Article:

Li, Kunpeng, Li, Degui orcid.org/0000-0001-6802-308X, Liang, Zhongwen et al. (1 more author) (2017) Estimation of Semi-Varying Coefficient Models with Nonstationary Regressors. Econometric Reviews. pp. 354-369. ISSN 0747-4938

https://doi.org/10.1080/07474938.2015.1114563

Estimation of Semi-Varying Coefficient Models with Nonstationary Regressors

Kunpeng Li[1], **Degui Li**[2], **Zhongwen Liang**[3] **and Cheng Hsiao**[4*]

[1] *School of Economcs & Management, Capital University of Economics & Business, China*

[2] *Department of Mathematics, University of York, United Kingdom*

[3] *Department of Economics, University at Albany, SUNY, U.S.A.*

[4] *Department of Economics, University of Southern California, U.S.A.*

□ *We study a semi-varying coefficient model where the regressors are generated by the multivariate unit root I(1) processes. The influence of the explanatory vectors on the response variable satisfies the semiparametric partially linear structure with the nonlinear component being functional coefficients. A semiparametric estimation methodology with the first-stage local polynomial smoothing is applied to estimate both the constant coefficients in the linear component and the functional coefficients in the nonlinear component. The asymptotic distribution theory for the proposed semiparametric estimators is established under some mild conditions, from which both the parametric and nonparametric estimators are shown to enjoy the well-known super-consistency property. Furthermore, a simulation study is conducted to investigate the finite sample performance of the developed methodology and results.*

JEL Classifications: C13, C14, C22.

Keywords: Functional coefficients, local polynomial fitting, semiparametric estimation, super-consistency, unit root process.

Abbreviated Title: Semi-Varying Coefficient Models.

*The corresponding author: Cheng Hsiao, Email: chsiao@usc.edu

1. INTRODUCTION

In this paper, we are interested in a partially linear varying coefficient model defined by

$$Y_t = X_{1t}^{\top}\gamma + X_{2t}^{\top}\beta(Z_t) + u_t, \tag{1.1}$$

where X_{1t} is a d_1-dimensional I(1) vector, X_{2t} is a d_2-dimensional I(1) vector, Z_t is a scalar stationary (or I(0)) variable, u_t is a stationary error term, γ is a $d_1 \times 1$ vector of constant parameters and $\beta(\cdot)$ is a d_2-dimensional vector of unspecified smooth functions. The notation "$\top$" denotes transpose of a vector (or matrix). Model (1.1) provides a very flexible framework in nonstationary time series analysis, and it covers various linear and nonlinear time series models with nonstationarity. For example, when $\beta(Z_t) \equiv \beta$, (1.1) reduces to a linear cointegration model which has been systematically investigated by existing literature such as Phillips (1986), Phillips and Durlauf (1986), Park and Phillips (1988, 1989) and Saikkonen (1995). When $\gamma = 0$, (1.1) becomes a functional coefficient model with nonstationarity, which has been studied by Cai *et al.* (2009), Xiao (2009) and Sun and Li (2011). The advantage of the functional coefficient structure in the nonparametric component of model (1.1) is that it could attenuate the so-called "curse of dimensionality" problem in nonparametric estimation when the dimension of the predictors is larger than three.

The main focus of this paper is to consider semiparametric estimation for both the parameter γ and the functional coefficient $\beta(\cdot)$, and then derive the associated asymptotic theory. In Section 2 below, we will use a so-called profile likelihood approach with first-stage local polynomial fitting to estimate the proposed model. In the case of independent or stationary weakly dependent observations, the profile likelihood methodology has been commonly used to estimate semiparametric varying coefficient models, see, for example, Fan and Huang (2005), Zhou and Liang (2009), Li *et al.* (2011) and the references therein. However, to the best of our knowledge, there is few work of extending such idea to the nonstationary time series case. Chen *et al.* (2012) considered the profile least squares estimation for the partially linear model through the null recurrent Markov chain framework. However, the model in Chen *et al.* (2012) is less general than model (1.1), and it is difficult to verify the null recurrent Markov property in practical applications.

The main challenge of deriving the asymptotic theory in this paper is the lack of uniform consistency results for the local polynomial estimators in the context of the functional coefficient models with nonstationarity. Hence, in this paper, we establish such uniform consistency result (see, for example, the argument in the proof of Proposition A.1), which is critical in our derivation and of independent interest. Under some mild conditions, we then establish the asymptotic distribution theory for the proposed semiparametric estimators. We show that the estimator for the parameter in the linear component enjoys the well-known super-consistency property with n-convergence rate, which is similar to that in parametric linear and nonlinear cointegration models (c.f., Park and Phillips, 1988, 1989, 2001). However, such super-consistency result is fundamentally different from the parametric convergence rate in Chen *et al.* (2012), who could only derive the root-n rate in the context of partially linear models with regressor being null recurrent Markov chain. Meanwhile, similar to Cai *et al.* (2009), Xiao (2009) and Phillips *et al.* (2013), we can also show that the convergence rate for the nonparametric estimator is faster than the root-nh rate which is common in the stationary case. Our results complement existing literature on nonparametric and semiparametric estimation for nonstationary time series (see, for example, Park and Hahn, 1999; Juhl and Xiao, 2005; Cai *et al.*, 2009; Wang and Phillips, 2009a, 2009b; Xiao 2009; Chen *et al.*, 2010; Sun and Li, 2011; and Chen *et al.*, 2012). Furthermore, a simulation study is conducted to illustrate the finite sample performance of the proposed methodology as well as the super-consistency results.

The rest of this paper is organized as follows. The semiparametric estimation methodology is given in Section 2. The asymptotic theory for the proposed method is provided in Section 3. The simulation study is conducted in Section 4. Section 5 concludes the paper. The mathematical proofs of the asymptotic results are given in an appendix.

2. SEMIPARAMETRIC ESTIMATION METHOD

As mentioned above, when $(Y_t, X_t^\top, Z_t)$ is stationary with $X_t^\top = (X_{1t}^\top, X_{2t}^\top)$, the profile likelihood estimation methodology as well as the related asymptotic properties have been extensively studied for the semi-varying coefficient model (1.1), see, for example, Fan and Huang (2005) and Zhou and Liang (2009). In this paper, we will extend such methodology to the nonstationary time series case, which is an important feature for economic data. To avoid confusion, throughout the paper, we let γ_0 and $\beta_0(\cdot)$ be the true parameters and functional coefficients.

Let $\mathbf{e} = \Big(\mathbf{I}_{d_2}, \mathbf{N}_{d_2\times qd_2}\Big)$, where $\mathbf{I}_{d_2}$ is a $d_2 \times d_2$ identity matrix and $\mathbf{N}_{d_2\times qd_2}$ is a $d_2 \times qd_2$ null matrix. Define

$$Z_{st,h} = (Z_s - Z_t)/h, \ \ K_{h,st} = K(Z_{st,h}), \ \ Q_{s,t} = \Big[1, (Z_s - Z_t), \cdots, (Z_s - Z_t)^q\Big]^\top,$$

and $\mathbf{G}_h = \mathrm{diag}\Big(1, h, \cdots, h^q\Big) \otimes \mathbf{I}_{d_2}$, where h is a bandwidth, $K(\cdot)$ is a kernel function and $\otimes$ denotes the Kronecker product. We next adopt the local polynomial approach (Fan and Gijbels, 1996) to estimate the functional coefficient $\beta_0(\cdot)$ when γ is given. Assuming that $\beta_0(\cdot)$ has q-th order continuous derivative ($q \geq 1$), we have the following Taylor expansion for the functional coefficient:

$$\beta_0(z) \approx \beta_0(z_0) + \beta_0'(z_0)(z - z_0) + \cdots + \beta_0^{(q)}(z_0)\frac{(z - z_0)^q}{q!}$$

for z in a small neighborhood of z_0. The local polynomial estimate of

$$M_\beta(\cdot) = \Big[\beta_0(\cdot), \beta_0'(\cdot), \cdots, \frac{\beta_0^{(q)}(\cdot)}{q!}\Big]^\top$$

at point z_0 for given γ, is defined by minimizing the weighted loss function (with respect to A):

$$L_n(A \mid \gamma) = \sum_{t=1}^n \Big\{Y_t - X_{1t}^\top \gamma - A\Big[Q_t(z_0) \otimes X_{2t}\Big]\Big\}^2 K\Big(\frac{Z_t - z_0}{h}\Big),$$

where A is a $(q+1)d_2$-dimensional row vector and $Q_t(z_0) = \Big[1, (Z_t - z_0), \cdots, (Z_t - z_0)^q\Big]^\top$. Then, by some elementary calculations, the local polynomial estimate of $\beta_0(\cdot)$ at Z_t for given γ is

$$\begin{aligned}
\widetilde{\beta}(Z_t, \gamma) &= \mathbf{e}\Big[\sum_{s=1}^n K_{h,st} Q_{s,t} Q_{s,t}^\top \otimes X_{2s} X_{2s}^\top\Big]^{-1} \sum_{s=1}^n K_{h,st} Q_{s,t} \otimes X_{2s}(Y_s - X_{1s}^\top \gamma) \\
&= \mathbf{e}\mathbf{G}_h\Big[\sum_{s=1}^n K_{h,st} Q_{s,t} Q_{s,t}^\top \otimes X_{2s} X_{2s}^\top\Big]^{-1} \mathbf{G}_h \mathbf{G}_h^{-1} \sum_{s=1}^n K_{h,st} Q_{s,t} \otimes X_{2s}(Y_s - X_{1s}^\top \gamma) \\
&= \mathbf{e}\Big[n^{-2} \sum_{s=1}^n K_{h,st} \mathbf{G}_h^{-1}\Big(Q_{s,t} Q_{s,t}^\top \otimes X_{2s} X_{2s}^\top\Big)\mathbf{G}_h^{-1}\Big]^{-1}\Big[n^{-2} \sum_{s=1}^n K_{h,st} \mathbf{G}_h^{-1}(Q_{s,t} \otimes X_{2s})(Y_s - X_{1s}^\top \gamma)\Big] \\
&= \mathcal{A}_{2t} - \mathcal{A}_{1t}\gamma \qquad (2.1)
\end{aligned}$$

where

$$
\begin{aligned}
\mathcal{A}_{1t} &= \mathbf{e}\mathcal{S}_t^{-1}\Big[n^{-2}\sum_{s=1}^{n} K_{h,st}\mathbf{G}_h^{-1}(Q_{s,t}\otimes X_{2s})X_{1s}^{\top}\Big],\\
\mathcal{A}_{2t} &= \mathbf{e}\mathcal{S}_t^{-1}\Big[n^{-2}\sum_{s=1}^{n} K_{h,st}\mathbf{G}_h^{-1}(Q_{s,t}\otimes X_{2s})Y_s\Big],\\
\mathcal{S}_t &= n^{-2}\sum_{s=1}^{n} K_{h,st}\mathbf{G}_h^{-1}\Big(Q_{s,t}Q_{s,t}^{\top}\otimes X_{2s}X_{2s}^{\top}\Big)\mathbf{G}_h^{-1},
\end{aligned}
$$

and in the second equality above we used the fact that $\mathbf{e}\mathbf{G}_h = \mathbf{e}$.

It is easy to see that $\widetilde{\beta}(Z_t,\gamma)$ can be seen as a function of the unknown parameter γ. Then, replacing $\beta(Z_t)$ by $\widetilde{\beta}(Z_t,\gamma)$ in model (1.1) and then applying ordinary least squares (OLS) method, we obtain the estimator of γ_0:

$$
\widehat{\gamma} = \left[\sum_{t=1}^{n}(X_{1t}-\mathcal{A}_{1t}^{\top}X_{2t})(X_{1t}-\mathcal{A}_{1t}^{\top}X_{2t})^{\top}\right]^{-1}\sum_{t=1}^{n}(X_{1t}-\mathcal{A}_{1t}^{\top}X_{2t})(Y_t - X_{2t}^{\top}\mathcal{A}_{2t}). \tag{2.2}
$$

With $\widehat{\gamma}$ replacing γ in (2.1), we obtain a feasible local polynomial estimator of $\beta_0(z)$ by

$$
\widehat{\beta}(z) = \mathcal{A}_2(z) - \mathcal{A}_1(z)\widehat{\gamma}, \tag{2.3}
$$

where $\mathcal{A}_1(z)$ and $\mathcal{A}_2(z)$ are defined as $\mathcal{A}_{1t}$ and $\mathcal{A}_{2t}$ with Z_t being replaced by z. The asymptotic properties of $\widehat{\gamma}$ and $\widehat{\beta}(z)$ will be given in Section 3 below.

3. ASYMPTOTIC THEORY

Before giving the asymptotic distribution theory for both $\widehat{\gamma}$ and $\widehat{\beta}(z)$, we first introduce some regularity conditions. Let $X_{1t} = X_{1,t-1} + x_{1t}$ and $X_{2t} = X_{2,t-1} + x_{2t}$, where x_{1t} and x_{2t} are stationary and weakly dependent random vector processes which will be specified later. Without loss of generality, we assume that $X_{10} = \mathbf{0}$ and $X_{20} = \mathbf{0}$, where $\mathbf{0}$ is a null vector whose dimension may vary from place to place. Hereafter, let $\|\cdot\| = \|\cdot\|_2$ denote the Euclidean norm.

Assumption 1. Let $w_t = (x_t^{\top},\ u_t)^{\top}$ with $x_t^{\top} = (x_{1t}^{\top},\ x_{2t}^{\top})$. For some $p_1 > p_2 > 2$, $\Big\{(w_t^{\top}, Z_t)\Big\}$ is a strictly stationary and strongly mixing sequence with zero mean and mixing coefficients $\alpha_m = O\Big(m^{-p_1p_2/(p_1-p_2)}\Big)$ and $\mathrm{E}\Big[\|w_t\|^{p_1} + |Z_t|^{p_1}\Big] < \infty$. In addition, there exists a positive definite matrix Ω such that $\frac{1}{n}\mathrm{E}\Big[(\sum_{t=1}^{n} w_t)(\sum_{t=1}^{n} w_t)^{\top}\Big] \to \Omega$, and Z_t has a compact support $\mathcal{S}_Z$.

Assumption 2. Let $(u_t, \mathcal{F}_{nt}, 1 \le t \le n)$ be a martingale difference sequence with $\mathrm{E}(u_t|\mathcal{F}_{n,t-1})$=0 a.s. and $\mathrm{E}(u_t^2|\mathcal{F}_{n,t-1}) = \sigma_u^2$ a.s., where $\mathcal{F}_{nt} = \sigma\Big\{x_{s_1}, Z_{s_1}, u_{s_2} : 1 \le s_1 \le n,\ 1 \le s_2 \le t\Big\}$.

Assumption 3. The function $\beta(z)$ has $(q+1)$-th order continuous derivatives when z is in the compact support of Z_t.

Assumption 4. The density function of Z_t, $f_Z(z)$, is positive and bounded away from infinity and zero, and has second-order continuous derivative when z is in the compact support of Z_t. Furthermore, the joint density function of (Z_1, Z_{s+1}), $f(u,v;s)$, is bounded for all $s \ge 1$.

Assumption 5. $K(\cdot)$ is continuous probability density function with a compact support.

Assumption 6. Let $nh^{q+1} \to 0$ and $(nh)/\log n \to \infty$ as $n \to \infty$.

Consider the partial sum process defined by $B_n(s) = n^{-1/2}\sum_{t=1}^{[ns]} w_t$ with w_t being defined in Assumption 1 and $0 \le s \le 1$, where $[a]$ denotes the largest integer less than or equal to a. By Assumption 1 and the multivariate invariance principle for $B_n(s)$ (c.f., Phillips and Durlauf, 1986), we have $B_n(s) \Rightarrow B(s)$, where $B(\cdot)$ is a multivariate Brownian motion with $\mathrm{E}\Big[B(1)B^\top(1)\Big] = \Omega$. We can further decompose $B_n(s)$ as $\Big[B_{1n}^\top(s), B_{2n}^\top(s), B_{3n}(s)\Big]^\top$, where $B_{1n}(s) = n^{-1/2}\sum_{t=1}^{[ns]} x_{1t}$, $B_{2n}(s) = n^{-1/2}\sum_{t=1}^{[ns]} x_{2t}$ and $B_{3n}(s) = n^{-1/2}\sum_{t=1}^{[ns]} u_t$. Then, we have $B_{jn}(s) \Rightarrow B_j(s)$ such that $B(s) = \Big[B_1^\top(s), B_2^\top(s), B_3(s)\Big]^\top$. The restriction of martingale differences on the error term u_t in Assumption 2 is to facilitate our proofs, and it can be relaxed at the cost of more lengthy proofs. In particular, we can relax the conditional homoskedastic condition to the heteroskedastic case, and similar asymptotic theory would still hold with modified proofs. The smoothness conditions in Assumptions 3–5 ensure that the local polynomial estimation and some uniform consistency results (c.f., Masry, 1996; and Hansen, 2008) are applicable, and such conditions are critical in our proofs and commonly used in the literature on nonparametric estimation such as Fan and Gijbels (1996) and Li and Racine (2007). Assumption 6 imposes some restrictions on the bandwidth h. The first bandwidth condition $nh^{q+1} = o(1)$ is imposed to ensure the asymptotic bias of the local polynomial estimators is asymptotically negligible and thus the n-convergence rate can be obtained for the parametric estimator $\widehat{\gamma}$. In particular, when the local linear approach ($q = 1$) is applied, we can further relax such condition to $nh^2 = O(1)$. The second bandwidth condition $(nh)/\log n \to \infty$ is common to apply the uniform consistency result of nonparametric kernel-based estimation.

Define

$$\Sigma_0 = \int \left[B_1(s) - W^\top(B_1, B_2) B_2(s) \right]^{\otimes 2} ds,$$

where $W(B_1, B_2) = \left[\int_0^1 B_2(s) B_2^\top(s) ds \right]^{-1} \int_0^1 B_2(s) B_1^\top(s) ds$ and $B^{\otimes 2} = BB^\top$ for any matrix B. Define

$$\Sigma_1(s) = B_1(s) - W^\top(B_1, B_2) B_2(s).$$

The next theorem gives the asymptotic distribution of $\widehat{\gamma}$ defined in (2.2).

Theorem 3.1. *Suppose that Assumptions 1–6 are satisfied and Σ_0 is non-singular. Then, we have*

$$n(\widehat{\gamma} - \gamma_0) \Rightarrow \Sigma_0^{-1} \int_0^1 \Sigma_1(s) dB_3(s). \tag{3.1}$$

The above theorem shows that the estimator $\widehat{\gamma}$ enjoys the super-consistency property in the context of semi-varying coefficient cointegration models, which can be seen as an extension of some existing results for the parametric cointegration models (c.f, Park and Phillips, 1988, 1989, 2001). However, the non-standard asymptotic distribution on the right hand side of (3.1) would make the associated statistical inference more difficult than that in the stationary case. We conjecture that the techniques developed in Park and Phillips (1988) can be generalized to the semiparametric setting in this paper and will consider this in the future study.

By using Theorem 3.1, we can also derive the asymptotic distribution for $\widehat{\beta}(z)$ defined in (2.3). To simplify the presentation, we only consider the case of $d_2 = 1$. The extension to the case of $d_2 > 1$ is straightforward. Let $\mu_j = \int s^j K(s) ds$, $\nu_j = \int s^j K^2(s) ds$, $\Delta(\mu)$ be a $(q+1) \times (q+1)$ matrix with the (i, j)-th element being μ_{i+j-2}, and $\Gamma(\nu)$ be $(q+1) \times (q+1)$ matrix with the (i, j)-th element being ν_{i+j-2}. Define

$$b_z = \frac{h^{q+1} \beta_0^{(q+1)}(z)}{(q+1)!} \mathbf{e}_1 \Delta^{-1}(\mu) (\mu_{q+1}, \cdots, \mu_{2q+1})^\top$$

and

$$\Sigma_z = \frac{\sigma_u^2 \mathbf{e}_1 \Delta^{-1}(\mu)\Gamma(\nu)\Delta^{-1}(\mu)\mathbf{e}_1^\top}{f_Z(z)\int_0^1 B_2^2(s)ds},$$

where $\mathbf{e}_1 = (1, 0, \cdots, 0)$ is of dimension $(q+1)$.

Theorem 3.2. *Suppose that the conditions of Theorem 3.1 are satisfied. Then, we have*

$$n\sqrt{h}\Big[\widehat{\beta}(z) - \beta_0(z) - b_z\Big] \Rightarrow \mathrm{MN}(\Sigma_z), \tag{3.2}$$

where $\mathrm{MN}(\Sigma_z)$ *is a mixed normal distribution with zero mean and conditional covariance matrix* Σ_z.

The mixed normal distribution in Theorem 3.2 means that the estimator has an asymptotic normal distribution conditional on the random variable which is involved in Σ_z. By using the first bandwidth condition in Assumption 6, we can further show that the asymptotic bias term in the above theorem is asymptotically negligible. Hence, the asymptotic distribution in (3.2) can be simplified to

$$n\sqrt{h}\Big[\widehat{\beta}(z) - \beta_0(z)\Big] \Rightarrow \mathrm{MN}(\Sigma_z).$$

We can find that the above convergence rate is faster than the root-nh rate in stationary case, which is consistent with the findings in Cai *et al.* (2009), Xiao (2009) and Sun and Li (2011).

4. SIMULATION STUDY

In this section, we give a simulated example to illustrate the proposed methodology and theory. Consider the model

$$Y_t = X_{1t}\gamma_0 + X_{2t}\beta_0(Z_t) + u_t, \quad t = 1, 2, \cdots, n, \tag{4.1}$$

where $\gamma_0 = 2$ and $\beta_0(z) = \sin(\pi z)$, $u_t \overset{i.i.d.}{\sim} \mathrm{N}(0, 0.5^2)$, and $\{Z_t\}$ is generated by the AR(1) model:

$$Z_t = 0.5Z_{t-1} + z_t \ \text{ with } \ z_t \overset{i.i.d.}{\sim} \mathrm{N}(0, 0.5^2),$$

and $\{z_t\}$ is independent of $\{u_t\}$. It is easy to check that $\{Z_t\}$ is stationary and α-mixing dependent with geometric decaying coefficient. For the generation of $\{X_t\}$ with $X_t = (X_{1t}, X_{2t})^\top$, we consider the following two cases:

(i) $\{X_t\}$ is generated by $X_t = 0.5X_{t-1} + x_t$, where

$$x_t = (x_{1t}, x_{2t})^\top \overset{i.i.d.}{\sim} \mathrm{N}\Big((0,0)^\top, \mathrm{diag}(1,1)\Big).$$

(ii) $\{X_t\}$ is generated by $X_t = X_{t-1} + x_t$, where x_t is generated as in Case (i).

It is easy to show that $\{X_t\}$ defined in Case (i) is stationary and α-mixing dependent, whereas $\{X_t\}$ defined in Case (ii) is nonstationary I(1). In this simulation, we consider the sample size $n = 300$ and 600 with replication number $N = 200$. For simplicity, we use the local linear smoother (which corresponds to the local polynomial smoother with $q = 1$) to estimate the coefficient function $\beta_0(\cdot)$ with the standard normal kernel function, and the bandwidth is chosen by using the cross-validation method.

To investigate the performance of the proposed semiparametric estimation methods for the above two cases, we calculate the bias for the parametric estimate as well as the mean squared errors for the nonparametric estimate when the replication is 200. Let

$$\mathrm{Bias}(\gamma) = \frac{1}{N}\sum_{j=1}^{N}\mathrm{Bias}_j(\lambda), \ \ \mathrm{Bias}_j(\gamma) = \widehat{\gamma}(j) - \gamma_0, \tag{4.2}$$

and

$$\mathrm{MSE}(\beta) = \frac{1}{N}\sum_{j=1}^{N}\mathrm{MSE}_j(\beta), \ \ \mathrm{MSE}_j(\beta) = \frac{1}{n}\sum_{t=1}^{n}\Big[\widehat{\beta}(Z_t, j) - \beta_0(Z_t)\Big]^2, \tag{4.3}$$

where $\widehat{\gamma}(j)$ and $\widehat{\beta}(\cdot, j)$ are the resulting parametric and nonparametric estimates in the j-th simulation, $1 \le j \le 200$. Meanwhile, in Table 1 below, we also provide the standard errors of the parametric estimates in the 200 replications denoted by $\mathrm{SE}(\gamma)$ which can be used to verify the super-consistency of the parametric estimation in the finite sample case, and the standard errors of $\mathrm{MSE}_j(\beta)$ in the 200 replications denoted by $\mathrm{SE}(\beta)$.

The bias and standard errors for the parametric estimates and the mean squared errors for the nonparametric estimates are given in Table 1, and the plots for the nonparametric estimation with sample size 600 are given in Figure 1. From Table 1, we have the following conclusions. (1) It is easy to find that the standard errors of the parametric estimates in Case (ii) are smaller than those in Case (i).

In particular, for Case (i), the ratio of the standard error when the sample size is 300 and that when the sample size is 600, equals to 2.4644/1.7600 = 1.4002, close the theoretical value $\sqrt{2}$; and for Case (ii), the ratio of the standard error when the sample size is 300 and that when the sample size is 600, equals to 7.7832/3.2729 = 2.3781, close the theoretical value 2. This finding is consistent with the super-consistency result for the parametric estimation in Theorem 3.1. (ii) The values of mean squared errors for the local linear estimates in Case (ii) are smaller than those in Case (i), which indicates that the convergence of the nonparametric estimation is faster in the nonstationary case. (iii) The performance of the parametric and nonparametric estimators improve as the sample size increases from 300 to 600.

Table 1. *Bias and standard errors for the parametric estimate and mean squared errors for the nonparametric estimate*

	n=300		n=600	
	Case (i)	Case (ii)	Case (i)	Case (ii)
Bias(γ)	-7.5667($\times 10^{-4}$)	1.5710($\times 10^{-4}$)	-0.4854($\times 10^{-4}$)	1.3429($\times 10^{-4}$)
SE(γ)	2.4644($\times 10^{-2}$)	7.7832($\times 10^{-3}$)	1.7600($\times 10^{-2}$)	3.2729($\times 10^{-3}$)
MSE(β)	2.8109($\times 10^{-2}$)	1.6770($\times 10^{-2}$)	1.4600($\times 10^{-2}$)	1.0400($\times 10^{-2}$)
SE(β)	1.5174($\times 10^{-2}$)	0.2975($\times 10^{-2}$)	0.4700($\times 10^{-2}$)	0.1100($\times 10^{-2}$)

5. CONCLUSION

In this paper, we consider a semi-varying coefficient model where the regressors are generated by the multivariate unit root I(1) processes, which provides a flexible framework to model the impact of nonstationary explanatory vectors on the response variable and covers some commonly-used parametric and nonparametric cointegration models. We apply a semiparametric estimation methodology with the first-stage local polynomial smoothing to estimate both the constant coefficients in the linear component and the functional coefficients in the nonlinear component. Through developing the uniform consistency result for the local polynomial estimators in the context of the functional coefficient models with nonstationarity which is of independent interest, we establish the asymptotic distribution

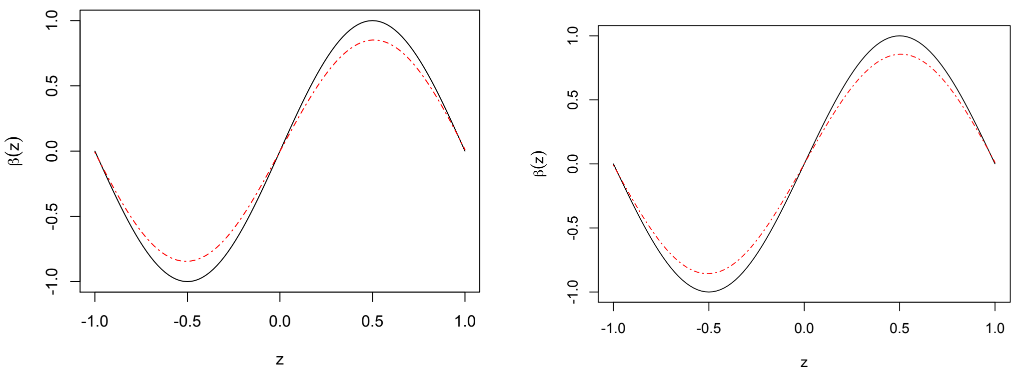

Figure 1. The left plot is the local linear estimated coefficient function for Case (i) and the right plot is the local linear estimated coefficient function for Case (ii). The solid line is the true function and the dashed line is the estimated function.

theory for the developed semiparametric estimators, from which both the parametric and nonparametric estimators are shown to enjoy the well-known super-consistency property. In particular, to derive the super-consistency result for the parametric estimator, we need to undersmooth the functional coefficient (see the first bandwidth restriction in Assumption 6) which is common in the stationary case. Our asymptotic results substantially generalize some existing results in the context of parametric and nonparametric cointegration models. Furthermore, a simulation study is conducted to investigate the finite sample performance of the developed methodology and asymptotic results.

6. Acknowledgments

The authors would like to thank the Editor, the Associate Editor and two referees for their valuable and constructive comments and suggestions. The first author acknowledges the financial support from the Natural Science Foundation of China under Grant Number 71201031. The second author acknowledges the financial support from the Australian Research Council Discovery Early Career Researcher Award under Grant Number DE120101130.

Appendix: Proofs of the main results

In this appendix, we give the proofs of the theoretical results given in Section

3. Throughout the proof, "≡" means "is defined as".

Proof of Theorem 3.1. Note that

$$
\begin{aligned}
\widehat{\gamma}-\gamma_0 &= \Big[\sum_{t=1}^{n}(X_{1t}-\mathcal{A}_{1t}^\top X_{2t})(X_{1t}-\mathcal{A}_{1t}^\top X_{2t})^\top\Big]^{-1}\sum_{t=1}^{n}(X_{1t}-\mathcal{A}_{1t}^\top X_{2t})(Y_t-X_{2t}^\top\mathcal{A}_{2t})-\gamma_0\\
&= \Big[\sum_{t=1}^{n}(X_{1t}-\mathcal{A}_{1t}^\top X_{2t})(X_{1t}-\mathcal{A}_{1t}^\top X_{2t})^\top\Big]^{-1}\sum_{t=1}^{n}(X_{1t}-\mathcal{A}_{1t}^\top X_{2t})(Y_t-X_{1t}^\top\gamma_0)+\\
&\quad\Big[\sum_{t=1}^{n}(X_{1t}-\mathcal{A}_{1t}^\top X_{2t})(X_{1t}-\mathcal{A}_{1t}^\top X_{2t})^\top\Big]^{-1}\sum_{t=1}^{n}(X_{1t}-\mathcal{A}_{1t}^\top X_{2t})X_{2t}^\top\Big(\mathcal{A}_{1t}\gamma_0-\mathcal{A}_{2t}\Big)\\
&= \Big[\sum_{t=1}^{n}(X_{1t}-\mathcal{A}_{1t}^\top X_{2t})(X_{1t}-\mathcal{A}_{1t}^\top X_{2t})^\top\Big]^{-1}\sum_{t=1}^{n}(X_{1t}-\mathcal{A}_{1t}^\top X_{2t})u_t-\\
&\quad\Big[\sum_{t=1}^{n}(X_{1t}-\mathcal{A}_{1t}^\top X_{2t})(X_{1t}-\mathcal{A}_{1t}^\top X_{2t})^\top\Big]^{-1}\sum_{t=1}^{n}(X_{1t}-\mathcal{A}_{1t}^\top X_{2t})\widetilde{u}_t+\\
&\quad\Big[\sum_{t=1}^{n}(X_{1t}-\mathcal{A}_{1t}^\top X_{2t})(X_{1t}-\mathcal{A}_{1t}^\top X_{2t})^\top\Big]^{-1}\sum_{t=1}^{n}(X_{1t}-\mathcal{A}_{1t}^\top X_{2t})X_{2t}^\top\Big[\beta_0(Z_t)-\overline{\beta}(Z_t)\Big],
\end{aligned}
$$

where $\widetilde{u}_t = \mathbf{e}\mathcal{S}_t^{-1}\Big[n^{-2}\sum_{s=1}^{n}K_{h,st}\mathbf{G}_h^{-1}(Q_{s,t}\otimes X_{2s})u_s\Big]$, $\overline{\beta}(Z_t)=\mathbf{e}\overline{M}_\beta(Z_t)$ with $\overline{M}_\beta(Z_t) = \mathcal{S}_t^{-1}\Big[n^{-2}\sum_{s=1}^{n}K_{h,st}\mathbf{G}_h^{-1}(Q_{s,t}\otimes X_{2s})X_{2s}^\top\beta_0(Z_s)\Big]$ and $\mathbf{e}$ is defined in Section 2.

To further simplify the presentation, we define

$$
\begin{aligned}
\mathcal{B}_{1n} &= \frac{1}{n^2}\sum_{t=1}^{n}\Big(X_{1t}-\mathcal{A}_{1t}^\top X_{2t}\Big)\Big(X_{1t}-\mathcal{A}_{1t}^\top X_{2t}\Big)^\top,\\
\mathcal{B}_{2n} &= \frac{1}{n}\sum_{t=1}^{n}(X_{1t}-\mathcal{A}_{1t}^\top X_{2t})u_t,\\
\mathcal{B}_{3n} &= \frac{1}{n}\sum_{t=1}^{n}(X_{1t}-\mathcal{A}_{1t}^\top X_{2t})X_{2t}^\top\widetilde{u}_t,\\
\mathcal{B}_{4n} &= \frac{1}{n}\sum_{t=1}^{n}(X_{1t}-\mathcal{A}_{1t}^\top X_{2t})X_{2t}^\top\Big[\beta_0(Z_t)-\overline{\beta}(Z_t)\Big].
\end{aligned}
$$

Then, Theorem 3.1 can be proved through the following three propositions. ■

Proposition A.1. *Under the conditions of Theorem 3.1, we have*

$$
\mathcal{B}_{1n}\Rightarrow\int\Big[B_1(s)-W^\top(B_1,B_2)B_2(s)\Big]^{\otimes 2}ds=\Sigma_0, \tag{A.1}
$$

where $W(B_1,B_2)$ is defined in Section 3.

Proof. Let $\mathbf{K}_{h,s}(z) = \frac{1}{h} K_{h,s}(z) Q_s(z) Q_s^\top(z)$, where

$$K_{h,s}(z) = K(Z_{s,h}(z)), \;\; Z_{s,h}(z) = (Z_s - z)/h, \;\; Q_{s,z} = \Big[1, (Z_s - z), \cdots, (Z_s - z)^q\Big]^\top.$$

We have

$$\begin{aligned}
\frac{1}{h}\mathcal{S}_n(z) &\equiv \mathbf{G}_h^{-1}\Big[n^{-2}\sum_{s=1}^n \mathbf{K}_{h,s}(z) \otimes X_{2s}X_{2s}^\top\Big]\mathbf{G}_h^{-1} \\
&= \mathbf{G}_h^{-1}\Big\{n^{-2}\sum_{s=1}^n \mathrm{E}\Big[\mathbf{K}_{h,s}(z)\Big] \otimes X_{2s}X_{2s}^\top\Big\}\mathbf{G}_h^{-1} + \\
&\quad\; \mathbf{G}_h^{-1}\Big(n^{-2}\sum_{s=1}^n \Big\{\mathbf{K}_{h,sz}(z) - \mathrm{E}\Big[\mathbf{K}_{h,sz}(z)\Big]\Big\} \otimes X_{2s}X_{2s}^\top\Big)\mathbf{G}_h^{-1}. \\
&\equiv \mathcal{S}_{n1}(z) + \mathcal{S}_{n2}(z),
\end{aligned}$$

where $\mathbf{G}_h$ is defined in Section 2,

$$\begin{aligned}
\mathcal{S}_{n1}(z) &= \mathbf{G}_h^{-1}\Big\{n^{-2}\sum_{s=1}^n \mathrm{E}\Big[\mathbf{K}_{h,s}(z)\Big] \otimes X_{2s}X_{2s}^\top\Big\}\mathbf{G}_h^{-1}, \\
\mathcal{S}_{n2}(z) &= \mathbf{G}_h^{-1}\Big(n^{-2}\sum_{s=1}^n \eta_{h,s}(z) \otimes X_{2s}X_{2s}^\top\Big)\mathbf{G}_h^{-1}, \\
\eta_{h,s}(z) &= \mathbf{K}_{h,s}(z) - \mathrm{E}\Big[\mathbf{K}_{h,s}(z)\Big].
\end{aligned}$$

By Assumptions 4 and 5, we have, uniformly in $z \in \mathcal{S}_Z$,

$$\mathrm{E}\Big[\mathbf{K}_{h,s}(z)\Big] = f_Z(z)\Delta(\mu) + o(1),$$

where $\Delta(\mu)$ is a $(q+1)\times(q+1)$ matrix with the (i,j)-th element being μ_{i+j-2}. On the other hand, by Assumption 1, we can prove that

$$n^{-2}\sum_{s=1}^n X_{2s}X_{2s}^\top = n^{-1}\sum_{s=1}^n \frac{X_{2s}}{\sqrt{n}} \cdot \frac{X_{2s}^\top}{\sqrt{n}} \Rightarrow \int_0^1 B_2(r)B_2^\top(r)dr = O_P(1).$$

Noting that $f_Z(z)$ is bounded away from infinity and zero for $z \in \mathcal{S}_z$, we have, uniformly in $z \in \mathcal{S}_Z$,

$$\frac{1}{f_Z(z)}\mathcal{S}_{n1}(z) - \Delta(\mu) \otimes \Big(n^{-2}\sum_{s=1}^n X_{2s}X_{2s}^\top\Big) = o_P(1). \tag{A.2}$$

We next prove that $\mathcal{S}_{n2}(z)$ is $o_P(1)$ uniformly for $z \in \mathcal{S}_Z$. Let

$$Q_s^*(z) = \Big[1, \frac{Z_s - z}{h}, \cdots, \frac{(Z_s - z)^q}{h^q}\Big]^\top$$

and $\eta^*_{h,s}(z)$ be defined as $\eta_{h,s}(z)$ with $Q_s(z)$ replaced by $Q^*_s(z)$. Then, we can show that

$$\mathcal{S}_{n2}(z) = \frac{1}{n^2}\sum_{s=1}^{n}\eta^*_{h,s}(z)\otimes X_{2s}X_{2s}^{\top}.$$

As in Theorem 1 of Masry (1996), we can prove that

$$\sup_{l\geq 0}\sup_{z\in\mathcal{S}_Z}\mathrm{Var}\Big[\sum_{s=l+1}^{l+m}\eta^*_{h,s}(z)\Big] = O\Big(\frac{m}{h}\Big)$$

for all $m \geq 1$. For some $0<\delta<1$, set $N=[1/\delta]$, $s_k=[kn/N]+1$, $s^*_k = s_{k+1}-1$, and $s^{**}_k = \min\{s^*_k, n\}$. Let $U_{n,s} = X_{2s}X_{2s}^{\top}/n$ for any $1\leq s\leq n$ and $U_n(r) = U_{n,[nr]}$ for any $r\in[0,1]$. Following the proof of Theorem 3.3 of Hansen (1992), we have

$$\begin{aligned}
\sup_{z\in\mathcal{S}_Z}\|\mathcal{S}_{n2}(z)\| &= \sup_{z\in\mathcal{S}_Z}\Big\|\frac{1}{n^2}\sum_{s=1}^{n}\eta^*_{h,s}(z)\otimes U_{n,s}\Big\| = \sup_{z\in\mathcal{S}_Z}\Big\|\frac{1}{n^2}\sum_{k=0}^{N-1}\sum_{s=s_k}^{s^{**}_k}\eta^*_{h,s}(z)\otimes U_{n,s}\Big\| \\
&\leq \sup_{z\in\mathcal{S}_Z}\Big\|\frac{1}{n^2}\sum_{k=0}^{N-1}\sum_{s=s_k}^{s^{**}_k}\eta^*_{h,s}(z)\otimes U_{n,s_k}\Big\| + \sup_{z\in\mathcal{S}_Z}\Big\|\frac{1}{n^2}\sum_{k=0}^{N-1}\sum_{s=s_k}^{s^{**}_k}\eta^*_{h,s}(z)\otimes\Big(U_{n,s}-U_{n,s_k}\Big)\Big\| \\
&\leq \sup_{z\in\mathcal{S}_Z}\frac{1}{n^2}\sum_{k=0}^{N-1}\Big\|\sum_{s=s_k}^{s^{**}_k}\eta^*_{h,s}(z)\Big\|\cdot\Big\|U_{n,s_k}\Big\| + \sup_{z\in\mathcal{S}_Z}\frac{1}{n^2}\sum_{k=0}^{N-1}\sum_{s=s_k}^{s^{**}_k}\Big\|\eta^*_{h,s}(z)\Big\|\cdot\|U_{n,s}-U_{n,s_k}\| \\
&\leq \frac{1}{n}\sum_{k=0}^{N-1}\sup_{z\in\mathcal{S}_Z}\|\sum_{s=s_k}^{s^{**}_k}\eta^*_{h,s}(z)\|\sup_{0\leq r\leq 1}\|U_n(r)\| \\
&\quad + \sup_{|r_1-r_2|\leq\delta}\|U_n(r_1)-U_n(r_2)\|\cdot\sup_{z\in\mathcal{S}_Z}\frac{1}{n}\sum_{k=0}^{N-1}\sum_{s=s_k}^{s^{**}_k}\Big\|\eta^*_{h,s}(z)\Big\| \\
&\equiv \mathcal{S}_{n,21}+\mathcal{S}_{n,22}.
\end{aligned}$$

Note that $\sup_{0\leq r\leq 1}\|U_n(r)\| = O_p(1)$ as $U_n(r)\Rightarrow B_2(r)B_2^{\top}(r)$ by Assumption 1. Furthermore, following the argument in the proof of Theorem in Masry (1996), we have

$$\begin{aligned}
&\frac{1}{n}\sum_{k=0}^{N-1}\sup_{z\in\mathcal{S}_Z}\Big\|\sum_{s=s_k}^{s^*_k}\eta_{h,s}(z)\Big\| \leq \frac{N}{n}\sup_{0\leq k\leq N-1}\sup_{z\in\mathcal{S}_Z}\Big\|\sum_{s=s_k}^{s^*_k}\eta_{h,s}(z)\Big\| \\
&\leq \sup_{1\leq s\leq n}\sup_{z\in\mathcal{S}_Z}\Big\|\frac{1}{\delta n}\sum_{i=s}^{s+\delta n}\eta_{h,s}(z)\Big\| = o_P(1),
\end{aligned}$$

which implies that

$$\mathcal{S}_{n,21} = \sup_{0\leq r\leq 1}\|U_n(r)\|\cdot o_P(1) = o_P(1). \tag{A.3}$$

It is easy to see that uniformly for $z \in \mathcal{S}_Z$,

$$\frac{1}{n}\sum_{k=0}^{N-1}\sum_{s=s_k}^{s_k^{**}}\left\|\eta_{h,s}(z)\right\| = O_P(1),$$

which implies that

$$\mathcal{S}_{n,22} = \sup_{|r_1-r_2|\le\delta}\left\|U_n(r_1)-U_n(r_2)\right\|\cdot O_P(1) = o_P(1) \tag{A.4}$$

by letting $\delta \to 0$.

By using (A.3) and (A.4), we have shown that $\mathcal{S}_{n2}(z) = o_P(1)$ uniformly in $z \in \mathcal{S}_Z$, which, together with (A.2), leads to

$$\frac{1}{hf_Z(Z_t)}\mathcal{S}_n(Z_t) - \Delta(\mu)\otimes\left(n^{-2}\sum_{s=1}^{n}X_{2s}X_{2s}^{\top}\right) = o_P(1) \tag{A.5}$$

Similarly, we can also prove that, for any $t = 1,\cdots,n$,

$$\frac{1}{n^2hf_Z(Z_t)}\sum_{s=1}^{n}K_{h,st}\mathbf{G}_h^{-1}(Q_{s,t}\otimes X_{2s})X_{1s}^{\top} - (\mu_0,\mu_1,\cdots,\mu_q)^{\top}\otimes\left(n^{-2}\sum_{s=1}^{n}X_{2s}X_{1s}^{\top}\right) = o_P(1). \tag{A.6}$$

Then, by the definition of $\mathcal{A}_{1t}$, we can show that

$$\mathcal{A}_{1,t} - W_n(B_1,B_2) = o_P(1), \quad \text{uniformly in } t = 1,\cdots,n, \tag{A.7}$$

where

$$W_n(B_1,B_2) = \left[\mathbf{e}_1\Delta^{-1}(\mu)(\mu_0,\mu_1,\cdots,\mu_q)^{\top}\right]\otimes\left[\left(n^{-2}\sum_{s=1}^{n}X_{2s}X_{2s}^{\top}\right)^{-1}\left(n^{-2}\sum_{s=1}^{n}X_{2s}X_{1s}^{\top}\right)\right],$$

$\mathbf{e} = (1,0,\cdots,0)^{\top}$. By standard algebraic calculation, we have $\mathbf{e}_1\Delta(\mu)^{-1}\Gamma(\mu) \equiv 1$. Noting that $W_n(B_1,B_2) \Rightarrow W(B_1,B_2)$, by the definition of $\mathcal{B}_{1n}$ and (A.7), we can complete the proof of Proposition A.1. ■

Proposition A.2. *Under the conditions of Theorem 3.1, we have* $\mathcal{B}_{4n} = O_P(nh^{q+1})$.

Proof. For $\overline{M}_\beta(Z_t) = \mathcal{S}_t^{-1}\left[n^{-2}\sum_{s=1}^{n}K_{h,st}\mathbf{G}_h^{-1}(Q_{s,t}\otimes X_{2s})X_{2s}^{\top}\beta_0(Z_s)\right]$ and

$$M_\beta(Z_t) = \left[\beta(Z_t),\beta'(Z_t),\cdots,\frac{\beta^{(q)}(Z_t)}{q!}\right]^{\top},$$

we have

$$
\begin{aligned}
& \left\|\overline{M}_\beta(Z_t)-M_\beta(Z_t)\right\| \\
= \ & \left\|S_t^{-1}\left\{\frac{1}{n^2}\sum_{s=1}^n K_{h,st}G_h^{-1}(Q_{s,t}\otimes X_{2s})X_{2s}^\top\Big[\beta_0(Z_s)-\sum_{i=0}^q\frac{\beta_0^{(i)}(Z_t)}{i!}(Z_s-Z_t)^i\Big]\right\}\right\| \\
\leq \ & \frac{1}{n^2}\left\|hS_t^{-1}\right\|\sum_{s=1}^n\left\|\mathbf{G}_h^{-1}(Q_{s,t}\otimes X_{2s})\right\|\cdot\left\|X_{2s}\right\|\cdot\left\|\Big[\beta_0(Z_s)-\sum_{i=0}^q\frac{\beta_0^{(i)}(Z_t)}{i!}(Z_s-Z_t)^i\Big]h^{-1}K_{h,st}\right\| \\
= \ & O_P(h^{q+1}),
\end{aligned}
$$

where we use

$$
\mathrm{E}\left\|\Big[\beta_0(Z_s)-\sum_{i=0}^q\frac{\beta_0^{(i)}(Z_t)}{i!}(Z_s-Z_t)^i\Big]h^{-1}K_{h,st}\right\|=O(h^{q+1})
$$

in the last equality. Thus, we obtain that

$$
\overline{M}_\beta(Z_t)-M_\beta(Z_t)=O_P(h^{q+1}) \tag{A.8}
$$

uniformly for $t=1,\cdots,n$. Thus, we can further prove that

$$
\sup_{1\leq t\leq n}\left\|\beta_0(Z_t)-\overline{\beta}(Z_t)\right\|=\sup_{1\leq t\leq n}\left\|\mathbf{e}M_\beta(Z_t)-\mathbf{e}\overline{M}_\beta(Z_t)\right\|=O_P(h^{q+1}). \tag{A.9}
$$

Then, we have

$$
\begin{aligned}
\|\mathcal{B}_{4n}\| \ &\leq \ \sup_{1\leq t\leq n}\left\|\beta_0(Z_t)-\overline{\beta}(Z_t)\right\|\frac{1}{n}\sum_{t=1}^n\left\|X_{1t}-\mathcal{A}_{1t}^\top X_{2t}\right\|\cdot\left\|X_{2t}\right\| \\
&= \ O_P(h^{q+1})\cdot O_P(n)=O_P(nh^{q+1}),
\end{aligned}
$$

as $\left\|X_{1t}-\mathcal{A}_{1t}^\top X_{2t}\right\|+\left\|X_{2t}\right\|=O_P(n^{1/2})$. We then complete the proof of Proposition A.2. ■

Proposition A.3. *Under the conditions of Theorem 3.1, we have*

$$
\mathcal{B}_{2n}-\mathcal{B}_{3n}\Rightarrow\int_0^1\Big[B_1(r)-W^\top(B_1,B_2)B_2(r)\Big]dB_3(r), \tag{A.10}
$$

where $W(B_1,B_2)$ is defined in Section 3.

Proof. Observe that

$$
\begin{aligned}
\mathcal{B}_{3n} \ &= \ \frac{1}{n}\sum_{t=1}^n(X_{1t}-\mathcal{A}_{1t}^\top X_{2t})X_{2t}^\top\widetilde{u}_t, \\
&= \ \frac{1}{n}\sum_{t=1}^n(X_{1t}-\mathcal{A}_{1t}^\top X_{2t})X_{2t}^\top\left\{\mathbf{e}\mathcal{S}_t^{-1}\left[\frac{1}{n^2}\sum_{s=1}^nK_{h,st}\mathbf{G}_h^{-1}(Q_{s,t}\otimes X_{2s})u_s\right]\right\} \\
&= \ \frac{1}{n}\sum_{s=1}^n\left[\frac{1}{n^2}\sum_{t=1}^n(X_{1t}-\mathcal{A}_{1t}^\top X_{2t})X_{2t}^\top\left(\mathbf{e}\mathcal{S}_t^{-1}K_{h,st}\mathbf{G}_h^{-1}(Q_{s,t}\otimes X_{2s})\right)\right]u_s. \quad (\text{A.11})
\end{aligned}
$$

Similar to the proof of Proposition A.1, we can prove that, uniformly in $s = 1, \cdots, n$,

$$\begin{aligned}
&\frac{1}{n^2}\sum_{t=1}^{n}(X_{1t} - \mathcal{A}_{1t}^{\top}X_{2t})X_{2t}^{\top}\left[\mathbf{e}\mathcal{S}_t^{-1}K_{h,st}\mathbf{G}_h^{-1}(Q_{s,t}\otimes\frac{X_{2s}}{\sqrt{n}})\right] \qquad \text{(A.12)}\\
= \;& \frac{1}{n^2}\sum_{t=1}^{n}\left[X_{1t} - W_n^{\top}(B_1,B_2)X_{2t}\right]X_{2t}^{\top}\left[\mathbf{e}\mathcal{S}_t^{-1}K_{h,st}\mathbf{G}_h^{-1}(Q_{s,t}\otimes\frac{X_{2s}}{\sqrt{n}})\right] + o_P(1)\\
= \;& \frac{1}{n^2}\sum_{t=1}^{n}\left[X_{1t} - W_n^{\top}(B_1,B_2)X_{2t}\right]X_{2t}^{\top}\left\{\mathbf{e}\left[\Delta(\mu)\otimes\left(n^{-2}\sum_{t=1}^{n}X_{2t}X_{2t}^{\top}\right)\right]^{-1}\left[m(\mu)\otimes\frac{X_{2s}}{\sqrt{n}}\right]\right\} + o_P(1),
\end{aligned}$$

where $m(\mu) = (\mu_0, \mu_1, \cdots, \mu_q)^{\top}$.

Let

$$V_{ns} = \frac{1}{n^2}\sum_{t=1}^{n}\left[X_{1t} - W_n^{\top}(B_1,B_2)X_{2t}\right]X_{2t}^{\top}\left\{\mathbf{e}\left[\Delta(\mu)\otimes\left(n^{-2}\sum_{t=1}^{n}X_{2t}X_{2t}^{\top}\right)\right]^{-1}\left[m(\mu)\otimes\frac{X_{2s}}{\sqrt{n}}\right]\right\},$$

and

$$\Theta_{ns} = \frac{1}{n^2}\sum_{t=1}^{n}(X_{1t} - \mathcal{A}_{1t}^{\top}X_{2t})X_{2t}^{\top}\left[\mathbf{e}\mathcal{S}_t^{-1}K_{h,st}\mathbf{G}_h^{-1}(Q_{s,t}\otimes\frac{X_{2s}}{\sqrt{n}})\right] - V_{ns}.$$

By (A.11) and (A.12), we can write that

$$\mathcal{B}_{3n} = \frac{1}{\sqrt{n}}\sum_{s=1}^{n}V_{ns}u_s + \frac{1}{\sqrt{n}}\sum_{s=1}^{n}\Theta_{ns}u_s, \qquad \text{(A.13)}$$

where $\Theta_{ns} = o_P(1)$ uniformly in $s = 1, \cdots, n$.

For any $\epsilon_1 > 0$ and $\epsilon_2 > 0$, we have

$$\begin{aligned}
&\mathrm{P}\left\{\left\|\frac{1}{\sqrt{n}}\sum_{s=1}^{n}\Theta_{ns}u_s\right\| > \epsilon_1\right\}\\
= \;& \mathrm{P}\left\{\left\|\frac{1}{\sqrt{n}}\sum_{s=1}^{n}\Theta_{ns}u_s\right\| > \epsilon_1,\ \max_s\|\Theta_{ns}\| > \epsilon_2\right\} +\\
& \mathrm{P}\left\{\left\|\frac{1}{\sqrt{n}}\sum_{s=1}^{n}\Theta_{ns}u_s\right\| > \epsilon_1,\ \max_s\|\Theta_{ns}\| \le \epsilon_2\right\}\\
\le \;& \mathrm{P}\left\{\left\|\frac{1}{\sqrt{n}}\sum_{s=1}^{n}\Theta_{ns}u_s\right\| > \epsilon_1,\ \max_s\|\Theta_{ns}\| \le \epsilon_2\right\} + \mathrm{P}\left\{\max_s\|\Theta_{ns}\| > \epsilon_2\right\}\\
= \;& \frac{\mathrm{E}\left\{\left\|\sum_{s=1}^{n}\Theta_{ns}u_s\right\|^2 I\left(\max_s\|\Theta_{ns}\| \le \epsilon_2\right)\right\}}{n\epsilon_1^2} + o(1). \qquad \text{(A.14)}
\end{aligned}$$

By Assumptions 1 and 2, we can show that

$$\frac{1}{n}\mathrm{E}\Big\{\Big\|\sum_{s=1}^{n}\Theta_{ns}u_s\Big\|^2 I\Big(\max_s\|\Theta_{ns}\|\le\epsilon_2\Big)\Big\}=o(1) \tag{A.15}$$

by letting $\epsilon_2\to 0$. Further, we can see that

$$\begin{aligned}
V_{ns} &= \frac{1}{n^2}\sum_{t=1}^{n}\Big[X_{1t}-W_n^\top(B_1,B_2)X_{2t}\Big]X_{2t}^\top\mathbf{e}\Big[\Delta(\mu)\otimes\Big(\frac{1}{n^2}\sum_{t=1}^{n}X_{2t}X_{2t}^\top\Big)\Big]^{-1}\Big[m(\mu)\otimes\frac{X_{2s}}{\sqrt{n}}\Big]\\
&= \frac{1}{n^2}\sum_{t=1}^{n}\Big[X_{1t}-\Big[\Big(\frac{1}{n^2}\sum_{s=1}^{n}X_{2s}X_{2s}^\top\Big)^{-1}\Big(\frac{1}{n^2}\sum_{s=1}^{n}X_{2s}X_{1s}^\top\Big)\Big]^\top X_{2t}\Big]X_{2t}^\top\Big[\Big(\frac{1}{n^2}\sum_{t=1}^{n}X_{2t}X_{2t}^\top\Big)^{-1}\frac{X_{2s}}{\sqrt{n}}\Big]\\
&= \Big[\frac{1}{n^2}\sum_{t=1}^{n}X_{1t}X_{2t}^\top-\Big(\frac{1}{n^2}\sum_{s=1}^{n}X_{1s}X_{2s}^\top\Big)\Big(\frac{1}{n^2}\sum_{s=1}^{n}X_{2s}X_{2s}^\top\Big)^{-1}\frac{1}{n^2}\sum_{t=1}^{n}X_{2t}X_{2t}^\top\Big]\Big[\Big(\frac{1}{n^2}\sum_{t=1}^{n}X_{2t}X_{2t}^\top\Big)^{-1}\frac{X_{2s}}{\sqrt{n}}\Big]\\
&\equiv 0.
\end{aligned} \tag{A.16}$$

By using (A.13)–(A.16), we prove that

$$\mathcal{B}_{3n}=o_P(1). \tag{A.17}$$

Since

$$\begin{aligned}
\mathcal{B}_{2n} &= \frac{1}{n}\sum_{s=1}^{n}(X_{1s}-\mathcal{A}_{1s}^\top X_{2s})u_s\\
&= \frac{1}{\sqrt{n}}\sum_{s=1}^{n}\Big(\frac{X_{1s}}{\sqrt{n}}-W_n(B_1,B_2)^\top\frac{X_{2s}}{\sqrt{n}}\Big)u_s+\frac{1}{n}\sum_{s=1}^{n}\Phi_{ns}u_s\\
&\equiv \frac{1}{\sqrt{n}}\sum_{s=1}^{n}U_{ns}^*u_s+\frac{1}{n}\sum_{s=1}^{n}\Phi_{ns}u_s,
\end{aligned}$$

where $U_{ns}^*=\frac{X_{1s}}{\sqrt{n}}-W_n(B_1,B_2)^\top\frac{X_{2s}}{\sqrt{n}}$, $\Phi_{ns}=\frac{X_{1s}}{\sqrt{n}}-\mathcal{A}_{1t}^\top\frac{X_{2s}}{\sqrt{n}}-U_{ns}^*$, and by (A.7) we have $\|\Phi_{ns}\|=o_P(1)$ uniformly in $s=1,\cdots,n$. Similar to (A.14) and (A.15), we then have that $\frac{1}{n}\sum_{s=1}^{n}\Phi_{ns}u_s=o_P(1)$.

By (A.17) and the definition of $\mathcal{B}_{2n}$, we have

$$\mathcal{B}_{2n}-\mathcal{B}_{3n}=\frac{1}{\sqrt{n}}\sum_{s=1}^{n}U_{ns}^*u_s+o_P(1),$$

which leads to (A.10). We thus complete the proof of Proposition A.3. ■

We next turn to the proof of Theorem 3.2. Similar to Lemma 2 in Xiao (2009) and Lemma 2 in Gu and Liang (2013), we have the following joint convergence result.

Proposition A.4. *Under Assumptions 1-6 in Section 3, we have that for $r \in [0,1]$*

$$\begin{pmatrix} \frac{1}{\sqrt{n}} \sum_{s=1}^{[nr]} x_{2s} \\ \frac{1}{\sqrt{nh}} \sum_{s=1}^{[nr]} u_s K_{h,s}(z) \end{pmatrix} \Longrightarrow \begin{pmatrix} B_2(r) \\ B_{u,z}(r) \end{pmatrix} \equiv \tilde{B}(r), \tag{A.18}$$

where $\tilde{B}(\cdot)$ is a multivariate Brownian motion with zero mean and variance-covariance matrix

$$\Omega(z) = \begin{pmatrix} \Omega_{22} & 0 & \cdots & 0 \\ 0 & \nu_0 f_Z(z)\sigma_u^2 & \cdots & \nu_q f_Z(z)\sigma_u^2 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \nu_q f_Z(z)\sigma_u^2 & \cdots & \nu_{2q} f_Z(z)\sigma_u^2 \end{pmatrix},$$

and $\Omega_{22} = \sum_{k=1}^{\infty} \mathrm{E}\Big[x_{2t}x_{2,t+k}^{\top}\Big]$.

Proof of Theorem 3.2. Note that d_2 is assumed to be 1 without loss of generality. From (2.3), we have that

$$\widehat{\beta}(z) = \mathcal{A}_2(z) - \mathcal{A}_1(z)\widehat{\gamma},$$

where

$$\begin{aligned} \mathcal{A}_1(z) &= \mathbf{e}_1 \mathcal{S}^{-1}(z)\Big[\frac{1}{n^2}\sum_{s=1}^{n} K_{h,s}(z)\mathbf{G}_h^{-1} Q_s(z) X_{2s} X_{1s}^{\top}\Big], \\ \mathcal{A}_2(z) &= \mathbf{e}_1 \mathcal{S}^{-1}(z)\Big[\frac{1}{n^2}\sum_{s=1}^{n} K_{h,s}(z)\mathbf{G}_h^{-1} Q_s(z) X_{2s} Y_s\Big], \\ \mathcal{S}(z) &= \frac{1}{n^2}\sum_{s=1}^{n} K_{h,s}(z)\mathbf{G}_h^{-1} Q_s(z) Q_s^{\top}(z) X_{2s} X_{2s} \mathbf{G}_h^{-1}, \end{aligned}$$

and as in the proof of Proposition A.1, $Z_{s,h}(z) = (Z_s - z)/h$, $K_{h,s}(z) = K(Z_{s,h}(z))$, $Q_s(z) = \Big[1, (Z_s - z), \cdots, (Z_s - z)^q\Big]^{\top}$. As $d_2 = 1$, it is easy to see that $\mathbf{G}_h = \mathrm{diag}\Big(1, h, \cdots, h^q\Big)$ and $\mathbf{e}$ is reduced to be $\mathbf{e}_1$.

Then we have

$$
\begin{aligned}
\widehat{\beta}(z)-\beta_0(z) &= \mathcal{A}_2(z)-\mathcal{A}_1(z)\widehat{\gamma}-\beta_0(z)\\
&= \mathbf{e}_1\mathcal{S}^{-1}(z)\Big[\frac{1}{n^2}\sum_{s=1}^{n}K_{h,s}(z)\mathbf{G}_h^{-1}Q_s(z)X_{2s}(Y_s-X_{1s}^{\top}\widehat{\gamma})\Big]-\beta_0(z)\\
&= \mathbf{e}_1\mathcal{S}^{-1}(z)\Big(\frac{1}{n^2}\sum_{s=1}^{n}K_{h,s}(z)\mathbf{G}_h^{-1}Q_s(z)X_{2s}\Big\{X_{1s}^{\top}\gamma_0+X_{2s}[\beta_0(Z_s)-\beta_0(z)]+u_s-X_{1s}^{\top}\widehat{\gamma}\Big\}\Big)\\
&= \mathbf{e}_1\mathcal{S}^{-1}(z)\Big[\frac{1}{n^2}\sum_{s=1}^{n}K_{h,s}(z)\mathbf{G}_h^{-1}Q_s(z)X_{2s}u_s\Big]+\\
&\quad\ \mathbf{e}_1\mathcal{S}^{-1}(z)\Big[\frac{1}{n^2}\sum_{s=1}^{n}K_{h,s}(z)\mathbf{G}_h^{-1}Q_s(z)X_{2s}X_{1s}^{\top}\Big](\gamma_0-\widehat{\gamma})+\\
&\quad\ \mathbf{e}_1\mathcal{S}^{-1}(z)\Big\{\frac{1}{n^2}\sum_{s=1}^{n}K_{h,s}(z)\mathbf{G}_h^{-1}Q_s(z)X_{2s}^2\Big[\beta_0(Z_s)-\sum_{j=1}^{q}\beta_0^{(j)}(z)\frac{(Z_s-z)^j}{j!}\Big]\Big\}\\
&\equiv \mathbf{e}_1\mathcal{S}^{-1}(z)\mathcal{C}_1(z)+\mathbf{e}_1\mathcal{S}^{-1}(z)\mathcal{C}_2(z)+\mathbf{e}_1\mathcal{S}^{-1}(z)\mathcal{C}_3(z).
\end{aligned}
$$

We first consider the convergence result for $\mathcal{S}(z)$. From (A.5), we have that

$$\frac{1}{hf_Z(z)}\mathcal{S}(z)-\Delta(\mu)\Big(\frac{1}{n^2}\sum_{s=1}^{n}X_{2s}X_{2s}^{\top}\Big)=o_P(1). \tag{A.19}$$

We next consider $\mathcal{C}_1(z)$. Following Theorem 3.1 of Hansen (1992) and (A.18), by noting that $w_{js}(z)=Z_{s,h}^{j}(z)u_sh^{-1/2}K_{h,s}(z)$ is a martingale difference sequence with respect to $\mathcal{F}_{ns}$ defined in Assumption 2, we can see that for the typical element of $nh^{-1/2}\mathcal{C}_1(z)$

$$\frac{1}{\sqrt{nh}}\sum_{s=1}^{n}\Big(\frac{Z_s-z}{h}\Big)^jK_{h,s}(z)\frac{X_{2s}}{\sqrt{n}}u_s\Rightarrow\int_0^1B_2(s)dB_{uz,j}(s),\ \ j=0,1,\cdots,q,$$

where $B_{uz}(s)=\Big[B_{uz,0}(s),B_{uz,1}(s),\cdots,B_{uz,q}(s)\Big]^{\top}$. Furthermore, we have that

$$nh^{-1/2}\mathcal{C}_1(z)\Rightarrow\sigma_uf_Z^{1/2}(z)\Gamma^{1/2}(\nu)\int_0^1B_2(s)dW(s),$$

where $\Gamma(\nu)$ is defined in Section 3 and $W(s)$ is a $(q+1)$-dimensional standard Brownian motion. Taking together with (A.19), we obtain

$$n\sqrt{h}\mathbf{e}_1\mathcal{S}^{-1}(z)\mathcal{C}_1(z)\Rightarrow\sigma_u\frac{\mathbf{e}_1\Delta^{-1}(\mu)\Gamma^{1/2}(\nu)}{f_Z^{1/2}(z)\int_0^1B_2^2(s)ds}\int_0^1B_2(s)dW(s). \tag{A.20}$$

Similar as the proof of (A.2), for the typical element of $\mathcal{C}_2(z)$, we have

$$\begin{aligned}\frac{1}{n^2}\sum_{s=1}^{n} Z_{s,h}^{j}(z)K_{h,s}(z)X_{2s}X_{1s}^{\top} &= \frac{1}{n}\sum_{s=1}^{n}\left\{Z_{s,h}^{j}(z)K_{h,s}(z)-\mathrm{E}\left[Z_{s,h}^{j}(z)K_{h,s}(z)\right]\right\}\frac{X_{2s}}{\sqrt{n}}\frac{X_{1s}^{\top}}{\sqrt{n}}+\\ &\quad \frac{1}{n}\sum_{s=1}^{n}\mathrm{E}\left[Z_{s,h}^{j}(z)K_{h,s}(z)\right]\frac{X_{2s}}{\sqrt{n}}\frac{X_{1s}^{\top}}{\sqrt{n}}\\ &= O_P(h).\end{aligned}$$

Thus, we have

$$\frac{1}{n^2}\sum_{s=1}^{n} K_{h,s}(z)\mathbf{G}_h^{-1}Q_s(z)X_{2s}X_{1s}^{\top} = O_P(h).$$

From Theorem 3.1, $\widehat{\gamma}-\gamma = O_P(n^{-1})$, then we have

$$n\sqrt{h}\mathbf{e}_1\mathcal{S}^{-1}(z)\mathcal{C}_2(z) = O_P(n\sqrt{h}h^{-1}hn^{-1}) = O_P(\sqrt{h}) = o_P(1). \tag{A.21}$$

We finally consider $\mathcal{C}_3(z)$. For the typical element of $h^{-1}\mathcal{C}_3(z)$, following similar arguments as (A.3) and (A.4) in the proof of Proposition A.1, we have

$$\begin{aligned}&\frac{1}{n}\sum_{s=1}^{n}\frac{X_{2s}^2}{n}\left(\frac{Z_s-z}{h}\right)^{j}B_q(Z_s,z)h^{-1}K_{h,s}(z)\\ =\ & \frac{1}{n}\sum_{s=1}^{n}\frac{X_{2s}^2}{n}\mathrm{E}\left[\left(\frac{Z_s-z}{h}\right)^{j}B_q(Z_s,z)h^{-1}K_{h,s}(z)\right]+\\ & \frac{1}{n}\sum_{s=1}^{n}\frac{X_{2s}^2}{n}\left\{\left(\frac{Z_s-z}{h}\right)^{j}B_q(Z_s,z)h^{-1}K_{h,sz}-\mathrm{E}\left[\left(\frac{Z_s-z}{h}\right)^{j}B_q(Z_s,z)h^{-1}K_{h,s}(z)\right]\right\}\\ =\ & \frac{h^{q+1}f_Z(z)\mu_{j+q+1}}{(q+1)!}\beta_0^{(q+1)}(z)\int_0^1 B_2^2(s)ds+o_P(h^{q+1}),\end{aligned}$$

where $B_q(Z_s,z)=\beta_0(Z_s)-\sum_{j=1}^{q}\beta_0^{(j)}(z)\frac{(Z_s-z)^j}{j!}$. Hence, we obtain

$$h^{-1}\mathcal{C}_3(z)=\frac{h^{q+1}f_Z(z)\beta_0^{(q+1)}(z)}{(q+1)!}\left(\mu_{q+1},\mu_{q+2},\cdots,\mu_{2q+1}\right)^{\top}\int_0^1 B_2^2(s)ds+o_P(h^{q+1}).$$

Therefore, we have

$$\mathbf{e}_1\mathcal{S}^{-1}(z)\mathcal{C}_3(z)-\frac{h^{q+1}\beta_0^{(q+1)}(z)}{(q+1)!}\mathbf{e}_1\Delta^{-1}(\mu)\left(\mu_{q+1},\mu_{q+2},\cdots,\mu_{2q+1}\right)^{\top}=o_P(h^{q+1}). \tag{A.22}$$

Combining (A.20)–(A.22) and using the continuous mapping theorem (Billingsley, 1999), we have

$$
\begin{aligned}
& n\sqrt{h}\left[\widehat{\beta}(z) - \beta_0(z) - b_z\right] \\
= \;& n\sqrt{h}\Big\{\mathbf{e}_1\mathcal{S}^{-1}(z)\mathcal{C}_1(z) + \mathbf{e}_1\mathcal{S}^{-1}(z)\mathcal{C}_2(z) + \left[\mathbf{e}_1\mathcal{S}^{-1}(z)\mathcal{C}_3(z) - b_z\right]\Big\} \\
= \;& n\sqrt{h}\mathbf{e}_1\mathcal{S}^{-1}(z)\mathcal{C}_1(z) + o_P(1) \\
\Rightarrow \;& \sigma_u \frac{\mathbf{e}_1\Delta^{-1}(\mu)\Gamma^{1/2}(\nu)}{f_Z^{1/2}(z)\int_0^1 B_2^2(s)ds}\int_0^1 B_2(s)dW(s) \\
\equiv \;& \mathrm{MN}(\Sigma_z),
\end{aligned}
$$

where

$$
b_z = \frac{h^{q+1}\beta_0^{(q+1)}(z)}{(q+1)!}\mathbf{e}_1\Delta^{-1}(\mu)\left(\mu_{q+1}, \mu_{q+2}, \cdots, \mu_{2q+1}\right)^\top
$$

and

$$
\Sigma_z = \frac{\sigma_u^2\mathbf{e}_1\Delta^{-1}(\mu)\Gamma(\nu)\Delta^{-1}(\mu)\mathbf{e}_1^\top}{f_Z(z)\int_0^1 B_2^2(s)ds}.
$$

This completes the proof of Theorem 3.2. ■

REFERENCES

Billingsley, P. (1999) *Convergence of Probability Measures* (2nd Edition). Wiley Series in Probability and Statistics.

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

Chen, J., D. Li, and L. Zhang (2010) Robust estimator in a nonlinear cointegration model. *Journal of Multivariate Analysis* 101, 706-717.

Chen, J., J. Gao, and D. Li (2012) Estimation in semiparametric regression with nonstationary regressors. *Bernoulli* 18, 678-702.

Fan, J., and I. Gijbels (1996) *Local Polynomial Modelling and Its Applications*. London: Chapman & Hall/CRC.

Fan. J., and T. Huang (2005) Profile likelihood inference on semiparametric varying coefficient partially linear models. *B*ernoulli 11, 1031-1059.

Gu, J., and Z. Liang (2013) Testing cointegration relationship in a semiparametric varying coefficient model. *Journal of Econometrics*, forthcoming.

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.

Li, D., J. Chen, and Z. Lin (2011) Statistical inference in partially time-varying coefficient models. *Journal of Statistical Planning and Inference* 141, 995-1013.

Li, Q., and J.S. Racine (2007) Nonparametric Econometrics: Theory and Practice, Princeton University Press, Princeton.

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 S.B. Hahn (1999) Cointegrating regressions with time varying coefficients. *E*conometric Theory 15, 664-703.

Park, J. Y., and P. C. B. Phillips (1988) Statistical inference in regression with integrated processes: part 1. *E*conometric Theory 4, 468-497.

Park, J. Y., and P. C. B. Phillips (1989) Statistical inference in regression with integrated processes: part 2. *E*conometric Theory 5, 95-131.

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

Phillips, P.C.B. (1986) Optimal inference in cointegrated systems. *Econometrica* 59, 283-306.

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

Phillips, P.C.B., D. Li, and J. Gao (2013) Estimating smooth structural change in cointegration models. Cowles Foundation Discussion Paper No. 1910, Yale University.

Saikkonen, P. (1995) Problems with the asymptotic theory of maximum likelihood estimation in integrated and cointegrated systems. *Econometric Theory* 11, 888-911.

Sun, Y., and Q. Li (2011) Data-driven bandwidth selection for nonstationary semiparametric models. *Journal of Business and Economic Statistics* 29, 541-551.

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

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

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

Zhou, Y., and H. Liang (2009) Statistical inference for semiparametric varying-coefficient partially linear models with error-prone linear covariates. *The Annals of Statistics* 37, 427-458.