Functional-coefficient models for nonstationary time series data

Zongwu Cai\textsuperscript{a,b,c,}\textsuperscript{,*}, Qi Li\textsuperscript{d,e}, Joon Y. Park\textsuperscript{d}

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

\textbf{A B S T R A C T}

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 dependent data. However, little is known about the behavior with nonstationary (in particular, integrated with order one, denoted by I(1)) data, which have predominantly been modeled linearly. The early nonparametric asymptotic analyses with nonstationary data include Phillips and Park (1998), Park and Hahn (1999), Chang and Martínez-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, 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, Karlsten 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

\begin{equation}
Y_t = \beta(Z_t)^T X_t + \epsilon_t, \quad 1 \leq t \leq n, \tag{1.1}
\end{equation}

where $Y_t$, $Z_t$ and $\epsilon_t$ are scalar, $X_t = (X_{t1}, \ldots, X_{td})^T$ 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
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 \(l(0)\)) 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 \(\varepsilon_t\) 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 \(Z_t = l(1)\), 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 \(Z_t = l(0)\) and \(X_t = l(1), \text{model (1.1)}\) 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 \(X_t = (1, X_t')\). Further, if the coefficient vector associated with \(X_t\) is a vector of constants, say \(\gamma\), then the varying coefficient model reduces to a partially linear model

\[
E(Y_t | X_t, Z_t) = \beta_1(Z_t) + X_t' \gamma;
\]

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 \(l(1)\) and \(Z_t\) is strictly stationary. For expository simplicity, we re-express (1.1) as the following varying coefficient model

\[
Y_t = \beta(Z_t)' X_t + \varepsilon_t = \beta_1(Z_t)' X_{t1} + \beta_2(Z_t)' X_{t2} + \varepsilon_t,
\]

where \(X_{t1}, X_{t2}, \text{and } \varepsilon_t\) are stationary, \(X_{t2}\) is an \(l(1)\) vector, \(\beta(Z_t) = (\beta_1(Z_t)', \beta_2(Z_t)')'\), and \(X_t = (X_{t1}', X_{t2}')\), where \(X_{t2}\) is a \(d_1 \times 1\) 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 \(E(\varepsilon_t | X_t, Z_t) = 0\) which implies that \(X_t\) and \(Z_t\) are uncorrelated with \(\varepsilon_t\). 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 \(X_{t2} = (P_t, P_t', I_t)\) (and no \(X_{t1}\)), where \(P_t\) and \(P_t'\) are the price levels of the domestic and a foreign country, \(I_t\) is the exchange rate between the domestic and the foreign currencies, and \(Z_t = l_t - I_t^*\) is the difference between the domestic interest rate \(l_t\) and the foreign interest rate \(l_t^*\). Then if \(Y_t\) is an \(l(0)\) variable, we say that \(P_t\), \(P_t'\) and \(I_t\) 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 \(\beta\) 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 approximate as \(\hat{\beta}(z) + \beta^{(1)}(z)(Z_t - z)\) to approximate \(\beta(Z_t)\), where \(\beta^{(1)}(z) = d^2\beta(z)/dz^2\). Define

\[
\hat{\beta}(z) = \text{argmin}_{\theta_0, \theta_1} \sum_{t=1}^n [Y_t - \theta_0' X_t - (Z_t - \theta_1' X_t)^2]
\]

\[
\times K_h(Z_t - z),
\]

where \(K_h(u) = h^{-1} K(u/h) K(1)\) is a kernel function satisfying Assumption A3 below, \(\theta_0 = \beta(z)\) estimates \(\beta(z)\), and \(\theta_1 = \beta^{(1)}(z)\) estimates \(\beta^{(1)}(z)\). Then, \(\beta(z)\) and \(\beta^{(1)}(z)\) can be expressed as

\[
\hat{\beta}(z) = \left( \sum_{t=1}^n X_t (Z_t - z) \right)^{\otimes 2} K_h(0)
\]

\[
\times \sum_{t=1}^n \left( X_t (Z_t - z) \right) Y_t K_h(Z_t - z),
\]

where \(A^{\otimes 2} = AA'\) for a vector or matrix \(A\).

2.2. Notations and assumptions

Since \(X_{t2}\) is a vector of \(l(1)\) processes, it can be re-expressed as \(X_{t2} = X_{t2-1} + \eta_t = X_{t2-1} + \sum_{s=1}^{t-1} \eta_s\) for \(t \geq 1\), where \(\{\eta_t\} = \{l(l)\}\) process with mean zero and variance \(\Omega_\eta\). Then,

\[
X_{t2} / \sqrt{n} \Rightarrow X_{t2} / \sqrt{n} = X_{t2} / \sqrt{n} + \frac{1}{\sqrt{n}} \sum_{s=1}^{t-1} \eta_s = \frac{1}{\sqrt{n}} \sum_{s=1}^{[t-1]} \eta_s
\]

where \(l = t/n\) 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 \(\eta_t\) and \(\rho\)-mixing \(\eta_t\), respectively, generalizes in an obvious way to the multivariate cases and leads to

\[
\frac{X_{t2} / \sqrt{n}}{\sqrt{n}} \Rightarrow W_{n,2}(r) \quad \text{as } n \rightarrow \infty,
\]

where \(W_{n,2}(\cdot)\) is a \(d_2\)-dimensional Brownian motion on \([0, 1]\) with covariance matrix \(\Sigma_\eta\) and “\(\Rightarrow\)” represents weak convergence. In particular, it follows from Merlevède et al. (2006) that (2.4) holds if \(\eta_t\) is a stationarity \((\alpha, \gamma)\)-mixing sequence satisfying, for example,

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

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

\[
\frac{1}{n} \sum_{t=1}^{n} \Gamma(X_{t2} / \sqrt{n}) \Rightarrow \int_0^1 \Gamma(W_{n,2}(s))ds \quad \text{as } n \rightarrow \infty,
\]

where \(\Rightarrow\) denotes the convergence in distribution, so that, for \(I = 1, 2\),

\[
\frac{1}{n} \sum_{t=1}^{n} (X_{t2} / \sqrt{n})^{\otimes I} \Rightarrow \int_0^1 [W_{n,2}(r)]^{\otimes I} dr \equiv W_{n,2}^{(I)} \quad \text{as } n \rightarrow \infty;
\]

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

\[
\sup_{0 \leq r \leq 1} ||X_{t2} / \sqrt{n} - W_{n,2}(r)|| = O(n^{-\delta_0} \log^{1+\delta_0}(n))
\]
almost surely, where $\| \cdot \|$ is the usual $L_2$ norm in $\mathbb{R}^{d_k}$, $\theta_\ast = (1/2) - 1/(2 + \delta_*)$ and $\lambda_\ast = \lambda_\ast(\delta_\ast) > 0$ is a function of $\delta_\ast$, provided that $\{\eta_t\}$ is a stationary strong mixing sequence satisfying that, for some $\gamma_* > 2 + \delta_\ast$ with $0 < \delta_\ast \leq 2$,

$$E|\eta_1|^{\gamma_*} < \infty, \quad \text{and} \quad \sum_{n=1}^{\infty} \alpha(n)^{1/2 + \delta_\ast - 1/\gamma_*} < \infty;$$

(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 $\gamma_\ast$-th moment ($\gamma_* > 2$), then the right hand side of (2.7) becomes $o(n^{-\delta_*})$, where $\theta_\ast = (1/2) - 1/\gamma_\ast$; see, e.g., Csörgö and Révész (1981, p. 107). We assume throughout the paper that the sequence $\{\eta_t\}$ is stationary $\alpha$-mixing and satisfies either (2.5) or (2.8). Note that either (2.5) or (2.8) ensures that $\lim_{n \to \infty} \text{Var}(n^{-1/2} \sum_{t=1}^{n} \eta_t)$ exists and is finite by Davydov’s inequality for an $\alpha$-mixing process; see, e.g., Corollary A.2 in Hall and Heyde (1980).

Next, we give regularity conditions for the asymptotic distribution of $\hat{\beta}(z)$. We introduce the following notations. Let $f_j(z)$ denote the marginal density of $Z_j$, Define $M_j(z) = E[X_{jX}^\top | Z_i = z]$ for $k = 1$ and 2. Also, set, for $j \geq 0$, $\mu_j(K) = \int_{-\infty}^{\infty} v_j(K(v))dv$, and $v_j(K) = \int_{-\infty}^{\infty} v_j(K^2(v))dv$. Further, let

$$S(z) = \begin{bmatrix} M_2(z) & M_1(z)W_{n,2}^{(1)T} \\ W_{n,2}^{(1)}M_1(z)^T & W_{n,2}^{(2)} \end{bmatrix},$$

(2.9)

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

**Assumptions:**

A1. $\beta(z)$ is twice continuously differentiable in $z$ for all $z \in \mathcal{A}$.

A2. $M_j(z)$ is positive-definite and continuous in a neighborhood of $z$. $f_j(z)$ is continuously differentiable in a neighborhood of $z$ for $j = 1, 2$.

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. $\epsilon_1$ has a finite fourth moment, $E(\epsilon_1 | X_1, Z_i) = 0$, and $E(\epsilon_1^2 | X_1, Z_i) = \sigma_\epsilon^2$ is a positive constant.

A6. $(\{X_i, Z_t, \epsilon_t, \eta_t\}; t \geq 1)$ is a strictly $\alpha$-mixing stationary process with the $\delta_\ast$-th moment ($\delta_\ast > 2$). $E[|X_1|^2 | Z_i = z] \leq C_1 < \infty$ with $\delta_\ast > \delta_1 > 0$ and $\alpha(t) = O(t^{-\delta_1})$ for some $\delta_1 > \min\{\delta_t \delta_1/(\delta_2 - \delta_1), \delta_2, 2\delta_2/(\delta_2 - \delta_1)\}$ where $\delta_t = \delta_t(\delta_t \epsilon_1 - \delta_1 - \delta_2)$ for some $\delta_t$ satisfying $\delta_t/(\delta_t - 1) < \delta_2$. Also, $\|\eta_t\|_{\infty} = E[|\eta_1|^2]^{1/2} < \infty$ with $\eta_0 = \delta_0 \delta_{t}\epsilon_1(\delta_2 - \delta_1)$ for some $1 < \delta_0 < \delta_2$. Further, $\sup_{t} E[\eta_t^2 I_{\{\epsilon_{t+1} = 1\}} | Z_{t+1} = z] \leq C_2 < \infty$.

A7. $f(z_0, z, x_0, x_1; s) \leq M < \infty$ for $s \geq 1$, where $f(z_0, z, x_0, x_1; s)$ is the conditional density of $(Z_0, Z_t)$ given $(X_0 = x_0, X_1 = x_1)$ and $\sigma^2(X_1, Z_i), \text{i.e.}$, the conditional variance is only a function of the stationary covariates $(X_i, Z_i)$. However, it is technically difficult to let it also be a function of the nonstationary covariate $X_2$. 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 $k$ and is slightly stronger than the usual condition $n h \to \infty$. For optimal bandwidth selection (i.e., $h = c n^{-\gamma}$ for $0 < \gamma < 1, c > 0$), A8 is automatically satisfied for $\delta_1 \geq 2(1+\gamma)/(1-\gamma)$ and it is still fulfilled for $2 < \delta_1 \leq 2(1+\gamma)/(1-\gamma)$ if $\delta_1$ satisfies $\delta_1 < \delta_2 \leq 2\delta_1/(1+\gamma) - (1-\gamma)$. 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 $\hat{\beta}(z)$, we define $D_n = \text{diag}(l_{11}, \sqrt{n} l_{12})$, and $C_\theta(z) = \mu_2(K) \beta_2^{(2)}(z)/2$, where $l_{ij}$ 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

$$\frac{\sqrt{nh}}{\theta_1} \left[ \hat{\beta}_1(z) - \frac{1}{2} \beta_1(z) - h^2 \mu_2(K) \beta_2^{(2)}(z) \right] \xrightarrow{d} \text{MN}(\Sigma_\theta(z)),

\text{where } \text{MN}(\Sigma(z)) \text{ is a mixed normal distribution with mean zero and conditional covariance matrix given by } \Sigma_\theta(z) = \sigma^2_{\theta}(K) \Sigma(z)^{-1}/f_2(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 $X_2$), then Theorem 2.1 reduces to

$$\frac{\sqrt{nh}}{\theta_1} \left[ \hat{\beta}_1(z) - \frac{1}{2} \beta_1(z) - h^2 \mu_2(K) \beta_2^{(2)}(z) \right] \xrightarrow{d} \text{N}(0, \Sigma_{\theta_1,0}(z)).

\text{where } \Sigma_{\theta_1,0}(z) = \sigma^2_{\theta}(K)M_2(z)^{-1}/f_2(z), \text{ which is non-stochastic and is exactly the same as that in Cai et al. (2000). Further, from }\text{Theorem 2.1, we see that Var}(\hat{\beta}_1(z)) \text{ has a faster rate of convergence than that of Var}(\hat{\beta}_1(z)) \text{ due to the fact that }X_2 \text{ is nonstationary }\text{and }\Sigma(z) = \sigma^2(z) \text{ rather than }\Sigma_\theta(z). \text{ This is similar to the linear model case. However, the local fitting method renders the asymptotic variance of }\hat{\beta}_1(z) \text{ to be of the order }\text{O}(n h^{-1}) \text{ rather than the linear model case of }\text{O}(n^{-2}). \text{ One can easily derive the integrated asymptotic mean squared error (IAMSE) for }\hat{\beta}_1(z)

$$

$$\text{IAMSE}_1 = \text{IAMSE}(\hat{\beta}_1) = \frac{1}{2} \mu_2(K) \beta_2^{(2)}(z) + \frac{\text{tr}(\Sigma_{\theta_1}(z))}{nh} \omega(z)dz$$

$$\forall j = 1 \text{ and } 2, \text{ where } \omega(z) \text{ is a non-negative weight function, }\Sigma_{\theta_1}(z) \text{ is the upper-left corner sub-matrix of }\Sigma_\theta(z), \text{ and }\Sigma_{\beta_1}(z) \text{ is the lower-right corner sub-matrix of }\Sigma_\theta(z). \text{ By minimizing IAMSE}_1 \text{ with respect to }h, \text{ we obtain the optimal bandwidth for }\hat{\beta}_1(z), \text{ which is }\hat{h}_{j,\text{opt}} = \int \text{tr}(\Sigma_{\beta_1}(z))\omega(z)dz$$

$$\times \left[ \mu_2(K) \beta_2^{(2)}(z) \right]^{1/2} \omega(z)dz \xrightarrow{n \downarrow 0} 1/\sqrt{5}.$$ 

(211)

With the above choice of $\hat{h}_{j,\text{opt}}$, we see that IAMSE has an order of $O(n^{-4/5})$. Hence, IAMSE$_2$ for $\hat{\beta}_2(z)$ has an order $O(n^{-8/5})$ which is
smaller than 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 $n^{-2/5}$, then $\hat{\beta}_2(z) - \beta_2(z)$ reaches the optimal convergent rate but $\hat{\beta}_1(z)$ is under-smoothed. Therefore, the first step is to estimate $\hat{\beta}_2(z) = (\hat{\beta}_1(z)^T, \hat{\beta}_2(z)^T)^T$ 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 $h_2 = c_2 n^{-1/5}$ for some positive constant $c_2$. We know that the resulting estimator $\hat{\beta}_2(z)$ has an optimal convergence rate as $\hat{\beta}_2(z) - \beta_2(z) = O_p(n^{-4/5})$, and the asymptotic distribution of $\hat{\beta}_2(z)$ is given by
\[
\sqrt{n}h_2 \left[ \hat{\beta}_2(z) - \beta_2(z) = -\frac{1}{2} h_2^2 \mu_2(K) \beta_2''(z) \right] 
\]

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 $\hat{\beta}_1(z)$ is not optimal with the choice of $h_2 = c_2 n^{-1/5}$ ($c_2 > 0$). Therefore, we suggest that at the second step, one should re-estimate $\hat{\beta}_1(\cdot)$ with $\hat{\beta}_2(Z_t)$ replaced by $\hat{\beta}_2(Z_t)$ obtained at the first step. That is, we replace $\hat{\beta}_2(Z_t)$ by $\hat{\beta}_2(Z_t)$ in (2.1) to obtain
\[
Y_t \equiv Y_t - \hat{\beta}_2(z) X_{t2} = \beta_1(z)^T X_{t1} + \epsilon_t',
\]
where $\epsilon_t' = \epsilon_t + [\hat{\beta}_2(z_t) - \beta_2(z_t)]^T X_{t2}$. Then, we can apply the local linear method to estimate $\hat{\beta}_1(z)$. Let $(\hat{\beta}_{1,2step}(z), \hat{\beta}_{1,2step}^{(1)}(z))^T$ represent the resulting estimator of $(\beta_1(z)^T, \beta_1^{(1)}(z)^T)^T$, which is given by
\[
(\hat{\beta}_{1,2step}(z), \hat{\beta}_{1,2step}^{(1)}(z)) = \left[ \sum_{t=1}^{n} \left( X_{t1} (Z_t - z) \right)^{\otimes 2} K_h(z_t - z) \right]^{-1} \times \left[ \sum_{t=1}^{n} \frac{X_{t1}}{(Z_t - z) X_{t1}} Y_t K_h(z_t - z) \right],
\]
where $h_1$ is the bandwidth used at this step for estimating $\hat{\beta}_1(z)$. We will show in Theorem 2.2 below that the asymptotic distribution of $\hat{\beta}_{1,2step}(z)$ is the same as the case when $\hat{\beta}_2(Z_t)$ were known; that is $\hat{\beta}_{1,2step}(z) - \hat{\beta}_1(z) = O_p(n h_1^{-1/2}) = O_p(n^{-2/5})$ 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 $\hat{\beta}_1(\cdot)$, 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 $\hat{\beta}_2(z) - \beta_2(z) = O_p(n^{-4/5})$ if $h_2 = c_2 n^{-1/5}$, 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
\[
\sup_{z \in D} | \hat{\beta}_2(z) - \beta_2(z) | = O_p(a_n),
\]
where $D$ is a compact support of $f(z)$ (by assuming that $f(z)$ has a compact support $D$) and $a_n$ satisfies $a_n \rightarrow 0$ 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 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
\[
\sqrt{n} h_1 \left[ \hat{\beta}_{1,2step}(z) - \beta_1(z) - \frac{1}{2} h_1^2 \mu_2(K) \beta_2''(z) + o_p(h_1^2) \right] 
\]

\[\xrightarrow{d} N \left( 0, \Sigma_{\beta_1,0}(z) \right), \]

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 $\hat{\beta}_{1,2step}$ 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,
\[
Y_t = \beta(Z_t)^T X_t + \epsilon_t, \quad 1 \leq t \leq n,
\]
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
\[
(\hat{\beta}(z), \hat{\beta}^{(1)}(z)) = \left[ \sum_{t=1}^{n} \left( X_t (Z_t - z) \right)^{\otimes 2} K_h(z_t - z) \right]^{-1} \times \left[ \sum_{t=1}^{n} \frac{X_t}{(Z_t - z) X_t} Y_t K_h(z_t - z) \right],
\]
where 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 $l(1)$ process, $Z_t$ can be expressed as $Z_t = \sum_{i=1}^{t} u_i + Z_0$, where $u_t$ is a mixing process with mean zero and variance $\sigma_u^2$. We assume that $\lim_{n \to \infty} \text{Var}(n^{-1/2} \sum_{i=1}^{n} u_i)$ is a finite positive constant. Consider the simple case that $Z_0 = 0$ and $u_t$ is a white noise. Then, $Z_t \sim (0, \sigma_u^2)$ and $(Z_t - z)/\sqrt{t} \sim (-z t^{-1/2}, \sigma_u^2)$. 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 $(u_t)$ is a linear process: $u_t = \sum_{j=0}^{\infty} \psi_j \omega_j$, where $\omega_j$ is a white noise with mean zero and $\sigma_\omega^2 = \text{Var}(\omega_j) < \infty$, and $\{\psi_j\}$ satisfies the following conditions, for some $0 < \psi < 1$, \[
abla \sum_{j=0}^{\infty} |c_j|^r < \infty, \quad \text{and} \quad \sum_{j=0}^{\infty} |c_j| = 1. \quad (3.3)
\]

Then, $\sigma_u^2 = \text{Var}(u_t) = \sigma_\omega^2 \sum_{j=0}^{\infty} |c_j|$ and $\text{Cov}(u_t, u_{t+r}) = \sigma_\omega^2 \sum_{j=0}^{\infty} |c_j| \psi_j$ for any $s$ and $t$.

Let $\rho_{u,t}(t - s)$ and $\rho_{u,t}(t - s)$ be the vector and matrix of autocorrelation coefficients between $X_t$ and $u_t$, and between $V_t$ and $u_t$, respectively, where $V_t = X_t \psi_t$. Then we have $\sum_{j=0}^{\infty} |\rho_{u,t}(s)| < \infty$ and $\sum_{j=0}^{\infty} |\rho_{u,t}(s)| < \infty$ for all $0 < j \leq d$, $t \geq 0$, and $(i,j)$th component of $X_t$ and $V_t$, respectively. The autocorrelation coefficient between $V_{t+k}$ and $Z_t$ is $O(t^{-1/2})$, which goes to zero as $t \to \infty$, because the correlation coefficient between $V_{t+k}$ and $Z_t$ is $\rho(V_{t+k}, Z_t) = \frac{\text{Cov}(V_{t+k}, Z_t)}{\sqrt{\text{Var}(V_{t+k}) \text{Var}(Z_t)}} \leq \frac{\sum_{j=0}^{\infty} |\rho_{u,t}(s)|^2}{\sqrt{\text{Var}(V_{t+k}) \text{Var}(Z_t)}} = O(t^{-1/2})$, for any fixed $s$ and $t$.

where $\sigma_{u,t}^2 = \text{Var}(V_{t+k})$.

Next, define $z_{i,k} = t^{-1/2}(Z_{t+k} - Z_t)$ and set $f_{i,k}(\cdot)$ to represent the density of $z_{i,k}$. Also, use $f_{i,k}(\cdot)$ to represent the joint density function of $(z_{i,j}, z_{i,k})$. Further, set $b_{i,k,t}(\cdot, \cdot)$ to be the smallest $\sigma$-field generated by $\{V_t, X_t, Z_t\}_{t=0}^{\infty}$. We make the following assumptions.

**Assumptions:**

C.1. $E(\tilde{\xi}_t[X_t, Z_t, Z_{t-1}]) = 0$, $E(\tilde{\xi}_t[X_t, Z_{t-1}, \tilde{\xi}_{t-j}]) = \sigma_u^2 \bar{E}(\tilde{\xi}_t[X_t, Z_{t-1}]) < C a.s.$, and $(\{X_t, u_t\})$ is a stationary and mixing process satisfying constraints as imposed by (2.5) and (2.8), where $\sigma_u^2$, $C$ and $\sigma_u^2$ are finite positive constants. Also, $\lim_{n \to \infty} \text{Var}(n^{-1/2} \sum_{i=1}^{n} u_i)$ exists and is finite.

C.2. $E(\bar{W}_{t,j}^{2q}) < \infty$ for some $q > 3$ and for all $1 \leq i < j \leq d$.

C.3. Both $f_{i,k}(\cdot)$ and $f_{i,t,s}(\cdot)$ have bounded continuous derivative functions (for all $t$, $s$, and fixed $x$). Also, assume that $b_{i,t,s}(\cdot, \cdot)$ has bounded continuous derivative functions (for all $t$, $s$, and $z$).

C.4. $E(V_t|Z_t)$ has the following expression.

\[
E(V_t|Z_t) = E(V_t) + \delta_t \bar{g}(\cdot).
\]

where $E[\delta_t \bar{g}(\cdot)] = 0$, $E[|\delta_t \bar{g}(\cdot)|^2] = O(1)$, $\delta_t = O(t^{-1/2})$, $\bar{g}(\cdot)$ has the same dimension as $V_t$, $\bar{g}(\cdot)$ denotes the $(i,j)$th component of $\bar{g}(\cdot)$ and satisfies the assumption $\sup_{i,j} \bar{f}(\cdot) \leq C(u)$ with $u$ being a continuous function (for all $1 \leq i, j \leq d$). Further, for $t > s$, assume that $E(V_t|Z_s, Z_{s-1}) = E(V_t) + \delta_t \bar{g}(\cdot, Z_s, Z_{s-1})$, where $g(\cdot, \cdot, \cdot)$ is of dimension $d \times d$, $E[\delta_t \bar{g}(\cdot, Z_s, Z_{s-1})] = O(1)$, and $\delta_t = O(t^{-1/2})$. Also, $O(s^{-1/2})$ for $t > s$.

C.5. $K_t(\cdot)$ is a symmetric and continuous density function, supported by $]-1, 1[$.

C.6. $h^{p+q-2} \to \infty$, where $p = q/(q - 1) < 3/2$ and $q > 3$ is given in Appendix C1.
Remark 3.1. The asymptotic properties for \( \hat{\beta}^{(1)}(z) \) 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 \( h^2B_p(z) \) 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 \( \hat{\beta}(z) \) is of the order \( O\left((n^{-1/2})^{-1}\right) \), 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

\[
\text{IAMSE} = \int \left[ \text{tr}(\Sigma_t) n^{-1/2} h^{-1} + \frac{h^2}{4} \mu_y^2(K) \left[ \beta(\Sigma_t) \right]_2^2 \right] M(z)dz.
\]

Minimizing the IAMSE with respect to \( h \) gives the optimal bandwidth \( h_{opt} = cn^{-1/4} \) for some \( c > 0 \), which is much larger than that for the stationary case; see (2.11).

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

\[
P(Z_t \geq z) = P(Z_t > a \sqrt{n}) \rightarrow P(W_t(r) > a) = 1 - \Phi(a/\sqrt{\sigma_n}) > 0,
\]

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 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 \( n^{1/4} h^{1/2} \beta(\cdot)(a/\sqrt{n}) = O(1) \) for any \( a \), then, we have

\[
\sqrt{n^{1/2} h} \left[ \hat{\beta}(a \sqrt{n}) - \beta(a \sqrt{n}) - h^2 B_p(a \sqrt{n}) \right] \overset{d}{\rightarrow} M \left( \Sigma_1, a \right),
\]

where \( M \left( \Sigma_1, a \right) \) is a mixed normal distribution with mean zero and conditional covariance \( \Sigma_1 = \sigma^2 \gamma \sigma_i \nu_i \Sigma(K) \left( E(X X^T) \right) I(1, a/\sigma_i)^{-1} \).

Remark 3.2. Comparing Theorem 3.2 with Theorem 3.1, we observe that the asymptotic variance of \( \hat{\beta}(\cdot) \) 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 \( M \) be a matrix of finite dimension, and \( b_n \) be a sequence of non-stochastic real numbers, we write \( |M| = O(b_n) \) to mean that \( |M| = O(b_n) \) for each \( i \) and \( j \), where \( M_{ij} \) is the \( (i, j) \)th component of \( M \). Similarly, \( |M| = O(b_n) \) means that \( |M| = O(b_n) \) for all \( i \) and \( j \), and \( \sup_i |M| = O(b_n) \) means that \( \sup_i |M| = O(b_n) \) for all \( i \) and \( j \). Also, we write \( M^2 = O(b_n) \) to mean that \( M_{ij} = O(b_n) \) for all \( i \) and \( j \). Finally, we use \( M^{x \times x} \) to denote the real-valued \( x \times x \) matrices and \( D[0, 1] \) 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 \((U_{in}, \Xi_{in}) : 1 \leq t \leq n; n \geq 1\)

where \( U_{in} \) is a \( k \times m \) matrix and \( \Xi_{in} \) is an \( m \times 1 \) vector. We transform these arrays into random elements on \([0, 1]\) by \( U_{in} = U_{in}(\Xi_{in}) \) and \( \Xi_{in} = \Xi_{in} \) for \( 0 \leq s \leq 1 \). Also, define \( \epsilon_{in} = \Xi_{in} - \Xi_{in-1} \). We can then define the stochastic integral

\[
\int_0^s U_{in}(\Xi_{in})d\Xi_{in}(s) = \sum_{j=1}^{[in]} U_{ij}\epsilon_{in}.
\]

Let \( BM(\Omega) \) denote the vector Brownian motion with covariance matrix \( \Omega \). Finally, we define \( \Pi_1 = \sum_{i=1}^{[in]} \epsilon_i \), where \( \{\epsilon_i\} \) satisfies the following assumption:

Assumption M1: For some \( p > 2 \), \( \{\epsilon_i\} \) is a mean zero and strong mixing sequence with mixing coefficient satisfying

\[
sup_{t \geq 0} E[|\epsilon_i|^p] \leq C < \infty.
\]

In addition, \( E(W_{in}W^T_{in}) / n \rightarrow \Omega \) as \( n \rightarrow \infty \),

Let \( W_{in} = n^{-1/2}W_{in} \) and set \( \Xi_t = \sigma(U_{in}, \epsilon_t : t \leq i) \) to be the smallest \( \sigma \)-field containing the past history of \( U_{in}, \epsilon_t \) for all \( n = 1, 2, \ldots, i \).

Then, \( \{\epsilon_i, \Xi_t\} \) 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, \( \sup_{0 \leq t \leq 1} |A^*_t(\epsilon)| = O_p(1) \), and \( (U_{in}, W_{in}) \rightarrow (U, W) \) in \( D_{\text{conv}} \times \text{conv}[0, 1] \), then

\[
\int_0^s U_{in}dW_{in} \rightarrow \int_0^s U^{-}dW, \quad \text{equivalently,}
\]

\[
\int_0^s U_{in}dW_{in} \rightarrow \int_0^s U^{-}dW.
\]

where \( U^{-} \) is a cadlag process, \( W \) is \( BM(\Omega) \) and \( \Omega \) is defined in Assumption M1.
Proof. This is Theorem 3.1 of Hansen (1992). \( \square \)

**Lemma A.2**. Suppose \( U_n \Rightarrow U \) in \( D_{\text{mix}}(0, 1) \) and \( U(\cdot) \) is almost surely continuous. For a random sequence \( \{e_t\} \) and a sequence of nondecreasing \( \sigma \)-field \( \{\mathcal{F}_t\} \) to which \( \{e_t\} \) is adapted, assume that \( \sup_t \mathbb{E}[e_t^2] < \infty \), then

\[
\sup_{0 \leq s \leq 1} \mathbb{E} \left( \sum_{j=1}^{m} U_{nj} e_j \right)^p \to 0.
\]

**Proof.** See Theorem 3.3 of Hansen (1992). \( \square \)

**Lemma A.3.** Let \( w_i = \sqrt{n} K_1(Z_i - z_i) e_{i,j} z_{i-j}^2 \) and \( U_{nj} = X_i / \sqrt{n} \).

Set \( \mathcal{F}_t = \sigma(U_{nj}, w_i : i \leq t) \) to be the smallest \( \sigma \)-field containing the past history of \((U_{nj}, w_i)\) for all \( n \) and \( i \leq t \). For any \( 0 \leq r \leq 1 \), define \( \Lambda^*_r(t) = \sup_n \left| \frac{\sum_{j=1}^{n} \eta_j \xi_j - \sum_{j=1}^{n} \xi_j}{\sqrt{n}} \right| \) for all \( n \) and \( i \) satisfying \( \eta_j \xi_j \) and \( \xi_j \) is \( \sqrt{n} \)-mixing and \( \Lambda^*_r(t) \) converges to zero in probability uniformly. To this effect, we first show that \( \eta_j \xi_j - \mathbb{E}(\eta_j \xi_j) \) is an \( L_{\delta_0} \)-mixingale for \( \delta_0 \) given in Assumption A6 satisfying \( 1 < \delta_0 < \delta < 2 \). Note that for the definition of a mixingale sequence, we refer to the paper by McLeish (1975). Indeed, by Minkowski's inequality, for any \( m \geq 1 \),

\[
\| \mathbb{E}_{m(n)}(\eta_j \xi_j - \mathbb{E}(\eta_j \xi_j)) \|_{\delta_0} \leq \sum_{k=1}^{\infty} \mathbb{E}_{m(n)} \left| \sum_{j=1}^{n} \eta_j \xi_{j+k} - \mathbb{E}(\eta_j \xi_{j+k}) \right| \| \xi_j \|_{\delta_0}
\]

\[
= \sum_{k=1}^{\infty} (\cdots) + \sum_{k=m+1}^{\infty} (\cdots) \equiv I_1 + I_2.
\]

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

\[
\mathbb{E}_{m(n)}(\eta_j \xi_j - \mathbb{E}(\eta_j \xi_j)) \|_{\delta_0} \leq C \mathbb{E}^{1/\delta_0 - 1/2} \| \theta_j \|_{\delta_0} \| \xi_j \|_{\delta_0} \leq C \mathbb{E}^{1/\delta_0 - 1/2} \| \eta_j \|_{\delta_0} \| \xi_j \|_{\delta_0} \| \xi_j \|_{\delta_0} \| \xi_j \|_{\delta_0}
\]

\[
\leq C \left( \frac{\| \eta_j \|_{\delta_0} \| \xi_j \|_{\delta_0} \| \xi_j \|_{\delta_0}}{\delta_0} \right) \| \theta_j \|_{\delta_0} \| \xi_j \|_{\delta_0} \| \xi_j \|_{\delta_0} \| \xi_j \|_{\delta_0} \| \xi_j \|_{\delta_0}
\]

\[
\leq C \left( \frac{\| \eta_j \|_{\delta_0} \| \xi_j \|_{\delta_0} \| \xi_j \|_{\delta_0}}{\delta_0} \right) \| \theta_j \|_{\delta_0} \| \xi_j \|_{\delta_0} \| \xi_j \|_{\delta_0} \| \xi_j \|_{\delta_0} \| \xi_j \|_{\delta_0}
\]

so that

\[
I_1 = \sum_{k=1}^{\infty} \mathbb{E}_{m(n)}(\eta_j \xi_j - \mathbb{E}(\eta_j \xi_j)) \|_{\delta_0} \leq C m \mathbb{E}^{1/\delta_0 - 1/2} \| \theta_j \|_{\delta_0} \| \xi_j \|_{\delta_0} \| \xi_j \|_{\delta_0} \| \xi_j \|_{\delta_0}
\]

For \( I_2 \) in (A.5), one obtains

\[
\mathbb{E}_{m(n)}(\eta_j \xi_j - \mathbb{E}(\eta_j \xi_j)) \|_{\delta_0} \leq \mathbb{E}_{m(n)} \left( \eta_j \xi_j - \mathbb{E}(\eta_j \xi_j) \right)^2 + \mathbb{E}(\eta_j \xi_j)^2 \|_{\delta_0} \leq C \left( \frac{\| \eta_j \|_{\delta_0} \| \xi_j \|_{\delta_0} \| \xi_j \|_{\delta_0}}{\delta_0} \right) \| \theta_j \|_{\delta_0} \| \xi_j \|_{\delta_0} \| \xi_j \|_{\delta_0} \| \xi_j \|_{\delta_0} \| \xi_j \|_{\delta_0}
\]

\[
\leq C \left( \frac{\| \eta_j \|_{\delta_0} \| \xi_j \|_{\delta_0} \| \xi_j \|_{\delta_0}}{\delta_0} \right) \| \theta_j \|_{\delta_0} \| \xi_j \|_{\delta_0} \| \xi_j \|_{\delta_0} \| \xi_j \|_{\delta_0} \| \xi_j \|_{\delta_0}
\]

Then

\[
I_2 = \sum_{k=m+1}^{\infty} \mathbb{E}_{m(n)}(\eta_j \xi_j - \mathbb{E}(\eta_j \xi_j)) \|_{\delta_0} \leq C \sum_{k=m+1}^{\infty} \mathbb{E}^{1/\delta_0 - 1/2} \| \theta_j \|_{\delta_0} \| \xi_j \|_{\delta_0} \| \xi_j \|_{\delta_0} \| \xi_j \|_{\delta_0} \| \xi_j \|_{\delta_0}
\]

Hence, (A.5) becomes

\[
\mathbb{E}_{m(n)}(\eta_j \xi_j - \mathbb{E}(\eta_j \xi_j)) \|_{\delta_0} \leq C \mathbb{E}^{1/\delta_0 - 1/2} \| \theta_j \|_{\delta_0} \| \xi_j \|_{\delta_0} \| \xi_j \|_{\delta_0} \| \xi_j \|_{\delta_0} \| \xi_j \|_{\delta_0}
\]

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

\[
\frac{1}{n} \sup_{0 \leq r \leq 1} \left| \frac{1}{n} \sum_{i=1}^{[n]} \eta_i \xi_i - \mathbb{E}(\eta_i \xi_i) \right| \to 0.
\]

This, together with (A.3), concludes that

\[
\sup_{0 \leq r \leq 1} \left| \frac{1}{n} \sum_{i=1}^{[n]} \eta_i \xi_i \right| \to 0.
\]

Therefore, by (A.4),

\[
\sup_{0 \leq r \leq 1} \left| \frac{1}{n} \sum_{i=1}^{[n]} \eta_i \xi_i \right| \to 0,
\]

\[
\sup_{0 \leq r \leq 1} \left| \frac{1}{n} \sum_{i=1}^{[n]} \eta_i \xi_i \right| \to 0.
\]

This completes the proof of Lemma A.3. \( \square \)

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

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

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

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

with $j = 0, 1, 2$.

$S_{n,j}(z) = \frac{1}{n} \sum_{t=1}^{n} K_h(Z_t - z) Z_{t,z,h}^{j} (D_n^{-1} X_t)^{\otimes 2}$. Now, to facilitate the analysis of $S_{n,j}(z)$, we first express $S_n(z)$ as

$$
S_n(z) = \begin{pmatrix} F_{n,j,0}(z) \\ F_{n,j,1}(z) \end{pmatrix} \begin{pmatrix} F_{n,j,0}(z) \\ F_{n,j,2}(z) \end{pmatrix},
$$

where

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

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

and

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

For $l = 1, 2$, define

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

By noting that $X_t$ and $Z_t$ are stationary and using the standard change-of-variable and a Taylor’s expansion argument, we have

$$
E \left[ F_{n,l}(z) \right] = E \left[ Z_{t,z,h} X_t^{\otimes l} K_h(Z_t - z) \right] = f_t(z) M_z \mu_{t}(l) + o(1).
$$

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

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

Therefore,

$$
F_{n,l}(z) = f_t(z) M_z \mu_{t}(l) + o_p(1), \quad (A.8)
$$

so that

$$
F_{n,j,0}(z) = F_{n,j,2}(z) = f_t(z) M_z \mu_{t}(2) + o_p(1). \quad (A.9)
$$

Let $\sigma^* = \sigma(X_t, Z_t \mid t \leq l)$ be the smallest $\sigma$-field containing the past history of $(X_t, Z_t)$, and denote by $\varepsilon_t = K_h(Z_t - z) Z_{t,z,h} X_t - E[K_h(Z_t - z) Z_{t,z,h} X_t]$. Then, similar to (A.7), it is easy to verify that

$$
\sup_{t \geq 0} \text{Var} \left( \sum_{i=t+1}^{t+m} \varepsilon_t \right) = O(m/h) \quad (A.10)
$$

for any $m \geq 1$. Define $U_{n,t} = X_{2t} / \sqrt{n}$ for any $1 \leq t \leq n$ and $U_{n,r} = U_{n,|n|}$ for any $r \in [0, 1]$. For any small $0 < \delta < 1$, set $N = \lceil 1/\delta \rceil$, $t_k = \lceil k n/N \rceil + 1$, $t_k^* = t_k - 1$, and $t_k^* = \min(t_k^*, n)$. Then,

$$
\left\lVert \frac{1}{n} \sum_{t=1}^{n} U_{n,t} \varepsilon_t \right\rVert \leq \left\lVert \frac{1}{n} \sum_{t=k_0}^{n} U_{n,t} \varepsilon_t \right\rVert + \left\lVert \frac{1}{n} \sum_{t=k_0}^{n} U_{n,t} \varepsilon_t \right\rVert \leq \sup_{0 \leq k \leq 1} |U_{n,k}| \frac{1}{n} \sum_{t=k}^{n-1} \sum_{t=k_0}^{n} \varepsilon_t.
$$

Since $U_{n,t}$ converges weakly to a Brownian motion, it is clear that

$$
\sup_{0 \leq k \leq 1} |U_{n,k}| \rightarrow 0.
$$

Therefore, as $n \rightarrow \infty$,

$$
\frac{1}{n} \sum_{t=1}^{n} U_{n,t} \varepsilon_t = o_p(1) + \sup_{0 \leq k \leq 1} \left| U_{n,k} - U_{n,\infty} \right| o_p(1).
$$

Note that as $n \rightarrow \infty$,

$$
\sup_{0 \leq k \leq 1} \left| U_{n,k} - U_{n,\infty} \right| \rightarrow 0.
$$

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

$$
\frac{1}{n} \sum_{t=1}^{n} U_{n,t} \varepsilon_t = o_p(1),
$$

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

$$
F_{n,j,1}(z) = E \left[ K_h(Z_t - z) Z_{t,z,h} X_t \right] = f_t(z) M_z \mu_{t}(1) + o_p(1), \quad (A.11)
$$

Similarly,

$$
F_{n,j,2}(z) = f_t(z) \mu_{t}(2) W_{\mu_{t}(2)} + o_p(1). \quad (A.12)
$$

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

$$
S_{n,j}(z) = f_t(z) \mu_{t}(K) S(z) + o_p(1). \quad (A.13)
$$

By noting that $\mu_{t}(K) = 1$ and $\mu_{t}(K) = 0$, we immediately obtain from (A.13) that

$$
S_n(z) = f_t(z) \left[ \begin{pmatrix} 1 \\ 0 \end{pmatrix} \mu_{t}(K) \right] \otimes S(z) + o_p(1). \quad (A.14)
$$
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

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

where

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

with

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

$$\times \{ \beta(Z_t) - \beta(z) - (Z_t - z) \beta^{(1)}(z) \},$$

and

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

Define,

$$G_{n,0}(z) = n^{-1} \sum_{t=1}^{n} K_h(Z_t - z) X_t^T \{ \beta_1(Z_t) - \beta_1(z) - (Z_t - z) \beta^{(1)}_1(z) \},$$

$$G_{n,1}(z) = n^{-1} \sum_{t=1}^{n} K_h(Z_t - z) X_t \{ \beta_2(Z_t) - \beta_2(z) - (Z_t - z) \beta^{(1)}_2(z) \},$$

$$G_{n,2}(z) = n^{-1} \sum_{t=1}^{n} K_h(Z_t - z) \{ \beta_1(Z_t) - \beta_1(z) - (Z_t - z) \beta^{(1)}_1(z) \},$$

and

$$G_{n,3}(z) = n^{-1} \sum_{t=1}^{n} K_h(Z_t - z) \{ \beta_2(Z_t) - \beta_2(z) - (Z_t - z) \beta^{(1)}_2(z) \},$$

so that

$$B_n(z) = \left( G_{n,0}(z) + G_{n,1}(z) + G_{n,3}(z) \right). \tag{A.18}$$

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

$$E[G_{n,0}(z)] = h^2 f_z(z) M_2(z) \left[ \frac{\mu_2(K)}{2} \beta^{(2)}_1(z) \right] \{ 1 + o_p(1) \}$$

and

$$\text{Var}[G_{n,0}(z)] = o(1),$$

so that

$$G_{n,0}(z) = h^2 f_z(z) M_2(z) \left[ \frac{\mu_2(K)}{2} \beta^{(2)}_1(z) \right] \{ 1 + o_p(1) \}. \tag{A.19}$$

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

$$G_{n,1}(z) = h^2 f_z(z) M_1(z) W_{\eta,2}^{(1)} \left[ \frac{\mu_2(K)}{2} \beta^{(2)}_1(z) \right] \{ 1 + o_p(1) \},$$

$$G_{n,2}(z) = h^2 f_z(z) M_1(z) W_{\eta,2}^{(1)} \sqrt{n} \left[ \frac{\mu_2(K)}{2} \beta^{(2)}_1(z) \right] \{ 1 + o_p(1) \},$$

and

$$G_{n,3}(z) = h^2 f_z(z) W_{\eta,2}^{(2)} \sqrt{n} \left[ \frac{\mu_2(K)}{2} \beta^{(2)}_1(z) \right] \{ 1 + o_p(1) \}.$$ 

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

$$B_n(z) = h^2 f_z(z) S(z) D_n \left[ \frac{\mu_2(K)}{2} \beta^{(2)}_1(z) \right] \{ 1 + o_p(1) \}. \tag{A.20}$$

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

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

Therefore,

$$D_n^{-1} I_3 = h^2 B_n(z) + o_p(h^2). \tag{A.21}$$

Finally, we consider $I_4$. Define

$$T_n(z) = \sqrt{n} \sum_{t=1}^{n} K_h(Z_t - z) \epsilon_t D_n^{-1} X_t = \left( T_{n,1}(z) \right)^T \left( T_{n,2}(z) \right)$$

with

$$T_{n,1}(z) = \sqrt{n} \sum_{t=1}^{n} K_h(Z_t - z) \epsilon_t X_t$$

and

$$T_{n,2}(z) = \sqrt{n} \sum_{t=1}^{n} K_h(Z_t - z) \epsilon_t X_t / \sqrt{n}.$$ 

By combining the above expressions with (A.16) and (A.20), we obtain

$$\sqrt{n} h D_n \left[ \hat{\beta}(z) - \beta(z) - h^2 B_n(z) + o_p(h^2) \right] = R_n(z)^{-1} T_n(z). \tag{A.22}$$

To prove the asymptotic normality of the left hand side of (A.22), 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

$$T_{n,1}(z) \rightarrow^{d} N(0, \sigma^2_v v_0(K) f_z(z) M_2(z)) = \sqrt{v_0(K) f_z(z)} W_{\epsilon,1}(1), \tag{A.23}$$

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

$$\sqrt{n} \sum_{t=1}^{n} K_h(Z_t - z) \epsilon_t \rightarrow^{d} N(0, \sigma^2_v v_0(K) f_z(z)) \rightarrow^{d} \sqrt{v_0(K) f_z(z)} W_{\epsilon,1}(1). \tag{A.24}$$

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

$$T_n(z) \rightarrow^{d} \sqrt{v_0(K) f_z(z)} \int_0^1 W_{\epsilon,2}(r) dW_{\epsilon,1}(r). \tag{A.25}$$

Since $W_{\epsilon,2}(\cdot)$ and $W_{\epsilon}(\cdot)$ are uncorrelated (because $Z_t$ and $\epsilon_t$ are uncorrelated), $\int_0^1 W_{\epsilon,2}(r) dW_{\epsilon}(r)$ has a mixed normal distribution, so that the conditional covariance of $\left( \int_0^1 W_{\epsilon,2}(r) dW_{\epsilon,1}(r) \right)$ is

$$\sigma^2_v \begin{pmatrix} M_2(z) W_{\eta,2}^{(1)} \end{pmatrix} \left( \int_0^1 M_1(z) W_{\eta,2}^{(1)} dW_{\eta,2} \right) = \sigma^2_v S(z).\tag{A.25}$$

Therefore, by Slussky's theorem, we have
\[ \sqrt{n} h_1 D_n \left[ \beta(z) - \beta(z) - B_\beta(z) + o_p(h_1^2) \right] \]
\[ \xrightarrow{d} f_z^{-1/2}(z) v_0^{-1/2}(K) S(z)^{-1} \int_0^1 W_{n,2}(r) dW_n(r). \]  
(A.26)

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

**Proof of Theorem 2.2.** To simplify the notation, in what follows, we drop the subscript "2" step in \( \hat{\beta}_{1,2,2}(z) \) and \( \hat{\beta}_{1,2,2}(z) \). First, define
\[ L_{n,j}(z) = \frac{1}{n} \sum_{t=1}^n Z_{t,h} X_{t1} X_{t1}^T K_{h_1}(Z_t - z). \]

Then, similar to (A.9), one can show that \( L_{n,j}(z) = f_c(z) M_2(z) \mu_j(z) + o_p(1). \)  

Therefore, 
\[ n^{-1} \left( \sum_{t=1}^n X_{t1} (Z_t - z) X_{t1} \right)^{\otimes 2} K_{h_1}(Z_t - z) = \left( \frac{1}{h_1} \right)^2 M_2(z) [1 + o_p(1)]. \]

By (2.14),
\[ \left( \hat{\beta}_1(z) - \beta_1(z) \right)(z) \otimes \left( \beta_1(z) - \beta_1(z) \right)(z) = \left( \sum_{t=1}^n \left( X_{t1} (Z_t - z) X_{t1} \right)^{\otimes 2} K_{h_1}(Z_t - z) \right)^{-1} \]
\[ \times \left[ \sum_{t=1}^n \left( Z_t - z \right) X_{t1} \right] \left[ \sum_{t=1}^n \left( Z_t - z \right) X_{t1} \right]^T \]
so that
\[ \beta_1(z) - \beta_1(z) = M_2(z) n^{-1} \sum_{t=1}^n X_{t1} \left[ \beta_1(z) - \beta_1(z) \right] \]
\[ - X_{t1}^T \beta_1(z)(Z_t - z) K_{h_1}(Z_t - z) \{1 + o_p(1)\} \]
\[ = M_2(z) n^{-1} \sum_{t=1}^n X_{t1} \left[ \beta_1(z) - \beta_1(z) \right] \]
\[ - \beta_1(z)(Z_t - z) K_{h_1}(Z_t - z) \{1 + o_p(1)\} \]
\[ + M_2(z) n^{-1} \sum_{t=1}^n X_{t1} \{1 + o_p(1)\} \]
\[ + M_2(z) n^{-1} \sum_{t=1}^n X_{t1} \left[ \beta_2(z) - \beta_2(z) \right] \]
\[ \times K_{h_1}(Z_t - z) \{1 + o_p(1)\} \]
\[ = J_1 + J_2 + J_3. \]

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
\[ J_1 = \frac{h_1^2}{2} \left\{ \mu_2(K) \right\}^2 [1 + o_p(1)] \]
and
\[ \sqrt{n} h_1 J_2 \xrightarrow{d} N(0, \Sigma_{\beta_1,o}(z)). \]

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
\[ |J_3| = o_p(h_1^2 \sqrt{n}) = o_p(h_1^2). \]

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

This proves the theorem. \( \square \)

**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
\[ S_n(z) = n^{-1/2} \sum_{t=1}^n K_{h_1}(Z_t - z) \left\{ \frac{1}{Z_{t,h,z}} \right\} \otimes X_t X_t^T \]
\[ = \left( \frac{S_n(0,z)}{S_n(z)} \right) \otimes \left( \frac{S_n(z)}{S_n(0,z)} \right) \]

with \( Z_{t,h} = (Z_t - z)/h \) and \( S_n(z) = n^{-1/2} \sum_{t=1}^n K_{h_1}(Z_t - z) X_t X_t^T \)

for \( j = 0, 1, 2 \), where \( K_{h}(v) = K(v)/h \).

Also, set
\[ \tilde{\mu}_j(z) = \frac{1}{\sqrt{n}} \sum_{t=1}^n K_{j}(Z_t - z) = \frac{\beta_n}{n} \sum_{t=1}^n K_{j}(\gamma_n^{-1} Z_t + x_0), \]

where \( \beta_n = \sqrt{n}/h, \gamma_0 = \sqrt{n}, \) and \( x_0 = -Z/\sqrt{n} \). Clearly, \( x_0 \rightarrow 0 \) for any fixed \( z \) and \( x_0 = -a \) if \( z = a \sqrt{n} \). Further, let \( \phi_j(x) = (\sqrt{2 \pi n})^{-1/2} \exp(-x^2/(2 n^2)) \) for \( x > 0 \). Finally, let \( o_{f}(1) \) denote the convergence in \( L_2 \) which implies \( o(1) \). We use the notation \( A_{1n} = A_{2n} + (s.o) \) to denote that \( A_{2n} \) has the same order as \( A_{1n} \) and \( (s.o) \) denotes the terms having orders smaller than \( A_{2n} \).

In what follows, we assume that \( Z \) satisfies (3.3). We present some preliminary results.

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

(i) \( \tilde{\mu}_j(z) \xrightarrow{p} \mu_1(K (L_1(0) / \sigma_a), \sigma_a), \) if \( z \) is fixed,

(ii) \( \tilde{\mu}_j(z) = O(1), \) and

(iii) \( E \left[ \tilde{\mu}_j(z)^p \right] = O(1 - 1/2 h^{-p}). \)

**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 \( \varepsilon > 0, \)
\[ \tilde{\mu}_j(z) = \frac{\mu_1(K_n)}{n} \sum_{i=1}^n \phi_j(y_i^{-1} Z_t + x_0) + o_{L_2}(1). \]

Since \( \phi_j(z) \) satisfies the Lipschitz condition and \( x_0 \rightarrow 0, \)
\[ \tilde{\mu}_j(z) = \frac{\mu_1(K_n)}{n} \sum_{i=1}^n \phi_j(y_i^{-1} Z_t) + o_{L_2}(1) \]
\[ = \frac{\mu_1(K_n)}{n} \sum_{i=1}^n \phi_j(W_i(t/n)) + o_{L_2}(1) \]
\[ = \tilde{\mu}_j(z) = \mu_1(K) \int_0^1 \phi_j(W_i(s)) ds + o_{L_2}(1). \]

in view of (3.7) and (2.7). By Lemma 9 of Jeganathan (2004), we have
\[ \tilde{\mu}_j(z) = \mu_j(K) L(1, 0)/\sigma_u + o_h(1) \]
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
\[
E \left[ \tilde{\mu}_j(z) \right] = n^{-1/2} \sum_{t=1}^n E \left[ K_{j,h}(Z_t - z) \right]
= n^{-1/2} h^{-1} \sum_{t=1}^n \int K_j(t^{1/2} u/h) f_{i,z}(u) du
= n^{-1/2} \sum_{t=1}^n h^{-1/2} \int K_j(v) f_{i,z}(h t^{-1/2} v) dv
\leq C n^{-1/2} \sum_{t=1}^n t^{-1/2} = O(1).
\]
Finally, recall that \( K_{j,h}(u) = h^{-1} K(u/h) \) and \( K_j = u^j K(u) \), it can be shown easily by the boundedness of \( f_{i,z}(\cdot) \) that
\[
E \left[ |K_{j,h}(Z_t - z)|^p \right] = h^{-p} \int |K_j(t^{1/2} u/h)|^p f_{i,z}(u) du
= t^{-1/2} h^{-p} \int |K_j(v)|^p f_{i,z}(t^{-1/2} h v) dv \leq C t^{-1/2} h^{-p}.
\]
This proves the lemma. \( \square \)

**Lemma B.2.** Under Assumptions given in Theorem 3.1, if \( z \) is fixed, we have
\[
S_{n,j}(z) = E(X(X'_T) [\tilde{\mu}_j(z)] + o_p(1) \rightarrow E(X(X'_T) \mu_j(K) L(1, 0)/\sigma_u).
\]
**Proof.** Recall that \( V_t = X(X'_T). \) 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 \( B_{1n,1} = E(V_t) n^{-1/2} \sum_{t=1}^n K_{j,h}(Z_t - z) = E(V_t|\tilde{\mu}_j(z)), \)
\[
B_{1n,2} = n^{-1/2} \sum_{t=1}^n [E(V_t|Z_t) - E(V_t)] K_{j,h}(Z_t - z)
= n^{-1/2} \sum_{t=1}^n \delta_t g_t(z) K_{j,h}(Z_t - z),
\]
and
\[
B_{1n,3} = n^{-1/2} \sum_{t=1}^n [V_t - E(V_t|Z_t)] K_{j,h}(Z_t - z)
\equiv n^{-1/2} \sum_{t=1}^n \zeta_t K_{j,h}(Z_t - z),
\]
where \( \zeta_t = V_t - E(V_t|Z_t), \) It follows from Lemma B.1 that \( B_{1n,1} \rightarrow \mu_j(K) E(V_t|L(1, 0)/\sigma_u). \) 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 \( g_t(z) \) as \( \delta_t = O(t^{-1/2}) \) (see Assumption C3), we have
\[
|B_{1n,2}| \leq n^{-1/2} \sum_{t=1}^n |\delta_t| |g_t(Z_t)| |K_{j,h}(Z_t - z)|
\leq C n^{-1/2} \sum_{t=1}^n t^{-1/2} C(Z_t) |K_{j,h}(Z_t - z)|
= C n^{-1/2} \sum_{t=1}^n t^{-1/2} C(Z_t - z + z) |K_{j,h}(Z_t - z)|.
\]
Since \( C(u) \) is continuous at \( z \) and \( K_j(\cdot) \) has a finite support, then \( C(u) \leq C_z \) for some \( C_z \) for all \( u's \) in a neighborhood of \( z. \)
Therefore,
\[
|B_{1n,2}| \leq C n^{-1/2} \sum_{t=1}^n t^{-1/2} |K_{j,h}(Z_t - z)| \rightarrow 0
\]
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 \( (z^2 \leq \text{means} x^2 \text{for any } i \text{ and } j, \) see the discussions below Lemma A.2 for our notation adoption)
\[
E[B_{1n,3}^2] = n^{-1} \sum_{t=1}^n E \left[ |\zeta_t|^2 |K_{j,h}(Z_t - z)| \right]
+ 2 n^{-1} \sum_{1 \leq t < s \leq n} E[|\zeta_t| \zeta_s K_{j,h}(Z_t - z) K_{j,h}(Z_s - z)]
\equiv B_{1n,31} + B_{1n,32}.
\]
Clearly, by Cauchy–Schwartz inequality, Lemma B.1, and Assumption C1, we have
\[
|B_{1n,31}| \leq C n^{-1} \sum_{1 \leq t \leq n} |E[|\zeta_t| |K_{j,h}(Z_t - z)|]| 
\leq C n^{-1} \sum_{1 \leq t \leq n} t^{-1/2} h^{-1/2} = o(1),
\]
since \( n h^{-p-2} \rightarrow \infty \) (from Assumption C5). For \( B_{1n,32}, \) by (3.5) and (3.6), we have
\[
E[|\zeta_t| |K_{j,h}(Z_t - z)|] = E[|V_t| Z_t, Z_s] - E[|V_t| Z_s] = \delta_t g_t(Z_t, Z_s) - \delta_t g_t(Z_s).
\]
It is easy to see that
\[
|B_{1n,32}| \leq C n^{-1} \sum_{1 \leq t \leq n} |E[|\zeta_t| |K_{j,h}(Z_t - z)|]| 
\leq C n^{-1} \sum_{1 \leq t \leq n} |E[|g_t(z)| |K_{j,h}(Z_t - z)|]| 
\leq C n^{-1} \sum_{1 \leq t \leq n} |E[|g_t(z)| |K_{j,h}(Z_t - z)|]| 
\equiv B_{1n,32,1} + B_{1n,32,2}.
\]
To evaluate \( B_{1n,32,1}, \) first, we consider the following quantity
\[
E[|g_t(z)| |K_{j,h}(Z_t - z)| K_{j,h}(Z_s - z)|V_s] = 
= \int b_{1,x,z}(t^{-1/2} h u, s^{-1/2} h v) g_{t,z}(z + hu, z + hv, V_z) \times \{V_s - E(V_s) - \delta_t g_t(z + hv)|K_j(u)K_j(v)| t^{-1/2} s^{-1/2} du dv
= \mu_j^2(K) t^{-1/2} s^{-1/2} b_{1,x,z}(0, 0) g_{t,z}(z, z, V_z) \times \{V_s - E(V_s) - \delta_t g_t(z)|1 + o(1)\}.
\]
Then, using (3.5) and (B.1), we have
\[
B_{1n,32,1} \leq C n^{-1} \sum_{1 \leq t \leq n} |E[|g_t(z)| |K_{j,h}(Z_t - z)|]| 
\leq C n^{-1} \sum_{1 \leq t \leq n} s^{-1} \sum_{t+1 \leq s \leq n} t^{-1/2} b_{1,x,z}(0, 0) \times E[|g_t(z, z, V_s) - E(V_s) - \delta_t g_t(z)|1 + o(1)\]
by Assumptions C2 and C3. Similarly, using the fact that $\delta_t = O(t^{-1/2})$, one can easily show that $B_{2n,32} = o(1)$. Thus, we have shown that $B_{2n,32} = o(1)$. By summarizing the above results, the lemma is proved. □

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

$$B_{2n} = h^2 B_0(z)E(X_i^2) \mu_2(z) + o_p(h^2)$$

$$= h^2 B_0(z)E(X_i^2) L(1,0)/\sigma_n + o_p(h^2).$$

**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 $B_{2n,1} = E(V_t) n^{-1/2} \sum_{i=1}^{n} \left[ \beta(Z_t) - \beta(z) - \beta^{(1)}(z)(Z_t - z) \right] K_0(Z_t - z)$, $B_{2n,2} = n^{-1/2} \sum_{i=1}^{n} \left[ E(V_t | Z_t) - E(V_t) \right] \left[ \beta(Z_t) - \beta(z) - \beta^{(1)}(z)(Z_t - z) \right] K_0(Z_t - z)$, and

$$B_{2n,3} = n^{-1/2} \sum_{i=1}^{n} \delta_i g_i(Z_t) \left[ \beta(Z_t) - \beta(z) - \beta^{(1)}(z)(Z_t - z) \right] K_0(Z_t - z),$$

and

$$B_{2n,3} = n^{-1/2} \sum_{i=1}^{n} \left[ V_t - E(V_t | Z_t) \right] \left[ \beta(Z_t) - \beta(z) - \beta^{(1)}(z)(Z_t - z) \right] K_0(Z_t - z)$$

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

$$B_{2n,1} = E(V_t) n^{-1/2} \sum_{i=1}^{n} \left[ \beta(Z_t) - \beta(z) - \beta^{(1)}(z)(Z_t - z) \right] K_0(Z_t - z)$$

$$= E(V_t) \frac{h^2}{2} \beta^{(2)}(z) \mu_2(z) + (s.o.)$$

$$= \frac{h^2}{2} E(V_t) L(1,0) \beta^{(2)}(z) \mu_2(K)/\sigma_n + o_p(h^2).$$

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

$$E \left[ B_{2n,2}^2 \right] = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} \left[ \beta(Z_t) - \beta(z) - \beta^{(1)}(z)(Z_t - z) \right] K_0(Z_t - z)$$

$$\times \left[ \beta(Z_t) - \beta(z) - \beta^{(1)}(z)(Z_t - z) \right] K_0(Z_t - z)$$

$$= \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} \delta_i \delta_j t^{-1/2} s^{-1/2}$$

$$\times \int \int f_{i,s,	au}(t^{-1/2}hu,s^{-1/2}hv)g_i(z + hu)g_j(z + hv)$$

$$\times \left[ \beta(z + hu) - \beta(z) - \beta^{(1)}(z)(hu) \right] K(u)$$

$$\times \left[ \beta(z + hv) - \beta(z) - \beta^{(1)}(z)(hv) \right] K(v)$$

$$\leq C \frac{h^2}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} \delta_i \delta_j t^{-1/2} s^{-1/2} f_{i,s,	au}(0,0)g_i(z)g_j(z)$$

$$= O(n^{-1/2}h^4) = o(h^4)$$

because $\delta_t = O(t^{-1/2})$ and $\delta_u = O(u^{-1/2})$, 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 $B_{2n,3}$, we decompose $E \left[ B_{2n,3}^2 \right]$ into two terms as follows:

$$E \left[ B_{2n,3}^2 \right] = \frac{1}{n} \sum_{i=1}^{n} E \left[ \epsilon_i^2 \right] \beta(\epsilon_i) - \beta(z)$$

$$\left( - \beta^{(1)}(z)(Z_t - z) \right)^2 K_0(Z_t - z)$$

$$+ 2 \sum_{i=1}^{n} E \left[ \epsilon_i \epsilon_j \beta(\epsilon_i) - \beta(z) - \beta^{(1)}(z)(Z_t - z) \right] K_0(Z_t - z)$$

$$\times \left( \beta(Z_t) - \beta(z) - \beta^{(1)}(z)(Z_t - z) \right) K_0(Z_t - z)$$

$$= B_{2n,31} + B_{2n,32}.$$

Similar to the evaluation of $B_{2n,31}$, by Cauchy–Schwarz inequality and assumption C1, one can show easily that

$$B_{2n,31} \leq n^{-1} \sum_{i=1}^{n} \| \epsilon_i \|_2^2 \| \beta(\epsilon_i) - \beta(z) - \beta^{(1)}(z)(Z_t - z) \|_2^2$$

$$\leq C n^{-1} \sum_{i=1}^{n} t^{-1/2} h^{1/2} \leq C (nh^{b/2} - 1) / nh$$

by Assumption C5. For $B_{2n,32}$, by analogy to $B_{2n,31}$, we have

$$B_{2n,32} \leq C (nh^{b/2} - 1) h^4 \sum_{i=1}^{n} \sum_{j=1}^{n} \| \delta_i \|_2 \delta_j \leq o(h^4)$$

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

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

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

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

$$V^* = \sigma_0^2 V_0(K)E(X_i^2) L(1,0)/\sigma_n.$$

**Proof.** Clearly, $E[B_{3n}] = 0$ because $E(\epsilon_i | X_t, Z_t) = 0$. Also, by the assumptions that $\| \epsilon_i \|$ is a martingale difference and $E(\epsilon_i^2 | X_t, Z_t) = \sigma_0^2$ (conditional homogenous errors), we conclude that the conditional variance of $n^{1/4} h^{1/2} B_{3n}$, given $(X_t, Z_t)$, is

$$V_{3n} = \sigma_0^2 h \sum_{i=1}^{n} X_i^2 K_0(z_t - z_t),$$

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

$$V_{3n} = \sigma_0^2 V_0(K) L(1,0) E(X_i^2) / \sigma_n + o_p(1).$$

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

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

This proves the lemma. □

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

$$S_n(z) = \left( \begin{array}{c} \text{S}_n,0(z) \\ \text{S}_n,1(z) \end{array} \right) \left( \begin{array}{c} \text{S}_n,1(z) \\ \text{S}_n,2(z) \end{array} \right)$$

$$= \left( \begin{array}{cc} 1 & 0 \\ h^2 \mu_2(K) \end{array} \right) \otimes E(X_i^2) L(1,0)/\sigma_n \left( 1 + o_1(1) \right).$$
which, by replacing \( Y_t \) in (3.2) by \( Y_t = X_t^\beta(Z_t) + \epsilon_t \), implies that
\[
\begin{align*}
\hat{\beta}(z) - \beta(z) &= \left[ E(X_t^T_1 L(1, 0)/\sigma_n) \right]^{-1} \\
\times &n^{-1/2} \sum_{t=1}^n X_t^T_1 \left[ \beta(Z_t) - \beta(z) - \beta^{(1)}(z)(Z_t - z) \right] \\
\times K_h(Z_t - z) + n^{-1/2} \sum_{t=1}^n X_t \epsilon_t K_h(Z_t - z) \left\{ 1 + o_p(1) \right\}
\end{align*}
\]
\[
\equiv \left[ E(X_t^T_1 L(1, 0)/\sigma_n) \right]^{-1} \{B_{2n} + B_{3n}\} \{1 + o_p(1)\}, \tag{B.2}
\]
where \( B_{2n} = n^{-1/2} \sum_{t=1}^n X_t^T_1 \left[ \beta(Z_t) - \beta(z) - \beta^{(1)}(z)(Z_t - z) \right] K_h(Z_t - z) \) 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
\[
\begin{align*}
n^{-1/4} h^{1/2} \left[ \hat{\beta}(z) - \beta(z) - h^2 \beta(z) + o_p(h^2) \right] \\
&= \sigma_n \left[ E(X_t^T_1) \right]^{-1} n^{-1/4} h^{1/2} B_{2n} \{1 + o_p(1)\} \overset{d}{\rightarrow} MN(\Sigma_1).
\end{align*}
\]
This completes the proof of Theorem 3.1. □

References


