Journal of Econometrics 148 (2009) 101–113

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Functional-coefficient models for nonstationary time series data[☆]

Zongwu Cai[a,b,c,*], Qi Li[d,e], Joon Y. Park[d]

[a] Department of Mathematics & Statistics, University of North Carolina at Charlotte, Charlotte, NC 28223, USA
[b] Department of Economics, University of North Carolina at Charlotte, Charlotte, NC 28223, USA
[c] Wang Yanan Institute for Studies in Economics, Xiamen University, Xiamen, Fujian 361005, China
[d] Department of Economics, Texas A&M University, College Station, TX 77843, USA
[e] Department of Economics, Tsinghua University, Beijing 100084, China

ARTICLE INFO

Article history:
Available online 15 October 2008

Keywords:
Nonstationary
Nonlinearity
Semiparametric estimation

ABSTRACT

This paper studies functional coefficient regression models with nonstationary time series data, allowing also for stationary covariates. A local linear fitting scheme is developed to estimate the coefficient functions. The asymptotic distributions of the estimators are obtained, showing different convergence rates for the stationary and nonstationary covariates. A two-stage approach is proposed to achieve estimation optimality in the sense of minimizing the asymptotic mean squared error. When the coefficient function is a function of a nonstationary variable, the new findings are that the asymptotic bias of its nonparametric estimator is the same as the stationary covariate case but convergence rate differs, and further, the asymptotic distribution is a mixed normal, associated with the local time of a standard Brownian motion. The asymptotic behavior at boundaries is also investigated.

Published by Elsevier B.V.

1. Introduction

Nonparametric estimation techniques offer numerous advantages relative to parametric techniques, due mainly to their flexibility and robustness to functional form misspecification, and have been embraced by applied researchers in social, behavioral and economic sciences. Asymptotic theory underlying nonparametric estimators and test statistics for many commonly used models has been well established for independent and identically distributed (iid) data as well as for weakly dependence data. However, little is known about the behavior with nonstationary (in particular, integrated with order one, denoted by I(1)) data, which have predominately been modeled linearly. The early nonparametric asymptotic analyses with nonstationary data include Phillips and Park (1998), Park and Hahn (1999), Chang and Martinez-Chombo (2003) and Juhl (2005). Phillips and Park (1998) and Juhl (2005) considered nonparametric estimation of regression models when the true data generating process is a linear unit root process, while the others considered the models linearized in the nonstationary variables. More recently,[1] Wang and Phillips (forthcoming, 2008) considered nonparametric estimation of a regression model with an I(1) regressor and Xiao (forthcoming) considered a varying coefficient model with I(1) regressors appearing in the parametric component of the model. Finally, Karlsen et al. (2007) considered nonparametric estimation of a regression model for a different (a more general) type of nonstationary processes, a subclass of the class of null recurrent Markov chains.

In this paper, we tackle a more general set-up for a class of semiparametric models with non-stationary covariates. Specifically, we focus on the popular varying coefficient regression model with some nonstationary covariates

$$Y_t = \beta(Z_t)^{\mathrm{T}} X_t + \varepsilon_t, \quad 1 \le t \le n, \tag{1.1}$$

where Y_t, Z_t and ε_t are scalar, $X_t = (X_{t1}, \ldots, X_{td})^{\mathrm{T}}$ is a vector of covariates with dimension d, $\beta(\cdot)$ is a $d \times 1$ column vector function, and the superscript T denotes transpose of a matrix. For ease notation, we assume that Z_t is univariate case. Extension to multivariate Z_t involves fundamentally no new ideas but

☆ We thank Peter Robinson for his constructive discussions and comments on an early version of this paper. We also thank an associate editor and two anonymous referees for their helpful comments that greatly improved the paper. Cai's research was supported, in part, by the National Science Foundation grant DMS-0404954 and funds provided by the University of North Carolina at Charlotte. Li's research was partially supported by the Private Enterprize Research Center, Texas A&M University. Park's research is partially supported by the National Science Foundation.

* Corresponding author at: Department of Mathematics & Statistics, University of North Carolina at Charlotte, Charlotte, NC 28223, USA.

E-mail addresses: zcai@uncc.edu (Z. Cai), qi@econmail.tamu.edu (Q. Li), jpark@tamu.edu (J.Y. Park).

[1] The first version of this paper was written independently of these recent works on nonparametric estimation of regression models with non-stationary covariates.

0304-4076/$ – see front matter. Published by Elsevier B.V.
doi:10.1016/j.jeconom.2008.10.003

complicated notations. We observe (Y_t, X_t, Z_t) for $t = 1, \ldots, n$. When $\{(X_t, Z_t, \varepsilon_t)\}$ is stationary (denoted by I(0)) or iid, various versions of (1.1) have been considered by many authors, including but not limited to, for example, Chen and Tsay (1993), Hastie and Tibshirani (1993), Cai et al. (2000), Li et al. (2002), and among others. When ε_t is stationary and $Z_t = t$, Eq. (1.1) has been tackled by Robinson (1989, 1991), Cai (2007) and Chen and Hong (2007) for stationary X_t, by Park and Hahn (1999) and Chang and Martinez-Chombo (2003) for nonstationary X_t, and by Cai and Wang (2008) for nearly integrated X_t. When $X_t = 1$ and Z_t is I(1), Eq. (1.1) becomes a standard univariate nonparametric regression model as considered by Wang and Phillips (forthcoming, 2008) and Karlsen et al. (2007). Finally, when Z_t is I(0) and X_t is I(1), model (1.1) reduces to the case considered by Xiao (forthcoming).

The advantage of a varying coefficient model specification, compared with an unrestricted nonparametric regression, is that it attenuates the "curse of dimensionality" problem. It also includes many popular semiparametric models as special cases. For example, when X_t contains a constant, say the first component $X_{t1} = 1$, we can write $X_t^{\mathrm{T}} = (1, \tilde{X}_t^{\mathrm{T}})$. Further, if the coefficient vector associated with $\tilde{X}_t$ is a vector of constants, say γ, then the varying coefficient model reduces to a partially linear model $E(Y_t|X_t, Z_t) = \beta_1(Z_t) + \tilde{X}_t^{\mathrm{T}}\gamma$; see, e.g., Robinson (1988).

The remainder of the paper is organized as follows: Section 2 discusses the case when Z_t is stationary. Here, local linear estimators of coefficient functions are developed, and their asymptotic properties are established. A two-step estimation procedure is also proposed when some covariates are nonstationary and the rest are stationary. Section 3 considers the case when Z_t is nonstationary. Nonparametric kernel smoothing of the coefficient functions is developed and its asymptotic behavior is investigated. Concluding remarks are presented in Section 4. Proofs of the main results of the paper are given in two Appendices.

2. Models with stationary Z_t

We consider first the case when some or all components of X_t are I(1) and Z_t is strictly stationary. For expositional simplicity, we re-express (1.1) as the following varying coefficient model

$$Y_t = \beta(Z_t)^{\mathrm{T}} X_t + \varepsilon_t = \beta_1(Z_t)^{\mathrm{T}} X_{t1} + \beta_2(Z_t)^{\mathrm{T}} X_{t2} + \varepsilon_t, \quad 1 \le t \le n, \tag{2.1}$$

where X_{t1}, Z_t, and ε_t are stationary, X_{t2} is an I(1) vector, $\beta(Z_t) = (\beta_1(Z_t)^{\mathrm{T}},\ \beta_2(Z_t)^{\mathrm{T}})^{\mathrm{T}}$, and $X_t = (X_{t1}^{\mathrm{T}},\ X_{t2}^{\mathrm{T}})^{\mathrm{T}}$, where X_{ti} is a $d_i \times 1$ vector, $i = 1, 2$, $d_1 + d_2 = d$, and the first component of X_{t1} is identically one. In what follows, we assume that $E(\varepsilon_t \mid X_t,\ Z_t) = 0$ which implies that X_t and Z_t are uncorrelated with ε_t. Note that Y_t is allowed to be stationary or nonstationary. For example, model (2.1) can be applied to the analysis of purchasing power of parity, in which $X_{t2}^{\mathrm{T}} = (P_t, P_t^*, E_t)$ (and no X_{t1}), where P_t and P_t^* are the price levels of the domestic and a foreign country, E_t is the exchange rate between the domestic and the foreign currencies, and $Z_t = I_t - I_t^*$ is the difference between the domestic interest rate I_t and the foreign interest rate I_t^*. Then if Y_t is an I(0) variable, we say that P_t, P_t^* and E_t are co-integrated with a varying coefficient co-integration vector $\beta(Z_t)$ which is a vector of smooth functions of Z_t. This setting is more general than the usual assumption that β is a vector of constant parameters in the usual purchasing power of parity analysis.

2.1. Local linear estimation

It is well known in the literature; see, e.g., Fan and Gijbels (1996), that a local linear fitting has several nice properties, over the classical Nadaraya–Watson (local constant) method, such as high statistical efficiency in an asymptotic minimax sense, design-adaptation, and automatic edge correction. We estimate $\beta(\cdot)$ using a local linear fitting from observations $\{(X_t,\ Z_t,\ Y_t)\}_{t=1}^n$. We assume throughout the paper that $\beta(\cdot)$ is twice continuously differentiable, so that for any given grid point z, we use a local approximation as $\beta(z) + \beta^{(1)}(z)\,(Z_t - z)$ to approximate $\beta(Z_t)$, where $\beta^{(s)}(z) = \mathrm{d}^s\beta(z)/\mathrm{d}z^s$. Define

$$\begin{pmatrix} \widehat{\theta}_0 \\ \widehat{\theta}_1 \end{pmatrix} = \operatorname{argmin}_{\theta_0, \theta_1} \sum_{t=1}^n \left[Y_t - \theta_0^{\mathrm{T}} X_t - (Z_t - z)\,\theta_1^{\mathrm{T}} X_t\right]^2 \times K_h(Z_t - z), \tag{2.2}$$

where $K_h(u) = h^{-1}K(u/h)$, $K(\cdot)$ is a kernel function satisfying Assumption A3 below, $\widehat{\theta}_0 = \widehat{\beta}(z)$ estimates $\beta(z)$, and $\widehat{\theta}_1 = \widehat{\beta}^{(1)}(z)$ estimates $\beta^{(1)}(z)$. Then, $\widehat{\beta}(z)$ and $\widehat{\beta}^{(1)}(z)$ can be expressed as

$$\begin{pmatrix} \widehat{\beta}(z) \\ \widehat{\beta}^{(1)}(z) \end{pmatrix} = \left[\sum_{t=1}^n \begin{pmatrix} X_t \\ (Z_t - z)\,X_t \end{pmatrix}^{\otimes 2} K_h(Z_t - z)\right]^{-1} \times \sum_{t=1}^n \begin{pmatrix} X_t \\ (Z_t - z)X_t \end{pmatrix} Y_t\,K_h(Z_t - z), \tag{2.3}$$

where $A^{\otimes 2} = A\,A^{\mathrm{T}}$ ($A^{\otimes 1} = A$) for a vector or matrix A.

2.2. Notations and assumptions

Since X_{t2} is a vector of I(1) processes, it can be re-expressed as $X_{t2} = X_{t-1,2} + \eta_t = X_{02} + \sum_{s=1}^t \eta_s$ $(t \ge 1)$, where $\{\eta_s\}$ is an I(0) process with mean zero and variance Ω_η. Then,

$$\frac{X_{[nr]2}}{\sqrt{n}} \equiv \frac{X_{t2}}{\sqrt{n}} = \frac{X_{0,2}}{\sqrt{n}} + \frac{1}{\sqrt{n}}\sum_{s=1}^t \eta_s = \frac{X_{0,2}}{\sqrt{n}} + \frac{1}{\sqrt{n}}\sum_{s=1}^{[nr]} \eta_s,$$

where $r = t/n$ and $[x]$ denotes the integer part of x. Under some regularity conditions, Donsker's theorem; see, for example, Theorems 14.1 and 19.2 in Billingsley (1999) for iid η_t and ρ-mixing η_t, respectively, generalizes in an obvious way to the multivariate cases and leads to

$$X_{[nr]2}/\sqrt{n} \Longrightarrow W_{\eta,2}(r) \quad \text{as } n \to \infty, \tag{2.4}$$

where $W_{\eta,2}(\cdot)$ is a d_2-dimensional Brownian motion on $[0,\ 1]$ with covariance matrix Σ_η and "$\Longrightarrow$" represents weak convergence. In particular, it follows from Merlevéde et al. (2006) that (2.4) holds if $\{\eta_t\}$ is a stationary strong (α-)mixing sequence satisfying, for some $\delta_0 > 0$,

$$E|\eta_t|^{2+\delta_0} < \infty, \quad \text{and} \quad \sum_{k=1}^\infty k^{(2+\delta_0)/\delta_0}\,\alpha(k) < \infty, \tag{2.5}$$

where $\alpha(\cdot)$ is the mixing coefficient; see, e.g., Hall and Heyde (1980) for more discussion on α-mixing process. Also, for any Borel measurable and totally Lebesgue integrable function $\Gamma(\cdot)$, one has

$$\frac{1}{n}\sum_{t=1}^n \Gamma(X_{[nr]2}/\sqrt{n}) \xrightarrow{d} \int_0^1 \Gamma(W_{\eta,2}(s))\mathrm{d}s \quad \text{as } n \to \infty,$$

where $\xrightarrow{d}$ denotes the convergence in distribution, so that, for $l = 1, 2$,

$$\frac{1}{n}\sum_{t=1}^n \left(X_{t2}/\sqrt{n}\right)^{\otimes l} \xrightarrow{d} \int_0^1 [W_{\eta,2}(r)]^{\otimes l}\mathrm{d}r \equiv W_{\eta,2}^{(l)} \quad \text{as } n \to \infty; \tag{2.6}$$

see Theorem 1.2 in Berkes and Horváth (2006) for details. Under stronger regularity conditions, (2.4) can be strengthened to the following strong approximation result

$$\sup_{0 \le r \le 1} \|X_{[nr]2}/\sqrt{n} - W_{\eta,2}(r)\| = O(n^{-\theta_*}\log^{\lambda_*}(n)) \tag{2.7}$$

almost surely, where $\|\cdot\|$ is the usual L_2 norm in $\Re^{d_2}$, $\theta_* = (1/2)-1/(2+\delta_*)$ and $\lambda_* = \lambda_*(\delta_*) > 0$ is a function of δ_*, provided that $\{\eta_t\}$ is a stationary strong mixing sequence satisfying that, for some $\gamma_* > 2+\delta_*$ with $0 < \delta_* \le 2$,

$$E|\eta_t|^{\gamma_*} < \infty, \quad \text{and} \quad \sum_{n=1}^{\infty} \alpha(n)^{1/(2+\delta_*)-1/\gamma_*} < \infty; \tag{2.8}$$

see Theorem 4.1 in Shao and Lu (1987) and Einmahl (1987) for details. It is not hard to see that assumption (2.8) is stronger than (2.5). This is not surprising, since strong approximation in (2.7) usually requires stronger assumptions than weak convergence as in (2.4). Finally, note that if $\{\eta_t\}$ is iid and has a finite γ_*-th moment ($\gamma_* > 2$), then the right hand side of (2.7) becomes $o(n^{-\theta_*})$, where $\theta_* = (1/2) - 1/\gamma_*$; see, e.g., Csörgő and Révés (1981, p. 107). We assume throughout the paper that the sequence $\{\eta_t\}$ is stationary α-mixing and satisfies either (2.5) or (2.8). Note that either (2.5) or (2.8) ensures that $\lim_{n\to\infty} \mathrm{Var}(n^{-1/2}\sum_{t=1}^n \eta_t)$ exists and is finite by Davydov's inequality for an α-mixing process; see, e.g., Corollary A.2 in Hall and Heyde (1980).

Next, we give regularity conditions for the asymptotic distribution of $\widehat{\beta}(z)$. We introduce the following notations. Let $f_z(z)$ denote the marginal density of Z_t. Define $M_k(z) = E\left[X_{t1}^{\otimes k} \mid Z_t = z\right]$ for $k = 1$ and 2. Also set, for $j \ge 0$, $\mu_j(K) = \int_{-\infty}^{\infty} v^j K(v)\mathrm{d}v$, and $\nu_j(K) = \int_{-\infty}^{\infty} v^j K^2(v)\mathrm{d}v$. Further, let

$$S(z) = \begin{pmatrix} M_2(z) & M_1(z)\, W_{\eta,2}^{(1)^{\mathrm{T}}} \\ W_{\eta,2}^{(1)}\, M_1(z)^{\mathrm{T}} & W_{\eta,2}^{(2)} \end{pmatrix}, \tag{2.9}$$

where $W_{\eta,2}^{(l)}$ is defined in (2.6). We make the following assumptions.

Assumptions:

A1. $\beta(z)$ is twice continuously differentiable in z for all $z \in \Re$.
A2. $M_k(z)$ is positive-definite and continuous in a neighborhood of z. $f(z)$ is continuously differentiable in a neighborhood of z and $f_z(z) > 0$.
A3. The kernel function $K(\cdot)$ is a symmetric and continuous density function, supported by $[-1, 1]$.
A4. The bandwidth h satisfies $h \to 0$ and $n\,h \to \infty$.
A5. ε_t has a finite fourth moment, $E(\varepsilon_t|X_t, Z_t) = 0$, and $E(\varepsilon_t^2|X_t, Z_t) = \sigma_\varepsilon^2$ is a positive constant.
A6. $\{(X_{t1}, Z_t, \varepsilon_t, \eta_t);\ t \ge 1\}$ is a strictly α-mixing stationary process with the δ_1-th moment ($\delta_1 > 2$). $E\left[|\varepsilon_t X_{t1}^2|^{\delta_2} \mid Z_t = z\right] \le C_1 < \infty$ with $\delta_2 > \delta_1$ and $\alpha(t) = O(t^{-\delta_3})$ for some $\delta_3 > \min\{\delta_2\delta_1/(\delta_2-\delta_1),\ \delta_5,\ 2\delta_6/(2-\delta_6)\}$, where $\delta_5 = \delta_4\delta_1/(\delta_4\delta_1 - \delta_1 - \delta_4)$ for some δ_4 satisfying $\delta_1/(\delta_1-1) < \delta_4 < 2$. Also, $\|\eta_t\|_{q_0} = [E|\eta_t|^{q_0}]^{1/q_0} < \infty$ with $q_0 = \delta_4\delta_6/(\delta_4-\delta_6)$ for some $1 < \delta_6 < \delta_4$. Further, $\sup_k E\left[\eta_1^2\,\varepsilon_{k+1}^2 \mid Z_{k+1} = z\right] \le C_2 < \infty$.
A7. $f(z_0, z_s|x_0,\ x_s;\ s) \le M < \infty$ for $s \ge 1$, where $f(z_0, z_s|x_0,\ x_s;\ s)$ is the conditional density of $(Z_0,\ Z_s)$ given $(X_{01} = x_0,\ X_{s1} = x_s)$.
A8. $n^{1/2-\delta_1/4}\,h^{\delta_1/\delta_2-1/2-\delta_1/4} = O(1)$.

We give some comments on the above conditions. Assumptions A1 and A2 are smoothness conditions. The requirement in Assumption A3 that $K(\cdot)$ be compactly supported is imposed for the sake of brevity of proofs, and can be removed at the cost of lengthier arguments. α-mixing is one of the weakest mixing conditions for weakly dependent stochastic processes. Stationary linear and nonlinear time series or Markov chains fulfilling certain (mild) conditions are α-mixing with exponentially decaying coefficients; see discussions and examples in Cai (2002a), Carrasco and Chen (2002) and Chen and Tang (2005). The conditional homoscedastic error Assumption in A5 can be relaxed to allow for conditional heteroscedasticity of the form $E(\varepsilon_t^2|X_t, Z_t) = \sigma^2(X_{1t}, Z_t)$, i.e., the conditional variance is only a function of the stationary covariates (X_{1t}, Z_t). However, it is technically difficult to let it also be a function of the nonstationary covariate X_{2t}. If $\alpha(\cdot)$ decays geometrically, then Assumption A6 is fulfilled with some standard moment conditions. Assumption A7 is a standard technical assumption. Clearly, Assumption A8 allows for choosing a wide range of h and is slightly stronger than the usual condition $n\,h \to \infty$. For optimal bandwidths selection (i.e., $h = c\,n^{-\gamma}$ for $0 < \gamma < 1, c > 0$), A8 is automatically satisfied for $\delta_1 \ge 2(1+\gamma)/(1-\gamma)$ and it is still fulfilled for $2 < \delta_1 < 2(1+\gamma)/(1-\gamma)$ if δ_2 satisfies $\delta_1 < \delta_2 \le 4\gamma\delta_1/[2(1+\gamma) - \delta(1-\gamma)]$. Conditions similar to Assumptions A6–A8 are also imposed by Cai et al. (2000) for the stationary data case.

2.3. Asymptotic properties

To establish the asymptotic property of $\widehat{\beta}(z)$, we define $D_n = \mathrm{diag}\{I_{d_1},\ \sqrt{n}\,I_{d_2}\}$, and $B_\beta(z) = \mu_2(K)\,\beta^{(2)}(z)/2$, where I_d is a $d \times d$ identity matrix. Detailed proof of the following Theorem is provided in Appendix A.

Theorem 2.1. *Under Assumptions* A1–A8, *we have*

$$\sqrt{nh}\,D_n\left[\widehat{\beta}(z) - \beta(z) - h^2 B_\beta(z)\right] \xrightarrow{d} MN(\Sigma_\beta(z)),$$

where $MN(\Sigma(z))$ is a mixed normal distribution with mean zero and conditional covariance matrix given by $\Sigma_\beta(z) = \sigma_\varepsilon^2\nu_0(K)S(z)^{-1}/f_z(z)$.

Here, a mixed normal distribution is defined as follows. Conditional on the random variable that appears at the asymptotic variance, the estimator has an asymptotic normal distribution, see Phillips (1989) and Phillips and Park (1998) for a formal definition of a mixed normal distribution. Clearly, if there is no nonstationary covariate (i.e., removing X_{t2}), then Theorem 2.1 reduces to

$$\sqrt{nh}\left[\widehat{\beta}_1(z) - \frac{1}{2}\,\beta_1(z) - h^2\mu_2(K)\beta_1^{(2)}(z)\right] \xrightarrow{d} N(0,\ \Sigma_{\beta_1,0}(z)), \tag{2.10}$$

where $\Sigma_{\beta_1,0}(z) = \sigma_\varepsilon^2\nu_0(K)M_2(z)^{-1}/f_z(z)$, which is non-stochastic and is exactly the same as that in Cai et al. (2000). Further, from Theorem 2.1, we see that $\mathrm{Var}(\widehat{\beta}_2(z))$ has a faster rate of convergence than that of $\mathrm{Var}(\widehat{\beta}_1(z))$ due to the fact that X_{t2} is nonstationary ($\sum_{t=1}^n X_{t2}^2 = O(n^2)$ rather than $O(n)$). This is similar to the linear model case. However, the local fitting method renders the asymptotic variance of $\widehat{\beta}_2(z)$ to be of the order $O((n^2h)^{-1})$ rather than the linear model case of $O(n^{-2})$. One can easily derive the integrated asymptotic mean squared error (IAMSE) for $\widehat{\beta}_j(z)$

$$\mathrm{IAMSE}_j = \mathrm{IAMSE}(\widehat{\beta}_j) = \int\left[\frac{h^4}{4}\,\mu_2^2(K)\|\beta_j^{(2)}(z)\|^2 + \frac{\mathrm{tr}(\Sigma_{\beta_j}(z))}{n^j\,h}\right]\omega(z)\mathrm{d}z$$

for $j = 1$ and 2, where $\omega(\cdot)$ is a non-negative weight function, $\Sigma_{\beta_1}(z)$ is the upper-left corner sub-matrix of $\Sigma_\beta(z)$, and $\Sigma_{\beta_2}(z)$ is the lower-right corner sub-matrix of $\Sigma_\beta(z)$. By minimizing IAMSE_j with respect to h, we obtain the optimal bandwidth for $\widehat{\beta}_j(z)$, which is

$$h_{j,opt} = \int \mathrm{tr}(\Sigma_{\beta_j}(z))\omega(z)\mathrm{d}z \times \left[\int \mu_2^2(K)\|\beta_j^{(2)}(z)\|^2\omega(z)\mathrm{d}z\right]^{-1/5} n^{-j/5}. \tag{2.11}$$

With the above choice of $h_{j,opt}$, we see that IAMSE_j has an order of $O(n^{-4j/5})$. Hence, IAMSE_2 for $\widehat{\beta}_2(z)$ has an order $O(n^{-8/5})$ which is smaller than $\text{IAMSE}_1 = O(n^{-4/5})$. Clearly, a single value of h can not make the estimation of both $\beta_1(\cdot)$ and $\beta_2(\cdot)$ optimal (in the sense of minimizing their asymptotic MSE). Therefore, to minimize the asymptotic MSE for each coefficient (function) estimate, an iterative estimation approach is needed. This issue is addressed next.

2.4. A two-step estimation procedure

As discussed in Section 2.3, the one-step estimation procedure based on (2.2) cannot minimize the asymptotic MSE for both $\beta_1(\cdot)$ and $\beta_2(\cdot)$. Therefore, we suggest a two-stage estimation procedure below. The idea is similar to the profile likelihood method; see, e.g., Cai (2002b,c) and Fan and Huang (2005), and is described as follows. If the bandwidth is taken to be of the order $n^{-2/5}$, $\widehat{\beta}_2(z) - \beta_2(z)$ reaches the optimal convergent rate but $\widehat{\beta}_1(z)$ is under-smoothed. Therefore, the first step is to estimate $\beta(z) = (\beta_1(z)^{\mathrm{T}}, \beta_2(z)^{\mathrm{T}})^{\mathrm{T}}$ with a value of h that is optimal for estimating $\beta_2(z)$. We use h_2 to denote it. For example, we can choose $h_2 = c_2\, n^{-2/5}$ for some positive constant c_2. We know that the resulting estimator $\widehat{\beta}_2(z)$ has an optimal convergence rate as $\widehat{\beta}_2(z)-\beta_2(z) = O_p(n^{-4/5})$, and the asymptotic distribution of $\widehat{\beta}_2(z)$ is given by

$$\sqrt{n^2 h_2}\left[\widehat{\beta}_2(z) - \beta_2(z) - \frac{1}{2} h_2^2 \mu_2(K) \beta_2^{(2)}(z)\right] \xrightarrow{d} MN(\Sigma_{\beta_2}(z)), \tag{2.12}$$

where $\Sigma_{\beta_2}(z)$ is previously defined as the lower-right corner $d_2 \times d_2$ sub-matrix of $\Sigma_\beta(z)$.

However, the corresponding estimator of $\beta_1(z)$ is not optimal with the choice of $h_2 = c_2\, n^{-2/5}$ ($c_2 > 0$). Therefore, we suggest that at the second step, one should re-estimate $\beta_1(\cdot)$ with $\beta_2(Z_t)$ replaced by $\widehat{\beta}_2(Z_t)$ obtained at the first step. That is, we replace $\beta_2(Z_t)$ by $\widehat{\beta}_2(Z_t)$ in (2.1) to obtain

$$Y_t^* \equiv Y_t - \widehat{\beta}_2(Z_t)^{\mathrm{T}} X_{t2} = \beta_1(Z_t)^{\mathrm{T}} X_{t1} + \varepsilon_t^*, \tag{2.13}$$

where $\varepsilon_t^* = \varepsilon_t + [\beta_2(Z_t) - \widehat{\beta}_2(Z_t)]^{\mathrm{T}} X_{t2}$. Then, we can apply the local linear method to estimate $\beta_1(z)$. Let $(\widehat{\beta}_{1,2\,step}(z)^{\mathrm{T}}, \widehat{\beta}_{1,2\,step}^{(1)}(z)^{\mathrm{T}})^{\mathrm{T}}$ represent the resulting estimator of $(\beta_1(z)^{\mathrm{T}}, \beta^{(1)}(z)^{\mathrm{T}})^{\mathrm{T}}$, which is given by

$$\begin{pmatrix} \widehat{\beta}_{1,2\,step}(z) \\ \widehat{\beta}_{1,2\,step}^{(1)}(z) \end{pmatrix} = \left[\sum_{t=1}^{n} \begin{pmatrix} X_{t1} \\ (Z_t - z) X_{t1} \end{pmatrix}^{\otimes 2} K_{h_1}(Z_t - z)\right]^{-1} \times \sum_{t=1}^{n} \begin{pmatrix} X_{t1} \\ (Z_t - z) X_{t1} \end{pmatrix} Y_t^* K_{h_1}(Z_t - z), \tag{2.14}$$

where h_1 is the bandwidth used at this step for estimating $\beta_1(z)$. We will show in Theorem 2.2 below that the asymptotic distribution of $\widehat{\beta}_{1,2\,step}(z)$ is the same as the case when $\beta_2(Z_t)$ were known; that is $\widehat{\beta}_{1,2\,step}(z) - \beta_1(z) = O_p((n\,h_1)^{-1/2}) = O_p(n^{-2/5})$ if $h_1 = c_1\, n^{-1/5}$ ($c_1 > 0$) and h_2 is as small as possible (see Theorem 2.2 later). Since the right hand side of (2.13) involves only $\beta_1(\cdot)$, one can use any data-driven method to select h_1 optimally, such as the nonparametric version of the Akaike information criterion type as in Hurvich et al. (1998) and Cai (2002b,c) or the plug-in method as in Ruppert et al. (1995).

It is clear from (2.12) that $\widehat{\beta}_2(z) - \beta_2(z) = O_p(n^{-4/5})$ if $h_2 = c_2\, n^{-2/5}$, and this pointwise convergence rate is optimal. To establish the asymptotic normality of the estimator given in (2.14), we might need a uniform convergence rate. Therefore, it is assumed that the initial estimator satisfies the following condition

$$\sup_{z \in \mathcal{D}} \left|\widehat{\beta}_2(z) - \beta_2(z)\right| = O_p(a_n), \tag{2.15}$$

where $\mathcal{D}$ is a compact support of $f_z(z)$ (by assuming that $f_z(z)$ has a compact support $\mathcal{D}$) and a_n satisfies $a_n \to 0$ with a certain convergence rate. Note that under some regularity conditions (see, e.g., the assumptions of Theorem 2 in Hansen (2008)), and by following the same arguments as in Hansen (2008), one can show[2] that (2.15) holds with $a_n = n^{-4/5}\log(n)$. Alternatively, an assumption similar to (2.15) is also imposed in Linton (2000) for iid samples, and Cai (2002c) for time series data to simplify the proof of the asymptotic results of a two-stage estimator. Now, the asymptotic normality for the proposed two-stage estimator is stated here and its proof is relegated to the Appendix.

Theorem 2.2. *Under assumption* A1–A8 *and* (2.15), *if* $h_2 = o(h_1\, n^{-1/4})$, *we have*

$$\sqrt{n\,h_1}\left[\widehat{\beta}_{1,2\,step}(z) - \beta_1(z) - \frac{1}{2} h_1^2\, \mu_2(K) \beta_1^{(2)}(z) + o_p(h_1^2)\right] \xrightarrow{d} N\left(0,\, \Sigma_{\beta_1,0}(z)\right),$$

where $\Sigma_{\beta_1,0}(z)$ *is defined in* (2.10).

Remark 2.1. Note that the consequences of (2.12) and Theorem 2.2 are that the convergence rates are optimal for estimating each coefficient function at each step. That is, at the first step, the optimal rate is obtained for estimating coefficient functions of non-stationary covariates and the second step is devoted to obtaining the optimal rate for estimating coefficient functions of the stationary covariates. Also, note that the result in Theorem 2.2 is exactly the same as that (see (2.10)) in Cai et al. (2000) for stationary case, which implies that $\widehat{\beta}_{1,2\,step}$ is "oracle" in the sense that its asymptotic distribution is the same as the case with a known $\beta_2(Z_t)$. Finally, we note that the order of h_2 should be smaller than its optimal value, which means we need to undersmooth the estimate of $\beta_2(Z_t)$ at the first step. This is a common phenomenon for a two-stage estimation method; see Cai (2002b,c). In practical implementation, we refer to the papers by Cai (2002b,c) for choosing the data-driven fashion bandwidths for a two-stage estimation.

3. Models with nonstationary Z_t

When Z_t is a nonstationary I(1) regressor, the asymptotic analysis is much involved. Therefore, we consider only the case that X_t is stationary in this section. The model is the same as given in (1.1) but now Z_t is nonstationary; that is,

$$Y_t = \beta(Z_t)^{\mathrm{T}} X_t + \varepsilon_t, \quad 1 \le t \le n, \tag{3.1}$$

where X_t is a $d \times 1$ vector of stationary variables, and Z_t is a univariate I(1) nonstationary variable. When $X_t = 1$, model (3.1) was considered by Karlsen et al. (2007) for the case that Z_t is a nonstationary process from a subclass of the class of null recurrent Markov chains, and by Wang and Phillips (forthcoming, 2008) for the case that Z_t is I(1), while all of them used the local constant fitting approach to estimate the nonparametric regression function.

We assume that $\beta(z)$ is twice continuously differentiable. Exactly similar to (2.3), we obtain the local linear estimator of $\beta(z)$, given by

$$\begin{pmatrix} \widehat{\beta}(z) \\ \widehat{\beta}^{(1)}(z) \end{pmatrix} = \left[\sum_{t=1}^{n} \begin{pmatrix} X_t \\ (Z_t - z) X_t \end{pmatrix}^{\otimes 2} K_h(Z_t - z)\right]^{-1} \times \sum_{t=1}^{n} \begin{pmatrix} X_t \\ (Z_t - z) X_t \end{pmatrix} Y_t\, K_h(Z_t - z). \tag{3.2}$$

[2] The detailed proof of the uniform convergence result as stated in (2.15) is available from the authors upon request.

Since Z_t is an I(1) process, Z_t can be expressed as $Z_t = \sum_{s=1}^{t} u_s + Z_0$, where u_t is a mixing process with mean zero and variance σ_u^2. We assume that $\lim_{n\to\infty} \mathrm{Var}(n^{-1/2}\sum_{t=1}^{n} u_t)$ is a finite positive constant. Consider the simple case that $Z_0 = 0$ and u_t is a white noise. Then, $Z_t \sim (0, t\sigma_u^2)$ and $(Z_t - z)/\sqrt{t} \sim (-zt^{-1/2}, \sigma_u^2)$. Note that $\{u_t\}$ is not required to be an independent process. Instead, in what follows, we assume that the process $\{(X_t, u_t)\}$ is a stationary mixing process and that $\{u_s\}$ is a linear process: $u_s = \sum_{j=0}^{\infty} c_j\, \omega_{s-j}$, where ω_j is a white noise with mean zero and $\sigma_\omega^2 = \mathrm{Var}(\omega_j) < \infty$, and $\{c_j\}$ satisfies the following conditions, for some $0 < \tau \le 1$,

$$\sum_{j=0}^{\infty} |c_j|^{\tau} < \infty, \quad \text{and} \quad \sum_{j=0}^{\infty} c_j = 1. \tag{3.3}$$

Then, $\sigma_u^2 = \mathrm{Var}(u_s) = \sigma_\omega^2 \sum_{j=0}^{\infty} c_j^2$ and $\mathrm{Cov}(u_s, u_{s+t}) = \sigma_\omega^2 \sum_{j=0}^{\infty} c_j\, c_{j+t}$ for any s and t.

Let $\rho_{x,u}(|t-s|)$ and $\rho_{v,u}(|t-s|)$ be the vector and matrix of autocorrelation coefficients between X_t and u_s, and between V_t and u_s, respectively, where $V_t \equiv X_t X_t^{\mathrm{T}}$. Then we assume that $\sum_{s=1}^{\infty} |\rho_{x_i,u}(s)| < \infty$ and $\sum_{s=1}^{\infty} |\rho_{v_{ij},u}(s)| < \infty$ for all $1 \le i, j \le d$, where $X_{t,i}$ and $V_{t,ij}$ are the ith and the (i,j)th component of X_t and V_t, respectively. Also, the correlation coefficient between $V_{t,ij}$ and Z_t is $O(t^{-1/2})$, which goes to zero as $t \to \infty$, because the correlation coefficient between $V_{t,ij}$ and Z_t is

$$\begin{aligned}\rho(V_{t,ij}, Z_t) &= \sum_{s=0}^{t} \mathrm{Cov}(V_{t,ij}, u_s)/[\sigma_{v,ij}\sigma_u \sqrt{t}] \\ &= \sum_{s=0}^{t} \rho_{v_{ij},u}(s)/\sqrt{t} = O(t^{-1/2}),\end{aligned} \tag{3.4}$$

where $\sigma_{v,ij}^2 = \mathrm{Var}(V_{t,ij})$. Next, define $\xi_{t,z} = t^{-1/2}(Z_t - z)$ and let $f_{t,z}(\cdot)$ denote the density of $\xi_{t,z}$. Also, we use $f_{t,s,z}(\cdot,\cdot)$ to represent the joint density function of $(\xi_{t,z}, \xi_{s,z})$. Further, set $b_{t,s,z}(\cdot,\cdot)$ to be the conditional density of $(\xi_{t,z}, \xi_{s,z})$, conditional on V_s. Let $\mathscr{F}_t$ be the smallest σ-field generated by $\{Y_s, X_s, Z_s\}_{s=-\infty}^{t}$. We make the following assumptions.

Assumptions:

C1. $E(\varepsilon_t|X_t, Z_t, \mathscr{F}_{t-1}) = 0$, $E(\varepsilon_t^2|X_t, Z_t, \mathscr{F}_{t-1}) = \sigma_\epsilon^2$, $E(\varepsilon_t^4|X_t, Z_t, \mathscr{F}_{t-1}) < C$ a.s., and $\{(X_t, u_t)\}$ is a stationary and mixing process satisfying constraints as imposed by (2.5) and (2.8), where σ_ϵ^2, C and σ_u^2 are finite positive constants. Also, $\lim_{n\to\infty} \mathrm{Var}(n^{-1/2}\sum_{t=1}^{n} u_t)$ exists and is finite.

C2. $E\left[|V_{t,ij}|^{2q}\right] < \infty$ for some $q > 3$ and for all $1 \le i, j \le d$.

C3. Both $f_{t,z}(\cdot)$ and $f_{t,s,z}(\cdot,\cdot)$ have bounded continuous derivative functions (for all t, s and fixed z). Also, assume that $b_{t,s,z}(\cdot,\cdot)$ has bounded continuous derivative functions (for all t, s, z).

C4. $E(V_t|Z_t)$ has the following expression.

$$E(V_t|Z_t) = E(V_t) + \delta_t g_t(Z_t), \tag{3.5}$$

where $E[\delta_t g_t(Z_t)] = 0$, $E[|g_t(Z_t)|^{2q}] = O(1)$, $\delta_t = O(t^{-1/2})$, $g_t(\cdot)$ has the same dimension as V_t. $g_{t,ij}(\cdot)$ denotes the (i,j)th component of $g_t(\cdot)$ and satisfies the assumption $\sup_t |g_{t,ij}(u)| \le C(u)$ with $C(u)$ being a continuous function (for all $1 \le i, j \le d$). Further, for $t > s$, assume that

$$E(V_t|Z_t, Z_s, V_s) = E(V_t) + \delta_{t,s} g_{t,s}(Z_t, Z_s, V_s), \tag{3.6}$$

where $g_{t,s}(.,.,.)$ is of dimension $d \times d$, $E[\delta_{t,s} g_{t,s}(Z_t, Z_s, V_s)] = 0$, $E[|g_{t,s}(Z_t, Z_s, V_s)|^{2q}] = O(1)$, and $\delta_{t,s} = O(t^{-1/2} + s^{-1/2}) = O(s^{-1/2})$ for $t > s$.

C5. $K(\cdot)$ is a symmetric and continuous density function, supported by $[-1, 1]$.

C6. $n\,h^{8p-2} \to \infty$, where $p = q/(q-1) < 3/2$ and $q > 3$ is given in Assumption C1.

Next, we discuss the above conditions. Condition C1 requires that ε_t is a martingale difference process with conditional homoskedastic variance and a finite fourth moment. The martingale difference assumption can be relaxed to a mixing process, and the conditional homoskedastic error can be loosen to the case that $E(\varepsilon_t^2|X_t, Z_t) = E(\varepsilon_t^2|X_t)$, with a lengthier proof. C2 implies that $E\left[|\zeta_t|^{2q}\right] = O(1)$, where $\zeta_t = V_t - E(V_t|Z_t)$. C3 is a very mild assumption. For condition C4, given that $E(V_t|Z_t)$ has finite second moment, it is natural to expect that the nonstationary variable Z_t should be at least associated with a factor $t^{-1/2}$ which appears in $E(V_t|Z_t)$. Therefore, $\delta_t = O(t^{-1/2})$ is not a restrictive assumption. For example, if $X_t = (1, \tilde{X}_t^{\mathrm{T}})^{\mathrm{T}}$, and $\tilde{X}_t$ and Z_t are jointly normal,[3] it is easy to see from the normal distribution theory and (3.4) that $E(\tilde{X}_t|Z_t) = E(\tilde{X}_t) + c_1 t^{-1}(Z_t - E(Z_t))$, so that $\delta_t = t^{-1/2}$ and $g_t(Z_t) = c_1 t^{-1/2}[Z_t - E(Z_t)]$ (c_1 is a finite constant). Also, if V_t is a χ_1^2 random variable, say $V_t = \tilde{X}_t^2$, and that $\tilde{X}_t$ and Z_t are jointly normal, then it is straightforward to show that $E(V_t|Z_t) = E(V_t) + c_2 t^{-2}[Z_t^2 - E(Z_t^2)] = E(V_t) + t^{-1} c_2 [t^{-1} Z_t^2 - \sigma_u^2]$ (c_2 is a matrix of constants). Hence, $\delta_t = t^{-1}$ and $g_t(Z_t) = c_2[t^{-1}Z_t^2 - \sigma_u^2]$. In the above examples, $g_t(Z_t)$ has finite moments of any order since Z_t is normally distributed with a finite variance. Further, by (3.4), $\mathrm{Var}(Z_t) = O(t)$, and $\mathrm{Var}(Z_s) = O(s)$, it follows that $\mathrm{Cov}(V_t, Z_t) = O(1)$, $\mathrm{Cov}(V_t, Z_s) = O(1)$, and $\mathrm{Cov}(V_t, V_s) = O(\rho_v(|t-s|))$, where $\rho_v(\cdot)$ is the autocorrelation coefficient of $\{V_t\}$. This, together with the normality assumption on (X_t, Z_t, Z_s, X_s), implies that $\delta_{t,s} = O(s^{-1/2} + t^{-1/2})$, which goes to 0 as $\min\{t, s\} \to \infty$. Moreover, by Assumption C2, (3.6), and C_r-inequality, it is easy to verify that $g_{t,s}(z, z, V_s)$ has the finite $2q$-th moment (V_s is random) for any fixed value of z, i.e., $E\{|g_{t,s}(z, z, V_t)|^{2q}\} = O(1)$ for all t, s and a fixed value of z. By (3.5) and (3.6), $E[g_t(Z_t)] = 0$ and $E[g_{t,s}(Z_t, Z_s, V_s)] = 0$. Finally, Assumption C6 is commonly imposed in the kernel estimation literature and Assumption C6 is satisfied for the optimal bandwidth $h = c\,n^{-1/10}$, $c > 0$ (see later).

As mentioned earlier since Z_t is an I(1) process, Z_t can be expressed as $Z_t = Z_{t-1} + u_t = Z_0 + \sum_{s=1}^{t} u_s$, where $\{u_s\}$ is a stationary process with mean zero and $\sigma_u^2 = \lim_{n\to\infty} \mathrm{Var}(n^{-1/2}\sum_{t=1}^{n} u_t) > 0$. Then, it follows from Donsker's theorem (see (2.4)) that under some regularity conditions, for $0 \le r \le 1$,

$$Z_{[nr]}/\sqrt{n} \Longrightarrow W_u(r), \tag{3.7}$$

where $W_u(\cdot)$ is a Brownian motion on $[0, 1]$ and $\sigma_u^{-1} W_u(r) = W(r)$ is a standard Brownian motion on $[0, 1]$. Next, we define the local time $L(t, x)$ for a standard Brownian motion as

$$L(t, x) = \lim_{\epsilon\to 0} \frac{1}{2\epsilon} \int_0^t I(|W(s) - x| \le \epsilon) ds, \tag{3.8}$$

where $W(\cdot)$ is a standard Brownian motion; see, e.g., Karatzas and Shreve (1991, p. 202) and Park and Phillips (1999) for details. Finally, we state our main result of this section below and the proof is relegated to Appendix B.

Theorem 3.1. *Under Assumptions* C1–C6, *we have*

$$\sqrt{n^{1/2}\,h}\left[\widehat{\beta}(z) - \beta(z) - h^2 B_\beta(z)\right] \xrightarrow{d} MN(\Sigma_1),$$

where $B_\beta(z) = \mu_2(K)\beta^{(2)}(z)/2$ *and* $MN(\Sigma_1)$ *is a mixed normal distribution with mean zero and conditional covariance* $\Sigma_1 = \sigma_\varepsilon^2\, \sigma_u\, \nu_0(K)\, [E(X_t X_t^{\mathrm{T}}) L(1, 0)]^{-1}$.

[3] Recall that $Z_t = \sum_{s=1}^{t} u_s + Z_0$, consider the simple case that $Z_0 = 0$ and u_t is iid normal $N(0, \sigma_u^2)$, then $Z_t \sim N(0, t\sigma_u^2)$. Here, we consider the case that $\tilde{X}_t$ and Z_t are jointly normally distributed.

Remark 3.1. The asymptotic properties for $\widehat{\beta}^{(1)}(z)$ can be obtained in a same way as that in Theorem 2.1 and they are omitted. By comparing the results in Theorems 2.1 and 3.1, our new findings are as follows: It is clear that $h^2 B_\beta(z)$ serves as the asymptotic bias, which is exactly the same as that for stationary case when one uses a local linear estimation method; see Theorem 2.1. This is not surprising since the asymptotic bias term comes from the local linear approximation. However, the asymptotic variance of $\widehat{\beta}(z)$ is of the order $O((n^{1/2}h)^{-1})$, which is larger than $O((nh)^{-1/2})$ for the stationary Z_t case as presented in Theorem 2.2. The integrated asymptotic mean squared error is given by

$$\text{IAMSE} = \int \left[\text{tr}(\Sigma_1)\, n^{-1/2} h^{-1} + \frac{h^4}{4}\, \mu_2^2(K) \left|\left|\beta^{(2)}(z)\right|\right|^2 \right] M(z)\mathrm{d}z.$$

Minimizing the IAMSE with respect to h gives the optimal bandwidth $h_{opt} = cn^{-1/10}$ for some $c > 0$, which is much larger than that for the stationary case; see (2.11).

Now, we consider the asymptotic behavior of $\widehat{\beta}(z)$ at boundaries. When Z_t is I(1), it follows from (3.7) that when $z = a\sqrt{n}$ ($a \neq 0$) and $r = t/n$,

$$\begin{aligned} P(Z_t \geq z) &= P(Z_t \geq a\sqrt{n}) \to P(W_u(r) \geq a) \\ &= 1 - \Phi(a/\sqrt{r}\sigma_u) > 0, \end{aligned}$$

where $\Phi(\cdot)$ is the distribution function of the standard normal random variable. This means that there is a great chance for $|Z_t|$ taking large values. Now the question is how the asymptotic behavior of the estimator looks like when z is large like $z = a\sqrt{n}$ for any fixed a. We offer the following asymptotic results at boundary $z = a\sqrt{n}$ for any fixed a. However, we do not provide the detailed proofs since they follow exactly the same arguments as those used in proving Theorem 3.1.

Theorem 3.2 (*Boundary Behavior*). *If Assumptions* C1–C5 *hold and* $C(\cdot)$ *in assumption* C3 *is bounded as well as* $n^{1/4} h^{5/2} \beta^{(2)}(a\sqrt{n}) = O(1)$ *for any a, then, we have*

$$\sqrt{n^{1/2}h}\left[\widehat{\beta}(a\sqrt{n}) - \beta(a\sqrt{n}) - h^2 B_\beta(a\sqrt{n})\right] \xrightarrow{d} MN(\Sigma_{1,a}),$$

where $MN(\Sigma_{1,a})$ *is a mixed normal distribution with mean zero and conditional covariance* $\Sigma_{1,a} = \sigma_\varepsilon^2 \sigma_u \nu_0(K)\, [E(X_t X_t^{\mathrm{T}}) L(1, a/\sigma_u)]^{-1}$.

Remark 3.2. Comparing Theorem 3.2 with Theorem 3.1, we observe that the asymptotic variance of $\widehat{\beta}(\cdot)$ at the boundary point differs from that at the interior point. This is different from its stationary counterpart; see Fan and Gijbels (1996) for the stationary case.

4. Discussion

In this paper, we studied the class of varying coefficient models with nonstationary time series data. We suggested using the local linear fitting scheme to estimate the nonparametric coefficient functions and derived the asymptotic properties of the proposed estimators. We would like to mention some interesting future research topics related to this paper. First, it would be very useful and important to discuss how to select data-driven (optimal) bandwidths theoretically and empirically. Secondly, an important extension would be to generalize the asymptotic analysis of this paper to the case where both Z_t and (or some of the components) X_t are nonstationary. Further, we conjecture that if some of X_t in (3.1) are the lagged variables, one might find some regularity conditions to show that Y_t generated by (3.1) is ergodic, so that it is stationary if it is assumed to be Markovian. We are currently exploring these issues. Finally, it is warranted to consider an extension to other types of nonstationarity such as nearly integrated processes; see, e.g., Torous et al. (2004), Campbell and Yogo (2006) and Polk et al. (2006), which has a potential application in finance, which is under investigation by Cai and Wang (2008).

Appendix A. Proofs of Theorems 2.1 and 2.2

In the remaining part of this paper, we denote by C a generic positive constant, which may take different values at different places. We often need to evaluate (probability) orders of some finite dimensional matrix random variables. Let M_t be a matrix of finite dimension, and b_n be a sequence of non-stochastic real numbers, we write $|M_t| = O_p(b_n)$ to mean that $|M_{t,ij}| = O_p(b_n)$ for each i and j, where $M_{t,ij}$ is the (i,j)th component of M_t. Similarly, $E|M_t| = O(b_n)$ means that $E|M_{t,ij}| = O(b_n)$ for all i and j; and $\sup_t |M_t| = O(b_n)$ means that $\sup_t |M_{t,ij}| = O(b_n)$ for all i and j. Also, we write $M_t^2 = O_p(b_n)$ to mean that $M_{t,ij}^2 = O_p(b_n)$ for all i and j. Finally, we use $M^{k\times m}$ to denote the real-valued $k\times m$ matrices and $\mathcal{D}[0,1]$ to represent the space of right-continuous with left limits (cadlag) functions on $[0,1]$ equipped with the Skorohod metric (as defined in Billingsley (1999, p. 124)).

Before we prove Theorems 2.1 and 2.2, we first give a few lemmas that will be used frequently in the proofs below. First, consider random arrays $\{(U_{nt}, Y_{nt}) : 1 \leq t \leq n; n \geq 1\}$, where U_{nt} is a $k \times m$ matrix and Y_{nt} is an $m \times 1$ vector. We transform these arrays into random elements on $[0,1]$ by $U_n(s) = U_{n[ns]}$ and $Y_n(s) = Y_{n[ns]}$ for $0 \leq s \leq 1$. Also, define $\epsilon_{nt} = Y_{nt} - Y_{n,t-1}$. We can then define the stochastic integral

$$\int_0^s U_n(r)\mathrm{d}Y_n(r) = \sum_{j=1}^{[ns]} U_{nj}\epsilon_{nj+1}.$$

Let $BM(\Omega)$ denote the vector Brownian motion with covariance matrix Ω. Finally, we define $W_t = \sum_{s=1}^t w_s$, where $\{w_s\}$ satisfies the following assumption:

Assumption M1: For some $p > \beta > 2$, $\{w_s\}$ is a mean zero and strong mixing sequence with mixing coefficient satisfying $\alpha(n) = O(n^{-p\beta/(p-\beta)})$ and $\sup_{t\geq 1} E[\|w_t\|^p] \leq C < \infty$. In addition, $E(W_n W_n^{\mathrm{T}})/n \to \Omega < \infty$ as $n \to \infty$.

Let $W_n(s) = n^{-1/2} W_{[ns]}$ and set $\mathscr{F}_i = \sigma(U_{nt}, w_t : t \leq i)$ to be the smallest σ-field containing the past history of (U_{nt}, w_t) for all n and $t \leq i$, and denote $E(X \mid \mathscr{F}_i)$ by $E_i(X)$. Since $W_n(s)$ might not be a martingale, we can use a martingale to approximate $W_n(s)$. To this end, define,

$$\epsilon_i = \sum_{k=0}^{\infty} [E_i(w_{i+k}) - E_{i-1}(w_{i+k})], \quad \text{and} \quad \zeta_i = \sum_{k=1}^{\infty} E_i(w_{i+k}).$$

Then, it is not hard to verify (see Hansen (1992, p. 492)) that $w_i = \epsilon_i + \zeta_{i-1} - \zeta_i$ with $E_{i-1}(\epsilon_i) = 0$, and for $0 \leq r \leq 1$,

$$\int_0^r U_n(s)\mathrm{d}W_n(s) = \int_0^r U_n(s)\mathrm{d}Y_n(s) + \Lambda_n^*(r),$$

where $Y_n(r) = \sum_{i=1}^{[nr]} \epsilon_i/\sqrt{n}$ and

$$\Lambda_n^*(r) = \frac{1}{\sqrt{n}} \sum_{i=1}^{[nr]} [U_{ni} U_{ni-1}]\zeta_i^{\mathrm{T}} - \frac{1}{\sqrt{n}}\, U_n(r)\, \zeta_{[nr]+1}^{\mathrm{T}}.$$

Then, $\{\epsilon_i, \mathscr{F}_i\}$ is a martingale difference sequence. Therefore, by Theorem 2.1 in Hansen (1992), we have the following result.

Lemma A.1. *Assume that Assumption* M1 *holds,* $\sup_{0\leq r\leq 1} |\Lambda_n^*(r)| = o_p(1)$*, and* $(U_n, W_n) \Longrightarrow (U, W)$ *in* $\mathcal{D}_{M^{km}\times\mathfrak{R}^m}[0,1]$*, then*

$$\int_0^s U_n \mathrm{d}Y_n \Longrightarrow \int_0^s U^- \mathrm{d}W, \quad \textit{equivalently,}$$

$$\int_0^s U_n \mathrm{d}W_n \Longrightarrow \int_0^s U^- \mathrm{d}W,$$

where U^- *is a cadlag process,* $W(s) = BM(\Omega)$*, and* Ω *is defined in Assumption* M1.

Proof. This is Theorem 3.1 of Hansen (1992). □

Lemma A.2. *Suppose $U_n \Longrightarrow U$ in $\mathcal{D}_{M^{km}}[0,1]$ and $U(\cdot)$ is almost surely continuous. For a random sequence $\{e_j\}$ and a sequence of nondecreasing σ-field $\{\mathcal{F}_j^e\}$ to which $\{e_j\}$ is adapted, assume that $\sup_j E|E(e_j|\mathcal{F}_{j-m}^e)| \to 0$ as $m \to \infty$. Then*

$$\sup_{0\le s\le 1}\left|n^{-1}\sum_{j=1}^{[ns]} U_{nj}\, e_j\right| \xrightarrow{p} 0.$$

Proof. See Theorem 3.3 of Hansen (1992). □

Lemma A.3. *Let $w_t = \sqrt{h}\, K_h(Z_t - z)\, \varepsilon_t\, Z_{t,z,h}^j$, and $U_{nt} = X_{t2}/\sqrt{n}$. Set $\mathcal{F}_t = \sigma(U_{ni}, w_i : i \le t)$ to be the smallest σ-field containing the past history of (U_{nt}, w_t) for all n and $i \le t$. For any $0 \le r \le 1$, define $\Lambda_n^*(r) \equiv n^{-1}\sum_{i=1}^{[nr]} \eta_i\,\zeta_i - n^{-1}\ X_{[nr]2}\,\zeta_{[nr]+1}$, where $\zeta_i = \sum_{k=1}^{\infty} E_i(w_{i+k})$ and η_t is from $X_{t2} = \sum_{s=1}^{t}\eta_s$. Then, we have*

$$\sup_{0\le r\le 1} |\Lambda_n^*(r)| = o_p(1).$$

Proof. It is easy to see that, any $p > 0$,

$$\|w_t\|_p = O(h^{-1/2+1/p}). \tag{A.1}$$

By the stationarity, we have, for any $k \ge 1$,

$$E[\eta_i\, w_{i+k}] = E[\eta_1\, w_{k+1}] = h^{1/2}\, E\left[\eta_1\, K_h(Z_{k+1} - z)\varepsilon_{k+1}\, Z_{k+1,z,h}^j\right].$$

By Davydov's inequality for an α-mixing and (A.1),

$$\begin{aligned}|E[\eta_1\, w_{k+1}]| &\le C\,\alpha^{1/p_1}(k)\,\|\eta_1\|_{p_2}\,\|w_{k+1}\|_{p_3}\\ &\le C\, h^{-1/2+1/p_3}\,\alpha^{1/p_1}(k)\end{aligned} \tag{A.2}$$

for any $p_2 > 1$ and $p_3 > 1$ satisfying $1/p_1 + 1/p_2 + 1/p_3 = 1$, which, in conjunction with Assumption A6, implies that for each i, as $n \to \infty$,

$$|E[\eta_i\,\zeta_i]| \le C\, h^{-1/2+1/\delta_4}\sum_{k=1}^{\infty}\alpha^{1/\delta_5}(k) \le C\, h^{-1/2+1/\delta_4} \to 0 \tag{A.3}$$

by setting $p_2 = \delta_1$ and $p_3 = \delta_4$, where δ_1, δ_4, and δ_5 are given in Assumption A6. Now, by Minkowski's inequality, McLeish' α-mixing inequality (see McLeish (1975)), Davydov's inequality, (A.1), and Assumption A6, for any $i \ge 1$,

$$\begin{aligned}\|\zeta_i\|_{\delta_1} &\le \sum_{k=1}^{\infty}\|E_i\,(w_{i+k})\|_{\delta_1} \le C\sum_{k=1}^{\infty}\alpha^{1/\delta_1-1/\delta_2}(k)\ \|w_{i+k}\|_{\delta_2}\\ &\le C\, h^{-1/2+1/\delta_2},\end{aligned}$$

which, together with Chebyshev's inequality and Assumption A8, implies that for any $\varepsilon > 0$,

$$\begin{aligned}P\left(\sup_{i\le n}|\zeta_i| > \sqrt{n}\,\varepsilon\right) &\le \varepsilon^{-\delta_1}\, n^{1-\delta/2}\, E|\zeta_i|^{\delta_1}\\ &\le C\, n^{1-\delta_1/2}\, h^{-\delta_1/2+\delta_1/\delta_2} \to 0.\end{aligned}$$

Therefore, we have

$$\begin{aligned}&\sup_{0\le r\le 1}\frac{1}{n}\,|X_{[nr]2}\,\zeta_{[nr]+1}|\\ &\le \sup_{0\le r\le 1}\frac{1}{\sqrt{n}}\,|X_{[nr]2}|\,\sup_{0\le r\le 1}\frac{1}{\sqrt{n}}\,|\zeta_{[nr]+1}| \xrightarrow{p} 0,\end{aligned} \tag{A.4}$$

since $\sup_{0\le t\le 1}|X_{[nr]2}| = O_p(\sqrt{n})$ by (2.4) and the continuous mapping theorem (see Theorem 2.7 in Billingsley (1999)). In view of (A.4), it suffices to show that the first term on the right hand side of $\Lambda_n^*(t)$ converges to zero in probability uniformly. To this effect, we first show that $\{\eta_i\,\zeta_i - E(\eta_i\,\zeta_i)\}$ is an L_{δ_6}-mixingale for δ_6 given in Assumption A6 satisfying $1 < \delta_6 < \delta_4 < 2$. Note that for the definition of a mixingale sequence, we refer to the paper by McLeish (1975). Indeed, by Minkowski's inequality, for any $m \ge 1$,

$$\begin{aligned}\|E_{i-m}[\eta_i\,\zeta_i - E(\eta_i\,\zeta_i)]\|_{\delta_6} &\le \sum_{k=1}^{\infty}\|E_{i-m}[\eta_i\, w_{i+k} - E(\eta_i\, w_{i+k})]\|_{\delta_6}\\ &= \sum_{k=1}^{m}(\cdots) + \sum_{k=m+1}^{\infty}(\cdots) \equiv I_1 + I_2.\end{aligned} \tag{A.5}$$

By McLeish' inequality, (A.2), and Assumption A6,

$$\begin{aligned}&\|E_{i-m}[\eta_i\, w_{i+k} - E(\eta_i\, w_{i+k})]\|_{\delta_6}\\ &\le C\,\alpha^{1/\delta_6-1/2}(m)[\|\eta_i\, w_{i+k}\|_2 + |E(\eta_i\, w_{i+k})|]\\ &\le C\,\alpha^{1/\delta_6-1/2}(m),\end{aligned}$$

so that

$$I_1 = \sum_{k=1}^{m}\|E_{i-m}[\eta_i\, w_{i+k} - E(\eta_i\, w_{i+k})]\|_{\delta_6} \le C\, m\,\alpha^{1/\delta_6-1/2}(m).$$

For I_2 in (A.5), one obtains

$$\begin{aligned}&\|E_{i-m}[\eta_i\, w_{i+k} - E(\eta_i\, w_{i+k})]\|_{\delta_6} \le \|E_{i-m}[\eta_i\, w_{i+k}]\|_{\delta_6} + |E(\eta_i\, w_{i+k})|\\ &\le \|\eta_i\, E_i[w_{i+k}]\|_{\delta_6} + |E(\eta_i\, w_{i+k})|\\ &\le \|\eta_i\|_{\delta_4\delta_6/(\delta_4-\delta_6)}\|E_i[w_{i+k}]\|_{\delta_4} + |E(\eta_i\, w_{i+k})|\\ &\le C\ \|\eta_i\|_{\delta_4\delta_6/(\delta_4-\delta_6)}\alpha^{1/\delta_4-1/2}(k)\ \|w_{i+k}\|_2 + |E(\eta_i\, w_{i+k})|,\end{aligned}$$

where δ_4 and δ_6 are given in Assumption A6. An application of (A.2) with $p_2 = \delta$ and $p_3 = 2$ and (A.1) with $p = 2$ leads to

$$\begin{aligned}&\|E_{i-m}[\eta_i\, w_{i+k} - E(\eta_i\, w_{i+k})]\|_{\delta_6}\\ &\le C\left[\|\eta_i\|_{q_0}\,\alpha^{1/\delta_4-1/2}(k) + \alpha^{1/2-1/\delta_1}(k)\right] \le C\,\alpha^{1/\delta_4-1/2}(k).\end{aligned}$$

Then,

$$I_2 = \sum_{k=m+1}^{\infty}\|E_{i-m}[\eta_i\, w_{i+k} - E(\eta_i\, w_{i+k})]\|_{\delta_6} \le C\sum_{k=m+1}^{\infty}\alpha^{1/\delta_4-1/2}(k).$$

Hence, (A.5) becomes

$$\begin{aligned}&\|E_{i-m}[\eta_i\,\zeta_i - E(\eta_i\,\zeta_i)]\|_{\delta_6}\\ &\le C\left[m\,\alpha^{1/\delta_6-1/2}(m) + \sum_{k=m+1}^{\infty}\alpha^{1/\delta_4-1/2}(k)\right] \to 0\end{aligned}$$

by Assumption A6 as $m \to \infty$. Therefore, the sequence $\{\eta_i\,\zeta_i - E(\eta_i\,\zeta_i)\}$ is a uniformly integrable L_1-mixingale. An application of Corollary to Theorem 3.3 in Hansen (1992) yields that

$$\frac{1}{n}\sup_{0\le r\le 1}\left|\sum_{i=1}^{[nr]}[\eta_i\,\zeta_i - E(\eta_i\,\zeta_i)]\right| \xrightarrow{p} 0.$$

This, together with (A.3), concludes that $\frac{1}{n}\sup_{0\le r\le 1}\left|\sum_{i=1}^{[nr]}\eta_i\,\zeta_i\right| \xrightarrow{p} 0$. Therefore, by (A.4),

$$\begin{aligned}\sup_{0\le r\le 1}|\Lambda_n^*(r)| &\le \frac{1}{n}\sup_{0\le r\le 1}\left|\sum_{i=1}^{[nr]}\eta_i\,\zeta_i\right| + \frac{1}{n}\sup_{0\le r\le 1}|X_{[nr]2}\,\zeta_{[nr]+1}|\\ &= o_p(1).\end{aligned}$$

This completes the proof of Lemma A.3. □

Proof of Theorem 2.1. First, note that the right hand side of (2.3) has the form of $A^{-1}B$, define $\mathcal{H}_n = \begin{pmatrix}1 & 0\\ 0 & h\end{pmatrix} \otimes D_n$, so that we can write $\mathcal{H}_n A^{-1} B = \mathcal{H}_n A^{-1} \mathcal{H}_n \mathcal{H}_n^{-1} B = [\mathcal{H}_n^{-1} A \mathcal{H}_n^{-1}]^{-1} \mathcal{H}_n^{-1} B$. Thus, $\widehat{\beta}(z)$ and $\widehat{\beta}^{(1)}(z)$ can be re-expressed as follows:

$$\mathcal{H}_n \begin{pmatrix} \widehat{\beta}(z) \\ \widehat{\beta}^{(1)}(z) \end{pmatrix} = S_n(z)^{-1} n^{-1} \sum_{t=1}^n K_h(Z_t - z) \times Y_t \begin{pmatrix} 1 \\ Z_{t,z,h} \end{pmatrix} \otimes \left(D_n^{-1} X_t\right), \tag{A.6}$$

where $Z_{t,z,h} = (Z_t - z)/h$ and

$$S_n(z) = n^{-1} \sum_{t=1}^n K_h(Z_t - z) \begin{pmatrix} 1 \\ Z_{t,z,h} \end{pmatrix}^{\otimes 2} \otimes \left(D_n^{-1} X_t\right)^{\otimes 2} = \begin{pmatrix} S_{n,0}(z) & S_{n,1}(z) \\ S_{n,1}(z) & S_{n,2}(z) \end{pmatrix}$$

with $j = 0, 1, 2,$

$$S_{n,j}(z) = \frac{1}{n} \sum_{t=1}^n K_h(Z_t - z) Z_{t,z,h}^j \left(D_n^{-1} X_t\right)^{\otimes 2}.$$

Now, to facilitate the analysis of $S_{n,j}(z)$, we first express $S_n(z)$ as

$$S_{n,j}(z) = \begin{pmatrix} F_{n,j,0}(z) & F_{n,j,1}(z) \\ F_{n,j,1}(z)^{\mathrm{T}} & F_{n,j,2}(z) \end{pmatrix},$$

where

$$F_{n,j,0}(z) = \frac{1}{n} \sum_{t=1}^n Z_{t,z,h}^j X_{t1} X_{t1}^{\mathrm{T}} K_h(Z_t - z),$$

$$F_{n,j,1}(z) = \frac{1}{n} \sum_{t=1}^n K_h(Z_t - z) Z_{t,z,h}^j X_{t1} X_{t2}^{\mathrm{T}}/\sqrt{n},$$

and

$$F_{n,j,2}(z) = \frac{1}{n} \sum_{t=1}^n Z_{t,z,h}^j K_h(Z_t - z) \left(X_{t2}/\sqrt{n}\right)^{\otimes 2}.$$

For $l = 1, 2$, define

$$F_{n,j,l}^*(z) = \frac{1}{n} \sum_{t=1}^n Z_{t,z,h}^j X_{t1}^{\otimes l} K_h(Z_t - z).$$

By noting that X_{t1} and Z_t are stationary and using the standard change-of-variable and a Taylor's expansion argument, we have

$$E\left[F_{n,j,l}^*(z)\right] = E\left[Z_{t,z,h}^j X_{t1}^{\otimes l} K_h(Z_t - z)\right] = f_z(z) M_l(z) \mu_j(K) + o(1).$$

By the kernel theory for the stationary mixing case; see Theorem 1 of Cai et al. (2000) for details, one can easily show that

$$\mathrm{Var}\left[F_{n,j,l}^*(z)\right] = O((n\,h)^{-1}) = o(1). \tag{A.7}$$

Therefore,

$$F_{n,j,l}^*(z) = f_z(z) M_l(z) \mu_j(K) + o_p(1), \tag{A.8}$$

so that

$$F_{n,j,0}(z) = F_{n,j,2}^*(z) = f_z(z) M_2(z) \mu_j(K) + o_p(1). \tag{A.9}$$

Let $\mathscr{F}_i^e = \sigma(X_{t1}, Z_i : t \le i)$ be the smallest σ-field containing the past history of (X_{t1}, Z_t), and denote by $e_t = K_h(Z_t - z) Z_{t,z,h}^j X_{t1} - E[K_h(Z_t - z) Z_{t,z,h}^j X_{t1}]$. Then, similar to (A.7), it is easy to verify that

$$\sup_{s \ge 0} \mathrm{Var}\left(\sum_{t=s+1}^{s+m} e_t\right) = O(m/h) \tag{A.10}$$

for any $m \ge 1$. Define $U_{nt} = X_{t2}/\sqrt{n}$ for any $1 \le t \le n$ and $U_n(r) = U_{n,[nr]}$ for any $r \in [0,\ 1]$. For any small $0 < \delta < 1$, set $N = [1/\delta]$, $t_k = [k\,n/N] + 1$, $t_k^* = t_{k+1} - 1$, and $t_k^{**} = \min\{t_k^*, n\}$. Then,

$$\begin{aligned}
\left|\frac{1}{n} \sum_{i=1}^n U_{ni} e_i\right| &= \left|\frac{1}{n} \sum_{k=0}^{N-1} \sum_{t=t_k}^{t_k^{**}} U_{nt} e_t\right| \\
&\le \left|\frac{1}{n} \sum_{k=0}^{N-1} U_{nt_k} \sum_{t=t_k}^{t_k^{**}} e_t\right| + \left|\frac{1}{n} \sum_{k=0}^{N-1} \sum_{t=t_k}^{t_k^{**}} [U_{nt} - U_{nt_k}] e_t\right| \\
&\le \frac{1}{n} \sum_{k=0}^{N-1} |U_{nt_k}| \left|\sum_{t=t_k}^{t_k^{**}} e_t\right| + \frac{1}{n} \sum_{k=0}^{N-1} \sum_{t=t_k}^{t_k^*} |U_{nt} - U_{nt_k}|\, |e_t| \\
&\le \sup_{0 \le s \le 1} |U_n(s)|\ \frac{1}{n} \sum_{k=0}^{N-1} \left|\sum_{t=t_k}^{t_k^*} e_t\right| \\
&\quad + \sup_{|r-s| \le \delta} |U_n(r) - U_n(s)|\ \frac{1}{n} \sum_{t=1}^n |e_t|.
\end{aligned}$$

Since $U_n(\cdot)$ converges weakly to a Brownian motion, it is clear that $\sup_{0 \le s \le 1} |U_n(s)| = O_p(1)$, while $\sum_{t=1}^n |e_t|/n = O_p(1)$ by using the same arguments as those used in the proof of (A.8). Further,

$$\begin{aligned}
E\left[\frac{1}{n} \sum_{k=0}^{N-1} \left|\sum_{t=t_k}^{t_k^*} e_t\right|\right] &\le \frac{N}{n} \sup_{0 \le k \le N-1} E\left|\sum_{t=t_k}^{t_k^*} e_t\right| \\
&\le \sup_{t \le n} E\left|\frac{1}{\delta\, n} \sum_{i=t}^{t+\delta\, n} e_i\right| \\
&\le C\,(\delta\, n\, h)^{-1/2} \to 0
\end{aligned}$$

by (A.10) as $n \to \infty$. Therefore, as $n \to \infty$,

$$\frac{1}{n} \sum_{i=1}^n U_{ni} e_i = o_p(1) + \sup_{|r-s| \le \delta} |U_n(r) - U_n(s)|\ O_p(1).$$

Note that as $n \to \infty$,

$$\sup_{|r-s| \le \delta} |U_n(r) - U_n(s)| \xrightarrow{d} \sup_{|r-s| \le \delta} |W_{\eta,2}(r) - W_{\eta,2}(s)| \xrightarrow{p} 0$$

as $\delta \to 0$. Hence (here we take the sequential limits: first let $n \to \infty$, and then let $\delta \to 0$),

$$\frac{1}{n} \sum_{i=1}^n U_{ni} e_i = o_p(1),$$

which, by combining (A.9), (2.6), and Lemma A.2, gives that

$$F_{n,j,1}(z) = E\left[K_h(Z_t - z) Z_{t,z,h}^j X_{t1}\right] \frac{1}{n} \sum_{t=1}^n U_{nt} + \frac{1}{n} \sum_{i=1}^n U_{ni} e_i = f_z(z) \mu_j(K) M_1(z) W_{\eta,2}^{(1)} + o_p(1). \tag{A.11}$$

Similarly,

$$F_{n,j,2}(z) = f_z(z) \mu_j(K) W_{\eta,2}^{(2)} + o_p(1). \tag{A.12}$$

Then, by plugging (A.9), (A.11) and (A.12) into $S_{n,j}(z)$, we have

$$S_{n,j}(z) = f_z(z) \mu_j(K) S(z) + o_p(1). \tag{A.13}$$

By noting that $\mu_0(K) = 1$ and $\mu_1(K) = 0$, we immediately obtain from (A.13) that

$$S_n(z) = f_z(z) \begin{pmatrix} 1 & 0 \\ 0 & \mu_2(K) \end{pmatrix} \otimes S(z) + o_p(1). \tag{A.14}$$

Let $R_n(z)^{-1}$ denote the upper-left corner $d \times d$ sub-matrix of $S_n(z)^{-1}$. From (A.14), we immediately obtain that

$$R_n(z)^{-1} = f_z(z)^{-1} S(z)^{-1} + o_p(1). \tag{A.15}$$

From (A.6), we have

$$D_n\left[\widehat{\beta}(z) - \beta(z)\right] \equiv I_3 + I_4, \tag{A.16}$$

where

$$I_3 = R_n(z)^{-1} B_n(z), \tag{A.17}$$

with

$$B_n(z) = n^{-1}\sum_{t=1}^{n} K_h(Z_t - z)\, D_n^{-1} X_t X_t^{\mathrm{T}} \times \{\beta(Z_t) - \beta(z) - (Z_t - z)\beta^{(1)}(z)\},$$

and

$$I_4 = R_n(z)^{-1}\, n^{-1} \sum_{t=1}^{n} K_h(Z_t - z)\, \varepsilon_t\, D_n^{-1} X_t.$$

Define,

$$\begin{aligned} G_{n,0}(z) &= n^{-1}\sum_{t=1}^{n} K_h(Z_t - z)\, X_{t1}^{\otimes 2} \\ &\quad \times \{\beta_1(Z_t) - \beta_1(z) - (Z_t - z)\beta_1^{(1)}(z)\}, \\ G_{n,1}(z) &= n^{-1}\sum_{t=1}^{n} K_h(Z_t - z)\, X_{t1}\, (X_{t2}/\sqrt{n})^{\mathrm{T}} \\ &\quad \times \{\beta_2(Z_t) - \beta_2(z) - (Z_t - z)\beta_2^{(1)}(z)\}, \\ G_{n,2}(z) &= n^{-1}\sum_{t=1}^{n} K_h(Z_t - z)\, (X_{t2}/\sqrt{n}) X_{t1}^{\mathrm{T}}\, \sqrt{n} \\ &\quad \times \{\beta_1(Z_t) - \beta_1(z) - (Z_t - z)\beta_1^{(1)}(z)\}, \end{aligned}$$

and

$$G_{n,3}(z) = n^{-1}\sum_{t=1}^{n} K_h(Z_t - z)\, (X_{t2}/\sqrt{n})^{\otimes 2}\, \sqrt{n} \times \{\beta_2(Z_t) - \beta_2(z) - (Z_t - z)\beta_2^{(1)}(z)\},$$

so that

$$B_n(z) = \begin{pmatrix} G_{n,0}(z) + G_{n,1}(z) \\ C_{n,2}(z) + G_{n,3}(z) \end{pmatrix}. \tag{A.18}$$

Similar to (A.9), by the kernel theory and an application of Taylor's expansion, it is easy to show that

$$E[G_{n,0}(z)] = h^2 f_z(z)\, M_2(z)\left[\frac{\mu_2(K)}{2}\beta_1^{(2)}(z)\right]\{1 + o(1)\}$$

and $\mathrm{Var}[G_{n,0}(z)] = o(1)$, so that

$$G_{n,0}(z) = h^2 f_z(z)\, M_2(z)\left[\frac{\mu_2(K)}{2}\beta_1^{(2)}(z)\right]\{1 + o_p(1)\}.$$

Further, following the proof of (A.11), we can easily show that

$$G_{n,1}(z) = h^2 f_z(z)\, M_1(z)\, W_{\eta,2}^{(1)^{\mathrm{T}}}\left[\frac{\mu_2(K)}{2}\beta_2^{(2)}(z)\right]\{1 + o_p(1)\},$$

$$G_{n,2}(z) = h^2 f_z(z)\, M_1(z)\, W_{\eta,2}^{(1)}\, \sqrt{n}\left[\frac{\mu_2(K)}{2}\beta_1^{(2)}(z)\right]\{1 + o_p(1)\},$$

and

$$G_{n,3}(z) = h^2 f_z(z)\, W_{\eta,2}^{(2)} \sqrt{n}\left[\frac{\mu_2(K)}{2}\beta_2^{(2)}(z)\right]\{1 + o_p(1)\}.$$

Plugging the above results into (A.18), we obtain

$$B_n(z) = h^2 f_z(z)\, S(z)\, D_n\left[\frac{\mu_2(K)}{2}\beta^{(2)}(z)\right]\{1 + o_p(1)\}. \tag{A.19}$$

Substituting (A.19) into (A.17) and using (A.15) lead to

$$I_3 = D_n\, h^2\, B_\beta(z)\, \{1 + o_p(1)\}.$$

Therefore,

$$D_n^{-1} I_3 = h^2 B_\beta(z) + o_p(h^2). \tag{A.20}$$

Finally, we consider I_4. Define

$$T_n(z) = \sqrt{\frac{h}{n}}\sum_{t=1}^{n} K_h(Z_t - z)\varepsilon_t\, D_n^{-1} X_t = \begin{pmatrix} T_{n,1}(z) \\ T_{n,2}(z) \end{pmatrix}$$

with $T_{n,1}(z) = \sqrt{\frac{h}{n}}\sum_{t=1}^{n} K_h(Z_t - z)\,\varepsilon_t\, X_{t1}$ and $T_{n,2}(z) = \sqrt{\frac{h}{n}}\sum_{t=1}^{n} K_h(Z_t - z)\,\varepsilon_t\, X_{t2}/\sqrt{n}$. By combining the above expressions with (A.16) and (A.20), we obtain

$$\sqrt{n\,h}\, D_n\left[\widehat{\beta}(z) - \beta(z) - h^2 B_\beta(z) + o_p(h^2)\right] = R_n(z)^{-1}\, T_n(z). \tag{A.21}$$

To prove the asymptotic normality of the left hand side of (A.21), it suffices to establish the asymptotic normality of $T_n(z)$. Note that $T_{n,1}$ only involves stationary variables. Hence, by the kernel estimation theory for stationary mixing data; see Theorem 2 of Cai et al. (2000) for details, we have

$$\begin{aligned} T_{n,1}(z) &\xrightarrow{d} N\left(0,\ \sigma_\varepsilon^2 \nu_0(K) f_z(z)\, M_2(z)\right) \\ &= \sqrt{\nu_0(K) f_z(z)}\, W_\varepsilon(1), \end{aligned} \tag{A.22}$$

where $W_\varepsilon(r)$ is a p_1-dimensional Brownian motion on $[0,\ 1]$ with covariance matrix $\sigma_\varepsilon^2 M_2(z)$. From (A.22) and note that the first element of X_{t1} is one, we immediately obtain

$$\begin{aligned} \sqrt{h/n}\sum_{t=1}^{n} K_h(Z_t - z)\, \varepsilon_t &\xrightarrow{d} N\left(0,\ \sigma_\varepsilon^2 \nu_0(K) f_z(z)\right) \\ &= \sqrt{\nu_0(K) f_z(z)}\, W_{\varepsilon,1}(1), \end{aligned} \tag{A.23}$$

where $W_{\varepsilon,1}(r)$ is the first element of $W_\varepsilon(r)$. Note that using the notations of Lemmas A.1–A.3, we have $U_n(r) = X_{[nr]2}/\sqrt{n}$, $W_n(r) = \sum_{s=1}^{[nr]} w_s/\sqrt{n}$ and $T_{n,2}(z) = \int_0^1 U_n(t)\mathrm{d}W_n(t)$. Hence, (A.22), in conjunction with Lemmas A.1 and A.3

$$T_{n,2}(z) \xrightarrow{d} \sqrt{\nu_0(K) f_z(z)}\int_0^1 W_{\eta,2}(r)\mathrm{d}W_{\varepsilon,1}(r). \tag{A.24}$$

Therefore, a combination of (A.22) and (A.24) leads to

$$T_n(z) \xrightarrow{d} \sqrt{\nu_0(K) f_z(z)}\begin{pmatrix} W_\varepsilon(1) \\ \int_0^1 W_{\eta,2}(r)\mathrm{d}W_{\varepsilon,1}(r) \end{pmatrix}.$$

Since $W_{\eta,2}(\cdot)$ and $W_\varepsilon(\cdot)$ are uncorrelated (because Z_t and ε_t are uncorrelated), $\int_0^1 W_{\eta,2}(r)\,\mathrm{d}W_\varepsilon(r)$ has a mixed normal distribution, so that the conditional covariance of $\begin{pmatrix} W_\varepsilon(1) \\ \int_0^1 W_{\eta,2}(r)\mathrm{d}W_{\varepsilon,1}(r) \end{pmatrix}$ is

$$\sigma_\varepsilon^2\begin{pmatrix} M_2(z) & M_1(z) W_{\eta,2}^{(1)^{\mathrm{T}}} \\ W_{\eta,2}^{(1)}\, M_1(z)^{\mathrm{T}} & W_{\eta,2}^{(2)} \end{pmatrix} = \sigma_\varepsilon^2 S(z). \tag{A.25}$$

Therefore, by Slusky's theorem, we have

$$\sqrt{n\,h}\,D_n\left[\widehat{\beta}(z)-\beta(z)-B_\beta(z)+o_p(h^2)\right]$$
$$\xrightarrow{d} f_z^{-1/2}(z)\nu_0^{1/2}(K)S(z)^{-1}\begin{pmatrix} W_\varepsilon(1)\\ \int_0^1 W_{\eta,2}(r)\mathrm{d}W_{\varepsilon,1}(r)\end{pmatrix}. \qquad (A.26)$$

It is easy to show using (A.25) that the conditional variance of the right hand side of (A.26) is $\Sigma_\beta(z)$ as given in Theorem 2.1. □

Proof of Theorem 2.2. To simplify the notation, in what follows, we drop the subscript "2 step" in $\widehat{\beta}_{1,2\,step}(z)$ and $\widehat{\beta}^{(1)}_{1,2\,step}(z)$. First, define

$$L_{n,j}(z)=\frac{1}{n}\sum_{t=1}^{n}Z^j_{t,z,h_1}\,X_{t1}\,X_{t1}^{\mathrm{T}}\,K_{h_1}(Z_t-z).$$

Then, similar to (A.9), one can show that

$$L_{n,j}(z)=f_z(z)\,M_2(z)\,\mu_j(K)+o_p(1). \qquad (A.27)$$

Therefore,

$$n^{-1}\sum_{t=1}^{n}\begin{pmatrix}X_{t1}\\(Z_t-z)X_{t1}\end{pmatrix}^{\otimes 2}K_{h_1}(Z_t-z)=\begin{pmatrix}L_{n,0}(z) & h_2\,L_{n,1}(z)^{\mathrm{T}}\\ h_1\,L_{n,1}(z) & h_1^2\,L_{n,2}(z)\end{pmatrix}$$
$$=\begin{pmatrix}1 & 0\\ 0 & h_1^2\end{pmatrix}\otimes M_2(z)\{1+o_p(1)\}.$$

By (2.14),

$$\begin{pmatrix}\widehat{\beta}_1(z)-\beta_1(z)\\ \widehat{\beta}^{(1)}_1(z)-\beta^{(1)}_1(z)\end{pmatrix}=\left[\sum_{t=1}^{n}\begin{pmatrix}X_{t1}\\(Z_t-z)\,X_{t1}\end{pmatrix}^{\otimes 2}K_{h_1}(Z_t-z)\right]^{-1}$$
$$\times\sum_{t=1}^{n}\begin{pmatrix}X_{t1}\\(Z_t-z)X_{t1}\end{pmatrix}$$
$$\times\left[Y_t^*-\beta_1(z)^{\mathrm{T}}X_{t1}-\beta_1^{(1)}(z)^{\mathrm{T}}X_{t1}(Z_t-z)\right]K_{h_1}(Z_t-z),$$

so that

$$\begin{aligned}\widehat{\beta}_1(z)-\beta_1(z)&=M_2(z)^{-1}\,n^{-1}\sum_{t=1}^{n}X_{t1}\left[Y_t^*-X_{t1}^{\mathrm{T}}\beta_1(z)\right.\\&\quad\left.-X_{t1}^{\mathrm{T}}\beta_1^{(1)}(z)(Z_t-z)\right]K_{h_1}(Z_t-z)\{1+o_p(1)\}\\&=M_2(z)^{-1}\,n^{-1}\sum_{t=1}^{n}X_{t1}X_{t1}^{\mathrm{T}}\left[\beta_1(Z_t)-\beta_1(z)\right.\\&\quad\left.-\beta_1^{(1)}(z)(Z_t-z)\right]K_{h_1}(Z_t-z)\{1+o_p(1)\}\\&\quad+M_2(z)^{-1}\sum_{t=1}^{n}X_{t1}\,\varepsilon_t\,K_{h_1}(Z_t-z)\{1+o_p(1)\}\\&\quad+M_2(z)^{-1}\,n^{-1}\sum_{t=1}^{n}X_{t1}X_{t2}^{\mathrm{T}}\left[\beta_2(Z_t)-\widehat{\beta}_2(Z_t)\right]\\&\quad\times K_{h_1}(Z_t-z)\{1+o_p(1)\}\\&\equiv J_1+J_2+J_3.\end{aligned}$$

Based on the kernel theory for the stationary mixing case; see Theorems 1 and 2 of Cai et al. (2000) for details, one can easily show that

$$J_1=\frac{h_1^2}{2}\,\mu_2(K)\,\beta_1^{(2)}(z)\{1+o_p(1)\}\quad\text{and}$$
$$\sqrt{n\,h_1}\,J_2\xrightarrow{d}N(0,\ \Sigma_{\beta_1,0}(z)).$$

Finally, similar to the proof of (2.15), by using the same arguments as those used in the proof of Theorem 1 in Cai (2002c), one can show that

$$|J_3|=O_p(h_2^2\sqrt{n})=o_p(h_1^2).$$

Hence,

$$\sqrt{n\,h_1}\left[\widehat{\beta}_1(z)-\beta_1(z)-\frac{h_1^2}{2}\,\mu_2(K)\,\beta_1^{(2)}(z)+o_p(h_1^2)\right]$$
$$=\sqrt{n\,h_1}J_2\xrightarrow{d}N(0,\ \Sigma_{\beta_1,0}(z)).$$

This proves the theorem. □

Appendix B. Proof of Theorem 3.1

Before we prove Theorem 3.1, we first provide some auxiliary results, which will be used sequently. Define, for any $j\geq 0$, $K_j(u)=u^j\,K(u)$. Then, it is easy to verify that similar to $K(\cdot)$, $K_j(\cdot)$ is continuous and has a compact support. Also, both $K_j(\cdot)$ and $K_j^2(\cdot)$ are integrable. Re-define $S_n(z)$ in the proof of Theorem 2.1 as follows

$$S_n(z)=n^{-1/2}\sum_{t=1}^{n}K_h(Z_t-z)\begin{pmatrix}1\\Z_{t,z,h}\end{pmatrix}^{\otimes 2}\otimes X_t\,X_t^{\mathrm{T}}$$
$$=\begin{pmatrix}S_{n,0}(z) & S_{n,1}(z)\\ S_{n,1}(z) & S_{n,2}(z)\end{pmatrix}$$

with $Z_{t,z,h}=(Z_t-z)/h$ and $S_{n,j}(z)=n^{-1/2}\sum_{t=1}^{n}K_{j,h}(Z_t-z)\,X_t\,X_t^{\mathrm{T}}$ for $j=0,1,2$, where $K_{j,h}(v)=K_j(v/h)/h$. Also, set,

$$\widehat{\mu}_j(z)=\frac{1}{\sqrt{n}}\sum_{t=1}^{n}K_{j,h}(Z_t-z)=\frac{\beta_n}{n}\sum_{t=1}^{n}K_j(\beta_n(\gamma_n^{-1}Z_t+x_n)),$$

where $\beta_n=\sqrt{n}/h$, $\gamma_n=\sqrt{n}$, and $x_n=-z/\sqrt{n}$. Clearly, $x_n\to 0$ for any fixed z and $x_n=-a$ if $z=a\sqrt{n}$. Further, let $\phi_\varepsilon(x)=(\sqrt{2\,\pi}\,\varepsilon)^{-1/2}\exp\left(-x^2/(2\,\varepsilon^2)\right)$ for any $\varepsilon>0$. Finally, let $o_{L_2}(1)$ denote the convergence in L_2 which implies $o_p(1)$. We use the notation $\mathcal{A}_{1n}=\mathcal{A}_{2n}+(s.o.)$ to denote that $\mathcal{A}_{2n}$ has the same order as $\mathcal{A}_{1n}$ and $(s.o.)$ denotes the terms having orders smaller than $\mathcal{A}_{2n}$. In what follows, we assume that Z_t satisfies (3.3). We present some preliminary results.

Lemma B.1. *Under Assumptions given in Theorem 3.1,*

$$\text{(i)}\quad\widehat{\mu}_j(z)\xrightarrow{p}\begin{cases}\mu_j(K)\,L(1,0)/\sigma_u, & \text{if } z \text{ is fixed},\\ \mu_j(K)\,L(1,a/\sigma_u)/\sigma_u, & \text{if } z=a\sqrt{n},\end{cases}$$

and for any $p>0$ and z,

(ii) $E\left[\widehat{\mu}_j(z)\right]=O(1)$, *and*

(iii) $E\left[|K_{j,h}(Z_t-z)|^p\right]=O(t^{-1/2}\,h^{1-p})$.

Proof. To establish the first assertion, we use some results from Jeganathan (2004). Indeed, by Proposition 6 and Lemma 7 of Jeganathan (2004), for each $\varepsilon>0$,

$$\widehat{\mu}_j(z)=\frac{\mu_j(K)}{n}\sum_{t=1}^{n}\phi_\varepsilon(\gamma_n^{-1}\,Z_t+x_n)+o_{L_2}(1).$$

Since $\phi_\varepsilon(z)$ satisfies the Lipschitz condition and $x_n\to 0$,

$$\widehat{\mu}_j(z)=\frac{\mu_j(K)}{n}\sum_{t=1}^{n}\phi_\varepsilon(\gamma_n^{-1}\,Z_t)+o_{L_2}(1)$$
$$=\frac{\mu_j(K)}{n}\sum_{t=1}^{n}\phi_\varepsilon(W_u(t/n))+o_{L_2}(1)$$

in view of (3.7) and (2.7). By Lemma 9 of Jeganathan (2004), we have

$$\widehat{\mu}_j(z)=\mu_j(K)\int_0^1\phi_\varepsilon(W_u(s))\mathrm{d}s+o_{L_2}(1).$$

An application of Proposition 11 of Jeganathan (2004) gives

$$\widehat{\mu}_j(z) = \mu_j(K)\, L(1,0)/\sigma_u + o_{L_2}(1)$$

as $\varepsilon \downarrow 0$. By the same token, it is easy to show the case of $x_n = -a$ ($z = a\sqrt{n}$). For assertion (ii), we have

$$\begin{aligned} E\left[\widehat{\mu}_j(z)\right] &= n^{-1/2}\sum_{t=1}^{n} E\left[K_{j,h}(Z_t - z)\right] \\ &= n^{-1/2}\, h^{-1}\sum_{t=1}^{n}\int K_j(t^{1/2}u/h) f_{t,z}(u)\mathrm{d}u \\ &= n^{-1/2}\sum_{t=1}^{n} t^{-1/2}\int K_j(v) f_{t,z}(ht^{-1/2}v)\mathrm{d}v \\ &\le Cn^{-1/2}\sum_{t=1}^{n} t^{-1/2} = O(1). \end{aligned}$$

Finally, recall that $K_{j,h}(u) = h^{-1}K_j(u/h)$ and $K_j(u) = u^j K(u)$, it can be shown easily by the boundedness of $f_{t,z}(\cdot)$ that

$$E\left[|K_{j,h}(Z_t - z)|^p\right] = h^{-p}\int |K_{j,h}(t^{1/2}u/h)|^p f_t(u)\mathrm{d}u$$
$$= t^{-1/2}h^{1-p}\int |K_j(v)|^p f_{t,z}(t^{-1/2}hv)\mathrm{d}v \le Ct^{-1/2}h^{1-p}.$$

This proves the lemma. □

Lemma B.2. *Under Assumptions given in Theorem 3.1, if z is fixed, we have*

$$S_{n,j}(z) = E(X_tX_t^{\mathrm{T}})[\widehat{\mu}_j(z)] + o_p(1) \xrightarrow{p} E(X_tX_t^{\mathrm{T}})\, \mu_j(K)\, L(1,0)/\sigma_u.$$

Proof. Recall that $V_t = X_tX_t^{\mathrm{T}}$. By adding and subtracting $E(V_t)$ and $E(V_t|Z_t)$ in $S_{n,j}(z)$, we decompose $S_{n,j}(z)$ into three terms as follows:

$$S_{n,j}(z) = B_{1n,1} + B_{1n,2} + B_{1n,3},$$

where $B_{1n,1} = E(V_t)\, n^{-1/2}\sum_t K_{j,h}(Z_t - z) = E(V_t)[\widehat{\mu}_j(z)]$,

$$\begin{aligned} B_{1n,2} &= n^{-1/2}\sum_t [E(V_t|Z_t) - E(V_t)]K_{j,h}(Z_t - z) \\ &= n^{-1/2}\sum_t \delta_t\, g_t(Z_t)\, K_{j,h}(Z_t - z), \end{aligned}$$

and

$$\begin{aligned} B_{1n,3} &= n^{-1/2}\sum_t [V_t - E(V_t|Z_t)]K_{j,h}(Z_t - z) \\ &\equiv n^{-1/2}\sum_t \zeta_t\, K_{j,h}(Z_t - z), \end{aligned}$$

where $\zeta_t = V_t - E(V_t|Z_t)$. It follows from Lemma B.1 that $B_{1n,1} \xrightarrow{p} \mu_j(K)\, E(V_t)L(1,0)/\sigma_u$. To show the lemma, it suffices to show that $B_{1n,2} = o_p(1)$ and $B_{1n,3} = o_p(1)$, respectively.

First, we show that $B_{1n,2} = o_p(1)$. By (3.5) and the boundedness of $g_t(\cdot)$ as well as $\delta_t = O(t^{-1/2})$ (see Assumption C3), we have

$$\begin{aligned} |B_{1n,2}| &\le n^{-1/2}\sum_{t=1}^{n} |\delta_t||g_t(Z_t)||K_{j,h}(Z_t - z)| \\ &\le C\, n^{-1/2}\sum_{t=1}^{n} t^{-1/2}C(Z_t)|K_{j,h}(Z_t - z)| \\ &= C\, n^{-1/2}\sum_{t=1}^{n} t^{-1/2}C(Z_t - z + z)|K_{j,h}(Z_t - z)|. \end{aligned}$$

Since $C(u)$ is continuous at z and $K_j(\cdot)$ has a finite support, then $C(u) \le C_z$ for some C_z for all u's in a neighborhood of z. Therefore,

$$|B_{1n,2}| \le C\, n^{-1/2}\sum_{t=1}^{n} t^{-1/2}|K_{j,h}(Z_t - z)| \xrightarrow{p} 0$$

by Lemma B.1 and Toeplitz lemma.

Next, we show that $B_{1n,3} = o_p(1)$. To do so, it suffices to show that $E[B_{1n,3}^2] = o(1)$. To this end, we have (ζ^2 below means $\zeta_{t,ij}^2$ for any i and j, see the discussions below Lemma A.2 for our notation adoption)

$$\begin{aligned} E[B_{1n,3}^2] &= n^{-1}\sum_{t=1}^{n} E\left[\zeta_t^2 K_{j,h}^2(Z_t - z)\right] \\ &\quad + 2\, n^{-1}\sum_{1\le s<t\le n} E[\zeta_t\zeta_s K_{j,h}(Z_t - z)K_{j,h}(Z_s - z)] \\ &\equiv B_{1n,31} + B_{1n,32}. \end{aligned}$$

Clearly, by Cauchy–Schwartz inequality, Lemma B.1, and Assumption C1, we have

$$\begin{aligned} |B_{1n,31}| &\le n^{-1}\sum_{t=1}^{n} \|\zeta_t\|_{2q}^2 \|K_{j,h}(Z_t - z)\|_{2p}^2 \\ &\le C\, n^{-1}\sum_t t^{-1/2p}\, h^{1/p-2} \\ &\le C\left(nh^{4p-2}\right)^{-1/2p} = o(1), \end{aligned}$$

since $n\, h^{4p-2} \to \infty$ (from Assumption C5). For $B_{1n,32}$, by (3.5) and (3.6), we have

$$\begin{aligned} E(\zeta_t|Z_t, Z_s, V_s) &= E(V_t|Z_t, Z_s, V_s) - E(V_t|Z_t) \\ &= \delta_{t,s}g_{t,s}(Z_t, Z_s, V_s) - \delta_t g_t(Z_t). \end{aligned} \tag{B.1}$$

It is easy to see that

$$\begin{aligned} |B_{1n,32}| &\le C\, n^{-1}\sum_{1\le s<t\le n} |E\{E[\zeta_t|Z_t, Z_s, V_s]\zeta_s K_{j,h}(Z_t - z)K_{j,h}(Z_s - z)\}| \\ &\le \frac{C}{n}\sum_{1\le s<t\le n} |\delta_{t,s}|\ |E\{g_{t,s}(Z_t, Z_s, V_s)\zeta_s \\ &\qquad \times K_{j,h}(Z_t - z)K_{j,h}h(Z_s - z)\}| \\ &\quad + \frac{C}{n}\sum_{1\le s<t\le n} |\delta_t|\ |E[g_t(Z_t)\zeta_s K_{j,h}(Z_t - z)K_{j,h}(Z_s - z)]| \\ &\equiv B_{1n,32,1} + B_{1n,32,2}. \end{aligned}$$

To evaluate $B_{1n,32,1}$, first, we consider the following quantity

$$\begin{aligned} &E[g_{t,s}(Z_t, Z_s, V_s)\zeta_s K_{j,h}(Z_t - z)K_{j,h}(Z_s - z)|V_s] \\ &\quad = \int b_{t,s,z}(t^{-1/2}hu, s^{-1/2}hv)g_{t,s}(z + hu, z + hv, V_s) \\ &\qquad \times [V_s - E(V_s) - \delta_s g_s(z + hv)]K_j(u)K_j(v)|\, t^{-1/2}s^{-1/2}\mathrm{d}u\mathrm{d}v \\ &\quad = \mu_j^2(K)\, t^{-1/2}s^{-1/2}b_{t,s,z}(0,0)g_{t,s}(z, z, V_s) \\ &\qquad \times [V_s - E(V_s) - \delta_s g_s(z)]\{1 + o_p(1)\}. \end{aligned}$$

Then, using (3.5) and (B.1), we have

$$\begin{aligned} B_{1n,32,1} &\le \frac{C}{n}\sum_{1\le s<t\le n} |\delta_{t,s}|\, |E\{E[g_{t,s}(Z_t, Z_s, V_s)\, \zeta_s\, K_{j,h}(Z_t - z) \\ &\quad \times K_{j,h}(Z_s - z)|V_s]\}| \\ &\le C\, n^{-1}\sum_{s=1}^{n-1} s^{-1}\sum_{t=s+1}^{n} t^{-1/2}b_{t,s,z}(0,0) \\ &\quad \times E[|\, g_{t,s}(z, z, V_s)\, \{V_s - E(V_s) - \delta_s g_s(z)\}\{1 + o(1)\}\, |] \\ &= o(1) \end{aligned}$$

by Assumptions C2 and C3. Similarly, using the fact that $\delta_t = O(t^{-1/2})$, one can easily show that $B_{1n,32,2} = o(1)$. Thus, we have shown that $B_{1n,32} = o(1)$. By summarizing the above results, the lemma is proved. □

Lemma B.3. *Under Assumptions given in Theorem 3.1, then,*

$$B_{2n} = h^2 B_\beta(z) E(X_t X_t^{\mathrm{T}})[\widehat{\mu}_2(z)] + o_p(h^2)$$
$$= h^2 B_\beta(z) E(X_t X_t^{\mathrm{T}}) L(1,0)/\sigma_u + o_p(h^2).$$

Proof. The proof is similar to that for Lemma B.2. By adding and subtracting terms ($E(V_t)$ and $E(V_t|Z_t)$), we can decompose B_{2n} into three terms as

$$B_{2n} \equiv B_{2n,1} + B_{2n,2} + B_{2n,3},$$

where $B_{2n,1} = E(V_t)\, n^{-1/2} \sum_{t=1}^n \left[\beta(Z_t) - \beta(z) - \beta^{(1)}(z)(Z_t - z)\right] K_h(Z_t - z)$,

$$\begin{aligned} B_{2n,2} &= n^{-1/2} \sum_{t=1}^n [E(V_t|Z_t) - E(V_t)]\,[\beta(Z_t) - \beta(z) \\ &\quad - \beta^{(1)}(z)(Z_t - z)]\, K_h(Z_t - z) \\ &= n^{-1/2} \sum_{t=1}^n \delta_t\, g_t(Z_t) \left[\beta(Z_t) - \beta(z) - \beta^{(1)}(z)(Z_t - z)\right] \\ &\quad \times K_h(Z_t - z), \end{aligned}$$

and

$$\begin{aligned} B_{2n,3} &= n^{-1/2} \sum_t [V_t - E(V_t|Z_t)] \left[\beta(Z_t) - \beta(z) - \beta^{(1)}(z)(Z_t - z)\right] \\ &\quad \times K_h(Z_t - z) \\ &= n^{-1/2} \sum_t \zeta_t \left[\beta(Z_t) - \beta(z) - \beta^{(1)}(z)(Z_t - z)\right] K_h(Z_t - z). \end{aligned}$$

Next, we show that $B_{2n,1}$ contributes an asymptotic bias term and $B_{2n,2}$ and $B_{2n,3}$ are a higher order term like $o_p(h^2)$. First, we consider $B_{2n,1}$. By Lemma B.1, we have

$$\begin{aligned} B_{2n,1} &= E(V_t)\, n^{-1/2} \sum_{t=1}^n \left[\beta(Z_t) - \beta(z) - \beta^{(1)}(z)(Z_t - z)\right] \\ &\quad \times K_h(Z_t - z) \\ &= E(V_t)\, \frac{h^2}{2}\, \beta^{(2)}(z)\, \widehat{\mu}_2(z) + (s.o.) \\ &= \frac{h^2}{2}\, E(V_t)\, L(1,0)\, \beta^{(2)}(z)\, \mu_2(K)/\sigma_u + o_p(h^2). \end{aligned}$$

It remains to show that $B_{2n,2} = o_p(h^2)$ and $B_{2n,3} = o_p(h^2)$. First, we consider $B_{2n,2}$. Using (3.5) and the change of variable, we have

$$\begin{aligned} E\left[B_{2n,2}^2\right] &= \frac{1}{n} \sum_{t=1}^n \sum_{s=1}^n \delta_t \delta_s E\{g_t(Z_t) g_s(Z_s) \\ &\quad \times \left[\beta(Z_t) - \beta(z) - \beta^{(1)}(z)(Z_t - z)\right] K_h(Z_t - z) \\ &\quad \times \left[\beta(Z_s) - \beta(z) - \beta^{(1)}(z)(Z_s - z)\right] K_h(Z_s - z)\} \\ &= \frac{1}{n} \sum_{t=1}^n \sum_{s=1}^n \delta_t \delta_s\, t^{-1/2} s^{-1/2} \\ &\quad \times \int\!\!\int f_{t,s,z}(t^{-1/2} hu, s^{-1/2} hv) g_t(z + hu) g_s(z + hv) \\ &\quad \times [\beta(z + hu) - \beta(z) - \beta^{(1)}(z)\, h\, u] K(u) \\ &\quad \times [\beta(z + hv) - \beta(z) - \beta^{(1)}(z)\, h\, v] K(v) \mathrm{d}u \mathrm{d}v \\ &\le C\, \frac{h^4}{n} \sum_{t=1}^n \sum_{s=1}^n |\delta_t|\, |\delta_s|\, t^{-1/2} s^{-1/2} f_{t,s,z}(0,0) g_t(z) g_s(z) \\ &= O(n^{-1} h^4) = o(h^4) \end{aligned}$$

because $\delta_t = O(t^{-1/2})$ and $\delta_s = O(s^{-1/2})$, which implies that $B_{2n,2} = o_p(h^2)$. Finally, it suffices to show that $E\left[B_{2n,3}^2\right] = o(h^4)$. Similar to the evaluation of $B_{1n,3}$, we decompose $E\left[B_{2n,3}^2\right]$ into two terms as follows:

$$\begin{aligned} E\left[B_{2n,3}^2\right] &= \frac{1}{n} \sum_{t=1}^n E\left[\zeta_t^2 \{\beta(Z_t) - \beta(z) \right. \\ &\quad \left. - \beta^{(1)}(z)(Z_t - z)\}^2 K_h^2(Z_t - z)\right] \\ &\quad + \frac{2}{n} \sum_{1 \le s < t \le n} E[\zeta_t \zeta_s \{\beta(Z_t) - \beta(z) \\ &\quad - \beta^{(1)}(z)(Z_t - z)\} K_h(Z_t - z) \\ &\quad \times \{\beta(Z_s) - \beta(z) - \beta^{(1)}(z)(Z_s - z)\} K_h(Z_s - z)] \\ &\equiv B_{2n,31} + B_{2n,32}. \end{aligned}$$

Similar to the evaluation of $B_{1n,31}$, by Cauchy–Schwartz inequality and assumption C1, one can show easily that

$$\begin{aligned} |B_{2n,31}| &\le n^{-1} \sum_{t=1}^n \|\zeta_t\|_{2q}^2 \|[\beta(Z_t) - \beta(z) \\ &\quad - \beta^{(1)}(z)(Z_t - z)] K_h(Z_t - z)\|_{2p}^2 \\ &\le C\, n^{-1} \sum_t t^{-1/2p}\, h^{1/p} \le C \left(n h^{4p-2}\right)^{-1/2p} h^2 \\ &= C \left(n h^{8p-2}\right)^{-1/2p} h^4 = o(h^4) \end{aligned}$$

by Assumption C5. For $B_{2n,32}$, by analogy to $B_{1n,32}$, we have

$$B_{2n,32} \le C\, (nh^2)^{-1}\, h^4 \sum_{s=1}^{n-1} \sum_{t=s+1}^n (|\delta_{t,s}| + |\delta_s|)\, t^{-1/2} s^{-1/2} = o(h^4)$$

by Assumptions C2 and C3. This completes the proof of Lemma B.3. □

Lemma B.4. *Under Assumptions given in Theorem 3.1, then,*

$$n^{3/4} h^{1/2} B_{3n} \xrightarrow{d} MN(V^*),$$

where $MN(V^*)$ is a mixed normal with mean zero and covariance matrix

$$V^* = \sigma_\varepsilon^2 \nu_0(K) E(X_t X_t^{\mathrm{T}}) L(1,0)/\sigma_u.$$

Proof. Clearly, $E[B_{3n}] = 0$ because $E(\epsilon_t | X_t, Z_t) = 0$. Also, by the assumptions that $\{\epsilon_t\}$ is a martingale difference and $E(\varepsilon_t^2 | X_t, Z_t) = \sigma_\varepsilon^2$ (conditional homogenous errors), we conclude that the conditional variance of $n^{1/4} h^{1/2} B_{3n}$, given $\{(X_t, Z_t)\}$, is

$$V_{3n} = \frac{\sigma_\varepsilon^2\, h}{\sqrt{n}} \sum_{t=1}^n X_t X_t^{\mathrm{T}} K_h^2(Z_t - z).$$

Similar to the proof of Lemma B.2, we can show that

$$V_{3n} = \sigma_\varepsilon^2\, \nu_0(K)\, L(1,0)\, E(X_t X_t^{\mathrm{T}})/\sigma_u + o_p(1).$$

Finally, by virtue of a central limit theorem for a martingale difference (see, e.g., Hall and Heyde (1980, p. 58)),

$$n^{1/4} h^{1/2} B_{3n} \xrightarrow{d} MN(V^*).$$

This proves the lemma. □

Proof of Theorem 3.1. By Lemma B.1, we have

$$\begin{aligned} S_n(z) &= \begin{pmatrix} S_{n,0}(z) & S_{n,1}(z) \\ S_{n,1}(z) & S_{n,2}(z) \end{pmatrix} \\ &= \begin{pmatrix} 1 & 0 \\ 0 & h^2\, \mu_2(K) \end{pmatrix} \otimes E(X_t X_t^{\mathrm{T}})\, L(1,0)/\sigma_u \{1 + o_p(1)\}, \end{aligned}$$

which, by replacing Y_t in (3.2) by $Y_t = X_t^{\mathrm{T}}\beta(Z_t) + \epsilon_t$, implies that

$$\widehat{\beta}(z) - \beta(z) = \left[E(X_t X_t^{\mathrm{T}})\, L(1,0)/\sigma_u\right]^{-1} \times \left\{ n^{-1/2} \sum_{t=1}^{n} X_t X_t^{\mathrm{T}} \left[\beta(Z_t) - \beta(z) - \beta^{(1)}(z)(Z_t - z)\right] \times K_h(Z_t - z) + n^{-1/2} \sum_{t=1}^{n} X_t \epsilon_t K_h(Z_t - z) \right\} \{1 + o_p(1)\}$$

$$\equiv \left[E(X_t X_t^{\mathrm{T}})\, L(1,0)/\sigma_u\right]^{-1} \{B_{2n} + B_{3n}\}\{1 + o_p(1)\}, \qquad \text{(B.2)}$$

where $B_{2n} = n^{-1/2} \sum_t X_t X_t^{\mathrm{T}} \left[\beta(Z_t) - \beta(z) - \beta^{(1)}(z)(Z_t - z)\right] K_h(Z_t - z)$ and $B_{3n} = n^{-1/2} \sum_{t=1}^{n} X_t\, \epsilon_t\, K_h(Z_t - z)$. The asymptotic behaviors of B_{2n} and B_{3n} are derived in Lemmas B.3 and B.4. Therefore, combining Lemmas B.3 and B.4 with (B.2), we obtain that

$$n^{1/4} h^{1/2} \left[\widehat{\beta}(z) - \beta(z) - h^2 B_\beta(z) + o_p(h^2)\right] = \sigma_u \left[L(1,0)\, E(X_t X_t^{\mathrm{T}})\right]^{-1} n^{1/4} h^{1/2} B_{3n}\{1 + o_p(1)\} \xrightarrow{d} MN(\Sigma_1).$$

This completes the proof of Theorem 3.1. □

References

Berkes, I., Horváth, L., 2006. Convergence of integral functionals of stochastic processes. Econometric Theory 22, 304–322.
Billingsley, P., 1999. Convergence of Probability Measures, 2nd ed. Wiley, New York.
Cai, Z., 2002a. Regression quantile for time series. Econometric Theory 18, 169–192.
Cai, Z., 2002b. Two-step likelihood estimation procedure for varying-coefficient models. Journal of Multivariate Analysis 82, 189–209.
Cai, Z., 2002c. A two-stage approach to additive time series models. Statistica Neerlandica 56, 415–433.
Cai, Z., 2007. Trending time varying coefficient time series models with serially correlated errors. Journal of Econometrics 137, 163–188.
Cai, Z., Fan, J., Yao, Q., 2000. Functional-coefficient regression models for nonlinear time series. Journal of the American Statistical Association 95, 941–956.
Cai, Z., Wang, Y., 2008. Instability of predictability of stock returns. Working Paper. Department of Mathematics and Statistics, University of North Carolina at Charlotte.
Campbell, J.Y., Yogo, M., 2006. Efficient tests of stock return predictability. Journal of Financial Economics 81, 27–60.
Carrasco, M., Chen, X., 2002. Mixing and moment properties of various GARCH and stochastic volatility models. Econometric Theory 18, 17–39.
Chang, Y., Martinez-Chombo, E., 2003. Electricity demand analysis using cointegration and error-correction models with time varying parameters: The Mexican case. Working Paper. Department of Economics, Texas A&M University.
Chen, B., Hong, Y., 2007. Testing for smooth structural changes in time series models via nonparametric regression. Working Paper. Department of Economics, Cornell University.
Chen, S.X., Tang, C.Y., 2005. Nonparametric inference of value at risk for dependent financial returns. Journal of Financial Econometrics 3, 227–255.
Chen, R., Tsay, R.S., 1993. Functional-coefficient autoregressive models. Journal of the American Statistical Association 88, 298–308.
Csörgő, M., Révés, P., 1981. Strong Approximation in Probability and Statistics. Academic Press, New York.
Einmahl, U., 1987. A useful estimate in the multidimensional invariance principle. Probability Theory and Related Fields 76, 81–101.
Fan, J., Huang, T., 2005. Profile likelihood inferences on semiparametric varying-coefficient partially linear models. Bernoulli 11, 1031–1057.
Fan, J., Gijbels, I., 1996. Local Polynomial Modeling and its Applications. Chapman and Hall, London.
Hall, P., Heyde, C.C., 1980. Martingale Limit Theory and its Applications. Academic Press, New York.
Hansen, B.E., 1992. Convergence to stochastic integrals for dependent heterogeneous processes. Econometric Theory 8, 489–500.
Hansen, B.E., 2008. Uniform convergence rates for kernel estimation with dependent data. Econometric Theory 24, 726–748.
Hastie, T., Tibshirani, R., 1993. Varying coefficient models (with discussions). Journal of the Royal Statistical Society, Series B 55, 757–796.
Hurvich, C.M., Simonoff, J.S., Tsai, C.L., 1998. Smoothing parameter selection in nonparametric regression using an improved Akaike information criterion. Journal of the Royal Statistical Society, Series B 60, 271–293.
Jeganathan, P., 2004. Convergence of functionals of sums of random variables to local times of fractional stable motions. The Annals of Probability 32, 1771–1795.
Juhl, T., 2005. Functional coefficient models under unit root behavior. Econometrics Journal 8, 197–213.
Karatzas, I., Shreve, S.E., 1991. Brownian Motion and Stochastic Calculus, 2nd ed. Springer-Verlag, New York.
Karlsen, H.A., Myklebust, T., Tjøstheim, D., 2007. Nonparametric estimation in a nonlinear cointegration type model. Annals of Statistics 35, 252–299.
Li, Q., Huang, C., Li, D., Fu, T., 2002. Semiparametric smooth coefficient models. Journal of Business and Economic Statistics 20, 412–422.
Linton, O.B., 2000. Efficient estimation of generalized additive nonparametric regression models. Econometric Theory 16, 502–523.
McLeish, D.L., 1975. A maximal inequality and dependent strong laws. The Annals of Probability 3, 829–839.
Merlevéde, F., Peligrad, M., Utev, S., 2006. Recent advances in invariance principles for stationary sequences. Probability Surveys 3, 1–36.
Park, J.Y., Hahn, S,B., 1999. Cointegrating regressions with time varying coefficients. Econometric Theory 15, 664–703.
Phillips, P.C.B., 1989. Partially identified econometric models. Econometric Theory 5, 181–240.
Park, J.Y., Phillips, P.C.B., 1999. Asymptotics for nonlinear transformations of integrated time series. Econometric Theory 15, 269–298.
Phillips, P.C.B., Park, J., 1998. Nonstationary density and kernel autoregression. Unpublished manuscript.
Polk, C., Thompson, S., Vuolteenaho, T., 2006. Cross-sectional forecasts of the equity premium. Journal of Financial Economics 81, 101–141.
Robinson, P.M., 1988. Root-n consistent semiparametric regression. Econometrica 56, 931–954.
Robinson, P.M., 1989. Nonparametric estimation of time-varying parameters. In: Hackl, P. (Ed.), Statistical Analysis and Forecasting of Economic Structural Change. Springer-Verlag, Berlin, pp. 253–264.
Robinson, P.M., 1991. Time-varying nonlinear regression. In: Hackl, P. (Ed.), Statistical Analysis and Forecasting of Economic Structural Change. Springer-Verlag, Berlin, pp. 179–190.
Ruppert, D., Sheather, S.J., Wand, M.P., 1995. An effective bandwidth selection for local least squares regression. Journal of the American Statistical Association 90, 1257–1270.
Shao, Q., Lu, C., 1987. Strong approximation for partial sums of weakly dependent random variables. Scientia Sinica 15, 576–587.
Torous, W., Valkanov, R., Yan, S., 2004. On predicting stock returns with nearly integrated explanatory variables. The Journal of Business 77, 937–966.
Wang, Q., Phillips, P.C.B., 2006. Asymptotic theory for local time density estimation and nonparametric cointegrating regression. Econometric Theory (forthcoming).
Wang, Q., Phillips, P.C.B., 2008. Structural nonparametric cointegrating regression. Manuscript.
Xiao, Z., 2007. Functional coefficient co-integration models. Journal of Econometrics (forthcoming).