Proc. Indian Acad. Sci. (Math. Sci.) (2021) 131:21
https://doi.org/10.1007/s12044-021-00614-x

Statistical inference for single-index-driven varying-coefficient time series model with explanatory variables

JINGWEN HUANG[1] and DEHUI WANG[2,*]

[1]School of Mathematics, Jilin University, Changchun 130012, China
[2]School of Economics, Liaoning University, Shenyang 110036, China
*Corresponding author. E-mail: wangdehui@lnu.edu.cn

MS received 2 April 2019; revised 30 June 2020; accepted 30 September 2020

Abstract. Varying-coefficient time series model has gained wide attention because of its flexibility and interpretability. This article considers the single-index-driven varying-coefficient time series model with explanatory variables. It can be seen as a generalization of the autoregressive model with explanatory variables by changing the coefficient of autoregressive part to a single-index structure, or a generalization of the classical linear model by putting a single-index-driven varying-coefficient autoregressive structure into the model. We adopt local linear smoothing and least square methods separately based on an iterative algorithm to estimate unknown link function and parameters. The estimator for the nonparametric part is proved to be asymptotically normal at any fixed point, and the estimators for the parametric part are derived to be asymptotically normal as well. Some simulation studies are carried out to illustrate the model and finite sample performances of the estimators, and a real data example is also conducted.

Keywords. Varying-coefficient time series model; single-index-driven; local linear smoothing; asymptotic normality; bandwidth selection.

Mathematics Subject Classification. Primary: 62F12, 62M10; Secondary: 60G10.

1. Introduction

As one of the branches of statistics, time series analysis plays an important role in various fields, such as finance, ecology, economics, medical science, etc. Due to its feasibility and efficiency in dealing with records that are collected over time, in the last several decades, time series analysis has experienced many exciting developments. And we focus on some advancement in semiparametric models for time series.

It is well known that Yule [30] proposed the second-order autoregressive model to analyze Wolfer's sunspot numbers, which is the foundation and origin of the modern time series. Besides, the autoregressive model with explanatory variables [1,19] has been studied widely and can be applied to many economic relationships. With the contribution of Box and Jenkins [3], the framework of autoregressive moving average (ARMA) model is established completely. The ARMA model and its variations are popular in handling time series data. However, linear models have limitations and are often unrealistic in applications. It

Published online: 24 June 2021

is a must to explore new models to explain nonlinear features. As we have seen, Engle [10] introduced the autoregressive conditional heteroskedastic (ARCH) model to estimate the means and variances of inflation in the UK, Bollerslev [2] defined the generalized autoregressive conditional heteroskedastic (GARCH) model by generalizing the ARCH model, Tong [23] first mentioned the threshold autoregressive (TAR) model and described systematically in [24], Granger and Andersen [14] discussed the now familiar bilinear time series model, Haggan and Ozaki [15] established the exponential autoregressive (EXPAR) model, Nicholls and Quinn [22] had investigated the random coefficient autoregressive (RCA) model in detail, and so on, which all belong to parametric nonlinear time series analysis. On the other hand, to overcome the difficulty in choosing suitable models in real applications and to take full advantage of advances in computing and nonparametric regression analysis, there have been many articles concerning the nonparametric nonlinear time series, such as the functional coefficient autoregressive (FAR) model of [7], the nonlinear additive autoregressive model with exogenous variables (NAARX) of [8], the functional coefficient linear (FL) model under dependence of [26], the functional-coefficient regression models for nonlinear time series of [5], the additive model for dependent data of [4], the vector functional coefficient autoregressive (VFCAR) model of [16], and the multivariate functional coefficient model for vector time series data of [18]. Yet, the worst thing in nonparametric modelling is probably the so-called "curse of dimensionality", which renders the accuracy of estimation unreasonable when the dimension of the covariate is high. To cope with the "curse of dimensionality" problem, many efficient models to reduce dimension have been studied, the main of which are the semiparametric models, including the single-index models.

The class of single-index model was first introduced by Ichimura [17] and has gained extensive attention in the literature, see for example, [6,25] and the like. Besides, there have been articles where the single-indexing specification is applied to dependent data. Xia and Li [27] analysed the single-index coefficient regression model under dependence, Xia *et al* [28] extended the partially linear single-index model to time series and, Xue and Pang [29] studied the single-index varying-coefficient model in the context of strictly stationary and strongly mixing sequences. In this paper, we will try to combine the single-index structure with the time series model and explore the single-index-driven varying-coefficient time series model with explanatory variables of the following form:

$$Y_t = g_0(\beta_0^\top X_t)Y_{t-1} + \theta_0^\top Z_t + \varepsilon_t, \tag{1.1}$$

where Y_t is the observable time series, $(X_t, Z_t) \in \mathbf{R}^p \times \mathbf{R}^q$ is a covariate, $(\beta_0, \theta_0) \in \mathbf{R}^p \times \mathbf{R}^q$ is an unknown parameter vector with $\|\beta_0\| = 1$ (where $\|\cdot\|$ denotes the Euclidean metric) with the first non-zero component positive for identifiability, $g_0(\cdot)$ is an unknown link function, and the model error ε_t is independent of (X_t, Z_t, Y_{t-1}) with $E(\varepsilon_t) = 0$ and $\mathrm{Var}(\varepsilon_t) = \sigma^2 < \infty$.

Model (1.1) is flexible enough to contain many other important statistical models as special cases. For example, model (1.1) reduces to the functional-coefficient autoregressive model if $p = 1$, $\beta_0 = 1$ and $\theta_0 = 0$. Model (1.1) becomes the functional-coefficient autoregressive model with explanatory variables when $p = 1$ and $\beta_0 = 1$. Moreover, model (1.1) is the single-index-driven varying-coefficient autoregressive model with $\theta_0 = 0$, which is a special case of single-index varying-coefficient model considered by Xue and Pang [29]. Therefore, it is worthwhile to deal with this kind of semiparametric model for time series.

Model (1.1) is a natural combination and generalization of the single-index-driven varying-coefficient time series model and the classical linear model. In this model, the autoregressive component is driven by a single-index structure to show the impact of covariate X_t on response variate Y_t, while the linear part represents that of another covariate Z_t. The setting like this is acceptable due to the fact that the response variable of interest is frequently suspected to be influenced by large amounts of covariates in dealing with some real problems. Clearly, the model is easy to interpret in practical applications because of its features of the traditional linear model, the autoregressive model, the single-index model and the varying-coefficient model. However, literature like the model (1.1) is limited, let alone the articles concerning time series. So far, we just know that Feng and Xue [13] firstly obtained the partially linear single-index varying coefficient model and studied the model based on the strictly stationary and strongly mixing sequence, and Zhao *et al.* [32] developed a stepwise estimation procedure for it considering the same sequence. Consequently, the work to develop the statistical inference of unknown function and parameters for the single-index-driven varying-coefficient time series model with explanatory variables is meaningful.

In this paper, we provide the estimation of $g_0(\cdot)$, β_0 and θ_0, and establish the asymptotic theory of the estimators. The approach is first to estimate the unknown functional coefficient $g_0(\cdot)$ and its derivative $g_0'(\cdot)$ employing the local linear method by pretending the unknown parameter (β_0, θ_0) to be known. Then, we use the least square method to obtain the estimators of unknown index parameter β_0 and unknown linear parameter θ_0. In addition, the asymptotic normality for the estimators are derived to illustrate the efficiency of the estimation procedure. Details can be found below.

The rest of this article is organized as follows. In Section 2, we elaborate on the estimation methods and the corresponding algorithm. Section 3 gives some large sample results of the proposed estimators. Section 4 discusses the bandwidth selection problem. In Section 5, simulation studies are provided to evaluate the estimation methodology. A real data example is presented in Section 6. And the concluding remarks are made in Section 7. For the sake of presentation, the proofs are all relegated to the Appendix.

2. Methodology

Suppose that $\{Y_t, X_t, Z_t\}_{t=1}^T$ is a sample generated from model (1.1) and we are ready to estimate the function $g_0(\cdot)$ and the parameter (β_0, θ_0).

2.1 *Estimation*

To estimate the unknown link function $g_0(\cdot)$, we can utilize the local linear regression technique in [11]. For any W in a small neighborhood of w, it follows from a Taylor expansion that $g_0(W) \approx \sum_{j=0}^n \frac{g_0^{(j)}(w)}{j!}(W-w)^j$. Based on the idea of local polynomial estimation, we choose the typical order $n = 1$ to locally approximate the smooth function $g_0(\cdot)$. That is,

$$g_0(W) \approx g_0(w) + g_0'(w)(W-w) \triangleq a + b(W-w).$$

Then for given parameters β and θ, find the initial estimators $\hat{g}(w;\beta,\theta) = \hat{a}$ and $\hat{g}'(w;\beta,\theta) = \hat{b}$ by minimizing the sum of weighted squares

$$\sum_{t=1}^{T}\{Y_t - [a + b(\beta^\top X_t - w)]Y_{t-1} - \theta^\top Z_t\}^2 K_h(\beta^\top X_t - w), \tag{2.1}$$

with respect to a and b, where $K_h(\cdot) = K(\cdot/h)/h$, $K(\cdot)$ is a kernel function and $h = h(T) > 0$ is a bandwidth. According to the theory of least squares, we get the estimators as shown below:

$$(\hat{g}(w;\beta,\theta), h\hat{g}'(w;\beta,\theta))^\top = R_T^{-1}(w;\beta,\theta)\eta_T(w;\beta,\theta), \tag{2.2}$$

where

$$R_T(w;\beta,\theta) = \begin{pmatrix} R_{T,0}(w;\beta,\theta) & R_{T,1}(w;\beta,\theta) \\ R_{T,1}(w;\beta,\theta) & R_{T,2}(w;\beta,\theta) \end{pmatrix}$$

and

$$\eta_T(w;\beta,\theta) = \begin{pmatrix} \eta_{T,0}(w;\beta,\theta) \\ \eta_{T,1}(w;\beta,\theta) \end{pmatrix}$$

with

$$R_{T,j}(w;\beta,\theta) = \frac{1}{T}\sum_{t=1}^{T} Y_{t-1}^2 \left(\frac{\beta^\top X_t - w}{h}\right)^j K_h(\beta^\top X_t - w),\ \ j = 0, 1, 2, 3$$

and

$$\eta_{T,j}(w;\beta,\theta) = \frac{1}{T}\sum_{t=1}^{T}(Y_t - \theta^\top Z_t)Y_{t-1}\left(\frac{\beta^\top X_t - w}{h}\right)^j K_h(\beta^\top X_t - w),\ \ j = 0, 1.$$

Now that we have got $\hat{g}(\cdot;\beta,\theta)$, the estimator for (β_0, θ_0) can be obtained by minimizing the sum of squared errors

$$\sum_{t=1}^{T}[Y_t - \hat{g}(\beta^\top X_t;\beta,\theta)Y_{t-1} - \theta^\top Z_t]^2,\ \text{subject to } \beta^\top\beta = 1, \tag{2.3}$$

with respect to β and θ.

With the estimator $(\hat{\beta}, \hat{\theta})$, we can define the final estimator of $g_0(\cdot)$ as $\hat{g}(w;\hat{\beta},\hat{\theta})$ and the estimator of σ^2 as

$$\hat{\sigma}^2 = \frac{1}{T}\sum_{t=1}^{T}[Y_t - \hat{g}(\hat{\beta}^\top X_t;\hat{\beta},\hat{\theta})Y_{t-1} - \hat{\theta}^\top Z_t]^2.$$

2.2 *Implementation*

In fact, the estimation procedure for $g_0(\cdot)$, β_0 and θ_0 is an iterative algorithm, which can be decomposed as follows:

Step 0. The initialization step is to select an initial value $(\hat{\beta}_1, \hat{\theta})$ by fitting a parametric generalized linear model, or fitting a linear model by taking $g_0(\cdot)$ as the identity function. Set $\hat{\beta} = \hat{\beta}_1/\|\hat{\beta}_1\|$ to make sure that $\|\hat{\beta}\| = 1$. Or, just specify the initial estimator as some reasonable numerical value.

Step 1. Calculate $\hat{g}(\cdot;\hat{\beta},\hat{\theta})$ and $\hat{g}'(\cdot;\hat{\beta},\hat{\theta})$ by expression (2.2), with $(\hat{\beta},\hat{\theta})$ replacing (β,θ) in all formulae.

Step 2. Minimize

$$\sum_{t=1}^{T}\{Y_t-[\hat{g}(\hat{\beta}^\top X_t;\hat{\beta},\hat{\theta})+\hat{g}'(\hat{\beta}^\top X_t;\hat{\beta},\hat{\theta})(\beta^\top X_t-\hat{\beta}^\top X_t)]Y_{t-1}-\theta^\top Z_t\}^2 \tag{2.4}$$

with respect to β and θ to get updated $(\hat{\beta},\hat{\theta})$. It is clear that expression (2.4) is an approximation of expression (2.3) to make the iterative process available in practice.

Step 3. Repeat Step 1 and Step 2 until convergence and the iterative estimator of (β,θ) can be obtained.

Step 4. With the estimated value $(\hat{\beta},\hat{\theta})$ from Step 3, we can get the final estimator of $g_0(\cdot)$ as $\hat{g}(\cdot;\hat{\beta},\hat{\theta})=\hat{a}$ by performing the Step 1.

Remark 2.1. It should be noticed that the expression (2.1) and (2.3) are similar superficially, while the essence behinds them is quite different. We search the minimum point of the expression (2.1) locally to get the estimators of $g_0(\cdot)$ and $g_0'(\cdot)$, but the estimator of (β,θ) is obtained by globally minimize the expression (2.3).

Remark 2.2. To perform the algorithm, one needs to consider the kernel function $K(\cdot)$ and the bandwidth h. Just as Fan and Gijbels [11] and Fan and Yao [12] put it, the choice of the kernel function is not important to the estimators theoretically and empirically, so, in this paper, the Epanechnikov kernel function, which is widely recommended, is used directly. As for the bandwidth which is rather critical, it will be discussed carefully in Section 4.

3. Asymptotic properties

Next, the asymptotic behaviors of the estimators described in Section 2 will be studied. To state them, the following regularity conditions are essential:

- (C1) $\{Y_t, X_t, Z_t\}_{t=1}^T$ is a strictly stationary and strongly mixing sequence with $\alpha(T)=O(\rho^T)(0<\rho<1)$ as the mixing coefficient.
- (C2) The conditional density functions $f_{W_1|Y_1}(w_1|y_1)$, $f_{W_1,W_t|Y_1,Y_t}(w_1,w_t|y_1,y_t)$, $f_{W_1|Y_1,Z_1}(w_1|y_1,z_1)$ and $f_{W_1,W_t|Y_1,Y_t,Z_1,Z_t}(w_1,w_t|y_1,y_t,z_1,z_t)$ are bounded for all $t>1$, where $W_t=\beta_0^\top X_t$ for $t\geq 1$.
- (C3) For some $l>0$, $E|Y_t|^l<\infty$, $E\|X_t\|^l<\infty$, $E\|Z_t\|^l<\infty$, and $E|\varepsilon_t|^l<\infty$.
- (C4) The density function $f_0(w)$ of $\beta_0^\top X_t$ satisfies the Lipschitz condition of order 1 on $\mathcal{W}$, and is bounded away from 0 for $w\in\mathcal{W}$.
- (C5) The kernel function $K(x)$ with compact support $\{x:-1<x<1\}$ is a symmetric probability density function with a bounded derivative and finite moments of all orders, i.e., $\int|x|^jK(x)\mathrm{d}x\leq A<\infty$. Denote that $\mu_j=\int x^jK(x)\mathrm{d}x$ $(j=0,1,2,3)$ and $\nu_j=\int x^jK^2(x)\mathrm{d}x$ $(j=0,1,2,3)$.
- (C6) The link function $g_0(w)$ has bounded and continuous derivatives up to order 2 on $\mathcal{W}$.
- (C7) The conditional expectation $E(X_t|\beta_0^\top X_t=w)$ and $E(Z_t|\beta_0^\top X_t=w)$ are Lipschitz continuous of order 1 on $\mathcal{W}$.

(C8) $Th^2/\log^2 T \to \infty$, $Th^4 \log T \to 0$, $\log T/Th^3 \to 0$.

Remark 3.1. The above conditions are commonly used in the literature of nonparametric analysis. (C1) and (C2) are made just as in [27], and the mixing coefficient can be weakened to $\alpha(T) = O(T^{-\kappa})$ for some $\kappa > 0$. (C4), (C5) and (C6) are the general smoothness conditions for the single-index models. (C7) is to ensure the existence of the limiting variance for the estimator $(\hat{\beta}, \hat{\theta})$. (C3) and (C8) are necessary conditions for the asymptotic properties of the estimators.

Then, the finite sample performances of estimators both for the unknown function $g_0(\cdot)$ and the unknown parameter (β, θ) are presented. From the results shown below we can see that the estimators satisfy the usual asymptotic properties and are efficient. Related proofs can be seen in the Appendix.

Firstly, the result for the estimator of unknown function $g_0(\cdot)$, $\hat{g}(\cdot)$, is derived. Obviously, the following theorem demonstrates the fact that the estimator $\hat{g}(\cdot)$ is asymptotic efficient as the sample size $T \to \infty$ and the bandwidth $h = h(T) \to 0$ $(T \to \infty)$.

Theorem 3.1. *When β and θ are known constants or estimated to the order $O_P(T^{-1/2})$, under the conditions* $(C1)$–$(C6)$ *and* $(C8)$, *we have*

$$\sqrt{Th}[\hat{g}(w; \hat{\beta}, \hat{\theta}) - g_0(w) - \frac{1}{2}h^2 g_0''(w)\mu_2] \xrightarrow{L} N(0, \gamma^{-1}\nu_0\sigma^2 f_0^{-1}(w)),$$

where $\gamma = E(Y_{t-1}^2)$.

Besides, for the estimator of unknown parameter (β_0, θ_0), $(\hat{\beta}, \hat{\theta})$, we have the following result as the sample size $T \to \infty$ and the bandwidth $h = h(T) \to 0$ $(T \to \infty)$.

Theorem 3.2. *Assuming that the estimators $\hat{\beta}$ and $\hat{\theta}$ are in a $\sqrt{T}$ neighborhood of β_0 and θ_0, that is, $\hat{\beta} - \beta_0 = O_P(T^{-1/2})$ and $\hat{\theta} - \theta_0 = O_P(T^{-1/2})$, then, under the conditions* $(C1)$–$(C8)$, *we have*

$$\sqrt{T}\begin{pmatrix}\hat{\beta} - \beta_0 \\ \hat{\theta} - \theta_0\end{pmatrix} \xrightarrow{L} N(0, \sigma^2 B^{-1} V (B^{-1})^\top),$$

where

$$V = \begin{pmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{pmatrix}, \quad A_t = \begin{pmatrix} g_0'(\beta_0^\top X_t) Y_{t-1} X_t \\ Z_t \end{pmatrix},$$

$$B = E\left[A_t A_t^\top\right] - E\left[A_t \begin{pmatrix} g_0'(\beta_0^\top X_t) Y_{t-1} E(X_t|\beta_0^\top X_t) \\ \gamma^{-1} Y_{t-1} E(Y_{t-1}) E(Z_t|\beta_0^\top X_t) \end{pmatrix}^\top\right],$$

with

$$V_{11} = \gamma E\{[g_0'(\beta_0^\top X_t)]^2 [X_t - E(X_t|\beta_0^\top X_t)][X_t - E(X_t|\beta_0^\top X_t)]^\top\},$$
$$V_{12} = E(Y_{t-1}) E\{g_0'(\beta_0^\top X_t)[X_t - E(X_t|\beta_0^\top X_t)][Z_t - E(Z_t|\beta_0^\top X_t)]^\top\},$$

$$V_{21} = E(Y_{t-1})E\{g_0'(\beta_0^\top X_t)[Z_t - E(Z_t|\beta_0^\top X_t)][X_t - E(X_t|\beta_0^\top X_t)]^\top\},$$
$$V_{22} = \{1-\gamma^{-1}[E(Y_{t-1})]^2\}E(Z_t Z_t^\top) + \gamma^{-1}[E(Y_{t-1})]^2 E\{[Z_t - E(Z_t|\beta_0^\top X_t)] [Z_t - E(Z_t|\beta_0^\top X_t)]^\top\}.$$

4. Bandwidth selection

Choosing an optimal bandwidth is of significance in the nonparametric estimation procedure and we would like to describe it specially here. The most simple and intuitive way is to select the bandwidth for $(\hat{g}(\cdot), \hat{g}'(\cdot))$ subjectively by visually examining the estimators and the fitting curve. We can also select the bandwidth automatically by minimizing a suitable risk driven by data. For model (1.1), the least squares delete-one cross-validation (CV) method and the modified multi-fold cross-validation (MCV) criterion can be employed to obtain the proper bandwidth.

The CV statistic is given by

$$cv(h) = \frac{1}{T}\sum_{t=1}^{T}[Y_t - \hat{g}_h^{(-t)}(X_t^\top \hat{\beta}_h^{(-t)})Y_{t-1} - Z_t^\top \hat{\theta}_h^{(-t)}]^2, \tag{4.1}$$

where $\hat{g}_h^{(-t)}(\cdot)$, $\hat{\beta}_h^{(-t)}$, and $\hat{\theta}_h^{(-t)}$ are respectively the "leave-one-out" version of $\hat{g}(\cdot)$, $\hat{\beta}$ and $\hat{\theta}$. The optimal bandwidth $\hat{h}_{\text{opt}}$ in this case is the one minimizing the expression (4.1).

The MCV statistic is given by

$$\text{mcv}(h) = \frac{1}{K}\sum_{k=1}^{K}\frac{1}{m}\sum_{t=T-km+1}^{T-km+m}[Y_t - \hat{g}_h^{(T-km)}(X_t^\top \hat{\beta}_h^{(T-km)})Y_{t-1}^* - Z_t^\top \hat{\theta}_h^{(T-km)}]^2, \ T > Km, \tag{4.2}$$

where K and m are given positive integers representing the number and the length of subseries for the sample $\{Y_t, X_t, Z_t\}_{t=1}^T$ respectively, and $\hat{g}_h^{(T-km)}(\cdot)$, $\hat{\beta}_h^{(T-km)}$, $\hat{\theta}_h^{(T-km)}$ are estimated using the subseries $\{Y_t, X_t, Z_t\}_{t=1}^{T-km}$ with the bandwidth $h[T/(T-km)]^{1/5} \propto (T-km)^{-1/5}$. Minimize expression (4.2) with respect to h and we acquire the optimal bandwidth $\hat{h}_{\text{opt}}$.

What should be mentioned is that, in the algorithm introduced above, Step 1 to Step 3 are aimed at estimating the parameter (β, θ), while the Step 4 is to get the estimator of function $g_0(\cdot)$. So, the different bandwidths, denoted as h_1 and h_2, are chosen separately for Step 1 and Step 4. For the CV method, according to Carroll *et al.* [6] and Lin *et al* [20], the recommended bandwidths are $\hat{h}_1 = \hat{h}_{\text{opt}} \times T^{1/5} \times T^{-1/3} = \hat{h}_{\text{opt}} \times T^{-2/15}$, $\hat{h}_2 = \hat{h}_{\text{opt}}$. When it comes to the MCV criterion, $\hat{h}_1 = \hat{h}_{\text{opt}} \times T^{-1/20} \times (\log T)^{-1/2}$ and $\hat{h}_2 = \hat{h}_{\text{opt}}$ are suggested by Xue and Pang [29].

Remark 4.1. It should be pointed out that Y_{t-1}^* in expression (4.2) is not equal to the true value of Y_{t-1}, and should be forecasted via recursive calculations.

Remark 4.2. According to Fan and Yao [12], the data-driven bandwidth selectors designed for independent data, such as the ordinary CV method, perform poorly for the dependent data because of the correlation structure. So it is better to choose the MCV method to select bandwidth for model (1.1).

5. Simulation studies

In this section, some simulation examples are considered to support the model (1.1), the iterative algorithm, and the asymptotic performances of estimators obtained previously. Furthermore, the bias (Bias), the standard deviation (SD), the mean squared error (MSE), the Q–Q plot and the probability density plot are used to assess the estimator of unknown parameter (β_0, θ_0). And the estimator of unknown function $g_0(\cdot)$ is assessed by the root mean squared error (RMSE), where RMSE $= \left[\frac{1}{T_{\text{grid}}}\sum_{t=1}^{T_{\text{grid}}}(\hat{g}(w_t) - g_0(w_t))^2\right]^{1/2}$, and $\{w_t\}_{t=1}^{T_{\text{grid}}}$ are some regular grid points.

The observations are generated from the simulated model of the form

$$Y_t = g_0(\beta_{01}X_{t1} + \beta_{02}X_{t2})Y_{t-1} + \theta_{01}Z_{t1} + \theta_{02}Z_{t2} + \varepsilon_t.$$

Example 1. $\beta_0 = (\beta_{01}, \beta_{02})^\top = (1/2, \sqrt{3}/2)^\top$, $\theta_0 = (\theta_{01}, \theta_{02})^\top = (1, 2)^\top$, $g_0(w) = w^2$, $X_t = (X_{t1}, X_{t2})^\top$ is a two-dimensional covariate with independent components uniformly distributed on $[-0.5, 0.5]^2$, the components of $Z_t = (Z_{t1}, Z_{t2})^\top$ are independent and obey the normal distribution $N(0, 1)$, and $\varepsilon_t \sim N(0, 0.3^2)$.

Example 2. $\beta_0 = (\beta_{01}, \beta_{02})^\top = (1/\sqrt{5}, 2/\sqrt{5})^\top$, $\theta_0 = (\theta_{01}, \theta_{02})^\top = (1.5, 4)^\top$, $g_0(w) = \sin(2\pi w)$, $X_t = (X_{t1}, X_{t2})^\top$ is a two-dimensional random variable with independent uniform $[-0.5, 0.5]$ components, $Z_t = (Z_{t1}, Z_{t2})^\top$ is identical to the multivariate normal distribution $N_2(0, \Sigma)$ with $\Sigma = (\sigma_{ij})_{2\times 2}$ and $\sigma_{ij} = 4 \times 0.5^{|i-j|}$ $(i, j = 1, 2)$, and $\varepsilon_t \sim N(0, 0.6^2)$.

For convenience, we selected suitable bandwidths by subjectively inspecting the estimation accuracy in the simulations. And as stated in Section 2, the Epanechnikov kernel, i.e., $K(x) = 0.75(1 - x^2)$ if $|x| \le 1$ and 0 otherwise, is used. The sample size is set to $T = 100, 200, 300, 400, 500$ and all the cases were performed with 500 runs under the R software. Simulated results are reported in the following tables and figures.

Tables 1 and 2 present some main statistical measures for estimators of the index parameter $\beta_0 = (\beta_{01}, \beta_{02})^\top$ and the linear parameter $\theta_0 = (\theta_{01}, \theta_{02})^\top$, which were obtained by 500 replications with different sample sizes. It can be found that the Bias, the SD and the MSE decrease when the sample size T is increasing.

When the sample size is $T = 500$, the histograms and the Q–Q plots in Figures 1 and 2 show that empirically the estimators of parameter (β_0, θ_0) are asymptotically normal. And the fitted curve of functional coefficient $g_0(\cdot)$ is close to the real curve as can be seen from Figure 3. The estimators of σ^2 for Example 1 and Example 2 are 0.08661 and 0.35811, respectively. Besides, it is noted in Table 3 that all the RMSE of $\hat{g}(\cdot)$ are small and decrease as the sample size increases.

Now, to demonstrate the impact of replication, different runs are conducted under the R software, which are denoted as $R = 500, 1000, 2000, 5000, 10000$. The simulation results with the sample size $T = 300$ are given in Tables 4, 5 and 6. It can be concluded from them that the SD, the MSE, the RMSE and the $\hat{\sigma}^2$ are almost similar for different replications, and the results for replication of 500 are comparable with that of 10000.

Generally speaking, the results verify the fact that the estimation procedure used in this paper works well on the condition that the bandwidths are chosen properly.

Table 1. Simulation results for the examples with 500 runs.

	T	$\hat{\beta}_1$				$\hat{\beta}_2$			
		Mean	Bias	SD	MSE	Mean	Bias	SD	MSE
1	100	0.44513	− 0.05487	0.23728	0.05920	0.85987	− 0.00616	0.07945	0.00634
	200	0.49827	− 0.00173	0.12717	0.01614	0.85625	− 0.00978	0.04933	0.00252
	300	0.50685	0.00685	0.09065	0.00825	0.85635	− 0.00967	0.03964	0.00166
	400	0.50323	0.00323	0.08587	0.00737	0.85913	− 0.00690	0.03606	0.00135
	500	0.50262	0.00262	0.06559	0.00430	0.86146	− 0.00457	0.03109	0.00099
2	100	0.43610	− 0.01111	0.04334	0.00200	0.89882	0.00440	0.00764	0.00008
	200	0.44090	− 0.00631	0.00800	0.00010	0.89751	0.00308	0.00394	0.00002
	300	0.44428	− 0.00293	0.00416	0.00003	0.89587	0.00145	0.00206	0.00001
	400	0.44591	− 0.00130	0.00226	0.00001	0.89507	0.00065	0.00113	0.00000
	500	0.44636	− 0.00085	0.00166	0.00000	0.89485	0.00042	0.00083	0.00000

Table 2. Simulation results for the examples with 500 runs.

	T	$\hat{\theta}_1$				$\hat{\theta}_2$			
		Mean	Bias	SD	MSE	Mean	Bias	SD	MSE
1	100	0.99951	−0.00049	0.03078	0.00095	1.99842	−0.00158	0.03192	0.00102
	200	0.99993	−0.00007	0.02119	0.00045	2.00229	0.00229	0.02096	0.00044
	300	0.99932	−0.00068	0.01734	0.00030	2.00040	0.00040	0.01776	0.00031
	400	0.99933	−0.00067	0.01615	0.00026	1.99973	−0.00027	0.01534	0.00024
	500	0.99978	−0.00022	0.01292	0.00017	2.00029	0.00029	0.01330	0.00018
2	100	1.49164	−0.00836	0.12406	0.01543	4.00839	0.00839	0.12281	0.01512
	200	1.49733	−0.00267	0.05896	0.00348	4.00287	0.00287	0.05763	0.00332
	300	1.50168	0.00168	0.03378	0.00114	3.99895	−0.00105	0.03427	0.00117
	400	1.50118	0.00118	0.02125	0.00045	3.99828	−0.00172	0.02011	0.00041
	500	1.50084	0.00084	0.01747	0.00031	3.99824	−0.00176	0.01600	0.00026

6. Real data example

We now illustrate the model and the estimation methods by considering the Beijing multi-site air-quality data set, which is available from the website (https://archive.ics.uci.edu/ml/datasets). According to Zhang *et al.* [31], winter (December–February of the following year) has the highest concentration followed by autumn (September–November) and spring (March–May), and summer (June–August) is the season with the lowest $PM_{2.5}$ on the seasonal variation. For Beijing, each season provides a homogeneous weather pattern. Therefore, the data set observed hourly between December 1, 2016 and February 28, 2017 from pollutant monitoring site Dingling will be used here. The response variable is $Y_t - PM_{2.5}$ (in μg/m^3), which is stationary based on the time series plot and the ADF test,

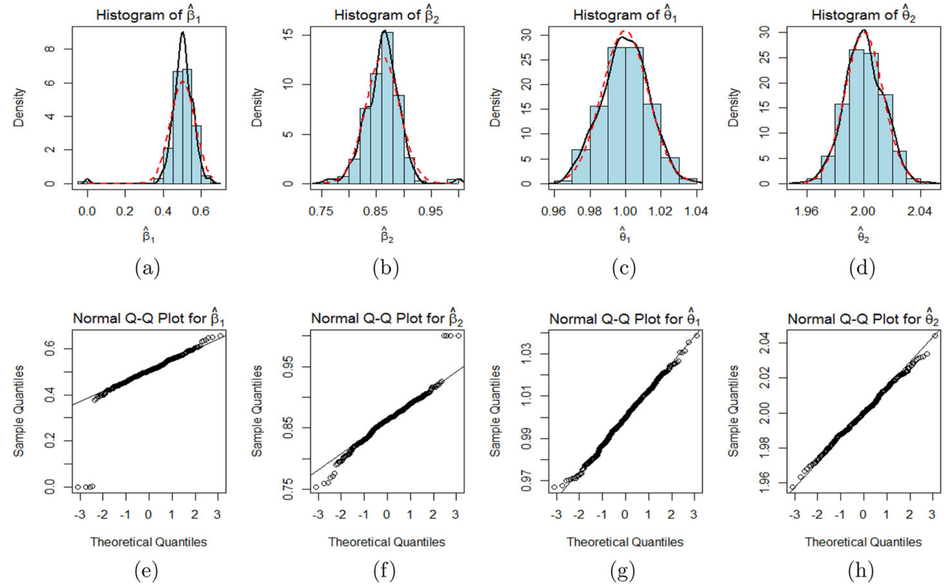


Figure 1. Simulation results for Example 1 when the sample size is $T = 500$. **(a)–(d)** The histograms of 500 estimators for every parameter, the estimated curve of density (solid curve), and the curve of normal density (dashed curve). **(e)–(h)** The Q–Q plots of 500 estimators for every parameter.

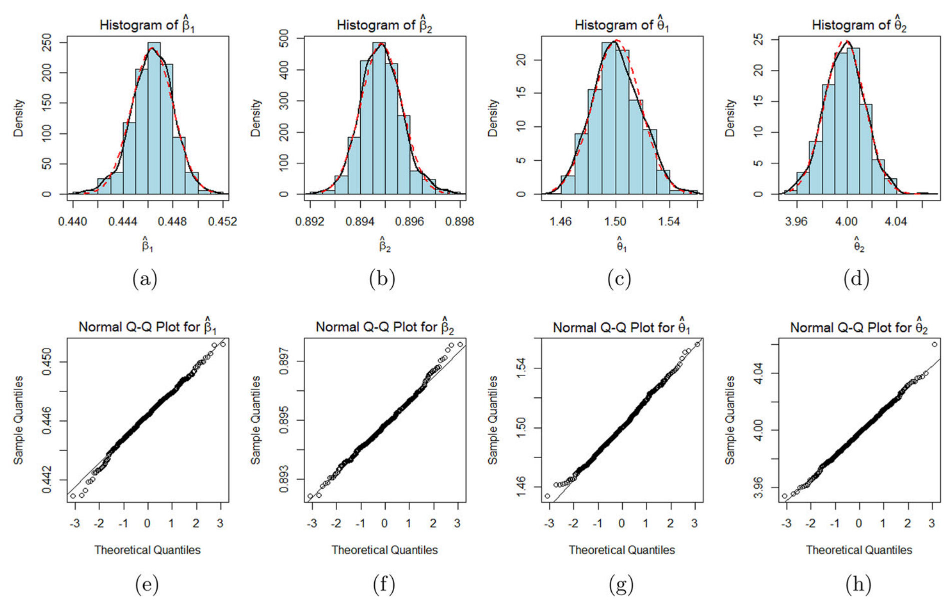
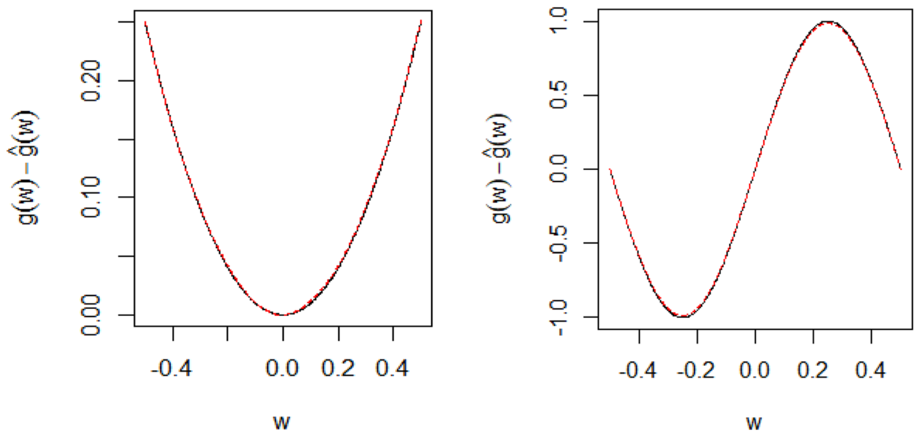


Figure 2. Simulation results for Example 2 when the sample size is $T = 500$. **(a)–(d)** The histograms of 500 estimators for every parameter, the estimated curve of density (solid curve), and the curve of normal density (dashed curve). **(e)–(h)** The Q–Q plots of 500 estimators for every parameter.



Example 1: $g_0(w) = w^2$ Example 2: $g_0(w) = sin(2\pi w)$

Figure 3. The real curve (solid curve) and the fitted curve (dashed curve) of functional coefficient $g_0(\cdot)$ for the examples when the sample size is $T = 500$.

Table 3. Simulation results of $\hat{\sigma}^2$ and RMSE with 500 runs.

	T	100	200	300	400	500
$\hat{\sigma}^2$	1	0.08029	0.08301	0.08422	0.08588	0.08661
	2	0.48397	0.36903	0.35791	0.35571	0.35811
RMSE	1	0.04186	0.02929	0.02532	0.02037	0.01865
	2	0.04037	0.01939	0.01482	0.01331	0.01239

and is autocorrelated as can be seen from the ACF plot in Figure 4(a). Thus, the effect of the $PM_{2.5}$ at time $t-1$ on the $PM_{2.5}$ at time t will be examined by the autoregressive structure. The covariates consist of X_{t1}-sulfur dioxide (SO_2 in $\mu g/m^3$), X_{t2}-nitrogen dioxide (NO_2 in $\mu g/m^3$), X_{t3}-carbon monoxide (CO in $\mu g/m^3$), X_{t4}-ozone (O_3 in $\mu g/m^3$), Z_{t1}-dew point temperature (DEWP in degree Celsius) and Z_{t2}-wind speed (WSPM in m/s), which are selected based on the correlation between the response variable and these covariates. Specifically, the dew point reflects both relative humidity and air temperature, as described in Zhang *et al.* [31]. The main goal in this paper is to study the effect of chemical pollutants (X_{t1}, X_{t2}, X_{t3} and X_{t4}) and environmental factors (Z_{t1} and Z_{t2}) on the $PM_{2.5}$. To avoid the curse of dimensionality, the single-index structure is used to model the relationship between chemical pollutants and the $PM_{2.5}$. As for the environmental factors, the simple linear combination is used. That is, we will analyze the data by fitting the single-index-driven varying-coefficient time series model with explanatory variables, which is $Y_t = g(\beta^\top X_t)Y_{t-1} + \theta^\top Z_t + \varepsilon_t$, with $\beta^\top X_t = \beta_1 X_{t1} + \beta_2 X_{t2} + \beta_3 X_{t3} + \beta_4 X_{t4}$ and $\theta^\top Z_t = \theta_1 Z_{t1} + \theta_2 Z_{t2}$.

In analysis, the data is divided into training data and testing data. The training data from December 1, 2016 to January 31, 2017, which includes about 70 percentage of the data, is to fit the model. And the testing data from January 31, 2017 to February 28, 2017, which includes about 30 percentage of the data, is to predict the value of response variable. To

Table 4. Simulation results of $\hat{\sigma}^2$ and RMSE with $T = 300$.

	R	500	1000	2000	5000	10000
$\hat{\sigma}^2$	1	0.08422	0.08416	0.08469	0.08440	0.08453
	2	0.35791	0.35598	0.35575	0.35616	0.35589
RMSE	1	0.02532	0.02585	0.02550	0.02596	0.02582
	2	0.01482	0.01491	0.01504	0.01501	0.01499

Table 5. Simulation results for the examples with $T = 300$.

	R	$\hat{\beta}_1$				$\hat{\beta}_2$			
		Mean	Bias	SD	MSE	Mean	Bias	SD	MSE
1	500	0.50685	0.00685	0.09065	0.00825	0.85635	−0.00967	0.03964	0.00166
	1000	0.50512	0.00512	0.10523	0.01109	0.85552	−0.01050	0.04333	0.00199
	2000	0.50394	0.00394	0.09939	0.00989	0.85709	−0.00894	0.03968	0.00165
	5000	0.50389	0.00389	0.10891	0.01188	0.85582	−0.01021	0.04260	0.00192
	10000	0.50194	0.00194	0.11441	0.01309	0.85629	−0.00974	0.04169	0.00183
2	500	0.44428	−0.00293	0.00416	0.00003	0.89587	0.00145	0.00206	0.00001
	1000	0.44417	−0.00305	0.00405	0.00003	0.89593	0.00150	0.00201	0.00001
	2000	0.44412	−0.00309	0.00415	0.00003	0.89595	0.00153	0.00206	0.00001
	5000	0.44412	−0.00309	0.00413	0.00003	0.89596	0.00153	0.00205	0.00001
	10000	0.44410	−0.00311	0.00417	0.00003	0.89596	0.00154	0.00207	0.00001

Table 6. Simulation results for the examples with $T = 300$.

	R	$\hat{\theta}_1$				$\hat{\theta}_2$			
		Mean	Bias	SD	MSE	Mean	Bias	SD	MSE
1	500	0.99932	−0.00068	0.01734	0.00030	2.00040	0.00040	0.01776	0.00031
	1000	0.99957	−0.00043	0.01790	0.00032	1.99989	−0.00011	0.01761	0.00031
	2000	0.99959	−0.00041	0.01793	0.00032	1.99952	−0.00048	0.01801	0.00032
	5000	0.99964	−0.00036	0.01791	0.00032	1.99951	−0.00049	0.01763	0.00031
	10000	0.99983	−0.00017	0.01770	0.00031	1.99975	−0.00025	0.01784	0.00032
2	500	1.50168	0.00168	0.03378	0.00114	3.99895	−0.00105	0.03427	0.00117
	1000	1.49998	−0.00002	0.03312	0.00110	3.99923	−0.00077	0.03471	0.00120
	2000	1.49904	−0.00096	0.03416	0.00117	4.00062	0.00062	0.03409	0.00116
	5000	1.50029	0.00029	0.03373	0.00114	4.00032	0.00032	0.03328	0.00111
	10000	1.49961	−0.00039	0.03407	0.00116	4.00007	0.00007	0.03330	0.00111

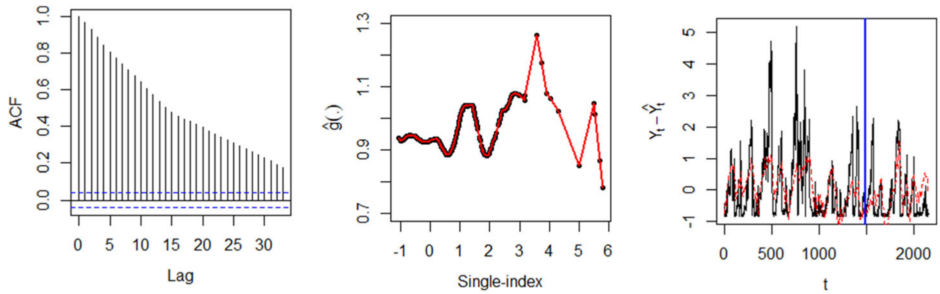


Figure 4. Results for real data example. **(a)** ACF plot of the $PM_{2.5}$ data from December 1, 2016 to February 28, 2017. **(b)** The estimated function curve along with points. **(c)** Sample path of the $PM_{2.5}$ data from December 1, 2016 to February 28, 2017, with the solid curve for real data and the dashed curve for fitted and forecasted data. The vertical line is used to distinguish between the fitted and forecasted data.

eliminate the impact of scales and order of magnitude, the data is standardized, and the MSE is used to evaluate the fitted model. With the MCV criterion for bandwidth selection, we get the estimators of parameters $\hat{\beta} = (0.99773, 0.03027, 0.01067, 0.05915)$ and $\hat{\theta} = (0.02664, -0.03016)$. Clearly, the covariate Z_{t1} has a positive effect on the $PM_{2.5}$, and the result can be explained in reality. Based on Zhang *et al.* [31], we know that the higher dew point reflects more humidity and warmer temperature, which encourages secondary generation of $PM_{2.5}$. Thus, the $PM_{2.5}$ increases as the DEWP increases. Conversely, the covariate Z_{t2} has a negative effect on the $PM_{2.5}$. We know that the location of Dingling is in the north of Beijing. Due to the surroundings of Dingling, it seems that the wind can bring more cooler and drier air from the cleaner north which reduces the $PM_{2.5}$, rather than the warmer and more humid air from the more polluted south that increases the $PM_{2.5}$. So, the conclusion that the stronger wind can result in the lower $PM_{2.5}$ is acceptable. The estimator of function $g(\cdot)$ is presented in Figure 4(b). From the figure, we can see that the $PM_{2.5}$ has the trend to increase with the chemical pollutants index $(\hat{\beta}^{\top} X_t)$ increasing, and it means that the fitted model can somehow capture the feature of the data. Finally, the fitted (MSE: 0.69659) and forecasted (MSE: 0.45832) values for Y_t are shown in Figure 4(c), which also verifies the performance.

7. Conclusions

In recent years, many different and valuable ideas have proposed for the autoregressive model with explanatory variables. Unfortunately, the research on single-index-driven varying-coefficient time series model with explanatory variables is rare. This paper mainly studied the estimation for the model. The estimators of unknown parameters and function are derived and they are proved to be efficient satisfying asymptotic properties. Furthermore, simulated examples and the real data example with proper bandwidths confirm that the methods to estimate and the model are successful and available.

Certainly, the assumption of root-T consistency for parameters when analyzing properties is a nonnegligible limitation, which needs further work. In addition, potential issues of single-index-driven varying-coefficient time series model with explanatory variables include the application to missing data, the problem of testing, forecasting and controlling, etc. What also deserves investigation is the single-index-driven varying-coefficient time series model with functional-coefficient explanatory variables.

Acknowledgements

This work is supported by National Natural Science Foundation of China (Nos 11871028, 11731015, 11901053) and Natural Science Foundation of Jilin Province (No. 20180101216JC).

Appendix

Denote $\mathcal{B} = \{\beta \in \mathbf{R}^p : \|\beta\| = 1$, and the first non-zero component of β is positive$\}$, then β_0 is an inner point of the compact set $\mathcal{B}$. Meanwhile, the assumptions $\|\hat{\beta} - \beta_0\| = O_P(T^{-1/2})$ and $\|\hat{\theta} - \theta_0\| = O_P(T^{-1/2})$ are required for the following proofs as in [6,20].

Proof of Theorem A.1. As $\|\hat{\beta} - \beta_0\| = O_P(T^{-1/2})$, it is easy to get that $\hat{g}(w; \hat{\beta}, \hat{\theta}) - \hat{g}(w; \beta_0, \hat{\theta}) = O_P(T^{-1/2})$ by Lagrange mean-value theorem. Note that

$$\begin{aligned}&\sqrt{Th}[\hat{g}(w; \hat{\beta}, \hat{\theta}) - g_0(w)]\\&\quad = \sqrt{Th}[\hat{g}(w; \hat{\beta}, \hat{\theta}) - \hat{g}(w; \beta_0, \hat{\theta}) + \hat{g}(w; \beta_0, \hat{\theta}) - g_0(w)]\\&\quad = \sqrt{Th}[\hat{g}(w; \hat{\beta}, \hat{\theta}) - \hat{g}(w; \beta_0, \hat{\theta})] + \sqrt{Th}[\hat{g}(w; \beta_0, \hat{\theta}) - g_0(w)], \qquad (A.1)\end{aligned}$$

then we only need to illustrate the asymptotic property of $\hat{g}(w; \beta_0, \hat{\theta})$. Minimize

$$\sum_{t=1}^{T}\{Y_t - [a + b(\beta_0^\top X_t - w)]Y_{t-1} - \hat{\theta}^\top Z_t\}^2 K_h(\beta_0^\top X_t - w),$$

with respect to (a, b), and $(\hat{g}(w; \beta_0, \hat{\theta}), \hat{g}'(w; \beta_0, \hat{\theta}))$ can be obtained, whose form is similar to that of (2.2) with (β, θ) replaced by $(\beta_0, \hat{\theta})$.

According to Doukhan *et al.* [9], a strongly mixing stationary sequence is an ergodic sequence, so that $\{Y_t, t \geq 1\}$ is a strictly stationary and ergodic sequence. Therefore, the second moment of Y_{t-1} is constant, denoted as $\gamma = E(Y_{t-1}^2)$, because of the constant mean and constant variance in stationary sequences.

Based on the Lemma 1 in [29], we can get that, for each $j = 0, 1, 2, 3$,

$$\begin{aligned}&\frac{1}{T}\sum_{t=1}^{T}\left\{Y_{t-1}^2\left(\frac{\beta_0^\top X_t - w}{h}\right)^j K_h(\beta_0^\top X_t - w)\right.\\&\quad\left. - E\left[Y_{t-1}^2\left(\frac{\beta_0^\top X_t - w}{h}\right)^j K_h(\beta_0^\top X_t - w)\right]\right\}\\&\quad = O\left(\left\{\frac{\log T}{Th}\right\}^{1/2}\right), \text{ a.s.,}\end{aligned}$$

uniformly for $w \in \mathcal{W}$, where,

$$\begin{aligned} E&\left[Y_{t-1}^2\left(\frac{\beta_0^\top X_t - w}{h}\right)^j K_h(\beta_0^\top X_t - w)\right] \\ &= E(Y_{t-1}^2)E\left[\left(\frac{\beta_0^\top X_t - w}{h}\right)^j K_h(\beta_0^\top X_t - w)\right] \\ &= \gamma f_0(w)\mu_j + O(h), \end{aligned}$$

i.e.,

$$R_{T,j}(w;\beta_0,\hat{\theta}) = \gamma f_0(w)\mu_j + O\left(\left\{\frac{\log T}{Th}\right\}^{1/2} + h\right),\ \text{a.s.} \tag{A.2}$$

So, it is followed immediately that

$$R_T(w;\beta_0,\hat{\theta}) = R(w) + O\left(\left\{\frac{\log T}{Th}\right\}^{1/2} + h\right),\ \text{a.s.,}$$

where $R(w) = \gamma f_0(w)\text{diag}\{1, \mu_2\}$. Using the fact that

$$(A + hB)^{-1} = A^{-1} - hA^{-1}BA^{-1} + O(h^2),$$

we have

$$R_T^{-1}(w;\beta_0,\hat{\theta}) = R^{-1}(w) + O\left(\left\{\frac{\log T}{Th}\right\}^{1/2} + h\right),\ \text{a.s.,} \tag{A.3}$$

uniformly for $w \in \mathcal{W}$. Let

$$\begin{aligned} &\eta_{T,j}^*(w;\beta_0,\hat{\theta}) \\ &\quad = \frac{1}{T}\sum_{t=1}^T [Y_t - \hat{\theta}^\top Z_t - g_0(\beta_0^\top X_t)Y_{t-1}]Y_{t-1}\left(\frac{\beta_0^\top X_t - w}{h}\right)^j \\ &\quad K_h(\beta_0^\top X_t - w),\ j = 0, 1 \end{aligned}$$

and

$$\eta_T^*(w;\beta,\hat{\theta}) = \begin{pmatrix} \eta_{T,0}^*(w;\beta_0,\hat{\theta}) \\ \eta_{T,1}^*(w;\beta_0,\hat{\theta}) \end{pmatrix}.$$

As

$$E\left\{[Y_t - \hat{\theta}^\top Z_t - g_0(\beta_0^\top X_t)Y_{t-1}]Y_{t-1}\left(\frac{\beta_0^\top X_t - w}{h}\right)^j K_h(\beta_0^\top X_t - w)\right\}$$

$$= E\left\{[Y_t - \theta_0^\top Z_t + \theta_0^\top Z_t - \hat{\theta}^\top Z_t - g_0(\beta_0^\top X_t)Y_{t-1}]Y_{t-1}\right.$$
$$\left.\left(\frac{\beta_0^\top X_t - w}{h}\right)^j K_h(\beta_0^\top X_t - w)\right\}$$
$$= E\left[\varepsilon_t Y_{t-1}\left(\frac{\beta_0^\top X_t - w}{h}\right)^j K_h(\beta_0^\top X_t - w)\right]$$
$$+E\left[(\theta_0 - \hat{\theta})^\top Z_t Y_{t-1}\left(\frac{\beta_0^\top X_t - w}{h}\right)^j K_h(\beta_0^\top X_t - w)\right]$$
$$= O(T^{-1/2}),$$

by Lemma 1 in [29] and arguments similar to that in the previous proof, we can show that

$$\eta^*_{T,j}(w; \beta_0, \hat{\theta}) = O\left(\left\{\frac{\log T}{Th}\right\}^{1/2} + T^{-1/2}\right), \text{ a.s.,}$$

uniformly for $w \in \mathcal{W}$. Using Taylor's expansion for $g_0(\beta_0^\top X_t)$ at w, it is presented that

$$\eta_{T,j}(w; \beta_0, \hat{\theta}) - \eta^*_{T,j}(w; \beta_0, \hat{\theta})$$
$$= R_{T,j}(w; \beta_0, \hat{\theta})g_0(w) + hR_{T,j+1}(w; \beta_0, \hat{\theta})g_0'(w)$$
$$+\frac{1}{2}h^2 R_{T,j+2}(w; \beta_0, \hat{\theta})g_0''(w) + o(h^2), \text{ a.s.,}$$

so that

$$\eta_T(w; \beta_0, \hat{\theta}) - \eta^*_T(w; \beta_0, \hat{\theta}) = R_T(w; \beta_0, \hat{\theta})\begin{pmatrix} g_0(w) \\ hg_0'(w)\end{pmatrix}$$
$$+\frac{1}{2}h^2 g_0''(w)\begin{pmatrix} R_{T,2}(w; \beta_0, \hat{\theta}) \\ R_{T,3}(w; \beta_0, \hat{\theta})\end{pmatrix} + o(h^2), \text{ a.s.} \tag{A.4}$$

Thus, from the formulae (A.2), (A.3) and (A.4) above, it follows that

$$\begin{pmatrix} \hat{g}(w; \beta_0, \hat{\theta}) - g_0(w) \\ h[\hat{g}'(w; \beta_0, \hat{\theta}) - g_0'(w)]\end{pmatrix} = R^{-1}(w)\eta^*_T(w; \beta_0, \hat{\theta})$$
$$+\frac{1}{2}h^2 g_0''(w)\begin{pmatrix} \mu_2 \\ \frac{\mu_3}{\mu_2}\end{pmatrix} + o(T^{-1/2} + h^2), \text{ a.s.}$$

Clearly,

$$\hat{g}(w; \beta_0, \hat{\theta}) - g_0(w) = \gamma^{-1} f_0^{-1}(w)\eta^*_{T,0}(w; \beta_0, \hat{\theta})$$
$$+\frac{1}{2}h^2 g_0''(w)\mu_2 + o_P(T^{-1/2} + h^2),$$

uniformly for $w \in \mathcal{W}$.

By simple calculation, we have

$$\begin{aligned}
&\sqrt{Th}\eta^*_{T,0}(w; \beta_0, \hat{\theta}) \\
&= \sqrt{Th}\frac{1}{T}\sum_{t=1}^{T}[Y_t - \hat{\theta}^\top Z_t - g_0(\beta_0^\top X_t)Y_{t-1}]Y_{t-1}K_h(\beta_0^\top X_t - w) \\
&= \sqrt{Th}\left\{\frac{1}{T}\sum_{t=1}^{T}[Y_t - \theta_0^\top Z_t - g_0(\beta_0^\top X_t)Y_{t-1}]Y_{t-1}K_h(\beta_0^\top X_t - w)\right. \\
&\quad \left. + O_P(T^{-1/2})\right\} \\
&= \sqrt{Th}[\frac{1}{T}\sum_{t=1}^{T}\varepsilon_t Y_{t-1}K_h(\beta_0^\top X_t - w)] + o_P(1) \\
&= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\{\sqrt{h}\varepsilon_t Y_{t-1}K_h(\beta_0^\top X_t - w) - \sqrt{h}E[\varepsilon_t Y_{t-1}K_h(\beta_0^\top X_t - w)]\} \\
&\quad + o_P(1).
\end{aligned}$$

By Slutsky's theorem and Theorem 4.4 of [21], it can be obtained immediately that

$$\sqrt{Th}\eta^*_{T,0}(w; \beta_0, \hat{\theta}) \xrightarrow{L} N(0, \gamma\nu_0\sigma^2 f_0(w)).$$

Consequently, we can derive

$$\sqrt{Th}\left[\hat{g}(w; \beta_0, \hat{\theta}) - g_0(w) - \frac{1}{2}h^2 g_0''(w)\mu_2\right] \xrightarrow{L} N(0, \gamma^{-1}\nu_0\sigma^2 f_0^{-1}(w)),$$

and hence, from (A.1),

$$\sqrt{Th}\left[\hat{g}(w; \hat{\beta}, \hat{\theta}) - g_0(w) - \frac{1}{2}h^2 g_0''(w)\mu_2\right] \xrightarrow{L} N(0, \gamma^{-1}\nu_0\sigma^2 f_0^{-1}(w)).$$

The proof is complete. □

Proof of Theorem A.2. By (2.3), with λ as the Lagrange multiplier, it is easy to know that $(\hat{\beta}, \hat{\theta})$ is the solution to

$$\begin{aligned}
&\lambda\begin{pmatrix}\hat{\beta}\\0\end{pmatrix} + \frac{1}{\sqrt{T}}\sum_{t=1}^{T}[Y_t - \hat{g}(\hat{\beta}^\top X_t; \hat{\beta}, \hat{\theta})Y_{t-1} \\
&\quad - \hat{\theta}^\top Z_t]\begin{pmatrix}\hat{g}'(\hat{\beta}^\top X_t; \hat{\beta}, \hat{\theta})Y_{t-1}X_t\\ Z_t\end{pmatrix} = 0, \qquad \text{(A.5)}
\end{aligned}$$

which can be rewritten as

$$\lambda\begin{pmatrix}\hat{\beta}\\0\end{pmatrix} + \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\varepsilon_t\begin{pmatrix}\hat{g}'(\hat{\beta}^\top X_t; \hat{\beta}, \hat{\theta})Y_{t-1}X_t\\ Z_t\end{pmatrix}$$

$$-\frac{1}{\sqrt{T}}\sum_{t=1}^{T}[(\hat{\theta}-\theta_0)^\top Z_t]\begin{pmatrix}\hat{g}'(\hat{\beta}^\top X_t;\hat{\beta},\hat{\theta})Y_{t-1}X_t\\ Z_t\end{pmatrix}$$
$$-\frac{1}{\sqrt{T}}\sum_{t=1}^{T}[\hat{g}(\hat{\beta}^\top X_t;\hat{\beta},\hat{\theta})-g_0(\beta_0^\top X_t)]Y_{t-1}$$
$$\begin{pmatrix}\hat{g}'(\hat{\beta}^\top X_t;\hat{\beta},\hat{\theta})Y_{t-1}X_t\\ Z_t\end{pmatrix}=0.$$

Through direct calculation, we find that

$$\lambda\begin{pmatrix}\hat{\beta}\\ 0\end{pmatrix}+\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\varepsilon_t A_t-\frac{1}{\sqrt{T}}\sum_{t=1}^{T}[(\hat{\theta}-\theta_0)^\top Z_t]A_t$$
$$-\frac{1}{\sqrt{T}}\sum_{t=1}^{T}[\hat{g}(\hat{\beta}^\top X_t;\hat{\beta},\hat{\theta})-g_0(\beta_0^\top X_t)]Y_{t-1}A_t$$
$$+\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\varepsilon_t B_t-\frac{1}{\sqrt{T}}\sum_{t=1}^{T}[(\hat{\theta}-\theta_0)^\top Z_t]B_t$$
$$-\frac{1}{\sqrt{T}}\sum_{t=1}^{T}[\hat{g}(\hat{\beta}^\top X_t;\hat{\beta},\hat{\theta})-g_0(\beta_0^\top X_t)]Y_{t-1}B_t=0,$$

where,

$$A_t=\begin{pmatrix}g_0'(\beta_0^\top X_t)Y_{t-1}X_t\\ Z_t\end{pmatrix},\quad B_t=\begin{pmatrix}[\hat{g}'(\hat{\beta}^\top X_t;\hat{\beta},\hat{\theta})-g_0'(\beta_0^\top X_t)]Y_{t-1}X_t\\ Z_t\end{pmatrix}.$$

So, we have the equation

$$\lambda\begin{pmatrix}\hat{\beta}\\ 0\end{pmatrix}+\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\varepsilon_t A_t-\frac{1}{\sqrt{T}}\sum_{t=1}^{T}[(\hat{\theta}-\theta_0)^\top Z_t]A_t$$
$$-\frac{1}{\sqrt{T}}\sum_{t=1}^{T}[\hat{g}(\hat{\beta}^\top X_t;\hat{\beta},\hat{\theta})-g_0(\beta_0^\top X_t)]Y_{t-1}A_t$$
$$+o_P(1)=0. \tag{A.6}$$

Obviously,

$$\begin{aligned}
&\hat{g}(\hat{\beta}^\top X_t;\hat{\beta},\hat{\theta})-g_0(\beta_0^\top X_t)\\
&=\hat{g}(\hat{\beta}^\top X_t;\hat{\beta},\hat{\theta})-\hat{g}(\beta_0^\top X_t;\hat{\beta},\hat{\theta})+\hat{g}(\beta_0^\top X_t;\hat{\beta},\hat{\theta})-g_0(\beta_0^\top X_t)\\
&=\hat{g}(\beta_0^\top X_t;\hat{\beta},\hat{\theta})+\hat{g}'(\beta_0^\top X_t;\hat{\beta},\hat{\theta})(\hat{\beta}-\beta_0)^\top X_t+o_p((\hat{\beta}-\beta_0)^\top X_t)\\
&\quad-\hat{g}(\beta_0^\top X_t;\hat{\beta},\hat{\theta})+\hat{g}(\beta_0^\top X_t;\hat{\beta},\hat{\theta})-g_0(\beta_0^\top X_t)\\
&=\hat{g}'(\beta_0^\top X_t;\hat{\beta},\hat{\theta})(\hat{\beta}-\beta_0)^\top X_t+\hat{g}(\beta_0^\top X_t;\hat{\beta},\hat{\theta})-g_0(\beta_0^\top X_t)+o_P(T^{-1/2})\\
&=g_0'(\beta_0^\top X_t)(\hat{\beta}-\beta_0)^\top X_t+\hat{g}(\beta_0^\top X_t;\hat{\beta},\hat{\theta})-g_0(\beta_0^\top X_t)
\end{aligned}$$

$$
\begin{aligned}
&+[\hat{g}'(\beta_0^\top X_t;\hat{\beta},\hat{\theta}) - g_0'(\beta_0^\top X_t)](\hat{\beta}-\beta_0)^\top X_t + o_P(T^{-1/2})\\
&= g_0'(\beta_0^\top X_t)(\hat{\beta}-\beta_0)^\top X_t + \hat{g}(\beta_0^\top X_t;\hat{\beta},\hat{\theta}) - g_0(\beta_0^\top X_t) + o_P(T^{-1/2}).
\end{aligned} \tag{A.7}
$$

Substituting (A.7) into (A.6), we can obtain

$$
\begin{aligned}
&\lambda\begin{pmatrix}\hat{\beta}\\0\end{pmatrix} + \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\varepsilon_t A_t - \frac{1}{\sqrt{T}}\sum_{t=1}^{T}[(\hat{\theta}-\theta_0)^\top Z_t]A_t\\
&\quad -\frac{1}{\sqrt{T}}\sum_{t=1}^{T} g_0'(\beta_0^\top X_t)(\hat{\beta}-\beta_0)^\top X_t Y_{t-1}A_t\\
&\quad -\frac{1}{\sqrt{T}}\sum_{t=1}^{T}[\hat{g}(\beta_0^\top X_t;\hat{\beta},\hat{\theta}) - g_0(\beta_0^\top X_t)]Y_{t-1}A_t + o_P(1) = 0,
\end{aligned}
$$

meaning that

$$
\begin{aligned}
&\lambda\begin{pmatrix}\hat{\beta}\\0\end{pmatrix} + \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\varepsilon_t A_t\\
&\quad -\frac{1}{\sqrt{T}}\sum_{t=1}^{T}A_tA_t^\top\begin{pmatrix}\hat{\beta}-\beta_0\\ \hat{\theta}-\theta_0\end{pmatrix} - \frac{1}{\sqrt{T}}\sum_{t=1}^{T}[\hat{g}(\beta_0^\top X_t;\hat{\beta},\hat{\theta})-g_0(\beta_0^\top X_t)]Y_{t-1}A_t\\
&\quad + o_P(1) = 0.
\end{aligned}
$$

Using the Ergodic theorem, we have

$$
\begin{aligned}
&\lambda\begin{pmatrix}\hat{\beta}\\0\end{pmatrix} + \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\varepsilon_t A_t - \sqrt{T}A\begin{pmatrix}\hat{\beta}-\beta_0\\ \hat{\theta}-\theta_0\end{pmatrix}\\
&\quad -\frac{1}{\sqrt{T}}\sum_{t=1}^{T}[\hat{g}(\beta_0^\top X_t;\hat{\beta},\hat{\theta}) - g_0(\beta_0^\top X_t)]Y_{t-1}A_t + o_P(1) = 0,
\end{aligned} \tag{A.8}
$$

where

$$
A = E[A_tA_t^\top].
$$

On the other hand, following the estimation procedure, $(\hat{a},\hat{b}) \equiv (\hat{g}(w;\hat{\beta},\hat{\theta}), \hat{g}'(w;\hat{\beta},\hat{\theta}))$ is the minimizer of

$$
\sum_{t=1}^{T}\{Y_t - [a + b(\hat{\beta}^\top X_t - w)]Y_{t-1} - \hat{\theta}^\top Z_t\}^2 K_h(\hat{\beta}^\top X_t - w),
$$

then, $(\hat{a},\hat{b})$ satisfies the formula

$$
\frac{1}{T}\sum_{t=1}^{T}\{Y_t - [\hat{a} + h\hat{b}(\hat{\beta}^\top X_t - w)/h]Y_{t-1} - \hat{\theta}^\top Z_t\}Y_{t-1}
$$

$$\begin{pmatrix} 1 \\ (\hat{\beta}^\top X_t - w)/h \end{pmatrix} K_h(\hat{\beta}^\top X_t - w) = 0, \tag{A.9}$$

via Taylor expansion and using the conditions on h, we get

$$\begin{aligned} &\frac{1}{T}\sum_{t=1}^{T}\{Y_t - [a + b(\beta_0^\top X_t - w)]Y_{t-1} - \theta_0^\top Z_t\}Y_{t-1} \\ &\begin{pmatrix} 1 \\ (\beta_0^\top X_t - w)/h \end{pmatrix} K_h(\beta_0^\top X_t - w) \\ &-B_{T1}\begin{pmatrix} \hat{a} - a \\ h(\hat{b} - b) \end{pmatrix} - B_{T2}(\hat{\beta} - \beta_0) - B_{T3}(\hat{\theta} - \theta_0) \\ &+o_P(T^{-1/2}) + O_P(h^2) = 0, \end{aligned}$$

where

$$\begin{aligned} B_{T1} &= \frac{1}{T}\sum_{t=1}^{T} Y_{t-1}^2 K_h(\beta_0^\top X_t - w)\begin{pmatrix} 1 & \frac{\beta_0^\top X_t - w}{h} \\ \frac{\beta_0^\top X_t - w}{h} & (\frac{\beta_0^\top X_t - w}{h})^2 \end{pmatrix}, \\ B_{T2} &= \frac{1}{T}\sum_{t=1}^{T} g_0'(w) Y_{t-1}^2 K_h(\beta_0^\top X_t - w)\begin{pmatrix} X_t^\top \\ X_t^\top \frac{\beta_0^\top X_t - w}{h} \end{pmatrix} \end{aligned}$$

and

$$B_{T3} = \frac{1}{T}\sum_{t=1}^{T} Y_{t-1} K_h(\beta_0^\top X_t - w)\begin{pmatrix} Z_t^\top \\ Z_t^\top \frac{\beta_0^\top X_t - w}{h} \end{pmatrix}.$$

Similarly, from Lemma 1 in [29], we provide the asymptotic counterparts of $B_{T,j} (j = 1, 2, 3)$ as follows:

$$\begin{aligned} B_{T1} &= \gamma f_0(w)\mathrm{diag}\{1, \mu_2\} + O_P\left(\left\{\frac{\log T}{Th}\right\}^{1/2} + h\right), \\ B_{T2} &= g_0'(w)\left[\begin{pmatrix} \gamma f_0(w) E(X_t^\top | \beta_0^\top X_t = w) \\ 0 \end{pmatrix} + O_P\left(\left\{\frac{\log T}{Th}\right\}^{1/2} + h\right)\right] \end{aligned}$$

and

$$B_{T3} = \begin{pmatrix} E(Y_{t-1}) f_0(w) E(Z_t^\top | \beta_0^\top X_t = w) \\ 0 \end{pmatrix} + O_P\left(\left\{\frac{\log T}{Th}\right\}^{1/2} + h\right).$$

Thus,

$$\begin{pmatrix} \hat{a} - a \\ h(\hat{b} - b) \end{pmatrix} = \frac{1}{T}\sum_{t=1}^{T} \varepsilon_t Y_{t-1}\gamma^{-1} f_0^{-1}(w)\begin{pmatrix} 1 \\ \frac{\beta_0^\top X_t - w}{h\mu_2} \end{pmatrix} K_h(\beta_0^\top X_t - w)$$

$$-\gamma^{-1} f_0^{-1}(w) \begin{pmatrix} g_0'(w) E(X_t^\top | \beta_0^\top X_t = w) \gamma f_0(w) \\ 0 \end{pmatrix} (\hat{\beta} - \beta_0)$$
$$-\gamma^{-1} f_0^{-1}(w) \begin{pmatrix} E(Z_t^\top | \beta_0^\top X_t = w) E(Y_{t-1}) f_0(w) \\ 0 \end{pmatrix} (\hat{\theta} - \theta_0)$$
$$+ o_P(T^{-1/2}).$$

Then, it can be shown that

$$\begin{aligned} \hat{g}(w; \hat{\beta}, \hat{\theta}) - g_0(w) = &\frac{1}{T} \sum_{t=1}^{T} \gamma^{-1} f_0^{-1}(w) Y_{t-1} K_h(\beta_0^\top X_t - w) \varepsilon_t \\ &- g_0'(w)(\hat{\beta} - \beta_0)^\top E(X_t | \beta_0^\top X_t = w) \\ &- \gamma^{-1} E(Y_{t-1})(\hat{\theta} - \theta_0)^\top E(Z_t | \beta_0^\top X_t = w) \\ &+ o_P(T^{-1/2}). \end{aligned} \tag{A.10}$$

Substituting (A.10) into (A.8) and applying the Ergodic theorem at the same time, we get

$$\begin{aligned} &\lambda \begin{pmatrix} \hat{\beta} \\ 0 \end{pmatrix} + \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \varepsilon_t A_t - \sqrt{T} B \begin{pmatrix} \hat{\beta} - \beta_0 \\ \hat{\theta} - \theta_0 \end{pmatrix} \\ &- \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \frac{1}{T} \sum_{l=1}^{T} [\gamma^{-1} f_0^{-1}(\beta_0^\top X_t) Y_{l-1} Y_{t-1} K_h(\beta_0^\top X_l - \beta_0^\top X_t) A_t \varepsilon_l] \\ &+ o_P(1) = 0, \end{aligned} \tag{A.11}$$

where

$$B = E[A_t A_t^\top] - E\left[A_t \begin{pmatrix} g_0'(\beta_0^\top X_t) Y_{t-1} E(X_t | \beta_0^\top X_t) \\ \gamma^{-1} Y_{t-1} E(Y_{t-1}) E(Z_t | \beta_0^\top X_t) \end{pmatrix}^\top \right].$$

Handle the fourth term in (A.11) by interchanging the summations and we have

$$-\frac{1}{\sqrt{T}} \sum_{l=1}^{T} \varepsilon_l Y_{l-1} \frac{1}{T} \sum_{t=1}^{T} [\gamma^{-1} f_0^{-1}(\beta_0^\top X_t) Y_{t-1} K_h(\beta_0^\top X_l - \beta_0^\top X_t) A_t].$$

Furthermore, by the Ergodic theorem, the term is equivalent asymptotically to

$$-\frac{1}{\sqrt{T}} \sum_{l=1}^{T} \varepsilon_l Y_{l-1} E[\gamma^{-1} f_0^{-1}(\beta_0^\top X_t) Y_{t-1} K_h(\beta_0^\top X_l - \beta_0^\top X_t) A_t],$$

i.e.,

$$-\frac{1}{\sqrt{T}} \sum_{l=1}^{T} \varepsilon_l Y_{l-1} \begin{pmatrix} g_0'(\beta_0^\top X_l) E(X_l | \beta_0^\top X_l) \\ \gamma^{-1} E(Y_{t-1}) E(Z_l | \beta_0^\top X_l) \end{pmatrix}. \tag{A.12}$$

For convenience, let

$$P_\beta = \begin{pmatrix} I - \beta_0\beta_0^\top & 0 \\ 0 & I \end{pmatrix}.$$

Combining (A.11) and (A.12), and multiplying by P_β, we obtain

$$\begin{aligned} &P_\beta B\sqrt{T}\begin{pmatrix} \hat{\beta} - \beta_0 \\ \hat{\theta} - \theta_0 \end{pmatrix} \\ &\quad = \frac{1}{\sqrt{T}}\sum_{t=1}^{T} \varepsilon_t P_\beta \begin{pmatrix} g_0'(\beta_0^\top X_t)Y_{t-1}[X_t - E(X_t|\beta_0^\top X_t)] \\ Z_t - \gamma^{-1}E(Y_{t-1})Y_{t-1}E(Z_t|\beta_0^\top X_t) \end{pmatrix} + o_P(1). \end{aligned} \tag{A.13}$$

By Slutsky's theorem and Theorem 4 of [9], we verify that

$$\sqrt{T}\begin{pmatrix} \hat{\beta} - \beta_0 \\ \hat{\theta} - \theta_0 \end{pmatrix} \xrightarrow{L} N(0, \sigma^2 B^{-1} V (B^{-1})^\top),$$

where

$$V = \begin{pmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{pmatrix}$$

with

$$\begin{aligned} V_{11} &= \gamma E\{[g_0'(\beta_0^\top X_t)]^2 [X_t - E(X_t|\beta_0^\top X_t)][X_t - E(X_t|\beta_0^\top X_t)]^\top\}, \\ V_{12} &= E(Y_{t-1})E\{g_0'(\beta_0^\top X_t)[X_t - E(X_t|\beta_0^\top X_t)][Z_t - E(Z_t|\beta_0^\top X_t)]^\top\}, \\ V_{21} &= E(Y_{t-1})E\{g_0'(\beta_0^\top X_t)[Z_t - E(Z_t|\beta_0^\top X_t)][X_t - E(X_t|\beta_0^\top X_t)]^\top\}, \\ V_{22} &= \{1 - \gamma^{-1}[E(Y_{t-1})]^2\}E(Z_t Z_t^\top) + \gamma^{-1}[E(Y_{t-1})]^2 E\{[Z_t \\ &\qquad - E(Z_t|\beta_0^\top X_t)][Z_t - E(Z_t|\beta_0^\top X_t)]^\top\}. \end{aligned}$$

This completes the proof of Theorem 3.2. □

References

[1] Anderson T W and Rubin H, The asymptotic properties of estimates of the parameters of a single equation in a complete system of stochastic equations, *Ann. Math. Stat.* **21** (1950) 570–582

[2] Bollerslev T, Generalized autoregressive conditional heteroskedasticity, *J. Econometrics* **31** (1986) 307–327

[3] Box G E P and Jenkins G M, Time Series Analysis, Forecasting, and Control (1970) (San Francisco: Holden-Day)

[4] Cai Z and Fan J, Average regression surface for dependent data, *J. Multivariate Anal.* **75** (2000) 112–142

[5] Cai Z, Fan J and Yao Q, Functional-coefficient regression models for nonlinear time series, *J. Amer. Statist. Assoc.* **95** (2000) 941–956

[6] Carroll R J, Fan J, Gijbels I and Wand M P, Generalized partially linear single-index models, *J. Amer. Statist. Assoc.* **92** (1997) 477–489

[7] Chen R and Tsay R S, Functional-coefficient autoregressive models, *J. Amer. Statist. Assoc.* **88** (1993a) 298–308
[8] Chen R and Tsay R S, Nonlinear Additive ARX Models, *J. Amer. Statist. Assoc.* **88** (1993b) 955–967
[9] Doukhan P, Massart P and Rio E, The functional central limit theorem for strongly mixing processes, *Ann. Inst. H. Poincare Probab. Statist.* **30** (1994) 63–82
[10] Engle R F, Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation, *Econometrica* **50** (1982) 987–1007
[11] Fan J and Gijbels I, Local Polynomial Modelling and its Applications (1996) (London: Chapman and Hall)
[12] Fan J and Yao Q, Nonlinear Time Series: Nonparametric and Parametric Methods (2003) (New York: Springer)
[13] Feng S and Xue L, Model detection and estimation for single-index varying coefficient model, *J. Multivariate Anal.* **139** (2015) 227–244
[14] Granger C W J and Andersen A P, An Introduction to Bilinear Time Series Models (1978) (Göttingen: Vandenhoeck & Ruprecht)
[15] Haggan V and Ozaki T, Modelling nonlinear random vibrations using an amplitude-dependent autoregressive time series model, *Biometrika* **68** (1981) 189–196
[16] Harvill J L and Ray B K, Functional coefficient autoregressive models for vector time series, *Comput. Statist. Data Anal.* **50** (2006) 3547–3566
[17] Ichimura H, Semiparametric least squares (SLS) and weighted SLS estimation of single-index models, *J. Econometrics* **58** (1993) 71–120
[18] Jiang J, Multivariate functional-coefficient regression models for nonlinear vector time series data, *Biometrika* **101** (2014) 689–702
[19] Koopmans T C, Statistical Inference in Dynamic Economic Models (Ed.) (1950) (New York: Wiley)
[20] Lin H, Zhang R, Shi J, Liu J and Liu Y, A new local estimation method for single index models for longitudinal data, *J. Nonparametr. Stat.* **28** (2016) 644–658
[21] Masry E and Tjøstheim D, Nonparametric estimation and identification of nonlinear ARCH time series: Strong convergence and asymptotic normality, *Econometric Theory* **11** (1995) 258–289
[22] Nicholls D F and Quinn B G, Random Coefficient Autoregressive Models: An Introduction, Lecture Notes in Statistics, 11 (1982) (New York: Springer)
[23] Tong H, Discussion of a paper by A. J. Lawrance and N. T. Kottegoda, *J. Roy. Statist. Soc. Ser. A* **140** (1977) 34–35
[24] Tong H and Lim K S, Threshold autoregression, limit cycles and cyclical data, *J. R. Stat. Soc. Ser. B Stat. Methodol.* **42** (1980) 245–292
[25] Wong H, Ip W and Zhang R, Varying-coefficient single-index model, *Comput. Statist. Data Anal.* **52** (2008) 1458–1476
[26] Xia Y and Li W K, On the estimation and testing of functional-coefficient linear models, *Statist. Sinica* **9** (1999a) 735–757
[27] Xia Y and Li W K, On single-index coefficient regression models, *J. Amer. Statist. Assoc.* **94** (1999b) 1275–1285
[28] Xia Y, Tong H and Li W K, On extended partially linear single-index models, *Biometrika* **86** (1999) 831–842
[29] Xue L and Pang Z, Statistical inference for a single-index varying-coefficient model, *Stat. Comput.* **23** (2013) 589–599
[30] Yule G U, On a method of investigating periodicities in disturbed series with special reference to Wolfer's sunspot numbers, *Philos. Trans. Roy. Soc. London Ser. A* **226** (1927) 267–298
[31] Zhang S, Guo B, Dong A, He J, Xu Z and Chen S X, Cautionary tales on air-quality improvement in Beijing, *Proc. R. Soc. A* **473** (2017) 20170457

[32] Zhao Y, Xue L and Feng S, Estimation for a partially linear single-index varying-coefficient model, *Comm. Statist. Simulation Comput.* (2019 Online), https://doi.org/10.1080/03610918.2019.1680691

COMMUNICATING EDITOR: Gopal Basak