Nonlinear estimators with integrated regressors but without exogeneity

Robert M. de Jong*
Department of Economics
217 Marshall Hall
Michigan State University
East Lansing, MI 48824

Preliminary, incomplete

March 2002

Abstract

This paper analyzes nonlinear cointegrating regressions as have been recently analyzed in a paper by Park and Phillips in *Econometrica*. I analyze the consequences of removing Park and Phillips' exogeneity assumption, which for the special case of a linear model would imply the asymptotic validity of the least squares estimator for linear cointegrating regressions. For the linear model, the unlikeliness of such an exogeneity assumption to hold in practice has inspired the "fully modified" technique, the "leads and lags" technique, and Park's "canonical regressions". In this paper, a "fully modified" type technique is proposed for nonlinear cointegrating regressions. The mathematical tool for proving this result is a new so-called "convergence to stochastic integrals" result. This result is proven for objects that are summations of a stationary random variable times an asymptotically homogeneous function of an integrated process. The increments of the integrated process are allowed to be correlated with the stationary random variable. This result is derived by extending results by Chan and Wei and by de Jong and Davidson.

*The comments of Peter Phillips and Alex Maynard on an earlier version of this paper are gratefully acknowledged.

1 Introduction

This paper analyzes nonlinear least squares estimation of the nonlinear cointegrating regression

$$y_t = f(x_t, \theta_0) + u_t \tag{1}$$

where u_t is assumed to be stationary and the scalar $x_t = \sum_{j=1}^{t} v_t$ is assumed to be an integrated process. For the case of a linear response function $f(x, \theta) = x'\theta$, the analysis of such cointegrating regressions is well-established. It can be shown that if x_t satisfies an exogeneity condition with respect to u_t, mixed normality of the nonlinear least squares estimator results. This implies that t-values, F-tests and confidence intervals will be asymptotically correct for the linear model if the exogeneity assumption holds.

In practice however, the exogeneity assumption is very unlikely to hold. ¿From the Granger-Engle representation theorem, it follows that typically, no exogeneity assumption can be expected to hold in a linear cointegrating regression. This motivates, for linear models, three techniques that are capable of dealing with cointegrating regression errors u_t that are possibly correlated with v_t. Those three techniques are the so-called "leads and lags" technique (see Saikkonen (1991), Phillips and Loretan (1991), and Stock and Watson (1993)), the "fully modified" technique (see Phillips and Hansen (1990)), and Park's canonical regressions (see Park (1992)).

For nonlinear least squares estimation of a nonlinear cointegrating regression, asymptotic theory under an exogeneity assumption is relatively recent. Park and Phillips (2001) have analyzed nonlinear cointegrating regressions, using tools developed in Park and Phillips (1999). Park and Phillips (1999) prove convergence results for objects such as

$$a_T \sum_{t=1}^{T} F(x_t) \tag{2}$$

for an appropriate scaling a_T and functions $F(.)$ that are in three different function classes (integrable, asymptotically homogeneous, and explosive functions). Note that the central difficulty in establishing asymptotic results for the above statistic is the fact that the integrated process has not been rescaled by $T^{-1/2}$. These results provide the building blocks for the development of the theory in Park and Phillips (2001). Moreover, this theory has also been applied in Park and Phillips (2000) and Chang (2000).

One aspect of Park and Phillips' (2001) theory is that it uses an exogeneity assumption that for the linear model would imply validity of standard least squares inference results. In this paper, I analyze the asymptotic consequences of the failure of the exogeneity assumption, and in addition I propose a "fully modified" type technique for nonlinear least squares estimation of a cointegrating regression. For the special case of the linear model, the proposed technique simplifies to the usual "fully modified" estimator as proposed by Phillips and Hansen (1990).

The mathematical problem solved in Section 2 of this paper, that will be the key to my results, is establishing the limit distribution for statistics of the form

$$T^{-1/2}\sum_{t=1}^{T-1} u_{t+1} H(T^{-1/2}\sum_{l=1}^{t} v_l) \tag{3}$$

where $H(.)$ is a continuously differentiable function. This result is essentially extended to provide a limit theory for statistics such as

$$\sum_{t=1}^{T-1} u_{t+1} F(\sum_{l=1}^{t} v_l) \tag{4}$$

where $F(.)$ is a function that is supposed to satisfy certain conditions (i.e. $F(.)$ is supposed to be *asymptotically homogeneous*, in the terminology of Park and Phillips (1999)), $Ev_t = Eu_t = 0$, and u_t and v_t are allowed to be weakly dependent and correlated among each other. This problem has been considered for the case of u_t and v_t that are uncorrelated (Park and Phillips (1999, 2001)), but to the best of this author's knowledge, no attempt has ever been made to solve the more general problem that is tackled in this paper.

If the u_t are independent of the v_t and if (u_t, v_t) is i.i.d., we can intuitively reason as follows. Define $\sigma_u = Eu_t^2$ and $\sigma_v = Ev_t^2$. The statistic of Equation (3), conditional on the v_t, may converge to a normal with a variance $\sigma_u^2 T^{-1}\sum_{t=1}^{T} H(T^{-1/2}\sum_{l=1}^{t} v_l)^2$, which satisfies

$$\sigma_u^2 T^{-1}\sum_{t=1}^{T} H(T^{-1/2}\sum_{l=1}^{t} v_l)^2 \xrightarrow{d} \sigma_u^2 \int_0^1 H(V(r))^2 dr \tag{5}$$

where $U(.)$ is the limit Brownian motion of $T^{-1/2}\sum_{l=1}^{[Tr]} u_l$, and $V(.)$ is the limit Brownian motion of $T^{-1/2}\sum_{l=1}^{[Tr]} v_l$. Therefore, it seems reasonable to expect convergence of the above statistic to

$$N(0,\sigma_u^2) \times (\int_0^1 H(V(r))^2 dr)^{1/2} \tag{6}$$

where both terms are independent. This intuition can be formalized, and the above limit distribution can then also be described as

$$\int_0^1 H(V(r)) dU(r), \tag{7}$$

which is a so-called stochastic integral. Park and Phillips (1999, 2001) analyzed nonlinear estimators in the presence of nonlinearities, and they derive limit distributions of the above type by formalizing the heuristic argument that I presented above, and they do not need the total independence requirement. However, their results are unable to handle u_t and v_t that are correlated, and I will show that the limit distribution of the estimator that they propose will be different in that case.

The analytical tool used in the proof is an extension of the "convergence to stochastic integrals" proof of De Jong and Davidson (2000) to the present situation. The proof in De Jong and Davidson (2000) again extends a proof by Chan and Wei (1988).

In Section 3 of this paper, the result of Equation (3) is used to calculate the limit distribution of the least squares estimator when the error is stationary and the regressor is of the form $|\sum_{l=1}^{t} v_l|^{\gamma}$ - i.e., a power of an integrated process - where it is assumed that $\gamma > 1$.

In Section 4 of this paper, the result of Equation (4) is applied to the nonlinear cointegrating regressions as studied by Park and Phillips (2001). By applying the second result described above, I calculate the limit distribution for nonlinear least squares estimators for the case where the long-run regression error and the increments of the integrated regressor are possibly correlated. The limit distribution of the nonlinear least squares estimator, as in the standard linear case, implies that standard least squares procedures will give invalid inference results in the general case where correlation is allowed. In Section 5 of this paper, I propose a "fully modified" type technique for nonlinear cointegrating regressions that will make it possible to conduct valid inference in nonlinear cointegrating regressions.

2 Mathematical results

In order to establish the results of this section, I will need assumptions on the function $F(.)$ and on the weak dependence structure of (u_t, v_t). In order to limit the amount of dependence that will be allowed, we need the concept of *near epoch dependence*. An array of random variables y_{Tt} is called near epoch dependent (NED) on w_t if

$$\sup_{t,T} \| y_{Tt} - E(y_{Tt}|w_{t-m}, \ldots, w_{t+m}) \|_2 \leq \nu(m) \to 0 \tag{8}$$

as $m \to \infty$. Typically, we need to make mixing (or independence) assumptions about w_t in order to be able to derive useful results such as laws of large numbers of central limit theorems. I will not define α-mixing or ϕ-mixing here, but instead note that Gallant and White (1988), Pötscher and Prucha (1991) and Davidson (1994) contain a large amount of information about these dependence concepts.

The weak dependence assumption on (u_t, v_t) that I need in this section is the following:

Assumption 1

1. *(u_t, v_t) is stationary.*

2. *For some $p > 2$, $E|u_t|^p + E|v_t|^p < \infty$.*

3. *$Eu_t = 0$ and $Ev_t = 0$.*

4. *(u_t, v_t) is L_2-NED of size -1 on w_t, where w_t is an α-mixing array of size $-2p/(p-2)$, or (u_t, v_t) is L_2-NED of size -1 on w_t, where w_t is a ϕ-mixing array of size $-p/(p-1)$.*

Define $V_T(r) = T^{-1/2} \sum_{t=1}^{[rT]} v_t$ and $U_T(r) = T^{-1/2} \sum_{t=1}^{[rT]} u_t$. I will equip $D[0,1]^2$ with the uniform metric, and by the Skorokhod representation, we can without loss of generality assume that $\sup_{r\in[0,1]}(|V_T(r) - V(r)| + |U_T(r) - U(r)|) \to 0$ almost surely.
The assumption on $F(.)$ that I need is the following:

Assumption 2

1. *$F : \mathbb{R} \longrightarrow \mathbb{R}$ is continuous on $\mathbb{R}$.*

2. *$F(\lambda x) = \kappa(\lambda)H(x) + R(x, \lambda)$ where $H(.)$ is locally integrable and $R(.)$ is such that*

   (a) *$|R(x, \lambda)| \leq a(\lambda)P(x)$, where $\limsup_{\lambda\to\infty} a(\lambda)/\kappa(\lambda) = 0$ and $P(.)$ is locally integrable; or*

   (b) *$|R(x, \lambda)| \leq b(\lambda)Q(\lambda x)$, where $\limsup_{\lambda\to\infty} b(\lambda)/\kappa(\lambda) < \infty$ and $Q(.)$ is locally integrable and vanishes at infinity, i.e., $Q(x) \to 0$ as $|x| \to \infty$.*

3. *$H'(.)$ is continuous on $\mathbb{R}$.*

4. *For any sequence δ_T such that $\delta_T \to 0$ and $n \to \infty$,*

$$\limsup_{T\to\infty} \kappa(T^{1/2})^{-2} \sup_{|x|\leq K} \sup_{x':|x-x'|\leq\delta_T} |F(T^{1/2}x) - F(T^{1/2}x')|^2 = 0.$$

Park and Phillips (1999) give a wide range of examples of asymptotically homogeneous functions. Distribution functions, polynomials, logarithmic functions and summations and products of such functions will all be within the class of asymptotically homogeneous functions. Integrable functions, periodic functions and functions with explosive behavior as the argument goes to infinity - such as the exponential - are outside the class of asymptotically

homogeneous functions.
Let

$$G_T = T^{-1/2}(\kappa(T^{1/2}))^{-1}\sum_{t=1}^{T-1} u_{t+1}F(\sum_{l=1}^{t} v_l). \tag{9}$$

First, we need the following lemma; this is a partial result that that can be specialized towards the main theorems of this paper. For a sequence k_T that is to be specified later but satisfies $k_T \to \infty$ and $k_T/T \to 0$ as $T \to \infty$, define $r_j = j/k_T$ and $T_j = \max(1, [Tj/k_T])$ for $j = 1, \ldots, k_T$.

Lemma 1 *Under Assumptions 1 and 2,*

$$G_T = \int_0^1 H(V(r))dU(r) + J_T + o_P(1), \tag{10}$$

where

$$J_T = \kappa(T^{1/2})^{-1}\sum_{j=1}^{k_T}\sum_{t=T_{j-1}}^{n_j-1}(U_T((t+1)/T) - U_T(t/T))(F(T^{1/2}V_T(t/T)) - F(T^{1/2}V_T(r_{j-1}))). \tag{11}$$

By finding the limit distribution of J_T, the following result can be proven:

Theorem 1 *Under Assumptions 1 and 2,*

$$G_T \xrightarrow{d} \int_0^1 H(V(r))dU(r) + \Lambda \int_0^1 H'(V(r))dr \tag{12}$$

and

$$G_T - \Lambda T^{-1}\sum_{t=1}^{T} H'(V_T(t/T)) \xrightarrow{d} \int_0^1 H(V(r))dU(r), \tag{13}$$

where

$$\Lambda = \lim_{T\to\infty}\sum_{l=1}^{T} Eu_{T+1}v_l. \tag{14}$$

The proof of Theorem 1 is based on finding a limit result for

$$T^{-1/2}\sum_{t=1}^{T-1} u_{t+1}H(T^{-1/2}\sum_{l=1}^{t} v_l), \tag{15}$$

and therefore, in the proof of Theorem 1, the following result can be found too:

Theorem 2 *Under Assumptions 1, 2, and 3,*

$$T^{-1/2}\sum_{t=1}^{T-1} u_{t+1}H(T^{-1/2}\sum_{l=1}^{t} v_l) \xrightarrow{d} \int_0^1 H(V(r))dU(r) + \Lambda \int_0^1 H'(V(r))dr \tag{16}$$

and

$$T^{-1/2}\sum_{t=1}^{T-1} u_{t+1}H(T^{-1/2}\sum_{l=1}^{t} v_l) - T^{-1}\sum_{t=1}^{T} H'(V_T(t/T)) \xrightarrow{d} \int_0^1 H(V(r))dU(r), \tag{17}$$

where

$$\Lambda = \lim_{T\to\infty} T^{-1}\sum_{t=1}^{T}\sum_{l=1}^{t} Eu_{t+1}v_l = \lim_{T\to\infty}\sum_{l=1}^{T} Eu_{T+1}v_l. \tag{18}$$

A striking feature of the above theorems is that only Λ comes up in the limit distribution as a measure of the correlation between the u_t and the v_t. Intuitively, one might have expected that introducing nonlinearities in $F(.)$ might cause the limit distribution of the G_T statistic to depend on measure of nonlinear correlation as well, but this turns out to be not the case. Park and Phillips (1999,2001) derive results similar to those of Theorems 1 and 2 for the case $\Lambda = 0$, but they use martingale difference and homoscedasticity assumptions. It is immediate from the above result that no homoscedasticity assumptions are needed to derive the limit distribution as long as $\Lambda = 0$, and therefore their homoscedasticity assumptions can be removed using the above theorem.

3 Powers of integrated processes as regressors

In this section, I consider the linear model

$$y_t = \theta_0 x_t + u_t \tag{19}$$

where $x_t = |\sum_{j=1}^{t} v_t|^\gamma$ for some $\gamma > 1$ and (u_t, v_t) is a weakly dependent but possibly correlated process. Let $\hat{\theta}_T$ denote the least squares estimator for θ_0 (assuming no constant has been included in the regression). For this situation, the following result can be proven:

Theorem 3 *Assume that (u_t, v_t) satisfies Assumption 1, and consider $\hat{\theta}_T$ and y_t and x_t as above. Then*

$$T^{1/2+\gamma/2}(\hat{\theta}_T - \theta_0) = \frac{T^{-1/2}\sum_{t=1}^{T} u_t |T^{-1/2}\sum_{j=1}^{t} v_t|^{\gamma}}{T^{-1}\sum_{t=1}^{T} |T^{-1/2}\sum_{j=1}^{t} v_t|^{2\gamma}}$$

$$\xrightarrow{d} \frac{\int_0^1 |V(r)|^{\gamma} dU(r) + \Lambda\gamma \int_0^1 \operatorname{sgn}(V(r))|V(r)|^{\gamma-1} dr}{\int_0^1 |V(r)|^{2\gamma} dr} \tag{20}$$

where

$$\Lambda = \lim_{T\to\infty} T^{-1} \sum_{t=1}^{T} \sum_{l=1}^{t} E u_t v_l. \tag{21}$$

¿From the above theorem, it can be concluded that it is incorrect in general to use regressors that are powers of integrated processes as regressors and still assume standard inference will remain valid. Of course, this fact is well-known for the case where the regressor is an integrated process.
The conjecture presents itself that the above result will remain valid for all $\gamma > 0$. The problem with the proof of that result is the requirement in Theorem 2 that $H'(.)$ be continuous, which in the setting of the above theorem translates into the requirement that $\gamma|x|^{\gamma-1}$ be continuous, and this will only hold for $\gamma \geq 1$. The author expects however that it will turn out to be possible to prove such a result.

4 Nonlinear regression with integrated regressors

In linear cointegrating relations, the result of Theorem 2 for $H(x) = x$ can be used, and the standard results for deriving the mathematics of linear cointegrating regressions can be obtained. Then, the observation can be made that if somehow the correlation between the error and the regressor can be made to disappear asymptotically, the limit distribution will be mixed normal, implying that standard least squares is valid. This is the basis of techniques such as the "fully modified" technique by Phillips and Hansen (1990), Park's "canonical regressions", and by the so-called leads-and-lags technique. The goal of this section is to obtain a similar technique in the situation of a nonlinear cointegrating regression

$$y_t = f(x_t, \theta_0) + u_t. \tag{22}$$

Examples of such nonlinear cointegrating regressions can be found in Park and Phillips (2001); one simple example is the case where

$$y_t = \theta_{01} + \theta_{02} a(x_t) + u_t. \tag{23}$$

For example, setting $a(x) = x^2$, using the analysis below, it is straightforward to analyze a linear cointegrating regression when the regressor itself is not I(1), but instead the regressor equals the square of an I(1) process.
In order to analyze nonlinear cointegrating relations, it will be necessary to borrow several results and definitions from Park and Phillips (2001). The treatment below uses the machinery from Park and Phillips, except that the limit theorem proven in this paper is now applied to Park and Phillips' setting. The first definition I need is that of H-regularity:

Definition 1 *Let*

$$F(\lambda x, \pi) = \kappa(\lambda, \pi) H(x, \pi) + R(x, \lambda, \pi) \tag{24}$$

where κ is nonsingular. We say that F is H-regular on Π if:

(a.) *H is regular on Π,*

(b.) *$R(x, \lambda, \pi)$ is of order smaller than $\kappa(\lambda, \pi)$ for all $\pi \in \Pi$.*

We call κ the asymptotic order *and H the* limit homogeneous function *of F. If κ does not depend upon π, then F is said to be H_0-regular.*

For the definition of "regular" and "of order smaller than" the reader is referred to Park and Phillips (2001).
Define

$$\dot{f} = (\partial f(x, \theta)/\partial \theta_i), \qquad \ddot{f} = (\partial^2 f(x, \theta)/\partial \theta_i \partial \theta_j) \tag{25}$$

to be all vectors, arranged by the lexicographic ordering of their indices, analogously to Park and Phillips (2001). Let $\dot{h}$ denote the limit homogeneous function of H-regular $\dot{f}$. The asymptotic orders of $\dot{f}$ and $\ddot{f}$ are denoted by $\dot{\kappa}$ and $\ddot{\kappa}$ respectively. Using the above definitions and assumptions, the following theorem easily follows:

Theorem 4 *Let Assumptions 1 and 2 hold, and let F be specified as in Definition 1. If F is H-regular on a compact set Π, then as $T \to \infty$*

$$T^{-1} \kappa(T^{1/2}, \pi)^{-1} \sum_{t=1}^{T} F(x_t, \pi) \xrightarrow{as} \int_0^1 H(V(r)) dr \tag{26}$$

uniformly in $\pi \in \Pi$. *Moreover, if* $F(.,.)$ *is* H*-regular, then*

$$T^{-1/2}\kappa(T^{1/2},\pi)^{-1}\sum_{t=1}^{T}F(x_t,\pi)u_t \xrightarrow{d} \int_0^1 H(V(r),\pi)dU(r) + \Lambda\int_0^1 H'(V(r),\pi)dr \quad (27)$$

as $T \to \infty$, *where* $H'(v,r)$ *denotes* $(\partial/\partial v)H(v,r)$.

Using the above theorem as a tool and invoking the results of Park and Phillips (2001), the following result for the nonlinear least squares estimator can be shown:

Theorem 5 *Let Assumption 1 hold. Assume*

a. f *is* H_0*-regular on* Θ *with asymptotic order* κ *and limit homogeneous function* h. *In addition, assume that* $\kappa(\lambda)$ *is bounded away from zero as* $\lambda \to \infty$ *and that for all* $\theta \neq \theta_0$ *and* $\delta > 0$, $\int_{|s|\leq\delta}(h(s,\theta) - h(s,\theta_0))^2 ds > 0$;

b. $\dot{f}$ *and* $\ddot{f}$ *are* H*-regular on* Θ;

c. $\| (\dot{\kappa}\otimes\dot{\kappa})^{-1}\kappa\ddot{\kappa} \| < \infty$, *and*

d. $\int_{|s|\leq\delta}\dot{h}(s,\theta_0)\dot{h}(s,\theta_0)'ds > 0$ *for all* $\delta > 0$.

Then

$$T^{1/2}\dot{\kappa}(T^{1/2})'(\hat{\theta}_T - \theta_0)$$

$$\xrightarrow{d} \left(\int_0^1 \dot{h}(V(r),\theta_0)\dot{h}(V(r),\theta_0)'dr\right)^{-1}\left(\int_0^1 \dot{h}(V(r),\theta_0)dU(r) + \Lambda\int_0^1 (\partial/\partial V)\dot{h}(V(r),\theta_0)dr\right) \quad (28)$$

as $T \to \infty$.

The practical implication of the above theorem is that, like in the linear case, a nonlinear cointegrating regression will not yield valid inference results when the nonlinear regression is treated in the "naive" way (i.e., assuming that the integration property of the regressor is of no consequence). We cannot assume that standard errors and F-tests will remain valid in the absence of exogeneity assumptions. This is because Λ appears in the expression for the asymptotic variance and because the $U(.)$ and $V(.)$ processes are not independent Gaussian processes. Therefore, it is desirable to find a "fully modified" type technique for nonlinear cointegrating regressions, as is developed in the next section.

5 Fully modified estimation for nonlinear cointegrating regressions

The "fully modified" procedure for linear cointegrating regressions consists of two elements. One element in the proof of the asymptotic validity of the "fully modified" procedure is to essentially make the limiting Brownian motions of the error process and the regressor process independent, while the second element is to make the expression involving Λ in Equation (28) disappear. To achieve the first goal, let

$$\Sigma^* = \sum_{k=-\infty}^{\infty} \begin{pmatrix} Eu_t u_{t-k} & Eu_t v_{t-k} \\ Eu_{t-k} v_t & Ev_t v_{t-k} \end{pmatrix} = \begin{pmatrix} \Sigma^*_{11} & \Sigma^*_{12} \\ \Sigma^*_{21} & \Sigma^*_{22} \end{pmatrix}. \tag{29}$$

If we now define

$$y_t^\dagger = y_t - \Sigma^*_{21}(\Sigma^*_{22})^{-1} v_t = f(x_t, \theta_0) + (u_t - \Sigma^*_{21}(\Sigma^*_{22})^{-1} v_t), \tag{30}$$

and treat $y_t^\dagger$ as our "new" y variable (note that $y_t^\dagger$ is observable), it follows that the "new error" in the above equation has become

$$\alpha_t = u_t - \Sigma^*_{21}(\Sigma^*_{22})^{-1} v_t. \tag{31}$$

Letting $A_T(r) = T^{-1/2} \sum_{t=1}^{[rT]} \alpha_t$, then for this α_t we have

$$EA_T(r)U_T(r) = ET^{-1} \sum_{t=1}^{[rT]} \sum_{s=1}^{[rT]} \alpha_t v_s$$

$$= ET^{-1} \sum_{t=1}^{[rT]} \sum_{s=1}^{[rT]} (u_t - \Sigma^*_{21}(\Sigma^*_{22})^{-1} v_t) v_s \longrightarrow r\Sigma^*_{21} - r\Sigma^*_{21}(\Sigma^*_{22})^{-1}\Sigma^*_{22} = 0. \tag{32}$$

Therefore, the limit Brownian motions generated by partial sums of the α_t and the v_t are asymptotically independent. Analogously to the linear case however, it is not sufficient to simply use $y_t^\dagger$ instead of y_t in the nonlinear least squares minimization objective function. This is because by Theorem 4,

$$T^{-1/2}\kappa(T^{1/2}, \pi)^{-1} \sum_{t=1}^{T} F(x_t, \pi)\alpha_t \xrightarrow{d} \int_0^1 H(V(r), \pi) dA(r) + \Lambda \int_0^1 H'(V(r), \pi) dr \tag{33}$$

where

$$\Lambda = \lim_{T\to\infty} T^{-1} \sum_{t=1}^{T} \sum_{l=1}^{t} E\alpha_t v_l = \sum_{k=0}^{\infty} E\alpha_t v_{t-k}, \tag{34}$$

and note that while by construction,

$$\sum_{k=-\infty}^{\infty} E\alpha_t v_{t-k} = 0, \tag{35}$$

it is not necessary that $\Lambda = \sum_{k=0}^{\infty} E\alpha_t v_{t-k} = 0$. Therefore, I propose the nonlinear fully modified least squares estimator $\tilde{\theta}_T$ that results from solving

$$0 = T^{-1/2}\dot{\kappa}(T^{1/2})' \sum_{t=1}^{T}(y_t^{\dagger} - f(x_t,\theta))(\partial/\partial\theta)f(x_t,\theta) + \hat{\Lambda}T^{-1}\sum_{t=1}^{T}(\partial/\partial\theta)f'(x_t,\hat{\theta}_T), \tag{36}$$

where $\hat{\theta}_T$ is the standard nonlinear least squares estimator for θ and $\hat{\Lambda}$ is some heteroscedasticity and autocorrelation consistent covariance matrix estimate for Λ. Note that for the standard linear cointegrating regression with a scalar x variable we have $f(x,\theta) = \theta_1 + \theta_2 x$, implying that the above problem simplifies to solving

$$0 = T^{-1}\sum_{t=1}^{T}[(y_t^{\dagger} - \theta_1 - \theta_2 x_t)(1, x_t) + (0, \hat{\Lambda}), \tag{37}$$

which returns the "fully modified" estimator as proposed in Phillips and Hansen (1990). For the "nonlinear fully modified" estimator $\tilde{\theta}_T$, the following result can be obtained:

Theorem 6 *Under the assumptions of Theorem 5,*

$$T^{1/2}\dot{\kappa}(T^{1/2})'(\tilde{\theta}_T - \theta_0)$$

$$\xrightarrow{d} (\int_0^1 \dot{h}(V(r),\theta_0)\dot{h}(V(r),\theta_0)'dr)^{-1}(\int_0^1 \dot{h}(V(r),\theta_0)dA(r)) \tag{38}$$

as $T \to \infty$.

The above theorem implies that the nonlinear fully modified estimator $\tilde{\theta}_T$ produces asymptotically valid inference results. This is because $A(.)$ and $V(.)$ are independent Gaussian processes, and therefore $\tilde{\theta}_T$ and its t-values and F-tests are asymptotically normally distributed conditional on $V(.)$.

While the above result shows that the fully modified technique can be extended towards nonlinear cointegrating regressions, it remains an open question whether or not the same can be done with the "leads and lags" technique and/or Park's canonical regressions.

Mathematical appendix

Proof of Lemma 1:

Recall that by assumption,

$$\sup_{r\in[0,1]} (|V_T(r) - V(r)| + |U_T(r) - U(r)|) \leq \delta_T = o(1)$$

almost surely, where δ_T is a deterministic sequence. It is assumed that the sequence k_T satisfies $k_T \to \infty$, $k_T/T \to 0$, and

$$k_T \kappa(T^{1/2})^{-2} \sup_{|x|\leq K} \sup_{x':|x-x'|\leq \delta_T} |F(T^{1/2}x) - F(T^{1/2}x')|^2 = o(1).$$

In addition, define

$$I_1 = \int_0^1 H(V(r))dU(r),$$

$$I_2 = \kappa(T^{1/2})^{-1} \int_0^1 F(T^{1/2}V(r))dU(r),$$

$$P_T = \kappa(T^{1/2})^{-1} \sum_{j=1}^{k_T} \int_{r_{j-1}}^{r_j} F(T^{1/2}V(r_{j-1}))dU(r) = \sum_{j=1}^{k_T} F(T^{1/2}V(r_{j-1}))(U(r_j) - U(r_{j-1})),$$

and

$$G_T^* = \kappa(T^{1/2})^{-1} \sum_{j=1}^{k_T} F(T^{1/2}V_T(r_{j-1}))(U_T(r_j) - U_T(r_{j-1})).$$

Write

$$G_T = (G_T - G_T^*) + (G_T^* - P_T) + (P_T - I_2) + (I_2 - I_1) + I_1.$$

I first argue that

$$|I_2 - I_1| \xrightarrow{p} 0 \tag{39}$$

and then that

$$|P_T - I_2| \xrightarrow{p} 0. \tag{40}$$

Then, I show that

$$|G_T^* - P_T| \xrightarrow{p} 0 \tag{41}$$

and then the lemma is completed by observing that

$$U_T(r_j) - U_T(r_{j-1}) = \sum_{t=T_{j-1}}^{T_j-1} (U_T((t+1)/T) - U_T(t/T))$$

and therefore

$$G_T = \kappa(T^{1/2})^{-1} \sum_{j=1}^{k_T} \sum_{t=T_{j-1}}^{T_j-1} F(T^{1/2}V_T(t/T))(U_T((t+1)/T) - U_T(t/T))$$

implying that

$$G_T - G_T^*$$

$$= \kappa(T^{1/2})^{-1} \sum_{j=1}^{k_T} \sum_{t=T_{j-1}}^{T_j-1} (U_T((t+1)/T) - U_T(t/T))(F(T^{1/2}V_T(t/T)) - F(T^{1/2}V_T(r_{j-1}))) = J_T.$$

To show the result of Equation (39), note that

$$E(I_1 - I_2)^2 = E(\int_0^1 (F(T^{1/2}V(r))\kappa(T^{1/2})^{-1} - H(V(r)))dU(r))^2$$

$$= E(\kappa(T^{1/2})^{-1} \int_0^1 R(V(r), T^{1/2})dU(r))^2$$

$$= \kappa(T^{1/2})^{-2} \int_0^1 ER(V(r), T^{1/2})^2 dr$$

$$\leq \kappa(T^{1/2})^{-2} \int_0^1 a(T^{1/2})^2 P(V(r))^2 dr \xrightarrow{p} 0$$

by Assumption for class H_1 (NOTE: how about H_2?). To show the result of Equation (40), first note that, for all $\delta > 0$,

$$\limsup_{T\to\infty} P(|P_T - I_2| > \delta)$$

$$\leq \limsup_{K\to\infty}(\limsup_{T\to\infty} P(|P_T - P_{TK}| > \delta/3) + \limsup_{T\to\infty} P(|P_{TK} - I_{T\varepsilon K}| > \delta/3)$$

$$+ \limsup_{T\to\infty} P(|I_{TK} - I_2| > \delta/3))$$

where

$$I_{TK} = \kappa(T^{1/2})^{-2}\sum_{j=1}^{k_T}\int_{r_{j-1}}^{r_j} F(T^{1/2}V(r))$$

$$\times I(\sup_{r\in[r_{j-1},r_j]}|V(r) - V(r_{j-1})| \leq Kk_T^{-1/2})I(\sup_{r\in[r_{j-1},r_j]}|V(r)| < K)dU(r)$$

and

$$P_{TK} = \kappa(T^{1/2})^{-2}\sum_{j=1}^{k_T}\int_{r_{j-1}}^{r_j} F(T^{1/2}V(r_{j-1}))$$

$$\times I(\sup_{r\in[r_{j-1},r_j]}|V(r) - V(r_{j-1})| < Kk_T^{-1/2})I(\sup_{r\in[r_{j-1},r_j]}|V(r)| < K)dU(r).$$

This is because

$$\limsup_{K\to\infty}\limsup_{T\to\infty} P(|P_T - P_{T\varepsilon K}| > 0)$$

$$\leq \limsup_{K\to\infty}\limsup_{T\to\infty} P(\sup_{r,r':|r-r'|\leq \max_j |r_j - r_{j-1}|}|V(r) - V(r')| \geq Kk_T^{-1/2})$$

$$+ \limsup_{K\to\infty} P(\sup_{r\in[0,1]}|V(r)| \geq K) = 0$$

where the last result follows because $V(k_T^{-1}r)$ is distributed identically to $k_T^{-1/2}V(r)$, and a similar argument holds for $|I_{TK} - I_2|$. Next, note that

$$|P_{T\varepsilon K} - I_{T\varepsilon K}|$$

$$= \kappa(T^{1/2})^{-1}\sum_{j=1}^{k_T}\int_{r_{j-1}}^{r_j}(F(T^{1/2}V(r_{j-1})) - F(T^{1/2}V(r)))I(\sup_{r\in[r_{j-1},r_j]}|V(r) - V(r_{j-1})| < Kk_T^{-1/2})$$

$$\times I(\sup_{r\in[r_{j-1},r_j]}|V(r)| < K)dU(r),$$

and therefore

$$E|I_{T\varepsilon K} - P_{T\varepsilon K}|^2$$

$$= \kappa(T^{1/2})^{-2} \sum_{j=1}^{k_T} \int_{r_{j-1}}^{r_j} E(F(T^{1/2}V(r)) - F(T^{1/2}V(r_{j-1})))^2$$

$$\times I(\sup_{r\in[r_{j-1},r_j]} |V(r) - V(r_{j-1})| < Kk_T^{-1/2}) I(\sup_{r\in[r_{j-1},r_j]} |V(r)| < K)dr$$

$$\leq \kappa(T^{1/2})^{-2} \sup_{|x|\leq K} \sup_{x':|x-x'|<Kk_T^{-1/2}} |F(T^{1/2}x) - F(T^{1/2}x')|^2 \to 0$$

as $T \to \infty$ by Assumption 2.4.
In order to show the result of Equation (41) and complete the proof of this lemma, set $a_j = \kappa(T^{1/2})^{-1}T(T^{1/2}V_T(r_j))$, $\alpha_j = \kappa(T^{1/2})^{-1}F(T^{1/2}V(r_j))$, $b_j = U_T(r_j)$, and $\beta_j = U(r_j)$, and note that by summation by parts, as in Davidson (1994),

$$G_T^* - P_n = \sum_{j=1}^{k_T}(a_{j-1} - \alpha_{j-1})(b_j - b_{j-1}) + \alpha_k(b_k - \beta_k) - \sum_{j=1}^{k_T}(b_j - \beta_j)(\alpha_j - \alpha_{j-1}).$$

It will be shown that all three terms converge to 0 in probability. For the first term, we have

$$[\sum_{j=1}^{k_T}(a_{j-1} - \alpha_{j-1})(b_j - b_{j-1})]^2$$

$$= [\kappa(T^{1/2})^{-1} \sum_{j=1}^{k_T}(F(V_T(r_{j-1})) - F(V(r_{j-1})))(U_T(r_j)) - U_T(r_{j-1}))]^2$$

$$\leq \kappa(T^{1/2})^{-2} \sum_{j=1}^{k_T}(F(V_T(r_{j-1})) - F(V(r_{j-1})))^2 \sum_{j=1}^{k_T}(U_T(r_j)) - U_T(r_{j-1}))^2.$$

Next, note that

$$E\sum_{j=1}^{k_T}(U_T(r_j)) - U_T(r_{j-1}))^2 = O(1)$$

because Assumption 1 implies that $E(U_T(r_j)) - U_T(r_{j-1}))^2 \leq C(r_j - r_{j-1})$, and therefore it suffices to show that

$$\kappa(T^{1/2})^{-2} \sum_{j=1}^{k_T} (F(T^{1/2} V_T(r_{j-1})) - F(T^{1/2} V(r_{j-1})))^2 = o_P(1).$$

The latter result holds because $\sup_{r \in [0,1]} |V_T(r)| = O_P(1)$ and therefore with arbitrarily large probability,

$$\kappa(T^{1/2})^{-2} \sum_{j=1}^{k_T} (F(T^{1/2} V_T(r_{j-1})) - F(T^{1/2} V(r_{j-1})))^2$$
$$\leq k_T \kappa(T^{1/2})^{-2} \sup_{|x| \leq K} \sup_{x' : |x - x'| \leq \delta_T} |F(T^{1/2} x) - F(T^{1/2} x')|^2 = o(1)$$

by assumption. A similar argument holds for

$$\sum_{j=1}^{k_T} (b_j - \beta_j)(\alpha_j - \alpha_{j-1}).$$

Finally, note that

$$\alpha_k (b_k - \beta_k) = \kappa(T^{1/2})^{-1} F(T^{1/2} V(1))(U_T(1) - U(1)) = O_P(\delta_T) = o_P(1),$$

which completes the proof of this lemma. □

Next, I state two lemmas that will be needed to complete the proof of Theorem 1.

Lemma 2 *Assume that $x_t \in \mathcal{F}_t$, and assume that for $l \geq m_T$ and $1 \leq t \leq n$, x_t satisfies*

$$\| E(x_{Tt} | \mathcal{F}_{t-l}) \|_2 \leq a_T \gamma_T(l)$$

and

$$(\theta_{2T} - \theta_{1T}) a_T^2 \sum_{l=0}^{\infty} \gamma_T(l)^2 (\log(l+1))^2 \to 0. \tag{42}$$

Then if

$$m_T^2 \sum_{t=1}^{T} E x_{Tt}^2 \to 0,$$

we have

$$\sum_{t=\theta_{1T}}^{\theta_{2T}} x_{Tt} \xrightarrow{p} 0.$$

Proof of Lemma 2:

Note that for all positive sequences m_T such that $m_T \to \infty$ as $n \to \infty$,

$$\sum_{t=1}^{T} x_{Tt} = \sum_{t=1}^{T}(x_{Tt} - E(x_{Tt}|\mathcal{F}_{t-m_T})) + \sum_{t=1}^{T} E(x_{Tt}|\mathcal{F}_{t-m_T}),$$

and by Lemma 2.1 of Hall and Heyde (1980),

$$\| \sum_{t=1}^{T} E(x_{Tt}|\mathcal{F}_{t-m_T}) \|_2^2 \leq C(\theta_{2n} - \theta_{1n})a_T^2 \sum_{l=m_T}^{\infty} \gamma_T(l)^2(\log(l+1))^2 \to 0$$

by assumption. Also,

$$\| \sum_{t=1}^{T}(x_{Tt} - E(x_{Tt}|\mathcal{F}_{t-m_T})) \|_2 = \| \sum_{j=0}^{m_T-1} \sum_{t=1}^{T}(E(x_{Tt}|\mathcal{F}_{t-j}) - E(x_{Tt}|\mathcal{F}_{t-j-1})) \|_2$$

$$\leq \sum_{j=0}^{m_T-1} \| \sum_{t=1}^{T}(E(x_{Tt}|\mathcal{F}_{t-j}) - E(x_{Tt}|\mathcal{F}_{t-j-1})) \|_2$$

$$\leq 2m_T(\sum_{t=1}^{T} Ex_{Tt}^2)^{1/2} \to 0$$

by assumption. □

Lemma 3 *If $z_t = (u_t, w_t)$ is L_2-NED on an alpha-mixing sequence v_t such that $E(z_t|v_t, v_{t-1}, \ldots) = z_t$, $|z_t| \leq B_T$, and with NED sequence $\nu(m)$, then for $m \geq 0$*

$$\| E(u_t w_t|\mathcal{F}_{t-2m}) - Eu_t w_t \|_2 \leq 4B_T\nu(m) + 6B_T \| u_t \|_p \alpha(m)^{1-1/p}.$$

Proof:

Define $\mathcal{V}_{t-m}^t = \sigma(\{v_t, \ldots, v_{t-m}\})$, and note that

$$E(u_t w_t | \mathcal{F}_{t-2m}) = E((u_t - E(u_t | V_{t-m}^t)) w_t | \mathcal{F}_{t-2m})$$

$$+ E(E(u_t | \mathcal{V}_{t-m}^t))(w_t - E(w_t | V_{t-m}^t) | \mathcal{F}_{t-2m}) + E(E(u_t | \mathcal{V}_{t-m}^t) E(w_t | V_{t-m}^t) | \mathcal{F}_{t-2m}),$$

and therefore

$$\| E(u_t w_t | \mathcal{F}_{t-2m}) - E u_t w_t \|_2$$

$$\leq 2 \| (u_t - E(u_t | \mathcal{V}_{t-m}^t)) w_t \|_2 + 2 \| (w_t - E(w_t | \mathcal{V}_{t-m}^t)) E(u_t | \mathcal{V}_{t-m}^t) \|_2$$

$$+ \| E(E(w_t | \mathcal{V}_{t-m}^t) E(u_t | \mathcal{V}_{t-m}^t) | \mathcal{F}_{t-2m}) - E(E(w_t | \mathcal{V}_{t-m}^t) E(u_t | \mathcal{V}_{t-m}^t)) \|_2$$

$$\leq 2 B_T \nu(m) + 2 B_T \nu(m) + 6 B_T \alpha(m)^{1/2 - 1/p},$$

where the last inequality follows by noting that $E(w_t | \mathcal{V}_{t-m}^t) E(u_t | \mathcal{V}_{t-m}^t)$ is alpha-mixing and by Theorem 17.5 of Davidson (1994). For the uniform mixing case, a similar result holds, but for this case we need the inequality for uniform mixing processes from Theorem 17.5 of Davidson (1994). □

Proof of Theorem 1:

(NOTE: the theorem below is incomplete and only constitutes a proof for Theorem 2. However it should not be too much trouble to extend the argument to provide a proof of Theorem 1 !)

Given the result of Lemma 1, it suffices to determine the limit distribution of J_T. Given a k_T sequence that satisfies $k_T = o(T^{1/2})$ and the earlier conditions, we can always find a positive integer-valued sequence m_T such that $m_T \to \infty$, yet increases slowly enough such that $k_T m_T = o(T^{1/2})$ and $m_T \leq \min_{1 \leq j \leq k_T}(T_j - T_{j-1} - 1) = T/k_T - 1$. Because of that last requirement, we can write

$$J_T = \sum_{j=1}^{k_T} \sum_{t=T_{j-1}}^{T_j - 1} (U_T((t+1)/T) - U_T(t/T))(H(V_T(t/T)) - H(V_T(r_{j-1})))$$

$$= \sum_{j=1}^{k_T} \sum_{t=T_{j-1}}^{T_{j-1}+m_T-1} (U_T((t+1)/T) - U_T(t/T))(H(V_T(t/T)) - H(V_T(r_{j-1})))$$

$$+ \sum_{j=1}^{k_T} \sum_{t=T_{j-1}+m_T}^{T_j-1} (U_T((t+1)/T) - U_T(t/T))(H(V_T((t-m_T)/T)) - H(V_T(r_{j-1})))$$

$$+ \sum_{j=1}^{k_T} \sum_{t=T_{j-1}+m_T}^{T_j-1} (U_T((t+1)/T) - U_T(t/T))(H(V_T(t/T)) - H(V_T((t-m_T)/T)))$$

$$= S_1 + S_2 + S_3,$$

say. In order to find the limit distribution of $S_1 + S_2 + S_3$, we need the following lemma's:

Lemma 4 *Under Assumptions 2.1 and 1,*

$$S_1 \xrightarrow{p} 0.$$

Proof of Lemma 4:

Note that we can choose a K_ε such that

$$\limsup_{T\to\infty} P(\sup_{r\in[0,1]} |V_T(r)| \le K_\varepsilon) > 1 - \varepsilon.$$

This implies that with arbitrarily large probability,

$$S_1 = \sum_{j=1}^{k_T} \sum_{t=T_{j-1}}^{T_{j-1}+m_T-1} (U_T((t+1)/T) - U_T(t/T))(H(V_T(t/T)) - H(V_T(r_{j-1})))$$

$$\le T^{-1/2} \sum_{j=1}^{k_T} \sum_{t=T_{j-1}}^{T_{j-1}+m_T-1} |u_{t+1}| 2 \sup_{|x|\le K_\varepsilon} |T(x)|$$

$$= O_P(m_T T^{-1/2} k_T) = o_P(1)$$

if $k_T m_T = o(T^{1/2})$ and because of continuity of $T(.)$. □

For S_2, we have the following result:

Lemma 5 *Under Assumptions 2.1 and 1,*

$$S_2 \xrightarrow{p} 0.$$

Proof of Lemma 5:

Assume that, in addition to the earlier requirements on m_T, m_T diverges slowly enough to give, for all $K > 0$,

$$m_T^2 \sup_{|x|\leq K} \sup_{x':|x-x'|\leq 2\delta_T+74k_T^{-1/2}(\log(k_T^{-1}))^{1/2}} |T(x) - T(x')|^2 \to 0.$$

Note that for all $K > 0$, S_2 is asymptotically equivalent to

$$\sum_{j=1}^{k_T} \sum_{t=T_{j-1}+m_T}^{T_j-1} (U_T((t+1)/T) - U_T(t/T))(H(V_T((t-m_T)/T)) - H(V_T(r_{j-1})))I_{1tm_Tj}I_{2tm_Tj}$$

where

$$I_{1tm_Tj} = I(|V_T((t-m_T)/T) - V_T(r_{j-1})| \leq 2\delta_T + 74k_T^{-1/2}(\log(k_T^{-1}))^{1/2})$$

and

$$I_{2tm_Tj} = I(|V_T((t-m_T)/T)| \leq K)I(|V_T(r_{j-1})| \leq K)$$

because

$$\sup_{t\in[T_{j-1}+m_T,T_j-1],j\in[1,k_T]} |V_T((t-m_T)/T) - V_T(r_{j-1})|$$

$$\leq \sup_{r,r':|r-r'|\leq \max_j |r_j-r_{j-1}|} |V_T(r) - V_T(r')|$$

$$\leq \sup_{r,r':|r-r'|\leq k_T^{-1}} |V(r) - V(r')| + 2\delta_T$$

$$\leq 2\delta_T + 74k_T^{-1/2}|\log(k_T^{-1})|^{1/2}$$

almost surely because $|V(r) - V(r')| \leq 74|r-r'|^{1/2}|\log|r-r'||^{1/2}$ almost surely (see e.g. Pollard (1984, p. 146). Also,

$$\limsup_{T\to\infty} P(\exists t \in [1,n] : |V_T(t/T)| > K) = \limsup_{T\to\infty} P(\sup_{r\in[0,1]} |V_T(r)| > K) \to 0$$

as $K \to \infty$. Next, define

$$\mathcal{F}_t = \sigma(\{(w_t, v_t), (w_{t-1}, v_{t-1}), \ldots\}),$$

and note that the summands

$$(U_T((t+1)/T) - U_T(t/T))(H(V_T((t-m_T)/T)) - H(V_T(r_{j-1})))I_{1tm_Tj}I_{2tm_Tj}$$

satisfy

$$\| E[(W_n((t+1)/T) - W_n(t/T))(H(V_T((t-m_T/T)) - H(V_T(r_{j-1})))I_{1tm_Tj}I_{2tm_Tj}|\mathcal{F}_{t-l}] \|_2$$

$$= \| (H(V_T((t-m_T)/T)) - H(V_T(r_{j-1})))I_{1tm_Tj}I_{2tm_Tj}E[(U_T((t+1)/T) - U_T(t/T))|\mathcal{F}_{t-l}] \|_2$$

$$\leq 2T^{-1/2} \sup_{|x|\leq K} \sup_{x':|x-x'|\leq 2\delta_T + 74k_T^{-1/2}|\log(k_T^{-1})|^{1/2}} |T(x) - T(x')|\psi(l)$$

for $l \geq m_T$, where the $\psi(l)$ sequence denotes the L_2-mixingale sequence of w_t - which by Assumption 1 is of size -1/2 - and because $I_{1tm_Tj} \in \mathcal{F}_{t-m_T}$ and $I_{2tm_Tj} \in \mathcal{F}_{t-m_T}$. Now, by Lemma 2, it follows that $S_2 \xrightarrow{p} 0$ because the requirement of Equation (42) is met and in addition,

$$m_T^2 \sum_{j=1}^{k_T} \sum_{t=T_{j-1}+m_T}^{T_j-1} E(U_T((t+1)/T) - U_T(t/T))^2(H(V_T((t-m_T)/T)) - H(V_T(r_{j-1})))^2 I_{1tm_Tj}I_{2tm_Tj}$$

$$\leq m_T^2 T^{-1} \sum_{j=1}^{k_T} \sum_{t=T_{j-1}+m_T}^{T_j-1} \sup_{|x|\leq K} \sup_{x':|x-x'|\leq 2\delta_T + 74k_T^{-1/2}|\log(k_T^{-1})|^{1/2}} |T(x) - T(x')|^2 Ew_{t+1}^2$$

$$\leq m_T^2 \sup_{|x|\leq K} \sup_{x':|x-x'|\leq 2\delta_T + 74k_T^{-1/2}|\log(k_T^{-1})|^{1/2}} |T(x) - T(x')|^2 Ew_t^2 \to 0$$

as $n \to \infty$ by assumption. □

In view of the above lemmas, the result of the following lemma now suffices to complete the proof of Theorem 1:

Lemma 6 *Under Assumptions 2.3 and 1,*

$$S_3 \xrightarrow{p} \Lambda \int_0^1 T'(U(r))dr.$$

Proof of Lemma 6:

Assume that in addition to the earlier requirements, m_T also satisfies

$$m_T \sup_{|x|\le K} \sup_{x':|x-x'|\le 4\delta_T+148k_T^{-1/2}|\log(k_T^{-1})|^{1/2}} |H'(x)-H'(x')| \to 0$$

for all $K>0$. Note that by the Taylor series expansion,

$$\sum_{j=1}^{k_T} \sum_{t=T_{j-1}+m_T}^{T_j-1} (U_T((t+1)/T)-U_T(t/T))(H(V_T(t/T))-H(V_T((t-m_T)/T)))$$

$$=\sum_{j=1}^{k_T} \sum_{t=T_{j-1}+m_T}^{T_j-1} (U_T((t+1)/T)-U_T(t/T))(V_T(t/T)-V_T((t-m_T)/T))H'(\tilde{V}_{Ttm_T})$$

where $|V_T(t/T)-\tilde{V}_{Ttm_T}| \le |V_T(t/T)-V_T((t-m_T)/T)|$. The last statistic is asymptotically equivalent to

$$S_2' = \sum_{j=1}^{k_T} \sum_{t=T_{j-1}+m+1}^{T_j} (U_T((t+1)/T)-U_T(t/T))(V_T(t/T)-V_T((t-m_T)/T))H'(V_T(r_{j-1})).$$

To see this, first note that

$$|V_T(r_{j-1})-\tilde{V}_{Ttm_T}| \le |V_T(r_{j-1})-V_T(t/T)| + |V_T(t/T)-V_T((t-m_T)/T)|$$

$$\le 2 \sup_{r,r':|r-r'|\le \max_j |r_j-r_{j-1}|} |V_T(r)-V_T(r')| \le 4\delta_T + 148k_T^{-1/2}|\log(k_T^{-1})|^{1/2}$$

almost surely by the same argument as in the proof of Lemma 5. With the above result, again noting that

$$\sup_{r\in[0,1]} |V_T(r)| \le K_\varepsilon$$

with arbitrarily large probability, it follows that

$$|S_2 - S_2'|$$

$$\le \sum_{j=1}^{k_T} \sum_{t=T_{j-1}+m_T}^{T_j-1} |U_T((t+1)/T)-U_T(t/T)||V_T(t/T)-V_T((t-m_T)/T)||H'(\tilde{V}_{Ttm_T})-H'(V_T(r_{j-1}))|$$

$$\leq 2 \sup_{x,x':|x-x'|\leq 4\delta_T+148k_T^{-1/2}|\log(k_T^{-1})|^{1/2}} |T'(x)-T'(x')|\times$$

$$\sum_{j=1}^{k_T}\sum_{t=T_{j-1}+m_T}^{T_j-1}|U_T((t+1)/T)-U_T(t/T)||V_T(t/T)-V_T((t-m_T)/T)|, \tag{43}$$

and the second term in Equation (43) is $O_P(m_T)$ because

$$E\sum_{j=1}^{k_T}\sum_{t=T_{j-1}+m_T}^{T_j-1}|U_T((t+1)/T)-U_T(t/T)||V_T(t/T)-V_T((t-m_T)/T)|$$

$$=ET^{-1}\sum_{j=1}^{k_T}\sum_{t=T_{j-1}+m_T}^{T_j-1}|u_{t+1}|\sum_{k=0}^{m_T}|v_{t-k}|\leq m_T \parallel u_t \parallel_2 \parallel v_t \parallel_2$$

and therefore, the expression from Equation (43) converges to zero because by assumption,

$$m_T \sup_{x,x':|x-x'|\leq 4\delta_T+148k_T^{-1/2}|\log(k_T^{-1})|^{1/2}} |T'(x)-T'(x')|\to 0.$$

Next, note that

$$S_2' = \sum_{j=1}^{k_T}H'(U_T(r_{j-1}))\sum_{t=T_{j-1}+m_T}^{T_j-1}\sum_{k=0}^{m_T-1}(U_T((t+1)/T)-U_T(t/T))(V_T((t-k)/T)-V_T((t-k-1)/T))$$

$$=T^{-1}\sum_{j=1}^{k_T}H'(V_T(r_{j-1}))\sum_{t=T_{j-1}+m_T}^{T_j-1}\sum_{k=0}^{m_T-1}(u_{t+1}v_{t-k}-Eu_{t+1}v_{t-k})$$

$$+T^{-1}\sum_{j=1}^{k_T}H'(V_T(r_{j-1}))\sum_{t=T_{j-1}+m_T}^{T_j-1}\sum_{k=0}^{m_T-1}Eu_{t+1}v_{t-k}. \tag{44}$$

The first term converges to 0 here. To see this, note that for n large enough, we have

$$T^{-1}\sum_{j=1}^{k_T}H'(V_T(r_{j-1}))\sum_{t=T_{j-1}+m_T}^{T_j-1}\sum_{k=0}^{m_T-1}(u_{t+1}v_{t-k}-Eu_{t+1}v_{t-k})$$

$$= T^{-1}\sum_{j=1}^{k_T} H'(V_T(r_{j-1})) \sum_{t=T_{j-1}+m_T}^{T_j-1} \sum_{k=0}^{m_T-1} (u_{t+1}^{B_T} v_{t-k}^{B_T} - E u_{t+1}^{B_T} v_{t-k}^{B_T}) + o_P(1)$$

where

$$u_t^{B_T} = u_t I(|u_t| \le B_T) + B_T I(u_t > B_T) - B_T I(u_t < -B_T)$$

and

$$v_t^{B_T} = v_t I(|v_t| \le B_T) + B_T I(v_t > B_T) - B_T I(v_t < -B_T),$$

and $B_T = m_T^{1/(p-2)+\eta}$ for some $\eta > 0$, and m_T is assumed to satisfy, in addition to the earlier requirements,

$$m_T^{3+2/(p-2)+2\eta} k_T T^{-1} = m_T^3 B_T^2 k_T T^{-1} \to 0.$$

This is because $\sup_{r\in[0,1]} |H'(V_T(r))| = O_P(1)$ and

$$ET^{-1}\sum_{j=1}^{k_T} |\sum_{t=T_{j-1}+m_T}^{T_j-1} \sum_{k=0}^{m_T-1} u_{t+1}v_{t-k} - u_{t+1}^{B_T} v_{t-k}^{B_T}|$$

$$\le T^{-1}\sum_{j=1}^{k_T} \sum_{t=T_{j-1}+m_T}^{T_j-1} \sum_{k=0}^{m_T-1} \| u_{t+1} I(|u_{t+1}| > B_T) \|_2 \| v_{t-k} I(|v_{t-k}| > B_T) \|_2$$

$$\le T^{-1}\sum_{j=1}^{k_T} \sum_{t=T_{j-1}+m_T}^{T_j-1} \sum_{k=0}^{m_T-1} (E|u_{t+1}|^p B_T^{2-p})^{1/2} (E|v_{t-k}|^p B_T^{2-p})^{1/2}$$

$$\le m_T B_T^{2-p} (E|v_t|^p)^{1/2} (E|u_t|^p)^{1/2} = O(m_T^{\eta(2-p)}) = o(1)$$

by assumption. In order to show that the first term in Equation (44) converges to zero, it now only remains to show that

$$T^{-1}\sum_{j=1}^{k_T} H'(V_T(r_{j-1})) \sum_{t=T_{j-1}+m_T}^{T_j-1} \sum_{k=0}^{m_T-1} (u_{t+1}^{B_T} v_{t-k}^{B_T} - E u_{t+1}^{B_T} v_{t-k}^{B_T}) \xrightarrow{p} 0.$$

To show this, it suffices to show that

$$\| k_T T^{-1} \sum_{t=T_{j-1}+m_T}^{T_j-1} \sum_{k=0}^{m_T-1} (u_{t+1}^{B_T} v_{t-k}^{B_T} - E u_{t+1}^{B_T} v_{t-k}^{B_T}) \|_2 \le c_T \to 0. \tag{45}$$

Now, it is well-known that $(u_t^{B_T}, v_t^{B_T})$ is again near epoch dependent with the same $\nu(m)$ sequence as (u_t, v_t). Therefore, by Lemma 3,

$$\| k_T T^{-1} \sum_{k=0}^{m_T-1} E(u_{t+1}^{B_T} v_{t-k}^{B_T} | \mathcal{F}_{t-l}) - E u_{t+1}^{B_T} v_{t-k}^{B_T} \|_2$$

$$\leq 4 k_T T^{-1} m_T B_T \nu(l - m_T) + 6 k_T T^{-1} B_T m_T \| u_t \|_p \alpha(l - m_T)^{1/2 - 1/p}$$

for $l \geq m_T$, implying that we can set $\gamma_T(l) = \alpha(l - m_T)^{1/2-1/p} + \nu(l - m_T)$ and $a_T = C k_T m_T T^{-1} B_T$ for Lemma 2, and note in addition that

$$\| k_T T^{-1} \sum_{k=0}^{m_T-1} u_{t+1}^{B_T} v_{t-k}^{B_T} \|_2 \leq k_T T^{-1} m_T B_T \| v_t \|_2 .$$

Therefore,

$$(n/k_T) a_T^2 \sum_{l=m_T}^{\infty} \gamma_T(l)^2 (\log(l+1))^2$$

$$= (n/k_T)(C m_T k_T T^{-1} B_T)^2 \sum_{l=m_T}^{\infty} (\alpha(l - m_T)^{1/2-1/p} + \nu(l - m_T))^2 (\log(l+1))^2$$

$$\leq C^2 n k_T m_T^2 T^{-2} B_T^2 (\log(m_T))^2 \sum_{l=1}^{\infty} (\alpha(l)^{1/2-1/p} + \nu(l))^2 (\log(l+1))^2$$

$$= O(T^{-1} k_T m_T^2 (\log(m_T))^2 B_T^2) = o(1)$$

by assumption, noting that the summation is finite by Assumption 1.4. In addition,

$$m_T \sum_{t=T_{j-1}+m_T}^{T_j-1} E x_{Tt}^2 = O(m_T (T/k_T)(k_T T^{-1} m_T B_T)^2) = O(m_T^3 k_T B_T^2 T^{-1}) = o(1)$$

by assumption, implying that the conditions of Lemma 2 are satisfied (note that Lemma 2 provides a c_n sequence (as defined in Equation (45)) that does not depend on j). This leaves

$$T^{-1} \sum_{j=1}^{k_T} H'(V_T(r_j)) \sum_{t=T_{j-1}+m_T}^{T_j-1} \sum_{k=0}^{m_T-1} E u_{t+1} v_{t-k}.$$

Define $\rho(k) = Eu_t v_{t-k}$ (which is possible by stationarity), and note that $\sum_{k=0}^{\infty} |\rho(k)| < \infty$ by Assumption 1. Then the last statistic equals

$$T^{-1} \sum_{j=1}^{k_T} H'(V_T(r_j)) \sum_{t=T_{j-1}+m_T}^{T_j-1} \sum_{k=0}^{m_T-1} \rho(k+1)$$

$$= T^{-1} \sum_{j=1}^{k_T} T'(U_T(r_j))(T_j - T_{j-1} - m_T) \sum_{k=0}^{m_T-1} \rho(k+1)$$

$$= T^{-1} \sum_{j=1}^{k_T} T'(U_T(j/k_T))(T/k_T - m_T) \sum_{k=0}^{m_T-1} \rho(k+1)$$

$$= o_P(1) + k_T^{-1} \sum_{j=1}^{k_T} T'(U_T(j/k_T)) \sum_{k=0}^{\infty} \rho(k+1) \xrightarrow{p} \Lambda \int_0^1 T'(U(r))dr$$

where the last result follows from the continuous mapping theorem. □

Proof of Theorem 3:

Theorem 3 follows from a simple application of Theorem 2.

Proof of Theorem 4:

Theorem 4 combines Theorem 1 with Theorem 3.3 of Park and Phillips (2001).

Proof of Theorem 5:

To be completed

Proof of Theorem 6:

To be completed

References

Chan, N.H. and C.Z. Wei (1988), Limiting distributions of least-squares estimates of unstable autoregressive processes, Annals of Statistics 16, 367-401.

Chang, Y. (2000), Nonlinear IV unit root tests in panels with cross-sectional dependency, working paper, Rice University.

Davidson, J. (1994), *Stochastic limit theory.* Oxford: Oxford University Press.

De Jong, R.M. and J. Davidson (1998), The functional central limit theorem and weak convergence to stochastic integrals I: weakly dependent and fractionally integrated processes, Econometric Theory 16, 621-642.

Gallant, A.R. and H. White (1988), *A unified theory of estimation and inference for nonlinear dynamic models.* New York: Basil Blackwell.

Park, J.Y. (1992), Canonical cointegrating regressions, Econometrica 60, 119-143.

Phillips, P.C.B. (1986), Understanding spurious regressions in econometrics, Journal of Econometrics 33, 311-340.

Park, J.Y. and P.C.B. Phillips (1999), Asymptotics for nonlinear transformations of integrated time series, Econometric Theory 15, 269-298.

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

Park, J.Y. and P.C.B. Phillips (2000), Nonstationary binary choice, Econometrica 68, 1249-1280.

Phillips and Hansen (1990), Statistical inference in instrumental variables regression with I(1) variables, Review of Economic Studies, 99-125.

Pötscher, B.M. and I.R. Prucha (1991), Basic structure of the asymptotic theory in dynamic nonlinear econometric models, part 1: consistency and approximation concepts, Econometric reviews 10, 125-216.

Saikkonen, P. (1991), Asymptotically efficient estimation of cointegrating regressions, Econometric Theory 7, 1-21.

Phillips, P. and Loretan (1991), Estimating long-run equilibria, Review of Economic Studies 58, 407-436.

Stock, J. and Watson (1993), Testing for common trends, Journal of the American Statistical Association 83, 1097-1107.