ANNALS OF ECONOMICS AND FINANCE **16-2**, 353–369 (2015)

Local Constant Kernel Estimation of a Partially Linear Varying Coefficient Cointegration Model

Luya Wang*

School of Banking and Finance, University of International Business and Economics

Zhongwen Liang

Department of Economics, University at Albany, SUNY, Albany, NY 12222, USA

Juan Lin

Department of Finance, School of Economics, and Wang Yanan Institute for Studies in Economics, Xiamen University, China

Qi Li†

Department of Economics, Texas A&M University
College Station, TX 77843, USA
School of Economics and Management
Tsinghua University, Beijing, 100084, China
E-mail: qi@econmail.tamu.edu

In this paper, we consider a partially linear varying coefficient cointegration model. We focus on the estimation of constant coefficients. We derive the saymptotic result for the local constant kernel estimator, which complements the results in Li, Li, Liang and Hsiao (2013) where the local polynomial estimation methods are studied. However, Li et al. (2013) impose stronger conditions to rule out the local constant estimation due to technical difficulties. We give the full treatment of the local constant method in this paper based on a novel proof. From the simulation results reported in the paper, we show that the local constant and local linear estimators perform similarly, but the local constant method requires less data. Also, in fnite sample applications the local linear estimation could suffer from the matrix singularity problem.

Key Words: Varying coefficient model; Partially linear model; Nonstationary; Cointegration.

1529-7373/2015

JEL Classification Numbers: C14, C32.

1. INTRODUCTION

Over the last three decades, there has been a surge of interest on both nonparametric econometric models and nonstationary time series. Recently, the nonparametric nonstationary models have drawn much attention, which requires the tools and results from handling both nonparametric models and models with non-stationary data. Similar as linear models, the results of nonparametric models involving integrated processes are quite different from those of regressions with stationary time series. In this paper, we consider a partially linear varying coefficient model

$$Y_t = X_{1t}'\gamma + X_{2t}'\beta(Z_t) + u_t, \qquad t = 1, \ldots, T, \tag{1}$$

where X_{1t} and X_{2t} are both multivariate integrated processes, and Z_t and u_t are univariate stationary processes.

A partially linear model enjoys the advantage of the direct economic interpretability of parametric coefficients and flexibility of modeling. It also alleviates the problem of "curse of dimensionality". Therefore, there is a large literature on theoretical analysis and empirical applications of partially linear models in econometrics and economics. In general, it is tempting to consider a standard partially linear model

$$Y_t = X_{1t}'\gamma + g(X_{2t}, Z_t) + u_t, \qquad t = 1, \ldots, T, \tag{2}$$

where g is an unknown smooth function which has both multivariate integrated process X_{2t} and a stationary process Z_t as arguments. However, the asymptotic theory of unit root processes which we will use in this paper is based the properties of Brownian motions. Up to now, the asymptotic theory of nonparametric cointegrating regression is entirely based on the recurrent property of Brownian motions (see e.g., Karlsen, Myklebust and Tjøstheim (2007), Park and Phillips (2001), Wang and Phillips (2009a), Wang and Phillips (2009b)). Unfortunately, Brownian motions are not recurrent when the dimension is greater than or equal to three. Also, the local time theory of the Brownian motions is not available for the dimension greater or equal to two. Thus the above-mentioned papers only consider the scalar case of integrated process in the nonparametric regression. In

* Luya Wang's research is supported by the Scientific Research Foundation of Graduate School of UIBE (No.201423).

† The corresponding author. Qi Li's research is partially supported by National Nature Science Foundation of China Key Project Grant # 71133001.

order to circumvent this problem and consider the multivariate case, we impose the additivity structure in the nonparametric part with varying coefficients. The varying coefficient cointegration models themselves are also very useful. They are considered by Cai, Li, and Park (2009) and Xiao (2009) recently. Another related paper is Juhl and Xiao (2005), who studied a partially linear autoregressive model with nonstationarity. Also, Li et al. (2014) consider the semiparametric estimation of the same model we investigate and use the local polynomial kernel method to estimate the finite dimensional parameter and the infinite dimensional function. However, the conditions they impose rule out the local constant kernel estimation due to the technical difficulty. Compared with the local polynomial kernel estimation, the local constant method is easier to implement and less likely to encounter singularity problems in finite sample applications.

In this paper, we adopt the standard framework for regression involving multivariate integrated processes from Park and Phillips (1988) and Park and Phillips (1989). We focus on the estimation of the constant coefficients γ. We construct our estimator from the local constant kernel method, and derive the T consistent asymptotic result for our estimator.

The rest of the paper is organized as follows. We discuss the model and conditions in Section 2. We give the local constant estimation and derive the asymptotic results in the Section 3. Monte Carlo simulation results are reported in Section 4, which shows that our estimator performs well in the finite sample applications. The proofs are relegated to the appendix.

2. A PARTIALLY LINEAR COINTEGRATION MODEL

We consider the partially linear cointegration model mentioned in introduction as following,

$$Y_t = X_{1t}'\gamma + X_{2t}'\beta(Z_t) + u_t, \qquad t = 1, \ldots, T, \tag{3}$$

where X_{1t} is a $d_1 \times 1$ vector of $I(1)$ variables, γ is a $d_1 \times 1$ vector of constant coefficients, X_{2t} is an $I(1)$ random vector with the dimension of $d_2 \times 1$, Z_t and u_t are stationary scalar processes (i.e., $I(0)$ variables), and $\beta(\cdot)$ is a smooth but unspecified function-valued vector with the dimension $d_2 \times 1$. The prime denotes the transpose.

Let $X_{1t} = X_{1,t-1} + \epsilon_t$ and $X_{2t} = X_{2,t-1} + v_t$, where ϵ_t and v_t are weakly dependent stationary vector processes which will be more specific later. Also, the initial points of the processes will not have impact on our asymptotic results, therefore, we assume X_{10} and X_{20} to be any $O_p(1)$ random variables including constants. We define $w_t' = (\epsilon_t', v_t', u_t)$, and construct a partial sum process $B_T(r) = T^{-1/2}\sum_{t=1}^{[Tr]} w_t$, where $[a]$ denotes the largest integer less than or equal to a. We require a multivariate invariance principle holds for $B_T(r)$ (see e.g. Phillips and Durlauf (1986)), then we have $B_T(r) \Rightarrow B(r)$ for $r \in [0,1]$, where $B(\cdot)$ is a Brownian motion, and " $\Rightarrow$ " represents weak convergence. We decompose $B_T(r)$, $B(r)$ and its variance covariance matrix Ω correspondingly with w_t, that is, $B_T(r)' = (B_{1T}(r)', B_{2T}(r)', B_{3T}(r))$, $B(r)' = (B_1(r)', B_2(r)', B_3(r))$ and

$$\Omega = \begin{pmatrix} \Omega_{11} & \Omega_{12} & \Omega_{13} \\ \Omega_{21} & \Omega_{22} & \Omega_{23} \\ \Omega_{31} & \Omega_{32} & \Omega_{33} \end{pmatrix} = \lim_{T\to\infty} T^{-1} E\left(\sum_{t=1}^{T} w_t \sum_{s=1}^{T} w_s'\right) = \Sigma + \Gamma + \Gamma',$$

where

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} & \Sigma_{13} \\ \Sigma_{21} & \Sigma_{22} & \Sigma_{23} \\ \Sigma_{31} & \Sigma_{32} & \Sigma_{33} \end{pmatrix} = \lim_{T\to\infty} T^{-1} \sum_{t=1}^{T} E(w_t w_t'),$$

and

$$\Gamma = \begin{pmatrix} \Gamma_{11} & \Gamma_{12} & \Gamma_{13} \\ \Gamma_{21} & \Gamma_{22} & \Gamma_{23} \\ \Gamma_{31} & \Gamma_{32} & \Gamma_{33} \end{pmatrix} = \lim_{T\to\infty} \sum_{t=2}^{T} \sum_{j=1}^{t-1} E(w_j w_t').$$

As $T \to \infty$, by the continuous mapping theorem, we have (see e.g. Park and Phillips (1988) Lemma 2.1 (c) or Phillips and Durlauf (1986))

$$T^{-1} \sum_{t=1}^{T} \left(\frac{X_{it}}{\sqrt{T}}\right)\left(\frac{X_{it}}{\sqrt{T}}\right)' \Rightarrow \int_0^1 B_i(r) B_i(r)' dr \stackrel{def}{=} B_{(i)}, \qquad \text{for } i = 1, 2, \tag{4}$$

and

$$T^{-1} \sum_{t=1}^{T} \left(\frac{X_{2t}}{\sqrt{T}}\right)\left(\frac{X_{1t}}{\sqrt{T}}\right)' \Rightarrow \int_0^1 B_2(r) B_1(r)' dr \stackrel{def}{=} B_{(2,1)}. \tag{5}$$

Also, we have (see e.g. Park and Phillips (1988) Lemma 2.1 (e))

$$T^{-1} \sum_{t=1}^{T} X_{1t} u_t \Rightarrow \int_0^1 B_1 dB_3 + \Delta_{13} \qquad \text{and} \qquad T^{-1} \sum_{t=1}^{T} X_{2t} u_t \Rightarrow \int_0^1 B_2 dB_3 + \Delta_{23}, \tag{6}$$

where $\Delta_{13} = \Sigma_{13} + \Gamma_{13}$ and $\Delta_{23} = \Sigma_{23} + \Gamma_{23}$. The joint weak convergence of (4), (5) and (6) also holds.

Since (3) is a partially linear varying coefficient model, the estimation involves nonparametric kernel method. We first give some notations and assumptions. Let $f(z)$ be the probability density function of Z_t. Denote $K_{h,tz} = h^{-1} K((Z_t - z)/h)$, where $K(\cdot)$ is a kernel density function such that $\int K(u) du = 1$ and h is the smoothing parameter satisfying $h \to 0$, as

$T \to \infty$. Also, write $\beta^{(1)}(z) = \frac{d\beta(z)}{dz}$, and define $\mu_j = \int_{-\infty}^{\infty} u^j K(u)du$ and $\nu_j = \int_{-\infty}^{\infty} u^j K^2(u)du$, $j = 0, 1, 2$.

We make the following assumptions:

Assumption 1. *For some $p > \tau > 2$, $w_t = (\epsilon_t', v_t', u_t)$ is a strictly stationary, strong mixing sequence with zero mean and mixing coefficients α_m of size $-p\tau/(p-\tau)$ and $\sup_{t\geq 1} \|w_t\|_p = M < \infty$. In addition,*

$$T^{-1}E(\sum_{t=1}^{T} w_t \sum_{t=1}^{T} w_t') \to \Omega < \infty \text{ as } T \to \infty,$$

where Ω is a positive definite matrix.

Assumption 2. *$(u_t, \mathcal{F}_t, 1 \leq t \leq T)$ is a martingale difference sequence with $E(u_t^2|\mathcal{F}_{t-1}) \overset{a.s.}{\to} \sigma_u^2$ as $t \to \infty$, where $\mathcal{F}_t$ is the σ-algebra generated by $\{\epsilon_s, v_s, Z_s\}$, $1 \leq s \leq T$ and $\{u_s\}$, $s \leq t$.*

Assumption 3. *Z_t has a compact support $\mathcal{S}_z$. Z_t is a strictly stationary process with mixing coefficients α_m of size $-\gamma\delta/(\gamma-\delta)$ and $\sup_t \|Z_t\|_\gamma < M < \infty$, where $\gamma > \delta > 2$. Also, $\|\beta(Z_t)\|_{2+\varpi} < M < \infty$ for all t and some $\varpi > 0$. $\beta(z)$ is three times differentiable, and $\beta(z)$ and all its derivatives are bounded uniformly over z in the domain of Z_t.*

Assumption 4. *$f(z)$ has bounded and continuous derivatives up to third order uniformly over z in the domain of Z_t, and $\inf_{z\in\mathcal{S}_z} f(z) > 0$. Also $f(u, v; l_1)$ is bounded for all $l_1 \geq 1$ where $f(u, v; l_1)$ is the joint density function of (Z_0, Z_{l_1}) evaluated at $(Z_0, Z_{l_1}) = (u, v)$.*

Assumption 5. *$K(\cdot)$ is a bounded probability density function, which is symmetric around zero. $\int |K(u)|du \leq M_1 < \infty$, $\int u^{2p}K(u)du < \infty$, and $\int u^{2p}K^2(u)du < \infty$. For some $M_2 < \infty$ and $M_3 < \infty$, either $K(u) = 0$ for $|u| > M_3$ and for any $u_1, u_2 \in \mathbf{R}$, $|K(u_1) - K(u_2)| \leq M_2|u_1 - u_2|$, or $K(\cdot)$ is differentiable, $|(\partial/\partial u)K(u)| \leq M_2$, and for some $\iota > 1$, $|(\partial/\partial u)K(u)| \leq M_2|u|^{-\iota}$ for $|u| > M_3$.*

Assumption 6. *$h \to 0$, $(Th)/\ln T \to \infty$ and $Th^2 \to 0$, as $T \to \infty$.*

Remark 2.1. Assumption 1 is the same as Assumption 1 given in Hansen (1992) which ensures an invariance principle, (4), (5) and (6) and their joint convergence to hold (see also Phillips and Durlauf (1986)). Assumption 2 is a strict exogeneity condition which gives $\Delta_{13} = 0$ and $\Delta_{23} = 0$, since $\Delta_{13} = \Sigma_{13} + \Gamma_{13} = \lim_{T\to\infty}\sum_{t=1}^{T} E(\epsilon_t u_t) + \lim_{T\to\infty}\sum_{t=2}^{t}\sum_{j=1}^{t-1} E(\epsilon_t u_t) = \lim_{T\to\infty}\sum_{t=1}^{T} E(\epsilon_t E(u_t|\mathcal{F}_{t-1})) + \lim_{T\to\infty}\sum_{t=2}^{t}\sum_{j=1}^{t-1} E(\epsilon_t E(u_t|\mathcal{F}_{t-1})) = 0$ by the law of iterated expectations and similarly it holds for Δ_{23}. Assumptions 3-6 ensure the uniform convergence results hold as in Hansen (2008).

3. LOCAL CONSTANT KERNEL ESTIMATION

Similar to Robinson (1988), we propose to use a profile least squares approach to estimate γ. First we treat γ as if it were known and rewrite (3) as

$$Y_t - X_{1t}'\gamma = X_{2t}'\beta(Z_t) + u_t. \tag{7}$$

Then we could estimate $\beta(Z_t)$ by the local constant kernel method, i.e.,

$$\tilde{\beta}_{lc}(Z_t) = \left[\sum_s X_{2s}X_{2s}'K_{h,st}\right]^{-1}\left[\sum_s X_{2s}(Y_s - X_{1s}'\gamma)K_{h,st}\right] \stackrel{def}{=} A_{2t} - A_{1t}\gamma, \tag{8}$$

where

$$A_{1t} = [\sum_s X_{2s}X_{2s}'K_{h,st}]^{-1}\sum_s X_{2s}X_{1s}'K_{h,st}, \tag{9}$$

$$A_{2t} = [\sum_s X_{2s}X_{2s}'K_{h,st}]^{-1}\sum_s X_{2s}Y_sK_{h,st}, \tag{10}$$

$K_{h,st} = h^{-1}K((Z_s - Z_t)/h)$ is the kernel function, and h is the smoothing parameter. However, it should be mentioned that $\tilde{\beta}_{lc}(Z_t)$ defined in (8) is infeasible as it depends on the unknown parameter γ. We will provide a feasible estimator for $\beta(Z_t)$ after we get a consistent estimator of γ.

Replacing $\beta(Z_t)$ by $\tilde{\beta}_{lc}(Z_t)$ from (8) and re-arranging terms, we obtain

$$Y_t - X_{2t}'A_{2t} = (X_{1t}' - X_{2t}'A_{1t})\gamma + \eta_t, \tag{11}$$

where $\eta_t \equiv Y_t - X'_{2t}A_{2t} - (X'_{1t} - X'_{2t}A_{1t})\gamma$. Applying the OLS method to the above equation leads to

$$\hat{\gamma}_{lc} = \left[\sum_t (X'_{1t} - X'_{2t}A_{1t})'(X'_{1t} - X'_{2t}A_{1t})\right]^{-1} \sum_t (X'_{1t} - X'_{2t}A_{1t})'(Y_t - X'_{2t}A_{2t}). \tag{12}$$

Hence, we get the local constant kernel estimator for γ. The next theorem gives the asymptotic distribution of $\hat{\gamma}_{lc}$.

THEOREM 1. *Define* $\Omega_{1,2}(r) = B_1(r)' - B_2(r)'B_{(2)}^{-1}B_{(2,1)}$. *Under Assumptions 1 to 6, we have*

$$T(\hat{\gamma}_{lc} - \gamma) \xrightarrow{d} \left[\int_0^1 (\Omega_{1,2})^{\otimes 2} dr\right]^{-1} \int_0^1 \Omega'_{1,2} dB_3(r),$$

where $B_{(2)} = \int_0^1 B_2(r)B_2(r)'dr$, $B_{(2,1)} = \int_0^1 B_2(r)B_1(r)'dr$, *and* $A^{\otimes 2} = AA'$ *for any matrix* A.

Then a feasible estimator of $\beta(z)$ is given by

$$\hat{\beta}_{lc}(z) = \left[\sum_{s=1}^T K_{h,sz}X_{2s}X'_{2s}\right]^{-1} \sum_{s=1}^T K_{h,sz}X_{2s}(Y_s - X'_{1s}\hat{\gamma}_{lc}), \tag{13}$$

where $K_{h,sz} = h^{-1}K((Z_s - z)/h)$. The asymptotic distribution of $\hat{\beta}_{lc}(z)$ is similar to that is given in Cai et al. (2009) and Li et al. (2014).

4. MONTE CARLO SIMULATIONS

We conduct some Monte Carlo simulation experiments to show the finite sample performance of our estimator. We consider the following two data generating processes (DGP):

$$DGP1: \quad Y_t = X_{1t}\gamma + X_{2t}\beta_1(Z_t) + u_t,$$
$$DGP2: \quad Y_t = X_{1t}\gamma + X_{2t}\beta_2(Z_t) + u_t,$$

where $\beta_1(z) = 2 + z^3$, $\beta_2(z) = 1 + sin(6\pi z)$, $X_{1t} = X_{1,t-1} + v_{1t}$, $X_{2t} = X_{2,t-1} + v_{2t}$, $Z_t = w_{t-1} + w_t$, v_{1t}, v_{2t} and u_t are i.i.d. with $N(0,1)$, and w_t are i.i.d. with uniform[0,1]. We choose the sample sizes as $T =$ 50, 100, 200 and 400, respectively. We compute the square-root of the average squared error (RASE) for γ and the RASE for $\hat{\beta}_i(\cdot)$ $i = 1, 2$ as follows: for each replication we compute $RASE_{\gamma,j} = |\hat{\gamma}_j - \gamma|$, $RASE_{\beta,j} =$

$\sqrt{T^{-1}\sum_{t=1}^{T}[\hat{\beta}_i(Z_t)-\beta_i(Z_t)]^2}$. Then we obtain the average of $RASE_{\gamma,j}$ and $RASE_{\beta,j}$ over the 2000 replications.

We use the standard normal kernel, and the smoothing parameters are first selected by an ad-hoc method: $h_{ad-hoc} = z_{sd}T^{-1/\alpha}$ (where we choose $\alpha = 1.8$ to satisfy the T consistent condition for $\hat{\gamma}$ with the local constant method, and for the local linear method we use $\alpha = 2.8$ due to the matrix singularity problem), where z_{sd} is the sample standard deviation of $\{Z_t\}_{t=1}^{T}$. We also select h by using the least squares cross validation (LS-CV) method and we denote it by $\hat{h}_{CV}$. We report simulation results for RASEs for $\hat{\gamma}$ and β_j $(j = 1, 2)$ in Table 1 and Table 2, respectively.

TABLE 1.

RASE of the estimation of γ

	Local constant				Local linear			
	DGP1		DGP2		DGP1		DGP2	
n	h_{ad-hoc}	h_{CV}	h_{ad-hoc}	h_{CV}	h_{ad-hoc}	h_{CV}	h_{ad-hoc}	h_{CV}
50	0.0762	0.0745	0.1001	0.0910	0.0664	0.0657	0.1685	0.0973
100	0.0358	0.0350	0.0416	0.0397	0.0308	0.0305	0.0810	0.0384
200	0.0183	0.0181	0.0203	0.0197	0.0162	0.0160	0.0335	0.0176
400	0.0095	0.0093	0.0099	0.0098	0.0083	0.0083	0.0131	0.0089

TABLE 2.

RASE of the estimation of β_1 and β_2

	Local constant				Local linear			
	β_1		β_2		β_1		β_2	
n	h_{ad-hoc}	h_{CV}	h_{ad-hoc}	h_{CV}	h_{ad-hoc}	h_{CV}	h_{ad-hoc}	h_{CV}
50	0.1757	0.1630	0.3463	0.3463	0.2262	0.1873	0.6003	0.5686
100	0.0988	0.0940	0.2015	0.1870	0.1020	0.0855	0.4623	0.3333
200	0.0608	0.0567	0.1157	0.1081	0.0518	0.0501	0.3377	0.3025
400	0.0350	0.0324	0.0660	0.0660	0.0280	0.0299	0.2340	0.1535

From the tables we can see that the estimator $\hat{\gamma}$ is convergent with T^{-1} rate, and $\hat{\beta}_1(\cdot)$ and $\hat{\beta}_2(\cdot)$ are consistent estimators. In general, local linear estimators perform better than local constant estimators. However, when the local constant estimators are used together with the LS-CV method for the smoothing parameters selection, the local constant estimators also perform closely with the local linear estimators. The advantage of the local constant estimator is that it is easier to compute, and requires less data. Further, we find that the local linear estimation could suffer severely from the matrix singularity problem.

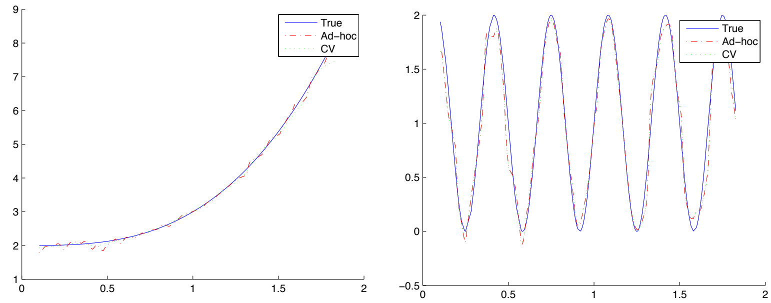

FIG. 1. Plots of the local constant estimation of β_1 and β_2

APPENDIX A

LEMMA 1. *Under the Assumptions 1 and 2, we have that*

$$B_T(r) = \begin{pmatrix} \frac{1}{\sqrt{T}} \sum_{t=1}^{[Tr]} \epsilon_t \\ \frac{1}{\sqrt{T}} \sum_{t=1}^{[Tr]} v_t \\ \frac{1}{\sqrt{T}} \sum_{t=1}^{[Tr]} u_t \end{pmatrix} \Rightarrow \begin{pmatrix} B_1(r) \\ B_2(r) \\ B_3(r) \end{pmatrix} = B(r), \tag{A.1}$$

where $B(r) = (B_1(r)', B_2(r)', B_3(r))'$ is a Brownian motion with covariance matrix

$$\Omega = \begin{pmatrix} \Omega_{11} & 0 & 0 \\ 0 & \Omega_{22} & 0 \\ 0 & 0 & \sigma_u^2 \end{pmatrix}.$$

Lemma 1 is a standard result which could be found in Phillips and Durlauf (1986) or Hansen (1992).

First, we strengthen the weak convergence result $B_T(r) \Rightarrow B(r)$ in Lemma 1 to a strong convergence result. By Skorohod-Dudley-Wichura representation theorem (e.g., Shorack and Wellner, 1986, Rmk. 2, p. 49), on an extended probability space, there exists a distributionally equivalent sequence $B_T^*(r)$ such that

$$\begin{pmatrix} \frac{X_{1t}}{\sqrt{T}} \\ \frac{X_{2t}}{\sqrt{T}} \\ \frac{\sum_{t=1}^{[Tr]} u_t}{\sqrt{T}} \end{pmatrix} = \begin{pmatrix} T^{-1/2}\sum_{t=1}^{[Tr]} \epsilon_t \\ T^{-1/2}\sum_{t=1}^{[Tr]} v_t \\ T^{-1/2}\sum_{t=1}^{[Tr]} u_t \end{pmatrix} =_d B_T^*(r) \overset{a.s.}{\to} \begin{pmatrix} B_1(r) \\ B_2(r) \\ B_3(r) \end{pmatrix} \tag{A.2}$$

where $A =_d B$ denotes that A and B follow the same distribution. Since we are only interested in weak convergence, we will not distinguish two random elements which have the same distribution.

Let β_s denote $\beta(Z_s)$. Substituting Y_s by $Y_s = X_{1s}'\gamma + X_{2s}'\beta_s + u_s$ in A_{2t} of (10) leads to

$$A_{2t} = A_{1t}\gamma + A_{3t} + A_{4t}, \tag{A.3}$$

where

$$\begin{aligned} A_{1t} &= [\sum_s X_{2s}X_{2s}'K_{h,st}]^{-1}\sum_s X_{2s}X_{1s}'K_{h,st}, \\ A_{3t} &= [\sum_s X_{2s}X_{2s}'K_{h,st}]^{-1}\sum_s X_{2s}X_{2s}'\beta_s K_{h,st}, \\ A_{4t} &= [\sum_s X_{2s}X_{2s}'K_{h,st}]^{-1}\sum_s X_{2s}u_s K_{h,st}. \end{aligned}$$

Combining (3) and (A.3), we obtain

$$Y_t - X_{2t}'A_{2t} = (X_{1t}' - X_{2t}'A_{1t})\gamma + X_{2t}'\beta_t - X_{2t}'A_{3t} + u_t - X_{2t}'A_{4t}.$$

Substituting the above result into (12) gives

$$\begin{aligned} \hat{\gamma}_{lc} - \gamma &= \left[\sum_t (X_{1t}' - X_{2t}'A_{1t})'(X_{1t}' - X_{2t}'A_{1t})\right]^{-1} \\ &\times \sum_t (X_{1t}' - X_{2t}'A_{1t})' \left[u_t + X_{2t}'(\beta_t - A_{3t}) - X_{2t}'A_{4t}\right]. \end{aligned} \tag{A.4}$$

We give a lemma before we give the proof of Theorem 1.

Lemma 2. *Under the Assumptions 1 to 6, we have*

(i) $B_{1T} \overset{def}{=} T^{-2}\sum_t (X_{1t}' - X_{2t}'A_{1t})'(X_{1t}' - X_{2t}'A_{1t})$
$\Rightarrow \int_0^1 [B_1(r)' - B_2(r)'B_{(2)}^{-1}B_{(2,1)}]^{\otimes 2} dr,$

(ii) $B_{2T} \stackrel{def}{=} T^{-1}\sum_t (X'_{1t} - X'_{2t}A_{1t})'u_t$
$\Rightarrow \int_0^1 [B_1(r)' - B_2(r)'B_{(2)}^{-1}B_{(2,1)}]'dB_3(r)$,

(iii) $B_{3T} \stackrel{def}{=} T^{-1}\sum_t (X'_{1t} - X'_{2t}A_{1t})'X'_{2t}(\beta_t - A_{3t}) = o_p(1)$,

(iv) $B_{4T} \stackrel{def}{=} -T^{-1}\sum_t (X'_{1t} - X'_{2t}A_{1t})'X'_{2t}A_{4t} = o_p(1)$.

Also, the joint convergence of (i), (ii), (iii) and (iv) hold by the joint convergence in probability.

Proof of (i): Let $K_{h,sz} = \frac{1}{h}K\left(\frac{Z_s - z}{h}\right)$. We have

$$\begin{aligned}\frac{1}{T^2}\sum_{s=1}^{T} X_{2s}X'_{2s}K_{h,sz} &= \frac{1}{T^2}\sum_{s=1}^{T} X_{2s}X'_{2s}E[K_{h,sz}] \\ &+ \frac{1}{T^2}\sum_{s=1}^{T} X_{2s}X'_{2s}[K_{h,sz} - E[K_{h,sz}]] \\ &\stackrel{def}{=} C_{1T}(z) + C_{2T}(z). \end{aligned} \tag{A.5}$$

Following Proposition 1 of Masry (1996), we have that $\sup_{z\in\mathcal{S}_z} |E[K_{h,sz}] - f(z)| = O_p(h^2)$.

Since $\frac{1}{T^2}\sum_{s=1}^{T} X_{s2}X'_{s2} = \int_0^1 B_2B'_2dr + o_p(1)$, we have

$$\sup_{z\in\mathcal{S}_z} \|C_{1T}(z) - f(z)\int_0^1 B_2B'_2dr\| = o_p(1). \tag{A.6}$$

Following the exact same steps in the proof of Theorem 1 in Gu and Liang (2014), we have that

$$\sup_{z\in\mathcal{S}_z} \|C_{2T}(z)\| = o_p(1). \tag{A.7}$$

Further, we have

$$\begin{aligned}\frac{1}{T^2}\sum_{s=1}^{T} X_{2s}X'_{1s}K_{h,sz} &= \frac{1}{T^2}\sum_{s=1}^{T} X_{2s}X'_{1s}E[K_{h,sz}] \\ &+ \frac{1}{T^2}\sum_{s=1}^{T} X_{2s}X'_{1s}[K_{h,sz} - E[K_{h,sz}]] \\ &\stackrel{def}{=} C_{3T}(z) + C_{4T}(z). \end{aligned} \tag{A.8}$$

Similar as (A.6) and (A.7), we have

$$\sup_{z\in\mathcal{S}_z}\|C_{3T}(z)-f(z)\int_0^1 B_2B_1'dr\| \;=\; o_p(1), \qquad (A.9)$$
$$\sup_{z\in\mathcal{S}_z}\|C_{4T}(z)\| \;=\; o_p(1). \qquad (A.10)$$

Now we consider A_{1t}. By (A.6), (A.7), (A.9) and (A.10), we have

$$\begin{aligned} A_{1t} &= [T^{-2}\sum_s X_{2s}X_{2s}'K_{h,st}]^{-1}T^{-2}\sum_s X_{2s}X_{1s}'K_{h,st} \\ &= [C_{1T}(Z_t)+C_{2T}(Z_t)]^{-1}\,[C_{3T}(Z_t)+C_{4T}(Z_t)] \\ &= \Big[[B_{(2)}f(Z_t)]^{-1}+O_p(\sup_{z\in\mathcal{S}_z}\|C_{1T}(z)-f(z)B_{(2)}\|+\sup_{z\in\mathcal{S}_z}\|C_{2T}(z)\|)\Big]\Big[B_{(2,1)}f(Z_t) \\ &\quad +O_p(\sup_{z\in\mathcal{S}_z}\|C_{3T}(z)-f(z)B_{(2,1)}\|+\sup_{z\in\mathcal{S}_z}\|C_{4T}(z)\|)\Big] \\ &= B_{(2)}^{-1}B_{(2,1)}+o_p(1). \end{aligned} \qquad (A.11)$$

Hence,

$$\begin{aligned} B_{1T} &= T^{-2}\sum_t (X_{1t}-X_{2t}'A_{1t})(X_{1t}-X_{2t}'A_{1t})' \\ &\Rightarrow \int_0^1 [B_1(r)'-B_2(r)'B_{(2)}^{-1}B_{(2,1)}]^{\otimes 2}dr. \end{aligned}$$

Proof of (ii): We have

$$\begin{aligned} B_{2T} &= T^{-1}\sum_t (X_{1t}'-X_{2t}'A_{1t})'u_t \\ &= T^{-1}\sum_t \Big[X_{1t}'-X_{2t}'B_{(2)}^{-1}B_{(2,1)}\Big]'u_t+T^{-1}\sum_t\Big[X_{2t}'(A_{1t}-B_{(2)}^{-1}B_{(2,1)})\Big]'u_t \\ &\overset{def}{=} C_{5T}+C_{6T}. \end{aligned}$$

It is easy to see that from (6) we have

$$C_{5T}\Rightarrow \int_0^1 [B_1(r)'-B_2(r)'B_{(2)}^{-1}B_{(2,1)}]'dB_3(r). \qquad (A.12)$$

Next, we discuss C_{6T}. Let $L_t=(A_{1t}-B_{(2)}^{-1}B_{(2,1)})'X_{2t}/\sqrt{T}$, then $C_{6T}=T^{-1/2}\sum_t L_tu_t$. Denote $L_T(r)=L_{[Tr]}/\sqrt{T}$, then we have that $L_T(r)\xrightarrow{p}0$

by (A.2) and (A.10). From Assumption 2, we have that $\{u_t\}$ is a martingale difference process with respect to the filtration $\{\mathcal{F}_t\}$ and L_t is adapted to $\mathcal{F}_t$. Following Theorem 2.2 of Kurtz and Protter (1991), we have that $C_{6T} = T^{-1/2}\sum_t L_t u_t \xrightarrow{p} 0$. Therefore, $C_{6T} = o_p(1)$.

Proof of (iii): We have

$$\begin{aligned}
B_{3T} &= T^{-1}\sum_t (X'_{1t} - X'_{2t}A_{1t})' X'_{2t}(\beta_t - A_{3t}) \\
&= T^{-1}\sum_t (X_{1t}X'_{2t} - A'_{1t}X_{2t}X'_{2t})[T^{-2}\sum_s X_{2s}X'_{2s}K_{h,st}]^{-1}T^{-2}\sum_s X_{2s}X'_{2s}(\beta_t - \beta_s)K_{h,st} \\
&= T^{-1}\sum_{t\neq s}\sum (X_{1t}X'_{2t} - A'_{1t}X_{2t}X'_{2t})[T^{-2}\sum_l X_{2l}X'_{2l}K_{h,lt}]^{-1}T^{-2}X_{2s}X'_{2s}(\beta_t - \beta_s)K_{h,st} \\
&= T^{-1}\sum_{t>s}\sum (X_{1t}X'_{2t} - A'_{1t}X_{2t}X'_{2t})[T^{-2}\sum_l X_{2l}X'_{2l}K_{h,lt}]^{-1}T^{-2}X_{2s}X'_{2s}(\beta_t - \beta_s)K_{h,st} \\
&\quad + T^{-1}\sum_{t>s}\sum (X_{1s}X'_{2s} - A'_{1s}X_{2s}X'_{2s})[T^{-2}\sum_l X_{2l}X'_{2l}K_{h,ls}]^{-1}T^{-2}X_{2t}X'_{2t}(\beta_s - \beta_t)K_{h,ts} \\
&\stackrel{def}{=} B_{3T,1} + B_{3T,2}.
\end{aligned}$$

Denote $\xi_s = (f(Z_t))^{-1}\beta^{(1)}(Z_s)\frac{(Z_t - Z_s)}{h}K_{h,st}$. Let $\mathcal{G}_s = \sigma(\epsilon_l, v_l, 1 \le l \le T, Z_t, t \le s)$ be the smallest sigma-field generated by $\{\epsilon_l, v_l\}$, $1 \le l \le T$, and Z_t, $t \le s$, and denote $E(X|\mathcal{G}_s)$ by $E_s X$. Define

$$\zeta_s = \sum_{k=0}^{\infty}(E_s\xi_{s+k} - E_{s-1}\xi_{s+k}), \quad z_s = \sum_{k=1}^{\infty} E_s\xi_{s+k}.$$

Then we can see that ζ_s is a martingale difference sequence with respect to the filtration $\mathcal{G}_s$ and

$$\xi_s = \zeta_s + z_{s-1} - z_s, \quad E_{s-1}\zeta_s = 0.$$

We have

$$\begin{aligned}
&B_{3T,1}\\
=\;& T^{-1}\sum_{t>s}\sum(X_{1t}X'_{2t}-A'_{1t}X_{2t}X'_{2t})[T^{-2}\sum_l X_{2l}X'_{2l}K_{h,lt}]^{-1}T^{-2}X_{2s}X'_{2s}(\beta_t-\beta_s)K_{h,st}\\
=\;& \sqrt{T}h\sum_s T^{-2}\sum_{t>s}(X_{1t}X'_{2t}-A'_{1t}X_{2t}X'_{2t})\frac{X_{2s}}{\sqrt{T}}\frac{X'_{2s}}{\sqrt{T}}[T^{-2}\sum_l X_{2l}X'_{2l}K_{h,lt}]^{-1}f(Z_t)\frac{\xi_s}{\sqrt{T}}\\
&+T^{-1}\sum_t(X_{1t}X'_{2t}-A'_{1t}X_{2t}X'_{2t})T^{-2}\sum_{s<t}X_{2s}X'_{2s}[T^{-2}\sum_l X_{2l}X'_{2l}K_{h,lt}]^{-1}[(\beta_t-\beta_s)K_{h,st}-hf(Z_t)\xi_s]\\
=\;& \sqrt{T}h\sum_s T^{-2}\sum_{t>s}(X_{1t}X'_{2t}-A'_{1t}X_{2t}X'_{2t})\frac{X_{2s}}{\sqrt{T}}\frac{X'_{2s}}{\sqrt{T}}[T^{-2}\sum_l X_{2l}X'_{2l}K_{h,lt}]^{-1}f(Z_t)\frac{\zeta_s}{\sqrt{T}}\\
&+\sqrt{T}h\sum_s T^{-2}\sum_{t>s}(X_{1t}X'_{2t}-A'_{1t}X_{2t}X'_{2t})\frac{X_{2s}}{\sqrt{T}}\frac{X'_{2s}}{\sqrt{T}}[T^{-2}\sum_l X_{2l}X'_{2l}K_{h,lt}]^{-1}f(Z_t)\frac{z_{s-1}}{\sqrt{T}}\\
&-\sqrt{T}h\sum_s T^{-2}\sum_{t>s}(X_{1t}X'_{2t}-A'_{1t}X_{2t}X'_{2t})\frac{X_{2s}}{\sqrt{T}}\frac{X'_{2s}}{\sqrt{T}}[T^{-2}\sum_l X_{2l}X'_{2l}K_{h,lt}]^{-1}f(Z_t)\frac{z_s}{\sqrt{T}}\\
&+T^{-1}\sum_t(X_{1t}X'_{2t}-A'_{1t}X_{2t}X'_{2t})T^{-2}\sum_{s<t}X_{2s}X'_{2s}[T^{-2}\sum_l X_{2l}X'_{2l}K_{h,lt}]^{-1}[(\beta_t-\beta_s)K_{h,st}-hf(Z_t)\xi_s]\\
\equiv\;& B_{3T,1,1}+B_{3T,1,2}+B_{3T,1,3}+B_{3T,1,4}. \qquad \text{(A.13)}
\end{aligned}$$

Since A_{1t} and $[T^{-2}\sum_l X_{2l}X'_{2l}K_{h,lt}]^{-1}f(Z_t)$ are asymptotically adapted to $\mathcal{G}_s$ and

$$T^{-2}\sum_{t>s}(X_{1t}X'_{2t}-A'_{1t}X_{2t}X'_{2t})\frac{X_{2s}}{\sqrt{T}}\frac{X'_{2s}}{\sqrt{T}}[T^{-2}\sum_l X_{2l}X'_{2l}K_{h,lt}]^{-1}f(Z_t)\xrightarrow{p}0,$$

following Theorem 2.2 of Kurtz and Protter (1991), we have

$$\sum_s T^{-2}\sum_{t>s}(X_{1t}X'_{2t}-A'_{1t}X_{2t}X'_{2t})\frac{X_{2s}}{\sqrt{T}}\frac{X'_{2s}}{\sqrt{T}}\frac{\zeta_s}{\sqrt{T}}\xrightarrow{p}0.$$

Hence, $B_{3T,1,1}=o_p(\sqrt{T}h)$.

Similar as in the proof of Theorem 3.1 in Hansen (1992), we have

$$\begin{aligned}
\|z_s\|_\beta &= \left\|\sum_{k=1}^{\infty}E_{s-1}\xi_{s+k}\right\| = \left\|\sum_{k=1}^{\infty}E_{s-1}\left((f(Z_t))^{-1}\beta^{(1)}(Z_t)O(h)\right)\right\|\\
&\le O(h)\sum_{k=1}^{\infty}6\alpha_k^{1/\beta-1/p}\|(f(Z_t))^{-1}\beta^{(1)}(Z_t)\|_p \le 6Ch\sum_{k=1}^{\infty}\alpha_k^{1/\beta-1/p}=O(h),
\end{aligned}$$

uniformly in s. Also, we have

$$\sum_s T^{-2}\sum_{t>s}\left\|(X_{1t}X'_{2t}-A'_{1t}X_{2t}X'_{2t})\frac{X_{2s}}{\sqrt{T}}\frac{X'_{2s}}{\sqrt{T}}[T^{-2}\sum_l X_{2l}X'_{2l}K_{h,lt}]^{-1}f(Z_t)\right\|=O_p(T),$$

and $\sup_{s,t} \|(\beta_t - \beta_s)K_{h,st} - hf(Z_t)\xi_s\| = O_p(h^2)$. Thus, $B_{3T,1,2} = O_p(Th^2)$, $B_{3T,1,3} = O_p(Th^2)$, and $B_{3T,1,4} = O_p(Th^2)$. Therefore,

$$B_{3T,1} = o_p(\sqrt{T}h) + O_p(Th^2) = o_p(1).$$

Similarly, we can show that $B_{3T,2} = o_p(\sqrt{T}h) + O_p(Th^2) = o_p(1)$.

Proof of (iv): Since

$$\begin{aligned} B_{4T} &= -T^{-1}\sum_t (X_{1t} - X_{2t}'A_{1t})'X_{2t}'A_{4t} = T^{-1}\sum_t (A_{1t}'X_{2t}X_{2t}' - X_{1t}'X_{2t}')A_{4t} \\ &= T^{-1}\sum_t (A_{1t}'X_{2t}X_{2t}' - X_{1t}'X_{2t}')C_{1T}(Z_t)^{-1}T^{-2}\sum_s X_{2s}u_sK_{h,st} + o_p(1), \end{aligned}$$

we can write $B_{4T} = B_{4T,1} + B_{4T,2} + o_p(1)$, where

$$\begin{aligned} B_{4T,1} &= T^{-3}\sum_t (A_{1t}'X_{2t}X_{2t}' - X_{1t}'X_{2t}')C_{1T}(Z_t)^{-1}X_{2t}u_tK(0), \\ B_{4T,2} &= T^{-3}\sum_t\sum_{s\neq t}(A_{1t}'X_{2t}X_{2t}' - X_{1t}'X_{2t}')C_{1T}(Z_t)^{-1}X_{2s}u_sK_{h,st}. \end{aligned}$$

It is easy to see that $E(B_{4T,1}) = 0$ and $E(B_{4T,1}^2) = T^{-2}$. Hence, $B_{4T,1} = O_p(T^{-1}) = o_p(1)$.

We can write $B_{4T,2}$ as

$$B_{4T,2} = T^{-3}\frac{T(T-1)}{2}\frac{1}{T(T-1)}\sum_t\sum_{s\neq t} H_{4T,ts} = T^{-3}\frac{T(T-1)}{2}U_{4T},$$

where $H_{4T,ts} = [(A_{1t}'X_{2t}X_{2t}' - X_{1t}'X_{2t}')C_{1T}(Z_t)^{-1}X_{2s}u_s + (A_{1s}'X_{2s}X_{2s}' - X_{1s}'X_{2s}')C_{1T}(Z_s)^{-1}X_{2t}u_t]K_{h,st}$. Then, U_{4T} is a second order U-statistic. Hence, we use a conditional Hoeffding decomposition as

$$\begin{aligned} U_{4T} &= E[H_{4T,ts}] + \frac{2}{T(T-1)}\sum_{t=1}^{T}\sum_{s\neq t}[H_{4T,t} - E(H_{4T,t})] \\ &\quad + \frac{2}{T(T-1)}\sum_{t=1}^{T}\sum_{s>t}^{T}\left[H_{4T,ts} - H_{4T,t} - H_{4T,s} + E(H_{4T,ts})\right], \end{aligned}$$

where $H_{4T,t} = E[H_{4T,ts}|w_t]$, $w_t = (\mathcal{F}^{\infty}, Z_t, u_t)$.

From Assumption 2, we have $E[H_{4T,ij}] = 0$. Also, we have that

$$
\begin{aligned}
& \frac{2}{T(T-1)} \sum_{t=1}^{T} \sum_{s \neq t} [H_{4T,t} - E(H_{4T,t})] \\
= & \frac{2}{T(T-1)} \sum_{t=1}^{T} \sum_{s \neq t} E[(A'_{1s} X_{2s} X'_{2s} - X'_{1s} X'_{2s}) C_{1T}(Z_s)^{-1} X_{2t} u_t K_{h,st} | w_t] \\
= & T \frac{2}{\sqrt{T}} \sum_{t=1}^{T} \left[\frac{1}{T} \sum_{s=1}^{T} \left((B_{(2)}^{-1} B_{(2,1)})' \frac{X_{2s}}{\sqrt{T}} \frac{X'_{2s}}{\sqrt{T}} - \frac{X'_{1s}}{\sqrt{T}} \frac{X'_{2s}}{\sqrt{T}} \right) f(Z_t)^{-1} B_{(2)}^{-1} f(Z_t) \right] \frac{X_{2t}}{\sqrt{T}} u_t + o_p(T) \\
= & \left[\frac{1}{T} \sum_{s=1}^{T} \left((B_{(2)}^{-1} B_{(2,1)})' \frac{X_{2s}}{\sqrt{T}} \frac{X'_{2s}}{\sqrt{T}} - \frac{X'_{1s}}{\sqrt{T}} \frac{X'_{2s}}{\sqrt{T}} \right) B_{(2)}^{-1} \right] T \frac{2}{\sqrt{T}} \sum_{t=1}^{T} \frac{X_{2t}}{\sqrt{T}} u_t + o_p(T) \\
= & o_p(T),
\end{aligned}
$$

since

$$
\begin{aligned}
& \frac{1}{T} \sum_{s=1}^{T} \left((B_{(2)}^{-1} B_{(2,1)})' \frac{X_{2s}}{\sqrt{T}} \frac{X'_{2s}}{\sqrt{T}} - \frac{X'_{1s}}{\sqrt{T}} \frac{X'_{2s}}{\sqrt{T}} \right) B_{(2)}^{-1} \\
\xrightarrow{p} & \left(B'_{(2,1)} B_{(2)}^{-1} \int_0^1 B_2 B'_2 dr - \int_0^1 B'_1 B'_2 dr \right) B_{(2)}^{-1} \xrightarrow{p} 0,
\end{aligned}
$$

and $Var[\frac{2}{\sqrt{T}} \sum_{t=1}^{T} \frac{X_{2t}}{\sqrt{T}} u_t] = O(1)$.

Moreover, we have

$$
\begin{aligned}
& Var\left[\frac{2}{T(T-1)} \sum_{t=1}^{T} \sum_{s>t}^{T} [H_{4T,ts} - H_{4T,t} - H_{4T,s} + E(H_{4T,ts})] \right] \\
= & T^3 \frac{2}{T^2(T-1)^2} \sum_{t=1}^{T} \sum_{s>t}^{T} E\left[T^{-3/2} H_{4T,ts} - T^{-3/2} H_{4T,t} - T^{-3/2} H_{4T,s} + T^{-3/2} E(H_{4T,ts}) \right]^2 \\
= & O(T^{-1}).
\end{aligned}
$$

Therefore, we have that $U_{4T} = o_p(T) + O_p(T^{-1/2})$ and $B_{4T,2} = T^{-3} \frac{T(T-1)}{2} U_{4T} = o_p(1)$. Further, $B_{4T} = o_p(1)$.

Proof of Theorem 1: From (A.3) and Lemma 2, we obtain that

$$
\begin{aligned}
& T(\hat{\gamma}_{lc} - \gamma) \\
= & B_{1T}^{-1} [B_{2T} + B_{3T} + B_{4T}] \\
= & \left(\int_0^1 [B_1(r)' - B_2(r)' B_{(2)}^{-1} B_{(2,1)}]^{\otimes 2} dr \right)^{-1} [\int_0^1 [B_1(r)' - B_2(r)' B_{(2)}^{-1} B_{(2,1)}]' dB_3(r) + o_p(1)] \\
\Rightarrow & \left(\int_0^1 [B_1(r)' - B_2(r)' B_{(2)}^{-1} B_{(2,1)}]^{\otimes 2} dr \right)^{-1} \int_0^1 [B_1(r)' - B_2(r)' B_{(2)}^{-1} B_{(2,1)}]' dB_3(r). \quad \text{(A.14)}
\end{aligned}
$$

This completes the proof.

REFERENCES

Andrews, D. W. K., 1988. Laws of large numbers for dependent non-identically distributed random variables. *Econometric Theory* **4**, 458-467.

Cai, Z., Q. Li, and J. Park, 2009. Functional-coefficient models for nonstationary time series data. *Journal of Econometrics* **148**, 101-113.

Gu, J. and Z. Liang, 2014. Testing cointegration relationship in a semiparametric varying coefficient model. *Journal of Econometrics* **178**, 57-70.

Hansen, B. E., 1992. Convergence to stochastic integrals for dependent heterogeneous processes. *Econometric Theory* **8**, 489-500.

Hansen, B. E., 2008. Uniform convergence rates for kernel estimation with dependent data. *Econometric Theory* **24**, 726-748.

Juhl, T. and Z. Xiao, 2005. Partially linear models with unit roots. *Econometric Theory* **21**, 877-906.

Karlsen, H. A., T. Myklebust, and D. Tjøstheim, 2007. Nonparametric estimation in a nonlinear cointegration type model. *Annals of Statistics* **35**, 252-299.

Kurtz, T. G. and P. Protter, 1991. Weak limit theorems for stochastic integrals and stochastic differential equations. *Annals of Probability* **19**, 1035-1070.

Li, K., D. Li, Z. Liang, and C. Hsiao, 2014. Estimation of semi-varying coefficient models with nonstationary regressors. *Econometric Reviews*, forthcoming.

Masry, E., 1996. Multivariate local polynomial regression for time series: uniform strong consistency and rates. *Journal of Time Series Analysis* **17**, 571-599.

Park, J. Y. and P. C. B. Phillips, 1988. Statistical inference in regressions with integrated processes: Part 1. *Econometric Theory* **4**, 468-497.

Park, J. Y. and P. C. B. Phillips, 1989. Statistical inference in regressions with integrated processes: Part 2. *Econometric Theory* **4**, 95-131.

Park, J. Y. and P. C. B. Phillips, 2001. Nonlinear regressions with integrated time series. *Econometrica* **69**, 117-161.

Phillips, P. C. B. and S. N. Durlauf, 1986. Multiple time series regression with integrated processes. *Review of Economic Studies* **53**, 473-495.

Robinson, P. M., 1988. Root-N-consistent semiparametric regression. *Econometrica* **56**, 931-954.

Shorack, G. R. and J. A. Wellner, 1986. Empirical Processes with Applications to statistics. Wiley.

Wang, Q. and P. C. B. Phillips, 2009. Asymptotic theory for local time density estimation and nonparametric cointegrating regression. *Econometric Theory* **25**, 710-738.

Wang, Q. and P. C. B. Phillips, 2009. Structural nonparametric cointegrating regression. *Econometrica* **77**, 1901-1948.

Xiao, Z., 2009. Functional-coefficient cointegration models. *Journal of Econometrics* **152**, 81-92.