Functional Coefficient Models for Nearly (Possibly Weakly) $I(1)$ Processes

Mirza Trokić

$^a$Department of Economics, Bilkent University, Ankara 06800, Turkey

Abstract

The focus of this article is on nonlinear time varying coefficient models when the covariates and coefficient components are weakly, nearly, or possibly purely integrated time series processes. Local linear fitting is used to derive coefficient estimators along with their asymptotic distributions. The rates of convergence for the estimators is shown to differ based on whether stationary, weakly, nearly, or purely integrated covariates are being modelled. Similar conclusions also hold for the derived optimal bandwidth parameters.

Keywords: Nonstationary, Nonlinear, Semiparametric estimation, Asymptotic theory, Local time, Spatial time series, Weak unit root, Near unit root, Time series, Varying coefficient model

1. Introduction

Functional time varying coefficient models are becoming increasingly prominent amongst both theoretical and empirical econometricians. Particularly appealing features of these models are their capacity to attenuate the curse of dimensionality and their flexibility in accommodating nonlinear phenomena in economic and financial time series data. Indeed, much has been written on these models with stationary and deterministic trend components; see for example Robinson (1989), Chen and Tsay (1993), Park and Hahn (1999), Cai et al. (2000), and Cai (2007) among others. In contrast, theoretical consideration of time varying coefficient models with nonstationary covariates and varying coefficient components is still an open area of research. In this regard, the recent progression of contributions includes Xiao (2009) who studies the model with possibly integrated regressors and stationary varying coefficient components and Cai et al. (2009) who consider the case when the varying coefficient components are stationary and the regressors are possibly integrated. More recently still, the contributions of Gao and Phillips (2013) and Sun et al. (2013) extend these varying coefficient models to also accommodate possibly integrated series in both regressors and varying coefficient components. On the other hand, apart from a working paper by Cai and Wang (2008), very little has been done in the way of varying coefficient models with nearly integrated variables. Accordingly, in order to bridge this gap in the literature, the focus of this paper is on the popular varying coefficient regression model

$$ Y_t = \beta(Z_t)^\top X_t + \epsilon_t, \quad 1 \leq t \leq n $$

where $Y_t$ and $\epsilon_t$ are scalars, $Z_t$ is possibly nearly or purely $I(1)$, and $X_t = (X_{t1}, \ldots, X_{td})^\top$ is a $d$-dimensional vector of nearly, weakly in the sense of Park (2003), and purely $I(1)$ covariates. Conformingly, $\beta(\cdot)$ is a $d \times 1$ column vector function, and the superscript $\top$ denotes a matrix transpose. Although extending the model above to accommodate multivariate $Z_t$ is conceptually straightforward, it is nonetheless notationally cumbersome and so henceforth $Z_t$ is assumed to be univariate. Moreover, local linear estimation is used to derive coefficient estimates and their asymptotic properties are derived using technical results in Phillips

*I am grateful to Joon Park for valuable comments.

Email address: trokic@bilkent.edu.tr (Mirza Trokić)

The rest of the paper is organized as follows: the next section considers the time varying coefficient model in equation (1) when $Z_t$ is stationary and $X_t$ consists of stationary, nearly (possibly weakly) integrated, and purely integrated covariates. Section 3 analyzes the underlying model when $X_t$ is nearly (possibly purely) $I(1)$ and $X_t$ is a nearly integrated process. Section 4 concludes. All proofs are contained in the appendices.

2. Models with Stationary $Z_t$

The first model which is analyzed considers the case when a subset of $X_t$ is weakly (or nearly, possibly even purely) integrated in the sense of Park (2003), and $Z_t$ is strictly stationary. In particular, this model assumes that $X_t = [X_{t1}^T, X_{t2}^T, X_{t3}^T]^T$ where $X_{ti}$ is a $d_i$-dimensional column vector, $d_1 + d_2 + d_3 = d$, $X_{t1}$ is stationary with first component identically one, $X_{t2}$ is weakly (or nearly) $I(1)$, and $X_{t3}$ is a pure $I(1)$ process. The model will also assume that when $X_{t2}$ enters as a weakly integrated covariate it will enter as a univariate process with $d_2 = 0$. Moreover, if $X_{t2}$ is a nearly integrated process and $d_3 = 0$, it will be assumed that $X_{t2}$ enters as a multivariate process with $d_2 \geq 1$. On the other hand, when $X_{t2}$ is nearly integrated process and $d_2, d_3 \neq 0$, the model will assume that $d_2 = d_3 = 1$. Otherwise, if $d_2 = 0$ and $d_3 \neq 0$ then $X_{t3}$ enters as a univariate $I(1)$ process with $d_3 \geq 1$. In either scenario, the coefficient function is formally expressed as $\beta(Z_t) = [\beta_1(Z_t)^\top, \beta_2(Z_t)^\top, \beta_3(Z_t)^\top]^\top$ and the model in equation (1) is re-expressed as:

$$Y_t = \beta(Z_t)^\top X_t + \epsilon_t = \beta_1(Z_t)^\top X_{t1} + \beta_2(Z_t)^\top X_{t2} + \beta_3(Z_t)^\top X_{t3} + \epsilon_t, \quad 1 \leq t \leq n$$  \hspace{1cm} (2)

The model further assumes $\epsilon_t$’s to be innovations with respect to both $X_t$ and $Z_t$. This assumption which is formalized as $E(\epsilon_t|X_t, Z_t)$ says that $X_t$ and $Z_t$ are uncorrelated with $\epsilon_t$. Note further that $Y_t$ is allowed to be stationary or nonstationary.

2.1. Local Linear Estimation

A powerful technique for handling nonlinear statistical models is local linear fitting. As demonstrated in Fan and Gijbels (2003), Fan (2003), and Li and Racine (2007), local linear fitting is particularly appealing for its high statistical efficiency in an asymptotic minimax sense, design-adaptation, bias reduction, and automatic boundary effect correction. Accordingly, $\beta(\cdot)$ is estimated using local linear fitting from observations $\{(X_t, Z_t, Y_t)\}_{t=1}^n$. In particular, under the assumption that $\beta(\cdot)$ is twice continuously differentiable, $\beta(Z_t)$ is locally approximated as $\hat{\beta}(z) + \hat{\beta}^{(1)}(z)(Z_t - z)$ for any grid point $z$, where $\beta^{(s)}(z) = d^s \beta(z)/dz^s$. Furthermore, the vector of parameter estimates is defined as

$$\left(\hat{\theta}_0, \hat{\theta}_1\right) = \arg \min_{\theta_0, \theta_1} \sum_{t=1}^n \left[Y_t - \theta_0^\top X_t - (Z_t - z) \theta_1^\top X_t\right]^2 K_h(Z_t - z)$$  \hspace{1cm} (3)

where $K_h(u) = h^{-1}K(u/h)$ and $K(\cdot)$ is a kernel function satisfying Assumptions 5 (f) and (g), $\hat{\theta}_0 = \hat{\beta}(z)$ is an estimate of $\beta(z)$ and $\hat{\theta}_1 = \hat{\beta}^{(1)}(z)$ estimates $\beta^{(1)}(z)$. The latter two estimators can now be expressed as

$$\left(\hat{\beta}(z), \hat{\beta}^{(1)}(z)\right) = \left[\sum_{t=1}^n \frac{X_t^\top (Z_t - z) X_t}{K_h(Z_t - z)}\right]^{-1} \times \left[\sum_{t=1}^n \frac{X_t^\top Y_t}{K_h(Z_t - z)}\right]$$  \hspace{1cm} (4)

where $A^{\otimes 2} = AA^\top (A^{\otimes 1} = A)$ for any vector or matrix $A$. 

2
2.2. Notations and Assumptions

Consider first the weakly integrated process defined as

\[ X_t^2 = \alpha X_{t-1}^2 + \eta_t \]

where \( \alpha = 1 - m/n \), the sample size is \( n \), and \( m \) is a function of \( n \) satisfying \( m/n \to 0 \) as \( n \to \infty \). Moreover, \( \eta_t \) is assumed to be a mean zero \( I(0) \) linear process \( \eta_t = \pi(L)u_t = \sum_{k=0}^{\infty} \pi_k u_{t-k} \) with longrun variance \( \sigma^2_\eta = \pi(1)^2 \mathbb{E} u_t^2 \) and satisfying

**Assumptions 1.**

(a) \( \sum_{k=0}^{\infty} k|\pi_k| < \infty \) and \( \mathbb{E}|u_t|^p < \infty \) for some \( p > 2 \).

(b) The distribution of \( \{u_t\} \) is absolutely continuous with respect to the Lebesgue measure, and has characteristic function \( \phi_u \) for which \( \lim_{t \to \infty} t^3 \phi_u(t) = 0 \).

Note that this framework interprets \( X_t^2 \) as a general linear process with a weak unit root with the autoregressive-weakly-integrated-moving-average (ARWIMA) process as a special case. Moreover, this construction implies that although \( \alpha \) tends to unity for all functions of \( m \), the rate of convergence will depend on the exact functional form of \( m \). In particular, the process assumes that \( m \to \infty \) as \( n \to \infty \) and therefore whenever \( m \) is not a constant it is clear that \( \alpha \) tends to unity at a rate slower than \( n^{-1} \). This is in contrast to the classical near unit root case of Phillips (1987) where \( m = c \) for some constant \( c > 0 \) and \( \alpha \) converges to unity at rate \( n^{-1} \). When this is indeed the case, the model under consideration will assume the (possibly multivariate) form

\[ X_t^2 = (1 - c/n) X_{t-1}^2 + \eta_t \]

where \( \eta_t \) is a mean zero stationary strong mixing process with variance \( \Sigma_\eta \) and satisfying one of the following two conditions:

**Assumptions 2.**

(a) For some \( \delta_\circ > 0 \), \( \mathbb{E}|\eta_t|^{2+\delta_\circ} < \infty \) and \( \sum_{k=0}^{\infty} \alpha(k)k^{(2+\delta_\circ)/\delta_\circ} < \infty \).

(b) For some \( \gamma_\star > 2+\delta_\circ \) with \( 0 < \delta_\circ \leq 2 \) and \( \lambda_\star = \lambda_\star(\delta_\circ) > 0 \), \( \mathbb{E}|\eta_t|^\gamma_\star < \infty \) and \( \sum_{k=0}^{\infty} \alpha(k)^{1/(2+\delta_\circ) - 1/\gamma_\star} < \infty \).

Similarly, when \( \alpha = 1 \) and the pure \( I(1) \) process is modelled as the unit root process \( X_t^3 = X_{t-1,3} + \omega_t \), assume \( \{\omega_t\} \) has variance \( \Sigma_\omega \) and satisfies one of the following:

**Assumptions 3.**

(a) For some \( \delta_\bullet > 0 \), \( \mathbb{E}|\omega_t|^{2+\delta_\bullet} < \infty \) and \( \sum_{k=0}^{\infty} \alpha(k)k^{(2+\delta_\bullet)/\delta_\bullet} < \infty \).

(b) For some \( \gamma_\bullet > 2+\delta_\bullet \) with \( 0 < \delta_\bullet \leq 2 \) and \( \lambda_\bullet = \lambda_\bullet(\delta_\bullet) > 0 \), \( \mathbb{E}|\omega_t|^\gamma_\bullet < \infty \) and \( \sum_{k=0}^{\infty} \alpha(k)^{1/(2+\delta_\bullet) - 1/\gamma_\bullet} < \infty \).

Finally, observe the following processes associated with \( X_t^2 \) and \( X_t^3 \).
\[ V_{mn}(r) = n^{-1/2}X_{[nr],2} \]
\[ V_{cn}(r) = n^{-1/2}X_{[nr],2} \]
\[ V_{0n}(r) = n^{-1/2}X_{[nr],3} \]
\[ V_{n,m}(r) = \int_0^r \exp(-m(r-s))dV_{\eta,0}(s) \]
\[ V_{n,c}(r) = \int_0^r \exp(-C(r-s))dV_{\eta,0}(s) \]
\[ V_{\eta}(r) = \sqrt{m}V_{\eta,m}\left(\frac{r}{m}\right) \]

where \( r \in [0,1] \), \(|z|\) denotes the largest integer which does not exceed \( z \), \( C \) is a \( d_2 \)-dimensional diagonal matrix \( \text{diag} \{c, \ldots, c\} \), \( V_{\eta,0} \) is a \( d_2 \)-dimensional (\( d_2 = 1 \) in the case of \( V_{n,m} \)) multivariate (univariate) Brownian motion process with covariance matrix \( \Sigma_{\eta} \) which reduces to \( \sigma_n^2 \) in case of the weakly integrated construction, and \( V_{n,m} \) and \( V_{n,c} \) are respectively the univariate and \( d_2 \)-dimensional multivariate Ornstein-Uhlenbeck processes which collapse to Brownian motions on \([0,1]^{d_2}\) and \([0,1] \) respectively when \( m = c = 0 \). Finally, the normalized process \( V_{n}(r) \) is an Ornstein-Uhlenbeck process with unit parameter and stable marginal distribution which approaches \( N(0, \sigma_n^2/2) \) as \( r \to \infty \), and the density of which will be denoted as \( D \). Note further that for any fixed \( m > 0 \), as \( n \to \infty \),

\[ V_{mn} \to_d V_{n,m} \]
\[ V_{cn} \to_d V_{n,c} \]
\[ V_{0n} \to_d V_{\omega,0} \]

where \( \to_d \) denotes convergence in distribution and \( V_{\omega,0} \) is a Brownian motion on \([0,1]^{d_3}\) with covariance matrix \( \Sigma_{\omega} \). Both results can in fact be strengthened further under Assumptions 1 to 3, and appropriate probabilistic embeddings of \( V_{mn} (V_{cn}) \) and \( V_{n,m} (V_{n,c}) \) and \( V_{0n} \) and \( V_{\omega,0} \) respectively, on common respective probability spaces so that

\[ \sup_{0 \leq r \leq \infty} \left| V_{mn}(r) - V_{n,m}(r) \right| = o_p \left( n^{-1/2+1/p} \right) + o_p \left( mn^{-1} \right) \]
\[ \sup_{0 \leq r \leq \infty} \left| V_{cn}(r) - V_{n,c}(r) \right| = O_{a.s.} \left( n^{-1/2+1/(2+s)} \log^{\lambda}(n) \right) \]
\[ \sup_{0 \leq r \leq \infty} \left| V_{0n}(r) - V_{\omega,0}(r) \right| = O_{a.s.} \left( n^{-1/2+1/(2+s)} \log^{\lambda}(n) \right) \]

for large \( n \), uniformly in \( m \) such that \( m/n \to 0 \) as \( n \to \infty \), where \(|\cdot|\) is the uniform norm, and \( \|\cdot\| \) is the \( L_2 \) norm in \( \mathbb{R}^{d_2} \) and \( \mathbb{R}^{d_3} \) in equations (7b) and (7c), respectively. The particular appeal of these result is that \( V_{n,m} (V_{n,c}) \) approximates \( V_{mn} (V_{cn}) \) and \( V_{\omega,0} \) approximates \( V_{0n} \) with negligible error for all large \( n \). In other words, all asymptotics for \( X_{l2} \) can be derived from the asymptotics of functionals of \( V_{n,m} \) or \( V_{n,c} \), and similarly for \( X_{l3} \). Moreover, the asymptotic results which follow will frequently exploit the local times \( L_{\eta}(r,x) \) of the normalized process \( V_{\eta}(s) \) in equation (5d) and the local times \( L_{\omega}(r,x) \) of the Brownian motion process \( V_{\omega,0}(s) \). Roughly speaking, the local time measures the length of time, up to time \( r \), a diffusion process spends in an immediate neighbourhood of \( x \). Formally, it is defined as
\[ L_\eta(r, x) = \lim_{\delta \to 0} \frac{1}{2\delta} \int_0^r 1 \{|V_\eta(s) - x|\} ds \]
\[ L_\omega(r, x) = \lim_{\delta \to 0} \frac{1}{2\delta} \int_0^r 1 \{|V_\omega(0)(s) - x|\} ds \]

Moreover, the \( m \)-asymptotics which characterize the weakly integrated process of interest here will be handled using the function

\[ D_m(x) = \frac{1}{m} L_\eta(m, x) \]

which, as shown in Park (2003), satisfies

\[ D_m(x) = D(x) + o(m^{-1/2} \log m \log \log m) \quad \text{a.s.} \]

uniformly over any compact interval, and for any \( k > -1 \),

\[ \int_{-\infty}^{\infty} |x|^k D_m(x) dx \to \int_{-\infty}^{\infty} |x|^k D(x) dx \]

Consider next the class of asymptotically homogeneous functions studied in detail in Park and Phillips (1999, 2001); Park (2003).

**Definition 1.** Let \( F : \mathbb{R} \to \mathbb{R} \). \( F \) is said to be regular if, for any \( \delta > 0 \) sufficiently small, it satisfies that

1. for all \( |x| \geq \delta \) and \( |x - y| \leq \delta/2 \), \( |F(x) - F(y)| \leq K_{ab} |x - y| \) with \( K_{ab}(x) = K(1 + |x|^a)(1 + |x|^b) \), where \( a > 0, b < 0 \) and \( K \) are some constants not depending upon \( \delta \), and

2. for all \( |x| < \delta \), \( |F(x)| \leq K|x|^c \) with some constants \( c > -1 \) and \( K \) independent of \( \delta \).

\( F \) is said to be regular in the second order if both \( F \) and \( F^2 \) are regular.

**Definition 2.** A function \( H : \mathbb{R} \to \mathbb{R} \) is said to belong to the class of asymptotically homogeneously functions if and only if

\[ H(\lambda x) = \kappa(\lambda) F(x) + R(x, \lambda) \]

provided

1. \( F(\cdot) \) is regular in the second order, and

2. \( |R(x, \lambda)| \leq \nu(\lambda) Q(x) \) for all \( \lambda \) sufficiently large and for all \( x \) over any compact set where \( \nu(\lambda)/\kappa(\lambda) \to 0 \) as \( \lambda \to \infty \) and \( Q(\cdot) \) is regular in the second order.

Introduce next the series \( \{v_t\} \) and consider the following set of assumptions.
**Assumptions 4.** Let \( \{v_t\} \) be a martingale difference sequence with respect to some filtration \( \mathcal{F}_t \) such that
(a) \( X_{t2} \) is adapted to \( \mathcal{F}_{t-1} \), and
(b) \( E\left(v_t^2|\mathcal{F}_{t-1}\right) = \sigma^2 \) a.s. for all \( t \), and \( \sup_t E\left(|v_t|^{2+\delta}|\mathcal{F}_{t-1}\right) < \infty \) a.s. for some \( \delta > 0 \).

Under Assumptions 1 and 4, Theorem 3.8 in Park (2003) provides a useful convergence result for weakly integrated processes when \( m = o\left(n^{1-2/p} \wedge n^{2/3}\right) \) and \( H \) belongs to the class of asymptotically homogeneous functions with asymptotic order \( \kappa \) and limit homogeneous function \( F \). In particular, the result implies that when \( m \to \infty \) as \( n \to \infty \), then,

\[
\frac{1}{n} \left( \frac{n}{m} \right)^{-1} \sum_{t=1}^{n} H(X_{t2}) \to_p \int_{-\infty}^{\infty} F(x)D(x)dx \quad (8a)
\]
\[
\frac{1}{n} \left( \frac{n}{m} \right)^{-1} \sum_{t=1}^{n} H(X_{t2}v_t) \to_d N\left(0, \sigma_v^2 \int_{-\infty}^{\infty} F^2(x)D(x)dx\right) \quad (8b)
\]

An important implication of this result concerns the subclass of polynomial functions in the class of asymptotically homogeneous transformations. The two results of particular interest here arise when \( F(x) = x^l \) for \( l = 1, 2 \), in which case the homogeneity function \( \kappa \left(\frac{n}{m}\right)^{-1} = \left(mn^{-1}\right)^{l/2} \) and, as \( m, n \to \infty \),

\[
\frac{1}{n} \sum_{t=1}^{n} \left( \sqrt{mn^{-1}} X_{t2} \right)^{\otimes l} \to_p \int_{-\infty}^{\infty} x^l D(x)dx \equiv V^{(l)}_D(x) \quad (9)
\]

On the other hand, observe that if \( B(\cdot) \) is any Borel measurable and totally Lebesgue integrable function (see Berkes and Horváth (2006)), and in particular, if \( B(l) = (\cdot)^{\otimes l} \) for \( l = 1, 2 \), then as \( n \to \infty \),

\[
\frac{1}{n} \sum_{t=1}^{n} \left( \frac{1}{\sqrt{n}} X_{t2}^{(n)} \right)^{\otimes l} \to_d V^{(l)}_{\eta,m} \quad (10a)
\]
\[
\frac{1}{n} \sum_{t=1}^{n} \left( \frac{1}{\sqrt{n}} X_{t2}^{(n)} \right)^{\otimes l} \to_d V^{(l)}_{\eta,0} \quad (10b)
\]

Since the asymptotics to follow will also involve covariance asymptotics between two nearly integrated series or a nearly integrated and purely integrated series, consider also Theorem 1 of Phillips (2009). This result establishes that joint triangular sequences \( (x_{t,n}, y_{t,n}) \) which jointly converge weakly to continuous Gaussian processes \( (G_x(r), G_y(r)) \), satisfy the following relation for any \( c_n \to \infty \) s.t. \( c_n n \to 0 \) and \( r \in [0, 1] \):

\[
(c_n n^{-1} \sum_{t=1}^{[nr]} y_{t,n} g(c_n x_{t,n}) \to_d \int_{-\infty}^{\infty} g(s)ds \int_0^r G_y(p)dL_{G_x}(p, 0) \quad (11)
\]

where \( g(\cdot) \) is a Lebesgue integrable function on \( \mathbb{R} \) with nonzero energy and \( L_{G_x}(p, s) \) is the local time of the process \( G_x(t) \). In particular, observe that when \( c_n \) is a constant, say \( c_n = 1 \), the result implies that when \( (V_{cn}, V_{0n}) \) jointly converge weakly to \( (V_{\eta,c}, V_{\eta,0}) \), then

\[
(n^{-1} \sum_{t=1}^{[nr]} V_{cn} V_{0n} \to_d \int_0^1 V_{\eta,c}(r)V_{\eta,0}dr \equiv V_{\eta,c,\eta,0} \quad (12)
\]

The following set of assumptions are now imposed.
Assumptions 5.

(a) $\beta(z)$ is twice continuously differentiable in $z$ for all $z \in \mathbb{R}$.

(b) $A_k(z)$ is positive-definite and continuous in a neighbourhood of $z$. $f(z)$ is continuously differentiable in a neighbourhood of $z$ and $f_z(z) > 0$.

(c) $\epsilon_i$ has a finite fourth moment $E(\epsilon_i | X_i, Z_i) = 0$ and $E(\epsilon_i^2 | X_i, Z_i) = \sigma_i^2$ is a positive constant.

(d) $\{X_{11}, Z_t, \epsilon_t, \eta_t : t \geq 1 \}$ is a strictly $\alpha$-mixing stationary process with the $\delta_1$-th moment ($\delta_1 > 2$).

The assumptions above are similar to those imposed in Cai and Wang (2008) and Cai et al. (2009). In particular, Assumptions 5 (a) and (b) are smoothness conditions while (c) assumes that regression errors are conditionally homoskedastic. Assumptions 5 (d) is satisfied under standard moment conditions if $((c))$.

Consider next the regularity conditions required to establish the limiting distribution of $\hat{\beta}(z)$.

The kernel function $K(\cdot)$ is a symmetric and continuous density function with support $[-1,1]$.

The bandwidth $h$ satisfies $h \to 0$ and $nh \to \infty$.

(h) $n^{1/2-\delta_1/4}h^{\delta_1/2-1-\delta_1/4} = O(1)$

2.3. Asymptotic Properties

To develop the asymptotic properties of $\hat{\beta}(z)$, first let

$$D_{m,n} = \text{diag} \{ I_{d_1}, m^{-1/2}I_n^{1/2}I_{d_2} \} \text{ and } D_n = \text{diag} \{ I_{d_1}, n^{1/2}I_{d_2+d_3} \}$$

where $I_{d_1}$ is a $d_1 \times d_1$ identity matrix, and define $B_\beta(z) = \mu_2(K)\beta^{(2)}(z)/2$. Observe now the first major result.
Theorem 1. Under Assumptions 1 to 5, when

(a) $X_{t2}$ is a weakly integrated process with $m = o\left(n^{1 - \frac{1}{d}} \wedge n^{\frac{2}{d}}\right)$, $d_2 = 1$ and $d_1 = 0$, then

$$\sqrt{m} D_{m,n} \left( \hat{\beta}(z) - \beta(z) - h^2 B_\beta(z) \right) \rightarrow_d N(\Sigma_{\beta,W}(z))$$

where $N(\Sigma_{\beta,W}(z))$ is a multivariate normal distribution with mean zero and conditional covariance matrix given by $\Sigma_{\beta,W}(z) = \nu_0(K)S_W(z) - f_z(z)$.

(b) $X_{t2}$ is a nearly integrated process with $d_2 \geq 1$ and $d_3 = 0$, then

$$\sqrt{m} D_{n} \left( \hat{\beta}(z) - \beta(z) - h^2 B_\beta(z) \right) \rightarrow_d MN(\Sigma_{\beta,N}(z))$$

where $MN(\Sigma_{\beta,N}(z))$ is a mean zero mixed normal distribution with conditional covariance matrix $\Sigma_{\beta,N}(z) = \nu_0(K)\left(Q^T S_N(z)Q\right) - f_z(z)$.

Note that Theorem 1 nests several important results. When $d_2 = d_3 = 0$ the results of Cai et al. (2000) are recovered with $\Sigma_{\beta,W}(z) = \Sigma_{\beta,0,W}(z) = \sigma^2_2 \nu_0(K) M_2(z) - f_z(z)$. On the other hand, when $d_2 = 0$ and $d_3 \geq 1$ the results of Cai et al. (2009) emerge with $\Sigma_{\beta,N}(z) = \Sigma_{\beta,0,\beta_3,N}(z) = \sigma^2_3 \nu_0(K) S(z) - f_z(z)$.

Theorem 1 also lends insight into convergence rates for $V\left(\hat{\beta}_1\right)$, $V\left(\hat{\beta}_2\right)$, and $V\left(\hat{\beta}_3\right)$. In particular, the use of local linear fitting to estimate $\beta$ implies that $V\left(\hat{\beta}_1\right)$ is of order $O\left((nh)^{-1}\right)$, $V\left(\hat{\beta}_2\right)$ is of order $O\left(m(3n^2h)^{-1}\right)$ when $X_{t2}$ is weakly $I(1)$ and $O\left((n^2h)^{-1}\right)$ when $X_{t2}$ is nearly $I(1)$. The order of $V\left(\hat{\beta}_3\right)$ is $O\left((n^2h)^{-1}\right)$ when $X_{t3}$ is a pure $I(1)$ process. Since $m = o\left(n^{1 - \frac{1}{d}} \wedge n^{\frac{2}{d}}\right)$, it is clear that the rate of convergence of the variance of $\hat{\beta}_3$ is slower in the case of weakly integrated process than nearly integrated and purely integrated ones. Moreover, it is also clear that when $X_{t2}$ is a nearly integrated process the rates of convergence of the variances of $\hat{\beta}_2$ and $\hat{\beta}_3$ are the same.

Another popular approach to comparing the estimators is to consider their integrated asymptotic mean squared error (IAMSE). In this regard, note that for $i = 1, 2$ and $j = 1, 2, 3$, the IAMSEs for the model with weakly and nearly integrated covariates respectively, assume the forms

$$IAMSE_{W}(\hat{\beta}_i(z)) = \frac{1}{4} h^4 \mu_2^2(K) \int \|\beta_i^{(2)}(z)\|w(z)dz + (m^{-i+1}n^i)^{-1} \int tr(\Sigma_{\beta_i,W}(z))w(z)dz$$

for $i = 1, 2, \ldots$, and

$$IAMSE_{N}(\hat{\beta}_j(z)) = \begin{cases} \frac{1}{4} h^4 \mu_2^2(K) \int \|\beta_j^{(2)}(z)\|w(z)dz + (n^j h)^{-1} \int tr(\Sigma_{\beta_j,N}(z))w(z)dz & \text{for } j = 1, 2 \\ \frac{1}{4} h^4 \mu_2^2(K) \int \|\beta_j^{(2)}(z)\|w(z)dz + (n^j h)^{-1} \int tr(\Sigma_{\beta_j,N}(z))w(z)dz & \text{for } j = 3 \end{cases}$$
for some weight function $w(\cdot) \geq 0$, where $\Sigma_{\beta_i,W}(z)$ and $\Sigma_{\beta_j,N}(z)$ are submatrices of dimension $d_i$ and $d_j$ along the diagonals of $\Sigma_{\beta_i,W}(z)$ and $\Sigma_{\beta_j,N}(z)$ respectively, the first elements of which are indexed by $(d_i,d_i)$ and $(d_j,d_j)$. The optimal bandwidth can now be derived by minimizing the IAMSE with respect to $h$ and obtaining the minimizer $h^*_t$. It is easily verified that the optimal bandwidth in case of the weak unit root model is $h^*_t = 0$, since $\beta_t$ is a univariate near $I(1)$ process, satisfying $\xi_t$ is any non-negative constant and $n$ is the sample size. Recall that when $c_z = 0$, $Z_t$ models a pure $I(1)$ process and when $c_z > 0$, it generates a near $I(1)$ process. In either scenario, $\xi_t$ is assumed to be a mean zero $I(0)$ linear process satisfying

**Assumptions 6.**

(a) For some $\delta > 0$, $E|\xi_t|^{2+\delta} < \infty$ and $\sum_{k=0}^{\infty} \alpha(k)k^{(2+\delta)/\delta} < \infty$. 

\[ Z_t = (1 - c_z/n)Z_{t-1} + \xi_t \quad (14) \]
(b) For some $\gamma > 2 + \delta$ with $0 < \delta \leq 2$ and $\lambda = \lambda(\delta) > 0$, $E|\xi_t|^\gamma < \infty$ and $\sum_{k=0}^\infty \alpha(k)^{1/(2+\delta)-1/\gamma} < \infty$

Similarly, as in Section 2.2, for $c_2 > 0$, $X_t$ is a near $I(1)$ process

$$X_t = (1 - c_2/n)X_{t-1} + \eta_t$$ (15)

satisfying Assumptions 2. Furthermore, let $W_t = (X_t^T, Z_t)^T$ and consider a real matrix of coefficients $C_j = (c_{j,k}: 1 \leq k, l \leq d + 1)$ which satisfy $W_t = \sum_{j=0}^\infty C_j \tilde{\epsilon}_{t-j}$. Finally, impose the following set of assumptions.

**Assumptions 7.**

(a) $\{\epsilon_i\}$ is a sequence of IID continuous random vectors with $E(\epsilon_1) = 0$, a positive definite matrix $\Sigma_\epsilon$ and finite fourth order cumulants.

(b) $\int_{-\infty}^{\infty} |u| |\phi_\epsilon(u)| du < \infty$ where $\phi_\epsilon(u)$ is the characteristic function of $\epsilon_1$.

(c) $\int_{-\infty}^{\infty} p_\epsilon(x + y) - p_\epsilon(x) dx \leq c_\epsilon |y|$ for each $y$ and constant $c_\epsilon$, where $p_\epsilon(\cdot)$ is the density of $\epsilon_1$.

(d) $E(\|\epsilon_1\|^{2+\gamma}) < \infty$ for some $\gamma > 0$ such that $2\gamma^2 + 4\gamma - 5 > 0$.

(e) $\sum_{j=0}^\infty c_{j,kl} x^j$ for $|x| \leq 1$ and $c_{j,kl} = O(j^{-\gamma})$ as $j \to \infty$ and $\zeta_\epsilon > 1$ satisfies $\zeta_\epsilon + 1/2 > 1 + \zeta_\epsilon > 1$ with $\zeta$ defined in (d) above.

(f) $G = \sigma(\epsilon_1, \ldots, \epsilon_t, \epsilon_{t+1}, \epsilon_t, \ldots, \epsilon_{\infty})$ be a $\sigma$-field generated by $\{(\epsilon_i, \epsilon_j): 1 \leq i \leq t; -\infty \leq j \leq t + 1\}$ where $E(\epsilon_t | G_{t-1}) = 0$ almost surely (a.s.), $E(\epsilon_t^2 | G_{t-1}) = \sigma_\epsilon^2$ a.s., and $E(\epsilon_t^2 | G_{t-1}) < \infty$ a.s. for all $t \geq 2$ where $\sigma_\epsilon^2 > 0$.

(g) Let $W_{c,n}(r) = n^{-1/2} \sum_{t=1}^{[nr]} \epsilon_t$ and $W_n(r) = (V_{c,n}(r)^T, V_{\epsilon,n}(r)^T)^T$ where similar to equation (5a), $V_{c,n}(r) = n^{-1/2} X_{[nr]}$ and $V_{\epsilon,n}(r) = n^{-1/2} Z_{[nr]}$. There exists a Skorohod space $D[0,1]^{d+2}$ on which $(W_{c,n}(r), W_n(r)) \to_d (V_{c,0}(r), V_n(r))$ as $n \to \infty$, where $(V_{c,0}(r), V_n(r))$ is a vector stochastic process, $V_{c,0}(r)$ is a Brownian motion process, $V_n(r) = (V_{n,c}(r), V_{n,\epsilon}(r))$ and $V_{n,c}(r)$ and $V_{n,\epsilon}(r)$ are Ornstein-Uhlenbeck processes with parameters $c_\epsilon$ and $c_\epsilon$, respectively, when $c_\epsilon, c_\epsilon > 0$. When $c_\epsilon = 0$, $V_{\epsilon,0}(r)$ is a Brownian motion.

(h) $K(\cdot)$ is a continuous, symmetric, non-negative and bounded probability kernel function satisfying $\int \|u\| K(u) du < \infty$.

(i) Let $h \to 0$ and $nh \to 0$ as $n \to \infty$.

The assumptions above are a close adaptation of Assumption A.1 in Gao and Phillips (2013). In particular, assumptions (a) - (e) ensure that $W_t$ is stationary and $\alpha$-mixing and accommodates contemporaneous endogeneity between regressors, varying coefficient components, and regression residuals, whereas assumption (f) and (g) allow for heteroskedastic residuals $\epsilon_t$. It should also be noted that a similar set of assumptions also exists in Phillips (2009) and Wang and Phillips (2009) although the latter did not consider contemporaneous endogeneity between covariates, coefficient components, or residuals.

**Theorem 2.** Under Assumptions 2, 6 and 7, let $X_t$ and $Z_t$ be defined by equations (14) and (15) with $c_2 > 0$ and $c_2 \geq 0$. Then, as $n \to \infty$,

$$\sqrt{nh^{3/4}} \left( \hat{\beta}(z) - \beta(z) - h^2 B_\beta(z) \right) \to_d MN(\Sigma_\beta)$$
a mean zero mixed normal distribution with conditional covariance matrix

\[ \Sigma_\beta = \sigma^2_0 \nu_0(K) \int_0^1 V_{\eta,cz} V_{\eta,cz}^T dL_{\xi,\epsilon z} (1, 0) \]

There are several important remarks to note at this point. Observe in particular that Theorem 2 implies the asymptotic variance of \( \hat{\beta}(z) \) has order \( O\left(h^{-1}n^{-1/2}\right) \). This is clearly larger than \( O\left(h^{-1}n^{-1}ight) \) which arises in the case of stationary \( Z_t \). In fact, this is also in clear contrast to the case of stationary \( X_t \) and nonstationary \( Z_t \) considered in Cai et al. (2009) where the asymptotic variance of \( \hat{\beta}(z) \) has order \( O\left(h^{-1}n^{-1/2}\right) \). On the other hand, the asymptotic bias term \( h^2 B_\beta(z) \) remains the same in all three scenarios. This of course is to be expected since the bias term arises as the result of using local linear estimation. These observations readily lead to the IAMSE of \( \hat{\beta} \) given by

\[ IAMSE = \int \left( \frac{1}{4} h^4 \mu^2_2(K) \| \beta^{(2)}(z) \| + h^{-1}n^{-3/4} \text{tr} (\Sigma_\beta) \right) w(z) dz \]

for some non-negative weight function \( w(\cdot) \). Note further that minimizing the IAMSE with respect to \( h \) renders the optimal bandwidth \( h^\ast = cn^{-3/20} \) for some \( c > 0 \). Although this is significantly smaller than the optimal bandwidth derived when \( Z_t \) is stationary, it is nonetheless somewhat larger than the optimal bandwidth obtained in the case of stationary \( X_t \) and nonstationary \( Z_t \) in Cai et al. (2009).

4. Concluding Remarks

This paper has analyzed the time varying coefficients model when covariates are nearly (possibly weakly) integrated and time varying coefficients are stationary or nearly (possibly purely) integrated. Along similar lines of reasoning to Cai and Wang (2008) and Cai et al. (2009), time varying coefficient components in this note were estimated nonparametrically using the the local linear fitting scheme and their asymptotic properties were derived using local time asymptotics. In particular, when \( Z_t \) is stationary, the asymptotics of time varying coefficient depend on whether the covariates \( X_t \) are nearly or weakly integrated. Nevertheless, the rate of convergence of the variance of time varying coefficients remains the same regardless of whether \( X_t \) is a near or pure \( I(1) \) process. On the other hand, the asymptotic analysis of Section 3 with nearly integrated covariates and nearly (possibly purely) integrated time varying components produces estimators with variances of larger order than in the case of stationary \( X_t \) and \( Z_t \) and stationary \( X_t \) and pure \( I(1) \) time varying components considered in Cai et al. (2009). Similar conclusions also hold for the optimal bandwidth choices. In this regard, although not explicitly analyzed here, the two-step estimation procedure considered in Cai et al. (2009) to deal with optimal bandwidths of different orders in the stationary \( Z_t \) models clearly continues to hold with nearly (weakly) integrated covariates as well. It is also worth mentioning that this note has answered several appeals in the literature to develop a theory for time varying coefficient models when both the covariates and time varying components are nearly integrated process. To the best knowledge of this author, this article is the first such contribution. Finally, given the importance of nearly (weakly) integrated process in financial and macroeconomic modelling, it is warranted to encourage the use of methods developed here in relevant empirical studies.
References


Appendix A

Proof of Theorem 1. The proof follows the proof of Theorem 2.1 in Cai et al. (2009). In this regard, define $\mathcal{H}_{m,n} = \begin{pmatrix} 1 & 0 \\ 0 & h \end{pmatrix} \otimes D_{m,n}$ and $\mathcal{H}_n = \begin{pmatrix} 1 & 0 \\ 0 & h \end{pmatrix} \otimes D_n$ and note that

$$
\mathcal{H}_{m,n} \left( \theta(z) \right) = S_{m,n}(z)^{-1} n^{-1} \sum_{t=1}^{n} K_h(Z_t - z) 
$$

$$
\times Y_t \left( \frac{1}{Z_{t,z,h}} \right) \otimes (D_{m,n}^{-1} X_t) 
$$

$$
\mathcal{H}_n \left( \theta(z) \right) = S_n(z)^{-1} n^{-1} \sum_{t=1}^{n} K_h(Z_t - z) 
$$

$$
\times Y_t \left( \frac{1}{Z_{t,z,h}} \right) \otimes (D_n^{-1} X_t) 
$$

(16a)

(16b)

where $Z_{t,z,h} = h^{-1} (Z_t - z)$ and $S_{m,n}(z)$ and $S_n(z)$ may be partitioned as

$$
S_{m,n}(z) = \begin{pmatrix} S_{m,n,0}(z) & S_{m,n,1}(z) \\ S_{m,n,1}(z) & S_{m,n,2}(z) \end{pmatrix}
$$

$$
S_n(z) = \begin{pmatrix} S_n,0(z) & S_n,1(z) & S_n,4(z) \\ S_n,1(z) & S_n,2(z) & S_n,7(z) \\ S_n,4(z) & S_n,7(z) & S_n,5(z) \end{pmatrix}
$$

where for $j = 0, 1, 2, 4, 5, 7,$

$$
S_{m,n,j}(z) = n^{-1} \sum_{t=1}^{n} K_h(Z_t - z) Z_{t,z,h}^j (D_{m,n}^{-1} X_t)^{\otimes 2}
$$

$$
S_{n,j}(z) = n^{-1} \sum_{t=1}^{n} K_h(Z_t - z) Z_{t,z,h}^{j \mod 3} (D_n^{-1} X_t)^{\otimes 2}
$$

Note in fact that $S_{m,n,j}(z)$ and $S_{n,j}(z)$ can further be partitioned as

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

$$
S_{n,j}(z) = \begin{pmatrix} F_{n,j,0}(z) & F_{n,j,1}(z) & F_{n,j,4}(z) \\ F_{n,j,1}(z) & F_{n,j,2}(z) & F_{n,j,4}(z) \\ F_{n,j,4}(z) & F_{n,j,5}(z) & F_{n,j,5}(z) \end{pmatrix}
$$

where
\[ F_{m,n,j,0}(z) = F_{n,j,0}(z) = n^{-1} \sum_{t=1}^{n} Z_{t,z,h}^j K_h(Z_t - z)X_t^\top X_t \]
\[ F_{m,n,j,1}(z) = m^{1/2} F_{n,j,0}(z) = n^{-1} \sum_{t=1}^{n} K_h(Z_t - z)Z_{t,z,h}^j X_t^{1/2}n^{-1/2}X_t^\top \]
\[ F_{m,n,j,2}(z) = m F_{n,j,2}(z) = n^{-1} \sum_{t=1}^{n} Z_{t,z,h}^j K_h(Z_t - z) \left( m^{1/2}n^{-1/2}X_t \right)^\otimes 2 \]
\[ F_{n,j,3}(z) = n^{-1} \sum_{t=1}^{n} Z_{t,z,h}^j K_h(Z_t - z)X_t n^{-1/2}X_t^\top \]
\[ F_{n,j,4}(z) = n^{-1} \sum_{t=1}^{n} Z_{t,z,h}^j K_h(Z_t - z)n^{-1/2}X_t^2 n^{-1/2}X_t^\top \]
\[ F_{n,j,5}(z) = n^{-1} \sum_{t=1}^{n} Z_{t,z,h}^j K_h(Z_t - z) \left( n^{-1/2}X_t \right)^\otimes 2 \]

For \( l = 1,2 \), also define the quantity
\[ F_{n,j,l}^*(z) = n^{-1} \sum_{t=1}^{n} Z_{t,z,h}^j X_t^{\otimes l} K_h(Z_t - z) \]

Note further that \( X_t^{1/2} \) and \( Z_t \) are stationary. Accordingly, with a change-of-variables transformation applied to \( K_h(\cdot) \) and a Taylor expansion argument on the resulting density functions, the following holds.
\[ \mathbb{E} F_{n,j,l}^*(z) = \mathbb{E} \left( Z_{t,z,h}^j X_t^{\otimes l} K_h(Z_t - z) \right) = f_z(z)M_t(z)\mu_j \otimes 3(K) + o(1) \]
\[ \mathbb{V} F_{n,j,l}^*(z) = O((nh)^{-1}) = o(1) \]

where the last line above follows from Assumptions 5 (g). Accordingly,
\[ F_{n,j,l}^*(z) = f_z(z)M_t(z)\mu_j \otimes 3(K) + o_p(1) \quad (17) \]
and therefore
\[ F_{m,n,j,0}(z) = F_{n,j,0}(z) = F_{n,j,2}(z) = f_z(z)M_t(z)\mu_j \otimes 3(K) + o_p(1) \quad (18) \]

Consider next \( F_t^* = \sigma (X_{t1}, Z_t : t \leq 1) \) as the smallest \( \sigma \)-field containing the history of \( (X_{t1}, Z_t) \) and define \( \epsilon_t = K_h(Z_t - z)Z_{t,z,h}^j X_t - \mathbb{E} \left( K_h(Z_t - z)Z_{t,z,h}^j X_t \right) \). Also, let \( V_{mnt} = m^{1/2}n^{-1/2}X_t^2 \) and note that
\(V_{mn[nr]} = m^{1/2} V_{mn}(r)\) where \(V_{mn}(r)\) was defined in equation (5a) for any \(r \in [0, 1]\). Moreover, for \(0 \leq \delta \leq 1\), set \(N = \lceil 1/\delta \rceil\), \(t_k = \lceil kn/N \rceil + 1\), \(t_k^* = t_k + 1\), and \(t_k^{**} = \min \{t_k^*, n\}\). Then, replacing \(U_{nt}\) in the proof of Cai et al. (2009) with \(V_{mnt}\) here, yields the following result.

\[
\left| n^{-1} \sum_{i=1}^{n} V_{mn} e_i \right| = \left| n^{-1} \sum_{k=0}^{N-1} \sum_{t=t_k}^{t_k^*} V_{mnt} e_t \right| \\
\leq n^{-1} \sum_{k=0}^{N-1} \sum_{t=t_k}^{t_k^*} \left| \mathbb{E} (V_{mnk} V_{mnt} e_k e_t) \right| \\
= n^{-1} \sum_{k=0}^{N-1} \sum_{t=t_k}^{t_k^*} \left| \mathbb{E} (V_{mnk} V_{mnt} \mathbb{E} (e_k e_t | Z_k, Z_t)) \right|
\]

Next, note that

\[
n^{-1} \sum_{t=0}^{t_k^*} \left| \mathbb{E} (e_k e_t | Z_k, Z_t) \right| \leq \frac{N}{n} \sup_{0 \leq k \leq N-1} \mathbb{E} \left( t_k^{**} \sum_{t=t_k}^{t_k^*} e_t^2 | Z_t \right) \\
\leq \sup_{t \leq n} \mathbb{E} \left( \frac{1}{\delta n} \sum_{i=t}^{t+\delta n} e_i^2 \right) \\
\leq Ch^{-1}
\]

where the last line follows from the result

\[
\sup_{s \geq 0} \mathbb{V} \left( \sum_{t=s+1}^{s+a} e_t \right) = O(ah^{-1})
\]

Accordingly,

\[
n^{-1} \sum_{i=1}^{n} V_{mn} e_i \leq Ch^{-1} n^{-1} \sum_{k=0}^{N-1} \sum_{t=t_k}^{t_k^*} \left| \mathbb{E} (V_{mnk} V_{mnt}) \right| \\
\leq Ch^{-1} \sup_{t \leq n} \mathbb{E} \left( n^{-1} \sum_{i=1}^{n} V_{mnt}^2 \right)
\]

By equation (9), since \(n^{-1} \sum_{t=1}^{n} V_{mnt}^2\) converges weakly to \(V^{(2)}_D(x)\), the expectation in the last line above is \(O(m/n)\). Accordingly, the fact that \(m/n \rightarrow 0\) along with Assumptions 5 (g) ensures that

\[
n^{-1} \sum_{i=1}^{n} V_{mn} e_i = O \left( m(nh)^{-1} \right) \rightarrow 0
\]

Again, invoking equation (9) and the conclusion above, it follows that

\[
F_{m,n,j,1}(z) = \mathbb{E} \left( K_h(Z_e - z) Z_j^{t_e,1} X_i \right) n^{-1} \sum_{t=1}^{n} V_{mnt} + n^{-1} \sum_{i=1}^{n} V_{mn} e_i \\
= f_z(z) M_j(z) \mu_{j,1} (K) V^{(1)}_D + o_p(1) \quad (19)
\]
Similar reasoning also produces

\[ F_{m,n,j,2}(z) = f_z(z)\mu_j \mod 3(K)V_D^{(2)} + o_p(1) \]  

(20)

Consider now the case when \( m = c \) and \( X_{t2} \) is a nearly integrated process. In this case, let \( V_{\text{cont}} = n^{-1/2}X_{t2} \) and note that \( V_{\text{cont}} = V_{\text{m}}(r) \) where \( V_{\text{m}}(r) \) was also defined in equation (5a). In this case, equation (10a) implies that \( n^{-1} \sum_{i=1}^{n} V_{\text{cont}}^{(2)} \) converges weakly to \( V_{\eta,c}^{(2)}(x) \) and it is not difficult to show that

\[ n^{-1} \sum_{i=1}^{n} V_{\text{cont}} e_i = O((nh)^{-1}) \longrightarrow 0 \]

Accordingly, it follows from the above that

\[ F_{n,j,1}(z) = f_z(z)M_1(z)\mu_j \mod 3(K)V_{\eta,c}^{(1)} + o_p(1) \]  

(21a)

\[ F_{n,j,2}(z) = f_z(z)\mu_j \mod 3(K)V_{\eta,c}^{(2)} + o_p(1) \]  

(21b)

Similar reasoning was used in Cai et al. (2009) to derive \( F_{n,j,3} \) and \( F_{n,j,5} \) where

\[ F_{n,j,3}(z) = f_z(z)M_1(z)\mu_j \mod 3(K)V_{\omega,0}^{(1)} + o_p(1) \]  

(22a)

\[ F_{n,j,5}(z) = f_z(z)\mu_j \mod 3(K)V_{\omega,0}^{(2)} + o_p(1) \]  

(22b)

What remains to be shown is the limiting form of \( F_{n,j,4} \). To do so, define \( \tilde{e}_t = K_h(Z_t - z)Z_{t,z,h}^{\mod 3} - E(K_h(Z_t - z)Z_{t,z,h}^{\mod 3}) \) and \( V_{\text{cont}} = n^{-1/2}X_{t2}n^{-1/2}X_{t2}^{\top} \) and note that for \( r \in [0,1] \), the results in Phillips (2009) demonstrate that \( V_{\text{cont}} \) converges to \( V_{\eta,c,\omega,0} \) where the latter is defined in equation (12). Again, invoking the methodology used above, it can readily be shown that \( n^{-1} \sum_{i=1}^{n} V_{\text{cont}} \tilde{e}_i \longrightarrow 0 \) and therefore

\[ F_{n,j,4}(z) = f_z(z)\mu_j \mod 3(K)V_{\eta,c,\omega,0} + o_p(1) \]  

(23)

Noting that \( \mu_0(K) = 1 \) and \( \mu_1(K) = 0 \) and plugging equations (18) to (20) into \( S_{m,n,j}(z) \) and equations (18), (21a), (21b), (22a), (22b) and (23) into \( S_{n,j}(z) \) then yields

\[ S_{m,n}(z) = f_z(z) \begin{pmatrix} 1 & 0 \\ 0 & \mu_2(K) \end{pmatrix} \otimes S_W(z) + o_p(1) \]  

(24a)

\[ S_n(z) = f_z(z) \begin{pmatrix} 1 & 0 & 0 \\ 0 & \mu_2(K) & 0 \\ 0 & 0 & \mu_2(K) \end{pmatrix} \otimes S_N(z) + o_p(1) \]  

(24b)

Denote next by \( R_{m,n}(z)^{-1} \) and \( R_{n}(z)^{-1} \) the \( d \times d \) submatrix in the upper-left corner of \( S_{m,n}(z)^{-1} \) and \( S_n(z)^{-1} \) respectively. In fact, it follows from equations (24a) and (24b) that

\[ R_{m,n}(z)^{-1} = f_z(z)^{-1}S_W(z)^{-1} + o_p(1) \]  

(25a)

\[ R_{n}(z)^{-1} = f_z(z)^{-1}S_N(z)^{-1} + o_p(1) \]  

(25b)
Moreover, from equations (16a) and (16b) it follows that

\[
D_{m,n} \left( \hat{\beta}(z) - \beta(z) \right) \equiv E_{m,n,1} + E_{m,n,2} \quad (26a)
\]

\[
D_n \left( \hat{\beta}(z) - \beta(z) \right) \equiv E_{n,1} + E_{n,2} \quad (26b)
\]

where

\[
E_{m,n,1} = R_{m,n}(z)^{-1} B_{m,n}(z) \quad (27a)
\]

\[
E_{m,n,2} = n^{-1} \sum_{t=1}^{n} K_h(Z_t - z) \epsilon_t D_{m,n}^{-1} X_t \quad (27b)
\]

\[
E_{n,1} = R_n(z)^{-1} B_n(z) \quad (27c)
\]

\[
E_{n,2} = n^{-1} \sum_{t=1}^{n} K_h(Z_t - z) \epsilon_t D_{n}^{-1} X_t \quad (27d)
\]

and

\[
B_{m,n}(z) = n^{-1} \sum_{t=1}^{n} K_h(Z_t - z) X_t^2 D_{m,n}^{-1}
\]

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

\[
B_n(z) = n^{-1} \sum_{t=1}^{n} K_h(Z_t - z) X_t^2 D_{n}^{-1}
\]

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

Similar to \( S_{m,n}(z) \) and \( S_n(z) \), here \( B_{m,n}(z) \) and \( B_n(z) \) admit the following partitions

\[
B_{m,n}(z) = \begin{pmatrix}
G_{n,0}(z) + G_{m,n,1}(z) \\
G_{m,n,2}(z) + G_{m,n,3}(z)
\end{pmatrix}
\]

\[
B_n(z) = \begin{pmatrix}
G_{n,0}(z) + G_{n,1}(z) + G_{n,2}(z) \\
G_{n,3}(z) + G_{n,4}(z) + G_{n,5}(z) \\
G_{n,6}(z) + G_{n,7}(z) + G_{n,8}(z)
\end{pmatrix}
\]

where
\[ G_{n,0}(z) = n^{-1} \sum_{t=1}^{n} K_h(Z_t - z)X_{i1}^{\otimes 2} \times \left( \beta_1(Z_t) - \beta_1(z) - (Z_t - z)\beta_1^{(1)}(z) \right) \]

\[ G_{m,n,1}(z) = n^{-1} \sum_{t=1}^{n} K_h(Z_t - z)X_{i1}m^{1/2}n^{-1/2}X_{i2}^T \times \left( \beta_2(Z_t) - \beta_2(z) - (Z_t - z)\beta_2^{(1)}(z) \right) \]

\[ G_{m,n,2}(z) = n^{-1} \sum_{t=1}^{n} m^{-1/2}n^{1/2}K_h(Z_t - z)m^{1/2}n^{-1/2}X_{i2}X_{i1}^T \times \left( \beta_1(Z_t) - \beta_1(z) - (Z_t - z)\beta_1^{(1)}(z) \right) \]

\[ G_{m,n,3}(z) = n^{-1} \sum_{t=1}^{n} m^{-1/2}n^{1/2}K_h(Z_t - z)mn^{-1}X_{i2}^{\otimes 2} \times \left( \beta_2(Z_t) - \beta_2(z) - (Z_t - z)\beta_2^{(1)}(z) \right) \]

and similarly

\[ G_{n,1}(z) = n^{-1} \sum_{t=1}^{n} K_h(Z_t - z)X_{i1}n^{-1/2}X_{i2}^T \times \left( \beta_2(Z_t) - \beta_2(z) - (Z_t - z)\beta_2^{(1)}(z) \right) \]

\[ G_{n,2}(z) = n^{-1} \sum_{t=1}^{n} K_h(Z_t - z)X_{i1}n^{-1/2}X_{i3}^T \times \left( \beta_3(Z_t) - \beta_3(z) - (Z_t - z)\beta_3^{(1)}(z) \right) \]

\[ G_{n,3}(z) = n^{-1} \sum_{t=1}^{n} n^{1/2}K_h(Z_t - z)n^{-1/2}X_{i2}X_{i1}^T \times \left( \beta_1(Z_t) - \beta_1(z) - (Z_t - z)\beta_1^{(1)}(z) \right) \]

\[ G_{n,4}(z) = n^{-1} \sum_{t=1}^{n} n^{1/2}K_h(Z_t - z)n^{-1}X_{i2}^{\otimes 2} \times \left( \beta_2(Z_t) - \beta_2(z) - (Z_t - z)\beta_2^{(1)}(z) \right) \]

\[ G_{n,5}(z) = n^{-1} \sum_{t=1}^{n} n^{1/2}K_h(Z_t - z)n^{-1/2}X_{i2}n^{-1/2}X_{i3}^T \times \left( \beta_3(Z_t) - \beta_3(z) - (Z_t - z)\beta_3^{(1)}(z) \right) \]
\[ G_{n,6}(z) = n^{-1} \sum_{t=1}^{n} n^{1/2} K_h(Z_t - z) n^{-1/2} X_t \beta_1^T \]
\[ \times \left( \beta_1(Z_t) - \beta_1(z) - (Z_t - z) \beta_1^{(1)}(z) \right) \]
\[ G_{n,7}(z) = n^{-1} \sum_{t=1}^{n} n^{1/2} K_h(Z_t - z) n^{-1/2} X_t n^{-1/2} X_t^T \]
\[ \times \left( \beta_2(Z_t) - \beta_2(z) - (Z_t - z) \beta_2^{(1)}(z) \right) \]
\[ G_{n,8}(z) = n^{-1} \sum_{t=1}^{n} n^{1/2} K_h(Z_t - z) n^{-1} X_t^T \beta_3 \]
\[ \times \left( \beta_3(Z_t) - \beta_3(z) - (Z_t - z) \beta_3^{(1)}(z) \right) \]

Making use of Taylor’s expansion, similar arguments which yield equations (18) to (20) can also be used here to show that

\[ E \left(G_{n,0} \right) = h^2 f_x(z) M_2(z) \left( \frac{1}{2} \mu_2(K) \beta_1^{(2)}(z) \right) (1 + o(1)) \]
\[ \text{Var} \left(G_{n,0} \right) = o(1) \] (29)

where \( \text{Var} \) \( (.) \) is the variance operator. Accordingly,

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

and it can be shown that

\[ G_{m,n,1} = h^2 f_x(z) M_1(z) V_D^{(1)} \left( \frac{1}{2} \mu_2(K) \beta_2^{(2)}(z) \right) (1 + o_p(1)) \]
\[ G_{m,n,2} = h^2 f_x(z) V_D^{(1)} M_1(z) \left( \frac{1}{2} \mu_2(K) \beta_1^{(2)}(z) \right) n^{-1/2} \left( \frac{1}{2} \mu_2(K) \beta_1^{(2)}(z) \right) (1 + o_p(1)) \]
\[ G_{m,n,3} = h^2 f_x(z) V_D^{(2)} M_1(z) \left( \frac{1}{2} \mu_2(K) \beta_2^{(2)}(z) \right) n^{-1/2} \left( \frac{1}{2} \mu_2(K) \beta_2^{(2)}(z) \right) (1 + o_p(1)) \]

Similarly, it can also be demonstrated that

\[ G_{n,1} = h^2 f_x(z) M_1(z) V_{\eta,c}^{(1)} \left( \frac{1}{2} \mu_2(K) \beta_2^{(2)}(z) \right) (1 + o_p(1)) \]
\[ G_{n,2} = h^2 f_x(z) M_1(z) V_{\eta,c}^{(1)} \left( \frac{1}{2} \mu_2(K) \beta_2^{(2)}(z) \right) (1 + o_p(1)) \]
\[ G_{n,3} = h^2 f_x(z) V_{\eta,c}^{(1)} M_1(z) \left( \frac{1}{2} \mu_2(K) \beta_1^{(2)}(z) \right) n^{1/2} \left( \frac{1}{2} \mu_2(K) \beta_1^{(2)}(z) \right) (1 + o_p(1)) \]
\[ G_{n,4} = h^2 f_x(z) V_{\eta,c}^{(2)} n^{1/2} \left( \frac{1}{2} \mu_2(K) \beta_2^{(2)}(z) \right) (1 + o_p(1)) \]
\begin{align*}
G_{n,5} &= h^2 f_z(z) V_{n,c,\omega,0} n^{1/2} \left( \frac{1}{2} h_3(K) \beta_3^{(2)}(z) \right) (1 + o_p(1)) \\
G_{n,6} &= h^2 f_z(z) V_{n,\omega,0}^T M_1(z) n^{1/2} \left( \frac{1}{2} h_2(K) \beta_1^{(2)}(z) \right) (1 + o_p(1)) \\
G_{n,7} &= h^2 f_z(z) V_{n,c,\omega,0}^T n^{1/2} \left( \frac{1}{2} h_2(K) \beta_2^{(2)}(z) \right) (1 + o_p(1)) \\
G_{n,8} &= h^2 f_z(z) V_{n,\omega,0}^T n^{1/2} \left( \frac{1}{2} h_3(K) \beta_3^{(2)}(z) \right) (1 + o_p(1))
\end{align*}

Now, inserting $G_{m,n,i}$ and $G_{n,i}$ for $i = 1, \ldots, 6$ back into $B_{m,n}(z)$ and $B_n(z)$ respectively, implies that

\begin{align*}
B_{m,n}(z) &= h^2 f_z(z) SW(z) D_{m,n} \left( \frac{1}{2} h_2(K) \beta^{(2)}(z) \right) (1 + o_p(1)) \quad (30a) \\
B_n(z) &= h^2 f_z(z) SN(z) D_n \left( \frac{1}{2} h_2(K) \beta^{(2)}(z) \right) (1 + o_p(1)) \quad (30b)
\end{align*}

Moreover, noting equations (25a) and (25b) and inserting the above into equations (27a) and (27c) implies that

\begin{align*}
D^{-1}_{m,n} E_{m,n,1} &= h^2 B_\beta(z) + o_p(h^2) \\
D^{-1}_{n} E_{n,1} &= h^2 B_\beta(z) + o_p(h^2)
\end{align*}

Next, consider equations (26a) and (26b) and note that

\begin{align*}
E_{m,n,2} &= D_{m,n} \left( \hat{\beta}(z) - \beta(z) - D^{-1}_{m,n} E_{m,n,1} \right) \\
&= D_{m,n} \left( \hat{\beta}(z) - \beta(z) - h^2 B_\beta(z) + o_p(h^2) \right) \\
E_{n,2} &= D_n \left( \hat{\beta}(z) - \beta(z) - D^{-1}_n E_{n,1} \right) \\
&= D_n \left( \hat{\beta}(z) - \beta(z) - h^2 B_\beta(z) + o_p(h^2) \right)
\end{align*}

In this regard, define

\begin{align*}
T_{m,n}(z) &= \begin{pmatrix} T_{m,n,1}(z) \\ T_{m,n,2}(z) \end{pmatrix} = \sqrt{\frac{h}{n}} \sum_{t=1}^{n} K_h(Z_t - z) \epsilon_t D^{-1}_{m,n} X_t \\
T_n(z) &= \begin{pmatrix} T_{n,1}(z) \\ T_{n,2}(z) \end{pmatrix} = \sqrt{\frac{h}{n}} \sum_{t=1}^{n} K_h(Z_t - z) \epsilon_t D^{-1}_n X_t
\end{align*}

where
\[ T_{m,n,1} = T_{n,1} \]
\[ = \sqrt{\frac{m}{n}} \sum_{t=1}^{n} K_h(Z_t - z) \epsilon_t X_t \]
\[ T_{m,n,2} = m^{1/2} T_{n,2} \]
\[ = \sqrt{\frac{m}{n}} \sum_{t=1}^{n} K_h(Z_t - z) \epsilon_t m^{1/2} n^{-1/2} X_t \]
\[ T_{n,3} = \sqrt{\frac{h}{n}} \sum_{t=1}^{n} K_h(Z_t - z) \epsilon_t n^{-1/2} X_t \]

and note that
\[ \sqrt{n h D_{m,n}(\hat{\beta}(z) - \beta(z) - h^2 B_\beta(z) + o_p(h^2))} = R_{m,n}(z) \]
\[ \sqrt{n h D_{n}(\hat{\beta}(z) - \beta(z) - h^2 B_\beta(z) + o_p(h^2))} = R_n(z) \]

Asymptotic normality of equations (31a) and (31b) can now be proven by establishing asymptotic normality of \( T_{m,n}(z) \) and \( T_n(z) \). Since \( T_{n,1} = T_{m,n,1} \) contains only stationary variables, it follows from Cai et al. (2000) that
\[ T_{m,n,1}(z) = T_{n,1}(z) \xrightarrow{d} N \left( 0, \sigma_n^2 \nu_0(K) f_z(z) M_2(z) \right) \]
\[ = \sqrt{\nu_0(K) f_z(z)} W_{\epsilon,1}(1) \] (32)

where \( W_{\epsilon,1}(1) \) is a \( d_1 \)-dimensional Brownian motion on \([0,1]\) with covariance matrix \( \sigma_n^2 M_2(z) \). Moreover, since the first element of \( X_{t1} \) is unity, it follows immediately that
\[ \sqrt{\frac{h}{n}} \sum_{t=1}^{n} K_h(Z_t - z) \epsilon_t \xrightarrow{d} N \left( 0, \sigma^2 \nu_0(K) f_z(z) \right) \]
\[ = \sqrt{\nu_0(K) f_z(z)} W_{\epsilon,1}(1) \]

where \( W_{\epsilon,1}(r) \) is the first element of \( W_{\epsilon}(r) \). Moreover, by equation (8b) it now follows that
\[ T_{m,n,2} \xrightarrow{d} \sqrt{\nu_0(K) f_z(z)} \left( \int_{-\infty}^{\infty} x^2 D(x) dx \right)^{1/2} W_{\epsilon,1}(1) \] (33)

Now, putting equations (32) and (33) together implies that
\[ T_{m,n}(z) \xrightarrow{d} \sqrt{\nu_0(K) f_z(z)} \left( \int_{-\infty}^{\infty} x^2 D(x) dx \right)^{1/2} W_{\epsilon,1}(1) \]

It is now clear that \( T_{m,n}(z) \) has a multivariate normal distribution with the conditional covariance matrix
\[ \begin{pmatrix} W_{\epsilon,1}(1) \\ \left( \int_{-\infty}^{\infty} x^2 D(x) dx \right)^{1/2} W_{\epsilon,1}(1) \end{pmatrix} \] being
\[ \sigma^2 \begin{pmatrix}
M_2(z) & M_1(z)V_{(2)}^{(1)\top} \\
V_{(2)}^{(1)}M_1(z) & V_{(2)}^{(1)}M_1(z)^\top
\end{pmatrix} = \sigma^2 S_W(z) \]

On the other hand, invoking Lemma A.1 of Cai and Wang (2008) now implies that

\[ T_{n,2} \longrightarrow d \sqrt{\nu_0(K)} f_z(z) \int_{-\infty}^{\infty} V_{e,c}(r) dW_{e,1}(r) \] (34a)
\[ T_{n,3} \longrightarrow d \sqrt{\nu_0(K)} f_z(z) \int_{-\infty}^{\infty} V_{e,0}(r) dW_{e,1}(r) \] (34b)

Putting equations (32), (34a) and (34b) together implies therefore that

\[ T_n(z) \longrightarrow d \sqrt{\nu_0(K)} f_z(z) \begin{pmatrix}
W_e(1) \\
\int_{-\infty}^{\infty} V_{e,c}(r) dW_{e,1}(r) \\
\int_{-\infty}^{\infty} V_{e,0}(r) dW_{e,1}(r)
\end{pmatrix} \]

Since \( W_e(\cdot), V_{e,c}(\cdot) \) and \( V_{e,0}(\cdot) \) are mutually uncorrelated, it follows that \( T_n(z) \) has a mixed normal distribution with the conditional covariance matrix of \( \begin{pmatrix}
\int_{-\infty}^{\infty} V_{e,c}(r) dW_{e,1}(r) \\
\int_{-\infty}^{\infty} V_{e,0}(r) dW_{e,1}(r)
\end{pmatrix} \) given by

\[ \sigma^2 \begin{pmatrix}
A_2(z) & A_1(z)V_{(1)}^{(1)\top} \\
V_{(1)}^{(1)}A_1(z)^\top & V_{(1)}^{(1)}A_1(z)^\top
\end{pmatrix} \]

\[ = \sigma^2 S_N(z) \]

Finally, invoking Slutsky's theorem implies that

\[ \sqrt{nh} D_{m.n} \left( \hat{\beta}(z) - \beta(z) - h^2 B_{\hat{\beta}}(z) + o_p(h^2) \right) \]
\[ \longrightarrow d f_z^{-1/2}(z) \nu_0^{1/2}(K) S_W^{-1}(z) \begin{pmatrix}
W_e(1) \\
\int_{-\infty}^{\infty} x^2 D(x) dx \end{pmatrix}^{1/2} W_{e,1}(1) \]

and

\[ \sqrt{nh} D_n \left( \hat{\beta}(z) - \beta(z) - h^2 B_{\hat{\beta}}(z) + o_p(h^2) \right) \]
\[ \longrightarrow d f_z^{-1/2}(z) \nu_0^{1/2}(K) S_N^{-1}(z) \begin{pmatrix}
W_e(1) \\
\int_{-\infty}^{\infty} V_{e,c}(r) dW_{e,1}(r) \\
\int_{-\infty}^{\infty} V_{e,0}(r) dW_{e,1}(r)
\end{pmatrix} \]

Demonstrating that \( \Sigma_\beta(z) \) takes the form specified in Theorem 1 should be clear. This completes the proof. \( \square \)
Appendix B

The proof of Theorem 2 relies on several auxiliary results contained in the lemmas below. In this regard define \( K_{j,h}(v) = \left( \frac{\sigma}{n} \right)^2 h^{-1} K \left( \frac{\sigma}{n} \right) \), where \( K_{j,h}(\cdot) \) is continuous with compact support. Moreover, redefine \( S_n(z) \) as

\[
S_n(z) = \frac{1}{n^{3/2}} \sum_{t=1}^{n} K_h(Z_t - z) \left( \frac{1}{Z_{t,z,h}} \right) \otimes X_t^{\otimes 2}
\]

where as before \( S_{n,j} = n^{-3/2} \sum_{t=1}^{n} K_{j,h}(Z_t - z) X_t^{\otimes 2} \) for \( j = 0, 1, 2 \). Note further that \( S_{n,j} \) can be expressed as

\[
S_{n,j} = \frac{c_n}{n} \sum_{t=1}^{n} X_t^{\otimes 2} K_j \left( c_n (Z_t - z_n) \right)
\]

where \( c_n = n^{1/2} h^{-1}, X_{t,n} = n^{-1/2} X_t, Z_{t,n} = n^{-1/2} Z_t \), and \( z_n = n^{-1/2} z \). Observe here that \( z_n \rightarrow 0 \) for any fixed \( z \) while \( z_n = b \) if \( z = n^{1/2} b \) for any constant \( b \). The following set of results lead to the proof of Theorem 2.

**Lemma B1.** Let \( \{\vartheta_{nk}\}, \{\vartheta_{nk}^*\}, \) and \( \{\vartheta_{nk}^{**}\} \) be sequences of random variables. Let \( u_{k,n} \) be a process defined as

\[
u_{k,n} = f_n(\vartheta_{nk-1}, \ldots, \vartheta_{nk}; \vartheta_{nk-1}^*, \ldots, \vartheta_{nk}^*; \vartheta_{nk}^{**}, \ldots, \vartheta_{nk}^{**})
\]

where \( f_n(\cdot, \cdot, \cdot) \) is a real function of its components. Furthermore, let \( \{\mathcal{F}_{nk} : 1 \leq k \leq n\} \) be a sequence of increasing \( \sigma \)-fields such that \( \{\vartheta_{nk}, \mathcal{F}_{nk} : 1 \leq k \leq n\} \) is a martingale difference sequence and \( u_{k,n} \) is adapted to \( \mathcal{F}_{nk} \) for all \( 1 \leq k \leq n \) and \( n \geq 1 \).

(a) Let \( \{\vartheta_{nk+1}^*, \vartheta_{nk+1}, \mathcal{F}_{nk} : 1 \leq k \leq n\} \) be a martingale difference sequence where \( \{\vartheta_{nk}\} \) and \( \{\vartheta_{nk}^*\} \) satisfy the following conditions as \( n,m \rightarrow \infty \)

\[
\max_{m \leq k \leq n} \mathbb{E} \left( \vartheta_{nk+1}^* \mid \mathcal{F}_{nk} \right) - \sigma_{\vartheta}^2 \rightarrow 0 \quad \text{a.s.}
\]

\[
\max_{m \leq k \leq n} \mathbb{E} \left( \vartheta_{nk+1} \mid \mathcal{F}_{nk} \right) - \sigma_{\vartheta}^2 \rightarrow 0 \quad \text{a.s.}
\]

for some \( \sigma_{\vartheta}^2 > 0 \) and \( \sigma_{\vartheta}^2 > 0 \), and for some \( \delta > 0 \)

\[
\max_{m \leq k \leq n} \mathbb{E} \left( \left| \vartheta_{nk+1}^* \right|^{2+\delta} \mid \mathcal{F}_{nk} \right) + \mathbb{E} \left( \left| \vartheta_{nk+1}^* \right|^{2+\delta} \mid \mathcal{F}_{nk} \right) \leq c_{\vartheta} \quad \text{a.s.}
\]

(b) Let \( \{\vartheta_{nk}^* : j \geq 1\} \) be \( \mathcal{F}_{n,1} \)-measurable for each \( n \geq 1 \), and there exists a sequence of positive constants \( d_n \rightarrow \infty \) and a Gaussian process \( V_{\vartheta^*}(r) \) such that \( d_n^{-1} \sum_{j=1}^{n} \vartheta_{nk}^{**} \rightarrow d V_{\vartheta^*}(r) \) on \( D[0, \infty) \). Moreover, \( V_{\vartheta^*}(r) \) is assumed to be independent of \( V_{\vartheta^*}(r) \) where the latter is the weak limit \( n^{-1/2} \sum_{j=1}^{n} \vartheta_{nk+1} \rightarrow d V_{\vartheta^*}(r) \) on \( D[0, 1] \).

(c) Let \( \max_{1 \leq k \leq n} \left| u_{k,n} \right| = o_p(1) \) and \( n^{-1/2} \sum_{k=1}^{n} \left| u_{k,n} \right| \mathbb{E} \left( \left| \vartheta_{nk+1} \vartheta_{nk+1} \mid \mathcal{F}_{nk} \right| \right) = o_p(1) \).
(d) There exists a random variable $T(\theta^*, \theta^{**}) > 0$ such that $T_n^2 = \sum_{k=1}^{n} u_{k,n}^2 \to_d T^2(\theta^*, \theta^{**})$ as $n \to \infty$.

Then, it follows that $T_n^{-1} \sum_{k=1}^{n} u_{k,n} \vartheta_{n,k+1} \to N(0,1)$.

**Proof of Lemma B1.** The proof follows directly from an extension of the martingale CLT of Hall and Heyde (1980) in Theorem 2.1 of Wang (2011).

**Lemma B2.** Let $X_t$ and $Z_t$ be defined as in equations (14) and (15) and suppose Assumptions 2, 6 and 7 hold. For all $0 < c_x \leq n$ and $0 \leq c_z \leq n$, $j = 0, 1, 2$, $r \in [0, 1]$, and $c_n = \sqrt{n}h^{-1} \to \infty$ such that $c_n n^{-1} \to 0$ if $z$ is fixed,

$$
c_n n^{-1} \sum_{t=1}^{n} X_{t,n} X_{1,n}^\top K_j(c_n (Z_{t,n} - z_n)) \to_p \mu_j(K) \int_{0}^{1} V_{q,c_x} (r) V_{q,c_x} (r)^\top dL_{V_{q,c_x}} (1, 0)$$

$$
c_n n^{-1} \sum_{t=1}^{n} X_{t,n} X_{1,n}^\top K_j(c_n (Z_{t,n} - z_n)) \to_p \nu_j(K) \int_{0}^{1} V_{q,c_x} (r) V_{q,c_x} (r)^\top dL_{V_{q,c_x}} (1, 0)$$

When $z = n^{1/2}b$ the result continues to hold with $L_{V_{q,c_x}} (1, 0)$ replaced with $L_{V_{q,c_x}} (1, b)$.

**Proof of Lemma B2.** The results ensue immediately from Remark (b) of Theorem 1 of Phillips (2009) by noting that $g(x) = xx^\top$ is locally integrable and that Assumptions 6 and 7 satisfy Assumptions 2.2 - 2.4 of Phillips (2009). A similar result exists in Gao and Phillips (2013). Note further that when $c_z = 0$, $L_{V_{q,c_x}} (1, \cdot)$ is the local time of $Z_t$ when it’s a pure $I(1)$ process.

**Lemma B3.** Let $X_t$ and $Z_t$ be defined as in equations (14) and (15) and suppose Assumptions 2, 6 and 7 hold. For all $0 < c_x \leq n$ and $0 \leq c_z \leq n$, $j = 0, 1, 2$, $r \in [0, 1]$, and $c_n = \sqrt{n}h^{-1} \to \infty$ such that $c_n n^{-1} \to 0$ if $z$ is fixed,

$$
\sigma_c^2 \frac{h}{n^{3/2}} \sum_{t=1}^{n} X_{t} X_{t}^\top K_{j,h}^2 (Z_t - z) \to_d \sigma_c^2 \nu_j(K) \int_{0}^{1} V_{q,c_x} V_{q,c_x}^\top dL_{V_{q,c_x}} (1, 0)
$$

**Proof of Lemma B3.** The proof is inspired by results in Phillips (2009), Sun et al. (2013) and Gao and Phillips (2013). Let $p_t (\cdot | X)$ denote the conditional density of $Z_t$ given $X_t = X$ and define $q_t (\cdot | \cdot)$ as the conditional density of $\frac{X_t}{\sqrt{t}}$ given $\frac{X}{\sqrt{t}}$. Note further that a change of variables argument implies that $p_t (Z | X) = t^{-1/2} q_t \left( \frac{Z}{\sqrt{t}} \left| \frac{X}{\sqrt{t}} \right. \right)$. Next, recall that $X_t$ has dimension $d \times 1$, let $J = (J_1, \ldots, J_d)^\top$ be any vector of real numbers such that $J^\top J = 1$, and define $X_t^* = J^\top X_t$. Finally, recall the assumption that $\sigma_c^2 = E(\epsilon_t^2 | Z_t, X_t)$ and observe that
\[
\begin{align*}
\mathbb{E} \left( \frac{e_t^2}{n^3} \mid Z_t, X_t \right) &= \sum_{t=1}^{n} \mathbb{E} \left( X_t^{*4} K_{j,h}^2 (Z_t - z) \right) \\
&= \frac{\sigma_t^2}{n^3 h^2} \sum_{t=1}^{n} \mathbb{E} \left( X_t^{*4} K_j^2 \left( \frac{Z_t - z}{h} \right) \right) \\
&= \frac{\sigma_t^2}{n^3 h^2} \sum_{t=1}^{n} \mathbb{E} \left( X_t^{*} K_j \left( \frac{Z_t - z}{h} \right) \mid X_t \right) \\
&= \frac{\sigma_t^2}{n^3 h^2} \sum_{t=1}^{n} \mathbb{E} \left( X_t^{*4} \int_{-\infty}^{\infty} K_j^2 (y) q_t \left( \frac{y h + z}{\sqrt{t}} \right) \mid X_t \right) \frac{h}{\sqrt{t}} dy \\
&\leq \frac{\sigma_t^2}{n^3 h^2} \sum_{t=1}^{n} \frac{1}{\sqrt{t}} = O \left( \frac{1}{n^{1/2}} \right) = o(1) \quad (35)
\end{align*}
\]

Equation (35) follows since \( \mathbb{E} \left( X_t^{*4} \right) = O(1) \), the integral expression is bounded and \( \sqrt{n} h \rightarrow \infty \). Next, let \( p_{st} (\cdot \mid Z_s, X_s, X_t) \) denote the conditional density of \( Z_t - Z_s \) given \( Z_s, X_s, X_t \). Moreover, if \( q_{st} (\cdot \mid Z_s, X_s, X_t) \) denotes the conditional density of \( \frac{Z_t - Z_s}{\sqrt{t}} \) given \( Z_s, X_s, X_t \), then \( p_{st} (Z \mid Z_s, X_s, X_t) = (t-s)^{-1/2} q_{st} \left( \frac{Z_t - Z_s}{\sqrt{t-s}} \mid Z_s, X_s, X_t \right) \).

Similar reasoning to equation (35) now yields the covariance result below.

\[
\begin{align*}
\mathbb{E} \left( \frac{e_t^2}{n^3 h^2} \mid Z_t, X_t \right) &= \sum_{t=2}^{n} \sum_{s=1}^{t-1} \mathbb{E} \left( X_s^{*2} X_t^{*2} K_j \left( \frac{Z_s - z}{h} \right) K_j \left( \frac{Z_t - z}{h} \right) \right) \\
&= \frac{\sigma_t^2}{n^3 h^2} \sum_{t=2}^{n} \sum_{s=1}^{t-1} \mathbb{E} \left( X_s^{*2} X_t^{*2} K_j \left( \frac{Z_s - z}{h} \right) K_j \left( \frac{Z_t - z}{h} \right) \mid X_s, X_t \right) \\
&= \frac{\sigma_t^2}{n^3 h^2} \sum_{t=2}^{n} \sum_{s=1}^{t-1} \mathbb{E} \left( X_s^{*2} X_t^{*2} \int_{-\infty}^{\infty} K_j^2 (y) q_t \left( \frac{y h + z}{\sqrt{t}} \right) \mid X_s, X_t \right) \frac{h}{\sqrt{t}} dy \\
&\leq \frac{\sigma_t^2}{n^3} \sum_{t=2}^{n} \sum_{s=1}^{t-1} \frac{1}{\sqrt{t}} = O \left( \frac{1}{n^{3/2}} \right) = o(1) \quad (36)
\end{align*}
\]

Equations (35) and (36) now imply that

\[
\begin{align*}
\mathbb{E} \left( \frac{e_t^2}{n^3 h^2} \mid Z_t, X_t \right) &= \mathbb{E} \left( \sum_{t=1}^{n} X_t^{*} K_j \left( \frac{Z_t - z}{h} \right) \right)^2 \\
&= \frac{\sigma_t^2}{n^3 h^2} \sum_{t=1}^{n} \mathbb{E} \left( X_t^{*4} K_j^2 \left( \frac{Z_t - z}{h} \right) \right) \\
&\leq \frac{\sigma_t^2}{n^3 h^2} \sum_{t=2}^{n} \sum_{s=1}^{t-1} \mathbb{E} \left( X_s^{*2} X_t^{*2} K_j \left( \frac{Z_s - z}{h} \right) K_j \left( \frac{Z_t - z}{h} \right) \right) \\
&= o(1) \quad (37)
\end{align*}
\]
and Lemma B2 ensures that

$$\frac{\sigma^2}{n^{3/2}h} \sum_{t=1}^{n} X_t X_t' K_j^2 \left( \frac{Z_t - z}{h} \right) \rightarrow_d \sigma^2 \nu_j(K) \int_0^1 V_{\eta,c_x} V_{\eta,c_x}' dL_{V_{\eta,c_x}}(1,0)$$

(38)

$$\frac{\sigma^2}{n^{3/2}h} \sum_{t=1}^{n} X_t X_t' K_j^2 \left( \frac{Z_t - z}{h} \right) \rightarrow_d \sigma^2 \nu_j(K) \int_0^1 V_{\eta,c_x} V_{\eta,c_x}' dL_{V_{\eta,c_x}}(1,0)$$

(39)

where $V_{\eta,c_x}'$ is defined in the same way as $V_{\eta,c_x}$ when $\eta_t$ is replaced with $J^T \eta_t$. The result follows by noting that $K_{j,h}(v) = h^{-1} K_j(v/h)$.

Lemma B4. Let $X_t$ and $Z_t$ be defined as in equations (14) and (15) and suppose Assumptions 2, 6 and 7 hold. For all $0 < c_x \leq n$ and $0 \leq c_z \leq n$, $j = 0, 1, 2$, $r \in [0,1]$, and $c_n = \sqrt{nh^{-1}} \rightarrow \infty$ such that $c_n n^{-1} \rightarrow 0$ if $z$ is fixed,

$$\frac{1}{n^{3/2}} \sum_{t=1}^{n} X_t X_t' \left( \beta (Z_t) - \beta(z) - \beta^{(1)}(z) (Z_t - z) \right) K_{j,h} (Z_t - z)$$

$$\rightarrow_d h^2 B_\beta(z) \int_0^1 V_{\eta,c_x} V_{\eta,c_x}' dL_{V_{\eta,c_x}}(1,0) + o_p(h^2)$$

Proof of Lemma B4. The limiting form $h^2 B_\beta(z) \int_0^1 V_{\eta,c_x} V_{\eta,c_x}' dL_{V_{\eta,c_x}}(1,0)$ follows directly from Lemma B2 by similar arguments found in the proof of Lemma B3 and a Taylor’s theorem application to $\beta (Z_t) - \beta(z) - \beta^{(1)}(z) (Z_t - z)$ as in Equation (29). What remains is to demonstrate that the limit is $o_p(h^2)$. To do this it suffices to show that $E \left\{ \sum_{t=1}^{n} X_t^4 \left( \beta (Z_t) - \beta(z) - \beta^{(1)}(z) (Z_t - z) \right) K_{j,h} (Z_t - z) \right\}^2$ is $O_p( h^4 )$.

$$\frac{1}{n^3} \sum_{t=1}^{n} \mathbb{E} \left( X_t^4 \left( \beta (Z_t) - \beta(z) - \beta^{(1)}(z) (Z_t - z) \right)^2 K_{j,h}^2 (Z_t - z) \right)$$

$$= \frac{1}{n^3} \sum_{t=1}^{n} \mathbb{E} \left( X_t^4 \left( \beta (Z_t) - \beta(z) - \beta^{(1)}(z) (Z_t - z) \right)^2 K_j^2 \left( \frac{Z_t - z}{h} \right) \right)$$

$$= \frac{1}{n^3} \sum_{t=1}^{n} \mathbb{E} \left( X_t^4 \int_0^\infty \left( \beta (y h + z) - \beta(z) - \beta^{(1)}(z) (y h + z) \right) K_j^2 \left( \frac{y h + z}{\sqrt{t}} \right) q_t \left( \frac{y h + z}{\sqrt{t}} \right) dy \right)$$

$$= \frac{1}{n^3} \sum_{t=1}^{n} \mathbb{E} \left( X_t^4 \int_0^\infty \left( \frac{1}{2} y^2 h^2 \beta^{(2)}(z) + o(h^2) \right) K_j^2 \left( \frac{y h + z}{\sqrt{t}} \right) q_t \left( \frac{y h + z}{\sqrt{t}} \right) dy \right)$$

$$\leq C \frac{h^3 \beta^{(2)}(z) + o(h^3)}{n^3} \sum_{t=1}^{n} \frac{1}{\sqrt{t}} \left( h^3 \beta^{(2)}(z) + o(h^3) \right) O \left( \frac{1}{n^{3/2}} \right) = o(h^3)$$

(40)

for some positive constant $C$. Turning next to the covariance result, consider the following.
\[
\frac{1}{n^3h^2} \sum_{t=2}^{n} \sum_{s=1}^{t-1} E \left\{ X_t^2 X_t^2 \left( \beta(Z_s) - \beta(z) - \beta^{(1)}(z)(Z_s - z) \right) \right. \\
\times \left( \beta(Z_t) - \beta(z) - \beta^{(1)}(z)(Z_t - z) \right) K_j \left( \frac{Z_s - z}{h} \right) K_j \left( \frac{Z_t - z}{h} \right) \right. \\
= \frac{1}{n^3h^2} \sum_{t=2}^{n} \sum_{s=1}^{t-1} E \left\{ X_t^2 X_t^2 E \left( \beta(Z_s) - \beta(z) - \beta^{(1)}(z)(Z_s - z) \right) \right. \\
\times \left( \beta(Z_t) - \beta(z) - \beta^{(1)}(z)(Z_t - z) \right) K_j \left( \frac{Z_s - z}{h} \right) K_j \left( \frac{Z_t - z}{h} \right) \left| X_s, X_t \right. \right. \\
= \frac{1}{n^3h^2} \sum_{t=2}^{n} \sum_{s=1}^{t-1} E \left\{ X_t^2 X_t^2 \left( \beta(Z_s) - \beta(z) - \beta^{(1)}(z)(Z_s - z) \right) \right. \\
\times \left( \beta(Z_t) - \beta(z) - \beta^{(1)}(z)(Z_t - z) \right) K_j \left( \frac{Z_s - z}{h} \right) K_j \left( \frac{Z_t - z}{h} \right) \left| Z_s, X_s, X_t \right. \right. \\
= \frac{1}{n^3h^2} \sum_{t=2}^{n} \sum_{s=1}^{t-1} E \left\{ X_t^2 X_t^2 \int_{-\infty}^{\infty} \left( \beta(yh + z) - \beta(z) - \beta^{(1)}(z)yh \right) K_j(y) \\
\times \left\{ \int_{-\infty}^{\infty} \left( \beta((w + y)h + z) - \beta(z) - \beta^{(1)}(z)(w + y) \right) \\
\times K_j(w + y) q_{st} \left( \frac{wh}{\sqrt{T - s}} \left| X_s, X_t \right. \right. \right. \\
\times K_j \left( \frac{Z_s - z}{h} \right) K_j \left( \frac{Z_t - z}{h} \right) \left| Z_s, X_s, X_t \right. \right. \\
= \frac{1}{n^3h^2} \sum_{t=2}^{n} \sum_{s=1}^{t-1} E \left\{ X_t^2 X_t^2 \int_{-\infty}^{\infty} \left( \frac{1}{2} \beta^{(2)}(z) + o(h^2) \right) K_j(y) \\
\times \left\{ \int_{-\infty}^{\infty} \left( \frac{1}{2} (w + y)^2 h^2 \beta^{(2)}(z) + o(h^2) \right) \\
\times K_j(w + y) q_{st} \left( \frac{wh}{\sqrt{T - s}} \left| X_s, X_t \right. \right. \right. \\
\times K_j \left( \frac{Z_s - z}{h} \right) K_j \left( \frac{Z_t - z}{h} \right) \left| Z_s, X_s, X_t \right. \right. \\
\leq C \frac{h^4 \beta^{(2)}(z)^2 + o(h^4)}{n^3} \sum_{t=2}^{n} \sum_{s=1}^{t-1} \frac{1}{\sqrt{T - s}} = \left( \frac{1}{n^3} \right) \left( \frac{1}{n^3} \right) = o(h^4) \quad (41)
\]

for some positive constant \( C \). Observe next that Equations (40) and (41) imply that

\[
\frac{1}{n^3h^2} \sum_{t=1}^{n} E \left\{ X_t^2 \left( \beta(Z_t) - \beta(z) - \beta^{(1)}(z)(Z_t - z) \right) \right. \\
\times \left( \beta(Z_t) - \beta(z) - \beta^{(1)}(z)(Z_t - z) \right) K_j \left( \frac{Z_t - z}{h} \right) \right. \\
= \frac{1}{n^3h^2} \sum_{t=1}^{n} E \left\{ X_t^2 \left( \beta(Z_t) - \beta(z) - \beta^{(1)}(z)(Z_t - z) \right)^2 \right. \\
\times \left( \beta(Z_t) - \beta(z) - \beta^{(1)}(z)(Z_t - z) \right) K_j \left( \frac{Z_t - z}{h} \right) \right. \\
+ \frac{1}{n^3h^2} \sum_{t=2}^{n} \sum_{s=1}^{t-1} E \left\{ X_t^2 X_t^2 \left( \beta(Z_s) - \beta(z) - \beta^{(1)}(z)(Z_s - z) \right) \right. \\
\times \left( \beta(Z_t) - \beta(z) - \beta^{(1)}(z)(Z_t - z) \right) K_j \left( \frac{Z_s - z}{h} \right) K_j \left( \frac{Z_t - z}{h} \right) \\
= o(h^3) + o(h^4) = o(h^4) \quad (42)
\]
This completes the proof.

**Proof of Theorem 2.** Since \( \mu_0(K) = 1 \) and \( \mu_1(K) = 0 \), note that Lemma B2 implies that

\[
S_n(z) = \begin{pmatrix} S_{n,0}(z) & S_{n,1}(z) \\ S_{n,1}(z) & S_{n,2}(z) \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & \mu_z(K) \end{pmatrix} \otimes \int_0^1 V_{\eta,cz}(r)V_{\eta,cz}(r)^T dL_{V_{\eta,cz}}(1,0)
\]

Moreover, replacing \( Y_1 \) in equation (4) by \( Y_t = X_t^T \beta(Z_t) + \epsilon_t \) further implies that

\[
\hat{\beta}(z) - \beta(z) = \left( \int_0^1 V_{\eta,cz}(r)V_{\eta,cz}(r)^T dL_{V_{\eta,cz}}(1,0) \right)^{-1} \times \left\{ \frac{1}{n^{3/2}} \sum_{t=1}^n X_t X_t^T \left( \beta(Z_t) - \beta(z) - \beta^{(1)}(z)(Z_t - z) \right) K_h(Z_t - z) + \frac{1}{n^{3/2}} \sum_{t=1}^n X_t \epsilon_t K_h(Z_t - z) \right\} = \left( \int_0^1 V_{\eta,cz}(r)V_{\eta,cz}(r)^T dL_{V_{\eta,cz}}(1,0) \right)^{-1} (B_1 + B_2)
\]

where

\[
B_1 = \frac{1}{n^{3/2}} \sum_{t=1}^n X_t X_t^T \left( \beta(Z_t) - \beta(z) - \beta^{(1)}(z)(Z_t - z) \right) K_h(Z_t - z)
\]

\[
B_2 = \frac{1}{n^{3/2}} \sum_{t=1}^n X_t \epsilon_t K_h(Z_t - z)
\]

Moreover, note that Lemma B4 and Equation (43) now imply that

\[
\sqrt{hn^{3/2}} \left( \hat{\beta}(z) - \beta(z) - \hat{h}^2 \beta_z(z) + o_p(\hat{h}^2) \right) = \sqrt{hn^{3/2}} B_2 + o_p(1)
\]

and the theorem will follow if it can be shown that \( \sqrt{hn^{3/2}} B_2 \rightarrow_d MN(\Sigma_{B_2}) \) where by Lemma B3, \( \Sigma_{B_2} = \sigma^2 \nu_0(K) \int_0^1 V_{\eta,cz}^2 dL_{V_{\eta,cz}}(1,0) \). In this regard, recall from Lemma B2 that equation (39) implies that

\[
\sigma^2 \frac{h}{n^{3/2}} \sum_{t=1}^n X_t^* X_t^* K_{0,h}(Z_t - z) \rightarrow_d \sigma^2 \nu_0(K) \int_0^1 V_{\eta,cz}^2 dL_{V_{\eta,cz}}(1,0)
\]

where \( X_t^* = J^T X_t \) and \( J^* = (J_1^*, \ldots, J_d^*)^T \) is any real vector satisfying \( J^T J^* = 1 \). Using the Cramér-Wold device, it stands to argue that

\[
\sqrt{hn^{3/2}} B_2 = \frac{\sqrt{h}}{n^{3/4}} \sum_{t=1}^n X_t^* \epsilon_t K_h(Z_t - z) \rightarrow_d MN(0, \Sigma_{B_2})
\]
where equation (46) follows by Lemma B3 and Lemma B1. To see this, observe that Lemma B1 can be invoked here using the following notation:

\[ \varphi_{n,t+1} = \epsilon_t \]
\[ \varphi^*_{nt} = J^* \varepsilon_t \]
\[ \varphi^{**}_{nt} = J^* \varepsilon_{t-1} \]
\[ u_{t,n} = \frac{1}{\sqrt{hn^{1/2}}} \frac{1}{\sqrt{n}} X^*_t K \left( \frac{Z_t - z}{h} \right) \]

where \( J^* = (J_{1}^*, \ldots, J_{d}^*) \) is a real vector satisfying \( J^* J^* = 1 \) and \( X^*_t = J^* X_t \). Moreover, let \( \mathcal{F}_{n,t} = \sigma (\epsilon_1; \ldots; \epsilon_t, \varepsilon_{t-1}; \ldots) \) be generated by \( \{(\epsilon_i, \varepsilon_j) : 1 \leq i \leq t, -\infty < j \leq t\} \). Note further that Assumptions 7 and equation (45) imply Assumptions (i), (ii), (iii), and (iv) of lemma B1, respectively. The Kolmogorov inequality now implies that

\[ P \left( \max_{1 \leq t \leq n} |u_{t,n}| > \delta_u \right) \leq \frac{1}{\sqrt{nh}} \max_{1 \leq t \leq n} \mathbb{E} \left( n^{-1} X^*_t K^2 \left( \frac{Z_t - z}{h} \right) \right) \leq \frac{C}{\sqrt{nh}} = o(1) \]  
(47)

for any small \( \delta_u > 0 \). Note that equation (47) implies that \( \max_{1 \leq t \leq n} |u_{t,n}| = o_p(1) \). Lemma B1 now implies that equation (46) holds and this completes the proof. \( \square \)