



ELSEVIER

Journal of Econometrics 96 (2000) 39–73

JOURNAL OF  
Econometrics

www.elsevier.nl/locate/econbase

# Testing for stationarity-ergodicity and for comovements between nonlinear discrete time Markov processes

Valentina Corradi<sup>a</sup>, Norman R. Swanson<sup>b</sup>, Halbert White<sup>c,\*</sup>

<sup>a</sup>University of Pennsylvania, Philadelphia, PA 19104, USA

<sup>b</sup>Pennsylvania State University, University Park, PA 16802, USA

<sup>c</sup>University of California-San Diego, Department of Economics, 9500 Gilman Driver,  
La Jolla CA 92093-0508, USA

---

## Abstract

In this paper we introduce a class of nonlinear data generating processes (DGPs) that are first order Markov and can be represented as the sum of a linear plus a bounded nonlinear component. We use the concepts of geometric ergodicity and of linear stochastic comovement, which correspond to the linear concepts of integratedness and cointegratedness, to characterize the DGPs. We show that the stationarity test due to Kwiatkowski et al. (1992, *Journal of Econometrics*, 54, 159–178) and the cointegration test of Shin (1994, *Econometric Theory*, 10, 91–115) are applicable in the current context, although the Shin test has a different limiting distribution. We also propose a consistent test which has a null of linear cointegration (comovement), and an alternative of ‘non-linear cointegration’. Monte Carlo evidence is presented which suggests that the test has useful finite sample power against a variety of nonlinear alternatives. An empirical illustration is also provided. © 2000 Elsevier Science S.A. All rights reserved.

*JEL classification:* C12; C22

*Keywords:* Cointegration; Linear stochastic comovement; Markov processes; Nonlinearities

---

\* Corresponding author. Fax: (858)-534-7040.

E-mail address: hwhite@weber.ucsd.edu (H. White)

## 1. Introduction

In most econometric applications there is little theoretical justification for believing in the correctness of linear specifications when modeling economic variables. In consequence, nonlinear time series models have received increasing attention during the last few years (for example, see Tong (1990), Granger and Teräsvirta (1993), Granger (1995), Granger et al. (1997), Anderson and Vahid (1998), and the references contained therein). Nevertheless, when nonlinear models are specified, correct inference requires some knowledge as to whether the underlying data generating processes (DGPs) are stationary and ergodic in some appropriate sense, or instead have trajectories that explode with positive probability as the time span approaches infinity. In the linear case it is common to test for unit roots in order to check whether a series is integrated of order 1, denoted  $I(1)$ , or integrated of order 0, denoted  $I(0)$ , where a process is said to be  $I(d)$  if the scaled partial sum of its  $d$ th difference satisfies a functional central limit theorem (FCLT). If one has two or more  $I(1)$  series, it is common to test for cointegration in order to determine whether there exists a linear combination of the variables which is  $I(0)$ . However, the concepts of integratedness and cointegratedness typically apply to linear DGPs, in the sense that the conditional mean is assumed to be a linear function of a set of conditioning variables. In contrast, strictly convex or concave transformations of random walks have a unit root component, but they are not  $I(1)$ , in the sense that their first differences need not be short memory processes (Corradi, 1995).

In this paper we examine nonlinear DGPs that are first-order Markov and can be represented as the sum of a linear plus a bounded nonlinear component. For such DGPs, we exploit results by Chan (see Appendix to Tong, 1990) to obtain simple conditions for distinguishing between processes that are geometric ergodic (and thus strong mixing) and processes having trajectories that explode with positive probability as  $T \rightarrow \infty$ . Using these conditions, we replace the concept of cointegratedness with concept of *linear stochastic comovement*. Specifically, if  $X = (X_{i,t}, i = 1, 2, \dots, p, t = 1, 2, \dots, T)$  is a nonergodic Markov process in  $R^p$ , in the sense that the trajectories explode with positive probability as  $T \rightarrow \infty$ , but there exists an  $r$  dimensional linear combination, say  $\theta_0'X_t$ , with  $\theta_0$  a full column rank  $p \times r$  matrix ( $r < p$ ) that is ergodic in  $R^r$ , then there is *linear stochastic comovement* among the components of  $X$ . We use the term 'Markov process' to mean a process in which the state space is continuous and time is discrete. Note that our approach differs from that of Granger and Hallman (1991). According to their terminology, two long memory series, say  $X_t$  and  $Y_t$  have an *attractor* if there exists a linear combination of nonlinear functions of  $X_t$  and  $Y_t$ , say  $g(X_t) - h(Y_t)$  that is short memory. In contrast, we consider the case of linear combinations among the components of nonlinear and nonergodic Markov processes that form nonlinear and *ergodic* Markov processes. For our class of nonlinear DGPs, the null hypothesis of ergodicity and the null

hypothesis of no unit root can be formulated in the same way. Similarly, the null hypothesis of linear stochastic comovement and the null hypothesis of cointegration can be formulated in the same way. Thus, the presence of stochastic comovement implies the presence of cointegration among the linear components of our nonlinear models, and vice versa. Given this framework, one of our main goals is to propose a ‘nonlinear cointegration’ test, for which the null hypothesis is linear cointegration, and the alternative is nonlinear cointegration. The test which we propose is consistent against a wide variety of nonlinear alternative, including neural network models with sigmoidal activation functions (e.g. logistic cumulative distribution functions (cdf)). We show using a series of Monte Carlo experiments that our nonlinearity test has the ability to distinguish between a variety of linear and nonlinear models for moderate sample sizes.

As we typically do not have information concerning the precise form of the nonlinear component, we examine the effect that neglected nonlinearities have on tests for the null of stationarity (unit root) and for the null of cointegration (no cointegration). We note that in the presence of neglected nonlinearities, tests with a null hypothesis of integratedness, as well as tests with a null hypothesis of no cointegration, do not have easily determined limiting distributions. This is because in the presence of neglected non-linearities the innovation terms are no longer strong mixing and in general do not satisfy standard invariance principles. Consequently, standard unit root asymptotics no longer necessarily apply. Along these lines, we first examine the stationarity test proposed by Kwiatkowski et al. (1992). We show that this test has a well-defined limiting distribution under the null hypothesis of general nonlinear stationary-ergodic DGPs and has power not only against the alternative of integratedness, but also against alternatives involving a range of nonlinear-nonergodic processes. Second, we show that the Shin (1994) test for the null hypothesis of cointegration can be used to test for stochastic comovement, although the limiting distribution of the test is different. Interestingly, the ADF unit root and Johansen cointegration tests no longer have straightforward limiting distributions in general. Nevertheless, we show using a series of Monte Carlo experiments that these tests may still be reliable in practice, in the sense that they exhibit moderate bias and reasonable power (e.g. the empirical power is more than 0.5 for samples as small as 250 observations when the nonlinear component in our model is a logistic cdf).

The rest of the paper is organized as follows. In Section 2 we describe our set-up. In Section 3, we examine stationary-ergodicity and cointegration (comovement) tests. In Section 4 we propose a test for distinguishing between linear and nonlinear cointegration. In Section 5 we summarize the results from a series of Monte Carlo experiments, while Section 6 contains an empirical illustration. Section 7 concludes. All the proofs are collected in an Appendix. In the sequel,  $\Rightarrow$  denotes weak convergence;  $W$  denotes a standard Brownian motion.

## 2. Assumption and preliminary results

We start by considering the following DGP:

$$X_t = AX_{t-1} + g_0(\theta_0^1 X_{t-1}, \dots, \theta_0^r X_{t-1}) + \varepsilon_t, \quad (2.1)$$

where  $X_t: \Omega \rightarrow R^p$ ,  $t = 1, 2, \dots, T$  with  $(\Omega, \mathcal{F}, P)$  an underlying probability space, and  $\theta_0^i$  denotes the  $i$ th-column of  $\theta_0$ , a full column rank  $p \times r$  matrix, with  $r \leq p$ , and  $1 \leq j \leq r < p$ . Assume also that

- A1.  $\varepsilon_t$  is identically and independently distributed (iid), has a distribution which is absolutely continuous with respect to the Lebesgue measure in  $R^p$ , and has positive density everywhere.
- A2.  $E(\varepsilon_t) = 0$  and  $E(\varepsilon_t \varepsilon_t') = \Sigma$  where  $\Sigma$  is positive definite and  $E((\varepsilon_t' \varepsilon_t)^2) < \infty$ .
- A3.  $g_0: R^j \rightarrow R^p$  is bounded, Lipschitz continuous, and differentiable in the neighborhood of the origin. Furthermore,  $g_0$  is not everywhere equal to zero.

Although A3 is a somewhat strong assumption, it should be noted that a wide variety of nonlinearities are contained within the class of DGPs which we examine. For example various neural network models with sigmoidal activation functions satisfy A3. Examples include feedforward artificial neural networks with a single 'hidden unit' and either logistic or normal cdfs as activation functions (see, e.g. Kuan and White, 1994). Other examples of functional forms for  $g_0$  include modified exponential autoregressive models where  $g_0(x) = xe^{-x^2}$  (Tong, 1990, p. 129), and symmetric smooth transition autoregressive (STAR) type models where  $g_0(x) = x/(1 + e^{x^2})$ . On the other hand, the logistic STAR  $g_0(x) = x/(1 + e^{-x})$  and the exponential STAR ( $g_0(x) = x(1 - e^{x^2})$ ) (see, e.g. Teräsvirta and Anderson, 1992) are ruled out. Furthermore  $g_0$  may contain a constant term. Higher-order lag structures are allowed by a variant of (2.1), as a  $p$ -dimensional  $k$ -order Markov process can be written as a  $kp$ -dimensional first-order Markov process with a positive semi-definite  $\Sigma$  of rank  $p$ .

*Proposition 2.1.* For DGP (2.1), suppose that A1–A3 hold. If all of the eigenvalues of the matrix  $A$  are strictly less than one in absolute value, then  $X = (X_{i,t}; i = 1, \dots, p, t = 1, 2, \dots, T)$  is a geometric ergodic Markov process in  $R^p$ , with an invariant probability measure which is absolutely continuous with respect to the Lebesgue measure in  $R^p$ .

*Definition 2.2. (Linear stochastic comovement).* Assume that  $X$  is a nonergodic Markov process in  $R^p$ , in the sense that each component of  $X$  approaches infinity with positive probability as  $T \rightarrow \infty$ . Assume also that there exists some full column rank  $p \times r$  ( $r < p$ ) matrix,  $\theta_0 = (\theta_0^1, \dots, \theta_0^r)$ , such that  $\theta_0' X_t$  is an ergodic Markov process, in the sense that it has an invariant probability measure which is absolutely continuous with respect to the Lebesgue measure in

$R^r$ . Then, there exists *linear stochastic comovement* among the components of  $X$ , and  $\theta_0^i$  is the *ith comovement vector*,  $i = 1, \dots, r$ .

In the sequel we specialize Definition 2.2. to the DGP given in (2.1). For this reason, we use  $\theta_0$  in (2.1). Now let  $\Phi \equiv A - I$ , where  $I$  is the  $p \times p$  identity matrix. We assume either of the following:

- A4(i).  $\Phi = \alpha\theta_0'$ , where  $\alpha$  and  $\theta_0$  are full column rank  $p \times r$  matrices,  $r < p$ . The eigenvalues of  $\Phi$ , say  $\lambda_i$ , are such that  $-2 < \lambda_1 \leq \lambda_2 \dots \leq \lambda_r < 0$ .
- A4(ii).  $A = I$ .

*Proposition 2.3. Assume that (2.1), and A1–A4(i) hold. Then: (i)  $X$  is a nonergodic Markov process in  $R^p$  and  $P(\|X_T\| \rightarrow \infty) \equiv \eta > 0$  as  $T \rightarrow \infty$ .*

*(ii)  $\theta_0'X_t$  is a geometric ergodic Markov process in  $R^r$  which has an invariant probability measure,  $\pi$ , that is absolutely continuous with respect to the Lebesgue measure ( $\mu$ ) in  $R^r$ , and which has density  $\nu = d\pi/d\mu$ . Further,*

$$\theta_0'\Delta X_t = \theta_0'\Phi X_{t-1} + \theta_0'g_0(\theta_0^1 X_{t-1}, \dots, \theta_0^r X_{t-1}) + \theta_0'\varepsilon_t.$$

It follows that there is stochastic comovement among the components of  $X$  (from Definition 2.2).

*Proposition 2.4. Assume that (2.1), A1–A3, and A4(ii) hold. Then both  $X$  and  $\theta_0'X$  are nonergodic Markov processes in  $R^p$  and  $R^r$ , respectively. It also follows that*

$$P[\|X_T\| \rightarrow \infty] > 0 \quad \text{and} \quad P[\|\theta_0'X_T\| \rightarrow \infty] > 0.$$

The ergodicity of the process defined in (2.1) is implied by the stability of the associated deterministic dynamical system. This allows us to analyze the ergodicity of the stochastic system by examining the eigenvalues of  $A$  or  $\Phi$ . This means that the same conditions which ensure the ergodicity of the linear part of our model also ensure the ergodicity of the entire nonlinear process. Thus, for the processes which we are considering, the existence of stochastic comovement is equivalent to the existence of cointegration among the linear components in the model. In particular, the existence of  $r$  comovement vectors implies the existence of  $r$  cointegrating vectors in the linear part of the model. For brevity, we refer to this subsequently as just ‘cointegration’.

Observe that in the case where  $A - I = \Phi = \alpha\theta_0'$  there is a clear interpretation of the argument of  $g_0$  in terms of cointegrating vectors. On the other hand when  $A = I$ ,  $g_0$  depends on some generic linear combination of the  $X$ 's. Thus under A4(i), the nonlinear component is a geometric ergodic process, while under A4(ii), the nonlinear component is a nonlinear nonergodic process. In order to

keep a ‘continuous’ relationship between the arguments of  $g_0$  and  $A - I$ , we could consider the following variation of (2.1):

$$\Delta X_t = \alpha \theta'_0 X_{t-1} + g_0((\theta_0, 0)' X_{t-1}) + \varepsilon_t,$$

where  $\theta_0$  is a  $p \times r$  ( $r < p$ ) matrix,  $0$  is a  $p \times (p - r)$  matrix of zeroes, and  $\theta_0 = 0$  when  $r = 0$ . Thus, when  $r = 0$ ,  $X_t$  is an  $I(1)$  process with a deterministic trend component. It should be noted that all the theorems below hold for this special case. (This point was kindly communicated to us by Herman Bierens.)

The following facts will be frequently used in the paper.

*Fact 2.5* (From Athreya and Pantula, 1986, Theorem 1). *Geometric ergodic discrete time Markov processes are strong mixing. Further, the speed at which the mixing coefficient declines to zero is proportional to the speed at which the transition distribution converges to the invariant probability measure. Thus, when the transition distribution approaches the invariant probability measure at a geometric rate, the mixing coefficients also decay at a geometric rate.*

To ensure the next fact, we add another assumption.

A5.  $X_0$  is a random  $p$ -vector and  $\theta'_0 X_0$  is drawn from a density  $v$ , where  $v$  is the density associated with the invariant probability measure,  $\pi$ , as defined in Proposition 2.3(ii).

*Fact 2.6* (from Meyn, 1989). *For (2.1), if A1–A4(i) and A5 hold, then  $\theta'_0 X_t$  has density  $v$  for all  $t = 1, 2, \dots, T$ . Thus  $X$  is strictly stationary, in addition to being a geometric ergodic process (and thus strong mixing).*

### 3. Testing for stationarity–ergodicity and for linear stochastic comovement

We begin by considering the one-dimensional case, (i.e.  $p = 1$ ), and the test proposed by Kwiatkowski et al. (1992). Without loss of generality assume that  $\theta_0 = 1$ , so that (2.1) can be written as:

$$X_t = \alpha X_{t-1} + g_0(X_{t-1}) + \varepsilon_t \quad (3.1)$$

The null hypothesis considered by Kwiatkowski et al. (1992) is rather general and includes (3.1) when  $\alpha < 1$ , and A5 holds. However, the alternative is somewhat restrictive, as  $X_t$  is assumed to be an integrated time series characterized by the sum of a random walk component, a stationary (short memory) component, and possibly a time trend component. This alternative does not include nonlinear nonergodic DGPs such as (3.1), with  $\alpha = 1$ . For this case the first difference of  $X_t$  is not a strong mixing process, in general, as it displays ‘too much’ memory. Nevertheless, we show below that the statistic proposed by Kwiatkowski et al. (1992) does have power against (3.1) with  $\alpha = 1$ .

*Theorem 3.1. Assume that (3.1), and A1–A3 hold.*

(i) *If  $|\alpha| < 1$ , and A5 holds, then*

$$S_T = \frac{1}{s_{l_r}^2} \frac{1}{T^2} \sum_{t=1}^T \left( \sum_{j=1}^t (X_j - \bar{X}) \right)^2 \Rightarrow \int_0^1 V_r^2 dr,$$

where  $V_r = W_r - rW_1$ ,  $W_r = W(r)$ ,  $0 \leq r \leq 1$ ,  $\bar{X} = T^{-1} \sum_{t=1}^T X_t$ , and

$$s_{l_r}^2 = \frac{1}{T} \sum_{t=1}^T (X_t - \bar{X})^2 + \frac{2}{T} \sum_{t=1}^{l_r} \sum_{j=1}^{l_r} \left( 1 - \frac{t}{l_r + 1} \right) \\ \times \sum_{j=t+1}^T (X_j - \bar{X})(X_{j-t} - \bar{X}),$$

with  $l_r = o(T^{1/2})$ .

(ii) *If  $\alpha = 1$ , then*

$$P[S_T > C_T] \rightarrow 1 \quad \text{as } T \rightarrow \infty,$$

where  $C_T \rightarrow \infty$  and  $\frac{C_T l_r}{T} \rightarrow 0$ , as  $T \rightarrow \infty$ .

Part (i) of Theorem 3.1 ensures that the distribution under our null is exactly the same as the distribution of the Kwiatkowski et al. (1992) test statistic. Further, part (ii) of Theorem 3.1 ensures that under our alternative,  $S_T$  diverges at the same rate as does the Kwiatkowski et al. (1992) statistic under their alternative. However, note that our alternative is more general than their alternative of integratedness, as it includes DGPs consisting of a unit root component plus a possibly long memory component. Assumption A5, which ensures strict stationarity, can be relaxed and replaced by the weaker property of constancy of the first moment (i.e.  $E(X_t) = E(X)$  for all  $t$ ). If  $E(X_t)$  depends on  $t$ , though, the partial sums of  $X_t - \bar{X}$  do not necessarily satisfy a FCLT and so the numerator of the test statistic may diverge. Recently, Domowitz and El-Gamal (1993, 1997) have proposed a test for the null hypothesis of ergodicity. Their test is based on the convergence of Cesaro averages of the iterates from different initial densities, has the correct size under the null, regardless of whether the process is stationary or not, and has power against the alternative of nonergodicity given a maintained assumption of stationarity.

Now we turn to the case where  $p > 1$ , and examine the cointegration test of Shin (1994). Although our testing framework holds for arbitrary finite  $p$ , for simplicity we limit ourselves to the case of  $p = 2$ ,  $r = 0, 1$ . The extension to the case of  $p > 2$  gives no further insight into the effect of neglected nonlinearities. As mentioned in Section 2, the existence of stochastic comovement is equivalent

to the existence of cointegration, in the current context. Thus, tests that have (no) stochastic comovement under the null hypothesis are equivalent to tests that have (no) cointegration under the null. In this section we consider the following two DGPs:

$$\Delta X_t = g_0(\theta'_0 X_{t-1}) + \varepsilon_t \quad (3.2)$$

and

$$\Delta X_t = \Phi X_{t-1} + g_0(\theta'_0 X_{t-1}) + \varepsilon_t, \quad (3.3)$$

where  $\theta_0 = (-\theta_1, \theta_2)'$  and  $\Phi = \alpha\theta'_0$ , such that  $-2 < -\alpha_1\theta_1 + \alpha_2\theta_2 < 0$ . From (3.3) we have that

$$\theta'_0 \Delta X_t = (-\alpha_1\theta_1 + \alpha_2\theta_2)\theta'_0 X_{t-1} + \theta'_0 g_0(\theta'_0 X_{t-1}) + \theta'_0 \varepsilon_t. \quad (3.4)$$

Let

$$\theta_2^{-1}\theta'_0 = \gamma'_0 = (-\theta_2^{-1}\theta_1, 1) = (-\beta_0, 1)$$

and

$$\delta = -\alpha_1\theta_1 + \alpha_2\theta_2,$$

so that

$$\gamma'_0 \Delta X_t = \delta \gamma'_0 X_{t-1} + \gamma'_0 g_0(\theta'_0 X_{t-1}) + \gamma'_0 \varepsilon_t. \quad (3.5)$$

As  $\theta'_0 X_t$  is a geometric ergodic process,  $\gamma'_0 X_t$  is also a geometric ergodic process.

It is convenient to use a triangular representation of (3.3):

$$X_{1,t} = \sum_{j=1}^t v_{1,j} \quad \text{and} \quad X_{2,t} = \beta_0 X_{1,t} + v_{2,t}, \quad (3.6)$$

where

$$v_{1,t} = \alpha_1 \theta_2 \gamma'_0 X_{t-1} + g_{0,1}(\theta'_0 X_{t-1}) + \varepsilon_{1,t}, \quad (3.7)$$

$$v_{2,t} = (\delta + 1)\gamma'_0 X_{t-1} + \gamma'_0 g_0(\theta'_0 X_{t-1}) + \gamma'_0 \varepsilon_t, \quad (3.8)$$

and  $g_{0,i}$ ,  $i = 1, 2$ , denotes the  $i$ th component of  $g_0$ .

In order to test the null hypothesis of stochastic comovement, we use the statistic proposed by Shin (1994) for testing the null of cointegration. Although stochastic comovement and cointegration are equivalent concepts, we show that, unless  $E(v_{1,t}) = 0$  for all  $t$ , the distribution that we obtain under our null hypothesis differs from that obtained by Shin. The underlying intuition is that



unless  $E(v_{1,t}) = 0$ ,  $X_{1,t}$  displays a deterministic trend component. In fact, we can write  $X_{1,t}$  in (3.6) as

$$X_{1,t} = \pi t + \sum_{j=1}^t (v_{1,j} - \pi),$$

where  $E(v_{1,t}) = \pi$  for all  $t$ , as under A5,  $v_{1,t}$  is strictly stationary.  $X_{2,t}$  can be written in an analogous way. The role of the comovement vector is to ‘cancel out’ both stochastic and deterministic trends when the linear combination,  $X_{2,t} - \beta_0 X_{1,t}$ , is formed. Furthermore,  $X_{2,t} - \beta_0 X_{1,t}$  is an ergodic process. As  $X_{1,t}$  is dominated by its trend component, the estimator of the comovement vector is  $T^{3/2}$ -consistent. Also, the asymptotic distribution of  $T^{3/2}(\hat{\beta}_T - \beta_0)$ , where  $\hat{\beta}_T$  is the coefficient from the regression of  $X_{2,t}$  on  $X_{1,t}$  and a constant term (as in (3.6)), differs from the asymptotic distribution for the driftless case. Note that the only difference between  $X_{2,t}$  in (3.6) and Eq. (2.1) in Hansen (1992a) is that  $E(v_{2,t}) = \phi \neq 0$ , so that we need to introduce a constant term into our cointegrating regression.

*Theorem 3.2.* Assume that (3.3), and A1–A4(i), A5 hold, and that  $E(v_{1,t}) \neq 0$ . Then

$$T^{3/2}(\hat{\beta}_T - \beta_0) \Rightarrow \frac{12}{\pi} \sigma_{v_2} \left( \int_0^1 s \, dW_s - 1/2 W_1 \right),$$

where  $\sigma_{v_2}^2 = \text{Var}(v_{2,t})$ , as defined in (3.8), and  $W_s = W(s)$ ,  $0 \leq s \leq 1$ .

As in Theorem 3.1(i), we require only that the first moment of  $X_t$  is constant, and A5 ensures this. The same is also true for Theorem 3.3(i) below.

Note that  $\int_0^1 s \, dW_s - 1/2 W_1 = 1/2 W_1 - \int_0^1 W_s \, ds \sim N(0, 1/12)$ , as  $\int_0^1 W_s \, ds \sim N(0, 1/13)$  and  $\text{Cov}(1/2 W_1, \int_0^1 W_s \, ds) = 1/4$ . Thus, the limiting distribution in Theorem 3.2 is normal. The representation of the limiting distribution given in Theorem 3.2 is more convenient for computing the limiting distribution of the test for the null of stochastic comovement (cointegration). If instead  $E(v_{1,t}) = 0$  for all  $t$ , then

$$T(\hat{\beta}_T - \beta_0) \Rightarrow \left( \int_0^1 W_{1,s}^2 \, ds \right)^{-1} \left( \int_0^1 W_{1,s} \, dW_{2,s} \right) + \Delta_{12},$$

where  $\Delta_{12} = 0$  and  $E(W_{1,t}, W_{2,t}) = 0$  for all  $t$ , only if  $E(v_{1,t}, v_{2,t}) = 0$  for all  $t, s$ . Nevertheless, for a wide class of nonlinearities  $E(v_{1,t}) \neq 0$ , for all  $t$  (see below). Under the alternative of no stochastic comovement,  $\hat{\beta}_T$  is bounded in probability. This holds even though we do not obtain a limiting distribution. (The reason why we do not obtain a limiting distribution is because the partial sum of the

nonlinear component, in general, does not satisfy a standard invariance principle.)

We now show that the Shin (1994, Eq. (6))  $C_\mu$  test can be applied to test the null hypothesis of stochastic comovement, although the limiting distribution under the null is different. Let  $\hat{\xi}_t$  be the residual from the regression of  $X_{2,t}$  on  $X_{1,t}$  and a constant term.

*Theorem 3.3.* (i) Assume that (3.3), and A1–A4(i), A5 hold, and that  $E(v_{1,t}) \neq 0$ . Then,

$$\frac{1}{\hat{s}_{l_T}^2} \frac{1}{T^2} \sum_{t=1}^T \left( \sum_{j=1}^t \hat{\xi}_t \right)^2 \Rightarrow \int_0^1 Q_s^2 ds$$

where

$$Q_s = (W_s - sW_1) - 6 \left( \int_0^1 s dW_s - 1/2W_1 \right) (s^2 - s)$$

and

$$\hat{s}_{l_T}^2 = \frac{1}{T} \sum_{t=1}^T \xi_t^2 + \frac{2}{T} \sum_{t=1}^{l_T} \left( 1 - \frac{t}{l_T + 1} \right) \sum_{j=t+1}^T \hat{\xi}_j \hat{\xi}_{j-t}$$

where  $l_T = o(T^{1/2})$ .

(ii) Assume that (3.2), and A1–A3, A4(ii) hold. Then,

$$P \left[ \frac{1}{\hat{s}_{l_T}^2} \frac{1}{T^2} \sum_{t=1}^T \left( \sum_{j=1}^t \hat{\xi}_t \right)^2 > C_T \right] \rightarrow 1 \quad \text{as } T \rightarrow \infty,$$

where  $C_T \rightarrow \infty$  and  $C_T l_T / T \rightarrow 0$ , as  $T \rightarrow \infty$ .

From part (i) of Theorem 3.3 note that the asymptotic distribution under the null is a functional of only one Brownian motion  $W$ , where  $W$  is the weak limit, property rescaled, of the partial sums of  $v_{2,t} - E(v_{2,t}) = \gamma'_0 X_t - E(\gamma'_0 X_t)$ . Thus, the asymptotic behavior of the statistic is not affected by whether  $v_{1,t}$  and  $v_{2,t}$  in (3.7) and (3.8) are correlated or not. As mentioned above, this is due to the fact that the asymptotic behavior of  $X_{1,t}$  is dominated by its trend component. This differs from the linear case,  $g_0 = 0$ , in which the non-zero correlation between  $v_{1,t}$  and  $v_{2,t}$  results in a nuisance parameter in the limiting distribution (when  $\beta_0$  is estimated by OLS). The asymptotic critical values for the distribution given in Theorem 3.3(i) are reported below. The simulated critical values are based on sample size  $n = 2000$  and 20,000 replications.

## Linear stochastic comovement test critical values

Nominal size					
1%	5%	10%	90%	95%	99%
0.219	0.150	0.121	0.028	0.023	0.018

Note that our critical values are smaller than those reported in Shin (1994) for his  $C_\mu$  test statistic. This difference arises because of the nonlinear component in (3.3).

In practice we do not know whether  $E(v_{1,t}) = 0$  or not. If  $E(v_{1,t}) \neq 0$ , then the critical values tabulated above should be used. On the other hand, if  $E(v_{1,t}) = 0$ , then the asymptotic distribution of our test is the same as that of  $C_\mu$  in Shin (1994, Theorem 1), provided  $E(v_{1,t}v_{2,t}) = 0$ . Given these facts, we suggest applying the comovement test in the following way. If the test statistic is less than our critical value, accept the null of (cointegration) comovement. On the other hand if the test statistic is above Shin's critical value for  $C_\mu$ , we have evidence of the absence of cointegration (comovement). If the test statistic falls in the region between the two critical values (say the 'intermediate region'), construct the following statistic:  $d_T = T^{-1/2} \sum_{t=1}^T \hat{\sigma}_T^{-1} \Delta X_{1,t}$ , where  $\hat{\sigma}_T^2$  is a consistent estimator of the long run variance of  $T^{-1/2} \sum \Delta X_{1,t}$ . Under cointegration (comovement),  $d_T \Rightarrow N(0, 1)$  when  $E(v_{1,t}) = 0$ , and diverges at rate  $T^{1/2}$  when  $E(v_{1,t}) \neq 0$ . On the other hand, when there is no cointegration (comovement), we must distinguish between three cases. First, assume that  $E(v_{1,t}) = 0$ . Here, there are two cases: (i)  $E(v_{1,t}) = 0$  because there is no nonlinear component; or (ii)  $E(v_{1,t}) = 0$  because there is zero mean nonlinear component. Under (i),  $d_T$  is normally distributed (as  $v_{1,t}$  is a zero mean mixing process). Under (ii), the statistic is not normally distributed in general. Finally, assume that  $E(v_{1,t}) \neq 0$ . In this case, the estimator of the variance term in  $d_T$  diverges at a rate less than or equal to  $l_T^{1/2}$ , while the numerator of  $d_T$  diverges at a faster rate. These facts suggest a procedure for testing for cointegration (comovement) when the test statistic is in the 'intermediate region'. In particular, if we reject  $E(v_{1,t}) = 0$ , reject the null of cointegration (comovement). If we fail to reject  $E(v_{1,t}) = 0$ , then use Shin's critical values. (In order to use Shin's critical values in this case, an efficient estimator of the cointegrating vector should be constructed, as in Shin (1994).) Note that the use of  $d_1$  may result in the false rejection of  $E(v_{1,t}) = 0$  when there is no cointegration (comovement) and the nonlinear component has zero mean. However, in this case we still reject the null of cointegration (comovement), which is the correct inference.

#### 4. Testing for nonlinear cointegration

In this section we propose a test for the null of linear cointegration (comovement) against the alternative of nonlinear cointegration (comovement).

A number of papers which examine nonlinear cointegration have recently appeared. For example, Balke and Fomby (1997) propose a test for threshold cointegration. Also, Granger (1995) suggests testing for the null of linear cointegration by regressing the residuals from a standard cointegrating regression on their lagged values and a nonlinear function, and then performing a Lagrange Multiplier (LM) type test. A similar approach is examined by Swanson (1999) who regresses the first differences of the data on their lagged values and on a polynomial function of the cointegrating vector. He shows, using Monte Carlo experiments, that such tests have good power when  $g_0$  is a logistic cdf. Nevertheless, this class of tests does not have unit asymptotic power against general nonlinear alternatives. One reason for this is the LM tests are implemented using polynomial test functions, and the use of polynomials does not ensure test consistency.

In our test, cointegration is maintained under both the null and the alternative hypothesis. Let  $\hat{\eta}_t$  be the residual and  $\hat{\psi}$  be the slope coefficient from the least squares regression of  $\hat{\gamma}'_T X_t$  on  $\hat{\gamma}'_T X_{t-1}$  and a constant, where  $\hat{\gamma}'_T = (-\hat{\beta}_T, 1)$  and  $\hat{\beta}_T$  is the coefficient from the regression on  $X_{2,t}$  on  $X_{1,t}$  and a constant. Thus,

$$\hat{\eta}_t = \left( \hat{\gamma}'_T X_t - \frac{1}{T} \sum_{i=1}^T \hat{\gamma}'_T X_i \right) - \hat{\psi}_T \left( \hat{\gamma}'_T X_{t-1} - \frac{1}{T} \sum_{i=1}^T \hat{\gamma}'_T X_{i-1} \right). \tag{4.1}$$

Under the null of no nonlinearity, we have  $\sqrt{T}(\hat{\psi}_T - \psi_0) = O_p(1)$  and  $T(\hat{\beta}_T - \beta_0) = O_p(1)$ . Let

$$\eta_t = \left( \gamma'_0 X_t - \frac{1}{T} \sum_{i=1}^T \gamma'_0 X_i \right) - \psi_0 \left( \gamma'_0 X_{t-1} - \frac{1}{T} \sum_{i=1}^T \gamma'_0 X_i \right), \tag{4.2}$$

so that  $\eta_t$  is uncorrelated with any function of  $\gamma'_0 X_{t-1}$ . Under the alternative of nonlinearity of the type described in Section 2,  $\sqrt{T}(\hat{\psi}_T - \psi_*) = O_p(1)$ , where  $\psi_* \neq \psi_0$ ,  $T^{3/2}(\hat{\beta}_T - \beta_0) = O_p(1)$ , and

$$\eta_t = \left( \gamma'_0 X_t - \frac{1}{T} \sum_{\tau=1}^T \gamma'_0 X_\tau \right) - \psi_* \left( \gamma'_0 X_{t-1} - \frac{1}{T} \sum_{\tau=1}^T \gamma'_0 X_\tau \right).$$

Under the alternative,  $\eta_t$  includes the neglected nonlinear term  $(g_0(\theta'_2 X_{t-1}) - E(g_0(\theta'_2 X_{t-1})))$ , where  $\theta_2^{-1} \theta'_0 \equiv \gamma'_0 = (-\theta_2^{-1} \theta_1, 1)$ . Thus,  $\eta_t$  is correlated with some function of  $\gamma'_0 X_{t-1}$ . If we use as a test function, call it  $g$ , an exponential, as in Bierens (1990), or any other generically comprehensive test function, as described in Stinchcombe and White (1998, Section 3, hereafter SW), then under the alternative

$$E(\eta_t(g_0(\gamma'_0 X_{t-1} \tau) - E(g_0(\gamma'_0 X_{t-1} \tau)))) \neq 0$$

for all  $\tau \in \mathcal{T}$ , a subset of  $R$  whose complement  $\mathcal{T}^c$  has Lebesgue measure zero and is not dense in  $R$ . Given the  $\sqrt{T}$  consistency of  $\hat{\psi}_T$  and the  $T^{3/2}$  consistency of  $\hat{\beta}_T$  under the alternative, we also have

$$\frac{1}{T} \sum_{\tau=1}^T \left( \hat{\eta}_t \left( g(\hat{\gamma}'_T X_{t-1} \tau) - \frac{1}{T} \sum_{\tau=1}^T g(\hat{\gamma}'_T X_{t-1} \tau) \right) \right) \rightarrow M_\tau \quad \text{in Prob.}$$

where  $M_\tau \neq 0$ , for all  $\tau \in \mathcal{T}$ .

According to Theorem 3.10 in Stinchcombe and White (1998), if  $g$  is a real analytic function, then  $g$  delivers a consistent test, regardless of  $g_0$ , provided that  $g$  is not a polynomial. One natural choice for  $g$  is the logistic cdf, as it is a non-polynomial real analytic function.

As the parameter  $\tau$  is not identified under the null hypothesis, our tests falls into the class of tests with nuisance parameters present only under the alternative. Thus, although the asymptotic size of the test statistic is not affected by the actual value of  $\tau$ , the finite sample size will be affected, while the power will be affected both in finite samples and asymptotically. Consider the following two DGPs:

$$\Delta X_t = \Phi X_{t-1} + \varepsilon_t \tag{4.3}$$

and

$$\Delta X_t = \Phi X_{t-1} + g_0(\theta'_0 X_{t-1}) + \varepsilon_t, \tag{4.4}$$

where  $\theta_0 = \theta_2 \gamma_0$ . Assume that  $\Phi$  satisfies A4(i),  $g_0$  satisfies A3, and  $\varepsilon_t$  satisfies A1 and A2. From (4.3) note that

$$\Delta \gamma'_0 X_t = \delta \gamma'_0 X_{t-1} + \gamma'_0 \varepsilon_t,$$

and from (4.4) note that

$$\Delta \gamma'_0 X_t = \delta \gamma'_0 X_{t-1} + \gamma'_0 g_0(\theta'_0 X_{t-1}) + \gamma'_0 \varepsilon_t,$$

where  $\delta = -\alpha_1 \theta_1 + \alpha_2 \theta_2$ . The proposed test is based on the statistic:

$$\frac{1}{\sqrt{T}} \sum_{t=2}^T (g(\hat{\gamma}'_T X_{t-1} \tau) - \bar{g}) \hat{\eta}_t, \tag{4.5}$$

where  $\bar{g} = (1/T) \sum_{t=2}^T g(\hat{\gamma}'_T X_{t-1} \tau)$ . It is shown in the proof of Theorem 4.1 that (4.5) can be written as

$$\frac{1}{\sqrt{T}} \sum_{t=2}^0 ((g(\hat{\gamma}'_T X_{t-1} \tau) - \bar{g}) - M_{(\gamma'_0 X)}^{-1} M_{\gamma'_0 X} (\gamma'_0 X_{t-1} - \gamma'_0 \bar{X})) \eta_t + o_p(1), \tag{4.6}$$

where  $M_{(\gamma'_0 X)^2} = E((\gamma'_0 X_t - \gamma'_0 \bar{X})^2)$ ,  $M_{\gamma'_0 X_g} = E((\gamma'_0 X_t - \gamma'_0 \bar{X})(g(\gamma'_0 X_{t-1}\tau)\bar{g}))$ ,  $\bar{g} = (1/T)\sum_{t=2}^T g(\gamma'_0 X_{t-1}\tau)$ : From the central limit theorem for strong mixing processes (e.g. White, 1984, p. 124) and the asymptotic equivalence lemma, note that the limiting distribution of (4.6), when scaled by  $\sigma_T^2$  is a zero mean normal, where

$$\sigma_T^2 = \text{Var}\left(T^{-1/2} \sum_{t=2}^T ((g(\gamma'_0 X_{t-1}\tau) - E(g)) - M_{(\gamma'_0 X)^2}^{-1} M_{\gamma'_0 X_g} (\gamma'_0 X_{t-1} - \gamma'_0 \bar{X})) \eta_t\right), \tag{4.7}$$

and  $\sigma_T^2 \rightarrow \sigma_0^2$  as  $T \rightarrow \infty$ . A convenient estimator for  $\sigma_0^2$  is given by

$$\hat{\sigma}_{l_T}^2 = \frac{1}{T} \sum_{t=2}^T (\hat{\phi}_t^\tau)^2 + \frac{2}{T} \sum_{t=2}^{l_T} \left(1 - \frac{t}{l_T + 1}\right) \sum_{j=t+1}^T \hat{\phi}_t^\tau \hat{\phi}_{j-t}^\tau, \tag{4.8}$$

where

$$\begin{aligned} \hat{\phi}_t^\tau &= ((g(\hat{\gamma}'_T X_{t-1}\tau) - \bar{g})) \\ &- \left(\frac{1}{T} \sum_{t=2}^T (\hat{\gamma}'_T X_{t-1} - \hat{\gamma}'_T \bar{X})^2\right)^{-1} \left(\frac{1}{T} \sum_{t=2}^T (\hat{\gamma}'_T X_{t-1} - \hat{\gamma}'_T \bar{X})(g(\hat{\gamma}'_T X_{t-1}\tau) - \bar{g})\right) \\ &\times (\hat{\gamma}'_T X_{t-1} - \hat{\gamma}'_T \bar{X}) \hat{\eta}_t, \end{aligned} \tag{4.9}$$

with  $l_T = o(T^{1/2})$ . Assume also that

A6. The test function  $g$  is a real nonpolynomial analytic function, with bounded first two derivatives.

Note that various sigmoidal functions (e.g. the logistic cdf) satisfy A6.

*Theorem 4.1.* (i) Assume that (4.3), A1–A2, A4(i), and A6 hold. Define

$$m_T^\tau \equiv \frac{1}{\sqrt{T}} \frac{1}{\hat{\sigma}_{l_T}} \sum_{t=2}^T \hat{\eta}_t (g(\hat{\gamma}'_t X_{t-1}\tau) - \bar{g}).$$

Then

$$(m_T^\tau)^2 \Rightarrow \chi_1^2$$

for each  $\tau \in \mathcal{T}$ , and  $\hat{\sigma}_{l_T}^2$  defined in (4.8). The same result follows when  $g_0$  is a constant.

(ii) For DGP (4.4), suppose that A1–A4(ii), A5–A6 hold, and that  $g_0$  is not a constant. Then for each  $\tau \in \mathcal{T}$ ,

$$P[|m_T^\tau|^2 > C_T] \rightarrow 1 \quad \text{as } T \rightarrow \infty,$$

where  $C_T \rightarrow \infty$  and  $C_T l_T / T \rightarrow 0$ , as  $T \rightarrow \infty$ .

Note that strict stationarity is required only under the alternative, as the first moment of (4.3) is constant. In fact, in Theorem 4.1(ii), as in Theorems 3.1(i), 3.2 and 3.3(i) above, we impose A5 simply because it implies the constancy of the first moment. Also, note that the case where  $g_0$  is constant is covered by the null hypothesis.

In the current context we do not provide a ‘sup’ type result, as in Bierens (1990) and Stinchcombe and White (1998), for example. The intuitive reason for this is that the  $o_p(1)$  term in (4.6) holds pointwise in  $\tau$ , but not necessarily uniformly in  $\tau$ . As will become clear in the proof of Theorem 5.1, this is due to the fact that in the nonstationary case, we cannot invoke the usual uniform law of large numbers. We appeal instead to invariance principles and to results on convergence to stochastic integrals that hold pointwise in  $\tau$ , but not necessarily uniformly.

As mentioned above, although the asymptotic size of the test statistic is not affected by the choice of a particular  $\tau$ , the finite sample size and power are affected. There are at least two ways of addressing this issue. First, we can construct the statistic for different  $\tau$ 's and apply Bonferroni type bounds as in Lee et al. (1993), for example. Second, let  $\tau_1, \dots, \tau_p$  be chosen according to a particular design (e.g. randomly), and let  $\hat{G}$  be a consistent estimator of  $\text{Cov}(\phi_i^\tau, \phi_i^\tau)$ , so that  $\hat{G}$  is the matrix whose  $i, k$  element is given by

$$\frac{1}{T} \sum_{t=1}^T \hat{\phi}_i^\tau \hat{\phi}_i^{\tau_k} + \frac{2}{T} \sum_{t=1}^{l_T} \left(1 - \frac{t}{l_T + 1}\right) \sum_{j=t+1}^T \hat{\phi}_j^\tau \hat{\phi}_j^{\tau_k}$$

where  $\hat{\phi}_i^\tau$  is defined in (4.9), and  $\phi_i^\tau$  is defined as in (4.9), but with  $\hat{\gamma}'_T$  replaced with  $\gamma'_0$ . Then,  $(m_T^{\tau_1}, \dots, m_T^{\tau_p})' \hat{G}^{-1} (m_T^{\tau_1}, \dots, m_T^{\tau_p}) \Rightarrow \chi_p^2$ , for arbitrary and finite  $p$ . Whether the above result also holds for  $p = p_T$ , with  $p_T \rightarrow \infty$  at an appropriate rate as  $T \rightarrow \infty$ , is left for future research.

### 5. Monte Carlo results

In this section, a summary of Monte Carlo experiments based on the above test, and for samples of 100, 250, and 500 observations, is given. For the sake of brevity, much of the discussion focuses on the nonlinear cointegration (NLICI) test.

Before turning to our finite sample NLICI test results, it is worth reiterating that unit root and cointegration tests are not generally robust to the inclusion of

nonlinearities. As noted above, however, it turns out that the Kwiatkowski et al. (1992) stationarity and the Shin (1994) cointegration tests have well-defined limiting distributions and unit asymptotic power, in our context. The same cannot be said for the augmented Dickey–Fuller (ADF) unit root test. In particular, note that under the null hypothesis of a unit root, if we neglect to account for the nonlinear component, the error term is given by  $v_t = \varepsilon_t + g_0(\cdot)$ , where  $g_0(\cdot)$  is generally not a strong mixing process. Thus, standard unit root asymptotics no longer apply, and the usual limiting distribution of the ADF test statistic is typically no longer valid (see, e.g. Ermini and Granger, 1993). On the other hand, under the alternative of no unit root, the error term is a strong mixing process, so that we expect ADF tests to have reasonable power in large samples.

Table 1 reports the results from a Monte Carlo experiment based on the ADF test, using data generated according to

$$X_t = a + bX_{t-1} + cg_0(X_{t-1}) + \varepsilon_t,$$

where  $X_t$  is a scalar,  $\varepsilon_t$  is a scalar  $IN(0, 1)$  random variable,  $g_0(\cdot)$  is the logistic cdf, and  $a = 0$  (results for  $a \neq 0$  and for different  $g_0(\cdot)$  are qualitatively similar, and are available upon request from the authors). Notice that in the table,  $b$  varies from  $-0.9$  to  $1.0$ , so that empirical power of a variety of different parameterizations, and empirical size ( $b = 1$ ) is reported. Also, note that the parameter  $c$  is alternately  $-0.5$ ,  $-0.1$ ,  $0.1$ , and  $0.5$ . Here and below, results based on 5% nominal size tests are reported (results for 10% size tests are similar, and are not included for the sake of brevity). The finite sample power of the ADF test is good (power is always close to or equal to unity, except when  $|b| = 0.9$ ), as expected. The finite sample size of the ADF test is between 0.084 and 0.056, even for samples of 100 observations, and improves as we move from 100 to 500 observations, when  $\hat{\tau}_t$  is used. This suggests that within our context, the ADF test can still be used to signal the presence of a unit root, even when a bounded nonlinear component is added to the DGP, as long as  $\hat{\tau}_t$  is used. This is perhaps not surprising, as the mean of  $g_0(\cdot)$  is not generally zero. In summary, one might argue in favour of using  $\hat{\tau}_t$  in our context, as the finite sample size is closer to the nominal size than when  $\hat{\tau}$  and  $\hat{\tau}_\mu$  are used. Further, the finite sample power of the  $\hat{\tau}_t$  test is comparable to the power associated with the use of  $\hat{\tau}$  and  $\hat{\tau}_\mu$ , except when  $b = 0.9$ .

Table 2 reports the results from a Monte Carlo experiment based on the Johansen cointegration test, using data generated according to

$$\Delta X_t = d + eZ_{t-1} + fg_0(Z_{t-1}) + \varepsilon_t, \tag{5.1}$$

where  $X_t = (X_{1,t}, X_{2,t})'$  is a  $2 \times 1$  vector,  $\varepsilon_t$  is a  $2 \times 1$  vector whose components are distributed  $IN(0,1)$ ,  $Z_t = -X_{2,t}$  if  $e_1 = e_2 = 0$ , otherwise  $Z_t = X_{1,t} - X_{2,t}$ ,  $g_0(x) = (2/[1 + e^{-x}]) - 1$ ,  $d = (d_1, d_2)'$ ,  $d_1 = d_2 = 0.2$ ,  $e = (e_1, e_2)'$ , and



Table 1  
Augmented Dickey–Fuller test performance under neglected nonlinearity<sup>a</sup>

b	c	T = 100			T = 250			T = 500		
		$\hat{\tau}$	$\hat{\tau}_\mu$	$\hat{\tau}_\tau$	$\hat{\tau}$	$\hat{\tau}_\mu$	$\hat{\tau}_\tau$	$\hat{\tau}$	$\hat{\tau}_\mu$	$\hat{\tau}_\tau$
1.0	-0.5	0.056	0.086	0.084	0.054	0.103	0.086	0.050	0.115	0.098
1.0	-0.1	0.089	0.067	0.071	0.109	0.061	0.071	0.121	0.066	0.067
1.0	-0.5	0.016	0.053	0.067	0.009	0.032	0.057	0.003	0.027	0.054
1.0	-0.5	0.001	0.015	0.067	0.000	0.006	0.056	0.000	0.005	0.057
-0.9	-0.5	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
-0.6	-0.5	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
-0.3	-0.5	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
0.3	-0.5	1.000	0.999	0.996	1.000	1.000	1.000	1.000	1.000	1.000
0.6	-0.5	0.996	0.994	0.985	1.000	1.000	1.000	1.000	1.000	1.000
0.9	-0.5	0.737	0.727	0.529	0.998	0.999	0.994	1.000	1.000	1.000
-0.9	-0.1	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
-0.6	-0.1	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
-0.3	-0.1	1.000	1.000	0.999	1.000	1.000	1.000	1.000	1.000	1.000
0.3	-0.1	1.000	0.999	0.995	1.000	1.000	1.000	1.000	1.000	1.000
0.6	-0.1	0.999	0.989	0.979	1.000	1.000	1.000	1.000	1.000	1.000
0.9	-0.1	0.821	0.443	0.286	1.000	0.983	0.912	1.000	1.000	1.000
-0.9	0.1	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
-0.6	0.1	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
-0.3	0.1	1.000	1.000	0.999	1.000	1.000	1.000	1.000	1.000	1.000
0.3	0.1	1.000	0.999	0.993	1.000	1.000	1.000	1.000	1.000	1.000
0.6	0.1	1.000	0.985	0.973	1.000	1.000	1.000	1.000	1.000	1.000
0.9	0.1	0.633	0.286	0.185	0.997	0.906	0.712	1.000	1.000	0.997
-0.9	0.5	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
-0.6	0.5	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
-0.3	0.5	1.000	1.000	0.999	1.000	1.000	1.000	1.000	1.000	1.000
0.3	0.5	0.999	0.998	0.990	1.000	1.000	1.000	1.000	1.000	1.000
0.6	0.5	0.972	0.973	0.948	1.000	1.000	1.000	1.000	1.000	1.000
0.9	0.5	0.022	0.280	0.184	0.080	0.783	0.606	0.690	0.995	0.978

<sup>a</sup>Based on the augmented Dickey–Fuller (ADF) test, entries correspond to the empirical frequency of rejection of the null hypothesis of a unit root. Three versions of the test regressions are run: with no constant or linear deterministic trend ( $\hat{\tau}$ ), with a constant only ( $\hat{\tau}_\mu$ ), and with a constant and a linear deterministic trend ( $\hat{\tau}_\tau$ ). Data are generated according to the following process:  $X_t = a + bX_{t-1} + c\theta_0(X_{t-1}) + \varepsilon_t$ , where  $X_t$  is a scalar,  $\varepsilon_t$  is a scalar  $IN(0, 1)$  random variable,  $\theta_0(\cdot)$  is the logistic cdf, and  $a = 0.0$ . The first four rows of entries in the table report the empirical size of the test based on a 5% nominal size, while the remaining rows report the empirical power, also for a 5% nominal size test. All experiments are repeated for samples of  $T = 100, 250$ , and 500 observations. Results are based on 5000 Monte Carlo replications.

Table 2  
Johansen cointegration test performance under neglected nonlinearity<sup>a</sup>

$e_1$	$e_2$	$f_1$	$f_2$	$T = 100$		$T = 250$		$T = 500$	
				Trace 1	Trace 2	Trace 1	Trace 2	Trace 1	Trace 2
0.0	0.0	-2.0	0.0	0.720	0.086	0.748	0.084	0.743	0.074
0.0	0.0	-5.0	0.0	0.683	0.080	0.744	0.076	0.764	0.075
0.0	0.0	-2.0	2.0	0.287	0.957	0.016	0.935	0.000	0.947
0.0	0.0	-5.0	5.0	0.282	0.955	0.014	0.936	0.000	0.949
-0.2	0.2	-2.0	0.0	0.449	0.967	0.074	0.938	0.003	0.949
-0.2	0.4	-2.0	0.0	0.344	0.959	0.032	0.936	0.000	0.949
-0.2	0.6	-2.0	0.0	0.300	0.956	0.020	0.937	0.000	0.948
-0.2	0.2	-5.0	0.0	0.471	0.969	0.093	0.941	0.004	0.944
-0.2	0.4	-5.0	0.0	0.355	0.956	0.039	0.933	0.001	0.946
-0.2	0.6	-5.0	0.0	0.342	0.960	0.031	0.940	0.000	0.953
-0.2	0.2	-2.0	2.0	0.280	0.955	0.014	0.937	0.000	0.949
-0.2	0.4	-2.0	2.0	0.286	0.955	0.015	0.934	0.000	0.950
-0.2	0.6	-2.0	2.0	0.325	0.961	0.023	0.939	0.001	0.950
-0.2	0.2	-5.0	5.0	0.285	0.895	0.014	0.937	0.000	0.950
-0.2	0.4	-5.0	5.0	0.330	0.957	0.023	0.936	0.000	0.945
-0.2	0.6	-5.0	5.0	0.374	0.958	0.042	0.933	0.001	0.943

<sup>a</sup>Based on the Johansen trace test statistic, entries correspond to the empirical frequency of rejection of the null hypothesis no cointegration, in favor of a cointegrating space rank of unity. Two versions of the test statistic are constructed: with no constant or linear deterministic trend (Trace 1), and with a drift and linear deterministic trend in the levels, and a drift in the differences (Trace 2: this version of the test corresponds to Case 1 in Osterwald-Lenum (1992)). Data are generated according to the following process:  $\Delta X_t = d + eZ_{t+1} + fg_0(Z_{t-1} + \varepsilon_t)$ , where  $X_t = (X_{1,t}, X_{2,t})'$  is a  $2 \times 1$  vector,  $\varepsilon_t$  is a  $2 \times 1$  vector whose components are distributed  $IN(0, 1)$ , and  $Z_t = -X_{2,t}$  if  $e_1 = e_2 = 0$ , otherwise  $Z_t = X_{1,t} - X_{2,t}$ . Also,  $g_0(x) = (2/[1 + e^{-x}]) - 1$ ,  $d = (d_1, d_2)'$ ,  $d_1 = d_2 = 0.2$ ,  $e = (e_1, e_2)'$ , and  $f = (f_1, f_2)'$ . The first four rows of entries in the table report the empirical size of the test based on a 5% nominal size, while the remaining rows report the empirical power, also for a 5% nominal size test. All experiments are repeated for samples of  $T = 100, 250$ , and 500 observations. Results are based on 5000 Monte Carlo replications.

$f = (f_1, f_2)'$ . The values used for  $e$  are  $e_1 = e_2 = 0$  (empirical size), and  $e_1 = -0.2, e_2 = \{0.2, 0.4, 0.6\}$  (empirical power). This DGP is the same as that used in Park and Ogaki (1991), except that we also include a nonlinear component. Note that in Table 2 it is clear that the empirical power of the Johansen test is quite good (always above 0.895, even for samples of only 100 observations) only for the *Trace 2* test, which includes an intercept in the differenced vector autoregression (the intercept in the DGPs is nonzero). However, even for the *Trace 2* test, the empirical size is only *relatively* close to the nominal size (e.g. 0.086 for  $T = 100$ , and lower for higher values of  $T$ ) when the nonlinear component enters only one of the equations in the system (i.e.  $f_2 = 0$ ). Thus, the Johansen test performs more poorly when the complexity of the nonlinearity in

(5.1) is increased. This is perhaps not too surprising, given that the Johansen test is not valid in our context.

We now turn to a discussion of our results based on the NLCI test. In order to illustrate the performance of our test statistic under various scenarios, the results of three different experiments are reported. In all cases, the nonlinear function used in the construction of the NLCI test is  $g(x) = (2/(1 + e^{-x})) - 1$ . The data are generated according to (5.1), with

- Table 3:  $g_0(x) = g(x)$ ,  $d_1 = d_2 = 0.2$ ,  $f_1 = \{0, -2.0, -5.0\}$ ,  $f_2 = \{0, 2.0, 5.0\}$ ,  
 Table 4:  $g_0(x) = \sin(x)$ ,  $d_1 = d_2 = 0.2$ ,  $f_1 = \{0, -1.0, -2.0\}$ ,  $f_2 = \{0, 1.0, 2.0\}$ ,  
 Table 5:  $g_0(x) = \sin(x)$ , if  $|x| \leq \pi/2$ ,  $g_0(x) = g(x)$  if  $|x| > \pi/2$ ,  
 $d_1 = d_2 = 0.1$ ,  $f_1 = \{0, -2.0, -5.0\}$ ,  $f_2 = \{0, 2.0, 5.0\}$ .

Note that the experiment reported in Table 5 uses data which are generated according to two different forms of nonlinear error correction, depending on how far  $x$  is from the origin, and hence the DGP used is a type of threshold error correction model. However, note that in this case the nonlinear function is discontinuous at  $\pi/2$ , so that assumption A3 is not satisfied. Thus, the results in Table 5 can be interpreted as yielding evidence of the usefulness of the NLCI test for ‘modest’ departures from A3. Finally, various other parameterizations of the above DGPs were also examined and are omitted because the Monte Carlo results are similar. Also, the overall results did not change when  $\tau$  was varied. Thus, all reported results use  $\tau = 1$ . The results presented in Tables 3–5 are straightforward to interpret. For example, the finite sample power of the NLCI test is rather low for samples of 100 observations, and is lower when nonlinearity enters through only one equation (compare the last six rows of entries with the previous six rows, in each table). In particular, the finite sample power ranges from 0.105 to 0.487 across all DGPs, when  $f_2 = 0$  and  $l_T = 0$ . The power of the test increases, though, as the sample size increases, and for samples of 500 observations, the rejection frequency of the NLCI test has a lower bound of 0.836, across all parametrizations and DGPs, when  $l_T = 0$ . The empirical size of the test is reported in the first four rows of entries in Table 3. For  $l_T = 0$ , the empirical size ranges from 0.038 to 0.050 for 100 observations, and from 0.051 to 0.053 for 500 observations. Note also that for  $l_T3$ , the empirical size is low when 100 observations are used (the range is 0.018–0.024), but is much closer to the nominal size (the range of 0.045–0.047) when 500 observations are used.

## 6. Empirical illustration

Nonlinear models have been used in empirical studies with varying degrees of success in recent years. Examples of such models include smooth transition autoregressive models (Teräsvirta and Anderson, 1992), threshold autoregressive models (Pesaran and Potter, 1997) and Altissimo and Violante (1995),

Table 3  
Nonlinearity test performance:<sup>a</sup>  $g_0(x) = (2/(1 + e^{-x})) - 1$

$e$	$f_1$	$f_2$	$T = 100$			$T = 250$			$T = 500$		
			$l_T = l_{T1}$	$l_T = l_{T2}$	$l_T = l_{T3}$	$l_T = l_{T1}$	$l_T = l_{T2}$	$l_T = l_{T3}$	$l_T = l_{T1}$	$l_T = l_{T2}$	$l_T = l_{T3}$
0.2	0.0	0.0	0.038	0.034	0.018	0.053	0.052	0.044	0.055	0.054	0.047
0.4	0.0	0.0	0.046	0.038	0.024	0.052	0.045	0.037	0.053	0.051	0.045
0.6	0.0	0.0	0.050	0.038	0.023	0.048	0.042	0.037	0.051	0.048	0.045
0.2	-2.0	0.0	0.105	0.085	0.051	0.441	0.392	0.340	0.836	0.816	0.783
0.4	-2.0	0.0	0.108	0.084	0.050	0.514	0.460	0.396	0.891	0.879	0.847
0.6	-2.0	0.0	0.123	0.092	0.058	0.626	0.581	0.509	0.954	0.946	0.932
0.2	-5.0	0.0	0.192	0.254	0.176	0.822	0.826	0.795	0.998	0.998	0.998
0.4	-5.0	0.0	0.278	0.393	0.269	0.784	0.797	0.756	0.999	0.999	0.999
0.6	-5.0	0.0	0.465	0.568	0.408	0.744	0.770	0.711	0.999	0.999	0.998
0.2	-2.0	2.0	0.313	0.283	0.197	0.978	0.976	0.964	1.000	1.000	1.000
0.4	-2.0	2.0	0.315	0.307	0.221	0.980	0.978	0.970	1.000	1.000	1.000
0.6	-2.0	2.0	0.331	0.352	0.259	0.972	0.971	0.967	1.000	1.000	1.000
0.2	-5.0	5.0	0.321	0.304	0.220	0.971	0.969	0.960	1.000	1.000	1.000
0.4	-5.0	5.0	0.635	0.669	0.470	0.767	0.778	0.680	0.994	0.994	0.992
0.6	-5.0	5.0	0.615	0.634	0.423	0.805	0.806	0.710	0.992	0.992	0.988

<sup>a</sup>Based on the nonlinear cointegration test discussed above, entries correspond to the empirical frequency of rejection of the null hypothesis of linear cointegration, in favor of a finding of nonlinear cointegration. Test statistics are constructed for various sample sizes ( $T = 100, 250$ , and 500 observations), for 3 different values of the lag truncation parameter:  $l_1 = 0$ ,  $l_{T2} = \text{integer}[4(T/100)^{1/4}]$ , and  $l_{T3} = \text{integer}[12(T/100)^{1/6}]$ , and based on the nonlinear function:  $g(x) = [2/(1 + e^{-x})] - 1$ . Data are generated according to the following process:  $\Delta X_t = d + eZ_{t-1} + fg_0(Z_{t-1}) + \varepsilon_t$  where  $X_t = (X_{1,t}, X_{2,t})'$  is a  $2 \times 1$  vector,  $\varepsilon_t$  is a  $2 \times 1$  vector whose components are distributed  $\text{IN}(0, 1)$  and  $Z_t = X_{1,t} - X_{2,t}$ . Also,  $d = (d_1, d_2)'$ ,  $d_1 = d_2 = 0.2$ ,  $e = (e_1, e_2)'$ , and  $f = (f_1, f_2)'$ . The first three rows of entries in the table report the empirical size of the test based on a 5% nominal size, while the remaining rows report the empirical power, also for a 5% nominal size test. Results are based on 5000 Monte Carlo replications.

Table 4  
 Nonlinearity test performance:<sup>a</sup>  $g_0(x) = \sin(x)$

$e_2$	$f_1$	$f_2$	$T = 100$			$T = 250$			$T = 500$		
			$l_T = l_{T1}$	$l_T = l_{T2}$	$l_T = l_{T3}$	$l_T = l_{T1}$	$l_T = l_{T2}$	$l_T = l_{T3}$	$l_T = l_{T1}$	$l_T = l_{T2}$	$l_T = l_{T3}$
0.2	-2.0	0.0	0.341	0.287	0.184	0.808	0.749	0.699	0.984	0.976	0.970
0.4	-2.0	0.0	0.305	0.255	0.165	0.833	0.799	0.756	0.989	0.987	0.985
0.6	-2.0	0.0	0.311	0.269	0.171	0.867	0.850	0.819	0.993	0.993	0.992
0.2	-5.0	0.0	0.467	0.398	0.288	0.821	0.753	0.704	0.975	0.963	0.954
0.4	-5.0	0.0	0.483	0.423	0.321	0.874	0.841	0.817	0.988	0.985	0.983
0.6	-5.0	0.0	0.487	0.461	0.373	0.893	0.881	0.860	0.992	0.991	0.990
0.2	-2.0	2.0	0.510	0.423	0.313	0.853	0.795	0.747	0.986	0.977	0.971
0.4	-2.0	2.0	0.523	0.460	0.354	0.908	0.881	0.858	0.995	0.992	0.991
0.6	-2.0	2.0	0.555	0.522	0.430	0.933	0.924	0.910	0.995	0.994	0.994
0.2	-5.0	5.0	0.078	0.079	0.059	0.302	0.327	0.286	0.671	0.686	0.651
0.4	-5.0	5.0	0.179	0.179	0.134	0.524	0.532	0.493	0.864	0.870	0.851
0.6	-5.0	5.0	0.426	0.422	0.353	0.782	0.784	0.759	0.995	0.956	0.950

<sup>a</sup>See notes to Table 3. All rows report empirical power of the test.

Table 5  
 Nonlinearity test performance:<sup>a</sup>  $g_0(x) = \sin(x)$  if  $|x| \leq \pi/2$ , otherwise  $g_0(x) = [2/(1 + e^{-x})] - 1$

$e_2$	$f_1$	$f_2$	$T = 100$			$T = 250$			$T = 500$		
			$l_T = l_{T1}$	$l_T = l_{T2}$	$l_T = l_{T3}$	$l_T = l_{T1}$	$l_T = l_{T2}$	$l_T = l_{T3}$	$l_T = l_{T1}$	$l_T = l_{T2}$	$l_T = l_{T3}$
0.2	-2.0	0.0	0.146	0.121	0.080	0.521	0.493	0.442	0.923	0.919	0.903
0.4	-2.0	0.0	0.137	0.116	0.076	0.551	0.527	0.476	0.938	0.935	0.924
0.6	-2.0	0.0	0.142	0.120	0.076	0.589	0.561	0.504	0.953	0.949	0.939
0.2	-5.0	0.0	0.165	0.143	0.088	0.629	0.602	0.552	0.962	0.958	0.951
0.4	-5.0	0.0	0.235	0.223	0.160	0.770	0.768	0.744	0.981	0.981	0.979
0.6	-5.0	0.0	0.223	0.224	0.162	0.773	0.776	0.753	0.983	0.984	0.984
0.2	-2.0	2.0	0.334	0.314	0.223	0.860	0.856	0.838	0.993	0.993	0.992
0.4	-2.0	2.0	0.327	0.312	0.227	0.876	0.874	0.864	0.993	0.994	0.994
0.6	-2.0	2.0	0.339	0.336	0.249	0.879	0.879	0.864	0.995	0.995	0.995
0.2	-5.0	5.0	0.241	0.254	0.207	0.692	0.707	0.682	0.963	0.965	0.961
0.4	-5.0	5.0	0.304	0.404	0.302	0.805	0.832	0.807	0.983	0.987	0.985
0.6	-5.0	5.0	0.370	0.477	0.356	0.804	0.839	0.808	0.978	0.982	0.980

<sup>a</sup>See notes to Table 3.

nonlinear error correction models (Granger and Swanson, 1996), threshold error correction models (Balke and Fomby, 1997), and the references contained therein. In this section, we do not estimate new varieties of nonlinear models, but rather, we illustrate the use of the nonlinear error-correction test discussed above. In order to do this, we examine data on the term structure of interest rates, which have been kindly provided to us by Heather Anderson. The data consist of monthly nominal yield to maturity figures from the Fama Twelve Month Treasury Bill Term Structure File, for the period January 1970–December 1988. Six variables, denoted  $R1$ – $R6$ , are examined, and correspond to Treasury bills with one month to maturity, Treasury bills with two months to maturity, and so on, up to bills with 6 months to maturity. A detailed discussion of the data is given in Hall et al. (1992), as well as in Anderson (1997).

We consider three types of tests: (i) For the one-dimensional case, we construct both the ADF statistic for the null hypothesis of a unit root and the Kwiatkowski et al. (1992) test, described in Section 3, for the null of stationarity/ergodicity. (ii) For the two-dimensional case, we construct the Johansen ‘trace’ test statistic (1988, 1991) for the null of no cointegration, and the Shin (1994) test for the null of cointegration (comovement). For the latter cointegration test, we compare the results using both the critical values in Shin (1994) and the critical values reported above. (iii) Also for the two-dimensional case, we perform the test for nonlinear cointegration described in Section 5.

Test results are reported in Table 6. Panel A contains ADF and Kwiatkowski et al. (1992) test results, where  $l_T$  denotes the number of lags used in the computation of the estimated variance (see above). For the ADF test,  $\hat{\tau}_\mu$  is reported, although test regressions without a constant were also run for all variables, and our findings did not differ. Note that the outcomes of the ADF and the Kwiatkowski et al. (1992) tests agree. In particular, for all maturities, the unit root null hypothesis is not rejected (using the ADF test) while the null of stationarity ergodicity is consistently rejected (using the Kwiatkowski et al. (1992) test), regardless of the value of  $l_T$ .

Panel B reports results based on cointegration and comovement tests. For the sake of brevity, only bivariate combinations which include  $R1$  are reported on. Complete results are available from the authors. As mentioned above, the limiting distribution from Shin (1994) does not apply in the presence of neglected nonlinearity. Interestingly, even using the smaller critical values reported in Section 3 above, we fail to reject the null of stochastic comovement for 3 of 5 bivariate combinations, based on statistics constructed using  $l_T = 4$  and 8 (columns 5 and 6 of the table). Furthermore, for the other two bivariate combinations, use of the standard Shin (1994) critical values leads to a failure to reject at a 5% level (and in some cases a 1% level). These results agree with the theory posited by Hall et al. (1992) which suggests that any bivariate combination of our nominal interest rate series is cointegrated. Given these findings, it may be of interest to test the bivariate combinations for nonlinear

Table 6  
Empirical illustration: the term structure of interest rates<sup>a</sup>

Panel A: Unit root and stationarity ergodicity tests

Variable	ADF $\hat{\tau}_\mu$ (lags)	Kwiatkowski et al. (1992)								
		$l_T = 0$	1	2	3	4	5	6	7	8
R1	-1.60 (7)*	4.95*	2.55*	1.73*	1.33*	1.08*	0.92*	0.80*	0.71*	0.64*
R2	-1.61 (6)*	5.25*	2.68*	1.82*	1.39*	1.13*	0.95*	0.83*	0.73*	0.66*
R3	-1.64 (6)*	5.30*	2.71*	1.84*	1.40*	1.14*	0.96*	0.83*	0.74*	0.67*
R4	-1.67 (8)*	5.34*	2.72*	1.85*	1.41*	1.14*	0.97*	0.84*	0.74*	0.67*
R5	-1.68 (7)*	5.42*	2.76*	1.87*	1.43*	1.16*	0.98*	0.85*	0.75*	0.68*
R6	-1.43 (7)*	5.54*	2.82*	1.91*	1.46*	1.18*	1.00*	0.87*	0.77*	0.69*

Panel B: Cointegration and comovement tests

Variables	Johansen	Shin				NLCI			
		$l_T = 0$	1	4	8	$l_T = 0$	1	4	8
R1, R2	20.3*	0.44 <sup>e</sup>	0.41 <sup>e</sup>	0.38 <sup>e</sup>	0.34 <sup>e</sup>	4.16** (0.041)	3.78** (0.051)	3.52* (0.060)	2.88* (0.089)
R1, R3	24.6*	0.37 <sup>e</sup>	0.31 <sup>d</sup>	0.26 <sup>e</sup>	0.25 <sup>d</sup>	3.99** (0.045)	2.67* (0.102)	1.90 (0.168)	1.85 (0.173)
R1, R4	25.6*	0.25 <sup>d</sup>	0.21 <sup>d</sup>	0.17 <sup>e</sup>	0.16 <sup>c</sup>	3.55* (0.059)	3.00* (0.083)	2.61* (0.106)	2.45 (0.117)
R1, R5	27.8*	0.27 <sup>d</sup>	0.22 <sup>c</sup>	0.16 <sup>c</sup>	0.15 <sup>b</sup>	2.60* (0.106)	1.95 (0.162)	1.63 (0.201)	1.68 (0.194)
R1, R6	26.1*	0.40 <sup>e</sup>	0.30 <sup>d</sup>	0.20 <sup>c</sup>	0.18 <sup>c</sup>	2.50 (0.113)	1.85 (0.173)	1.52 (0.217)	1.62 (0.203)

<sup>a</sup>The data are monthly Treasury-Bill nominal yield to maturity figures for the period 1970 : 1–1988 : 12. R1 is the series for bills with one month to maturity, R2 is the series for bills with two months to maturity, and so on up until R6 which is the series for bills with 6 months to maturity. Panel A contains augmented Dickey–Fuller (ADF) and Kwiatkowski et al. (1992) test statistics (as discussed above). For the ADF tests, the ‘lag augmentations’ used is in brackets, chosen based on an examination of residual autocorrelations. All starred entries in Panel A correspond to evidence of a unit root (nonstationary-ergodicity) at the 5% level using critical values from Kwiatkowski et al. (1992) or MacKinnon (1991). In Panel B, the second column contains the Johansen (1988,1991) trace test statistics, where the associated vector autoregressions are estimated with a constant in the cointegrating relation, a linear deterministic trend in the data (results were the same without the deterministic trend), and 6 lags of each variable (similar results were found for the 12 lag case). Starred entries indicate rejection of the null hypothesis of no cointegration (in favor of cointegrating space rank of unity) using the 5% level critical value. The last 8 columns of Panel B contain Shin (1994) cointegration (comovement) and nonlinear error correction test statistics. For each of these two statistics, values are tabulated for  $l_T = 0, 1, 4, 8$ .

<sup>b</sup>For the Shin-type tests superscripts ‘b’ and ‘c’ denote failure to reject the null hypothesis of cointegration (comovement) using the 5% and 1% (respectively) critical values in Section 3 of the paper.

<sup>c</sup>is the same as footnote ‘b’ above.

<sup>d</sup>is the same as footnotes ‘b’ and ‘c’, but use the critical values of Shin (1994). For the nonlinear cointegration test (last 4 columns), values of the statistics,  $(m_j^*)^2$ , which is used in the modified Bonferroni bound of Hochberg (1988) defined as  $\alpha = \min_{j=1, \dots, m} (m-j+1)P_{(j)}$ , where  $P_{(j)}$  is the  $p$ -value of the test statistic, is reported. Here the values used for  $\tau$  are  $\tau = \{2.0, 5.0, 8.0, 10.0\}$ , so that  $m = 4$ . Modified Bonferroni bounds are given in brackets below statistic values. Rejection of the null of linear cointegration in favor of the alternative of nonlinear cointegration at a 5% and 10% size are denoted by superscripts \* and \*\*, respectively.

<sup>e</sup>is the same as footnote ‘d’ above.



error correction. The appropriate test statistics are reported in the last 4 columns of the table, with modified Bonferroni bounds given in brackets below statistic values (see footnote to Table 6). Entries superscripted  $^{**}$  denote rejection of the null hypothesis of linear cointegration (comovement) in favor of nonlinear cointegration at a 5% (10%) level. Thus, for the pair of series consisting of  $(R1, R2)$ , some evidence of nonlinear cointegration is found, regardless of the value of  $l_T$ . Weaker evidence (i.e. rejections for *some* of  $l_T$ ) of nonlinear cointegration is also found for  $(R1, R3)$ ,  $(R1, R4)$  and  $(R1, R5)$ . Based on a comparison of linear and nonlinear 1-step ahead forecast errors, Anderson (1997) finds evidence of nonlinear error correction among the variables considered here, consistent with our findings.

## 7. Conclusions

In this paper we introduce a class of nonlinear Markov processes characterized by the sum of a linear component plus a bounded nonlinear component. In the one-dimensional case, the ergodicity of the process is equivalent to the absence of a unitary or explosive root, and in the multidimensional case the existence of *linear stochastic comovement* is equivalent to the existence of cointegration.

We show that the statistic proposed by Kwiatkowski et al. (1992) has a well-defined limiting distribution under the null of general stationary-ergodic nonlinear processes, and has power not only against the alternative of integratedness, but also against the alternative of a more general nonlinear nonergodic process. We also show that the cointegration test statistic proposed by Shin (1994) is consistent, in our context, although the critical values of the test are quite different from those tabulated by Shin (1994) for the linear case. Finally, we propose a consistent test for the null hypothesis of linear cointegration against the alternative of nonlinear cointegration (NLCI). In a series of Monte Carlo experiments, we find that the NLCI test has good finite sample size and power. Further, in an illustration of the NLCI test in which we examine the term structure of interest rate, we find some evidence that bivariate models of interest rates of different maturities may be nonlinearly cointegrated.

## Acknowledgements

We are grateful to two anonymous referees, Filippo Altissimo, Frank Diebold, Rob Engle, Clive Granger, Jin Hahn, Ross Starr, Shinichi Sakata, Chor-Yiu Sin, Ruth Williams and to seminar participants at the University of Pennsylvania, Pennsylvania State University and the 1995 Winter Meetings of the Econometric Society for helpful comments on an earlier version of this

paper. We also owe special thanks to Herman Bierens for very useful suggestions, and to Heather Anderson for providing us with the data used in the illustration of our nonlinearity test. Swanson thanks the National Science Foundation and the Research and Graduate Studies Office at Pennsylvania State University for financial assistance. White's participation was supported by NSF grant SBR 9511253.

## Appendix A

*Proof of Proposition 2.1.* Note first that

$$\begin{aligned} X_t &= AX_{t-1} + (g_0(\theta_0^1 X_{t-1}, \dots, \theta_0^j X_{t-1}) - g_0(0)) + (\varepsilon_t + g_0(0)) \\ &= H(X_{t-1}) + (\varepsilon_t + g_0(0)). \end{aligned}$$

Assume that the associated deterministic system, say  $x_t$ , is given by  $x_t = H(x_{t-1})$ . The proposition follows from Theorem 4.3 of Tong (1990), once we have shown that his assumptions B1–B3 are satisfied. First, note that A1–A3 imply B2–B3. It remains to show that B1 is satisfied. As  $\lim_{\|x\| \rightarrow \infty} (g_0(x) - g_0(0))/\|x\| = 0$  from Theorem 1.3.5(a) in Kocic and Ladas (1993), it follows that  $H$  is asymptotically stable at large, so that B1 is satisfied.  $\square$

*Proof of Proposition 2.3.* (i) Let  $X_t = \sum_{j=0}^t A^{t-j} \varepsilon_j + \sum_{j=1}^t A^{t-j} g_0(\theta_0^1 X_{j-1}, \dots, \theta_0^j X_{j-1}) = \tilde{X}_t + \tilde{g}_{0,t}$ . As  $\tilde{X}_t$  is a linear component, from Johansen (1988) and from the Granger representation theorem (Engle and Granger, 1987), it follows that the Wold representation for  $\tilde{X}_t$  is  $\Delta \tilde{X}_t = C(L)\varepsilon_t$ . Using the Beveridge and Nelson (1981) decomposition it follows that  $\tilde{X}_t = C(1) \sum_{j=0}^t \varepsilon_j + C^*(L)\varepsilon_t$ , where  $C^*(L) = \sum_{j=0}^{\infty} (\sum_{i=j+1}^{\infty} C_i) L^j$ . Note also that  $\tilde{g}_{0,t-1} \leq O_p(T)$ , as it is the sum of  $T$  bounded components. Now,  $\tilde{X}_T/T^{1/2} \Rightarrow B_1$ , where  $B_1$  is a  $p$ -dimensional mean zero normal with covariance matrix equal to  $C(1)\Sigma C(1)'$ , and so is a non-degenerate random variable. If  $\tilde{g}_{0,T}/T^{1/2} \rightarrow \infty$  or  $\tilde{g}_{0,T}/T^{1/2} \rightarrow 0$ , then  $X_T \rightarrow \infty$  at rate  $T^{1/2}$  (if  $\tilde{X}_T$  is the component of higher order of probability) or at a rate faster than  $T^{1/2}$  (if the nonlinear component is of higher order than the linear component). Finally, consider the case in which  $\tilde{g}_{0,T}/T^{1/2} \Rightarrow G$ , where  $G$  is either a nondegenerate or a degenerate random variable. As  $B_1$  is a continuously distributed nondegenerate random variable,  $P(\omega: B_1(\omega) = -G(\omega)) = 0$ . The result follows directly.

(ii) The result follows from the fact that  $\theta_0^j X_t$  satisfies the assumptions of Proposition 2.1.  $\square$

*Proof of Proposition 2.4.* Using the arguments from the proof of Proposition 2.3, and by setting  $A = I$ , it follows that  $\|X_T\|$  diverges.

Furthermore,  $\theta'_0 X_t = \theta'_0 \sum_{j=1}^t \varepsilon_j + \theta'_0 \sum_{j=1}^t g_0(\theta_0^1, \dots, \theta_0^j X_{j-1})$ . Thus, noting that  $T^{-1/2}(\theta'_0 \sum_{j=1}^T \varepsilon_j) \Rightarrow \theta'_0 B_2$ , where  $B_2$  is a  $p$ -dimensional mean zero normal with covariance matrix equal to  $\Sigma$ , the result follows by the same argument used in the Proof of Proposition 2.3(i).  $\square$

*Proof of Theorem 3.1.* (i) First, note that

$$\begin{aligned} \frac{1}{\sqrt{T}} \sum_{j=1}^{[Tr]} (X_j - \bar{X}) &= \frac{1}{\sqrt{T}} \sum_{j=1}^{[Tr]} (X_j - a) - \frac{1}{\sqrt{T}} \sum_{j=1}^{[Tr]} (\bar{X} - a) \\ &= \frac{1}{\sqrt{T}} \sum_{j=1}^{[Tr]} (X_j - a) - \frac{1}{T} \sum_{j=1}^{[Tr]} \left( \frac{1}{\sqrt{T}} \sum_{t=1}^T (X_t - a) \right) \Rightarrow \sigma_X W_r - \sigma_X r W_1, \end{aligned}$$

where  $a = E(X_t)$  for all  $t$ ,  $\sigma_X^2 = \lim_{T \rightarrow \infty} \text{Var}(T^{-1/2} \sum_{t=1}^T (X_t - a)^2)$ , and  $W = (W_s, s \in [0, 1])$  is a standard Brownian motion. Also, by Fact 2.6,  $X_t$  is a strictly stationary mixing process, with mixing coefficients decaying at a geometric rate. Furthermore, given A2,  $X_t$  is fourth-order stationary. Thus, by Lemma 1 in Andrews (1991),  $X_t - E(X_t)$  satisfies his Assumption A. Given that  $\bar{X}$  is  $T^{1/2}$ -consistent for the true first moment, from Theorem 1(a) in Andrews (1991), it follows that  $\hat{s}_T^2 \rightarrow \sigma_X^2$  in *Prob.* The result then follows by the continuous mapping theorem.

(ii) We need to show that the numerator of  $S_T$ , say  $T^{-2} \sum_{t=1}^T (\sum_{j=1}^t (X_j - \bar{X}))^2$  explodes at a faster rate than does  $\hat{s}_T^2$ . First, consider the numerator. Under the alternative,  $X_t = \sum_{j=1}^t \varepsilon_j + \sum_{j=1}^t g_0(X_{j-1})$ . Thus,

$$\begin{aligned} \frac{1}{T^2} \sum_{t=1}^T \left( \sum_{j=1}^t (X_j - \bar{X}) \right)^2 &= \frac{1}{T^2} \sum_{t=1}^T \left( \sum_{j=1}^t \left( \sum_{i=1}^j (\varepsilon_i - \bar{\varepsilon}) \right. \right. \\ &\quad \left. \left. + \sum_{i=1}^j (g_0(X_{i-1}) - \bar{g}_0) \right) \right)^2 \\ &= \frac{1}{T^2} \sum_{t=1}^T \left( \sum_{j=1}^t \sum_{i=1}^j (\varepsilon_i - \bar{\varepsilon}) \right)^2 + \frac{1}{T^2} \sum_{t=1}^T \left( \sum_{j=1}^t \sum_{i=1}^j (g_0(X_{i-1}) - \bar{g}_0) \right)^2 \\ &\quad + \frac{2}{T^2} \sum_{t=1}^T \left( \sum_{j=1}^t \left( \sum_{i=1}^j (\varepsilon_i - \bar{\varepsilon}) \right) \left( \sum_{i=1}^j (g_0(X_{i-1}) - \bar{g}_0) \right) \right), \end{aligned}$$

where  $\bar{\varepsilon}$  and  $\bar{g}_0$  denote sample means. From Kwiatkowski et al. (1992, p. 168), we know that

$$\frac{1}{T^4} \sum_{t=1}^T \left( \sum_{j=1}^t \sum_{i=1}^j (\varepsilon_i - \bar{\varepsilon}) \right)^2 \Rightarrow \sigma_\varepsilon^2 \int_0^1 \left( \int_0^u W_s^\mu ds \right)^2 du,$$

where  $W_s^\mu = W_s - \int_0^1 W_\tau d\tau$ , and  $W$  is a standard Brownian motion. Now, as  $g_0$  is bounded, it follows that  $(1/T^4)\sum_{t=1}^T(\sum_{j=1}^t \sum_{i=1}^j (g_0(X_{i-1}) - \bar{g}_0))^2 = O_p(T)$ . Turning now to the denominator, note first from Kwiatkowski et al. (1992, p. 168), that it follows that

$$\frac{1}{Tl_T} \left( \frac{1}{T} \sum_{t=1}^T \left( \sum_{j=1}^t (\varepsilon_i - \bar{\varepsilon}) \right)^2 \right) + \frac{1}{Tl_T} \left( \frac{2}{T} \sum_{t=1}^{l_T} \left( 1 - \frac{t}{l_T + 1} \right) \sum_{j=t+1}^T \right. \\ \left. \times \left( \left( \sum_{i=1}^j (\varepsilon_i - \bar{\varepsilon}) \right) \left( \sum_{s=1}^{j-t} (\varepsilon_s - \bar{\varepsilon}) \right) \right) \right) \Rightarrow \sigma_\varepsilon^2 \int_0^1 (W_s^\mu)^2 ds.$$

In order to prove part (ii), it thus suffices to show that

$$\sum_{t=1}^T \left( \sum_{j=1}^t \sum_{i=1}^j (g_0(X_{i-1}) - \bar{g}_0) \right)^2 \sim T^2 \sum_{t=1}^T \left( \sum_{j=1}^t (g_0(X_{j-1}) - \bar{g}_0) \right)^2, \quad (A.1)$$

where  $\sim$  means ‘of the same order of probability’. Let  $M$  be the LHS of (A.1) and  $Q$  be the RHS of (A.1). Also, let  $g_{i,c} = g_0(X_i) - (1/T)\sum_{i=1}^T g_0(X_i)$ . Note that

$$M = \sum_{i=1}^T i^2 g_{1,c}^2 + \sum_{i=1}^{T-1} i^2 g_{2,c}^2 + \dots + \sum_{i=1}^{T-k} i^2 g_{k,c}^2 + \dots + g_{T,c}^2 \\ + 2 \sum_{i=1}^{T-1} i^2 g_{1,c} g_{2,c} + 2 \sum_{i=1}^{T-2} i^2 g_{1,c} g_{3,c} + \dots + 2 \sum_{i=1}^{T-k} i^2 g_{1,c} g_{k+1,c} \\ + \dots + 2 g_{1,c} g_{T,c} + \dots + 2 \sum_{i=1}^{T-k} i^2 g_{k,c} g_{k+1,c} + 2 \sum_{i=1}^{T-k-1} i^2 g_{k,c} g_{k+2,c} \\ + \dots + 2 g_{k,c} g_{T,c} + \dots + 2 g_{T-1,c} g_{T,c}.$$

Also,

$$Q = T g_{1,c}^2 + (T-1) g_{2,c}^2 + \dots + (T-k) g_{k,c}^2 + \dots + g_{T,c}^2 \\ + 2(T-1) g_{1,c} g_{2,c} + 2(T-2) g_{1,c} g_{3,c} + \dots + 2(T-k) g_{1,c} g_{k+1,c} \\ + \dots + 2 g_{1,c} g_{T,c} + \dots + 2(T-k) g_{k,c} g_{k+1,c} + 2(T-k-1) g_{k,c} g_{k+2,c} \\ + \dots + 2 g_{k,c} g_{T,c} + \dots + 2 g_{T-1,c} g_{T,c}.$$

By comparing  $M$  and  $Q$  term by term, observe that  $M \sim T^2 Q$ . By the same argument,

$$\begin{aligned} \frac{1}{T^2} \sum_{t=1}^T \left[ \sum_{j=1}^t \sum_{i=1}^j g_{i,c} \right]^2 &\sim l_T^{-1} \left[ \frac{1}{T} \left[ \sum_{t=1}^T \left( \sum_{j=1}^t g_{j,c} \right)^2 \right] \right. \\ &\quad \left. + \frac{2}{T} \left( \sum_{t=1}^{l_T} \left( 1 - \frac{t}{l_T + 1} \right) \sum_{j=t+1}^T \left( \sum_{i=1}^j g_{i,c} \right) \left( \sum_{i=1}^{j-t} g_{i,c} \right) \right) \right]. \end{aligned}$$

Thus, if the denominator explodes at rate  $T^{1+\eta} l_T$ ,  $\eta \in [0, 1]$ , the numerator will explode at rate  $T^{2+\eta}$ , and the ratio of the two will then explode at rate  $T l_T^{-1}$ . Note that  $\eta = 0$  is the case in which either the linear component is dominant or the two components, linear and nonlinear, are of the same order of probability. We have  $\eta > 0$  in the case in which the nonlinear component is dominant.  $\square$

*Proof of Theorem 3.2.* First, note that

$$T^{3/2}(\hat{\beta}_T - \beta_0) = \frac{(1/T^{3/2}) \sum_{t=1}^T (X_{1,t} - \bar{X}_1) u_{2,t}}{(1/T^3) \sum_{t=1}^T (X_{1,t} - \bar{X}_1)^2}$$

where  $u_{2,t} = v_{2,t} - \phi$ , and  $E(v_{1,t}) = \phi$ , for all  $t$ . Now,

$$\begin{aligned} \frac{1}{T^3} \sum_{t=1}^T (X_{1,t} - \bar{X}_1)^2 &= \frac{1}{T^3} \sum_{t=1}^T \left( \sum_{j=1}^t u_{1,j} - \frac{1}{T} \sum_{t=1}^T \left( \sum_{j=1}^t u_{1,t} \right) + \pi \left( t - \frac{T+1}{2} \right) \right)^2 \\ &= \frac{1}{T^3} \sum_{t=1}^T \left( \pi \left( t - \frac{T+1}{2} \right) \right)^2 + o_p(1), \end{aligned}$$

where  $u_{1,t} = v_{1,t} - \pi$ , and  $E(v_{1,t}) = \pi$ , for all  $t$ . Thus  $(1/T^3) \sum_{t=1}^T (X_{1,t} - \bar{X}_1)^2 \rightarrow \pi^2/12$  in Prob. Further,

$$\begin{aligned} \frac{1}{T^{3/2}} \sum_{t=1}^T (X_{1,t} - \bar{X}_1) u_{2,t} &= \frac{1}{T^{3/2}} \sum_{t=1}^T \left( \sum_{j=1}^t u_{1,j} - \frac{1}{T} \sum_{t=1}^T \left( \sum_{j=1}^t u_{1,j} \right) \right) u_{2,t} \\ &\quad + \frac{1}{T^{3/2}} \sum_{t=1}^T \pi \left( t - \frac{T+1}{2} \right) u_{2,t} \\ &= \frac{1}{T^{3/2}} \sum_{t=1}^T \pi \left( t - \frac{T+1}{2} \right) u_{2,t} + o_p(1) \\ &\Rightarrow \pi \sigma_{v_2} \int_0^1 s dW_s - \frac{\pi}{2} \sigma_{v_2} W_1. \end{aligned}$$

Thus,

$$T^{3/2}(\hat{\beta}_T - \beta_0) \Rightarrow \frac{12}{\pi} \sigma_{v_2} \left( \int_0^1 s \, dW_s - (1/2)W_1 \right).$$

*Proof of Theorem 3.3* (i) Let

$$\begin{aligned} \hat{\xi}_t &= (X_{2,t} - \bar{X}_2) - \hat{\beta}_T(X_{1,t} - \bar{X}_1) = (X_{2,t} - \bar{X}_2) - \beta_0(X_{1,t} - \bar{X}_1) \\ &\quad + (\hat{\beta}_T - \beta_0)(X_{1,t} - \bar{X}_1). \end{aligned}$$

Then,

$$\begin{aligned} \frac{1}{\sqrt{T}} \sum_{j=1}^t \hat{\xi}_j &= \frac{1}{\sqrt{T}} \sum_{j=1}^t ((X_{2,j} - \bar{X}_2) - \beta_0(X_{1,j} - \bar{X}_1)) \\ &\quad + T^{3/2}(\hat{\beta}_T - \beta_0) \frac{1}{T^2} \sum_{j=1}^t (X_{1,j} - \bar{X}_1) \\ &= \frac{1}{\sqrt{T}} \sum_{j=1}^t ((X_{2,t} - \bar{X}_2) - \beta_0(X_{1,t} - \bar{X}_1)) \\ &\quad + T^{3/2}(\hat{\beta}_T - \beta_0) \frac{1}{T^2} \sum_{j=1}^t \left( \pi j - \pi \frac{T+1}{2} \right) + o_p(1). \end{aligned}$$

Given Theorem 3.3(i), it follows that

$$\frac{1}{T^2} \sum_{t=1}^T \left( \sum_{j=1}^t \hat{\xi}_j \right)^2 \Rightarrow \sigma_{v_2}^2 \int_0^1 Q_s^2 \, ds,$$

where

$$Q_s = (W_s - sW_1) - 6 \left( \int_0^1 s \, dW_s - (1/2)W_1 \right) (s^2 - s).$$

As  $T^{3/2}(\hat{\beta}_T - \beta_0) = O_p(1)$ , and using the same argument used in the proof of Theorem 3.1(i), it follows that  $s_{T^*}^2 \rightarrow \sigma_{v_2}^2$  in *Prob.* The result follows.

(ii) Recall that

$$\hat{\beta}_T = \frac{\sum_{t=1}^T (X_{1,t} - \bar{X}_1)(X_{2,t} - \bar{X}_2)}{\sum_{t=1}^T (X_{1,t} - \bar{X}_1)^2}.$$

Given that  $X_{1,t}$  and  $X_{2,t}$  are of the same order of probability (note that the nonlinear components,  $g_{0,1}$ ,  $g_{0,2}$ , depend on the same argument,  $\theta'_0 X_t$ ),  $\hat{\beta}_T = O_p(1)$ . Thus,  $s_{T^*}^{-2} (1/T^2) \sum_{t=1}^T (\sum_{j=1}^t \hat{\xi}_j)^2$  will explode at rate  $Tl_T^{-1}$ , by the same argument used in the proof of Theorem 3.1(ii).  $\square$

*Proof of Theorem 4.1.* (i) Recall that  $\hat{\beta}_T$  is the slope coefficient from the regression of  $X_{2,t}$  on  $X_{1,t}$  and a constant, and  $\hat{\gamma}_T = (-\hat{\beta}_T, 1)$ . Let  $\hat{\eta}_t$  be defined as in Eq. (4.1), and  $\hat{g} = (1/T)\sum_{t=2}^T g(\hat{\gamma}'_T X_{t-1}\tau)$ . Finally, let  $\bar{g} = (1/T)\sum_{t=2}^T g(\gamma'_0 X_{t-1}\tau)$ . Given Eq. (4.2), note that

$$\begin{aligned} \frac{1}{\sqrt{T}} \sum_{t=2}^T \hat{\eta}_t(g(\hat{\gamma}'_T X_{t-1}\tau) - \hat{g}) &= \frac{1}{\sqrt{T}} \sum_{t=2}^T \hat{\eta}_t(g(\hat{\gamma}'_T X_{t-1}\tau) - \bar{g}) \\ &- (\hat{\beta}_T - \beta_0) \frac{1}{\sqrt{T}} \sum_{t=2}^T (X_{1,t} - \bar{X}_1)(g(\hat{\gamma}'_T X_{t-1}\tau) - \bar{g}) \\ &- \psi_0(\hat{\beta}_T - \beta_0) \frac{1}{\sqrt{T}} \sum_{t=2}^T (X_{1,t} - \bar{X}_1)(g(\hat{\gamma}'_T X_{t-1}\tau) - \bar{g}) \\ &- (\hat{\psi}_T - \psi_0) \frac{1}{\sqrt{T}} \sum_{t=2}^T (\gamma'_0 X_{t-1} - \gamma'_0 \bar{X})(g(\hat{\gamma}'_T X_{t-1}\tau) - \bar{g}) \\ &- (\hat{\psi}_T - \psi_0)(\hat{\beta}_T - \beta_0) \frac{1}{\sqrt{T}} \sum_{t=2}^T (X_{1,t-1} - \bar{X}_1)(g(\hat{\gamma}'_T X_{t-1}\tau) - \bar{g}). \end{aligned}$$

Now write

$$g(\hat{\gamma}'_T X_{t-1}\tau) - \hat{g} = (g(\hat{\gamma}'_0 X_{t-1}\tau) - \bar{g}) + (g(\hat{\gamma}'_T X_{t-1}\tau) - g(\hat{\gamma}'_0 X_{t-1}\tau) - (\hat{g} - \bar{g})).$$

Then,

$$\begin{aligned} \frac{1}{\sqrt{T}} \sum_{t=2}^T \hat{\eta}_t(g(\hat{\gamma}'_T X_{t-1}\tau) - \hat{g}) &= \frac{1}{\sqrt{T}} \sum_{t=2}^T \hat{\eta}_t(g(\hat{\gamma}'_0 X_{t-1}\tau) - \bar{g}) \\ &+ \frac{1}{\sqrt{T}} \sum_{t=2}^T \eta_t(g(\gamma'_0 X_{t-1}\tau) - g(\hat{\gamma}'_T X_{t-1}\tau) - (\bar{g} - \hat{g})) \\ &- (\hat{\beta}_T - \beta_0) \frac{1}{\sqrt{T}} \sum_{t=2}^T (X_{1,t} - \bar{X}_1)(g(\hat{\gamma}'_0 X_{t-1}\tau) - \bar{g}) \\ &- (\hat{\beta}_T - \beta_0) \frac{1}{\sqrt{T}} \sum_{t=2}^T (X_{1,t} - \bar{X}_1)(g(\hat{\gamma}'_T X_{t-1}\tau) - g(\hat{\gamma}'_0 X_{t-1}\tau) - (\hat{g} - \bar{g})) \\ &- \psi_0(\hat{\beta}_T - \beta_0) \frac{1}{\sqrt{T}} \sum_{t=2}^T (X_{t-1} - \bar{X}_1)(g(\hat{\gamma}'_0 X_{t-1}\tau) - \bar{g}) \\ &+ \psi_0(\hat{\beta}_T - \beta_0) \frac{1}{\sqrt{T}} \sum_{t=2}^T (X_{1,t-1} - \bar{X}_1)(g(\hat{\gamma}'_T X_{t-1}\tau) - g(\hat{\gamma}'_0 X_{t-1}\tau) - (\hat{g} - \bar{g})) \end{aligned}$$

$$\begin{aligned}
& - (\hat{\psi}_T - \psi_0) \frac{1}{\sqrt{T}} \sum_{t=2}^T \gamma'_0 (X_{t-1} - \bar{X}_1) (g(\gamma'_0 X_{t-1} \tau) - \bar{g}) \\
& + (\hat{\psi}_T - \psi_0) \frac{1}{\sqrt{T}} \sum_{t=2}^T (\gamma'_0 X_{t-1} - \gamma'_0 \bar{X}) (g(\hat{\gamma}'_T X_{t-1} \tau) - g(\gamma'_0 X_{t-1} \tau) - (\bar{g} - \bar{g})) \\
& - (\hat{\psi}_T - \psi_0) (\hat{\beta}_T - \beta_0) \frac{1}{\sqrt{T}} \sum_{t=2}^T (X_{1,t-1} - \bar{X}_1) (g(\gamma'_0 X_{t-1} \tau) - \bar{g}) \\
& + (\hat{\psi}_T - \psi_0) (\hat{\beta}_T - \beta_0) \frac{1}{\sqrt{T}} \sum_{t=2}^T (X_{1,t-1} - \bar{X}_1) (g(\hat{\gamma}'_T X_{t-1} \tau) - g(\gamma'_0 X_{t-1} \tau) \\
& - (\bar{g} - \bar{g})). \tag{A.2}
\end{aligned}$$

We want to show that

$$\begin{aligned}
& \frac{1}{\sqrt{T}} \sum_{t=2}^T \hat{\eta}_t (g(\hat{\gamma}'_T X_{t-1} \tau) - \bar{g}) = \frac{1}{\sqrt{T}} \sum_{t=2}^T \eta_t (g(\gamma'_0 X_{t-1} \tau) - \bar{g}) \\
& - (\hat{\psi}_T - \psi_0) \frac{1}{\sqrt{T}} \sum_{t=2}^T (\gamma'_0 X_{t-1} - \gamma'_0 \bar{X}) (g(\hat{\gamma}'_T X_{t-1} \tau) - \bar{g}) + o_p(1). \tag{A.3}
\end{aligned}$$

As

$$\sqrt{T} (\hat{\psi}_T - \psi_0) = \frac{(1/\sqrt{T}) \sum_{t=2}^T (\gamma'_0 X_{t-1} - \gamma'_0 \bar{X}) \eta_t}{(1/\sqrt{T}) \sum_{t=2}^T (\gamma'_0 X_{t-1} - \gamma'_0 \bar{X})^2} + o_p(1),$$

(A.3) implies (4.6) in the text. Recall that under  $H_0$ ,  $\sqrt{T}(\hat{\psi}_T - \psi_0) = O_p(1)$  and  $T(\hat{\beta}_T - \beta_0) = O_p(1)$ . As the ninth term on the RHS of (A.2) converges to zero faster than the fifth term, and the tenth term vanishes faster than the sixth, we can neglect them. Also, under  $H_0$ ,  $X_{1,t} = \sum_{j=1}^t v_{1,j}$  where  $E(v_{1,j}) = 0$  for all  $j$ . Note that when  $g_0$  is a constant, then  $T^{1/2}(\hat{\psi}_T - \psi_0) = O_p(1)$  and  $T^{3/2}(\hat{\beta}_T - \beta_0) = O_p(1)$  and  $E(v_{1,j}) \neq 0$  for all  $j$ . By twice applying the mean value theorem, it turns out that

$$\begin{aligned}
g(\hat{\gamma}'_T X_{t-1} \tau) - g(\gamma'_0 X_{t-1} \tau) &= (\hat{\beta}_T - \beta_0) X_{1,t} \tau Dg(\gamma'^* X_{t-1} \tau) \\
&= (\hat{\beta}_T - \beta_0) X_{1,t-1} \tau Dg(\gamma'_0 X_{t-1} \tau) \\
&\quad - (\hat{\beta} - \beta_0) (\beta^* - \beta_0) X_{1,t-1}^2 \tau^2 D^2 g(\gamma'^* X_{t-1} \tau),
\end{aligned}$$



where  $D^k g$  denotes the  $k$ th derivative of  $g$  with respect to its argument, and  $\gamma^* = (-\beta^*, 1)'$ , with  $\beta^* \in (\beta_0, \hat{\beta}_T)$ ,  $\gamma^{**} = (-\beta^{**}, 1)$ , and  $\beta^{**} \in (\beta^*, \beta_0)$ . Thus, the second term on the RHS of (A.2) can be written as

$$\begin{aligned} & T(\hat{\beta}_T - \beta_0) \frac{1}{T^{3/2}} \sum_{t=2}^T (X_{1,t-1} - \bar{X}_1) \tau \eta_t (Dg(\gamma'_0 X_{t-1} \tau) - \overline{Dg}_{\gamma_0}) \\ & - T(\hat{\beta}_T - \beta_0) T(\beta^* - \beta_0) \frac{1}{T^{5/2}} \sum_{t=2}^T \eta_t (X_{1,t-1} - \bar{X}_1)^2 \\ & \times \tau^2 (D^2 g(\gamma^{**} X_{t-1} \tau) - \overline{Dg}^2_{\gamma^{**}}), \end{aligned} \tag{A.4}$$

where  $\overline{Dg}_{\gamma_0}^k = (1/T) \sum D^k g(\gamma'_0 X_{t-1} \tau)$  and  $\overline{Dg}_{\gamma^{**}}^k = (1/T) \sum D^k g(\gamma^{**} X_{t-1} \tau)$ . Note that  $Dg(\gamma'_0 X_{t-1} \tau) - \overline{Dg}_{\gamma_0}$  is a zero mean strong mixing process, and that (from Fact 2.5) the usual size conditions required by the invariance principle hold. Also, by a similar argument as that used in Hansen (1992b, Proof of Theorem 4.1),

$$\frac{1}{T} \sum_{t=2}^T (X_{1,t-1} - \bar{X}_1) \tau \eta_t (Dg(\gamma'_0 X_{t-1} \tau) - \overline{Dg}_{\gamma_0}) = O_p(1),$$

so that the first term in (A.4) is  $O_p(T^{-1/2})$ . Also, note that the absolute value of the second term in (A.4) is majorized by

$$\begin{aligned} & T|\hat{\beta}_T - \beta_0| T|\beta^* - \beta_0| \sup_t |D^2 g(\gamma^{**} X_{t-1} \tau)| \\ & - \overline{Dg}_{\gamma^{**}}^2 \tau^2 \frac{1}{\sqrt{T}} \sup_t |\eta_t| \frac{1}{T^2} \sum_{t=2}^T (X_{1,t-1} - \bar{X}_1)^2. \end{aligned} \tag{A.5}$$

Now,  $T|\beta^* - \beta_0| \leq T|\hat{\beta}_T - \beta_0| = O_p(1)$  and  $(1/\sqrt{T}) \sup_t |\eta_t| = o_p(1)$ . It follows that the expression in (A.5) is  $o_p(1)$ , and thus the expression in (A.4) is also  $o_p(1)$ . Now, consider the eighth term on the RHS of (A.2), which can be written as

$$\begin{aligned} & (\hat{\psi}_T - \psi_0)(\hat{\beta}_T - \beta_0) \frac{1}{\sqrt{T}} \sum_{t=2}^T \tau X_{1,t-1} (\gamma'_0 X_{t-1} - \gamma'_0 \bar{X}) (Dg(\gamma'_0 X_{t-1} \tau) - \overline{Dg}_{\gamma_0}) \\ & - (\hat{\psi}_T - \psi_0)(\hat{\beta}_T - \beta_0)(\beta^* - \beta_0) \frac{1}{\sqrt{T}} \sum (X_{1,t-1} - \bar{X}_1)^2 \tau^2 \\ & \times (\gamma'_0 X_{t-1} - \gamma'_0 \bar{X}) (D^2 g(\gamma^{**} X_{1,t-1} \tau) - \overline{Dg}_{\gamma^{**}}^2) = o_p(1), \end{aligned}$$

using the same majorization argument which is used above. By an analogous argument, the fourth, fifth and sixth terms are  $o_p(1)$ . Given that  $T(\hat{\beta}_T - \beta_0) = O_p(1)$ , from the same argument used in the proof of Theorem 3.1(i)

we have that  $s_{t_r}^2 \rightarrow \sigma_0^2$  in *Prob.* The result follows. Note also that when  $g_0$  is a nonzero constant,  $T^{1/2}(\hat{\psi} - \psi_0) = O_p(1)$ ,  $T^{3/2}(\hat{\beta}_T - \beta_0) = O_p(1)$ , and  $E(v_{1,j}) \neq 0$  for all  $j$ . Thus, the same arguments used above apply when  $g_0$  is a nonzero constant.

(ii) Under the alternative,  $X_{1,t} = \pi t + \sum_{j=1}^t (v_{1,t} - \pi) = \pi t + \sum_{j=1}^t u_j$ . As  $X_{1,t}$  is dominated by the deterministic component,  $T^{3/2}(\hat{\beta}_T - \beta_0) = O_p(1)$ . Further,  $\sqrt{T}(\hat{\psi}_T - \psi^*) = O_p(1)$ , where  $\psi^* \neq \psi_0$ . Now all of the terms on the RHS of (A.2), except for the first and the seventh are  $o_p(1)$ , by an argument analogous to that used in part (i), and by the law of large numbers for strong mixing processes,  $(1/T)\sum_{t=2}^T (g(\gamma'_0 X_{t-1}\tau) - \bar{g}) - M_{\gamma'_0 X_{t-1}}^{-1} M_{\gamma'_0 X_{t-1}} M_{\gamma'_0 X_{t-1}} \eta_t \rightarrow M \neq 0$ . The result then follows.  $\square$

## References

- Altissimo, F., Violante, G.L., 1995. Persistence and Nonlinearity in GNP and Unemployment: an Endogeneous Delay Threshold VAR, Mimeo, University of Pennsylvania.
- Anderson, H.M., 1997. Transaction costs and nonlinear adjustment towards equilibrium in the US treasury bill market. *Oxford Bulletin of Economics and Statistics* 59, 465–480.
- Anderson, H.M., Vahid, F., 1988. Testing multiple equation systems for common nonlinear components. *Journal of Econometrics* 84, 1–36.
- Andrews, D., 1991. Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* 59, 817–858.
- Athreya, K.B., Pantula, S.G., 1986. Mixing properties of Harris chains and autoregressive processes. *Journal of Applied Probability* 23, 880–892.
- Balke, N.S., Fomby, T.B., 1997. Threshold cointegration. *International Economic Review* 38, 627–645.
- Beveridge, S., Nelson, C., 1981. A new approach to decomposition of economic time series into permanent and transitory components with particular attention to measurement of the business cycle. *Journal of Monetary Economics* 7, 151–174.
- Bierens, H.J., 1990. A consistent conditional moment test of functional form. *Econometrica* 58, 1443–1458.
- Corradi, V., 1995. Nonlinear transformations of integrated time series: a reconsideration. *Journal of Times Series Analysis* 16, 539–549.
- Domowitz, I., El-Gamal, M.A., 1993. A consistent test for stationary-ergodicity. *Econometric Theory* 9, 589–601.
- Domowitz, I., El-Gamal, M.A., 1997. A consistent nonparametric test of ergodicity for time series with applications. Mimeo, Northwestern University.
- Engle, R.F., Granger, C.W.J., 1987. Co-integration and error correction: representation, estimation and testing. *Econometrica* 55, 547–557.
- Ermini, L., Granger, C.W.J., 1993. Some generalization of the algebra of  $I(1)$  processes. *Journal of Econometrics* 58, 369–384.
- Granger, C.W.J., 1995. Modelling nonlinear relationships between extended-memory variables. *Econometrica* 63, 265–279.
- Granger, C.W.J., Hallman, J., 1991. Long memory series with attractors. *Oxford Bulletin of Economics and Statistics* 53, 11–26.
- Granger, C.W.J., Inoue, T., Morin, N., 1997. Nonlinear stochastic trends. *Journal of Econometrics* 81, 65–92.

- Granger, C.W.J., Swanson, N.R., 1996. Further developments in the study of cointegrated economic variables. *Oxford Bulletin of Economics and Statistics* 58, 537–553.
- Granger, C.W.J., Teräsvirta, T., 1993. *Modelling Non-linear Economic Relationships*. Oxford University Press, New York.
- Hall, A.D., Anderson, H.M., Granger, C.W.J., 1992. A cointegration analysis of treasury bill yields. *The Review of Economics and Statistics* 74, 116–126.
- Hansen, B.E., 1992a. Efficient estimation and testing of cointegrating vectors in the presence of deterministic trends. *Journal of Econometrics* 53, 87–121.
- Hansen, B.E., 1992b. Convergence to stochastic integrals for dependent and heterogeneous processes. *Econometric Theory* 8, 489–500.
- Hochberg, Y., 1988. A sharper Bonferroni procedure for multiple tests of significance. *Biometrika* 75, 800–802.
- Johansen, S., 1988. Statistical analysis of cointegration vectors. *Journal of Economics and Dynamics and Control* 12, 231–254.
- Johansen, S., 1991. Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models. *Econometrica* 59, 1551–1580.
- Kocic, V.L., Ladas, G., 1993. *Global Behavior of Nonlinear Difference Equations of Higher Order with Applications*. Kluwer, London.
- Kuan, C.-M., White, H., 1994. Artificial neural networks: an econometric perspective. *Econometric Reviews* 13, 1–91.
- Kwiatkowski, D., Phillips, P.C.B., Schmidt, P., Shin, Y., (KPSS), 1992. Testing the null hypothesis of stationarity against the alternative of unit root: how sure are we that economic series have a unit root? *Journal of Econometrics* 54, 159–178.
- Lee, T.H., White, H., Granger, C.W.J., 1993. Testing for neglected nonlinearity in time series models. a comparison of neural network methods and alternative tests. *Journal of Econometrics* 56, 269–290.
- MacKinnon, J.G., 1991. Critical values for cointegration tests in long-run economic relationships. In: Engle, R.F., Granger, C.W.J. (Eds.), *Long-run Economic Relationships: Readings in Cointegration*. Oxford University Press, New York.
- Meyn, S.P., 1989. Ergodic theorems for discrete stochastic systems using a stochastic Lyapunov function. *SIAM Journal of Control and Optimization* 27, 1409–1439.
- Park, J.Y., Ogaki, M., 1991. Inference in cointegrated models using VAR prewhitening to estimate shortrun dynamics. Working Paper Number 281, The Rochester Center for Economic Research.
- Pesaran, M.H., Potter, S.M., 1997. A floor and ceiling model of U.S. output. *Journal of Economic Dynamics and Control* 21, 661–695.
- Shin, Y., 1994. A residual based test of the null of cointegration against the alternative of no cointegration. *Econometric Theory* 10, 91–115.
- Stinchcombe, M., White, H., 1998. Consistent specification testing with nuisance parameters present only under the alternative. *Econometric Theory* 14, 295–324.
- Swanson, N.R., 1999. LM tests and nonlinear error correction in economic time series. *Statistica Neerlandica*, 53, 76–95.
- Teräsvirta, T., Anderson, H.M., 1992. Characterizing nonlinearities in business cycles using smooth transition autoregressive models. *Journal of Applied Econometrics* 7, S: S119–S136.
- Tong, H., 1990. *Non-linear Time Series: A Dynamical System Approach*. Oxford University Press, New York.
- White, H., 1984. *Asymptotic Theory for Econometricians*. Academic Press, New York.