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

Mirza Trokić[a]

[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 varyinf 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

$$Y_t = \beta(Z_t)^\top X_t + \epsilon_t, \quad 1 \le t \le n \tag{1}$$

where Y_t, and ϵ_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)

(2009), Wang and Phillips (2009) and Gao and Phillips (2013).

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 Z_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 $X_t = [X_{t1}^\top, X_{t2}^\top, X_{t3}^\top]^\top$ 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 $d_3 = 0$. 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 multivariate $I(1)$ process with $d_3 \geq 1$. In either scenario, the coefficient function is conformingly 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:

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

The model further assumes ϵ_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 ϵ_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 $\beta(z) + \beta^{(1)}(z)(Z_t - z)$ for any grid point z, where $\beta^{(s)} = d^s\beta(z)/dz^s$. Furthermore, the vector of parameter estimates is defined as

$$\begin{pmatrix} \theta_0 \\ \theta_1 \end{pmatrix} = \arg\min_{\theta_0, \theta_1} \sum_{t=1}^n \left[Y_t - \theta_0^\top X_t - (Z_t - z)\, \theta_1^\top X_t\right]^2 K_h(Z_t - z) \tag{3}$$

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

$$\begin{aligned} \begin{pmatrix} \widehat{\beta}(z) \\ \widehat{\beta}^{(1)}(z) \end{pmatrix} &= \left[\sum_{t=1}^n \begin{pmatrix} X_t \\ (Z_t - z)X_t \end{pmatrix}^{\otimes 2} K_h(Z_t - z)\right]^{-1} \\ &\times \sum_{t=1}^n \begin{pmatrix} X_t \\ (Z_t - z)X_t \end{pmatrix} Y_t K_h(Z_t - z) \end{aligned} \tag{4}$$

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

2.2. Notations and Assumptions

Consider first the weakly integrated process defined as $X_{t2} = \alpha X_{t-1,2} + \eta_t$ where $\alpha = 1 - m/n$, the sample size is n, and m is a function of n satisfying $m/n \longrightarrow 0$ as $n \longrightarrow \infty$. Moreover, η_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 $\sigma_\eta^2 = \pi(1)^2 \mathbf{E} u_t^2$ and satisfying

Assumptions 1.

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

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

Note that this framework interprets X_{t2} 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 α 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 $m \longrightarrow \infty$ as $n \longrightarrow \infty$ and therefore whenever m is not a constant it is clear that α 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 α converges to unity at rate n^{-1}. When this is indeed the case, the model under consideration will assume the (possibly multivariate) form $X_{t2} = (1 - c/n)X_{t-1,2} + \eta_t$ where η_t is a mean zero stationary strong mixing process with variance Σ_η and satisfying one of the following two conditions:

Assumptions 2.

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

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

Similarly, when $\alpha = 1$ and the pure $I(1)$ process is modelled as the unit root process $X_{t3} = X_{t-1,3} + \omega_t$, assume $\{\omega_t\}$ has variance Σ_ω and satisfies one of the following:

Assumptions 3.

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

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

Finally, observe the following processes associated with X_{t2} and X_{t3}.

$$V_{mn}(r) = n^{-1/2} X_{\lfloor nr \rfloor,2}$$
$$\text{or} \tag{5a}$$
$$V_{cn}(r) = n^{-1/2} X_{\lfloor nr \rfloor,2}$$
$$V_{0n}(r) = n^{-1/2} X_{\lfloor nr \rfloor,3} \tag{5b}$$
$$V_{\eta,m}(r) = \int_0^r \exp(-m(r-s)) dV_{\eta,0}(s)$$
$$\text{or} \tag{5c}$$
$$V_{\eta,c}(r) = \int_0^r \exp(-C(r-s)) dV_{\eta,0}(s)$$
$$V_\eta(r) = \sqrt{m} V_{\eta,m}\left(\frac{r}{m}\right) \tag{5d}$$

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

$$V_{mn} \longrightarrow_d V_{\eta,m} \tag{6a}$$
$$V_{cn} \longrightarrow_d V_{\eta,c} \tag{6b}$$
$$V_{0n} \longrightarrow_d V_{\omega,0} \tag{6c}$$

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

$$\sup_{0 \le r \le \infty} \left| V_{mn}(r) - V_{\eta,m}(r) \right| = o_p\left(n^{-1/2+1/p}\right) + O_p\left(mn^{-1}\right) \tag{7a}$$
$$\sup_{0 \le r \le \infty} \left|\left| V_{cn}(r) - V_{\eta,c}(r) \right|\right| = O_{a.s.}\left(n^{-1/2+1/(2+\delta_*)} \log^{\lambda_*}(n)\right) \tag{7b}$$
$$\sup_{0 \le r \le \infty} \left|\left| V_{0n}(r) - V_{\omega,0}(r) \right|\right| = O_{a.s.}\left(n^{-1/2+1/(2+\delta_*)} \log^{\lambda_*}(n)\right) \tag{7c}$$

for large n, uniformly in m such that $m/n \longrightarrow 0$ as $n \longrightarrow \infty$, where $|\cdot|$ is the uniform norm, and $\|\cdot\|$ is the L_2 norm in $\mathbf{R}^{d_2}$ and $\mathbf{R}^{d_3}$ in equations (7b) and (7c), respectively. The particular appeal of these result is that $V_{\eta,m}$ ($V_{\eta,c}$) 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_{t2} can be derived from the asymptotics of functionals of $V_{\eta,m}$ or $V_{\eta,c}$, and similarly for X_{t3}. 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

$$L_\eta(r,x) = \lim_{\delta\longrightarrow 0} \frac{1}{2\delta}\int_0^r \mathbf{1}\left\{\left|V_\eta(s) - x\right|\right\} ds$$
$$L_\omega(r,x) = \lim_{\delta\longrightarrow 0} \frac{1}{2\delta}\int_0^r \mathbf{1}\left\{\left|V_{\omega,0}(s) - x\right|\right\} ds$$

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 $k > -1$,

$$\int_{-\infty}^{\infty} |x|^k D_m(x)dx \longrightarrow_{a.s.} \int_{-\infty}^{\infty} |x|^k D(x)dx$$

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

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

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

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

Definition 2. *A function $H : \mathbf{R} \longrightarrow \mathbf{R}$ 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. *$\left|R(x,\lambda)\right| \leq \nu(\lambda)Q(x)$ for all λ sufficiently large and for all x over any compact set where $\nu(\lambda)/\kappa(\lambda) \longrightarrow 0$ as $\lambda \longrightarrow \infty$ 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) $\mathbf{E}\left(v_t^2|\mathcal{F}_{t-1}\right) = \sigma^2$ a.s. for all t, and $\sup_t \mathbf{E}\left(\left|v_t\right|^{2+\delta}\middle|\mathcal{F}_{t-1}\right) < \infty$ 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 process 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 κ and limit homogeneous function F. In particular, the result implies that when $m \longrightarrow \infty$ as $n \longrightarrow \infty$, then,

$$\frac{1}{n}\kappa\left(\sqrt{\frac{n}{m}}\right)^{-1}\sum_{t=1}^{n} H(X_{t2}) \longrightarrow_p \int_{-\infty}^{\infty} F(x)D(x)dx \tag{8a}$$

$$\frac{1}{n}\kappa\left(\sqrt{\frac{n}{m}}\right)^{-1}\sum_{t=1}^{n} H(X_{t2})v_t \longrightarrow_d \mathbf{N}\left(0, \sigma_u^2 \int_{-\infty}^{\infty} F^2(x)D(x)dx\right) \tag{8b}$$

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 $\kappa\left(\sqrt{\frac{n}{m}}\right)^{-1} = \left(mn^{-1}\right)^{l/2}$ and, as $m, n \longrightarrow \infty$,

$$\frac{1}{n}\sum_{t=1}^{n}\left(\sqrt{mn^{-1}}X_{t2}\right)^{\otimes l} \longrightarrow_p \int_{-\infty}^{\infty} x^l D(x)dx \equiv V_D^{(l)}(x) \tag{9}$$

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 $B(\cdot) = (\cdot)^{\otimes l}$ for $l = 1, 2$, then as $n \longrightarrow \infty$,

$$\frac{1}{n}\sum_{t=1}^{n}\left(\frac{1}{\sqrt{n}}X_{\lfloor nr\rfloor 2}\right)^{\otimes l} \longrightarrow_d \int_0^1 V_{\eta,m}(s)^{\otimes l}ds \equiv V_{\eta,c}^{(l)} \tag{10a}$$

$$\frac{1}{n}\sum_{t=1}^{n}\left(\frac{1}{\sqrt{n}}X_{\lfloor nr\rfloor 3}\right)^{\otimes l} \longrightarrow_d \int_0^1 V_{\omega,0}(s)^{\otimes l}ds \equiv V_{\omega,0}^{(l)} \tag{10b}$$

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 $c_n \longrightarrow \infty$ s.t. $c_n n \longrightarrow 0$ and $r \in [0, 1]$:

$$c_n n^{-1}\sum_{t=t=1}^{\lfloor nr\rfloor} y_{t,n} g\left(c_n x_{t,n}\right) \longrightarrow_d \int_{-\infty}^{\infty} g(s)ds \int_0^r G_y(p)dL_{G_x}(p, 0) \tag{11}$$

where $g(\cdot)$ is a Lebesgue integrable function on $\mathbf{R}$ 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 $(V_{\eta,c}, V_{\omega,0})$, then

$$n^{-1}\sum_{t=1}^{\lfloor nr\rfloor} V_{cn}V_{0n} \longrightarrow_d \int_0^1 V_{\eta,c}(r)V_{\omega,0}dr \equiv V_{\eta,c,\omega,0} \tag{12}$$

The following set of assumptions are now imposed.

Assumptions 5.

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

(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) ϵ_t *has a finite fourth moment* $\mathbf{E}\left(\epsilon_t \middle| X_t, Z_t\right) = 0$ *and* $\mathbf{E}\left(\epsilon_t^2 \middle| X_t, Z_t\right) = \sigma_\epsilon^2$ *is a positive constant.*

(d) $\{(X_{t1}, Z_t, \epsilon_t, \eta_t); t \geq 1\}$ *is a strictly* α*-mixing stationary process with the* δ_1*-th moment* ($\delta_1 > 2$). $E\left(|\epsilon_t X_{t1}^2|^{\delta_2} \middle| Z_t = z\right) \leq C_1 < \infty$ *with* $\delta_2 > \delta_1$ *and* $\alpha(t) = O\left(t^{-\delta_3}\right)$ *for some* $\delta_3 > \min\{\delta_2\delta_1/(\delta_2 - \delta_1), \delta_5, 2\delta_6/(2-\delta_6)\}$, *where* $\delta_5 = \delta_4\delta_1/(\delta_4\delta_1 - \delta_1 - \delta_4)$ *for some* δ_4 *satisfying* $\delta_1/(\delta_1 - 1) < \delta_4 < 2$. *Also,* $\|\eta_t\|_{q_0} = (\mathbf{E}|\eta_t|^{q_0})^{1/q_0} < \infty$ *with* $q_0 = \delta_4\delta_6/(\delta_4 - \delta_6)$ *for some* $1 < \delta_6 < \delta_4$. *Further,* $\sup_k \mathbf{E}\left(\eta_1^2 \epsilon_{k+1}^2 \middle| Z_{k+1} = z\right) \leq C_2 < \infty$.

(e) $f(z_0, z_s | x_0, x_s; s) \leq M \leq \infty$ *for* $s \geq 1$, *where* $f(z_0, z_s | x_0, x_s; s)$ *is the conditional density of* (Z_0, Z_s) *given* $(X_{01} = x_0, X_{s1} = x_s)$.

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

(g) The bandwidth h *satisfies* $h \longrightarrow 0$ *and* $nh \longrightarrow \infty$.

(h) $n^{1/2-\delta_1/4} h^{\delta_1/\delta_2 - 1/2 - \delta_1/4} = O(1)$

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 α-mixing is assumed to have geometrically decaying coefficients and is the weakest condition one can impose for weakly dependent stochastic processes. Assumptions 5 (e) is a technical assumption required for the proofs. Moreover, Assumptions 5 (f) is commonly imposed in the literature and implies that the kernel function $K(\cdot)$ is compactly supported (which can be relaxed at the expense of lengthier proofs) while (i) and (j) are assumptions on the bandwidth parameter and allow for a wide range of choices. Assumption (i) in particular is slightly stronger than the assumption $nh \longrightarrow \infty$ but it immediately satisfies the selection criterion $h = cn^{-\lambda}$ for $0 < \lambda < 1$ and $c > 0$ required for optimal bandwidths. See Cai et al. (2009) for details.

Consider next the regularity conditions required to establish the limiting distribution of $\widehat{\beta}(z)$. In this regard, denote by $f_z(z)$ the marginal density of Z_t and define the k^{th} conditional moment of X_{t1} with respect to $Z_t = z$ as $A_k(z) = \mathbf{E}\left(X_{t1}^{\otimes k} | Z_t = z\right)$, for $k = 1, 2$. Finally, for $j \geq 0$ define $\mu_j(K) = \int_{-\infty}^{\infty} v^j K(v) dv$ and $\nu_j(K) = \int_{-\infty}^{\infty} v^j K^2(v) dv$, and let

$$S_W(z) = \begin{pmatrix} A_2(z) & A_1(z) V_D^{(1)\top} \\ V_D^{(1)} A_1(z)^\top & V_D^{(2)} \end{pmatrix} \tag{13a}$$

$$S_N(z) = \begin{pmatrix} A_2(z) & A_1(z) V_{\eta,c}^{(1)\top} & A_1(z) V_{\omega,0}^{(1)} \\ V_{\eta,c}^{(1)} A_1(z)^\top & V_{\eta,c}^{(2)} & V_{\eta,c,\omega,0} \\ V_{\omega,0}^{(1)} A_1(z)^\top & V_{\eta,c,\omega,0} & V_{\omega,0}^{(2)} \end{pmatrix} \tag{13b}$$

2.3. Asymptotic Properties

To develop the asymptotic properties of $\widehat{\beta}(z)$, first let $D_{m,n} = \text{diag}\left\{I_{d_1}, m^{-1/2} n^{1/2} I_{d_2}\right\}$ and $D_n = \text{diag}\left\{I_{d_1}, n^{1/2} I_{d_2+d_3}\right\}$ where I_{d_i} is a $d_i \times d_i$ 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* $m = o\left(n^{1-\frac{1}{p}} \wedge n^{\frac{2}{3}}\right)$*,* $d_2 = 1$ *and* $d_3 = 0$*, then*

$$\sqrt{hn} D_{m,n} \left(\widehat{\beta}(z) - \beta(z) - h^2 B_\beta(z)\right) \longrightarrow_d \mathbf{N}\left(\Sigma_{\beta,W}(z)\right)$$

where $\mathbf{N}\left(\Sigma_{\beta,W}(z)\right)$ *is a multivariate normal distribution with mean zero and conditional covariance matrix given by* $\Sigma_{\beta,W}(z) = v_0(K) S_W(z)^{-1}/f_z(z)$

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

$$\sqrt{hn} D_n \left(\widehat{\beta}(z) - \beta(z) - h^2 B_\beta(z)\right) \longrightarrow_d \mathbf{MN}\left(\Sigma_{\beta,N}(z)\right)$$

a mean zero mixed normal distribution with conditional covariance matrix $\Sigma_{\beta,N}(z) = v_0(K) \left(Q^\top S_N(z) Q\right)^{-1}/f_z(z)$*, where* $Q^\top = [I_{d_1+d_2}, \mathbf{0}]$ *is a* $(d_1+d_2)\times d$ *matrix. When* $d_2 = d_3 = 1$ *then* $\Sigma_{\beta,N}(z) = v_0(K) S_N(z)^{-1}/f_z(z)$

Note that Theorem 1 nests several important results. When $d_2 = d_3 = 0$ the results of Cai et al. (2000) are recovered with $\Sigma_{\beta,W}(z) = \Sigma_{\beta_1,0,W}(z) = \sigma_\epsilon^2 \nu_0(K) M_2(z)^{-1} f_z^{-1}(z)$. On the other hand, when $d_2 = 0$ and $d_3 \geq 1$ the results of Cai et al. (2009) emerge with $\Sigma_{\beta,N}(z) = \Sigma_{\beta_1,0,\beta_3,N}(z) = \sigma_\epsilon^2 \nu_0(K) S(z)^{-1} f_z^{-1}(z)$ where

$$S(z) = \begin{pmatrix} A_2(z) & A_1(z) V_{\omega,0}^{(1)\top} \\ V_{\omega,0}^{(1)} A_1(z)^\top & V_{\omega,0}^{(2)} \end{pmatrix}$$

Theorem 1 also lends insight into convergence rates for $\mathbf{V}\left(\widehat{\beta}_1\right)$, $\mathbf{V}\left(\widehat{\beta}_2\right)$, and $\mathbf{V}\left(\widehat{\beta}_3\right)$. In particular, the use of local linear fitting to estimate β implies that $\mathbf{V}\left(\widehat{\beta}_1\right)$ is of order $O\left((nh)^{-1}\right)$, $\mathbf{V}\left(\widehat{\beta}_2\right)$ is of order $O\left(m\left(n^2 h\right)^{-1}\right)$ when X_{t2} is weakly $I(1)$ and $O\left(\left(n^2 h\right)^{-1}\right)$ when X_{t2} is nearly $I(1)$. The order of $\mathbf{V}\left(\widehat{\beta}_3\right)$ is $O\left(\left(n^2 h\right)^{-1}\right)$ when X_{t3} is a pure $I(1)$ process. Since $m = o\left(n^{1-\frac{1}{p}} \wedge n^{\frac{2}{3}}\right)$, it is clear that the rate of convergence of the variance of $\widehat{\beta_2}$ 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 $\widehat{\beta}_2$ and $\widehat{\beta}_3$ 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

$$\begin{aligned} IAMSE_W(\widehat{\beta}_i(z)) = & \frac{1}{4} h^4 \mu_2^2(K) \int \|\beta_i^{(2)}(z)\| w(z) dz \\ & + (m^{-i+1} n^i h)^{-1} \int \operatorname{tr}\left(\Sigma_{\beta_i,W}(z)\right) w(z) dz \end{aligned}$$

$$IAMSE_N(\widehat{\beta}_j(z)) = \begin{cases} \frac{1}{4} h^4 \mu_2^2(K) \int \|\beta_j^{(2)}(z)\| w(z) dz \\ +(n^j h)^{-1} \int \operatorname{tr}\left(\Sigma_{\beta_j,N}(z)\right) w(z) dz & \text{for } j = 1, 2 \\ \\ \frac{1}{4} h^4 \mu_2^2(K) \int \|\beta_3^{(2)}(z)\| w(z) dz \\ +(n^2 h)^{-1} \int \operatorname{tr}\left(\Sigma_{\beta_3,N}(z)\right) w(z) dz & \text{for } j = 3 \end{cases}$$

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 $\Sigma_{\beta,W}(z)$ and $\Sigma_{\beta,N}(z)$ 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 $h^\star$. It is easily verified that the optimal bandwidth in case of the weak unit root and near unit models are respectively given by

$$h^\star_{W,i} = \left(m^{-i+1}n^i\right)^{-1/5}\left(\int \mathrm{tr}\left(\Sigma_{\beta_i,W}(z)\right)w(z)dz\right)^{1/5} \times \left(\int \mu_2^2(K)\|\beta_i^{(2)}(z)\|w(z)dz\right)^{-1/5}$$

$$h^\star_{N,j} = \begin{cases} n^{-j/5}\left(\int \mathrm{tr}\left(\Sigma_{\beta_j,N}(z)\right)w(z)dz\right)^{1/5} \\ \times\left(\int \mu_2^2(K)\|\beta_j^{(2)}(z)\|w(z)dz\right)^{-1/5} & \text{for } j = 1,2 \\ \\ n^{-2/5}\left(\int \mathrm{tr}\left(\Sigma_{\beta_3,N}(z)\right)w(z)dz\right)^{1/5} \\ \times\left(\int \mu_2^2(K)\|\beta_3^{(2)}(z)\|w(z)dz\right)^{-1/5} & \text{for } j = 3 \end{cases}$$

The above implies that the minimal IAMSE has order $O\left(\left(m^{-i+1}n^i\right)^{-\frac{4}{5}}\right)$ which becomes $O\left(n^{-\frac{4}{5}\left(1-\frac{1-i}{p}\right)} \wedge n^{-\frac{4}{5}\frac{2+i}{3}}\right)$ for some $p > 2$ in case of the weak unit root model since the configuration assumes $m = o\left(n^{1-\frac{1}{p}} \wedge n^{\frac{2}{3}}\right)$. In contrast, the orders of the IAMSE for the near $I(1)$ model are $O\left(n^{-4j/5}\right)$ for $j = 1,2$ and $O\left(n^{-8/5}\right)$ for $j = 3$. In either model it is clear that a single optimal choice of h is not possible for all elements of $\beta(z)$. The reader is referred to Section 2.4 of Cai et al. (2009) for a two-step estimation procedure which guarantees optimal convergence rates for all elements of $\beta(z)$. Note in particular that in the nearly integrated model since $\beta_2(z)$ and $\beta_3(z)$ can be optimized with a single optimal bandwidth $h^\star_{N,2,3} = O\left(n^{-8/5}\right)$, the two-step procedure can be applied here as well with minimal adaptation.

3. Models with Integrated and Nearly Integrated Z_t

Establishing results when Z_t is an integrated or nearly integrated process can be quite complex. Technical details for general models are still under development and the working paper of Gao and Phillips (2013) in particular is developing limiting results for models with nonstationarity in both the regressors and the varying coefficient components. Accordingly, the approach here considers the model in equation (1) when Z_t is a univariate near $I(1)$ or univariate pure $I(1)$ process and X_t is a d-dimensional nearly integrated vector of covariates.

As in Section 2, the model in this section assumes $\beta(z)$ is twice continuously differentiable with its local linear estimator again given by equation (4). Moreover, since Z_t is a nearly (possibly purely) $I(1)$ process, it can be expressed as

$$Z_t = (1 - c_z/n)Z_{t-1} + \xi_t \tag{14}$$

where c_z 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, ξ_t is assumed to be a mean zero $I(0)$ linear process satisfying

Assumptions 6.

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

(b) *For some* $\gamma > 2+\delta$ *with* $0 < \delta \leq 2$ *and* $\lambda = \lambda(\delta) > 0$, $\mathbf{E}|\xi_t|^\gamma < \infty$ *and* $\sum_{k=0}^{\infty} \alpha(k)^{1/(2+\delta)-1/\gamma} < \infty$

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

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

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

Assumptions 7.

(a) $\{\varepsilon_i\}$ *is a sequence of IID continuous random vectors with* $\mathbf{E}(\varepsilon_1) = 0$, *a positive definite matrix* Σ_ε *and finite fourth order cumulants.*

(b) $\int_{-\infty}^{\infty} |u||\varphi_\varepsilon(u)|du < \infty$ *where* $\varphi_\varepsilon(u)$ *is the characteristic function of* ε_1.

(c) $\int_{-\infty}^{\infty} \left|p_\varepsilon(x+y) - p_\varepsilon(x)\right|dx \leq c_\varepsilon |y|$ *for each* y *and constant* c_ε, *where* $p_\varepsilon(\cdot)$ *is the density of* ε_1.

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

(e) $\sum_{j=0}^{\infty} c_{j,kl} x^j$ *for* $|x| \leq 1$ *and* $c_{j,kl} = O\left(j^{-\varsigma_\star}\right)$ *as* $j \longrightarrow \infty$ *and* $\varsigma_\star > 1$ *satisfies* $\varsigma_\star + 1/2 > 2 + \varsigma > 2/(\varsigma_\star - 1)$ *with* ς *defined in (d) above.*

(f) $\mathcal{G} = \sigma\left(\epsilon_t, \ldots, \epsilon_1; \varepsilon_{t+1}, \varepsilon_t, \ldots, \varepsilon_{-\infty}\right)$ *be a* σ*-field generated by* $\{(\epsilon_i, \varepsilon_j) : 1 \leq i \leq t; -\infty \leq j \leq t+1\}$ *where* $\mathbf{E}\left(\epsilon_t \mid \mathcal{G}_{t-1}\right) = 0$ *almost surely (a.s.),* $\mathbf{E}\left(\epsilon_t^2 \mid \mathcal{G}_{t-1}\right) = \sigma_\epsilon^2$ *a.s., and* $\mathbf{E}\left(\epsilon_t^4 \mid \mathcal{G}_{t-1}\right) < \infty$ *a.s. for all* $t \geq 2$ *where* $\sigma_\epsilon^2 > 0$.

(g) *Let* $W_{\epsilon,n}(r) = n^{-1/2}\sum_{t=1}^{\lfloor nr \rfloor} \epsilon_t$ *and* $W_n(r) = \left(V_{c_x n}(r)^\top, V_{c_z n}(r)\right)^\top$ *where similar to equation* (5a), $V_{c_x n}(r) = n^{-1/2} X_{\lfloor nr \rfloor}$ *and* $V_{c_z n}(r) = n^{-1/2} Z_{\lfloor nr \rfloor}$. *There exists a Skorohod space* $D[0,1]^{d+2}$ *on which* $(W_{\epsilon,n}(r), W_n(r)) \longrightarrow_d (V_{\epsilon,0}(r), V_w(r))$ *as* $n \longrightarrow \infty$, *where* $(V_{\epsilon,0}(r), V_w(r))$ *is a vector stochastic process,* $V_{\epsilon,0}(r)$ *is a Brownian motion process,* $V_w(r) = (V_{\eta,c_x}(r), V_{\xi,c_z}(r))$ *and* $V_{\eta,c_x}(r)$ *and* $V_{\xi,c_z}(r)$ *are Ornstein-Uhlenbeck processes with parameters* c_x *and* c_z *respectively, when* $c_x, c_z > 0$. *When* $c_z = 0$, $V_{\xi,0}(r)$ *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* $h \longrightarrow 0$ *and* $nh \longrightarrow 0$ *as* $n \longrightarrow \infty$.

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 α-mixing and accommodates contemporaneous endogeneity between regressors, varying coefficient components, and regression residuals, whereas assumption (f) and (g) allow for heteroskedastic residuals ϵ_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* $c_x > 0$ *and* $c_z \geq 0$. *Then, as* $n \longrightarrow \infty$,

$$\sqrt{hn^{3/4}}\left(\widehat{\beta}(z) - \beta(z) - h^2 B_\beta(z)\right) \longrightarrow_d \mathbf{MN}\left(\Sigma_\beta\right)$$

a mean zero mixed normal distribution with conditional covariance matrix

$$\Sigma_\beta = \sigma_\epsilon^2 \nu_0(K) \int_0^1 V_{\eta,c_x} V_{\eta,c_x}^\top dL_{V_{\xi,c_z}}(1,0)$$

There are several important remarks to note at this point. Observe in particular that Theorem 2 implies the asymptotic variance of $\widehat{\beta}(z)$ has order $O\left(h^{-1}n^{-3/4}\right)$. This is clearly larger than $O\left(h^{-1}n^{-1}\right)$ 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 $\widehat{\beta}(z)$ 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 $\widehat{\beta}$ given by

$$IAMSE = \int \left(\frac{1}{4} h^4 \mu_2^2(K) \|\beta^{(2)}(z)\| + h^{-1} n^{-3/4} \operatorname{tr}(\Sigma_\beta) \right) w(z) dz$$

for some non-negative weight function $w(\cdot)$. Note further that minimizing the IAMSE with respect to h renders the optimal bandwidth h as $h^\star = cn^{-3/20}$ 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

István Berkes and Lajos Horváth. Convergence of integral functionals of stochastic processes. *Econometric Theory*, 22(02): 304–322, 2006.

Zongwu Cai. Trending time-varying coefficient time series models with serially correlated errors. *Journal of Econometrics*, 136 (1):163–188, 2007.

Zongwu Cai and Y Wang. Instability of predictability of asset returns. Technical report, Working paper, Department of Mathematics and Statistics, University of North Carolina at Charlotte, 2008.

Zongwu Cai, Jianqing Fan, and Qiwei Yao. Functional-coefficient regression models for nonlinear time series. *Journal of the American Statistical Association*, 95(451):941–956, 2000.

Zongwu Cai, Qi Li, and Joon Y Park. Functional-coefficient models for nonstationary time series data. *Journal of Econometrics*, 148(2):101–113, 2009.

Rong Chen and Ruey S Tsay. Functional-coefficient autoregressive models. *Journal of the American Statistical Association*, 88(421):298–308, 1993.

J Fan and I Gijbels. *Local Polynomial Modelling and its Applications. 1996*. Chapman, Hall, London, 2003.

Jianquing Fan. *Nonlinear time series: nonparametric and parametric methods*. Springer, 2003.

Jiti Gao and Peter CB Phillips. Functional coefficient nonstationary regression. Technical report, Cowles Foundation for Research in Economics, Yale University, 2013.

Peter Hall and Christopher C Heyde. *Martingale limit theory and its application*. Academic press New York, 1980.

Qi Li and Jeffrey Scott Racine. *Nonparametric econometrics: Theory and practice*. Princeton University Press, 2007.

Joon Y Park. Weak unit roots. Technical report, Rice University, Department of Economics, 2003.

Joon Y Park and Sang B Hahn. Cointegrating regressions with time varying coefficients. *Econometric Theory*, 15(5):664–703, 1999.

Joon Y Park and Peter CB Phillips. Asymptotics for nonlinear transformations of integrated time series. *Econometric Theory*, 15(3):269–298, 1999.

Joon Y Park and Peter CB Phillips. Nonlinear regressions with integrated time series. *Econometrica*, 69(1):117–161, 2001.

Peter CB Phillips. Towards a unified asymptotic theory for autoregression. *Biometrika*, 74(3):535–547, 1987.

Peter CB Phillips. Local limit theory and spurious nonparametric regression. *Econometric Theory*, 25(6):1466, 2009.

Peter M Robinson. *Nonparametric estimation of time-varying parameters*. Springer, 1989.

Yiguo Sun, Zongwu Cai, and Qi Li. Semiparametric functional coefficient models with integrated covariates. *Econometric Theory*, 29(03):659–672, 2013.

Qiying Wang. Martingale limit theorems revisited and non-linear cointegrating regression. Technical report, Working Paper, 2011.

Qiying Wang and Peter CB Phillips. Asymptotic theory for local time density estimation and nonparametric cointegrating regression. *Econometric Theory*, 25(3):710, 2009.

Zhijie Xiao. Functional-coefficient cointegration models. *Journal of Econometrics*, 152(2):81–92, 2009.

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

$$\mathcal{H}_{m,n} \begin{pmatrix} \widehat{\beta}(z) \\ \widehat{\beta}^{(1)}(z) \end{pmatrix} = S_{m,n}(z)^{-1} n^{-1} \sum_{t=1}^{n} K_h(Z_t - z) \times Y_t \begin{pmatrix} 1 \\ Z_{t,z,h} \end{pmatrix} \otimes \left(D_{m,n}^{-1} X_t\right) \tag{16a}$$

$$\mathcal{H}_{n} \begin{pmatrix} \widehat{\beta}(z) \\ \widehat{\beta}^{(1)}(z) \end{pmatrix} = S_{n}(z)^{-1} n^{-1} \sum_{t=1}^{n} K_h(Z_t - z) \times Y_t \begin{pmatrix} 1 \\ Z_{t,z,h} \end{pmatrix} \otimes \left(D_{n}^{-1} X_t\right) \tag{16b}$$

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

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

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

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

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

$$S_{n,j}(z) = \begin{pmatrix} F_{n,j,0}(z) & F_{n,j,1}(z) & F_{n,j,3}(z) \\ F_{n,j,1}(z)^{\top} & F_{n,j,2}(z) & F_{n,j,4}(z) \\ F_{n,j,3}(z)^{\top} & F_{n,j,4}(z)^{\top} & F_{n,j,5}(z) \end{pmatrix}$$

where

$$
\begin{aligned}
F_{m,n,j,0}(z) &= F_{n,j,0}(z) \\
&= n^{-1}\sum_{t=1}^{n} Z_{t,z,h}^{j \bmod 3} K_h(Z_t - z) X_{t1} X_{t1}^{\top} \\
F_{m,n,j,1}(z) &= m^{1/2} F_{n,j,0}(z) \\
&= n^{-1}\sum_{t=1}^{n} K_h(Z_t - z) Z_{t,z,h}^{j \bmod 3} X_{t1} m^{1/2} n^{-1/2} X_{t2}^{\top} \\
F_{m,n,j,2}(z) &= m F_{n,j,2}(z) \\
&= n^{-1}\sum_{t=1}^{n} Z_{t,z,h}^{j \bmod 3} K_h(Z_t - z) \left(m^{1/2} n^{-1/2} X_{t2}\right)^{\otimes 2} \\
F_{n,j,3}(z) &= n^{-1}\sum_{t=1}^{n} Z_{t,z,h}^{j \bmod 3} K_h(Z_t - z) X_{t1} n^{-1/2} X_{t3}^{\top} \\
F_{n,j,4}(z) &= n^{-1}\sum_{t=1}^{n} Z_{t,z,h}^{j \bmod 3} K_h(Z_t - z) n^{-1/2} X_{t2} n^{-1/2} X_{t3}^{\top} \\
F_{n,j,5}(z) &= n^{-1}\sum_{t=1}^{n} Z_{t,z,h}^{j \bmod 3} K_h(Z_t - z) \left(n^{-1/2} X_{t3}\right)^{\otimes 2}
\end{aligned}
$$

For $l = 1, 2$, also define the quantity

$$
F_{n,j,l}^{\star}(z) = n^{-1}\sum_{t=1}^{n} Z_{t,z,h}^{j \bmod 3} X_{t1}^{\otimes l} K_h(Z_t - z)
$$

Note further that X_{t1} 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.

$$
\begin{aligned}
\mathbf{E}F_{n,j,l}^{\star}(z) &= \mathbf{E}\left(Z_{t,z,h}^{j \bmod 3} X_{t1}^{\otimes l} K_h(Z_t - z)\right) \\
&= f_z(z) M_l(z) \mu_{j \bmod 3}(K) + o(1) \\
\mathbf{V}F_{n,j,l}^{\star}(z) &= O((nh)^{-1}) \\
&= o(1)
\end{aligned}
$$

where the last line above follows from Assumptions 5 (g). Accordingly,

$$
F_{n,j,l}^{\star}(z) = f_z(z) M_l(z) \mu_{j \bmod 3}(K) + o_p(1) \tag{17}
$$

and therefore

$$
\begin{aligned}
F_{m,n,j,0}(z) = F_{n,j,0}(z) &= F_{n,j,2}^{\star}(z) \\
&= f_z(z) M_l(z) \mu_{j \bmod 3}(K) + o_p(1)
\end{aligned} \tag{18}
$$

Consider next $\mathcal{F}_i^e = \sigma\left(X_{t1}, Z_i : t \leq 1\right)$ as the smallest σ-field containing the history of (X_{t1}, Z_t) and define $e_t = K_h(Z_t - z) Z_{t,z,h}^{j \bmod 3} X_{t1} - \mathbf{E}\left(K_h(Z_t - z) Z_{t,z,h}^{j \bmod 3} X_{t1}\right)$. Also, let $V_{mnt} = m^{1/2} n^{-1/2} X_{t2}$ and note that

$V_{mn\lfloor nr \rfloor} = m^{1/2}V_{mn}(r)$ where $V_{mn}(r)$ was defined in equation (5a) for any $r \in [0,1]$. Moreover, for $0 \le \delta \le 1$, set $N = \lfloor 1/\delta \rfloor$, $t_k = \lfloor kn/N \rfloor + 1$, $t_k^\star = t_{k+1} - 1$, and $t_k^{\star\star} = \min\{t_k^\star, n\}$. Then, replacing U_{nt} in the proof of Cai et al. (2009) with V_{mnt} here, yields the following result.

$$\begin{aligned}
\left| n^{-1} \sum_{i=1}^{n} V_{mni} e_i \right| &= \left| n^{-1} \sum_{k=0}^{N-1} \sum_{t=t_k}^{t_k^{\star\star}} V_{mnt} e_t \right| \\
&\le n^{-1} \sum_{k=0}^{N-1} \sum_{t=t_k}^{t_k^{\star\star}} \left| \mathbf{E}\left(V_{mnk} V_{mnt} e_k e_t \right) \right| \\
&= n^{-1} \sum_{k=0}^{N-1} \sum_{t=t_k}^{t_k^{\star\star}} \left| \mathbf{E}\left(V_{mnk} V_{mnt} \mathbf{E}\left(e_k e_t | Z_k, Z_t \right) \right) \right|
\end{aligned}$$

Next, note that

$$\begin{aligned}
n^{-1} \sum_{k=0}^{N-1} \sum_{t=t_k}^{t_k^{\star\star}} \left| \mathbf{E}\left(e_k e_t | Z_k, Z_t \right) \right| &\le \frac{N}{n} \sup_{0 \le k \le N-1} \mathbf{E} \left(\sum_{t=t_k}^{t_k^{\star\star}} e_t^2 \middle| Z_t \right) \\
&\le \sup_{t \le n} \mathbf{E} \left(\frac{1}{\delta n} \sum_{i=t}^{t+\delta n} e_i^2 \middle| Z_i \right) \\
&\le C h^{-1}
\end{aligned}$$

where the last line follows from the result

$$\sup_{s \ge 0} \mathbf{V} \left(\sum_{t=s+1}^{s+a} e_t \right) = O(a h^{-1})$$

Accordingly,

$$\begin{aligned}
\left| n^{-1} \sum_{i=1}^{n} V_{mni} e_i \right| &\le C h^{-1} n^{-1} \sum_{k=0}^{N-1} \sum_{t=t_k}^{t_k^{\star\star}} \left| \mathbf{E}\left(V_{mnk} V_{mnt} \right) \right| \\
&\le C h^{-1} \sup_{t \le n} \mathbf{E} \left(n^{-1} \sum_{t=1}^{n} V_{mnt}^2 \right)
\end{aligned}$$

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

$$n^{-1} \sum_{i=1}^{n} V_{mni} e_i = O\left(m (nh)^{-1} \right) \longrightarrow 0$$

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

$$\begin{aligned}
F_{m,n,j,1}(z) &= \mathbf{E}\left(K_h(Z_t - z) Z_{t,z,h}^{j \bmod 3} X_{t1} \right) n^{-1} \sum_{t=1}^{n} V_{mnt} + n^{-1} \sum_{i=1}^{n} V_{mni} e_i \\
&= f_z(z) M_1(z) \mu_{j \bmod 3}(K) V_D^{(1)} + o_p(1) \qquad (19)
\end{aligned}$$

Similar reasoning also produces

$$F_{m,n,j,2}(z) = f_z(z)\mu_{j \bmod 3}(K)V_D^{(2)} + o_p(1) \tag{20}$$

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

$$n^{-1}\sum_{i=1}^{n} V_{cni}e_i = O\left((nh)^{-1}\right) \longrightarrow 0$$

Accordingly, it follows from the above that

$$F_{n,j,1}(z) = f_z(z)M_1(z)\mu_{j \bmod 3}(K)V_{\eta,c}^{(1)} + o_p(1) \tag{21a}$$

$$F_{n,j,2}(z) = f_z(z)\mu_{j \bmod 3}(K)V_{\eta,c}^{(2)} + o_p(1) \tag{21b}$$

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

$$F_{n,j,3}(z) = f_z(z)M_1(z)\mu_{j \bmod 3}(K)V_{\omega,0}^{(1)} + o_p(1) \tag{22a}$$

$$F_{n,j,5}(z) = f_z(z)\mu_{j \bmod 3}(K)V_{\omega,0}^{(2)} + o_p(1) \tag{22b}$$

What remains to be shown is the limiting form of $F_{n,j4}$. To do so, define $\tilde{e}_t = K_h(Z_t - z)Z_{t,z,h}^{j \bmod 3} - \mathbf{E}\left(K_h(Z_t - z)Z_{t,z,h}^{j \bmod 3}\right)$ and $V_{c0nt} = n^{-1/2}X_{t2}n^{-1/2}X_{t3}^{\top}$ and note that for $r \in [0,1]$, the results in Phillips (2009) demonstrate that $V_{c0n\lfloor nr\rfloor}$ 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 $n^{-1}\sum_{i=1}^{n} V_{c0ni}\tilde{e}_i \longrightarrow 0$ and therefore

$$\begin{aligned} F_{n,j,4}(z) &= \mathbf{E}\left(K_h(Z_t - z)Z^{j \bmod 3}\right)n^{-1}\sum_{t=1}^{n} V_{c0nt} + n^{-1}\sum_{i=1}^{n} V_{c0ni}e_i \\ &= f_z(z)\mu_{j \bmod 3}(K)V_{\eta,c,\omega,0} + o_p(1) \end{aligned} \tag{23}$$

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

$$S_{m,n}(z) = f_z(z)\begin{pmatrix}1 & 0\\ 0 & \mu_2(K)\end{pmatrix} \otimes S_W(z) + o_p(1) \tag{24a}$$

$$S_n(z) = f_z(z)\begin{pmatrix}1 & 0 & 0\\ 0 & \mu_2(K) & 0\\ 0 & 0 & \mu_2(K)\end{pmatrix} \otimes S_N(z) + o_p(1) \tag{24b}$$

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

$$R_{m,n}(z)^{-1} = f_z(z)^{-1}S_W(z)^{-1} + o_p(1) \tag{25a}$$

$$R_n(z)^{-1} = f_z(z)^{-1}S_N(z)^{-1} + o_p(1) \tag{25b}$$

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

$$D_{m,n}\left(\widehat{\beta}(z)-\beta(z)\right) \equiv E_{m,n,1}+E_{m,n,2} \tag{26a}$$

$$D_{n}\left(\widehat{\beta}(z)-\beta(z)\right) \equiv E_{n,1}+E_{n,2} \tag{26b}$$

where

$$E_{m,n,1} = R_{m,n}(z)^{-1}B_{m,n}(z) \tag{27a}$$

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

$$E_{n,1} = R_{n}(z)^{-1}B_{n}(z) \tag{27c}$$

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

and

$$\begin{aligned} B_{m,n}(z) &= n^{-1}\sum_{t=1}^{n}K_h(Z_t-z)X_t^{\otimes 2}D_{m,n}^{-1} \\ &\quad \times\left(\beta(Z_t)-\beta(z)-(Z_t-z)\beta^{(1)}(z)\right) \\ B_{n}(z) &= n^{-1}\sum_{t=1}^{n}K_h(Z_t-z)X_t^{\otimes 2}D_{n}^{-1} \\ &\quad \times\left(\beta(Z_t)-\beta(z)-(Z_t-z)\beta^{(1)}(z)\right) \end{aligned}$$

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

$$\begin{aligned} B_{m,n}(z) &= \begin{pmatrix} G_{n,0}(z)+G_{m,n,1}(z) \\ G_{m,n,2}(z)+G_{m,n,3}(z) \end{pmatrix} \\ B_{n}(z) &= \begin{pmatrix} G_{n,0}(z)+G_{n,1}(z)+G_{n,2}(z) \\ G_{n,3}(z)+G_{n,4}(z)+G_{n,5}(z) \\ G_{n,6}(z)+G_{n,7}(z)+G_{n,8}(z) \end{pmatrix} \end{aligned}$$

where

$$
\begin{aligned}
G_{n,0}(z) &= n^{-1}\sum_{t=1}^{n} K_h(Z_t - z)X_{t1}^{\otimes 2} \\
&\quad \times \Big(\beta_1(Z_t) - \beta_1(z) - (Z_t - z)\beta_1^{(1)}(z)\Big) \\
G_{m,n,1}(z) &= n^{-1}\sum_{t=1}^{n} K_h(Z_t - z)X_{t1} m^{1/2} n^{-1/2} X_{t2}^{\top} \\
&\quad \times \Big(\beta_2(Z_t) - \beta_2(z) - (Z_t - z)\beta_2^{(1)}(z)\Big) \\
G_{m,n,2}(z) &= n^{-1}\sum_{t=1}^{n} m^{-1/2} n^{1/2} K_h(Z_t - z) m^{1/2} n^{-1/2} X_{t2} X_{t1}^{\top} \\
&\quad \times \Big(\beta_1(Z_t) - \beta_1(z) - (Z_t - z)\beta_1^{(1)}(z)\Big) \\
G_{m,n,3}(z) &= n^{-1}\sum_{t=1}^{n} m^{-1/2} n^{1/2} K_h(Z_t - z) m n^{-1} X_{t2}^{\otimes 2} \\
&\quad \times \Big(\beta_2(Z_t) - \beta_2(z) - (Z_t - z)\beta_2^{(1)}(z)\Big)
\end{aligned}
$$

and similarly

$$
\begin{aligned}
G_{n,1}(z) &= n^{-1}\sum_{t=1}^{n} K_h(Z_t - z)X_{t1} n^{-1/2} X_{t2}^{\top} \\
&\quad \times \Big(\beta_2(Z_t) - \beta_2(z) - (Z_t - z)\beta_2^{(1)}(z)\Big) \\
G_{n,2}(z) &= n^{-1}\sum_{t=1}^{n} K_h(Z_t - z)X_{t1} n^{-1/2} X_{t3}^{\top} \\
&\quad \times \Big(\beta_3(Z_t) - \beta_3(z) - (Z_t - z)\beta_3^{(1)}(z)\Big) \\
G_{n,3}(z) &= n^{-1}\sum_{t=1}^{n} n^{1/2} K_h(Z_t - z) n^{-1/2} X_{t2} X_{t1}^{\top} \\
&\quad \times \Big(\beta_1(Z_t) - \beta_1(z) - (Z_t - z)\beta_1^{(1)}(z)\Big) \\
G_{n,4}(z) &= n^{-1}\sum_{t=1}^{n} n^{1/2} K_h(Z_t - z) n^{-1} X_{t2}^{\otimes 2} \\
&\quad \times \Big(\beta_2(Z_t) - \beta_2(z) - (Z_t - z)\beta_2^{(1)}(z)\Big) \\
G_{n,5}(z) &= n^{-1}\sum_{t=1}^{n} n^{1/2} K_h(Z_t - z) n^{-1/2} X_{t2} n^{-1/2} X_{t3}^{\top} \\
&\quad \times \Big(\beta_3(Z_t) - \beta_3(z) - (Z_t - z)\beta_3^{(1)}(z)\Big)
\end{aligned}
$$

$$
\begin{aligned}
G_{n,6}(z) &= n^{-1}\sum_{t=1}^{n} n^{1/2}K_h(Z_t-z)n^{-1/2}X_{t3}X_{t1}^{\top} \\
&\quad\times\Big(\beta_1(Z_t)-\beta_1(z)-(Z_t-z)\beta_1^{(1)}(z)\Big) \\
G_{n,7}(z) &= n^{-1}\sum_{t=1}^{n} n^{1/2}K_h(Z_t-z)n^{-1/2}X_{t3}n^{-1/2}X_{t2}^{\top} \\
&\quad\times\Big(\beta_2(Z_t)-\beta_2(z)-(Z_t-z)\beta_2^{(1)}(z)\Big) \\
G_{n,8}(z) &= n^{-1}\sum_{t=1}^{n} n^{1/2}K_h(Z_t-z)n^{-1}X_{t3}^{\otimes 2} \\
&\quad\times\Big(\beta_3(Z_t)-\beta_3(z)-(Z_t-z)\beta_3^{(1)}(z)\Big)
\end{aligned}
$$

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

$$
\begin{aligned}
\mathbf{E}\left(G_{n,0}\right) &= h^2 f_z(z)M_2(z)\left(\frac{1}{2}\mu_2(K)\beta_1^{(2)}(z)\right)(1+o(1)) \qquad (29)\\
\mathbf{V}\left(G_{n,0}\right) &= o(1)
\end{aligned}
$$

where $\mathbf{V}(\cdot)$ is the variance operator. Accordingly,

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

and it can be shown that

$$
\begin{aligned}
G_{m,n,1} &= h^2 f_z(z)M_1(z)V_D^{(1)\top}\left(\frac{1}{2}\mu_2(K)\beta_2^{(2)}(z)\right)(1+o_p(1)) \\
G_{m,n,2} &= h^2 f_z(z)V_D^{(1)}M_1(z)^{\top}m^{-1/2}n^{1/2}\left(\frac{1}{2}\mu_2(K)\beta_1^{(2)}(z)\right)(1+o_p(1)) \\
G_{m,n,3} &= h^2 f_z(z)V_D^{(2)}m^{-1/2}n^{1/2}\left(\frac{1}{2}\mu_2(K)\beta_2^{(2)}(z)\right)(1+o_p(1))
\end{aligned}
$$

Similarly, it can also be demonstrated that

$$
\begin{aligned}
G_{n,1} &= h^2 f_z(z)M_1(z)V_{\eta,c}^{(1)\top}\left(\frac{1}{2}\mu_2(K)\beta_2^{(2)}(z)\right)(1+o_p(1)) \\
G_{n,2} &= h^2 f_z(z)M_1(z)V_{\omega,0}^{(1)\top}\left(\frac{1}{2}\mu_2(K)\beta_3^{(2)}(z)\right)(1+o_p(1)) \\
G_{n,3} &= h^2 f_z(z)V_{\eta,c}^{(1)}M_1(z)^{\top}n^{1/2}\left(\frac{1}{2}\mu_2(K)\beta_1^{(2)}(z)\right)(1+o_p(1)) \\
G_{n,4} &= h^2 f_z(z)V_{\eta,c}^{(2)}n^{1/2}\left(\frac{1}{2}\mu_2(K)\beta_2^{(2)}(z)\right)(1+o_p(1))
\end{aligned}
$$

$$
\begin{aligned}
G_{n,5} &= h^2 f_z(z) V_{\eta,c,\omega,0} n^{1/2} \left(\frac{1}{2}\mu_3(K)\beta_3^{(2)}(z) \right) (1 + o_p(1)) \\
G_{n,6} &= h^2 f_z(z) V_{\omega,0}^{(1)} M_1(z)^\top n^{1/2} \left(\frac{1}{2}\mu_2(K)\beta_1^{(2)}(z) \right) (1 + o_p(1)) \\
G_{n,7} &= h^2 f_z(z) V_{\eta,c,\omega,0}^\top n^{1/2} \left(\frac{1}{2}\mu_2(K)\beta_2^{(2)}(z) \right) (1 + o_p(1)) \\
G_{n,8} &= h^2 f_z(z) V_{\omega,0}^{(2)} n^{1/2} \left(\frac{1}{2}\mu_3(K)\beta_3^{(2)}(z) \right) (1 + o_p(1))
\end{aligned}
$$

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

$$
B_{m,n}(z) = h^2 f_z(z) S_W(z) D_{m,n} \left(\frac{1}{2}\mu_2(K)\beta^{(2)}(z) \right) (1 + o_p(1)) \tag{30a}
$$

$$
B_n(z) = h^2 f_z(z) S_N(z) D_n \left(\frac{1}{2}\mu_2(K)\beta^{(2)}(z) \right) (1 + o_p(1)) \tag{30b}
$$

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

$$
\begin{aligned}
D_{m,n}^{-1} E_{m,n,1} &= h^2 B_\beta(z) + o_p(h^2) \\
D_n^{-1} E_{n,1} &= h^2 B_\beta(z) + o_p(h^2)
\end{aligned}
$$

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

$$
\begin{aligned}
E_{m,n,2} &= D_{m,n} \left(\widehat{\beta}(z) - \beta(z) - D_{m,n}^{-1} E_{m,n,1} \right) \\
&= D_{m,n} \left(\widehat{\beta}(z) - \beta(z) - h^2 B_\beta(z) + o_p(h^2) \right) \\
E_{n,2} &= D_n \left(\widehat{\beta}(z) - \beta(z) - D_n^{-1} E_{n,1} \right) \\
&= D_n \left(\widehat{\beta}(z) - \beta(z) - h^2 B_\beta(z) + o_p(h^2) \right)
\end{aligned}
$$

In this regard, define

$$
\begin{aligned}
T_{m,n}(z) &= \begin{pmatrix} T_{m,n,1}(z) \\ T_{m,n,2}(z) \end{pmatrix} = \sqrt{\frac{h}{n}} \sum_{t=1}^{n} K_h(Z_t - z)\epsilon_t D_{m,n}^{-1} X_t \\
T_n(z) &= \begin{pmatrix} T_{n,1}(z) \\ T_{n,2}(z) \\ T_{n,3}(z) \end{pmatrix} = \sqrt{\frac{h}{n}} \sum_{t=1}^{n} K_h(Z_t - z)\epsilon_t D_n^{-1} X_t
\end{aligned}
$$

where

$$
\begin{aligned}
T_{m,n,1} &= T_{n,1} \\
&= \sqrt{\frac{h}{n}} \sum_{t=1}^{n} K_h(Z_t - z)\epsilon_t X_{t1} \\
T_{m,n,2} &= m^{1/2} T_{n,2} \\
&= \sqrt{\frac{h}{n}} \sum_{t=1}^{n} K_h(Z_t - z)\epsilon_t m^{1/2} n^{-1/2} X_{t2} \\
T_{n,3} &= \sqrt{\frac{h}{n}} \sum_{t=1}^{n} K_h(Z_t - z)\epsilon_t n^{-1/2} X_{t3}
\end{aligned}
$$

and note that

$$
\sqrt{nh} D_{m,n} \left(\widehat{\beta}(z) - \beta(z) - h^2 B_\beta(z) + o_p(h^2) \right) = R_{m,n}(z)^{-1} T_{m,n}(z) \tag{31a}
$$

$$
\sqrt{nh} D_{n} \left(\widehat{\beta}(z) - \beta(z) - h^2 B_\beta(z) + o_p(h^2) \right) = R_{n}(z)^{-1} T_{n}(z) \tag{31b}
$$

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

$$
\begin{aligned}
T_{m,n,1}(z) = T_{n,1}(z) &\longrightarrow_d \mathbf{N}\left(0, \sigma_\epsilon^2 \nu_0(K) f_z(z) M_2(z)\right) \\
&= \sqrt{\nu_0(K) f_z(z)} W_\epsilon(1)
\end{aligned} \tag{32}
$$

where $W_\epsilon(1)$ is a d_1-dimensional Brownian motion on $[0,1]$ with covariance matrix $\sigma_\epsilon^2 M_2(z)$. Moreover, since the first element of X_{t1} is unity, it follows immediately that

$$
\begin{aligned}
\sqrt{\frac{h}{n}} \sum_{t=1}^{n} K_h(Z_t - z)\epsilon_t &\longrightarrow_d \mathbf{N}\left(0, \sigma_\epsilon^2 \nu_0(K) f_z(z)\right) \\
&= \sqrt{\nu_0(K) f_z(z)} W_{\epsilon,1}(1)
\end{aligned}
$$

where $W_{\epsilon,1}(r)$ is the first element of $W_\epsilon(r)$. Moreover, by equation (8b) it now follows that

$$
T_{m,n,2} \longrightarrow_d \sqrt{\nu_0(K) f_z(z)} \left(\int_{-\infty}^{\infty} x^2 D(x) dx \right)^{1/2} W_{\epsilon,1}(1) \tag{33}
$$

Now, putting equations (32) and (33) together implies that

$$
T_{m,n}(z) \longrightarrow_d \sqrt{\nu_0(K) f_z(z)} \begin{pmatrix} W_\epsilon(1) \\ \left(\int_{-\infty}^{\infty} x^2 D(x) dx \right)^{1/2} W_{\epsilon,1}(1) \end{pmatrix}
$$

It is now clear that $T_{m,n}(z)$ has a multivariate normal distribution with the conditional covariance matrix of $\begin{pmatrix} W_\epsilon(1) \\ \left(\int_{-\infty}^{\infty} x^2 D(x) dx \right)^{1/2} W_{\epsilon,1}(1) \end{pmatrix}$ being

$$\sigma_\epsilon^2 \begin{pmatrix} M_2(z) & M_1(z)V_D^{(1)^\top} \\ V_D^{(1)}M_1(z)^\top & V_D^{(2)} \end{pmatrix} = \sigma_\epsilon^2 S_W(z)$$

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

$$T_{n,2} \longrightarrow_d \sqrt{\nu_0(K)f_z(z)} \int_{-\infty}^{\infty} V_{\eta,c}(r)dW_{\epsilon,1}(r) \tag{34a}$$

$$T_{n,3} \longrightarrow_d \sqrt{\nu_0(K)f_z(z)} \int_{-\infty}^{\infty} V_{\omega,0}(r)dW_{\epsilon,1}(r) \tag{34b}$$

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

$$T_n(z) \longrightarrow_d \sqrt{\nu_0(K)f_z(z)} \begin{pmatrix} W_\epsilon(1) \\ \int_{-\infty}^{\infty} V_{\eta,c}(r)dW_{\epsilon,1}(r) \\ \int_{-\infty}^{\infty} V_{\omega,0}(r)dW_{\epsilon,1}(r) \end{pmatrix}$$

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

$$\sigma_\epsilon^2 \begin{pmatrix} A_2(z) & A_1(z)V_{\eta,c}^{(1)\top} & A_1(z)V_{\omega,0}^{(1)} \\ V_{\eta,c}^{(1)}A_1(z)^\top & V_{\eta,c}^{(2)} & V_{\eta,c,\omega,0} \\ V_{\omega,0}^{(1)}A_1(z)^\top & V_{\eta,c,\omega,0} & V_{\omega,0}^{(2)} \end{pmatrix} = \sigma_\epsilon^2 S_N(z)$$

Finally, invoking Slutsky's theorem implies that

$$\begin{aligned} &\sqrt{nh}D_{m,n}\left(\widehat{\beta}(z) - \beta(z) - h^2 B_\beta(z) + o_p(h^2)\right) \\ &\longrightarrow_d f_z^{-1/2}(z)\nu_0^{1/2}(K)S_W^{-1}(z) \begin{pmatrix} W_\epsilon(1) \\ \left(\int_{-\infty}^{\infty} x^2 D(x)dx\right)^{1/2} W_{\epsilon,1}(1) \end{pmatrix} \end{aligned}$$

and

$$\begin{aligned} &\sqrt{nh}D_n\left(\widehat{\beta}(z) - \beta(z) - h^2 B_\beta(z) + o_p(h^2)\right) \\ &\longrightarrow_d f_z^{-1/2}(z)\nu_0^{1/2}(K)S_N^{-1}(z) \begin{pmatrix} W_\epsilon(1) \\ \int_{-\infty}^{\infty} V_{\eta,c}(r)dW_{\epsilon,1}(r) \\ \int_{-\infty}^{\infty} V_{\omega,0}(r)dW_{\epsilon,1}(r) \end{pmatrix} \end{aligned}$$

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

Appendix B

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

$$\begin{aligned} S_n(z) &= \frac{1}{n^{3/2}} \sum_{t=1}^{n} K_h(Z_t - z) \begin{pmatrix} 1 \\ Z_{t,z,h} \end{pmatrix} \otimes X_t^{\otimes 2} \\ &= \begin{pmatrix} S_{n,0}(z) & S_{n,1}(z) \\ S_{n,1}(z) & S_{n,2}(z) \end{pmatrix} \end{aligned}$$

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

$$S_{n,j} = \frac{c_n}{n} \sum_{t=1}^{n} X_{t,n}^{\otimes 2} K_j \left(c_n \left(Z_{t,n} - z_n\right)\right)$$

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 $z_n \longrightarrow 0$ 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 $\{\vartheta_{nk}\}$, $\{\vartheta_{nk}^{\star}\}$, and $\{\vartheta_{nk}^{\star\star}\}$ be sequences of random variables. Let $u_{k,n}$ be a process defined as*

$$u_{k,n} = f_n(\vartheta_{n1}, \ldots, \vartheta_{nk}; \vartheta_{n1}^{\star}, \ldots, \vartheta_{nk}^{\star}; \vartheta_{n1}^{\star\star}, \ldots, \vartheta_{nk}^{\star\star})$$

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

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

$$\begin{aligned} \max_{m \leq k \leq n} \left|\mathbf{E}\left(\vartheta_{n,k+1}^{\star} \,\middle|\, \mathcal{F}_{n,k}\right) - \sigma_{\vartheta^{\star}}^2\right| &\longrightarrow 0 \ a.s. \\ \max_{m \leq k \leq n} \left|\mathbf{E}\left(\vartheta_{n,k+1} \mid \mathcal{F}_{n,k}\right) - \sigma_{\vartheta}^2\right| &\longrightarrow 0 \ a.s. \end{aligned}$$

for some $\sigma_{\vartheta}^2 > 0$ and $\sigma_{\vartheta^{\star}}^2 > 0$, and for some $\delta > 0$

$$\max_{m \leq k \leq n} \left(\mathbf{E}\left(\left|\vartheta_{n,k+1}\right|^{2+\delta} \,\middle|\, \mathcal{F}_{n,k}\right) + \mathbf{E}\left(\left|\vartheta_{n,k+1}^{\star}\right|^{2+\delta} \,\middle|\, \mathcal{F}_{n,k}\right)\right) < \infty \ a.s.$$

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

(c) Let $\max_{1 \leq k \leq n} \left|u_{k,n}\right| = o_p(1)$ and $n^{-1/2} \sum_{k=1}^{n} \left|u_{k,n}\right| \left|\mathbf{E}\left(\vartheta_{n,k+1}^{\star} \vartheta_{n,k+1} \,\middle|\, \mathcal{F}_{n,k}\right)\right| = o_p(1)$.

(d) There exists a random variable $T(\vartheta^\star, \vartheta^{\star\star}) > 0$ such that $T_n^2 = \sum_{k=1}^n u_{k,n}^2 \longrightarrow_d T^2(\vartheta^\star, \vartheta^{\star\star})$ as $n \longrightarrow \infty$.

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

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 $c_n = \sqrt{n} h^{-1} \longrightarrow \infty$ such that $c_n n^{-1} \longrightarrow 0$ if z is fixed,*

$$c_n n^{-1} \sum_{t=1}^n X_{t,n} X_{t,n}^\top K_j \left(c_n \left(Z_{t,n} - z_n\right)\right) \longrightarrow_p \mu_j(K) \int_0^1 V_{\eta,c_x}(r) V_{\eta,c_x}(r)^\top dL_{V_{\xi,c_z}}(1,0)$$

$$c_n n^{-1} \sum_{t=1}^n X_{t,n} X_{t,n}^\top K_j^2 \left(c_n \left(Z_{t,n} - z_n\right)\right) \longrightarrow_p \nu_j(K) \int_0^1 V_{\eta,c_x}(r) V_{\eta,c_x}(r)^\top dL_{V_{\xi,c_z}}(1,0)$$

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

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 $c_z = 0$, $L_{V_{\xi,0}}(1, \cdot)$ 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 $c_n = \sqrt{n} h^{-1} \longrightarrow \infty$ such that $c_n n^{-1} \longrightarrow 0$ if z is fixed,*

$$\sigma_\epsilon^2 \frac{h}{n^{3/2}} \sum_{t=1}^n X_t X_t^\top K_{j,h}^2 \left(Z_t - z\right) \longrightarrow_d \sigma_\epsilon^2 \nu_j(K) \int_0^1 V_{\eta,c_x} V_{\eta,c_x}^\top dL_{V_{\xi,c_z}}(1,0)$$

Proof of Lemma B3. The proof is inspired by results in Phillips (2009), Sun et al. (2013) and Gao and Phillips (2013). Let $p_t(\cdot \mid X)$ denote the conditional density of Z_t given $X_t = X$ and define $q_t(\cdot \mid \cdot)$ as the conditional density of $\frac{Z_t}{\sqrt{t}}$ given $\frac{X_t}{\sqrt{t}} = \frac{X}{\sqrt{t}}$. Note further that a change of variables argument implies that $p_t(Z \mid X) = t^{-1/2} q_t\left(\frac{Z}{\sqrt{t}} \,\middle|\, \frac{X}{\sqrt{t}}\right)$. 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 $X_t^\star = J^\top X_t$. Finally, recall the assumption that $\sigma_\epsilon^2 = \mathbf{E}\left(\epsilon_t^2 \,\middle|\, Z_t, X_t\right)$ and observe that

$$
\begin{aligned}
&\frac{\mathbf{E}\left(\epsilon_t^2 \mid Z_t, X_t\right)}{n^3} \sum_{t=1}^{n} \mathbf{E}\left(X_t^{\star 4} K_{j,h}^2\left(Z_t - z\right)\right) \\
&= \frac{\sigma_\epsilon^2}{n^3 h^2} \sum_{t=1}^{n} \mathbf{E}\left(X_t^{\star 4} K_j^2\left(\frac{Z_t - z}{h}\right)\right) \\
&= \frac{\sigma_\epsilon^2}{n^3 h^2} \sum_{t=1}^{n} \mathbf{E}\left(X_t^{\star 4} \mathbf{E}\left(K_j^2\left(\frac{Z_t - z}{h}\right) \middle| X_t\right)\right) \\
&= \frac{\sigma_\epsilon^2}{n^3 h^2} \sum_{t=1}^{n} \mathbf{E}\left(X_t^{\star 4} \int_{\infty}^{\infty} K_j^2(y)\, q_t\left(\frac{yh + z}{\sqrt{t}} \middle| X_t\right) \frac{h}{\sqrt{t}} dy\right) \\
&\leq \frac{\sigma_\epsilon^2}{n^3 h} \sum_{t=1}^{n} \frac{1}{\sqrt{t}} = O\left(\frac{1}{n^{5/2} h}\right) = o(1) \qquad (35)
\end{aligned}
$$

Equation (35) follows since $\mathbf{E}\left(X_t^{\star 4}\right) = O(1)$, the integral expression is bounded and $\sqrt{n}h \longrightarrow \infty$. Next, let $p_{st}\left(\cdot \mid Z_s, X_s, X_t\right)$ denote the conditional density of $Z_t - Z_s$ given Z_s, X_s, X_t. Moreover, if $q_{st}\left(\cdot \mid Z_s, X_s, X_t\right)$ denotes the conditional density of $\frac{Z_t - Z_s}{\sqrt{t-s}}$ given Z_s, X_s, X_t, then $p_{st}\left(Z \mid Z_s, X_s, X_t\right) = (t-s)^{-1/2} q_{st}\left(\frac{Z}{\sqrt{t-s}} \middle| Z_s, X_s, X_t\right)$. Similar reasoning to equation (35) now yields the covariance result below.

$$
\begin{aligned}
&\frac{\mathbf{E}\left(\epsilon_t^2 \mid Z_t, X_t\right)}{n^3 h^2} \sum_{t=2}^{n} \sum_{s=1}^{t-1} \mathbf{E}\left(X_s^{\star 2} X_t^{\star 2} K_j\left(\frac{Z_s - z}{h}\right) K_j\left(\frac{Z_t - z}{h}\right)\right) \\
&= \frac{\sigma_\epsilon^2}{n^3 h^2} \sum_{t=2}^{n} \sum_{s=1}^{t-1} \mathbf{E}\left(X_s^{\star 2} X_t^{\star 2} \mathbf{E}\left(K_j\left(\frac{Z_s - z}{h}\right) K_j\left(\frac{Z_t - z}{h}\right) \middle| X_s, X_t\right)\right) \\
&= \frac{\sigma_\epsilon^2}{n^3 h^2} \sum_{t=2}^{n} \sum_{s=1}^{t-1} \mathbf{E}\left(X_s^{\star 2} X_t^{\star 2} \mathbf{E}\left(K_j\left(\frac{Z_s - z}{h}\right) \mathbf{E}\left(K_j\left(\frac{Z_t - Z_s}{h} + \frac{Z_s - z}{h}\right) \middle| Z_s, X_s, X_t\right)\right)\right) \\
&= \frac{\sigma_\epsilon^2}{n^3 h^2} \sum_{t=2}^{n} \sum_{s=1}^{t-1} \mathbf{E}\left(X_s^{\star 2} X_t^{\star 2}\right) \int_{\infty}^{\infty} K_j(y) \left(\int_{\infty}^{\infty} K_j(w + y)\, q_{st}\left(\frac{wh}{\sqrt{t-s}} \middle| X_s, X_t\right) \frac{h}{\sqrt{t-s}} dw\right) \\
&\times q_t\left(\frac{yh + z}{\sqrt{s}} \middle| X_s, X_t\right) \frac{h}{\sqrt{s}} dy \\
&\leq \frac{\sigma_\epsilon^2}{n^3} \sum_{t=2}^{n} \sum_{s=1}^{t-1} \frac{1}{\sqrt{s}} \frac{1}{\sqrt{t-s}} = O\left(\frac{1}{n^{3/2}}\right) = o(1) \qquad (36)
\end{aligned}
$$

Equations (35) and (36) now imply that

$$
\begin{aligned}
&\frac{\mathbf{E}\left(\epsilon_t^2 \mid Z_t, X_t\right)}{n^3 h^2} \mathbf{E}\left(\sum_{t=1}^{n} X_t^{\star} K_j\left(\frac{Z_t - z}{h}\right)\right)^2 \\
&= \frac{\sigma_\epsilon^2}{n^3 h^2} \sum_{t=1}^{n} \mathbf{E}\left(X_t^{\star 4} K_j^2\left(\frac{Z_t - z}{h}\right)\right) \\
&+ \frac{\sigma_\epsilon^2}{n^3 h^2} \sum_{t=2}^{n} \sum_{s=1}^{t-1} \mathbf{E}\left(X_s^{\star 2} X_t^{\star 2} K_j\left(\frac{Z_s - z}{h}\right) K_j\left(\frac{Z_t - z}{h}\right)\right) \\
&= o(1) \qquad (37)
\end{aligned}
$$

and Lemma B2 ensures that

$$\frac{\sigma_\epsilon^2}{n^{3/2}h}\sum_{t=1}^{n} X_t X_t^\top K_j^2\left(\frac{Z_t - z}{h}\right) \longrightarrow_d \sigma_\epsilon^2 \nu_j(K)\int_0^1 V_{\eta,c_x}V_{\eta,c_x}^\top dL_{V_{\xi,c_z}}(1,0) \tag{38}$$

$$\frac{\sigma_\epsilon^2}{n^{3/2}h}\sum_{t=1}^{n} X_t^\star X_t^{\star\top} K_j^2\left(\frac{Z_t - z}{h}\right) \longrightarrow_d \sigma_\epsilon^2 \nu_j(K)\int_0^1 V_{\eta,c_x}^{\star 2} dL_{V_{\xi,c_z}}(1,0) \tag{39}$$

where $V_{\eta,c_x}^\star$ is defined in the same way as V_{η,c_x} when η_t is replaced with $J^\top \eta_t$. 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 $c_n = \sqrt{n}h^{-1} \longrightarrow \infty$ such that $c_n n^{-1} \longrightarrow 0$ if z is fixed,*

$$\frac{1}{n^{3/2}}\sum_{t=1}^{n} X_t X_t^\top \left(\beta(Z_t) - \beta(z) - \beta^{(1)}(z)(Z_t - z)\right) K_{jh}(Z_t - z)$$
$$\longrightarrow_d h^2 B_\beta(z)\int_0^1 V_{\eta,c_x}V_{\eta,c_x}^\top dL_{V_{\xi,c_z}}(1,0) + o_p(h^2)$$

Proof of Lemma B4. The limiting form $h^2 B_\beta(z)\int_0^1 V_{\eta,c_x}V_{\eta,c_x}^\top dL_{V_{\xi,c_z}}(1,0)$ 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 $\mathbf{E}\left\{\sum_{t=1}^n X_t^{\star 2}\left(\beta(Z_s) - \beta(z) - \beta^{(1)}(z)(Z_s - z)\right)K_j\left(\frac{Z_t - z}{h}\right)\right\}^2$ is $O_p\left(h^4\right)$.

$$\begin{aligned}
&\frac{1}{n^3}\sum_{t=1}^{n}\mathbf{E}\left(X_t^{\star 4}\left(\beta(Z_t) - \beta(z) - \beta^{(1)}(z)(Z_t - z)\right)^2 K_{j,h}^2(Z_t - z)\right)\\
&= \frac{1}{n^3h^2}\sum_{t=1}^{n}\mathbf{E}\left(X_t^{\star 4}\left(\beta(Z_t) - \beta(z) - \beta^{(1)}(z)(Z_t - z)\right)^2 K_j^2\left(\frac{Z_t - z}{h}\right)\right)\\
&= \frac{1}{n^3h^2}\sum_{t=1}^{n}\mathbf{E}\left(X_t^{\star 4}\mathbf{E}\left(\left(\beta(Z_t) - \beta(z) - \beta^{(1)}(z)(Z_t - z)\right)^2 K_j^2\left(\frac{Z_t - z}{h}\right)\,\middle|\, X_t\right)\right)\\
&= \frac{1}{n^3h^2}\sum_{t=1}^{n}\mathbf{E}\left(X_t^{\star 4}\int_{\infty}^{\infty}\left(\beta(yh + z) - \beta(z) - \beta^{(1)}(z)yh\right)^2 K_j^2(y)\, q_t\left(\frac{yh+z}{\sqrt{t}}\,\middle|\, X_t\right)\frac{h}{\sqrt{t}}dy\right)\\
&= \frac{1}{n^3h^2}\sum_{t=1}^{n}\mathbf{E}\left(X_t^{\star 4}\int_{\infty}^{\infty}\left(\frac{1}{2}y^2h^2\beta^{(2)}(z) + o\left(h^2\right)\right)^2 K_j^2(y)\, q_t\left(\frac{yh+z}{\sqrt{t}}\,\middle|\, X_t\right)\frac{h}{\sqrt{t}}dy\right)\\
&\leq C\frac{h^3\beta^{(2)}(z) + o\left(h^3\right)}{n^3}\sum_{t=1}^{n}\frac{1}{\sqrt{t}} = \left(h^3\beta^{(2)}(z) + o\left(h^3\right)\right)O\left(\frac{1}{n^{5/2}}\right) = o\left(h^3\right)
\end{aligned} \tag{40}$$

for some positive constant C. Turning next to the covariance result, consider the following.

$$\frac{1}{n^3h^2}\sum_{t=2}^{n}\sum_{s=1}^{t-1}\mathbf{E}\left\{X_s^{\star 2}X_t^{\star 2}\left(\beta\left(Z_s\right)-\beta(z)-\beta^{(1)}(z)\left(Z_s-z\right)\right)\right.$$
$$\left.\times\left(\beta\left(Z_t\right)-\beta(z)-\beta^{(1)}(z)\left(Z_t-z\right)\right)K_j\left(\frac{Z_s-z}{h}\right)K_j\left(\frac{Z_t-z}{h}\right)\right\}$$
$$=\frac{1}{n^3h^2}\sum_{t=2}^{n}\sum_{s=1}^{t-1}\mathbf{E}\left\{X_s^{\star 2}X_t^{\star 2}\mathbf{E}\left(\left(\beta\left(Z_s\right)-\beta(z)-\beta^{(1)}(z)\left(Z_s-z\right)\right)\right.\right.$$
$$\left.\left.\times\left(\beta\left(Z_t\right)-\beta(z)-\beta^{(1)}(z)\left(Z_t-z\right)\right)K_j\left(\frac{Z_s-z}{h}\right)K_j\left(\frac{Z_t-z}{h}\right)\ \middle|\ X_s,X_t\right)\right\}$$
$$=\frac{1}{n^3h^2}\sum_{t=2}^{n}\sum_{s=1}^{t-1}\mathbf{E}\left\{X_s^{\star 2}X_t^{\star 2}\mathbf{E}\left(\left(\beta\left(Z_s\right)-\beta(z)-\beta^{(1)}(z)\left(Z_s-z\right)\right)\right.\right.$$
$$\times\left(\beta\left(Z_t-Z_s+Z_s\right)-\beta(z)-\beta^{(1)}(z)\left(Z_t-Z_s+Z_s-z\right)\right)$$
$$\left.\left.K_j\left(\frac{Z_s-z}{h}\right)K_j\left(\frac{Z_t-Z_s}{h}+\frac{Z_s-z}{h}\right)\ \middle|\ Z_s,X_s,X_t\right)\right\}$$
$$=\frac{1}{n^3h^2}\sum_{t=2}^{n}\sum_{s=1}^{t-1}\mathbf{E}\left\{X_s^{\star 2}X_t^{\star 2}\int_{-\infty}^{\infty}\left(\beta\left(yh+z\right)-\beta(z)-\beta^{(1)}(z)yh\right)K_j\left(y\right)\right.$$
$$\times\left[\int_{-\infty}^{\infty}\left(\beta\left((w+y)h+z\right)-\beta(z)-\beta^{(1)}(z)\left(w+yh\right)\right)\right.$$
$$\left.\left.\times\ K_j\left(w+y\right)q_{st}\left(\frac{wh}{\sqrt{t-s}}\ \middle|\ X_s,X_t\right)\frac{h}{\sqrt{t-s}}dw\right]q_t\left(\frac{yh+z}{\sqrt{s}}\ \middle|\ X_s,X_t\right)\frac{h}{\sqrt{s}}dy\right\}$$
$$=\frac{1}{n^3h^2}\sum_{t=2}^{n}\sum_{s=1}^{t-1}\mathbf{E}\left\{X_s^{\star 2}X_t^{\star 2}\int_{-\infty}^{\infty}\left(\frac{1}{2}y^2h^2\beta^{(2)}(z)+o\left(h^2\right)\right)K_j\left(y\right)\right.$$
$$\times\left[\int_{-\infty}^{\infty}\left(\frac{1}{2}(w+y)^2h^2\beta^{(2)}(z)+o\left(h^2\right)\right)\right.$$
$$\left.\left.\times\ K_j\left(w+y\right)q_{st}\left(\frac{wh}{\sqrt{t-s}}\ \middle|\ X_s,X_t\right)\frac{h}{\sqrt{t-s}}dw\right]q_t\left(\frac{yh+z}{\sqrt{s}}\ \middle|\ X_s,X_t\right)\frac{h}{\sqrt{s}}dy\right\}$$
$$\leq C\frac{h^4\beta^{(2)}(z)^2+o\left(h^4\right)}{n^3}\sum_{t=2}^{n}\sum_{s=1}^{t-1}\frac{1}{\sqrt{s}}\frac{1}{\sqrt{t-s}}=\left(h^4\beta^{(2)}(z)^2+o\left(h^4\right)\right)O\left(\frac{1}{n^{3/2}}\right)=o\left(h^4\right)\tag{41}$$

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

$$\frac{1}{n^3h^2}\mathbf{E}\left\{\sum_{t=1}^{n}X_t^{\star 2}\left(\beta\left(Z_s\right)-\beta(z)-\beta^{(1)}(z)\left(Z_s-z\right)\right)K_j\left(\frac{Z_t-z}{h}\right)\right\}^2$$
$$=\frac{1}{n^3h^2}\sum_{t=1}^{n}\mathbf{E}\left\{X_t^{\star 4}\left(\beta\left(Z_t\right)-\beta(z)-\beta^{(1)}(z)\left(Z_t-z\right)\right)^2K_j^2\left(\frac{Z_t-z}{h}\right)\right\}$$
$$+\frac{1}{n^3h^2}\sum_{t=2}^{n}\sum_{s=1}^{t-1}\mathbf{E}\left\{X_s^{\star 2}X_t^{\star 2}\left(\beta\left(Z_s\right)-\beta(z)-\beta^{(1)}(z)\left(Z_s-z\right)\right)\right.$$
$$\left.\times\left(\beta\left(Z_t\right)-\beta(z)-\beta^{(1)}(z)\left(Z_t-z\right)\right)K_j\left(\frac{Z_s-z}{h}\right)K_j\left(\frac{Z_t-z}{h}\right)\right\}$$
$$=o(h^3)+o(h^4)=o(h^4)\tag{42}$$

This completes the proof. □

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

$$\begin{aligned} S_n(z) &= \begin{pmatrix} S_{n,0}(z) & S_{n,1}(z) \\ S_{n,1}(z) & S_{n,2}(z) \end{pmatrix} \\ &= \begin{pmatrix} 1 & 0 \\ 0 & \mu_2(K) \end{pmatrix} \otimes \int_0^1 V_{\eta,c_x}(r) V_{\eta,c_x}(r)^\top dL_{V_{\xi,c_z}}(1,0) \end{aligned}$$

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

$$\begin{aligned} \widehat{\beta}(z) - \beta(z) &= \left(\int_0^1 V_{\eta,c_x}(r) V_{\eta,c_x}(r)^\top dL_{V_{\xi,c_z}}(1,0) \right)^{-1} \\ &\times \left\{ \frac{1}{n^{3/2}} \sum_{t=1}^n X_t X_t^\top \left(\beta\left(Z_t\right) - \beta(z) - \beta^{(1)}(z) \left(Z_t - z\right) \right) K_h\left(Z_t - z\right) \right. \\ &+ \left. \frac{1}{n^{3/2}} \sum_{t=1}^n X_t \epsilon_t K_h\left(Z_t - z\right) \right\} \\ &\equiv \left(\int_0^1 V_{\eta,c_x}(r) V_{\eta,c_x}(r)^\top dL_{V_{\xi,c_z}}(1,0) \right)^{-1} (B_1 + B_2) \end{aligned} \tag{43}$$

where

$$\begin{aligned} B_1 &= \frac{1}{n^{3/2}} \sum_{t=1}^n X_t X_t^\top \left(\beta\left(Z_t\right) - \beta(z) - \beta^{(1)}(z) \left(Z_t - z\right) \right) K_h\left(Z_t - z\right) \\ B_2 &= \frac{1}{n^{3/2}} \sum_{t=1}^n X_t \epsilon_t K_h\left(Z_t - z\right) \end{aligned}$$

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

$$\sqrt{hn^{3/2}} \left(\widehat{\beta}(z) - \beta(z) - h^2 B_\beta(z) + o_p(h^2) \right) = \sqrt{hn^{3/2}} B_2 + o_p(1) \tag{44}$$

and the theorem will follow if it can be shown that $\sqrt{hn^{3/2}} B_2 \longrightarrow_d MN\left(\Sigma_{B_2}\right)$ where by Lemma B 3, $\Sigma_{B_2} = \sigma_\epsilon^2 \nu_0(K) \int_0^1 V_{\eta,c_x}^{\star 2} dL_{V_{\xi,c_z}}(1,0)$. In this regard, recall from Lemma B2 that equation (39) implies that

$$\sigma_\epsilon^2 \frac{h}{n^{3/2}} \sum_{t=1}^n X_t^\star X_t^{\star\top} K_{0,h}^2\left(Z_t - z\right) \longrightarrow_d \sigma_\epsilon^2 \nu_0(K) \int_0^1 V_{\eta,c_x}^{\star 2} dL_{V_{\xi,c_z}}(1,0) \tag{45}$$

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

$$\sqrt{hn^{3/2}} B_2 = \frac{\sqrt{h}}{n^{3/4}} \sum_{t=1}^n X_t^\star \epsilon_t^\star K_h\left(Z_t - z\right) \longrightarrow_d MN\left(0, \Sigma_{B_2}\right) \tag{46}$$

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:

$$\begin{aligned}
\vartheta_{n,t+1} &= \epsilon_t \\
\vartheta^{\star}_{nt} &= J^{*\top}\varepsilon_t \\
\vartheta^{\star\star}_{nt} &= J^{*\top}\varepsilon_{1-t} \\
u_{t,n} &= \frac{1}{\sqrt{hn^{1/2}}}\frac{1}{\sqrt{n}}X^{*}_t K\left(\frac{Z_t - z}{h}\right)
\end{aligned}$$

where $J^{*\top} = \left(J_1^{*\top}, \ldots, J_d^{*\top}\right)$ is a real vector satisfying $J^{*\top}J^{*} = 1$ and $X^{*}_t = J^{*\top}X_t$. Moreover, let $\mathcal{F}_{n,t} = \sigma\left(\epsilon_t, \ldots, \epsilon_1; \varepsilon_t, \varepsilon_{t-1}, \ldots\right)$ 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

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

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