{\gamma}}=\boldsymbol{\Sigma}_{\boldsymbol{\gamma}}(t_0)=\mathbf{D}^{-1}\boldsymbol{\Sigma}\mathbf{D}^{-1}.$$

Therefore, we obtain

$$\sqrt{nh}\left\{\hat{\boldsymbol{\xi}}(t_0)-\boldsymbol{\xi}(t_0)-\frac{h^2}{2}\boldsymbol{\Omega}_1^{-1}\boldsymbol{\Omega}_2\boldsymbol{\mu}_2\boldsymbol{\eta}'(t_0)\right\}\xrightarrow{\mathcal{D}}N(\mathbf{0},\boldsymbol{\Sigma}_{\boldsymbol{\xi}}),$$

where  $\boldsymbol{\xi}_0=\boldsymbol{\xi}(t_0)=\{\boldsymbol{\alpha}_1^T(t_0),\boldsymbol{\alpha}_2^T(t_0),\boldsymbol{\beta}_1^T(t_0),\boldsymbol{\beta}_2^T(t_0)\}^T$  and

$$\boldsymbol{\Sigma}_{\boldsymbol{\xi}}=\boldsymbol{\Sigma}_{\boldsymbol{\xi}}(t_0)=\mathbf{D}_1^{-1}\boldsymbol{\Sigma}_1\mathbf{D}_1^{-1}.$$

In practice, the variance-covariance matrix of  $(\hat{\boldsymbol{\xi}}(t_0)-\boldsymbol{\xi}_0)$  can be calculated by using the following formula:

$$(nh)^{-1}\mathbf{D}_{n,1}^{-1}\boldsymbol{\Sigma}_{n,1}\mathbf{D}_{n,1}^{-1}, \tag{5}$$

where

$$\mathbf{D}_{n,1}=\mathbf{D}_{n,1}(t_0)=\frac{1}{n-p}\sum_{i=p+1}^n\mathbf{G}_i\mathbf{G}_i^TK_h(t_i-t_0),$$

$$\boldsymbol{\Sigma}_{n,1}=\frac{h}{(n-p)\nu_0}\left(\sum_{i=p+1}^n\hat{\varepsilon}_i\mathbf{G}_iK_h(t_i-t_0)\right)\left(\sum_{j=p+1}^n\hat{\varepsilon}_j\mathbf{G}_jK_h(t_j-t_0)\right)^T,$$

and  $\hat{\varepsilon}_i, i=p+1, \dots, n$ , are residuals of model (3). The proof of Theorem 1 is provided in the Appendix.

For constant coefficients in the model (3), the corresponding estimators obtained from (4) are not  $\sqrt{n}$  consistent. To improve its convergence rate, we use the following formula (6) to obtain a global estimator, that is,

$$\begin{pmatrix} \tilde{\boldsymbol{\alpha}}_2 \\ \tilde{\boldsymbol{\beta}}_2 \end{pmatrix}=\int_{\mathcal{T}}\boldsymbol{\Gamma}(t_0)\begin{pmatrix} \hat{\boldsymbol{\alpha}}_2(t_0) \\ \hat{\boldsymbol{\beta}}_2(t_0) \end{pmatrix}dt_0, \tag{6}$$

where the weight matrix  $\boldsymbol{\Gamma}(t_0)$  satisfies  $\int_{\mathcal{T}}\boldsymbol{\Gamma}(t_0)dt_0=\mathbf{I}_{s \times s}$ , and  $\mathbf{I}$  is an identity matrix. Typically, we choose  $\boldsymbol{\Gamma}(t_0)$  to be the standardized inverse covariance matrix [37] of  $\hat{\boldsymbol{\alpha}}_2(t_0)$  and  $\hat{\boldsymbol{\beta}}_2(t_0)$ , which can be obtained from (5). Then, the finite sample variances of the global estimators  $(\tilde{\boldsymbol{\alpha}}_2^T, \tilde{\boldsymbol{\beta}}_2^T)^T$  can be calculated by the following formula (7):

$$\frac{1}{n-p}\sum_{i=p+1}^n\boldsymbol{\Gamma}(t_i)\mathbf{A}\mathbf{D}_n^{-1}(t_i)\boldsymbol{\Sigma}_n^{-1}(t_i)\mathbf{D}_n^{-1}(t_i)\mathbf{A}^T\boldsymbol{\Gamma}^T(t_i), \tag{7}$$

where  $\mathbf{A}$  is a  $(s_2+p_2) \times (p+s+s_1+p_1)$  matrix with  $a_{ij}$  as the  $(i, j)$  element of  $\mathbf{A}$ ; furthermore,  $a_{ij}=1$  for  $i=1, \dots, s_2+p_2, j=i+s_1+p_1$ ; otherwise,  $a_{ij}=0$ . Definitions of  $\mathbf{D}_n$  and  $\boldsymbol{\Sigma}_n$  can be seen in the Appendix.

### 2.4. Selection Best Bandwidth and Order

When using estimator (4) to obtain  $\hat{\boldsymbol{\gamma}}$ , bandwidth  $h$  and order  $p$  ( $p=p_1+p_2$ )

must be determined. Cheng *et al.* [26] proposed a Bayesian approach to determine bandwidth selection for local constant estimators of time-varying coefficients, Chen *et al.* [23] applied the cross-validation (CV) bandwidth selection to estimate the time-varying coefficient heterogeneous autoregressive model. Inspired by the idea of the average mean squared (AMS) criterion proposed by Cai *et al.* [8], we slightly modify this criterion to simultaneously select the best bandwidth  $h$  and the optimal order  $p$  of model (3). The procedure is as follows: First, we divide the sample data into  $m$  groups and denote the sample size of the  $q$ th group as  $n_q$ ,  $q = 1, \dots, m$ . Second, given the values of  $h$  and  $p$ , and based on data of removing the  $q$ th group, we estimate  $\gamma$  by using (4), denoted as  $\hat{\gamma}_q$ . Third, we calculate the estimated mean square error for sample data of the  $q$ th group, denoted as  $\text{AMS}_q(h, p)$ , and then we obtain the total mean squared error when  $q$  goes from 1 to  $m$ . Finally, the optimal  $h$  (*i.e.*,  $h_{\text{opt}}$ ) and  $p$  (*i.e.*,  $p_{\text{opt}}$ ) that minimize the (AMS) error has the following form:

$$(h_{\text{opt}}, p_{\text{opt}})^{\text{T}} = \arg \min_{h, q} \sum_{q=1}^m \text{AMS}_q(h, p),$$

where for  $q = 1, \dots, m$ ,

$$\text{AMS}_q(h, p) = \frac{1}{n_q - p} \sum_{i=1+(q-1)n_q+p}^{1+qn_q} \{Y_i - \hat{\gamma}^{\text{T}} \mathbf{Z}_i\}^2.$$

Note that in the second step, except for all covariates of  $\mathbf{X}_{i,s_1}$  and  $\mathbf{X}_{i,s_2}$ , we can also try to use subsets of  $\mathbf{X}_{i,s_1}$  and  $\mathbf{X}_{i,s_2}$  in  $\mathbf{Z}_i$  when we estimate  $\gamma$  by using (4). Obviously, different subsets of covariates of  $\mathbf{X}$  may result in different best combinations of  $h$  and  $p$ , and we take the combination of  $h$  and  $p$  corresponding to the minimum value of AMS for all subsets of covariates of interest as the optimal bandwidth and the order of the autoregressive part. Therefore, AMS can also be used to conduct variable selection of model (3).

### 3. Simulation Studies

We perform simulation studies to examine the finite-sample performances of the proposed estimators and consider the following partial time-varying coefficient regression and autoregressive mixed model:

$$Y_i = \alpha_1(t_i)X_{i,1} + \alpha_2 X_{i,2} + \beta_1(t_i)Y_{i-1} + \beta_2 Y_{i-2} + \varepsilon_i, \quad i = 3, 4, \dots, n, \quad (8)$$

where  $\alpha_1(t_i) = \sin(t_i)$ ,  $\alpha_2 = 0.3$ ,  $\beta_1(t_i) = 0.5 \exp(t_i)$ ,  $\beta_2 = -0.1$ ;  $X_{i,1}$  follows the autoregressive moving average model with  $p=1$  and  $q=1$ , *i.e.*, ARMA(1,1); and  $X_{i,2}$  is an autoregressive (AR) model of order 1, *i.e.*, AR(1). Typically,  $X_{i,1} = 0.1X_{i-1,2} + e_{i,1} + 0.3e_{i-1,1}$  and  $X_{i,2} = 0.2X_{i-1,2} + e_{i,2}$ . Both  $e_{i,1}$  and  $e_{i,2}$  are standard white noise series with mean 0 and variance 1, *i.e.*,  $e_{i,1} \sim WN(0,1)$  and  $e_{i,2} \sim WN(0,1)$ . Also,  $\varepsilon_i \sim WN(0,0.04)$  and that is independent of both  $X_{i,1}$  and  $X_{i,2}$ . We run 500 Monte Carlo simulations for 9 grid points at  $t_0 = 0.1, \dots, 0.9$  in the time interval  $[0,1]$  and examine the average of 500 simulations of the parameters in model (8) at these 9 points. We apply the



Gaussian kernel function with standard error 2 to the estimation of  $\gamma$ . Once  $\hat{\gamma}$  is obtained by (4), then  $\hat{\xi}_0$  is immediately known and its variance estimation can be calculated by (5). For constant parameter vector  $\alpha_2$  and  $\beta_2$ , we obtain their global estimators  $\tilde{\alpha}_2$  and  $\tilde{\beta}_2$  by using formula (6), and compute the corresponding variance estimation of  $\tilde{\alpha}_2$  and  $\tilde{\beta}_2$  by (7). For each simulation, we consider the sample size as  $n = 300$  and  $n = 800$ , respectively. According to simulation results under sample size 300 and 800, we find that estimation results are good when bandwidth is between 0.02 and 0.08. Thus, three bandwidths of  $h$  are chosen, that is,  $h = 0.03, 0.05$  and  $0.07$ . Finally, we report the following quantities under each  $h$  and  $n$  for  $\hat{\xi}_0$  in **Table 1** & **Table 2**: Monte Carlo average estimators of  $\hat{\xi}_0$ , Monte Carlo standard deviation of estimates (ESD), Monte Carlo average of estimated standard errors (ASE) and coverage probability of nominal 95% confidence intervals (CP).

**Table 1.** Estimation results for time-varying coefficient  $\alpha_1(t) = \sin(t)$  and  $\beta_1(t) = 0.5e^{-t}$  under different scenarios.

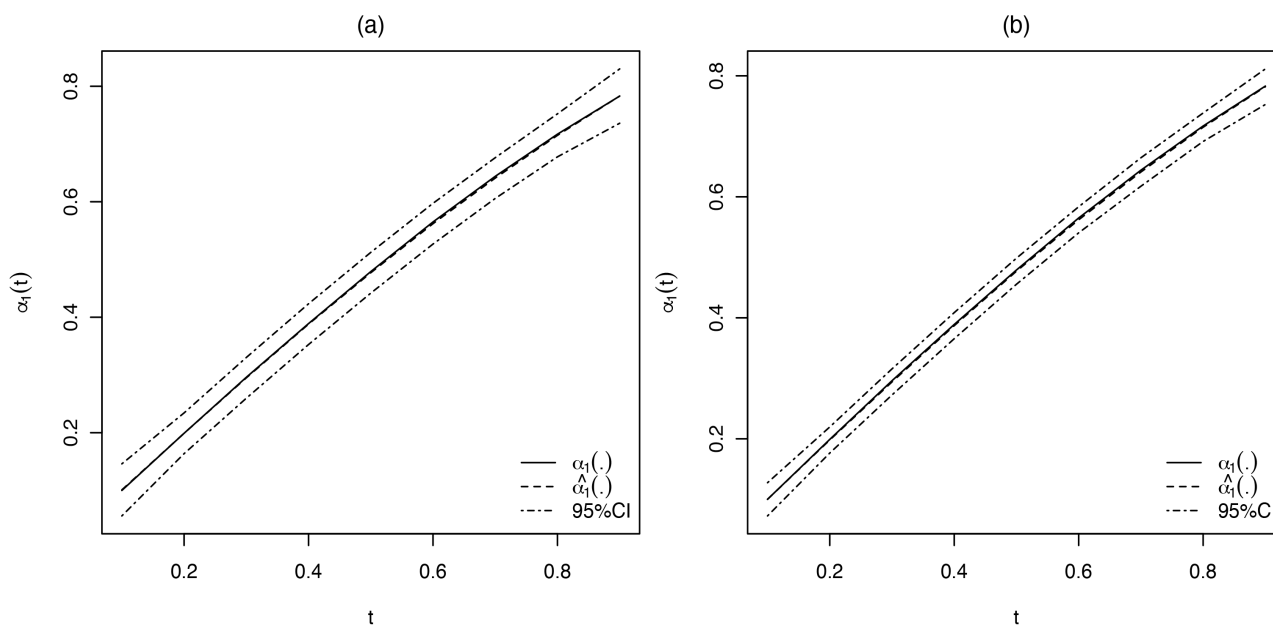
$n$	$t_0$	$h$	$\alpha_1(t_0)$	$\hat{\alpha}_1(t_0)$	ASE	ESD	CP (%)	$\beta_1(t_0)$	$\hat{\beta}_1(t_0)$	ASE	ESD	CP (%)
300	0.2	0.03	0.199	0.199	0.025	0.025	92.6	0.409	0.407	0.064	0.058	94.2
		0.05		0.199	0.020	0.018	93.8		0.406	0.051	0.046	92.0
		0.07		0.198	0.018	0.017	94.0		0.409	0.047	0.043	92.8
	0.4	0.03	0.389	0.389	0.026	0.023	93.2	0.335	0.333	0.049	0.046	92.2
		0.05		0.388	0.018	0.018	95.0		0.336	0.039	0.036	92.0
		0.07		0.386	0.015	0.015	94.0		0.338	0.031	0.031	94.0
	0.6	0.03	0.565	0.564	0.025	0.023	92.0	0.274	0.271	0.039	0.037	93.4
		0.05		0.562	0.018	0.018	93.4		0.274	0.031	0.029	92.6
		0.07		0.558	0.015	0.016	93.4		0.277	0.026	0.025	94.2
	0.8	0.03	0.717	0.716	0.024	0.024	93.4	0.225	0.223	0.032	0.031	92.2
		0.05		0.715	0.020	0.019	93.6		0.226	0.025	0.025	93.0
		0.07		0.713	0.017	0.018	95.4		0.225	0.023	0.023	94.8
800	0.2	0.03	0.199	0.199	0.014	0.014	95.4	0.409	0.407	0.039	0.036	92.6
		0.05		0.198	0.011	0.011	94.6		0.410	0.031	0.029	93.2
		0.07		0.197	0.011	0.010	95.2		0.410	0.027	0.026	95.2
	0.4	0.03	0.389	0.389	0.015	0.014	93.4	0.335	0.336	0.030	0.029	94.6
		0.05		0.387	0.012	0.011	92.6		0.337	0.024	0.023	94.8
		0.07		0.386	0.010	0.010	93.2		0.337	0.020	0.019	93.8
	0.6	0.03	0.565	0.565	0.016	0.015	94.0	0.274	0.272	0.023	0.023	94.2
		0.05		0.562	0.012	0.011	93.2		0.277	0.019	0.018	94.6
		0.07		0.560	0.010	0.010	93.2		0.276	0.016	0.015	93.4
	0.8	0.03	0.717	0.716	0.015	0.015	93.0	0.225	0.225	0.020	0.019	94.4
		0.05		0.715	0.012	0.012	95.8		0.225	0.016	0.016	94.8
		0.07		0.713	0.011	0.011	92.0		0.226	0.014	0.014	95.4

**Table 1** shows that both estimators of  $\alpha_1(\cdot)$  and  $\beta_1(\cdot)$  are approximate to their true values. The ASEs are close to the ESDs, which demonstrates the good performance of the variance estimation by Theorem 1. 95% confidence interval coverage probabilities are reasonably accurate, matching the nominal level. According to the estimation results from **Table 2**, biases of  $\tilde{\alpha}_2$  and  $\tilde{\beta}_2$  can be ignored, ASEs are still approximate to the ESDs and CPs of both estimators are close to 95%, which means good variance estimation of global estimators  $\tilde{\alpha}_2$  and  $\tilde{\beta}_2$ . Furthermore, when bandwidth  $h$  increases from 0.03 to 0.07, as expected, biases of  $\hat{\alpha}_1(\cdot)$  and  $\hat{\beta}_1(\cdot)$  tend to be larger, and corresponding ASEs and ESDs tend to be smaller. When the sample size  $n$  increases from 300 to 800, the ASEs and ESDs of all estimators decrease, which indicates that increasing the sample size can improve our proposed estimators.

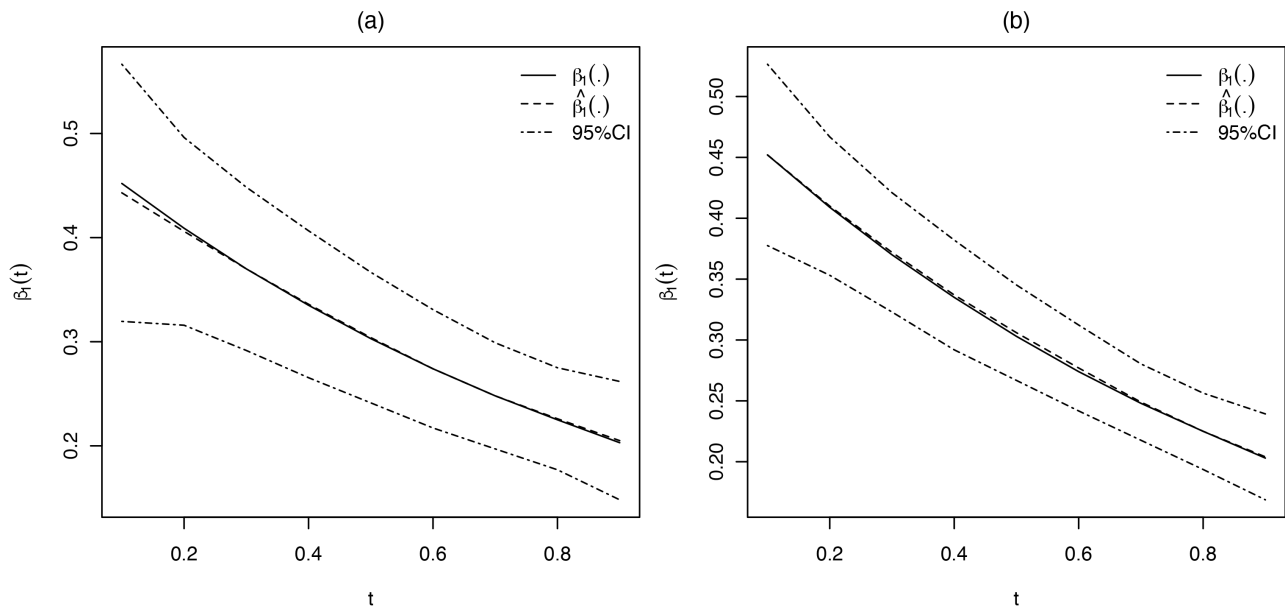
**Figure 1** & **Figure 2** present the true and estimated time-varying coefficients

**Table 2.** Estimation results for constant coefficients  $\alpha_2 = 0.3$  and  $\beta_2 = -0.1$  under different scenarios.

$n$	$h$	$\alpha_2$	$\tilde{\alpha}_2$	ASE	ESD	CP (%)	$\beta_2$	$\tilde{\beta}_2$	ASE	ESD	CP (%)
300	0.03	0.3	0.301	0.012	0.012	95.2	-0.1	-0.102	0.021	0.022	95.8
	0.05		0.300	0.012	0.013	96.0		-0.102	0.021	0.023	96.6
	0.07		0.300	0.012	0.013	97.2		-0.100	0.020	0.023	97.2
800	0.03	0.3	0.300	0.007	0.008	95.6	-0.1	-0.100	0.013	0.014	95.4
	0.05		0.300	0.007	0.008	97.4		-0.101	0.013	0.014	95.6
	0.07		0.300	0.007	0.008	97.2		-0.101	0.012	0.014	97.4



**Figure 1.** Estimated curves of  $\alpha_1(t) = \sin(t)$  with bandwidth  $h = 0.05$ . (a) Sample size  $n = 300$ ; (b) sample size  $n = 800$ . The solid lines are the true functions of  $\alpha_1(t)$ , the dashed lines are the estimates of  $\alpha_1(t)$ , and the dash-dotted lines are the 95% confidence intervals.



**Figure 2.** Estimated curves of  $\beta_1(t) = 0.5e^{-t}$  with bandwidth  $h = 0.05$ . (a) sample size  $n = 300$ ; (b) sample size  $n = 800$ . The solid lines are the true functions of  $\beta_1(t)$ , the dashed lines are the estimates of  $\beta_1(t)$ , and the dash-dotted lines are the 95% confidence intervals.

of model (8) as well as 95% pointwise confidence intervals. It can be intuitively seen that the estimated curves of  $\hat{\alpha}_1(\cdot)$  and  $\hat{\beta}_1(\cdot)$  are very close to the true curves, which again confirms the good performance of our proposed estimators.

#### 4. Real Data Analysis

In this Section, we use Lake Shasta inflow data [38] to illustrate the application value of our proposed model. The data includes 454 months of measured values for several climatic variables: air temperature (Temp), dew point (DewPt), cloud cover (CldCvr), wind speed (WndSpd), precipitation (Precip), and inflow (Inflow) at Lake Shasta, California, USA. We are interested in building models to predict Lake Shasta inflow based on these climate variables. We treat the lake water inflow (Inflow) as a time series, but it has both autocorrelation and heteroscedasticity through the Box-Ljung autocorrelation test and Lagrange Multiplier test. We let the dependent variable be the inflow of lake water. According to the suggestion by Shumway and Stoffer [38], it is better to model this variable after logarithmic transformation. Therefore, we set  $Y_i = \log(\text{Inflow}_i)$ ,  $i = 1, \dots, 454$  and then use the proposed partial time-varying coefficient regression and autoregressive mixed model (3) to model the relationship between inflows and the other five climate variables.

Before using (4) to obtain estimators, we need to determine the optimal bandwidth  $h$ , the autoregressive order  $p$  and which covariates have time-varying coefficient effects as well as which covariates have constant coefficient effects. In other words, we need to conduct variable selection. As we mentioned in Section 2.4, the criterion AMS can not only determine the optimal bandwidth  $h$  and the

autoregressive order  $p$ , but also conduct variable selection. In this analysis, when choosing a combination of covariates, *i.e.*, which variables should be put into  $\mathbf{X}_{i,s_1}$  and which variables are put into  $\mathbf{X}_{i,s_2}$ , as well as which lag terms of the dependent variable are put into  $\mathbf{Y}_{i,p_1}$  and  $\mathbf{Y}_{i,p_2}$ , respectively, one rule we follow is that the vectors  $\mathbf{X}_{i,s_1}$  and  $\mathbf{Y}_{i,p_1}$  cannot be empty simultaneously. Otherwise, model (3) will reduce to the traditional regression and autoregressive mixed model. The autoregressive order  $p$  we consider starts at 1 and increases sequentially. A rough criterion for  $p$  is that there is no autocorrelation and heteroscedasticity in the residuals of the model. When  $p$  is given, we determine the optimal combination of bandwidth  $h$  with a combination of covariates according to the value of AMS. If the estimator of a constant effect covariate selected by the AMS is insignificant, then we delete this covariate and calculate AMS again. By using this strategy, we can guarantee that all covariates with constant coefficients selected by the AMS are significant.

The procedure of using the proposed model to analyze real data is as follows: First, we let Temp, DewPt, CldCvr, WndSpd and Precip be five candidate covariates. When calculating AMS, according to the suggestion of Cai *et al.* [8], we choose  $m = 4$  and  $n_q = [0.1n]$ , where  $[\ ]$  denotes a rounding function, and the optimal bandwidth satisfies  $h \propto n^{-1/5}$ . Since the sample size of Shasta Lake inflow data is 454, we set the selection range of bandwidth  $h$  from 0.08 to 0.50. Combined with other rules mentioned in the previous paragraph, that is, the  $\mathbf{X}_{i,s_1}$  and  $\mathbf{Y}_{i,p_1}$  cannot be empty simultaneously, the estimation for covariates selected by the model with constant coefficients should be significant. Then, we consider that model (3) contains only 1, 2, 3, 4 and 5 of the candidate covariates and calculate the corresponding values of AMS based on given  $p$  and  $h$ . Next, by comparing AMS under all various combinations, we find the minimum value of AMS occurring at  $h = 0.15$ ,  $p = 4$ , and the covariate combination is

$$\mathbf{X}_{i,s_1} = (\text{CldCvr}_i, \text{WndSpd}_i)^T, \quad \mathbf{X}_{i,s_2} = \text{Precip}_i,$$

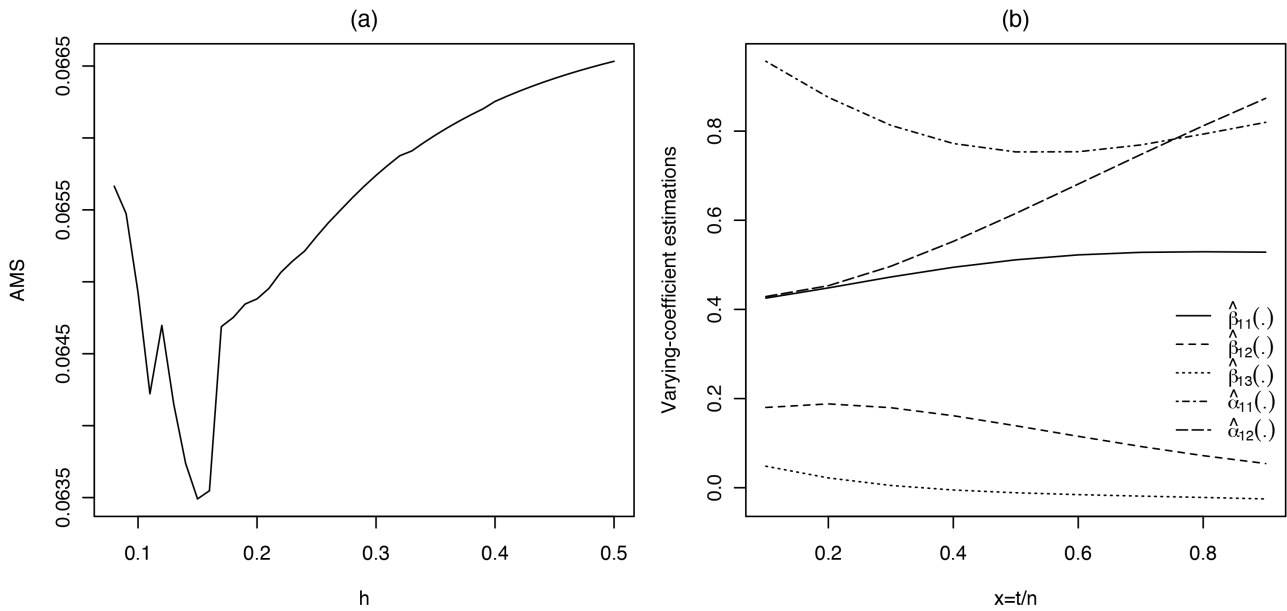
$$\mathbf{Y}_{i,p_1} = (\log(\text{Inflow}_{i-1}), \log(\text{Inflow}_{i-2}), \log(\text{Inflow}_{i-3}))^T \quad \text{and}$$

$\mathbf{Y}_{i,p_2} = \log(\text{Inflow}_{i-4})$ . Thus,  $(s_1, s_2, p_1, p_2) = (2, 1, 3, 1)$ . **Figure 3(a)** shows the AMS changes with different  $h$  under this combination in model (3).

Therefore, the final model to be estimated is:

$$Y_i = \alpha_{1,1}(t_i)\text{CldCvr}_i + \alpha_{1,2}(t_i)\text{WndSpd}_i + \alpha_{2,1}\text{Precip}_i + \beta_{1,1}(t_i)Y_{i-1} + \beta_{1,2}(t_i)Y_{i-2} + \beta_{1,3}(t_i)Y_{i-3} + \beta_{2,1}Y_{i-4} + \varepsilon_i, \quad (9)$$

where  $Y_i = \log(\text{Inflow}_i)$ . Once  $\mathbf{X}_{i,s_1}$ ,  $\mathbf{X}_{i,s_2}$ ,  $\mathbf{Y}_{i,p_1}$  and  $\mathbf{Y}_{i,p_2}$  are determined, we then use (4) to obtain local estimators and (6) to get global estimators of constant coefficients. To evaluate the prediction performance of model (9), we use the first 451 sample data to estimate the unknown parameters and obtain  $\hat{\varepsilon}_i \sim WN(0, 0.049)$ . The estimated time-varying coefficients  $\hat{\alpha}_{1,1}(\cdot)$ ,  $\hat{\alpha}_{1,2}(\cdot)$ ,  $\hat{\beta}_{1,1}(\cdot)$ ,  $\hat{\beta}_{1,2}(\cdot)$  and  $\hat{\beta}_{1,3}(\cdot)$  are presented in **Figure 3(b)**, and the estimated constant coefficients  $\tilde{\alpha}_{2,1}$  and  $\tilde{\beta}_{2,1}$  are reported in **Table 3**. Interestingly, considering the time-varying coefficient estimation results, it seems that the prediction



**Figure 3.** (a) AMS results under different bandwidths  $h$  for model (9); (b) time-varying coefficient estimation results of model (9).

**Table 3.** Constant coefficient estimation results of model (9).

Coefficient	Estimate	Std. Error	T value	Pr (> t )
$\alpha_{2,1}$	$1.852 \times 10^{-3}$	$0.155 \times 10^{-3}$	11.983	<0.001
$\beta_{2,1}$	0.141	0.036	3.930	<0.001

effect of lag order 2 of  $Y_i$  (i.e.,  $Y_{i-2}$ ) increases with time  $t_i$  while the lag order 1 of  $Y_i$  (i.e.,  $Y_{i-1}$ ) has no such characteristic, and effect of the lag order 3 of  $Y_i$  (i.e.,  $Y_{i-3}$ ) is close to zero. The time-varying effects from both  $CldCvr_i$  and  $WndSpd_i$  decrease with  $t_i$ .

The last three sample values are used to calculate the relative prediction error (RPE) defined as:

$$RPE = \left| \frac{\hat{Y}_i - Y_i}{Y_i} \right|,$$

where  $\hat{Y}_i = \log(\widehat{\text{Inflow}}_i)$  ( $i = 452, 453, 454$ ) are predicted by model (9).

**Table 4** shows the relative prediction error results of model (9) for the three forward steps. It can be seen that the average RPE of model (9) is only 2.6%. For comparison, **Table 4** also shows the 3-step forward prediction results of the lake inflow based on the FAR(p) model. According to the AMS criterion of Cai *et al.* [8], the optimal fitted FAR(p) model is obtained when  $h = 0.17$ ,  $p = 3$  and  $d = 1$ , that is,

$$Y_i = \hat{\alpha}_0(Y_{i-1}) + \hat{\alpha}_1(Y_{i-1})Y_{i-1} + \hat{\alpha}_2(Y_{i-1})Y_{i-2} + \hat{\alpha}_3(Y_{i-1})Y_{i-3}, \quad (10)$$

In addition, we also apply classical regression and autoregressive models to analyze the lake water inflow and find that residuals have heteroskedasticity. Therefore, we combine AR (p) model with the autoregressive conditional

**Table 4.** 3-step forward relative prediction error of  $\log(\text{Inflow}_i)$  under different models.

Forward step	True value	Model (9)	Model (10)	Model (11)
1	4.301	0.003	0.061	0.056
2	4.436	0.072	0.075	0.017
3	4.535	0.001	0.133	0.114
Average RPE		0.026	0.090	0.062

heteroskedasticity (ARCH) model, *i.e.*, AR(p)-ARCH(q) model, to fit the residuals. The final optimal AR(p)-ARCH(q) model is:

$$\begin{aligned}
 Y_i &= 1.335 + 0.460Y_{i-1} + 0.159Y_{i-2} + 0.415\text{CldCvr}_i \\
 &\quad + 0.212\text{WndSpd}_i + 0.002\text{Precip}_i + R_i, \\
 R_i &= -0.093R_{i-1} - 0.088R_{i-2} - 0.161R_{i-3} + \delta_i, \\
 \delta_i &= \phi_i\eta_i, \quad \eta_i \sim WN(0,1), \\
 \phi_i^2 &= 0.040 + 0.078\delta_i^2.
 \end{aligned} \tag{11}$$

By comparing the prediction results of the three models in **Table 4**, our proposed model (9) performs better than the other two models. Compared to model (10), model (9) takes into account the effects of covariates on lake inflow; compared to model (11), model (9) uses the advantages of the time-varying coefficient. Since our proposed model combines the strengths of the FAR(p) model and the classical regression and autoregressive mixed model, it has advantages on modeling time series data with complex correlations between sample components.

## 5. Conclusions and Future Works

In this article, we propose a partial time-varying coefficient regression and autoregressive mixed model, which can be regarded as an extension of traditional regression and autoregressive mixed model and time-varying coefficient regression model. The proposed model is very flexible and can handle both complex correlations between time series components and effects of other covariates, so it can improve model fitting when building relationships between complex time series and covariates. We apply the local polynomial expansion technique and the least squares estimation method to obtain local estimates of the parameter functions in the model. At the same time, we also propose a global estimator algorithm for the constant coefficients in the model. In addition, we derive the asymptotic normality of the proposed estimators and conduct simulation studies to examine their finite-sample performances, and find our proposed estimators perform well. Finally, we apply the model to analyze the relationships between the water inflow of Shasta Lake and the other five climatic variables.

As we mentioned in Section 1, local polynomial technique is widely used to estimate varying-coefficient models, including time-varying coefficient time series models. One advantage of this method is it has solid theory so that we can

derive out explicit biases of corresponding estimators. Furthermore, local polynomial method is easily implemented. Based on the local estimators, we can obtain global estimators for constant coefficients by using formula (6). Note that it is needed to select the best bandwidth when using local polynomial method. However, if we apply other smooth methods such as spline approach to estimate time-varying coefficients in (3), we also need to determine some turning parameters, e.g., knots and degree.

Although we can apply the local polynomial expansion technique and the least squares estimation method to obtain the local estimators as well as corresponding variances, how to test these estimated time-varying coefficients is another important problem. For example, the AMS criterion suggests that model (9) includes five time-varying coefficients, that is,  $\alpha_{1,1}(\cdot)$ ,  $\alpha_{1,2}(\cdot)$ ,  $\beta_{1,1}(\cdot)$ ,  $\beta_{1,2}(\cdot)$  and  $\beta_{1,3}(\cdot)$ , but **Figure 3(b)** shows that some estimated curves seem to be flat across time. Thus, it makes us wonder whether those estimated time-varying coefficients can be treated as constant. One way to assess this is to construct the confidence band to see whether a time-varying coefficient is a function or a constant. This method may be realized by obtaining weak Bahadur representations of estimators [29]. Another way is to develop a more objective and theoretical test following the simulation-assisted hypothesis testing procedure proposed in Zhang and Wu [21]. This will be our future research work.

## Acknowledgements

Dr. Cao's research is supported by Natural Science Foundation of Top Talent of SZTU (grant no. GDRC202135).

## Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

## References

- [1] Hamilton, J.D. (1994) *Time Series Analysis*. Princeton University Press, Princeton.
- [2] Box, G.E.P., Jenkins, G.M. and Reinsel, G.C. (2008) *Time Series Analysis: Forecasting and Control*. 4th Edition, John Wiley & Sons, New York.  
<https://doi.org/10.1002/9781118619193>
- [3] Gujarati, D.N. and Porter, D.C. (2010) *Essentials of Econometrics*. 4th Edition, McGraw-Hill/Irwin, New York.
- [4] Robinson, P.M. (1989) Nonparametric Estimation of Time-Varying Parameters. In: Hackl, P., Eds., *Statistical Analysis and Forecasting of Economic Structural Change*, Springer-Verlag, Berlin, 253-264. [https://doi.org/10.1007/978-3-662-02571-0\\_15](https://doi.org/10.1007/978-3-662-02571-0_15)
- [5] Robinson, P.M. (1989) Time-Varying Nonlinear Regression. In: Hackl, P. and Westlund, A.H., Eds., *Economic Structure Change*, Springer-Verlag, Berlin, 179-190.  
[https://doi.org/10.1007/978-3-662-06824-3\\_13](https://doi.org/10.1007/978-3-662-06824-3_13)
- [6] Chen, R. and Tsay, R.S. (1993) Functional-Coefficient Autoregressive Models. *Journal of the American Statistical Association*, **88**, 298-308.  
<https://doi.org/10.1080/01621459.1993.10594322>

- [7] Hoover, D.R., Rice, J.A., Wu, C.O. and Yang, L.P. (1998) Nonparametric Smoothing Estimates of Time-Varying Coefficient Models with Longitudinal Data. *Biometrika*, **85**, 809-822. <https://doi.org/10.1093/biomet/85.4.809>
- [8] Cai, Z., Fan, J. and Yao, Q. (2000) Functional-Coefficients Regression Models for Nonlinear Time Series. *Journal of the American Statistical Association*, **95**, 941-956. <https://www.tandfonline.com/doi/abs/10.1080/01621459.2000.10474284>  
<https://doi.org/10.1080/01621459.2000.10474284>
- [9] Cai, Z. (2007) Trending Time-Varying Coefficient Time Series Models with Serially Correlated Errors. *Journal of Econometrics*, **136**, 163-188. <https://doi.org/10.1016/j.jeconom.2005.08.004>
- [10] Liu, X., Cai, Z. and Chen, R. (2007) Functional Coefficient Seasonal Time Series Models with an Application of Hawaii Tourism Data. *Computational Statistics*, **30**, 719-744. <https://doi.org/10.1007/s00180-015-0574-x>
- [11] Cai, Z., Chen, L. and Fang, Y. (2015) Semiparametric Estimation of Partially Varying-Coefficient Dynamic Panel Data Models. *Econometric Reviews*, **34**, 695-719. <https://doi.org/10.1080/07474938.2014.956569>
- [12] Cai, Z., Fang, Y. and Tian, D. (2018) Assessing Tail Risk Using Expectile Regressions with Partially Varying Coefficients. *Journal of Management Science and Engineering*, **3**, 183-213. <https://doi.org/10.3724/SP.J.1383.304011>
- [13] Chen, Y., Chua, W.S. and Koch, T. (2018) Forecasting Day-Ahead High-Resolution Natural-Gas Demand and Supply in Germany. *Applied Energy*, **228**, 1091-1100. <https://doi.org/10.1016/j.apenergy.2018.06.137>
- [14] Tu, Y. and Wang, Y. (2019) Functional Coefficient Cointegration Models Subject to Time-Varying Volatility with an Application to the Purchasing Power Parity. *Oxford Bulletin of Economics and Statistics*, **86**, 1401-1423. <https://doi.org/10.1111/obes.12309>
- [15] Cai, Z., Fang, Y., Lin, M. and Su, J. (2019) Inferences for a Partially Varying Coefficient Model with Endogenous Regressors. *Journal of Business & Economic Statistics*, **37**, 158-170. <https://doi.org/10.1080/07350015.2017.1294079>
- [16] Yousuf, K. and Ng, S. (2021) Boosting High Dimensional Predictive Regressions with Time Varying Parameters. *Journal of Econometrics*, **224**, 60-87. <https://doi.org/10.1016/j.jeconom.2020.08.003>
- [17] Kalli, M. and Griffin, J.E. (2014) Time-Varying Sparsity in Dynamic Regression Models. *Journal of Econometrics*, **178**, 779-793. <https://doi.org/10.1016/j.jeconom.2013.10.012>
- [18] Feldkircher, M., Huber, F. and Kastner, G. (2017) Sophisticated and Small versus Simple and Sizeable: When Does It Pay off to Introduce Drifting Coefficients in Bayesian VARs? Department of Economics Working Paper Series 260, WU Vienna University of Economics and Business.
- [19] Bitto, A. and Fruhwirth-Schnatter, S. (2019) Achieving Shrinkage in a Time-Varying Parameter Model Framework. *Journal of Econometrics*, **210**, 75-97. <https://doi.org/10.1016/j.jeconom.2018.11.006>
- [20] Li, D., Chen, J. and Lin, Z. (2011) Statistical Inference in Partially Time-Varying Coefficient Models. *Journal of Statistical Planning and Inference*, **141**, 995-1013. <https://doi.org/10.1016/j.jspi.2010.09.004>
- [21] Zhang, T. and Wu, W.B. (2012) Inference of Time-Varying Regression Models. *Annals of Statistics*, **40**, 1376-1402. <https://doi.org/10.1214/12-AOS1010>
- [22] Fan, G.L., Liang, H.Y. and Wang, J.F. (2013) Statistical Inference for Partially



- Time-Varying Coefficient Errors-in-Variables Models. *Journal of Statistical Planning and Inference*, **143**, 505-519. <https://doi.org/10.1016/j.jspi.2012.08.017>
- [23] Chen, X.B., Gao, J., Li, D. and Silvapulle, P. (2018) Nonparametric Estimation and Forecasting for Time-Varying Coefficient Realized Volatility Models. *Journal of Business & Economic Statistics*, **36**, 88-100. <https://doi.org/10.1080/07350015.2016.1138118>
- [24] Kim, S., Zhao, Z. and Xiao, Z. (2018) Efficient Estimation for Time-Varying coefficient Logitudinal Models. *Journal of Nonparametric Statistics*, **30**, 680-702. <https://doi.org/10.1080/10485252.2018.1467415>
- [25] Cuaresma, J.C., Doppelhofer, G., Feldkircher, M. and Huber, F. (2019) Spillovers from US Monetary Policy: Evidence from a Time Varying Parameter Global Vector Autoregressive Model. *Journal of the Royal Statistical Society Series A: Statistics in Society*, **182**, 831-861. <https://doi.org/10.1111/rssa.12439>
- [26] Cheng, T., Gao, J. and Zhang, X. (2019) Bayesian Bandwidth Estimation in Nonparametric Time-Varying Coefficient Models. *Journal of Business & Economic Statistics*, **37**, 1-12. <https://doi.org/10.1080/07350015.2016.1255216>
- [27] Hu, L., Huang, T. and You, J. (2019) Two-Step Estimation of Time-Varying Additive Model for Locally Stationary Time Series. *Computational Statistics & Data Analysis*, **130**, 94-110. <https://doi.org/10.1016/j.csda.2018.08.023>
- [28] Li, D., Phillips, P.C.B. and Gao, J. (2019) Kernel-Based Inference in Time-Varying Coefficient Cointegrating Regression. *Journal of Econometrics*, **215**, 607-632. <https://doi.org/10.1016/j.jeconom.2019.10.005>
- [29] Karmakar, S., Richter, S. and Wu, W.B. (2022) Simultaneous Inference for Time-Varying Models. *Journal of Econometrics*, **227**, 408-428. <https://doi.org/10.1016/j.jeconom.2021.03.002>
- [30] Hastie, T. and Tibshirani, R. (1993) Varying-Coefficient Models. *Journal of the Royal Statistical Society. Series B (Methodological)*, **55**, 757-796. <https://doi.org/10.1111/j.2517-6161.1993.tb01939.x>
- [31] Ramsay, J.O. and Silverman, B.W. (1997) *Functional Data Analysis*. Springer-Verlag, New York. <https://doi.org/10.1007/978-1-4757-7107-7>
- [32] Brumback, B.A. and Rice, J.A. (1998) Smoothing Spline Models for the Analysis of Nested and Crossed Samples of Curves. *Journal of the American Statistical Association*, **93**, 961-976. <https://doi.org/10.1080/01621459.1998.10473755>
- [33] Huang, J., Wu, C.O. and Zhou, L. (2004) Polynomial Spline Estimation and Inference for Varying Coefficient Models with Longitudinal Data. *Statistica Sinica*, **14**, 763-788.
- [34] Aneiros-Perez, G. and Vieu, P. (2008) Nonparametric Time Series Prediction: A Semi-Functional Partial Linear Modeling. *Journal of Multivariate Analysis*, **99**, 834-857. <https://doi.org/10.1016/j.jmva.2007.04.010>
- [35] Fan, J. and Gijbels, I. (1996) *Local Polynomial Modelling and Its Applications*. Chapman & Hall, London.
- [36] De Boor, C (2001) *A Practical Guide to Splines*. Springer-Verlag, New York.
- [37] Tian, L., Zucker, D. and Wei, L.J. (2005) On the Cox Model with Time-Varying Regression Coefficients. *Journal of the American Statistical Association*, **100**, 172-183. <https://doi.org/10.1198/016214504000000845>
- [38] Shumway, R.H. and Stoffer, D. (2017) *Time Series Analysis and Its Applications*. 4th Edition, Springer-Verlag, New York. <https://doi.org/10.1007/978-3-319-52452-8>

### Appendix: Proof of Theorem 1

We first introduce some notations for the convenience. Denote

$$D_n = D_n(t_0) = \begin{pmatrix} D_{n,1} & D_{n,2} \\ D_{n,2}^T & D_{n,3} \end{pmatrix}, \quad \Sigma_n = \Sigma_n(t_0) = \begin{pmatrix} \Sigma_{n,1} & \Sigma_{n,2} \\ \Sigma_{n,2}^T & \Sigma_{n,3} \end{pmatrix},$$

where

$$D_{n,1} = D_{n,1}(t_0) = \frac{1}{n-p} \sum_{i=p+1}^n G_i G_i^T K_h(t_i - t_0),$$

$$D_{n,2} = D_{n,2}(t_0) = \frac{1}{n-p} \sum_{i=p+1}^n G_i V_i^T \left( \frac{t_i - t_0}{h} \right) K_h(t_i - t_0),$$

$$D_{n,3} = D_{n,3}(t_0) = \frac{1}{n-p} \sum_{i=p+1}^n V_i V_i^T \left( \frac{t_i - t_0}{h} \right)^2 K_h(t_i - t_0),$$

$$\Sigma_{n,1} = \Sigma_{n,1}(t_0) = \frac{h}{(n-p)V_0} \left( \sum_{i=p+1}^n \hat{\varepsilon}_i G_i K_h(t_i - t_0) \right) \left( \sum_{j=p+1}^n \hat{\varepsilon}_j G_j^T K_h(t_j - t_0) \right),$$

$$\Sigma_{n,2} = \Sigma_{n,2}(t_0) = \frac{h}{(n-p)V_0} \left( \sum_{i=p+1}^n \hat{\varepsilon}_i G_i K_h(t_i - t_0) \right) \left( \sum_{j=p+1}^n \hat{\varepsilon}_j V_j^T \left( \frac{t_j - t_0}{h} \right) K_h(t_j - t_0) \right),$$

$$\Sigma_{n,3} = \Sigma_{n,3}(t_0) = \frac{h}{(n-p)V_0} \left( \sum_{i=p+1}^n \hat{\varepsilon}_i V_i \left( \frac{t_i - t_0}{h} \right) K_h(t_i - t_0) \right) \left( \sum_{j=p+1}^n \hat{\varepsilon}_j V_j^T \left( \frac{t_j - t_0}{h} \right) K_h(t_j - t_0) \right).$$

Define

$$T_n = T_n(t_0) = \begin{pmatrix} T_{n,0}(t_0) \\ T_{n,1}(t_0) \end{pmatrix}, \quad T_n^* = T_n^*(t_0) = \begin{pmatrix} T_{n,0}^*(t_0) \\ T_{n,1}^*(t_0) \end{pmatrix},$$

where

$$T_{n,0} = T_{n,0}(t_0) = \frac{1}{n-p} \sum_{i=p+1}^n G_i K_h(t_i - t_0) Y_i,$$

$$T_{n,1} = T_{n,1}(t_0) = \frac{1}{n-p} \sum_{i=p+1}^n V_i \left( \frac{t_i - t_0}{h} \right) K_h(t_i - t_0) Y_i,$$

$$T_{n,0}^* = T_{n,0}^*(t_0) = \frac{1}{n-p} \sum_{i=p+1}^n G_i K_h(t_i - t_0) \varepsilon_i,$$

$$T_{n,1}^* = T_{n,1}^*(t_0) = \frac{1}{n-p} \sum_{i=p+1}^n V_i \left( \frac{t_i - t_0}{h} \right) K_h(t_i - t_0) \varepsilon_i.$$

Noting the definitions of  $T_n$  and  $T_n^*$ , it is easy to know

$$T_{n,0} - T_{n,0}^* = D_{n,1} \xi_0 + h D_{n,2} \eta_0 + \frac{h^2}{2} D_{n,1}^* \eta_0' + o_p(h^2),$$

$$T_{n,1} - T_{n,1}^* = D_{n,2}^T \xi_0 + h D_{n,3} \eta_0 + \frac{h^2}{2} D_{n,2}^* \eta_0' + o_p(h^2),$$

$$\xi_0 = \xi(t_0) = \{ \alpha_1^T(t_0), \alpha_2^T(t_0), \beta_1^T(t_0), \beta_2^T(t_0) \}^T,$$

$$\eta_0 = \eta(t_0) = \left( (\alpha_1'(t_0))^T, (\beta_1'(t_0))^T \right)^T,$$

$$\begin{aligned} \boldsymbol{\eta}'_0 &= \boldsymbol{\eta}'(t_0) = \left( (\boldsymbol{\alpha}''(t_0))^T, (\boldsymbol{\beta}''_1(t_0))^T \right)^T, \\ \mathbf{D}_{n,1}^* &= \mathbf{D}_{n,1}^*(t_0) = \frac{1}{n-p} \sum_{i=p+1}^n \mathbf{G}_i \mathbf{V}_i^T \left( \frac{t_i - t_0}{h} \right)^2 K_h(t_i - t_0), \\ \mathbf{D}_{n,2}^* &= \mathbf{D}_{n,2}^*(t_0) = \frac{1}{n-p} \sum_{i=p+1}^n \mathbf{V}_i \mathbf{V}_i^T \left( \frac{t_i - t_0}{h} \right)^3 K_h(t_i - t_0). \end{aligned}$$

After a simple combination, we can obtain

$$\mathbf{T}_n - \mathbf{T}_n^* = \mathbf{D}_n \mathbf{H} \boldsymbol{\gamma}_0 + \frac{h^2}{2} \begin{pmatrix} \mathbf{D}_{n,1}^* \boldsymbol{\eta}'(t_0) \\ \mathbf{D}_{n,2}^* \boldsymbol{\eta}'(t_0) \end{pmatrix} + o_p\left((t - t_0)^2\right). \tag{A.1}$$

Note that the equivalent form of formula (4) is:

$$\hat{\boldsymbol{\gamma}} = \mathbf{H}^{-1} \mathbf{D}_n^{-1} \mathbf{T}_n. \tag{A.2}$$

By (A.1) and (A.2), it is easy to know

$$\mathbf{H}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0) = \mathbf{D}_n^{-1} \mathbf{T}_n^* + \frac{h^2}{2} \mathbf{D}_n^{-1} \begin{pmatrix} \mathbf{D}_{n,1}^* \boldsymbol{\eta}'(t_0) \\ \mathbf{D}_{n,2}^* \boldsymbol{\eta}'(t_0) \end{pmatrix} + o_p(h^2). \tag{A.3}$$

Next, we discuss the large sample properties of  $\mathbf{T}_n^*$ . To facilitate the derivation, the following notation is introduced:  $\sigma^2(t_0, \mathbf{g}) = \text{Var}(Y | t = t_0, \mathbf{G} = \mathbf{g})$ ,  $u_i = \varepsilon_i / \sigma(t, \mathbf{G}_i)$ . For  $p+1 \leq k, l \leq n$ ,  $t, s \in [0, 1]$ , let

$$\begin{aligned} \mathbf{R}_{i,k}^{(1)}(t, s) &= \text{Cov}(\mathbf{G}_i u_i \sigma(t, \mathbf{G}_i), \mathbf{G}_k u_k \sigma(s, \mathbf{G}_k)), \\ \mathbf{R}_{i,k}^{(2)}(t, s) &= \text{Cov}(\mathbf{G}_i u_i \sigma(t, \mathbf{G}_i), \mathbf{V}_k u_k \sigma(s, \mathbf{G}_k)), \\ \mathbf{R}_{i,k}^{(3)}(t, s) &= \text{Cov}(\mathbf{V}_i u_i \sigma(t, \mathbf{G}_i), \mathbf{V}_k u_k \sigma(s, \mathbf{G}_k)). \end{aligned}$$

It can be seen from the above that for a stationary process  $\mathbf{R}_{l,l+k}^{(i)}(t, t) = \mathbf{R}_k^{(i)}(t), i = 1, 2, 3$ . Using a similar derivation from Cai (2007), we obtain

$$\begin{aligned} &nh \text{Var}(\mathbf{T}_{n,0}^*(t_0)) \\ &= \frac{nh}{(n-p)^2} \sum_{p+1 \leq i, j \leq n} \text{Cov}(\mathbf{G}_i u_i \sigma(t_i, \mathbf{G}_i), \mathbf{G}_j u_j \sigma(t_j, \mathbf{G}_j)) K_h(t_i - t_0) K_h(t_j - t_0) \\ &\approx \frac{h}{n-p} \sum_{i=p+1}^n E(\mathbf{G}_i \mathbf{G}_i^T u_i^2 \sigma^2(t_i, \mathbf{G}_i)) K_h^2(t_i - t_0) \\ &\quad + 2 \frac{h}{n-p} \sum_{p+1 \leq i < j \leq n} \text{Cov}(\mathbf{G}_i u_i \sigma(t_i, \mathbf{G}_i), \mathbf{G}_j u_j \sigma(t_j, \mathbf{G}_j)) K_h(t_i - t_0) K_h(t_j - t_0) \\ &\approx \frac{h}{n-p} \sum_{i=p+1}^n \mathbf{R}_{i,i}^{(1)}(t_i, t_i) K_h^2(t_i - t_0) \\ &\quad + \frac{2h}{n-p} \sum_{p+1 \leq i < j \leq n} \mathbf{R}_{i,j}^{(1)}(t_i, t_j) K_h(t_i - t_0) K_h(t_j - t_0) \\ &\approx \nu_0 \mathbf{R}_0^{(1)}(t_0) + 2\nu_0 \sum_{k=1}^{\infty} \mathbf{R}_k^{(1)}(t_0). \end{aligned} \tag{A.4}$$

Thus, we have  $nh \text{Var}(\mathbf{T}_{n,0}^*(t_0)) \rightarrow \boldsymbol{\Sigma}_1(t_0) = \left( \mathbf{R}_0^{(1)}(t_0) + 2 \sum_{k=1}^{\infty} \mathbf{R}_k^{(1)}(t_0) \right) \nu_0$ . Si-

ilarly, we obtain

$$\begin{aligned}
 nh\text{Var}(\mathbf{T}_{n,1}^*(t_0)) &\rightarrow \Sigma_3(t_0) = \left( \mathbf{R}_0^{(3)}(t_0) + 2 \sum_{k=1}^{\infty} \mathbf{R}_k^{(3)}(t_0) \right) \mathbf{v}_0, \\
 nh\text{Cov}(\mathbf{T}_{n,0}^*(t_0), \mathbf{T}_{n,1}^*(t_0)) &\rightarrow \Sigma_2(t_0) = \left( \mathbf{R}_0^{(2)}(t_0) + 2 \sum_{k=1}^{\infty} \mathbf{R}_k^{(2)}(t_0) \right) \mathbf{v}_0. \quad (\text{A.5})
 \end{aligned}$$

Next, we derive the properties of  $\mathbf{D}_n(t_0)$ . From the approximation of the Riemann summation of definite integrals, it is known that when  $0 < t_0 < 1$ , we have

$$\begin{aligned}
 \mathbf{D}_{n,1}(t_0) &= \frac{1}{n-p} \sum_{i=p+1}^n \mathbf{G}_i \mathbf{G}_i^T K_h(t_i - t_0) \approx \mathbf{\Omega}_1 \int_{t_0/h}^{(1-t_0)/h} K(u) du \approx \mathbf{\Omega}_1 \mu_0, \\
 \mathbf{D}_{n,2}(t_0) &= \frac{1}{n-p} \sum_{i=p+1}^n \mathbf{G}_i \mathbf{V}_i^T \left( \frac{t_i - t_0}{h} \right) K_h(t_i - t_0) \approx \mathbf{\Omega}_2 \int_{t_0/h}^{(1-t_0)/h} u K(u) du \approx \mathbf{\Omega}_2 \mu_1, \\
 \mathbf{D}_{n,3}(t_0) &= \frac{1}{n-p} \sum_{i=p+1}^n \mathbf{V}_i \mathbf{V}_i^T \left( \frac{t_i - t_0}{h} \right)^2 K_h(t_i - t_0) \approx \mathbf{\Omega}_3 \int_{t_0/h}^{(1-t_0)/h} u^2 K(u) du \approx \mathbf{\Omega}_3 \mu_2, \\
 \mathbf{D}_{n,1}^*(t_0) &= \frac{1}{n-p} \sum_{i=p+1}^n \mathbf{G}_i \mathbf{V}_i^T \left( \frac{t_i - t_0}{h} \right)^2 K_h(t_i - t_0) \approx \mathbf{\Omega}_2 \int_{t_0/h}^{(1-t_0)/h} u^2 K(u) du \approx \mathbf{\Omega}_2 \mu_2, \\
 \mathbf{D}_{n,2}^*(t_0) &= \frac{1}{n-p} \sum_{i=p+1}^n \mathbf{V}_i \mathbf{V}_i^T \left( \frac{t_i - t_0}{h} \right)^3 K_h(t_i - t_0) \approx \mathbf{\Omega}_3 \int_{t_0/h}^{(1-t_0)/h} u^3 K(u) du \approx \mathbf{\Omega}_3 \mu_3. \quad (\text{A.6})
 \end{aligned}$$

Since the kernel function  $K(\cdot)$  is symmetrical,  $\mu_0 = 1$ ,  $\mu_1 = \mu_3 = 0$ . Therefore,

$$\mathbf{B}_n(t_0) = \frac{h^2}{2} \mathbf{D}_n^{-1} \begin{pmatrix} \mathbf{D}_{n,1}^* \boldsymbol{\eta}'(t_0) \\ \mathbf{D}_{n,2}^* \boldsymbol{\eta}'(t_0) \end{pmatrix} \rightarrow \frac{h^2}{2} \begin{pmatrix} \mathbf{\Omega}_1^{-1} \mathbf{\Omega}_2 \mu_2 \boldsymbol{\eta}'(t_0) \\ \mathbf{0} \end{pmatrix} = \mathbf{b} = \mathbf{b}(t_0), \quad (\text{A.7})$$

when  $n \rightarrow \infty$ . Combining the results of (A.3)-(A.7), it is not difficult to obtain the conclusion of Theorem 1.