Hitotsubashi Journal of Economics 57 (2016), pp.111-138. © Hitotsubashi University

TESTING FOR LINEARITY IN REGRESSIONS WITH I(1) PROCESSES*

Yoichi Arai

National Graduate Institute for Policy Studies (GRIPS)
Minato-ku, Tokyo 106-8677, Japan
yarai@grips.ac.jp

Received August 2015; Accepted September 2015

Abstract

We propose a generalized version of the RESET test for linearity in regressions with I(1) processes against various nonlinear alternatives and no cointegration. The proposed test statistic for linearity is given by the Wald statistic and its limiting distribution under the null hypothesis is shown to be a χ^2 distribution with a “leads and lags” estimation technique. We show that the test is consistent against a class of nonlinear alternatives and no cointegration. Finite-sample simulations show that the empirical size is close to the nominal one and the test succeeds in detecting both nonlinearity and no cointegration.

Keywords: cointegration, I(1) processes, no cointegration, nonlinear cointegration, RESET test
JEL Classification Codes: C22, C32

I. *Introduction*

The objective of this paper is to study the relationships between economic variables in the context of regression models where explanatory variables are integrated of order one, I(1). It is well known that dynamic relationships, such as cost and production functions, are nonlinear. Many researchers have also found empirical evidence of nonlinearity in economic relationships (for example, see Granger and Teräsvirta 1993, Granger 1995 and the references contained therein). However, most of the econometric techniques for testing linearity and nonlinearity are developed for stationary variables and are not applicable for nonstationary variables, especially I(1) variables.

In studying linear relationships between I(1) economic variables, Granger (1983) and Engle and Granger (1987) introduced the concept of cointegration. Cointegration has been an intensive subject of research ever since. However, most results on cointegration provided so far have been restricted to cointegration in a linear sense. That is, most attentions has been paid

* I am especially grateful to Graham Elliott who motivated this paper. I also thank Clive Granger, Takeo Hoshi, Yixiao Sun, Halbert White and an anonymous referee for many helpful comments. The paper has benefited from presentations at Hitotsubashi University, Institute for Monetary and Economic Studies, Bank of Japan, Kyoto University and the University of Tokyo. This research was supported by Grants-in-Aid for Scientific Research No. 17730144 from the Japan Society for the Promotion of Science. Any remaining errors are my own.

only to linear relationships between I(1) variables. After Granger (1995) introduced the concept of nonlinear cointegration, some researchers began to pay attention to nonlinear relationships between nonstationary variables.

Since any relationships that are not linear can be called nonlinear, the concept of nonlinear cointegration is quite broad. Several specific types of nonlinear cointegration have been discussed by various authors. Park and Phillips (2000) established the limiting properties of nonstationary binary choice models where covariates are integrated of order one. Park and Phillips (2001) showed the limiting properties of nonlinear regression models with I(1) regressors. Chang et al. (2001) extend earlier work by Phillips and Hansen (1990) to nonlinear models with integrated time series. Hansen and Seo (2002) developed a test for threshold cointegration. They dealt with a model where a cointegrating vector changes according to the regime to which the error correction term belongs. Corradi et al. (2000) studied nonlinear relationships between variables that are first order Markov processes. Tøstheim (2012) provide an extensive survey on nonlinear and nonstationary time series. They considered an error correction-like system with a nonlinear component and proposed some tests to discriminate linear cointegration from nonlinear cointegration or no cointegration. However, these tests are directed toward specific kinds of nonlinear cointegration and may have low power against other alternatives. It is desirable that a test for linear cointegration be consistent with a wide variety of nonlinear relationships because we typically lack precise information about them in practice. Thus we seek a test for linearity in regressions with I(1) processes that is consistent with a wide variety of nonlinear alternatives as well as no cointegration.

In this paper we propose a generalized version of the RESET test for linearity in regressions with I(1) processes. Note that the linearity in regressions with I(1) processes we consider in this paper is equivalent to the linear cointegration of Engle and Granger (1987)[1]. In this sense we are trying to propose a test for the null hypothesis of linear cointegration. We cannot simply apply the RESET test directly to the present context because it is well known that the limiting distribution of the least squares estimators in regressions with (linear) I(1) processes generally involves second-order bias effects and these make standard statistical inference invalid without modification. In fact we show that second-order bias effects are still present when we use nonlinear transformations of integrated processes as regressors as in the formulation of the RESET test (see de Jong, 2002 for more general treatment on this issue). Thus we propose employing a "leads and lags" estimation technique by Saikkonen (1991) among others to get a test statistics that is free of nuisance parameters. With this modification we can show that the limiting distribution of the test statistic under the null hypothesis of linear cointegration is the χ^2 distribution with degrees of freedom that depend on the number of regressors. Moreover the test that we propose is consistent against a class of nonlinear alternatives. For example, our test for linearity can distinguish linear cointegration models from nonlinear cointegration models that involve the logarithmic function, any distribution type functions,[2] and polynomial functions of finite order. Further the test is consistent against the alternative of no cointegration.

One important feature of the test is that it allows for an endogenous regressor. That is, the

[1] For the rest of the paper we call the standard cointegration concept developed by Engle and Granger (1987) linear cointegration to distinguish it from nonlinear cointegration.

[2] We call bounded and monotonically increasing functions distribution type functions.

regressor may be correlated with the regression error. This not usually allowed in nonlinear regression models with I(1) processes as in Park and Phillips (2000, 2001), but special features of the RESET test enable us to accommodate it. With this generality we would potentially be able to apply the test to many empirical problems which would be excluded when we assume that a regressor is exogenous.

The approach of this article and the asymptotic distribution theory developed here are similar to those developed independently in closely related work by Hong and Phillips (2010), although this article is different in several key aspects. First, it considers simple regression while we consider multiple regression. Second, it allows for an endogenous regressor as our test does, however, it assumes that the regressor is predetermined as in Park and Phillips (2000, 2001) and Chang et al. (2001) although our test does not. Third, it employs a technique similar to the fully-modified OLS proposed by Phillips and Hansen (1990) to deal with second-order bias effects, while we extend the "leads and lags" estimation technique proposed by Saikkonen (1991).

The rest of the paper is organized as follows. Some assumptions and preliminary results are presented in section II. Section III explains our test for linearity and the power property of our test is examined in section IV. Section V gives some simulation evidence. We summarize some conclusions in Section VII. All proofs are in the Appendix.

A word on notation. For a vector $a=(a_i)$ "$||a||$" stands for the standard Euclidean norm, i.e., $||a||^2=\sum_i a_i^2$. When applied to a matrix, $||A||$ signifies the operator norm, i.e. $||A||=\sup_x ||Ax||/||x||$. We also use $||\cdot||$ to denote the supremum of a function. $||\cdot||_K$ stands for the supremum norm over a subset of K of its domain, $||f||_K=\sup_{x\in K}||f(x)||$. "$\Rightarrow$" denotes weak convergence with respect to the Skorohod metric (as defined in Billingsley (1968)). $[s]$ denotes the largest integer not exceeding s.

II. *Assumptions and Preliminary Results*

The regression model from which we derive a test statistic is driven by a sequence of innovation variables denoted by $\{u_t\}$ where u_t consists of a scalar time series u_{1t} and an $m\times 1$ vector time series $u_{2t}=(u_{21,t},u_{22,t},\ldots,u_{2m,t})$, i.e. $u_t=(u_{1t},u'_{2t})'$. We assume throughout that the innovation sequence $\{u_t\}$ satisfies the following assumption.

Assumption 2.1 *For some $p>\beta>2$, $\{u_t\}$ is a zero mean, strong mixing sequence with mixing coefficients α_m of size $-p\beta/(p-\beta)$ and $\sup_{t\geq 1}(\mathbf{E}|u_{1t}|^p+\sum_{i=1}^m \mathbf{E}|u_{2i,t}|^p)^{1/p}=C<\infty$. In addition, $(1/T)\mathbf{E}(U_T U'_T)\to\Omega$ as $n\to\infty$ where $U_i=\sum_{j=1}^i u_j$.*

For example, Assumption 2.1 permits u_t to be weakly dependent with possible heterogeneity. A wide variety of data generating processes satisfies Assumption 2.1, including invertible autoregressive moving average (ARMA) processes under general conditions. Assumption 2.1 is one of the common assumptions for innovation processes. We sometimes maintain the following assumption in addition to Assumption 2.1.

Assumption 2.2 *We assume u_t is a general linear process*

$$u_{1t}=\phi(L)e_{1,t}=\sum_{i=0}^{\infty}\phi_i e_{1,t-i} \quad u_{2t}=\Psi(L)e_{2,t}=\sum_{i=0}^{\infty}\Psi_i e_{2,t-i} \tag{1}$$

where ϕ_i is a scalar with $\phi_0=1$, Ψ_i is an $(m\times m)$ matrix with $\Psi_0=I_m$, $\{e_{1t}\}$ is a scalar sequence and $\{e_{2t}\}$ is an $(m\times 1)$ vector sequence. $e_t=(e_{1t},e'_{2t})'$ is iid with mean zero and covariance Σ_e.

(a) $\phi(1)$ *nonsingular,* $\sum_{i=0}^{\infty} i\|\phi_i\|<\infty$, *and* $\sup_{t\geq 1}\mathbf{E}\|e_{2t}\|^q<\infty$ *for some* $q>4$.

(b) $\Psi(1)$ *nonsingular,* $\sum_{i=0}^{\infty} i\|\Psi_i\|<\infty$, *and* $\mathbf{E}\|e_{2t}\|^r<\infty$ *for some* $r>8$. *e_{2t} has a distribution that is absolutely continuous with respect to Lebesgue measure and has a characteristic function $\psi(t)$ that satisfies* $\lim_{\|t\|\to\infty}\|t\|^{\xi}\psi(t)=0$ *for some* $\xi>0$.

In the following sections, u_{1t} serves as a regression error process and u_{2t} generates an integrated process. The nonsingularity and summability conditions for ϕ and Ψ in Assumption 2.2 are common in stationary time series analysis. Assumption 2.2 (b) states stronger conditions on the moment and characteristic function for e_{2t} than for e_{1t}. It will be needed when we deal with nonlinear transformations of integrated processes. Assumption 2.2 (b) is commonly imposed in nonlinear regression models with integrated regressors.[3] Processes that satisfy both Assumptions 2.1 and 2.2 include invertible ARMA models under general conditions. Note that $\{u_{1t}\}$ is allowed to have a general correlation structure with $\{u_{2t}\}$. This is usually not the case as in Park and Phillips (2000, 2001). We will return to this point in section IV.

Under Assumption 2.1, the sequence $\{u_t\}$ satisfies a multivariate invariance principle.

Lemma 2.1 (Wooldridge and White 1988)

$$T^{-1/2}\sum_{t=1}^{[Tr]}u_t\Rightarrow B(r),\ 0<r\leq 1,$$

where $B(r)=(B_1(r),B_2(r)')'$ is an $(m+1)$ dimensional Brownian motion with covariance matrix Ω. $B_1(r)$ and $B_2(r)=(B_{21},...,B_{2m})'$ denote Brownian motions of 1 *and m dimensions respectively.*
We assume that Ω can be written as

$$\Omega=\begin{bmatrix}\omega_{11} & \omega'_{21}\\ \omega_{21} & \Omega_{22}\end{bmatrix}=\lim_{T\to\infty}T^{-1}\mathbf{E}(U_TU'_T)=\Sigma+\Lambda+\Lambda',$$

where

$$\Sigma=\begin{bmatrix}\sigma_{11} & \sigma'_{21}\\ \sigma_{21} & \Sigma_{22}\end{bmatrix}=\lim_{T\to\infty}T^{-1}\sum_{t=1}^{T}\mathbf{E}(u_tu'_t),\ \Lambda=\begin{bmatrix}\lambda_{11} & \lambda_{12}\\ \lambda_{21} & \Lambda_{22}\end{bmatrix}=\lim_{T\to\infty}T^{-1}\sum_{t=2}^{T}\sum_{j=1}^{t-1}\mathbf{E}(u_ju'_t).$$

These notation will be used repeatedly throughout the paper. We assume that the covariance matrices ω_{11} and Ω_{22} of $B_1(r)$ and $B_2(r)$ are positive definite. It will often be convenient to write these and other stochastic processes on $[0,1]$ without the argument. Thus, we shall frequently use B, B_1 and B_2 in place of $B(r)$, $B_1(r)$, and $B_2(r)$.

[3] See Akonom (1993) and Park and Phillips (1999) for more details on these conditions.

Let an m-vector time series $\{x_t\}$ satisfy $x_t = x_{t-1} + u_{2t}$ where $x_t = (x_{1t},...,x_{mt})'$. Our results do not depend on the initialization x_0 as long as it is bounded in probability. For notational convenience we assume $x_0 = 0$. Let $x_t^{(j)} = (x_{1t}^j,...,x_{mt}^j)'$ and $B_2^{(j)} = (B_{21}^j,...,B_{2m}^j)'$ where j is a positive integer. In Lemma 2.2, we show the limiting distributions of some partial sums that will be needed to derive the limiting distributions of test statistics. The limit distributions are expressed as functions of Brownian motion. To simplify formulae, all integrals are understood to be taken over the interval [0, 1] unless otherwise stated, and integrals such as $\int B$ and $\int B^{(\kappa)}$ are understood to be taken with respect to Lebesgue measure.

The following lemma is very useful in the derivation of our result in the next section.

Lemma 2.2 *Let Assumption 2.1 hold with $\beta = \kappa + 1$. Then for $2 \leq i,j \leq \kappa$,* as $T \to \infty$

(a) $T^{-2}\sum_{t=1}^{T} x_t x_t' \Rightarrow \int B_2 B_2'$,

(b) $T^{-(i+j+2)/2}\sum_{t=1}^{T} x_t^{(i)} x_t^{(j)\prime} \Rightarrow \int B_2^{(i)} B_2^{(j)\prime}$,

(c) $T^{-1}\sum_{t=1}^{T} x_t u_{1,t+1} \Rightarrow \int B_2 dB_1 + \Lambda_{21}$,

(d) $T^{-1}\sum_{t=1}^{T} x_t u_{1t} \Rightarrow \int B_2 dB_1 + \Delta_{21}$,

(e) $T^{-(i+1)/2}\sum_{t=1}^{T} x_t^{(i)} u_{1,t+1} \Rightarrow \int B_2^{(i)} dB_1 + iD(B_2^{(i-1)})\Lambda_{21}$,

(f) $T^{-(i+1)/2}\sum_{t=1}^{T} x_t^{(i)} u_{1t} \Rightarrow \int B_2^{(i)} dB_1 + iD(B_2^{(i-1)})\Delta_{21}$,

where $\Delta_{21} = \Sigma_{21} + \Lambda_{21}$ and $D(B_2^{(i)}) = \mathrm{diag}\left[\int B_{21}^i, \int B_{22}^i, ..., \int B_{2m}^i\right]$.

Parts (a) − (d) of Lemma 2.2 are standard results that can be found in the literature (e.g., Phillips, 1987 and Park and Phillips, 1988) or can be derived easily from it. However, part (e) of Lemma 2.2 is nonstandard and part (f) is an extension of part (e). Part (e) can be considered as an extension of the results of Park and Phillips (1999, 2001) in the sense that we extend their results to a case where a regressor x_t is endogenous and multivariate. Recently de Jong, (2002) extended the results by Park and Phillips (1999, 2001) to accommodate general correlation structure between u_{1t} and u_{2t} under a different set of assumptions. However, we note that his result still deals with a scalar process u_{2t} rather than a multivariate process as considered in Lemma 2.2 although it includes results for general functional forms other than polynomials.

III. *Testing for Linearity in Regressions with I(1) Processes*

In this section we propose a generalized version of the RESET test for linearity in regression with I(1) processes. Consider the following regression model:

$$y_t = \gamma_0 + \gamma_1' x_t + \gamma_2' x_t^{(2)} + \gamma_3' x_t^{(3)} + \cdots + \gamma_\kappa' x_t^{(\kappa)} + u_{1t},\ t = 1,...,T. \tag{2}$$

where γ_0 is a scalar parameter, γ_i is an $(m \times 1)$ parameter vector for $1 \leq i \leq \kappa$, x_t, $x_t^{(j)}$ for

$2 \leq j \leq \kappa$, and u_{1t} are defined in the previous section. Our test is a generalized version of the test for functional misspecification proposed by Thursby and Schmidt (1977) that is a variant of the RESET test originally proposed by Ramsey (1969). If $\{x_t\}$ is stationary and $\{u_{1t}\}$ is normally distributed, the present situation reduces to that in Thursby and Schmidt (1977).

The idea of their test is that if there is functional misspecification, i.e. if the functional form is nonlinear, "the omitted portion of the regression is definitely a function of the included regressors." If this function is analytic, it can be expressed in a Taylor series expansion, involving powers and cross products of the explanatory variables. Hence they proposed to test whether coefficients of powers of the explanatory variables were zero or not. Since this justification does not depend on the property of the process $\{x_t\}$, it would be natural to expect that this test will work even if $\{x_t\}$ is I(1) as in our present situation. However, note that we are not claiming that our test is consistent against nonlinear cointegration because of this argument. We must prove consistency against a whole class of nonlinear alternatives and no cointegration and this is covered in the next section.

Another word on the regression model (2). We do not include cross products of the explanatory variables. Thursby and Schmidt (1977) found that they do not contribute to the power of their test very much through Monte Carlo experiments. Since the present situation is different from theirs, those cross products may contribute to the power of the test in the present circumstance. However, they are not included in the regression (2) in order to keep our theoretical development simple.

The null hypothesis of linearity or linear cointegration between y_t and x_t corresponds to

$$H_0: \quad \gamma_2 = \cdots = \gamma_\kappa = 0. \tag{3}$$

If the null hypothesis is true, the specification in (2) would correspond to "deterministic cointegration" as defined by Ogaki and Park (1997). The results that will be shown in this section can easily be extended to "stochastic cointegration" where nonzero deterministic time trends are present in (2). The null hypothesis (3) is to be tested against the alternative of nonlinear cointegration or no cointegration. In this section we will present the limiting property of the test under the null hypothesis of linear cointegration and establish the limiting property under the alternative of nonlinear cointegration and no cointegration in the next section. The next theorem characterizes the limiting distribution of the least squares estimator from the regression model (2) under the null hypothesis.

Theorem 3.1 *Suppose Assumption 2.1 holds with $\beta = \kappa + 1$. Then under the null hypothesis* (3) *(as $T \to \infty$)*

$$\Upsilon_T \begin{bmatrix} \hat{\gamma}_0 - \gamma_0 \\ \hat{\gamma}_1 - \gamma_1 \\ \hat{\gamma}_2 \\ \vdots \\ \hat{\gamma}_\kappa \end{bmatrix} \Rightarrow \begin{bmatrix} 1 & \int B_2' & \int B_2^{(2)\prime} & \cdots & \int B_2^{(\kappa)\prime} \\ \int B_2 & \int B_2 B_2' & \int B_2 B_2^{(2)\prime} & \cdots & \int B_2 B_2^{(\kappa)\prime} \\ \int B_2^{(2)} & \int B_2^{(2)} B_2' & \int B_2^{(2)} B_2^{(2)\prime} & \cdots & \int B_2^{(2)} B_2^{(\kappa)\prime} \\ \vdots & \vdots & \vdots & \ddots & \\ \int B_2^{(\kappa)} & \int B_2^{(\kappa)} B_2' & \int B_2^{(\kappa)} B_2^{(2)\prime} & \cdots & \int B_2^{(\kappa)} B_2^{(\kappa)\prime} \end{bmatrix}^{-1} \begin{bmatrix} B_1 \\ \int B_2 dB_1 + \Delta_{21} \\ \int B_2^{(2)} dB_1 + D(B_2)\Delta_{21} \\ \vdots \\ \int B_2^{(\kappa)} dB_1 + D(B_2^{(\kappa-1)})\Delta_{21} \end{bmatrix}$$

where $\Upsilon_T = \mathrm{diag}[T^{1/2}, TI_m, T^{3/2}I_m, \ldots, T^{(\kappa+1)/2}I_m]$

Before we move on to the development of our test statistic, some remarks are in order. As described above, our null hypothesis is that there exists a linear cointegration relationship between y_t and x_t. Therefore it would not be hard to imagine that the limiting properties of the least squares estimates of the present regression model under H_0 would share some characteristics with the least squares estimates of cointegrating vectors in standard cointegrated regression models.

First, a regressor x_t is allowed to be endogenous under Assumption 2.1, i.e. x_t can be correlated with the regression error u_{1t} as in linearly cointegrated regression models. When it is enogenous in stationary regression models, the least squares estimator fails to satisfy the conditions for consistency and therefore we typically employ an instrumental variable estimator to achieve consistency. However, one notable difference between stationary regression models and linearly cointegrated regression models is that the least squares estimator in the latter is still consistent for its population value (e.g., Stock, 1987, Park and Phillips, 1988, and Phillips and Hansen, 1990). In the present regression model, this is true and the least squares estimator is consistent even though we have an endogenous regressor, as shown in Theorem 3.1.

Second, we have second order bias effects such as Δ_{21}, $D(B_2)\Delta_{21}, \ldots$, $D(B_2^{(\kappa-1)})\Delta_{21}$ in the limiting distribution of Theorem 3.1 that are similar to the limiting properties of the cointegrating vectors in linear cointegrating models. We call this the second order bias because it does not have an effect on the consistency result but does have an effect on the limiting distribution (see Stock, 1987 and Phillips and Hansen, 1990). It arises because of the existence of contemporaneous and serial dependence between the regressors x_t and the regression error u_{1t}. This is directly analogous to the phenomenon that occurs in linearly cointegrated regression.

Next we propose our test statistic. There are two obstacles in the limiting distribution of the least squares estimator given in Theorem 3.1 in conducting a standard hypothesis testing procedure such as a χ^2 test. One is the existence of a nonzero covariance structure between B_1 and B_2 and another is that the limiting distribution of the least squares estimates depends not only on the property of the Brownian motion B_1 and B_2 but also on the nuisance parameter matrix Δ_{21}. These obstacles are same as those arising in linearly cointegrated models and the methods proposed to remove these obstacles in linearly cointegrated regression models can be extended to our regression model. Here we consider an estimation technique by Saikkonen (1991).[4]

Saikkonen (1991) proposed an efficient estimator that eventually removes the obstacles by adding leads and lags of Δx_t in linearly cointegrated regressions where $\Delta x_t = x_t - x_{t-1}$. We show that this "leads and lags" estimation technique works in our regression model (2).

Consider the following new regression model:

$$y_t = \gamma_0 + \gamma_1' x_t + \gamma_2' x_t^{(2)} + \gamma_3' x_t^{(3)} + \cdots + \gamma_\kappa' x_t^{(\kappa)} + \sum_{s=-K}^{K} \theta_s' \Delta x_{t-s} + v_t^*, \; t = 1, \ldots, T, \tag{4}$$

where θ_i is an $(m \times 1)$ parameter vector for $-K \leq i \leq K$. This is a regression model where leads

[4] Fully-modified least squares developed by Phillips and Hansen (1990) is also applicable to the present problem although it is not presented here. See de Jong (2004) and Hong and Phillips (2010) for fully-modified least squares in nonlinear regression models when a regressor is a scalar.

and lags of Δx_t are added to the regression model (2).[5] Note that the regression error here is not u_{1t} but v_t^*. The relationship between them is characterized below. To derive the limiting distribution of the least squares estimator of the regression model (4), we need to make the following assumption on the error process u_t in (2):

Assumption 3.1 (a) *$\{u_t\}$ is strictly stationary with the spectral density matrix $f_{uu}(\lambda)$ bounded away from zero so that $f_{uu}(\lambda)\geq \alpha I_n$, $\lambda\in[0,\pi]$, where $\alpha>0$.*
(b) *The covariance function of u_t is absolutely summable $\sum_{j=-\infty}^{\infty}\|\Gamma(j)\|<\infty$, where $\Gamma(j)=\mathbf{E}(u_t u'_{t+j})$ and $\|\cdot\|$ is the standard Euclidean norm.*
(c) *The fourth order cumulants of u_t, denoted by $c_{ijkl}(\mathrm{m}_1,\mathrm{m}_2,\mathrm{m}_3)$, satisfy $\sum_{m_1,m_2,m_3=-\infty}^{\infty}|c_{ijkl}(\mathrm{m}_1,\mathrm{m}_2,\mathrm{m}_3)|<\infty$.*

It is well known that we can deduce under Assumption 3.1 that $u_{1t}=\sum_{j=-\infty}^{\infty}\Pi_j u_{2,t-j}+v_t$ where $\sum_{j=-\infty}^{\infty}\|\Pi_j\|<\infty$ and v_t is a stationary process with the property that $\mathbf{E}(u_{2t}v_{t+k})=0,\ k=0,\pm1,\pm2,\cdots$. Furthermore, $2\pi f_{vv}(0)=\omega_{11}-\omega'_{21}\Omega_{22}^{-1}\omega_{21}$ where $f_{vv}(\lambda)$ is the spectral density of v at frequency λ. These are key properties that play important roles in proving the next theorem. Note that v_t^* in (4) can be represented as $v_t^*=v_t+\sum_{|j|>K}\Pi_j u_{2,t-j}$. If $\Pi_j=0$ for $|j|>K$, then v_t^* is strictly exogenous and we get the desired limiting properties of the coefficient estimator in (4). That is, there exist neither the second order bias effects nor the correlation between B_1 and B_2. However, this is not the case in general. Thus we also need to make an assumption on the truncation parameter K:

$$K^s\sum_{j>|K|}\|\Pi_j\|^2\to 0 \quad \text{for some } s\geq 5. \tag{5}$$

We must let $K\to\infty$ as $T\to\infty$. We choose the rate of $K=T^\delta$ such that $\frac{1}{s}<\delta<\frac{r}{2(2+3r)}$, where r is given by the moment condition for e_{2t} in Assumption 2.2. For example, invertible ARMA models satisfy Assumptions 2.2 and (5) for any finite r and s under general conditions. In this case δ can take any value between 0 and 1/6. The condition (5) is analogous to Assumption 5.1 of Chang et al. (2001) although the admissible values of δ are different. In fact, the condition (5) is more than necessary to derive the limiting distribution in Theorem 3.2, but it will be required when we deal with the limiting property under the alternative of nonlinear cointegration.

The regression model (4) leads to the following limiting distribution for the least squares estimator: Let $(\tilde{\gamma}_0,\tilde{\gamma}'_1,\tilde{\gamma}'_2,\ldots,\tilde{\gamma}'_K)'$ be the least squares estimator of $(\gamma_0,\gamma'_1,\gamma'_2,\ldots,\gamma'_K)'$ in the regression model (4).

Theorem 3.2 *Suppose that $\{w_t\}$ satisfies Assumption 2.1 with $\beta=\kappa+1$ where $w_t=(v_t,u'_{2t})'$. Also suppose Assumption 3.1 and the conditions* (5) *on the truncation parameter K hold. Then under the null hypothesis* (3) *(as $T\to\infty$)*

[5] It is possible to allow the leads and lags to be different as in Choi and Kurozumi (2012). We proceed with the same length K for simplicity.

$$\Upsilon_T\begin{bmatrix}\tilde{\gamma}_0-\gamma_0\\ \tilde{\gamma}_1-\gamma_1\\ \tilde{\gamma}_2\\ \vdots\\ \tilde{\gamma}_\kappa\end{bmatrix}\Rightarrow\begin{bmatrix}1 & \int B_2' & \int B_2^{(2)\prime} & \cdots & \int B_2^{(\kappa)\prime}\\ \int B_2 & \int B_2B_2' & \int B_2B_2'^{(2)\prime} & \cdots & \int B_2B_2^{(\kappa)\prime}\\ \int B_2^{(2)} & \int B_2^{(2)}B_2' & \int B_2^{(2)}B_2^{(2)\prime} & \cdots & \int B_2^{(2)}B_2^{(\kappa)\prime}\\ \vdots & \vdots & \vdots & \ddots & \\ \int B_2^{(\kappa)} & \int B_2^{(\kappa)}B_2' & \int B_2^{(\kappa)}B_2^{(2)\prime} & \cdots & \int B_2^{(\kappa)}B_2^{(\kappa)\prime}\end{bmatrix}^{-1}\begin{bmatrix}B_{11\cdot 2}\\ \int B_2dB_{11\cdot 2}\\ \int B_2^{(2)}dB_{11\cdot 2}\\ \vdots\\ \int B_2^{(\kappa)}dB_{11\cdot 2}\end{bmatrix}$$

where $B_{11\cdot 2}(r)=B_1(r)-\omega_{21}'\Omega_{22}^{-1}B_2(r)$ *with covariance matrix* $\omega_{11}-\omega_{21}'\Omega_{22}^{-1}\omega_{21}$.

There are two notable differences between the limiting distribution given in Theorem 3.1 and that in Theorem 3.2. First, the Brownian motions B_1 and B_2 in Theorem 3.1 are generally correlated, but the Brownian motions $B_{11\cdot 2}$ and B_2 in Theorem 3.2 are uncorrelated, implying independence due to the Gaussian properties of a Brownian motion. Second, the second order bias terms that are present in Theorem 3.1 vanish in Theorem 3.2. Therefore the limiting distribution in Theorem 3.2 is free of the obstacles mentioned above and so we can apply the standard hypothesis testing procedure. Thus Theorem 3.2 suggests that we use the Wald statistic. Let $\underline{\tilde{\gamma}}=(\tilde{\gamma}_0,\tilde{\gamma}_1',\tilde{\gamma}_2',\ldots,\tilde{\gamma}_\kappa')'$, $M=\sum_{t=1}^{T}X_tX_t'$, $X_t=(1,x_t',x_t^{(2)\prime},\ldots,x_t^{(\kappa)\prime})'$, $\widehat{\omega}_{11\cdot 2}$ be any consistent estimator of $\omega_{11}-\omega_{21}'\Omega_{22}^{-1}\omega_{21}$ (see Newey and West, 1987, Phillips, 1987 and Andrews, 1991 for a discussion of possible estimators), and

$$R=\begin{bmatrix}0 & 0 & I_m & 0 & 0 & \cdots & 0\\ 0 & 0 & 0 & I_m & 0 & \cdots & 0\\ 0 & 0 & 0 & 0 & I_m & \cdots & 0\\ \vdots & & & & & \ddots & \vdots\\ 0 & 0 & 0 & & & 0 & I_m\end{bmatrix}\quad m(\kappa-1)\times(m\kappa+1),$$

Construct the following statistic:

$$W_T=(R\underline{\tilde{\gamma}})'(\widehat{\omega}_{11\cdot 2}RM^{-1}R')^{-1}(R\underline{\tilde{\gamma}}).$$

The next theorem shows the limiting distribution of this statistic under the null hypothesis.

Theorem 3.3 *Suppose the conditions in Theorem 3. 2 are satisfied. Then under the null hypothesis* (3) *(as $T\to\infty$)*

$$W_T\Rightarrow\chi^2_{m(\kappa-1)}.$$

Theorem 3.3 shows that we can apply the standard χ^2 test procedure to our test. If $\Omega_{21}=0$, the test statistic based on the estimator considered in Theorem 3.1 has the limiting distribution given in Theorem 3.3. For example, this will occur when x_t is strictly exogenous and the driving process u_{2t} is independent of the regression error u_{1t}.

IV. *Power of The Test*

In this section we show that the proposed test is consistent against a class of nonlinear alternatives and no cointegration. First, we shall consider the types of nonlinear functions for which our test for linear cointegration is consistent. From the construction of the test statistic discussed in the previous section, it would be clear that the test is consistent against nonlinear cointegration involved in finite order polynomial functions of x_t. So now we are interested in for which types of nonlinear functions other than finite order polynomial functions the test is consistent. Consider the following alternative hypothesis:

$$H_1: \quad y_t = g_1(x_{1t}) + g_2(x_{2t}) + \cdots + g_m(x_{mt}) + u_{1t}, \tag{6}$$

where $g_i : R \to R$ is a nonlinear measurable function for $1 \leq i \leq m$. This structure is same as that considered in Chang et al. (2001). To prove consistency, we must investigate the limiting property of the test statistic under the alternative. This involves some nonlinear transformations of integrated variables. However, the limiting properties of nonlinear functions of integrated time series are fairly complicated. These remained unknown until Park and Phillips (1999) showed the limiting properties of nonlinear transformations of "scalar" integrated time series.[6] Unfortunately, analogous results for vector-valued integrated time series has not yet been proven. Since we use their results, we confine ourselves to alternatives that can be expressed by (6).

We consider the following two classes of functions treated in Park and Phillips (1999), the integrable class $\mathcal{T}(I)$ and the homogeneous class $\mathcal{T}(H)$

Definition 4.1 (Park and Phillips 1999) (a) *A transformation T on R is said to be regular if and only if* (i) *it is continuous in a neighborhood of infinity, and* (ii) *for any compact subset K on R given, there exist for each* $\varepsilon > 0$ *continuous functions* $\underline{T}_\varepsilon$, $\overline{T}_\varepsilon$, *and* $\delta_\varepsilon > 0$ *such that* $\underline{T}_\varepsilon(x) \leq T(y) \leq \overline{T}_\varepsilon(x)$ *for all* $|x-y| < \delta_\varepsilon$ *on K, and such that* $\int_K (\underline{T}_\varepsilon - \overline{T}_\varepsilon)(x)dx \to 0$ *as* $\varepsilon \to 0$.

(b) *A transformation T is said to be in Class (I), denoted by* $T \in \mathcal{T}(I)$, *if it is bounded and integrable.*

(c) *A transformation T is said to be in Class (H), denoted by* $T \in \mathcal{T}(H)$, *if and only if*

$$T(\lambda x) = \nu(\lambda)h(x) + R(x, \lambda) \tag{7}$$

where h is regular and $R(x,\lambda)$ *is of order smaller than* $\nu(\lambda)$. ν *and h are sometimes called the asymptotic order and the limit homogeneous function of T respectively.*

All homogeneous functions belong to $\mathcal{T}(H)$ as long as they are locally integrable. Other functions that belong to $\mathcal{T}(H)$ include polynomials of finite order, the logarithmic function and the distribution function of any random variable. Each of the two classes, $\mathcal{T}(I)$ and $\mathcal{T}(H)$ is closed under the operations of addition, subtraction, and multiplication (see Park and Phillips, 1999 for more details). In the following, if $g_i \in \mathcal{T}(H)$, we denote its asymptotic order by $\nu_i(\lambda)$

[6] Nonstationary binary choice models by Park and Phillips (2000) allows covariates to be multivariate. However, it essentially reduces to a scalar case by decomposing the covariates (See Park and Phillips, 2000 for more details).

and its limit homogeneous function by $h_i(x)$.

Before we develop the limiting property of the test statistic under the alternative (6), we show some useful results that are helpful in proving consistency of the test and that give some intuition about why the proposed test works. To do so, we assume either $g_i \in \mathcal{T}(I)$ or $g_i \in \mathcal{T}(H)$ for all i. If there exists at least one i such that $g_i \in \mathcal{T}(H)$, without loss of generality we let g_1 be the function that is in $\mathcal{T}(H)$ and dominates other g_i's belonging to $\mathcal{T}(H)$ asymptotically, i.e. for any $j \neq 1$ such that $g_j \in \mathcal{T}(H)$, $\dfrac{\nu_1(T^{1/2})}{\nu_j(T^{1/2})} \to \infty$ or $\dfrac{\nu_1(T^{1/2})}{\nu_j(T^{1/2})} \to$ constant. For part (b) in the next lemma, we must specify the consistent estimator for $\omega_{11\cdot 2}$ explicitly. We employ the semiparametric consistent estimator

$$\widehat{\omega}_{11\cdot 2} = T^{-1} \sum_{t=1}^{T} \widehat{v}_t^{*2} + 2T^{-1} \sum_{s=1}^{l} w_{sl} \sum_{t=s+1}^{T} \widehat{v}_t^* \widehat{v}_{t-s}^*$$

where $\widehat{v}_t^*$ is the residual obtained from the regression (4) and $w_{sl} = 1 - s/(l+1)$. This is one of the standard choices for a consistent estimator in the present context (see Phillips, 1987, Newey and West, 1987, and Andrews, 1991 for more discussions on this choice).

Lemma 4.1 *Let $\{u_t\}$ satisfy Assumption 2.1 with $\beta = \kappa + 1$, Assumptions 2.2 and 3.1 and the truncation parameter K satisfies* (5). *Also, suppose either $g_i \in \mathcal{T}(I)$ or $g_i \in \mathcal{T}(H)$ for all i, and the limit homogeneous function h_i is piecewise differentiable with a locally bounded derivative for i such that $g_i \in \mathcal{T}(H)$. In addition, assume that for some $q \geq 1$ there exists a grid $\{a_1, \ldots, a_q\}$, where $a_j < a_{j+1}$ for all $j = 1, \ldots, q-1$, such that h_i is continuous at any $x \in R \backslash \{a_1, \ldots, a_q\}$, and monotone on (a_{j-1}, a_j) for $j = 1, \ldots, q+1$ for i such that $g_i \in \mathcal{T}(H)$.*

(a) *Let $\underline{\tilde{\gamma}} = (\tilde{\gamma}_0, \tilde{\gamma}_1', \tilde{\gamma}_2', \ldots, \tilde{\gamma}_\kappa')'$ be the least squares estimator of $(\gamma_0, \gamma_1', \gamma_2', \ldots, \gamma_\kappa')'$ in the regression model (4).*
    (i) *If $g_i \in \mathcal{T}(H)$ with $T^{1/2}\nu_i(T^{1/2}) \to \infty$ as $T \to \infty$ for some i, then under the alternative, $\Upsilon_T \underline{\tilde{\gamma}} = O_p(T^{1/2}\nu_1(T^{1/2}))$,*
    (ii) *Otherwise, under the alternative* (6), $\Upsilon_T \underline{\tilde{\gamma}} = O_p(1)$,

(b) *Suppose $l \to \infty$ as $T \to \infty$ such that $l = o(T)$.*
    (i) *If $g_i \in \mathcal{T}(H)$ with $\nu_i(T^{1/2}) \to \infty$ as $T \to \infty$ for some i, then under the alternative (6), $\widehat{\omega}_{11\cdot 2} = O_p(l\nu_1^2(T^{1/2}))$,*
    (ii) *Otherwise, under the alternative (6), $\widehat{\omega}_{11\cdot 2} = O_p(l)$.*

The intuition behind Lemma 4.1 is clear if we consider some simple functional forms of g. For example, let's consider $g(x) = x^{5/2}$. This function belongs to $\mathcal{T}(H)$ with $\nu(T^{1/2}) = T^{5/4}$. Thus Lemma 4.1 (a) implies that $\Upsilon_T \underline{\tilde{\gamma}} = O_p(T^{7/4})$. In particular, we have $T^{3/2}\tilde{\gamma}_2 = O_p(T^{7/4})$ or equivalently $\tilde{\gamma}_2 = O_p(T^{1/4})$. That is, the coefficient estimator of $x_t^{(2)}$ diverges at the rate $T^{1/4}$ and this would be a signal of nonlinearity. Now consider $g(x) = x^{1/2}$. In this case, Lemma 4.1 implies $\tilde{\gamma}_2 = O_p(T^{-3/4})$. Thus $\tilde{\gamma}_2$ converges to zero but at much slower rate than under the null hypothesis because $\tilde{\gamma}_2 = O_p(T^{-3/2})$ under the null, as we can see in Theorem 3.2. If it converges to zero at a rate that is slow enough, it would be interpreted as a signal of nonlinearity and the next theorem tells us how slow it must be for the test to be consistent.

The following theorem tells us for which type of functions our test is consistent.

Theorem 4.1 *Let the conditions in Lemma 4.1 hold.*

(a) *If* $g_i \in \mathcal{T}(H)$ *with either* $\nu_i(T^{1/2}) \to \infty$ or $\nu_i(T^{1/2})$ *is constant as* $T \to \infty$ *for some* i, *then under the alternative (6),* $W_T = O_p(T/l)$,

(b) *If* $g_i \in \mathcal{T}(H)$ *for some* i *and if* $T^{1/2}\nu_1(T^{1/2}) \to \infty$ *and* $\nu_1(T^{1/2}) \to 0$ *as* $T \to \infty$, *then under the alternative (6),* $W_T = O_p(T\nu_1^2(T^{1/2})/l)$.

(c) *Otherwise, under the alternative (6),* $W_T = O_p(l^{-1})$.

First, Theorem 4.1 (a) shows that our test for linear cointegration is consistent against nonlinear cointegration if either $g_i \in \mathcal{T}(H)$ with $\nu(T^{1/2}) \to \infty$ or $\nu(T^{1/2})$ is constant as $T \to \infty$ for some i, i.e. consistency of the test can be achieved if there exists at least one function that satisfies the conditions of Theorem 4.1 (a). A class of functions that satisfy the conditions of Theorem 4.1 (a) includes $g_i(x) = |x|^k$ for $k > 0$ and $k \neq 1$, $g_i(x) = 1/(1+e^{-x})$, the logarithmic function $g_i(x) = \log|x|$, polynomial functions of finite order $g(x) = x^k + a_1 x^{k-1} + \cdots + a^k$ for $k > 1$. All distribution functions also satisfy the conditions of Theorem 4.1 (a).

Second, we can deduce from Theorem 4.1 (b) that if $g_i \in \mathcal{T}(H)$ for some i and $T^{1/2}\nu_1(T^{1/2}) \to \infty$, $\nu_1(T^{1/2}) \to 0$ and $T\nu_1^2(T^{1/2})/l \to \infty$ as $T \to \infty$, the test is still consistent but the test statistic diverges at a slower rate than in case (a). For example, this happens when all functions $g_i(x)$ for $1 \leq i \leq m$ decrease to zero as x goes to infinity, but at least one of them decreases to zero at a moderate rate as in a case where $g_1(x) = |x|^{-1/2}$ and we choose the lag truncation parameter l such that $l = o(T^{1/3})$. In this case, $W_T = O_p(T^{1/2}/l)$ and W_T diverges at an approximate rate of $T^{1/6}$ that is much slower than that for case (a). The argument above shows that a choice of the lag truncation parameter is crucial for case (b). If we choose l such that $l = O(T^{1/2})$, the test becomes inconsistent for $g_1(x) = |x|^{1/2}$. Thus we must be careful about the choice of l when we are especially interested in this type of nonlinear alternatives. A class of functions that satisfies the conditions of Theorem 4.1 (b) includes $g_i(x) = |x|^k$ for $-2/3 \leq k \leq 0$ when we choose $l = o(T^{1/3})$.

Finally, Theorem 4.1 (c) implies that our test is inconsistent if all functions $g_i(x)$ $(1 \leq i \leq m)$ decrease rapidly as x goes to infinity such as when $g_i(x) = |x|^k$ where $k < -2/3$ and we use the lag truncation parameter l such that $l = o(T^{1/3})$ or especially when all functions are integrable. This is expected from Lemma 4.1 because in this case $\tilde{\gamma}$ converges to zero at the same rate as it does under the null hypothesis.

One important characteristic of our test for linear cointegration is that it allows for an endogenous regressor as mentioned in the last section. Researchers who are familiar with nonlinear regression models with integrated regressors from Park and Phillips (2000, 2001) and Chang et al. (2001) may wonder why we can do this because all models mentioned here assume that x_t is predetermined and (u_{1t}, F_t) is a martingale difference sequence where $\{F_t\}$ is a natural filteration to which $\{u_{1t}\}$ is adapted. When this is the case, x_t is uncorrelated with u_{1t}, i.e. $\mathbf{E}(x_t u_{1t}) = 0$, which rules out an endogenous regressor x_t. In general, when we deal with the limiting properties of nonlinear models with integrated regressors x_t, we must investigate the limiting properties of a sample mean function such as $\sum_{t=1}^{T} f(x_t)$ and a covariance function such as $\sum_{t=1}^{T} f(x_t)u_{1t}$ for some nonlinear function f. An exogenous x_t is critical in deriving the limiting distribution of the covariance function in Park and Phillips (2000, 2001) and Chang et al. (2001). However, as shown in the proofs of Lemma 4.1 and Theorem 4.1, the covariance function which we need to deal with has a specific form where f is a polynomial function. Moreover its limiting properties is provided in Lemma 2.2 (f) without assuming exogeneity. Thus we can allow an endogenous regressor x_t in the alternative model of nonlinear cointegration. This is a very important assumption in practice. Suppose we are interested in investigating the nonlinear relationship between the exchange rate e_t and the fundamentals. We take output y_t and money m_t as proxies for fundamentals and post it a nonlinear relationship $e_t = f_1(y_t) + f_2(m_t) + u_{1t}$. In this model, it would be very unrealistic to assume that y_t and m_t are predetermined because it is usually the case that e_t, y_t and m_t interact simultaneously. Allowing a endogenous regressor will thus make it possible to apply our test of nonlinear cointegration to many economic problems.

Finally we show that the test is also consistent against the alternative of no cointegration. Suppose that the system of y_t and x_t is generated by the following:

$$y_t = y_{t-1} + u_{1t},\ x_t = x_{t-1} + u_{2t},\ t = 1,...,T. \tag{8}$$

In this case, there is neither linear nor nonlinear cointegration in the system and the present problem reduces to that *spurious regressions* as studied by Granger and Newbold (1974) and Phillips (1986). As we have done for the case of nonlinear cointegration, we first show the limiting properties of the normalized coefficient estimator and the long-run variance estimator and then show the limiting properties of the test statistics.

Lemma 4.2 *Suppose that the system of y_t and x_t is generated by (8). Also suppose that $\{u_t\}$ satisfies Assumption 2.1 with $\beta = \kappa + 1$. In addition, suppose that $\{w_t\}$ satisfies Assumptions 2.2 and 3.1 and the truncation parameter K satisfies* (5).

(a) *Let $\underline{\tilde{\gamma}} = (\tilde{\gamma}_0, \tilde{\gamma}'_1, \tilde{\gamma}'_2, ..., \tilde{\gamma}'_K)'$ be the least squares estimator of $(\gamma_0, \gamma'_1, \gamma'_2, ..., \gamma'_K)'$ in the regression model (4). Then as $T \to \infty$, $\Upsilon_T \tilde{\gamma} = O_p(T)$,*

(b) *Suppose $l \to \infty$* as *$T \to \infty$ such that $l = o(T^{1/4})$. Then as $T \to \infty$, $\widehat{\omega}_{11 \cdot 2} = O_p(lT)$.*

Theorem 4.2 *Suppose the conditions in Lemma 4.2 hold. Then as $T \to \infty$, $W_T = O_p(T/l)$,*

In other words the intuition in Lemma 4.1 applies to the case of nonlinear cointegration as well. $\tilde{\gamma}_2$ converges to zero, but at a rate $T^{1/2}$ that is much slower than under the null hypothesis. This slow rate serves as a signal of no cointegration and underlies the consistency of the test.

V. *Some Simulation Evidence*

In this section we show some simulation evidence to investigate the properties of our test in small samples. First we show the size properties. For the study of size properties, we use the following data generating process (DGP):

$$DGP_1: \quad y_t \quad = \quad 1.5x_t + u_{1t}, \quad x_t = x_{t-1} + u_{2t},$$

with $x_0=0$. $u_t=(u_{1t},u_{2t})'$ is generated from a simplified version of a MA process (1)

$$\begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix} = \begin{bmatrix} e_{1t} \\ e_{2t} \end{bmatrix} + \begin{bmatrix} \phi_1 & 0 \\ 0 & 0.5 \end{bmatrix} \begin{bmatrix} e_{1,t-1} \\ e_{2,t-1} \end{bmatrix},$$

where $e_t=(e_{1t},e_{2t})'$ is distributed as $NID(0,\Sigma)$ with

$$\Sigma = \begin{bmatrix} 1 & \sigma_{e,12} \\ \sigma_{e,12} & 1 \end{bmatrix},\ \sigma_{e,12}=0.8, 0.4, 0, -0.4, \text{ and } -0.8.$$

The test statistics are constructed as described in section III. The sizes of the test depend on sample size (T), the lag truncation parameter (l) used to estimate $\hat{\omega}_{11\cdot 2}$, and κ in equation (4). We consider three types of sample size ($T=100,200$ and 400) and four types of l. First three choices of l are $l0=0$, $l4=\left[4(T/100)^{1/4}\right]$ and $l12=\left[12(T/100)^{1/4}\right]$. These choices of l are used in many simulations (e.g. Schwert, 1989 and Kwiatkowski et al., 1992). The last choice of l, denoted l_A, is a truncated version of a data dependent choice by Andrews (1991)

$$l_A = \left[1.1447 \min\left(\left\{ \frac{4T\hat{\rho}^2}{(1-\hat{\rho})^2(1+\hat{\rho})^2} \right\}^{\frac{1}{3}}, \left\{ \frac{4T0.9^2}{(1-0.9)^2(1+0.9)^2} \right\}^{\frac{1}{3}} \right) \right] \tag{9}$$

where $\hat{\rho}$ is a coefficient estimated from the first order autoregression of $\hat{v}_t^*$. The truncation was made to avoid a choice of l which would make our test inconsistent.[7] We use a value of $\kappa=3$ for all experiments. We do not use a value of $\kappa>3$ because for values of κ that are greater than 3 the second moment matrix often becomes close to singular in small samples and therefore we may not be able to get accurate results. We choose $\kappa=3$ rather than $\kappa=2$ because, in terms of size-corrected power, the result using $\kappa=3$ dominates that obtained using $\kappa=2$. The number of leads and lags used to estimate the parameters in (4) is determined by Schwartz's Bayesian criterion[8] with a maximum lag length of 10.

Table 1 shows the size properties. Since we use the upper 5% critical value from a χ^2 distribution, the nominal size of the test is 0.05. For each experiment, the number of replications is 1000. The results of the simulation are summarized as follows: (i) The size of the test becomes closer to the nominal size as the sample size becomes larger. (ii) A nonzero correlation $\sigma_{e,12}$ between e_{1t} and e_{2t} causes moderate degrees of size distortion as opposed to cases where $\sigma_{e,12}=0$. (iii) The size of the test for positive ϕ_1 tends to be larger than the nominal size, while that for negative ϕ_1 tends to be smaller. This positive correlation between the size and the MA parameter ϕ_1 is also commonly observed in unit root tests (e.g. Schwert, 1989). (iv) The size of the test with $l0$ is overly sensitive to ϕ_1. This is a consequence of ignoring

---

[7] For example if $\hat{\rho}=O_p(T)$ as in spurious regression, it is easy to see that $l_A=O_p(T)$, violating the assumption in Lemma4.1. See Kurozumi (2002) for a similar problem in a different application

[8] We also tried different lag length selections such as AIC or general-to-specific procedures such as in Ng and Perron (1995). The results are not very different from those presented here. The lag length selection based on Schwartz's criterion looks only slightly better than others in terms of empirical size. Clearly this does not imply that Schwartz's criterion is the best method, and further investigation on the lag length selection is needed.

TABLE 1. SIZE OF THE TEST

ϕ_1	$\sigma_{e,12}$	T=100				T=200				T=400			
		$l0$	$l4$	$l12$	l_A	$l0$	$l4$	$l12$	l_A	$l0$	$l4$	$l12$	l_A
0.8	0.8	0.258	0.159	0.223	0.157	0.223	0.102	0.131	0.105	0.244	0.086	0.107	0.091
	0.4	0.269	0.138	0.189	0.142	0.249	0.098	0.132	0.105	0.230	0.086	0.101	0.088
	0	0.225	0.115	0.159	0.120	0.242	0.108	0.132	0.108	0.206	0.082	0.099	0.089
	−0.4	0.220	0.112	0.170	0.112	0.220	0.101	0.129	0.109	0.238	0.081	0.100	0.091
	−0.8	0.264	0.144	0.199	0.144	0.238	0.113	0.141	0.111	0.242	0.076	0.088	0.077
0.4	0.8	0.196	0.134	0.191	0.129	0.179	0.094	0.115	0.095	0.180	0.091	0.106	0.092
	0.4	0.198	0.131	0.186	0.130	0.196	0.089	0.126	0.090	0.165	0.079	0.097	0.077
	0	0.175	0.114	0.161	0.106	0.188	0.102	0.126	0.103	0.161	0.083	0.093	0.088
	−0.4	0.170	0.105	0.169	0.102	0.181	0.099	0.122	0.097	0.178	0.077	0.096	0.078
	−0.8	0.194	0.117	0.180	0.115	0.192	0.103	0.119	0.100	0.183	0.077	0.087	0.069
0	0.8	0.096	0.151	0.238	0.125	0.074	0.085	0.142	0.079	0.070	0.085	0.104	0.079
	0.4	0.083	0.126	0.184	0.106	0.079	0.092	0.138	0.083	0.063	0.072	0.094	0.068
	0	0.075	0.104	0.165	0.089	0.070	0.087	0.122	0.071	0.062	0.072	0.092	0.066
	−0.4	0.059	0.101	0.170	0.079	0.069	0.089	0.128	0.083	0.056	0.070	0.104	0.062
	−0.8	0.091	0.143	0.200	0.115	0.074	0.093	0.137	0.084	0.059	0.074	0.100	0.066
−0.4	0.8	0.007	0.118	0.247	0.121	0.001	0.054	0.141	0.067	0.001	0.049	0.092	0.056
	0.4	0.004	0.095	0.191	0.103	0.004	0.056	0.139	0.070	0.001	0.036	0.079	0.040
	0	0.002	0.065	0.156	0.070	0.000	0.040	0.112	0.052	0.000	0.037	0.079	0.048
	−0.4	0.002	0.086	0.207	0.086	0.000	0.042	0.116	0.063	0.000	0.030	0.087	0.045
	−0.8	0.009	0.101	0.227	0.102	0.000	0.051	0.122	0.067	0.001	0.040	0.084	0.051
−0.8	0.8	0.001	0.056	0.174	0.088	0.000	0.011	0.085	0.035	0.000	0.004	0.043	0.017
	0.4	0.001	0.049	0.147	0.069	0.000	0.010	0.089	0.036	0.000	0.002	0.038	0.012
	0	0.000	0.036	0.138	0.057	0.000	0.004	0.063	0.020	0.000	0.004	0.030	0.012
	−0.4	0.001	0.050	0.170	0.078	0.000	0.006	0.081	0.028	0.000	0.002	0.041	0.009
	−0.8	0.000	0.064	0.194	0.104	0.000	0.011	0.081	0.037	0.000	0.005	0.047	0.022

serial correlation when constructing $\widehat{\omega}_{11\cdot 2}$. (v) The size of the test with $l12$ tends to be larger than the nominal size. This generally results from using too many lags to construct $\widehat{\omega}_{11\cdot 2}$. (vi) The size of the test with $l4$ and l_A is close to the nominal size for moderate values of the MA parameter ϕ_1. Hence we recommend that applied researchers use the lag truncation choice either l_4 or l_A rather than $l0$ or $l12$.

Next we turn to power properties against several nonlinear alternatives. The nonlinear alternatives considered are as follows:

$$DGP_2: \; y_t = 5(\Psi(x_t) - 0.5) + u_{1t}, \; DGP_3: \; y_t = 1/|x_t|^{1/3} + u_{1t}, \; DGP_4: \; y_t = y_{t-1} + u_{1t},$$

where $\Psi(\cdot)$ is the cumulative distribution function of a normal random variable with mean zero and variance 6. Other assumptions on x_t and u_t are same as in DGP_1. DGP_2 satisfy the assumption in Theorem 4.1 (a) and DGP_3 does likewise for Theorem 4.1 (b). DGP_4 represents the case of no cointegration in Theorem 4.2.

Tables 2-4 show the size-corrected power properties at a 5% nominal level for DGP_2-DGP_4 respectively. We summarize the general results first and discuss some specific alternatives later. (i) The power of the test becomes better as the sample size becomes larger for all alternatives. (ii) As expected, a nonzero correlation $\sigma_{e,12}$ between e_{1t} and e_{2t} increases the

TABLE 2. SIZE-ADJUSTED POWER OF THE TEST, $g(x)=5(\Psi(x)-0.5)$

ϕ_1	$\sigma_{e,12}$	$T=100$				$T=200$				T=400			
		$l0$	$l4$	$l12$	l_A	$l0$	$l4$	$l12$	l_A	$l0$	$l4$	$l12$	l_A
0.8	0.8	0.644	0.625	0.575	0.628	0.927	0.922	0.912	0.915	0.998	0.998	0.999	0.998
	0.4	0.409	0.403	0.405	0.415	0.842	0.844	0.821	0.845	0.990	0.988	0.989	0.991
	0	0.384	0.383	0.373	0.389	0.812	0.798	0.791	0.795	0.971	0.970	0.968	0.970
	−0.4	0.459	0.437	0.412	0.429	0.821	0.830	0.793	0.808	0.990	0.989	0.988	0.989
	−0.8	0.656	0.622	0.580	0.628	0.926	0.921	0.916	0.922	1.000	1.000	0.999	1.000
0.4	0.8	0.769	0.752	0.713	0.752	0.964	0.961	0.951	0.958	1.000	1.000	1.000	1.000
	0.4	0.539	0.521	0.517	0.532	0.898	0.893	0.892	0.894	0.995	0.996	0.997	0.996
	0	0.530	0.516	0.482	0.510	0.884	0.874	0.853	0.870	0.989	0.988	0.989	0.988
	−0.4	0.577	0.561	0.530	0.556	0.892	0.884	0.874	0.877	0.997	0.996	0.996	0.996
	−0.8	0.743	0.715	0.663	0.720	0.954	0.953	0.951	0.953	1.000	1.000	1.000	1.000
0	0.8	0.819	0.813	0.785	0.815	0.984	0.981	0.980	0.983	1.000	1.000	1.000	1.000
	0.4	0.695	0.696	0.661	0.694	0.942	0.943	0.942	0.941	0.998	0.998	0.998	0.998
	0	0.657	0.653	0.622	0.673	0.933	0.931	0.923	0.933	0.996	0.996	0.995	0.996
	−0.4	0.713	0.692	0.664	0.703	0.939	0.934	0.928	0.935	0.999	0.999	0.999	0.999
	−0.8	0.819	0.812	0.780	0.817	0.979	0.976	0.974	0.979	1.000	1.000	1.000	1.000
−0.4	0.8	0.891	0.880	0.870	0.881	0.994	0.995	0.992	0.994	1.000	1.000	1.000	1.000
	0.4	0.809	0.802	0.789	0.800	0.975	0.974	0.971	0.974	0.999	0.999	0.999	0.999
	0	0.792	0.796	0.791	0.796	0.967	0.966	0.962	0.965	1.000	1.000	1.000	1.000
	−0.4	0.789	0.788	0.762	0.784	0.976	0.975	0.971	0.973	1.000	1.000	1.000	1.000
	−0.8	0.886	0.879	0.867	0.872	0.995	0.995	0.993	0.995	1.000	1.000	1.000	1.000
−0.8	0.8	0.917	0.923	0.919	0.923	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	0.4	0.857	0.863	0.880	0.879	0.991	0.991	0.991	0.991	1.000	1.000	1.000	1.000
	0	0.853	0.859	0.864	0.863	0.989	0.990	0.992	0.993	1.000	1.000	1.000	1.000
	−0.4	0.836	0.839	0.847	0.838	0.989	0.990	0.989	0.990	1.000	1.000	1.000	1.000
	−0.8	0.916	0.916	0.918	0.916	0.998	0.998	0.998	0.998	1.000	1.000	1.000	1.000

power of the test for all alternatives. (iii) The power of the test for positive ϕ_1 tends to be less powerful than the test with $\phi_1=0$, and that for negative ϕ_1 tends to be more powerful especially when $T=100$. (iv) The power of the test against nonlinear alternatives that satisfy the assumption in Theorem 4.1 (a) increases very quickly as the sample size grows, on the other hand when the nonlinear alternatives that satisfy the assumption in Theorem 4.1 (b), the power increases very slowly as Theorem 4.1 (b) predicts. (v) The power of the test with $l4$ is as powerful as that with l_A for all nonlinear alternatives. (vi) The power of the test for DGP_3 is more sensitive to the choice of l than that for DGP_2.

The second thing to note is that the test with l_A against the alternative of no cointegration suffers from a lack of power. This is because the truncated version of the lag length choice l_A tends to choose a longer lag length since $\hat{\rho}$ in the formula of l_A (9) is close to 1. As we see in Theorem 4.2, there exists a tradeoff between lag length l and power, so the test with l_A performs poorly against the alternative.

VI. *Concluding Remarks*

This paper has developed a testing procedure for linearity in regressions with I (1)

TABLE 3. SIZE-ADJUSTED POWER OF THE TEST, $g(x)=1/|x|^{1/3}$

ϕ_1	$\sigma_{e,12}$	$T=100$				$T=200$				T=400			
		$l0$	$l4$	$l12$	l_A	$l0$	$l4$	$l12$	l_A	$l0$	$l4$	$l12$	l_A
0.8	0.8	0.415	0.258	0.147	0.240	0.606	0.467	0.262	0.358	0.783	0.681	0.481	0.574
	0.4	0.216	0.150	0.130	0.154	0.377	0.323	0.199	0.289	0.566	0.486	0.365	0.434
	0	0.230	0.179	0.135	0.169	0.351	0.299	0.204	0.247	0.514	0.430	0.321	0.379
	−0.4	0.270	0.195	0.135	0.181	0.380	0.320	0.190	0.247	0.601	0.511	0.385	0.444
	−0.8	0.454	0.290	0.156	0.272	0.637	0.515	0.299	0.415	0.819	0.718	0.494	0.610
0.4	0.8	0.596	0.389	0.210	0.370	0.756	0.613	0.327	0.503	0.891	0.776	0.556	0.698
	0.4	0.331	0.219	0.150	0.219	0.533	0.440	0.274	0.392	0.741	0.641	0.485	0.585
	0	0.330	0.249	0.156	0.243	0.513	0.391	0.238	0.357	0.670	0.578	0.405	0.521
	−0.4	0.417	0.264	0.170	0.270	0.528	0.425	0.252	0.354	0.769	0.634	0.447	0.572
	−0.8	0.614	0.381	0.181	0.371	0.791	0.666	0.408	0.560	0.918	0.814	0.572	0.730
0	0.8	0.738	0.443	0.249	0.528	0.889	0.691	0.375	0.715	0.963	0.841	0.620	0.855
	0.4	0.549	0.331	0.166	0.410	0.744	0.582	0.351	0.625	0.900	0.801	0.573	0.838
	0	0.542	0.357	0.191	0.460	0.720	0.568	0.312	0.634	0.855	0.741	0.508	0.796
	−0.4	0.621	0.377	0.222	0.482	0.775	0.574	0.321	0.648	0.902	0.783	0.591	0.836
	−0.8	0.743	0.457	0.240	0.521	0.902	0.740	0.467	0.753	0.976	0.880	0.637	0.888
−0.4	0.8	0.900	0.564	0.311	0.620	0.977	0.815	0.463	0.862	0.997	0.936	0.691	0.965
	0.4	0.787	0.475	0.264	0.517	0.932	0.768	0.451	0.779	0.988	0.928	0.710	0.933
	0	0.811	0.552	0.305	0.594	0.925	0.779	0.451	0.788	0.979	0.904	0.647	0.907
	−0.4	0.828	0.512	0.252	0.573	0.949	0.800	0.437	0.799	0.986	0.925	0.715	0.934
	−0.8	0.888	0.549	0.299	0.625	0.977	0.835	0.532	0.886	0.998	0.947	0.695	0.975
−0.8	0.8	0.945	0.694	0.374	0.694	0.996	0.936	0.555	0.901	1.000	0.995	0.811	0.987
	0.4	0.898	0.636	0.336	0.608	0.992	0.926	0.549	0.839	1.000	0.991	0.807	0.971
	0	0.922	0.692	0.342	0.634	0.994	0.928	0.615	0.859	0.999	0.996	0.824	0.977
	−0.4	0.918	0.650	0.346	0.586	0.995	0.929	0.566	0.854	1.000	0.994	0.798	0.972
	−0.8	0.953	0.655	0.351	0.689	0.998	0.946	0.597	0.918	1.000	0.993	0.785	0.989

processes. We proposed the Wald test based on a generalization of the RESET test and we showed that the limiting distribution of the test statistic under the null of linearity is a χ^2 distribution when a "leads and lags" estimation technique is employed to construct it. We also showed that the test is consistent against both a class of nonlinear alternatives and no cointegration. The simulation experiment revealed that the proposed test has nice power properties against the functions considered. Finally, we applied our test for linearity to see whether relationships between exchange rates and fundamentals and found significant evidence against linearity.

APPENDIX A

Proof of Lemma 2.2: Proofs of (a) and (b) are trivial extensions of the results in Phillips (1987). Proof of (c) can be found in Hansen (1992). Part (c) is an extension of part (d) and its proof follows the argument of Durlauf (1986).

Proof of (e): This is a version of Theorem 4.2 of Hansen (1992). The case for $i=2$ is given by Theorem 4.2 of Hansen (1992). Thus we show the case for $i=3$. Cases where $i\geq 4$ can be proved by the same argument as in the case for $i=3$. Let $F_t=\sigma(u_s : s\leq t)$ be the smallest $\sigma-$field containing the past of $\{u_t\}$

TABLE 4. SIZE-ADJUSTED POWER OF THE TEST, NO COINTEGRATION

ϕ_1	$\sigma_{e,12}$	T=100				T=200				T=400			
		$l0$	$l4$	$l12$	l_A	$l0$	$l4$	$l12$	l_A	$l0$	$l4$	$l12$	l_A
0.8	0.8	0.795	0.502	0.250	0.302	0.882	0.669	0.377	0.314	0.940	0.801	0.545	0.402
	0.4	0.766	0.513	0.289	0.313	0.846	0.690	0.422	0.362	0.925	0.792	0.557	0.395
	0	0.769	0.547	0.317	0.343	0.862	0.684	0.406	0.319	0.923	0.796	0.558	0.403
	−0.4	0.802	0.554	0.318	0.347	0.856	0.675	0.383	0.312	0.927	0.800	0.580	0.432
	−0.8	0.789	0.530	0.263	0.313	0.866	0.692	0.415	0.336	0.926	0.799	0.568	0.419
0.4	0.8	0.819	0.540	0.267	0.328	0.903	0.688	0.379	0.340	0.948	0.802	0.547	0.410
	0.4	0.787	0.516	0.287	0.326	0.859	0.690	0.436	0.365	0.936	0.802	0.576	0.409
	0	0.794	0.551	0.314	0.346	0.886	0.697	0.405	0.341	0.932	0.796	0.558	0.414
	−0.4	0.826	0.554	0.316	0.358	0.871	0.677	0.384	0.324	0.940	0.801	0.571	0.421
	−0.8	0.828	0.538	0.262	0.329	0.892	0.721	0.446	0.367	0.937	0.802	0.568	0.436
0	0.8	0.860	0.506	0.246	0.338	0.930	0.679	0.364	0.366	0.962	0.808	0.538	0.439
	0.4	0.863	0.532	0.283	0.365	0.913	0.698	0.436	0.376	0.961	0.811	0.567	0.452
	0	0.868	0.567	0.312	0.403	0.926	0.717	0.410	0.376	0.958	0.807	0.558	0.456
	−0.4	0.886	0.558	0.316	0.382	0.924	0.693	0.404	0.377	0.962	0.806	0.577	0.465
	−0.8	0.871	0.534	0.259	0.343	0.929	0.719	0.438	0.395	0.958	0.804	0.571	0.461
−0.4	0.8	0.924	0.552	0.252	0.367	0.975	0.755	0.412	0.423	0.990	0.857	0.571	0.486
	0.4	0.921	0.556	0.301	0.385	0.971	0.734	0.441	0.422	0.990	0.845	0.605	0.492
	0	0.944	0.611	0.327	0.422	0.981	0.750	0.432	0.435	0.989	0.840	0.580	0.495
	−0.4	0.930	0.568	0.281	0.394	0.964	0.751	0.421	0.431	0.987	0.848	0.599	0.508
	−0.8	0.923	0.563	0.245	0.360	0.981	0.758	0.449	0.447	0.989	0.829	0.584	0.481
−0.8	0.8	0.903	0.572	0.266	0.519	0.984	0.823	0.458	0.683	0.995	0.925	0.668	0.785
	0.4	0.913	0.563	0.311	0.510	0.974	0.801	0.472	0.679	0.996	0.907	0.665	0.767
	0	0.930	0.590	0.284	0.542	0.986	0.818	0.487	0.681	0.997	0.919	0.693	0.766
	−0.4	0.910	0.559	0.273	0.488	0.970	0.821	0.455	0.652	0.995	0.917	0.647	0.762
	−0.8	0.904	0.562	0.270	0.526	0.981	0.824	0.471	0.675	0.993	0.892	0.662	0.744

for all t. We can decompose u_{1t} into two parts

$$u_{1t}=\varepsilon_t+z_{t-1}-z_t \tag{10}$$

where $\varepsilon_t=\sum_{k=0}^{\infty}(\mathrm{E}_t u_{1,t+k}-\mathrm{E}_{t-1}u_{1,t+k})$ and $z_t=\sum_{k=1}^{\infty}\mathrm{E}_t u_{1,t+k}$ where $\mathrm{E}_t(\cdot)=\mathbf{E}(\cdot|F_t)$. Note that $\{\varepsilon_t,F_t\}$ is a martingale difference sequence. By the decomposition (10) we have

$$T^{-2}\sum_{t=1}^{T}x_t^{(3)}u_{1,t+1}=T^{-2}\sum_{t=1}^{T}x_t^{(3)}\varepsilon_{1,t+1}+T^{-2}\sum_{t=1}^{T}x_t^{(3)}(z_t-z_{t+1}). \tag{11}$$

Applying Theorem 3.1 of Hansen (1992) gives

$$T^{-2}\sum_{t=1}^{T}x_t^{(3)}\varepsilon_{1,t+1}\Rightarrow\int B_2^{(3)}dB_1. \tag{12}$$

Note that no second order bias terms show up in the limit.

It remains to be shown that

$$T^{-2}\sum_{t=1}^{T}x_t^{(3)}(z_t-z_{t+1})\Rightarrow 3D(B_2^{(2)})\lambda_{21}. \tag{13}$$

Observe that

$$T^{-2}\sum_{t=1}^{T}x_t^{(3)}(z_t-z_{t+1})=T^{-2}\sum_{t=1}^{T}(x_t^{(3)}-x_{t-1}^{(3)})z_t-T^{-2}x_T^{(3)}z_{T+1}.$$

By the argument of Theorem 4.1 in Hansen (1992), we have $T^{-2}x_T^{(3)}z_{T+1}=o_p(1)$. Since a typical element of $x_t^{(3)}-x_{t-1}^{(3)}$ can be written as

$$x_{it}^3-x_{i,t-1}^3=(x_{i,t-1}+u_{2i,t})^3-x_{i,t-1}^3=3x_{i,t-1}^2u_{2i,t}+3x_{i,t-1}u_{2i,t}^2+u_{2i,t}^3,$$

we have

$$T^{-2}\sum_{t=1}^{T}x_t^{(3)}(z_t-z_{t+1})=3T^{-2}\sum_{t=1}^{T}(x_{t-1}^{(2)}\odot u_{2t})z_t+3T^{-2}\sum_{t=1}^{T}(x_{t-1}\odot u_{2t}^{(2)})z_t+T^{-2}\sum_{t=1}^{T}u_{2t}^{(3)}z_t+o_p(1),$$

where "$\odot$" is the element-by-element product and $u_{2t}^{(k)}=(u_{21,t}^k,\ldots,u_{2m,t}^k)'$. Note that

$$(x_{t-1}^{(2)}\odot u_{2t})z_t=\begin{bmatrix}x_{1,t-1}^2u_{21,t}\\ x_{2,t-1}^2u_{22,t}\\ \vdots\\ x_{m,t-1}^2u_{2m,t}\end{bmatrix}z_t=\tilde{D}(x_{t-1}^{(2)})u_{2t}z_t,$$

where $\tilde{D}(x_{t-1}^{(2)})=\mathrm{diag}[x_{1,t-1}^2,x_{2,t-1}^2,\ldots,x_{m,t-1}^2]$. Then we have

$$\begin{aligned}T^{-2}\sum_{t=1}^{T}x_t^{(3)}(z_t-z_{t+1})=3T^{-2}\sum_{t=1}^{T}\tilde{D}(x_{t-1}^{(2)})u_{2t}z_t+3T^{-2}\sum_{t=1}^{T}(x_{t-1}\odot u_{2t}^{(2)})z_t\\ +T^{-2}\sum_{t=1}^{T}u_{2t}^{(3)}z_t+o_p(1)\end{aligned}\tag{14}$$

First, the second and third terms on the right-hand side of (14) vanish in probability because we have by the Hölder's inequality that

$$\begin{aligned}\mathbf{E}\left|T^{-2}\sum_{t=1}^{T}(x_{t-1}\odot u_{2t}^{(2)})z_t\right|&\le T^{-3/2}\sum_{t=1}^{T}\left\|\frac{x_{t-1}}{\sqrt{T}}\odot u_{2t}^{(2)}\right\|_{4/3}\|z_t\|_4\\ &\le T^{-3/2}\sum T\left\|\frac{x_{t-1}}{\sqrt{T}}\right\|_4\|u_{2t}\|_4^2\|z_t\|_4\to 0\end{aligned}\tag{15}$$

and

$$\mathbf{E}\left|T^{-2}\sum_{t=1}^{T}u_{2t}^{(3)}z_t\right|\le T^{-2}\sum_{t=1}^{T}\|u_{2t}^{(3)}\|_{4/3}\|z_t\|_4=T^{-2}\sum_{t=1}^{T}\|u_{2t}\|_4^3\|z_t\|_4\to 0\tag{16}$$

since $\|x_{t-1}/\sqrt{T}\|_4$ is bounded as in the proof of Lemma 3.1 (f) of Chang et al. (2001), $\|u_{2t}\|_4^3<C$ by Assumption 1 and $\|z_t\|_4$ is uniformly bounded by the proof of Theorem 3.1 in Hansen (1992) where $\|a\|_r$ denotes the L^r-norm with subscript, defined by $\|a\|_r=(\sum_i\mathbf{E}|a_i|^r)^{1/r}$.

Second, note that the first term on the right-hand side of (14) can be written as

$$T^{-2}\sum_{t=1}^{T}\tilde{D}(x_{t-1}^{(2)})u_{2t}z_t=T^{-2}\sum_{t=1}^{T}\tilde{D}(x_{t-1}^{(2)})\lambda_{21}+T^{-2}\sum_{t=1}^{T}\tilde{D}(x_{t-1}^{(2)})(u_{2t}z_t-\lambda_{21}).$$

By Theorem 3.2 of Hansen (1992), the sequence $\{u_{2t}z_t-\lambda_{21}\}$ is an $L_{\beta/2}$-mixingale and $\tilde{D}(x_{t-1}^{(2)})=O_p(T)$.

Then, applying Theorem 3.3 of Hansen (1992), $|T^{-2}\sum_{t=1}^{T}\tilde{D}(x_{t-1}^{(2)})(u_{2t}z_t-\Lambda_{21})|\xrightarrow{p}0$. By the continuous mapping theorem we obtain

$$T^{-2}\sum_{t=1}^{T}\tilde{D}(x_{t-1})\Lambda_{21}\Rightarrow D(B_2)\Lambda_{21} \tag{17}$$

Thus combining (14)-(17) gives (13).

(11), (12), and (13) together establishes the result of Lemma 2.2 (e) when $i=3$. The proof for the case where $i\geq 4$ follows along the same line with appropriate moment conditions specified in Lemma 2.2. The proof for part (f) can be shown by combining that of part (e) and the argument used in Phillips (1988). □

Proof of Theorem 3.1: Observe that

$$\Upsilon_T\begin{pmatrix}\hat{\gamma}_0-\gamma_0\\ \hat{\gamma}_1-\gamma_1\\ \hat{\gamma}_2\\ \vdots\\ \hat{\gamma}_\kappa\end{pmatrix}\Rightarrow\Upsilon_T\begin{pmatrix}T & \sum_{t=1}^{T}x_t' & \sum_{t=1}^{T}x_t^{(2)\prime} & \cdots & \sum_{t=1}^{T}x_t^{(\kappa)\prime}\\ \sum_{t=1}^{T}x_t & \sum_{t=1}^{T}x_tx_t' & \sum_{t=1}^{T}x_tx_t^{(2)\prime} & \cdots & \sum_{t=1}^{T}x_tx_t^{(\kappa)\prime}\\ \sum_{t=1}^{T}x_t^{(2)} & \sum_{t=1}^{T}x_t^{(2)}x_t' & \sum_{t=1}^{T}x_t^{(2)}x_t^{(2)\prime} & \cdots & \sum_{t=1}^{T}x_t^{(2)}x_t^{(\kappa)\prime}\\ \vdots & \vdots & \vdots & \ddots & \\ \sum_{t=1}^{T}x_t^{(\kappa)} & \sum_{t=1}^{T}x_t^{(\kappa)}x_t' & \sum_{t=1}^{T}x_t^{(\kappa)}x_t^{(2)\prime} & \cdots & \sum_{t=1}^{T}x_t^{(\kappa)}x_t^{(\kappa)\prime}\end{pmatrix}^{-1}\Upsilon_T$$

$$\times\Upsilon_T^{-1}\begin{pmatrix}\sum_{t=1}^{T}u_{1t}\\ \sum_{t=1}^{T}x_tu_{1t}\\ \sum_{t=1}^{T}x_t^{(2)}u_{1t}\\ \vdots\\ \sum_{t=1}^{T}x_t^{(\kappa)}u_{1t}\end{pmatrix}.$$

Applying Lemma 2.2 to each element on the right hand side of the equation gives the required result. □

Proof of Theorem 3.2: Although we now have stationary regressors in (4), we may concentrate on $\tilde{\underline{\gamma}}$ without loss of generality because they are asymptotically orthogonal to the integrated parts of the model, $\frac{1}{T^{(i+2)/2}}\sum_{t=1}^{T}x_{jt}^{i}\Delta x_{k,t-s}\xrightarrow{p}0$, for $1\leq i\leq\kappa$, $1\leq j,k\leq m$, and any s by Lemma 3.1 of Chang et al. (2001). This asymptotic orthogonality of nonlinear transformations of integrated regressors to stationary regressors are analogous to that of (untransformed) integrated regressors to stationary regressors in linear cointegration models as noted in Chang et al. (2001).

Note that

$$\frac{1}{T^{1/2}}\sum_{j}^{[Tr]} w_j \Rightarrow \begin{bmatrix} B_{11\cdot 2}(r) \\ B_2(r) \end{bmatrix}$$

with covariance matrix

$$\begin{bmatrix} \omega_{11}-\omega_{21}'\Omega_{22}^{-1}\omega_{21} & 0 \\ 0 & \Omega_{22} \end{bmatrix}$$

Then by Lemma 2.2 (e) we have $T^{-(i+1)/2}\sum_{t=1}^{T} x_t^{(i)} v_t \Rightarrow \int B_2^{(i)} dB_{11\cdot 2}$ for $2 \le i \le \kappa$. Since similar arguments are used as in the proof of Theorem 4.1 in Saikkonen (1991), it suffices to show that for $2 \le i \le \kappa$ and $1 \le j \le m$

$$\frac{1}{T^{(i+1)/2}}\sum_{t=1}^{T} x_{jt}^i v_t^* = \frac{1}{T^{(i+1)/2}}\sum_{t=1}^{T} x_{jt}^i v_t + o_p(1). \tag{18}$$

Note that the number of observations now is $T-2K$, but we may use T instead of $T-2K$ without loss of generality. Note that

$$\left\| \frac{1}{T^{i/2}}\sum_{t=1}^{T} x_{jt}^i (v_t^* - v_t) \right\| \le \left\| \left(\frac{x_{jt}}{T^{1/2}}\right)^i \right\|_{E_j} \sum_{t=1}^{T} \| v_t^* - v_t \|$$

where

$$E_j = [\min_{0\le r\le 1} B_{2j}(r) - 1, \max_{0\le r\le 1} B_{2j}(r) + 1]. \tag{19}$$

It can be shown by the argument in Lemma A1 of Chang et al. (2001) that

$$\mathbf{E}(\|v_t^* - v_t\|^r) = O(K^{-rs/2}). \tag{20}$$

Then by (20) and the fact that $\|(x_{jt}/T^{1/2})^i\|_{E_i} = O_p(1)$ we can deduce $\frac{1}{T^{i/2}}\sum_{t=1}^{T} x_{jt}^i (v_t^* - v_t) = O_p(TK^{-s/2})$ as shown in Lemma A4 (b) of Chang et al. (2001), leading to $\frac{1}{T^{(i+1)/2}}\sum_{t=1}^{T} x_{jt}^i v_t^* = \frac{1}{T^{(i+1)/2}}\sum_{t=1}^{T} x_{jt}^i v_t + O_p(T^{1/2}K^{-s/2})$. Thus (18) follows if $\delta > 1/s$ since $K = T^{\delta}$. □

Proof of Theorem 3.3: Given the result of Theorem 3.2, applying Lemma 5.1 in Park and Phillips (1988) gives the required result. □

Proof of Lemma 4.1: In this proof and the subsequent proof, we frequently use results from Park and Phillips (2001) and Chang et al. (2001). In those citations, the space $D[0,1]$ is endowed with the uniform metric. However, we use "⇒" to imply weak convergence using the Skorohod metric in our proofs. This is possible because convergence in the uniform metric implies the convergence in the Skorohod metric.

Proof of (a): Note that we have stationary regressors in (4). Again we may concentrate on $\tilde{\underline{\gamma}}$ without loss of generality by the same reasoning as described in the proof of Theorem 3.2. Observe that

$$\Upsilon_T \tilde{\underline{\gamma}} = \Upsilon_T M^{-1} \Upsilon_T \times \Upsilon_T^{-1} \sum_{t=1}^{T} X_t y_t. \tag{21}$$

First, as we have seen in the proof of Theorem 3.1

$$\Upsilon_T M^{-1} \Upsilon_T = O_p(1). \tag{22}$$

Next we consider the second component of the right hand side of (21). Observe that under the alternative (6) we have

$$\Upsilon_T^{-1} \sum_{t=1}^{T} X_t y_t = \Upsilon_T^{-1} \sum_{t=1}^{T} g(x_t) X_t + \Upsilon_T^{-1} \sum_{t=1}^{T} X_t u_{1t} \tag{23}$$

where $g(x_t) \equiv \sum_{i=1}^{m} g_i(x_{it})$. The second term on the right hand side of (23) is $O_p(1)$ as we saw in the proof of Theorem 3.1. To analyze the first term on the right hand side of (23), we consider the asymptotic properties of $g_i \in \mathcal{T}(I)$ and $g_i \in \mathcal{T}(H)$ separately. For a function $g_i \in \mathcal{T}(I)$, it follows from Part (k) of Lemma 3.1 of Chang et al. (2001), for $0 \leq s \leq \kappa$ and $1 \leq j,k \leq m$, $\frac{1}{T^{(s+1)/2}} \sum_{t=1}^{T} g_i(x_{jt}) x_{kt}^s = O_p(1)$. Hence we have for a function $g_i \in \mathcal{T}(I)$

$$\Upsilon_T^{-1} \sum_{t=1}^{T} g_i(x_{it}) X_t = O_p(1). \tag{24}$$

It follows that if $g_i \in \mathcal{T}(I)$ for all i

$$\Upsilon_T^{-1} \sum_{t=1}^{T} X_t y_t = O_p(1). \tag{25}$$

Thus if $g_i \in \mathcal{T}(I)$ for all i, we can deduce by (21), (22) and (25) that $\Upsilon_T^{-1} \tilde{\underline{\gamma}} = O_p(1)$, giving one case of the result required for Part (ii) of Lemma 4.1 (a).

For a function $g_i \in \mathcal{T}(H)$ with asymptotic order ν_i and limit homogeneous function h_i, we have, from Part (l) of Lemma 3.1 in Chang et al. (2001) and Theorem 1 of de Jong (2004), for $0 \leq s \leq \kappa$ and $1 \leq j,k \leq m$, $\frac{1}{T^{(s/2+1)} \nu_i(T^{1/2})} \sum_{t=1}^{T} g_i(x_{jt}) x_{kt}^s = O_p(1)$. Thus we have for a function $g_i \in \mathcal{T}(H)$

$$\Upsilon_T^{-1} \sum_{t=1}^{T} g_i(x_{it}) X_t = O_p(T^{1/2} \nu_i(T^{1/2})). \tag{26}$$

Note that (26) holds for any $\nu_i(\cdot)$ that satisfies the assumptions of Lemma 4.1.

Now we consider a case where $g_i \in \mathcal{T}(H)$ for some i under the alternative (6). Remember that g_1 is the dominating function among functions belonging to $\mathcal{T}(H)$. Then it is clear from the argument above that the order of the dominating component in the first term of (23) is given by

$$\Upsilon_T^{-1} \sum_{t=1}^{T} g_1(x_{1t}) X_t = \begin{cases} O_p(T^{1/2} \nu_1(T^{1/2})) & \text{if } T^{1/2} \nu_1(T^{1/2}) \to \infty \text{ as } T \to \infty, \\ O_p(1) & \text{otherwise.} \end{cases}$$

Thus we get

$$\Upsilon_T^{-1} \sum_{t=1}^{T} g(x_t) X_t = \begin{cases} O_p(T^{1/2} \nu_1(T^{1/2})) & \text{if } T^{1/2} \nu_1(T^{1/2}) \to \infty \text{ as } T \to \infty, \\ O_p(1) & \text{otherwise,} \end{cases}$$

leading to

$$\Upsilon_T^{-1}\sum_{t=1}^{T}X_t y_t=\begin{cases}O_p(T^{1/2}\nu_1(T^{1/2})) & \text{if } T^{1/2}\nu_1(T^{1/2})\to\infty \text{ as } T\to\infty,\\ O_p(1) & \text{otherwise.}\end{cases} \tag{27}$$

Hence combining (21), (22) and (27) shows that

$$\Upsilon_T\tilde{\gamma}=\begin{cases}O_p(T^{1/2}\nu_1(T^{1/2})) & \text{if } T^{1/2}\nu_1(T^{1/2})\to\infty \text{ as } T\to\infty,\\ O_p(1) & \text{otherwise,}\end{cases}$$

giving the result required for Part (i) and the other case of Part (ii) in Lemma 4.1 (a).
Proof of (b): Recall that

$$\hat{\omega}_{11\cdot 2}=T^{-1}\sum_{t=1}^{T}\hat{v}_t^{*2}+2T^{-1}\sum_{s=1}^{l}w_{sl}\sum_{t=s+1}^{T}\hat{v}_t^*\hat{v}_{t-s}^* \tag{28}$$

where $\hat{v}_t^*$ is the residual obtained from the regression (4) and $w_{sl}=1-s/(l+1)$. First, consider the first term in (28). Let $\underline{\theta}=(\theta'_K,\theta'_{K-1},\theta'_{K-2},\ldots,\theta'_{-K+2},\theta'_{-K+1},\theta_{-K})'$ and

$$Z_t=(\Delta x'_{t+K},\Delta x'_{t+K-1},\Delta x'_{t+K-2},\ldots,\Delta x'_{t-K+2},\Delta x'_{t-K+1},\Delta x'_{t-K})'.$$

Also let $\tilde{\underline{\theta}}$ be the OLS estimator of $\underline{\theta}$ from (4). Observe that

$$\begin{aligned}T^{-1}\sum_{t=1}^{T}\hat{v}_t^{*2}&=T^{-1}\sum_{t=1}^{T}(y_t-X_t'\tilde{\underline{\gamma}}-Z_t'\tilde{\underline{\theta}})^2=T^{-1}\sum_{t=1}^{T}(g(x_t)+u_{1t}-X_t'\tilde{\underline{\gamma}}-Z_t'\tilde{\underline{\theta}})^2\\&=T^{-1}\sum_{t=1}^{T}g^2(x_t)+T^{-1}\sum_{t=1}^{T}u_{1t}^2+T^{-1}\tilde{\underline{\gamma}}'\left(\sum_{t=1}^{T}X_tX_t'\right)\tilde{\underline{\gamma}}+T^{-1}\tilde{\underline{\theta}}'\left(\sum_{t=1}^{T}Z_tZ_t'\right)\tilde{\underline{\theta}}\\&\quad+2\Bigg\{T^{-1}\sum_{t=1}^{T}g(x_t)u_{1t}-T^{-1}\tilde{\underline{\gamma}}'\sum_{t=1}^{T}g(x_t)X_t-T^{-1}\tilde{\underline{\theta}}'\sum_{t=1}^{T}g(x_t)Z_t\\&\quad-T^{-1}\tilde{\underline{\gamma}}'\sum_{t=1}^{T}X_tu_{1t}-T^{-1}\tilde{\underline{\theta}}'\sum_{t=1}^{T}Z_tu_{1t}+T^{-1}\tilde{\underline{\gamma}}'(\sum_{t=1}^{T}X_tZ_t')\tilde{\underline{\theta}}\Bigg\}\end{aligned} \tag{29}$$

First, we deal with the case where $g_i\in\mathcal{T}(H)$ with $\nu_i(T^{1/2})\to\infty$ as $T\to\infty$ for some i. Again remember that g_1 is the dominating function among functions that belong to $\mathcal{T}(H)$ We check the order of convergence of each term in (29). $T^{-1}\sum_{t=1}^{T}g^2(x_t)=O_p(\nu_1^2(T^{1/2}))$ by the argument in Part (a) of Lemma 4.1, $T^{-1}\sum_{t=1}^{T}u_{1t}^2=O_p(1)$ by the law of large numbers and $T^{-1}\sum_{t=1}^{T}g(x_t)u_{1t}=o_p(\nu(T^{1/2}))$ by Lemma 3.1 (f) of Chang et al. (2001) and Theorem 1 of de Jong (2004). It follows by the proof of Theorem 3.1 and Part (a) of Lemma 4.1 that

$$\begin{aligned}T^{-1}\tilde{\underline{\gamma}}'\left(\sum_{t=1}^{T}X_tX_t'\right)\tilde{\underline{\gamma}}&=(T^{-1/2}\Upsilon_T\tilde{\underline{\gamma}})'\left(\Upsilon_T^{-1}\sum_{t=1}^{T}X_tX_t'\Upsilon_T^{-1}\right)(T^{-1/2}\Upsilon_T\tilde{\underline{\gamma}})\\&=O_p(\nu_1(T^{1/2}))O_p(1)O_p(\nu_1(T^{1/2}))=O_p(\nu_1^2(T^{1/2})),\\T^{-1}\tilde{\underline{\gamma}}'\sum_{t=1}^{T}g(x_t)X_t&=(T^{-1/2}\Upsilon_T\tilde{\underline{\gamma}})'T^{-1/2}\sum_{t=1}^{T}g(x_t)\Upsilon_T^{-1}X_t\\&=O_p(\nu_1(T^{1/2}))O_p(\nu_1(T^{1/2}))=O_p(\nu_1^2(T^{1/2}))\end{aligned}$$

and

$$T^{-1}\underline{\tilde{\gamma}}'\sum_{t=1}^{T}X_t u_{1t}=(T^{-1/2}\Upsilon_T\underline{\tilde{\gamma}})'T^{-1/2}\sum_{t=1}^{T}\Upsilon_T^{-1}X_t u_{1t}$$
$$=O_p(\nu_1(T^{1/2}))O_p(T^{-1/2})=O_p(T^{-1/2}\nu_1(T^{1/2}))$$

Observe that

$$\begin{aligned}T^{-1}\underline{\tilde{\theta}}'\left(\sum_{t=1}^{T}Z_tZ_t'\right)\underline{\tilde{\theta}}&=\left\{\left(\sum_{t=1}^{T}Z_tZ_t'\right)^{-1}\sum_{t=1}^{T}Z_ty_t\right\}'\left(\sum_{t=1}^{T}Z_tZ_t'\right)\left\{\left(\sum_{t=1}^{T}Z_tZ_t'\right)^{-1}\sum_{t=1}^{T}Z_ty_t\right\}\\&=T^{-2}\left(\sum_{t=1}^{T}g(x_t)Z_t\right)'\left(T^{-1}\sum_{t=1}^{T}Z_tZ_t'\right)^{-1}\left(\sum_{t=1}^{T}g(x_t)Z_t\right)\\&\quad+2T^{-3/2}\left(\sum_{t=1}^{T}g(x_t)Z_t\right)'\left(T^{-1}\sum_{t=1}^{T}Z_tZ_t'\right)^{-1}\left(T^{-1/2}\sum_{t=1}^{T}Z_tu_{1t}\right)\\&\quad+T^{-1}\left(T^{-1/2}\sum_{t=1}^{T}Z_tu_{1t}\right)'\left(T^{-1}\sum_{t=1}^{T}Z_tZ_t'\right)^{-1}\left(T^{-1/2}\sum_{t=1}^{T}Z_tu_{1t}\right)\end{aligned}\tag{30}$$

since $y_t=g(x_t)+u_{1t}$. Note that $\left\|(T^{-1}\sum_{t=1}^{T}Z_tZ_t')^{-1}\right\|=O_p(1)$ and $\left\|T^{-1/2}\sum_{t=1}^{T}Z_tu_{1t}\right\|=O_p(K^{1/2})$ by the proof of Lemma 3 in Berk (1974). Also note that

$$\sum_{t=1}^{T}g(x_t)\Delta x_{t-k}=O_p\left(\nu_1(T^{1/2})T^{\frac{4+3r}{4(1+r)}}K^{\frac{r}{2(1+r)}}\right),$$

uniformly for $k=1,\ldots,K$ by Lemma A4 (b) of Chang et al. (2001). Then it follows by (30) that

$$\begin{aligned}T^{-1}\underline{\tilde{\theta}}'\left(\sum_{t=1}^{T}Z_tZ_t'\right)\underline{\tilde{\theta}}&=O_p(T^{-2})O_p\left(\nu_1(T^{1/2})T^{\frac{4+3r}{4(1+r)}}K^{\frac{r}{2(1+r)}}K\right)O_p(1)O_p\left(\nu_1(T^{1/2})T^{\frac{4+3r}{4(1+r)}}K^{\frac{r}{2(1+r)}}K\right)\\&\quad+O_p(T^{-3/2})O_p\left(\nu_1(T^{1/2})T^{\frac{4+3r}{4(1+r)}}K^{\frac{r}{2(1+r)}}K\right)O_p(1)O_p(K^{1/2})\\&\quad+O_p(T^{-1})O_p(K^{1/2})O_p(1)O_p(K^{1/2})\\&=O\left(\nu_1^2(T^{1/2})T^{\frac{2\delta(2+3r)-1}{2(1+r)}}\right)+O\left(\nu_1(T^{1/2})T^{\frac{2\delta(3+4r)-2-3r}{4(1+r)}}\right)+O(T^{\delta-1}),\end{aligned}\tag{31}$$

letting $K=T^{\delta}$. Thus we get $T^{-1}\underline{\tilde{\theta}}'\left(\sum_{t=1}^{T}Z_tZ_t'\right)\underline{\tilde{\theta}}=o_p(\nu_1^2(T^{1/2}))$, for $\delta<r/(2(2+3r))$. Similarly, we get $T^{-1}\underline{\tilde{\theta}}'\sum_{t=1}^{T}g(x_t)Z_t=o_p(\nu_1^2(T^{1/2}))$, $T^{-1}\underline{\tilde{\theta}}'\sum_{t=1}^{T}Z_tu_{1t}=o_p(\nu_1(T^{1/2}))$, and $T^{-1}\underline{\tilde{\gamma}}'\left(\sum_{t=1}^{T}X_tZ_t'\right)\underline{\tilde{\theta}}=o_p(\nu_1^2(T^{1/2}))$ for $\delta<r/(2(2+3r))$. Thus by the discussion above we can deduce that

$$T^{-1}\sum_{t=1}^{T}\hat{v}_t^{*2}=\begin{cases}O_p(\nu_1^2(T^{1/2})) & \text{if } \nu_1(T^{1/2})\to\infty \text{ as } T\to\infty,\\ O_p(1) & \text{otherwise,}\end{cases}\tag{32}$$

When $g_i\in\mathcal{T}(I)$ for all i, the similar arguments show that $T^{-1}\sum_{t=1}^{T}g^2(x_t)=o_p(1)$ by Theorem 5.1 of Park and Phillips (1999), $T^{-1}\sum_{t=1}^{T}g(x_t)u_{1t}=o_p(1)$ by Theorem 3.2 of Park and Phillips (2001), $T^{-1}\underline{\tilde{\gamma}}'\left(T^{-1}\sum_{t=1}^{T}X_tX_t'\right)\underline{\tilde{\gamma}}=o_p(1)$, $T^{-1}\underline{\tilde{\gamma}}'\sum_{t=1}^{T}g(x_t)X_t=o_p(1)$ by the proof of Theorem 3.1, Part (a) of Lemma 4.1 and Part (k) of Lemma 3.1 in Chang et al. (2001), $T^{-1}\underline{\tilde{\theta}}'\left(T^{-1}\sum_{t=1}^{T}Z_tZ_t'\right)\underline{\tilde{\theta}}=o_p(1)$ by Lemma A3 (b) of Chang et al. (2001). Thus we can deduce that a dominating term in (29) for this case is $T^{-1}\sum_{t=1}^{T}u_{1t}^2$ and its order $O_p(1)$, leading to

$$T^{-1}\sum_{t=1}^{T}\hat{v}_t^{*2}=O_p(1) \quad \text{if } g_i\in\mathcal{T}(I) \text{ for all } i. \tag{33}$$

Next we consider the cross product terms in (28). Observe that

$$\begin{aligned}
T^{-1}\sum_{t=1}^{T}\hat{v}_t^{*2}&=T^{-1}\sum_{t=1}^{T}(y_t-X_t'\tilde{\underline{\gamma}}-Z_t'\tilde{\underline{\theta}})(y_{t-s}-X_{t-s}'\tilde{\underline{\gamma}}-Z_{t-s}'\tilde{\underline{\theta}})\\
&=T^{-1}\sum_{t=1}^{T}\{g(x_t)+u_{1t}-X_t'\tilde{\underline{\gamma}}-Z_t'\tilde{\underline{\theta}}\}\{g(x_{t-s})+u_{1,t-s}-X_{t-s}'\tilde{\underline{\gamma}}-Z_{t-s}'\tilde{\underline{\theta}}\}\\
&=T^{-1}\sum_{t=1}^{T}g(x_t)g(x_{t-s})+T^{-1}\sum_{t=1}^{T}u_{1t}u_{1,t-s}\\
&\quad+T^{-1}\tilde{\underline{\gamma}}'\left(\sum_{t=1}^{T}X_tX_{t-s}'\right)\tilde{\underline{\gamma}}+T^{-1}\tilde{\underline{\theta}}'\left(\sum_{t=1}^{T}Z_tZ_{t-s}'\right)\tilde{\underline{\theta}}\\
&\quad+T^{-1}\sum_{t=1}^{T}g(x_t)u_{1,t-s}+T^{-1}\sum_{t=1}^{T}g(x_{t-s})u_{1t}-T^{-1}\tilde{\underline{\gamma}}'\left\{\sum_{t=1}^{T}g(x_t)X_{t-s}+\sum_{t=1}^{T}g(x_{t-s})X_t\right\}\\
&\quad-T^{-1}\tilde{\underline{\theta}}'\left\{\sum_{t=1}^{T}g(x_t)Z_{t-s}+\sum_{t=1}^{T}g(x_{t-s})Z_t\right\}-T^{-1}\tilde{\underline{\gamma}}'\left\{\sum_{t=1}^{T}X_tu_{1,t-s}+\sum_{t=1}^{T}X_{t-s}u_{1t}\right\}\\
&\quad-T^{-1}\tilde{\underline{\theta}}'\left\{\sum_{t=1}^{T}Z_tu_{1,t-s}+\sum_{t=1}^{T}Z_{t-s}u_{1t}\right\}+T^{-1}\tilde{\underline{\gamma}}'\left\{\sum_{t=1}^{T}X_tZ_{t-s}'\underline{\theta}+\sum_{t=1}^{T}X_{t-s}Z_t'\right\}\tilde{\underline{\theta}}.
\end{aligned}$$

A similar argument to the one used deriving (32) and (33) can be applied to each term to get

$$T^{-1}\sum_{t=s+1}^{T}\hat{v}_t^{*}\hat{v}_{t-s}^{*}=\begin{cases}O_p(\nu_1^2(T^{1/2})) & \text{if } g_i\in\mathcal{T}(H) \text{ with } \nu_i(T^{1/2})\to\infty \text{ as } T\to\infty \text{ for some } i\\ O_p(1) & \text{otherwise.}\end{cases} \tag{34}$$

except for the two terms, $T^{-1}\sum_{t=1}^{T}g(x_t)g(x_{t-x})$ and $\tilde{\underline{\gamma}}'(T^{-1}\sum_{t=1}^{T}X_tX_{t-s})\tilde{\underline{\gamma}}$. For the former, observe that when $g_i\in\mathcal{T}(H)$ for some i,

$$\frac{1}{T\nu_1^2(T^{1/2})}\sum_{t=1}^{T}g(x_t)g(x_{t-s})=T^{-1}\sum_{t=1}^{T}h_1(x_{1t})h_1(x_{1,t-s})+o_p(1) \tag{35}$$

by Theorem 3.3 of Park and Phillips (2001). We can also show that for large T

$$\begin{aligned}
T^{-1}\sum_{t=1}^{T}\|h_1(x_{1t})h_1(x_{1,t-s})\|&\le\left(T^{-1}\sum_{t=1}^{T}|h(x_{1t})|^2\right)^{1/2}\left(T^{-1}\sum_{t=1}^{T}|h(x_{1,t-s})|^2\right)^{1/2}\\
&\le\|h_1\|_E^2
\end{aligned} \tag{36}$$

where the first inequality follows from Cauchy-Schwartz inequality and E_i is defined in (19). Then it follows from (35) and (36) that

$$T^{-1}\sum_{t=1}^{T}g(x_t)g(x_{t-s})=\begin{cases}O_p(\nu_1^2(T^{1/2})) & \text{if } \nu_1(T^{1/2})\to\infty \text{ as } T\to\infty,\\ O_p(1) & \text{otherwise,}\end{cases}$$

When $g_i\in\mathcal{T}(I)$ for all i, we get from (36) that $T^{-1}\sum_{t=1}^{T}g(x_t)g(x_{t-s})=O_p(1)$. Applying the same argument to $(T^{-1}\sum_{t=1}^{T}X_tX_{t-s})$ gives $\tilde{\underline{\gamma}}'(T^{-1}\sum_{t=1}^{T}X_tX_{t-s})\tilde{\underline{\gamma}}=O_p(1)$ leading to (34).

Given (32), (33) and (34), the argument used in the proof of Theorem 3.1 of Phillips (1991) shows that

$$\widehat{\omega}_{11\cdot 2}=\begin{cases} O_p(l\nu_1^2(T^{1/2})) & \text{if } g_i\in\mathcal{T}(H) \text{ with } \nu_i(T^{1/2})\to\infty \text{ as } T\to\infty \text{ for some } i \\ O_p(l) & \text{otherwise.}\end{cases}$$

□

Proof of Theorem 4.1: Observe that

$$\begin{aligned} W_T&=(R\tilde{\underline{\gamma}})'(\widehat{\omega}_{11\cdot 2}RM^{-1}R')^{-1}(R\tilde{\underline{\gamma}}) \\ &=(R\Upsilon_T^{-1}\Upsilon_T\tilde{\underline{\gamma}})'(\widehat{\omega}_{11\cdot 2}R\Upsilon_T^{-1}\Upsilon_TM^{-1}\Upsilon_T\Upsilon_T^{-1}R')^{-1}(R\Upsilon_T^{-1}\Upsilon_T\tilde{\underline{\gamma}}) \\ &=(\Upsilon_T\tilde{\underline{\gamma}})'(R\Upsilon_T^{-1})'\left\{\widehat{\omega}_{11\cdot 2}(R\Upsilon_T^{-1})(\Upsilon_TM^{-1}\Upsilon_T)(\Upsilon_T^{-1}R')\right\}^{-1}(R\Upsilon_T^{-1})(\Upsilon_T\tilde{\underline{\gamma}}). \\ &=(\Upsilon_T\tilde{\underline{\gamma}})'(R'\tilde{\Upsilon}_T^{-1})\left\{\widehat{\omega}_{11\cdot 2}(\tilde{\Upsilon}_T^{-1}R)(\Upsilon_TM^{-1}\Upsilon_T)(R'\tilde{\Upsilon}_T^{-1})\right\}^{-1}(\tilde{\Upsilon}_T^{-1}R(\Upsilon_T\tilde{\underline{\gamma}}) \\ &=(R\Upsilon_T\tilde{\underline{\gamma}})'\{\widehat{\omega}_{11\cdot 2}R(\Upsilon_TM^{-1}\Upsilon_T)R'\}^{-1}(R\Upsilon_T\tilde{\underline{\gamma}}) \end{aligned} \tag{37}$$

where $\tilde{\Upsilon}_T$ is a lower-right $(\kappa-1)\times(\kappa-1)$ submatrix of Υ_T. If $g_i\in\mathcal{T}(H)$ with $\nu_i(T^{1/2})\to\infty$ as $T\to\infty$ for some i,

$$\begin{aligned} W_T&=(R\Upsilon_T\tilde{\underline{\gamma}})'\{\widehat{\omega}_{11\cdot 2}R(\Upsilon_TM^{-1}\Upsilon_T)R'\}^{-1}(R\Upsilon_T\tilde{\underline{\gamma}}) \\ &=O_p(T^{1/2}\nu_1(T^{1/2}))O_p(l^{-1}\nu_1^{-2}(T^{1/2}))O_p(1)O_p(T^{1/2}\nu_1(T^{1/2}))=O_p(T/l), \end{aligned}$$

by Lemma 4.1 and the proof of Theorem 3.1. If $g_i\in\mathcal{T}(H)$ with $\nu_1(T^{1/2})$ is constant as $T\to\infty$ for some i, using the same argument we get $W_T=O_p(T^{1/2})O_p(l^{-1})O_p(1)O_p(T^{1/2})=O_p(T/l)$. If $g_i\in\mathcal{T}(H)$ with $T^{1/2}\nu_1(T^{1/2})\to\infty$ and $\nu_1(T^{1/2})\to 0$ as $T\to\infty$, we also obtain by the same argument

$$W_T=O_p(T^{1/2}\nu_1(T^{1/2}))O_p(l^{-1})O_p(1)O_p(T^{1/2}\nu_1(T^{1/2}))=O_p(T\nu_1^2(T^{1/2})/l).$$

Otherwise $W_T=O_p(1)O_p(l^{-1})O_p(1)O_p(1)=O_p(l^{-1})$, giving the required result. □

Proof of Lemma 4.2: (a) Given the result of the proof of Lemma 4.1, it is sufficient to show that

$$\Upsilon_T\sum_{t=1}^{T}X_ty_t=O_p(T) \tag{38}$$

since the arguments in Lemma 4.1 can be applied to other parts of the proof. (38) can be easily proved by the application of the continuous mapping theorem $T^{-(i+3)/2}\sum_{t=1}^{T}x_t^{(i)}y_t\Rightarrow\int B_2^{(i)}B_1$.

(b) Given (a), the proof of (b) is completely analogous to that of Lemma 4.1 (b) and so it is omitted. □

Proof of Theorem 4.2: By (37), we have

$$W_T=(R\Upsilon_T\tilde{\underline{\gamma}})'\{\widehat{\omega}_{11\cdot 2}R(\Upsilon_TM^{-1}\Upsilon_T)R'\}^{-1}(R\Upsilon_T\tilde{\underline{\gamma}})=O_p(T)O_p(l^{-1}T^{-1})O_p(1)O_p(T))=O_p(T/l),$$

where the last equality is from Lemma 4.1. □

References

Akonom, J. (1993), "Comportement Asymptotique du Temps d'Occupation du Process Sus des Sommes Partielles," *Annale de l'Institut Henri Poincaré* 29, pp.57-81.

Andrews, Donald W.K. (1991), "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica* 59, pp.817-858.

Berk, K.N. (1974), "Consistent Autoregressive Spectral Estimates," *Annals of Statistics* 2, pp.489-502.

Billingsley, P. (1968), *Convergence of Probability Measures*, New York, John Wiley & Sons.

Chang, Y., J.Y. Park, and P.C.B. Phillips (2001), "Nonlinear Econometric Models with Cointegrated and Deterministically Trending Regressors," *Econometrics Journal* 4, pp.1-36.

Choi, I., and E. Kurozumi (2012), "Model Selection Criteria for the Leads-and-Lags Cointegrating Regression," *Journal of Econometrics* 169, pp.224-238.

Corradi, V., N.R. Swanson and H. White (2000), "Testing for Stationarity-Ergodicity and for Comovements between Nonlinear Discrete Time Markov Processes," *Journal of Econometrics* 96, pp.39-73.

de Jong, R.M. (2002), "Nonlinear Estimators with Integrated Regressors but without Exogeneity," mimeo.

de Jong, R.M. (2004), "Addendum to "Asymptotics for Nonlinear Transformations of Integrated Time Series"," *Econometric Theory* 20, pp.627-635.

Engle, R.F. and C.W.J. Granger (1987), "Co-Integration and Error Correction: Representation, Estimation and Testing," *Econometrica* 55, pp.547-557.

Granger, C.W.J. (1983), "Co-Integrated and Error-Correcting Models," Discussion Paper 83-13, Department of Economics, University of California, San Diego.

Granger, C.W.J. (1995), "Modelling Nonlinear Relationships between Extended-Memory Variables," *Econometrica* 63, pp.265-279.

Granger, C.W.J. and P. Newbold (1974), "Spurious Regressions in Econometrics," *Journal of Econometrics* 2, pp.111-120.

Granger, C.W.J. and T. Teräsvirta (1993), *Modelling Non–Linear Economic Relationships*, New York, Oxford University Press.

Hansen, B.E. (1992), "Convergence to Stochastic Integrals for Dependent Heterogeneous Processes," *Econometric Theory* 8, pp.489-500.

Hansen, B.E. and B. Seo (2002), "Testing for Two-Regime Threshold Cointegration in Vector Error Correction Models," *Journal of Econometrics* 110, pp.293-318.

Hong, S.H., and P.C.B. Phillips (2010), "Testing Linearity in Cointegrating Relations with an Application to Purchasing Power Parity," *Journal of Business & Economic Statistics* 28, pp.96-114.

Kurozumi, E. (2002), "Testing for Stationarity with a Break," *Journal of Econometrics* 108, pp.63-99.

Kwiatkowski, D., P.C.B. Phillips, P. Schmidt and Y. Shin (1992), "Testing the Null Hypothesis of Stationarity against the Alternative of a Unit Root," *Journal of Econometrics* 54, pp.159-178.

Newey, W.K., and K.D. West (1987), "A Simple, Positive Semidefinite, Heteroskedasticity and Autocorrelated Consistent Covariance Matrix," *Econometrica* 55, pp.586-591.

Ng, S. and P. Perron (1995), "Unit Root Tests in ARMA Models with Data-Dependent Methods for the Selection of the Truncation Lag," *Journal of the American Statistical Association* 90, pp.268-281.

Ogaki, M. and J.Y. Park (1997), "A Cointegration Approach to Estimating Preference Parameters," *Journal of Econometrics* 82, pp.107-134.

Park, J.Y., and P.C.B. Phillips (1988), "Statistical Inference in Regressions with Integrated Processes: Part 1," *Econometric Theory* 4, pp.468-498.

Park, J.Y. and P.C.B. Phillips (1999), "Asymptotics for Nonlinear Transformations of Integrated Time Series," *Econometric Theory* 15, pp.269-298.

Park, J.Y., and P.C.B. Phillips (2000), "Nonstationary Binary Choice," *Econometrica* 68, pp.1249-1280.

Park, J.Y., and P.C.B. Phillips (2001), "Nonlinear Regressions with Integrated Time Series," *Econometrica* 69, pp.117-161.

Phillips, P.C.B. (1986), "Understanding Spurious Regressions in Econometrics," *Journal of Econometrics* 33, pp.311-340.

Phillips, P.C.B. (1987), "Time Series Regression with a Unit Root," *Econometrica* 55, pp.277-301.

Phillips, P.C.B. (1988), "Weak Convergence of Sample Covariance Matrices to Stochastic Integrals via Martingale Approximations," *Econometric Theory* 4, pp.528-533.

Phillips, P.C.B. (1991), "Spectral Regression for Cointegrated Time Series," in *Nonparametric and Semiparametric Methods in Economics and Statistics,* ed. W. Barnett, New York, Cambridge University Press, pp.413-435.

Phillips, P.C.B. and B.E. Hansen (1990), "Statistical Inference in Instrumental Variables Regression with I(1) Processes," *Review of Economic Studies* 57, pp. 99-125.

Phillips, P.C.B., and S.N. Durlauf (1986), "Multiple Time Series Regression with Integrated Processes," *Review of Economic Studies* 53, pp.473-496.

Ramsey, J.B. (1969), "Tests for Specification Errors in Classical Linear Least-Squares Regression Analysis," *Journal of the Royal Statistical Society, Series B* 31, pp.350-371.

Saikkonen, P. (1991), "Asymptotically Efficient Estimation of Cointegration Regressions," *Econometric Theory* 7, pp.1-21.

Schwert, G.W. (1989), "Tests for Unit Roots: A Monte Carlo Investigation," *Journal of Business & Economic Statistics* 7, pp.147-159.

Stock, J.H. (1987), "Asymptotic Properties of Least Squares Estimators of Cointegrating Vectors," *Econometrica* 55, pp.1035-1056.

Thursby, J.G. and P. Schmidt (1977), "Some Properties of Tests for Specification Error in a Linear Regression Model," *Journal of the American Statistical Association* 72, pp.635-641.

Tøstheim, D. (2012), "Modelling Nonlinear Nonstationary Time Series," in *Handbook of Statistics,* ed. T.S. Rao, S.S. Rao, and C.R. Rao, vol. 30, Amsterdam, North Holland, pp.67-97.

Wooldridge, J.M., and H. White (1988), "Some Invariance Principles and Central Limit Theorems for Dependent Heterogeneous Processes," *Econometric Theory* 4, pp.210-230.