

Department of Economics

Testing for Neglected Nonlinearity in Cointegrating Relationships

Andrew P. Blake and George Kapetanios

Working Paper No. 508

February 2004

ISSN 1473-0278



Testing for Neglected Nonlinearity in Cointegrating Relationships*

Andrew P. Blake[†]
CCBS, Bank of England

and

George Kapetanios[‡]
Queen Mary, University of London

February 10, 2004

Abstract

This paper proposes pure significance tests for the absence of nonlinearity in cointegrating relationships. No assumption of the functional form of the nonlinearity is made. It is envisaged that the application of such tests could form the first step towards specifying a nonlinear cointegrating relationship for empirical modelling. The asymptotic and small sample properties of our tests are investigated, where special attention is paid to the role of nuisance parameters and a potential resolution using the bootstrap.

JEL Codes: C32, C45

Keywords: Cointegration, Nonlinearity, Neural Networks, Bootstrap

*This paper contains the results and analysis of the authors and should not be thought to reflect the views of either the Bank of England or Monetary Policy Committee members.

[†]Centre for Central Banking Studies, Bank of England, Threadneedle Street, London EC2R 8AH, UK. Email: Andrew.Blake@bankofengland.co.uk.

[‡]Department of Economics, Queen Mary, University of London, Mile End Road, London E1 4NS, UK. Email: G.Kapetanios@qmul.ac.uk

1 Introduction

In the past two decades long run relationships between economic time series, and in particular integrated time series, have been the focus of extensive theoretical and empirical analysis. In most of this work the assumption of linearity has been maintained. Recently, there has been increasing interest in the possibility that nonlinear models may be a fruitful avenue for further investigation. Nevertheless, so far the assumption of linearity has been relaxed only in the context of investigating convergence to long run linear relationships. So, for example, nonlinear specifications for the speed of adjustment to cointegrating relationships in vector error correction models have been repeatedly suggested in the empirical literature. By contrast, little work has been carried out to investigate the possibility that the actual cointegrating relationships themselves are nonlinear. On the other hand, a firm understanding of the econometric underpinnings of such a specification for a single cointegrating relationship has been achieved through the work of Park and Phillips (1999, 2001).

It is, of course, an open question whether such nonlinearity exists in observed data. Currently, a researcher would need to assume a particular functional form for any nonlinear cointegrating relationships. However, assuming a particular functional form may be problematic if economic theory does not provide any guidance in this respect. This paper proposes a pure significance test for the absence of nonlinearity in cointegrating relationships. No assumption on the functional form of the nonlinearity is made. It is envisaged that such a test would form the first step towards specifying a nonlinear cointegrating relationship for empirical modelling.

The test is based on neural networks and mirrors similar work done by the authors and others using neural networks in a stationary framework.¹ The paper is structured as follows: Section 2 describes and provides the

¹See, in particular, Lee, White, and Granger (1993), Teräsvirta, Lin, and Granger (1993) and Blake and Kapetanios (2003b).

theoretical underpinnings for our tests. Section 3 discusses their asymptotic properties. Section 4 provides a Monte Carlo study. Section 5 concludes.

2 The Setup

We propose the following general model following Chang, Park, and Phillips (2001):

$$y_t = \boldsymbol{\alpha}'\mathbf{x}_t + \boldsymbol{\gamma}'\mathbf{w}_t + f(\boldsymbol{\beta}, \mathbf{x}_t) + \epsilon_t \quad (1)$$

where \mathbf{x}_t is a vector of integrated regressors generated by:

$$\mathbf{x}_t = \mathbf{x}_{t-1} + \mathbf{v}_t \quad (2)$$

and \mathbf{w}_t is a vector of stationary regressors. We specify \mathbf{w}_t and \mathbf{v}_t as general linear processes given by:

$$\mathbf{w}_t = \boldsymbol{\phi}(L)\mathbf{u}_t = \sum_{i=1}^{\infty} \boldsymbol{\phi}_i \mathbf{u}_{t-i} \quad (3)$$

$$\mathbf{v}_t = \boldsymbol{\psi}(L)\mathbf{z}_t = \sum_{i=1}^{\infty} \boldsymbol{\psi}_i \mathbf{z}_{t-i} \quad (4)$$

and make the following assumptions:

Assumption 1. (i) $\boldsymbol{\phi}(1)$ is nonsingular. (ii) $\sum_{k=0}^{\infty} k^b \|\boldsymbol{\phi}_k\| < \infty$, $b > 1$.

Assumption 2. \mathbf{u}_t is i.i.d. with $E\|\mathbf{u}_t\|^r < \infty$ for some $r > 8$ and its distribution is absolutely continuous w.r.t. Lebesgue measure and has characteristic function $\phi(\lambda) = o(\|\lambda\|^{-\delta})$, $\delta > 0$.

Assumption 3. $\boldsymbol{\xi}_t = (\epsilon_t, \mathbf{u}'_{t+1}, \mathbf{z}'_{t+1})'$ is a stationary and ergodic martingale difference sequence with finite second moments $\boldsymbol{\Sigma}$ and $\sup_{t \geq 1} E(\|\boldsymbol{\xi}_t\|^r | \mathcal{F}_{t-1}) < \infty$, $r > 4$, where \mathcal{F}_{t-1} is the σ -field generated by $\boldsymbol{\xi}_t$.

We wish to test the null hypothesis that:

$$P\{E(y_t|\mathbf{x}_t, \mathbf{w}_t) = \boldsymbol{\alpha}'\mathbf{x}_t + \boldsymbol{\gamma}'\mathbf{w}_t\} = 1 \quad (5)$$

for some constant vector $(\boldsymbol{\alpha}', \boldsymbol{\gamma}')$. The alternative is then defined as:

$$P\{E(y_t|\mathbf{x}_t, \mathbf{w}_t) = \boldsymbol{\alpha}'\mathbf{x}_t + \boldsymbol{\gamma}'\mathbf{w}_t\} < 1 \quad (6)$$

for all $(\boldsymbol{\alpha}', \boldsymbol{\gamma}')$. Before proceeding with our analysis it is worth distinguishing between two classes of functions $f(\cdot)$. The first class is the class of integrable functions and the second is the class of asymptotically homogeneous functions as discussed in Park and Phillips (2001). For the purposes of empirical analysis these two classes of functions have very distinct implications. For example it is clear that when $f(\cdot)$ is an integrable function its effect will only be of relevance for periods where the processes \mathbf{x}_t are in the activation area of the function. These periods will be of order $T^{1/2}$. On the other hand asymptotically homogeneous functions will be of relevance for much longer periods.

For any aspect of nonlinearity testing the usefulness of neural networks arises out of their potential to approximate arbitrary nonlinear functions. The generic form of a neural network approximation applied in this context is given by:

$$y_t = \boldsymbol{\alpha}'\mathbf{x}_t + \boldsymbol{\gamma}'\mathbf{w}_t + \sum_{j=1}^q \beta_j \varphi(\delta_j, \mathbf{x}_t) \quad (7)$$

Clearly, a test of $\beta_1 = \dots = \beta_q = 0$ in:

$$y_t = \boldsymbol{\alpha}'\mathbf{x}_t + \boldsymbol{\gamma}'\mathbf{w}_t + \sum_{j=1}^q \beta_j \varphi(\delta_j, \mathbf{x}_t) + \epsilon_t \quad (8)$$

provides a test for neglected nonlinearity.

Artificial neural networks can approximate arbitrary continuous functions arbitrarily well. More specifically, a continuous function $f(z)$ can be arbitrarily well approximated in the supremum norm by $\sum_{i=1}^q g(z'_i)$ for finite q and

$z'_i = a_{i,0} + a'_{i,1}z$ if either (i) $g(\cdot)$ is sigmoidal, i.e. $g(\cdot)$ is non-decreasing with $\lim_{z \rightarrow -\infty} g(z) = 0$ and $\lim_{z \rightarrow \infty} g(z) = 1$ (Condition C1) or (ii) $g(\cdot)$ has non-zero Lebesgue measure expectation and is L_p bounded for some $p \geq 1$ (Condition C2). For more details see Hornik, Stinchcombe, and White (1989), Stinchcombe and White (1989) and Cybenko (1989).² These results are only a small subset of the available results in the literature on the approximation properties of artificial neural networks.

We need to choose a suitable function $\varphi(\boldsymbol{\delta}_j, \mathbf{x}_t)$. Note that the parameters, $\boldsymbol{\delta}_j$, are not identified under the null hypothesis. The most widely used neural network nonlinearity test suggested by Lee, White, and Granger (1993) specifies that $\varphi(\cdot, \cdot)$ in (8) is given by $\varphi(\boldsymbol{\delta}'\mathbf{x}_t)$ where $\varphi(\lambda)$ is the logistic function $\{1 + e^{-\lambda}\}^{-1}$. This is a monotonic function, with output bounded between 0 and 1. The elements of the coefficient vector $\boldsymbol{\delta}_j$ are randomly generated from a uniform distribution over (δ_l, δ_h) . This procedure addresses the problem of identifiability of the neural network model under the null hypothesis.

A number of other functions have been proposed and used in the literature to construct neural networks (see, e.g., Blake and Kapetanios, 2003b). The most common alternative is a radial basis function. In contrast to the logistic function a radial basis function (RBF) is a function which is monotonic about some centre. Define q centres by \mathbf{c}_j and a radius vector τ . We interest ourselves only in those functions that are monotonically decreasing about \mathbf{c}_j . For example, the Gaussian RBF is:

$$\varphi\{(\mathbf{c}'_j, \tau)', \mathbf{x}_t\} = e^{-\|\mathbf{x}_t - \mathbf{c}_j\|^2 / \tau^2}. \quad (9)$$

²More accurately, Hornik, Stinchcombe, and White (1989) show that an artificial neural network based on a function $g(\cdot)$ satisfying Condition C1 will approximate any continuous function uniformly on compacta, whereas Stinchcombe and White (1989) show the equivalent result for artificial neural networks based on functions $g(\cdot)$, satisfying Condition C2. The definition of approximation uniformly on compacta is as follows. A sequence of functions f_n converges to a function f uniformly on compacta if for all compact sets $K \subset \mathbb{R}^r$, $\sup_K |f_n(x) - f(x)| \rightarrow 0$.

By the monotonicity property, each RBF has maximum activation (of unity) when the input vector coincides with the j th centre independent of τ . Conversely, if the input vector is far enough away for the centre the activation is zero, controlled by τ . Other functional forms, such as the multiquadratic, have the same properties and can be used instead. See Campbell, Lo, and MacKinlay (1997) for an introduction to artificial neural networks in general which covers RBF networks or Bishop (1995) for a more thorough account. RBFs have been used in the econometric literature to test for neglected nonlinearity by Blake and Kapetanios (2003b). We suggest the use of an RBF neural network to test for neglected nonlinearity in long run relationships.

Of course, an identification problem, similar to that faced by Lee, White, and Granger (1993) arises when an RBF neural network is considered. Problems arise as we need to determine the centres (c) and radii (τ) for each RBF, and the number of ‘hidden units’ used. The problem arises since these parameters are not identified under the null hypothesis. We propose the use of data-based procedures. Several rules could be used for choosing the radii. It is common practice in the artificial neural network literature to use a fixed multiple of the maximum change from period t to period $t+1$, $t = 1, \dots, T$ of each input as the radius for that input (see Orr, 1995). Another option is to normalise the data and use unity for the radius. The centres are naturally determined. We allow T candidate hidden units by using all available observations themselves as possible centres to RBFs. Following Orr (1995), we rank the T RBFs in order of maximum reduction of the residual sum of squares in (8). To do that we estimate T regressions of the form (8) where each regression contains one RBF hidden unit. Then, we obtain the residual sum of squares from each regression and use these to rank the RBFs. Then, we successively add the ranked RBFs in (8) until we minimise an information criterion, where we choose one from those proposed by Akaike (1976) (AIC) and Schwarz (1978) (BIC). This determines q . In the Monte Carlo study we will use BIC.

We then test for the significance of the included hidden units. Note that the test cannot be carried out if no hidden units are chosen by the information criterion. We therefore do not consider the case of no hidden units and start with a minimum of one hidden unit. For consistent information criteria (i.e. criteria which pick the order of the model correctly in probability such as BIC) exactly one hidden unit will be chosen asymptotically in probability under the null hypothesis of linearity. If the information criterion search included the case of no hidden units then, asymptotically, consistent criteria would pick no hidden units with probability approaching one. We further note that, asymptotically, AIC, being inconsistent in model order selection in general, will pick more than one hidden units with non-zero probability under the null hypothesis.

We use a standard Wald test statistic to test the null hypothesis that $\beta_1 = \dots = \beta_q = 0$. This takes the form:

$$\frac{1}{\hat{\sigma}^2} \hat{\beta}' \{R'(W'W)^{-1}R\}^{-1} \hat{\beta} \quad (10)$$

where W is the matrix of regressors of (8) and a constant, R is the selector matrix for the coefficients of the hidden units, $\beta = (\beta_1, \dots, \beta_q)'$ and $\hat{\sigma}^2$ is the estimated variance of the residuals in (8). The test is asymptotically distributed as a χ_q^2 .

As an alternative to the Lee, White, and Granger (1993) test, Teräsvirta, Lin, and Granger (1993) (TLG) suggested an array of Lagrange Multiplier (LM) tests for linearity motivated from alternative hypotheses of neural network models. These tests which are similar to the test against STAR type nonlinearity were shown to have superior power properties to the neural network test of the Lee, White, and Granger (1993) test using Monte Carlo methods. The authors suggest that a procedure which tests for the significance of the squares, cubes and cross products of the original regressors should be powerful against a wide variety of departures from linearity. As an example, in a model with two regressors, $x_{1,t}$ and $x_{2,t}$, the joint significance of the following terms is tested: $x_{1,t}^2$, $x_{2,t}^2$, $x_{1,t}^3$, $x_{2,t}^3$, $x_{1,t}^2 x_{2,t}$, $x_{1,t} x_{2,t}^2$ and $x_{1,t} x_{2,t}$.

The choice of the regressors used in the TLG test is motivated as a third order Taylor expansion of the logistic artificial neural network model. As we do not wish to arbitrarily restrict analysis to third order Taylor expansions we also consider a second and a fourth order Taylor expansion. In particular, we consider a test where the order of the Taylor expansion is chosen by an information criterion which in our case is BIC. For more details on this approach see also Blake and Kapetanios (2003b).

3 Asymptotic Properties

We now discuss the asymptotic distribution of the tests. At this point we will provide a rigorous treatment of the asymptotic distribution for a bivariate model only. The regression used to construct the RBF test statistic is given by:

$$y_t = \beta x_t + \sum_{i=1}^m \zeta_i e^{-((x_t - c_i)/\tau)^2} + \epsilon_t \quad (11)$$

where c_i are assumed finite and $\tau \neq 0$. Then, this regression falls under the framework discussed by Park and Phillips (1999, 2001). We have the following theorem:

Theorem 1. *Under assumptions 1-3, for a bivariate model and assuming that m , c_i and τ are known in (11) the asymptotic distribution of the RBF test statistic is χ_m^2 .*

Proof. The function $\sum_{i=1}^m \zeta_i e^{-((\cdot - c_i)/\tau)^2}$ is I -regular as defined by Park and Phillips (2001) since it is bounded, integrable and satisfies the Lipschitz condition (b) of definition 3.3 of Park and Phillips (2001). Define:

$$\tilde{z}_t = \sum_{i=1}^m \zeta_i e^{-(x_t - c_i)/\tau)^2}$$

for fixed τ and c_i . Using Theorem 3.2 of Park and Phillips (2001) we have that:

$$\frac{1}{\sqrt{T}} \sum_{t=1}^T \tilde{z}_t^2 \xrightarrow{p} \left(\int_{-\infty}^{\infty} \left(\sum_{i=1}^m \zeta_i e^{-((s-c_i)/\tau)^2} \right)^2 ds \right) L(1,0) \quad (12)$$

and:

$$\frac{1}{\sqrt[4]{T}} \sum_{t=1}^T \tilde{z}_t \epsilon_t \xrightarrow{d} \left(\left(\int_{-\infty}^{\infty} \left(\sum_{i=1}^m \zeta_i e^{-((s-c_i)/\tau)^2} \right)^2 ds \right) L(1,0) \right)^{\frac{1}{2}} W_i(1) \quad (13)$$

where $W_i(r)$ is an independent set of Brownian motions which are also independent of those generated by the error processes ϵ_t and \mathbf{u}_t , and $L(t, s)$ is the standardised local time for a Brownian motion. The local time for a Brownian motion $W(r)$ with variance σ^2 is defined as:

$$L_W(t, s) = \lim_{\epsilon \rightarrow 0} \int_0^t 1\{s \leq W(r) < s + \epsilon\} \sigma^2 dr$$

and:

$$L(t, s) = \frac{1}{\sigma^2} L_W(t, s).$$

The probability density of $L(t, s)$ for $\sigma^2 = 1$ is given by:

$$Prob(L(t, s) \in dy) = \sqrt{2}/\sqrt{\pi t} e^{-(y+|s|)^2/2t} dy$$

so that:

$$Prob(L(1, 0) \in dy) = \sqrt{2}/\sqrt{\pi} e^{-y^2/2} dy.$$

From the above it follows that:

$$\sqrt[4]{T} \hat{\zeta}_2 \xrightarrow{d} \left(\left(\int_{-\infty}^{\infty} \left(e^{-((s-c)/\tau)^2} \right)^2 ds \right) L(1,0) \right)^{-1/2} W(1).$$

It easily follows from the above, using corollary 5.4 of Park and Phillips (2001) and Theorem 7 of Chang, Park, and Phillips (2001) that the Wald test of the hypothesis $\zeta_i = 0$ is asymptotically distributed as χ_m^2 . \square

However, a basic problem, with application of the above asymptotic theory is the fact that the RBF model parameters c_i and τ are neither known nor estimated by nonlinear least squares (NLS). Rather, the parameters c_i and τ are obtained by ad hoc methods to simplify the application of the test. If these parameters were known then the asymptotic distribution of β is straightforward as indicated by Theorem 1. If instead the parameters c and τ were estimated by NLS then the asymptotics would again be straightforward, but only under the alternative hypothesis. Under the null hypothesis, these parameters are not identified, leading to a Davies-type problem, named after the seminal work of Davies (1977) on underidentified parameters. We face a further difficulty. The method we use—with good reason—is neither of the above estimation cases. The parameter estimates are random variables but with unknown properties and unknown covariance matrix. Viewed over all parameters, the estimation is not NLS. However, it is clear that all parameter estimates are continuous functionals of the relevant Brownian motions. Moreover, the theory does not extend to multivariate models since in that case the additivity of the functions in the regression model is violated and the theory of Chang, Park, and Phillips (2001) does not go through. Our Monte Carlo simulations will clearly illustrate the first of these underlying problems.

Things are simpler for the TLG test. We can easily see that we do not need such complicated theory to deal with the bivariate model. We have the following theorem:

Theorem 2. *Under assumptions 1-3, for a bivariate model the asymptotic distribution of the TLG test statistic is given by (14) under the restriction that ϵ_t is uncorrelated with $(\mathbf{u}'_{t+1}, \mathbf{z}'_{t+1})'$.*

Proof. In the bivariate case the test statistics $W^{(i),ECM}$ can be written:

$$\begin{aligned} W^{(i),ECM} &= \sigma^{-2} \mathbf{y}' \mathbf{M} \mathbf{x}^{(i)} (\mathbf{x}^{(i)'} \mathbf{M} \mathbf{x}^{(i)})^{-1} \mathbf{x}^{(i)} \mathbf{M} \mathbf{y} \\ &= \sigma^{-2} \epsilon' \mathbf{M} \mathbf{x}^{(i)} (\mathbf{x}^{(i)'} \mathbf{M} \mathbf{x}^{(i)})^{-1} \mathbf{x}^{(i)} \mathbf{M} \epsilon \end{aligned}$$

where the second equality follows from the null hypothesis, $\mathbf{x}^{(i)} = (\mathbf{x}^2, \dots, \mathbf{x}^i)$, $\mathbf{x}^i = (x_1^i, \dots, x_T^i)'$, $\mathbf{y} = (y_1, \dots, y_T)'$, $\epsilon = (\epsilon_1, \dots, \epsilon_T)'$ and \mathbf{M} is the projection matrix of the stationary regressors and \mathbf{x} . We also define \mathbf{M}_x to be the projection matrix on \mathbf{x} only.

Anticipating the required rates of convergence, we need to obtain the asymptotic distribution of the two terms $A^{(i)}\mathbf{x}^{(i)'}\mathbf{M}\mathbf{x}^{(i)}A^{(i)}$ and $A^{(i)}\mathbf{x}^{(i)'}\mathbf{M}\epsilon$ where:

$$A^{(4)} = \begin{pmatrix} T^{-3/2} & 0 & 0 \\ 0 & T^{-2} & 0 \\ 0 & 0 & T^{-5/2} \end{pmatrix}$$

and $A^{(i)}$, $i = 2, 3$ are the $i - 1 \times i - 1$ top left hand corner submatrices of $A^{(4)}$. Note that $A^{(i)}\mathbf{x}^{(i)'}\mathbf{M}_x\mathbf{x}^{(i)}A^{(i)}$ has the same probability limit as $A^{(i)}\mathbf{x}^{(i)'}\mathbf{M}\mathbf{x}^{(i)}A^{(i)}$ and similarly for $A^{(i)}\mathbf{x}^{(i)'}\mathbf{M}_x\epsilon$, since the nonlinear regressors require faster rates for their asymptotics than the linear stationary regressors.

Then, by the continuous mapping theorem, we have that:

$$T^{-i-1} \sum_{t=1}^{[Tr]} x_{t-1}^{2i} \Rightarrow \sigma^{2i} \int_0^1 B(r)^{2i} dr.$$

where $B(r)$ is a Brownian motion independent of the one generated by ϵ_t . Further, by the martingale difference property of ϵ_t and by Theorem 2.1 of Kurtz and Protter (1991) we have that:

$$T^{-(i+1)/2} \sum_{t=1}^{[Tr]} x_{t-1}^i \epsilon_t \Rightarrow \sigma^i \int_0^1 B(r)^i dW(r).$$

Therefore:

$$\begin{aligned} W^{(i),ECM} &= \sigma^{-2} \mathbf{y}' \mathbf{M}_x \mathbf{x}^{(i)} A^{(i)} (A^{(i)} \mathbf{x}^{(i)'} \mathbf{M}_x \mathbf{x}^{(i)} A^{(i)})^{-1} A^{(i)} \mathbf{x}^{(i)} \mathbf{M}_x \mathbf{y} \\ &\Rightarrow \mathcal{W}_1^{(i)'} \mathcal{W}_2^{(i)-1} \mathcal{W}_1^{(i)} \equiv \mathcal{W}^{(i),ECM} \end{aligned} \quad (14)$$

where:

$$\mathcal{W}_1^{(4)} = \begin{pmatrix} \omega_{2,2} & \omega_{2,3} & \omega_{2,4} \\ \omega_{2,3} & \omega_{3,3} & \omega_{3,4} \\ \omega_{2,4} & \omega_{3,4} & \omega_{4,4} \end{pmatrix}, \quad \mathcal{W}_2^{(4)} = \begin{pmatrix} \omega'_2 \\ \omega'_3 \\ \omega'_4 \end{pmatrix}$$

$$\omega'_i = \int_0^1 B(r)^i dW(r) - \int_0^1 B(r)^{i+1} dr \left(\int_0^1 B(r)^2 dr \right)^{-1} \int_0^1 B(r) dW(r)$$

$$\omega_{i,j} = \int_0^1 B(r)^{i+j} dr - \int_0^1 B(r)^{i+1} dr \left(\int_0^1 B(r)^2 dr \right)^{-1} \int_0^1 B(r)^{j+1} dr$$

and $\mathcal{W}_1^{(i)}$, $i = 2, 3$ is the top left hand $i \times i$ submatrix of $\mathcal{W}_1^{(4)}$ and $\mathcal{W}_2^{(i)}$, $i = 2, 3$ is the $i \times 1$ top subvector of $\mathcal{W}_2^{(4)}$. \square

However, when a multivariate model is considered with more than one nonstationary regressor or if ϵ_t is correlated with $(\mathbf{u}'_{t+1}, \mathbf{z}'_{t+1})'$, the analysis of this theorem breaks down. For the multivariate case, cross products of the regressors appear and the fact that \mathbf{v}_t is cross correlated causes problems. Specifically, the covariances of the \mathbf{v}_t do not cancel out to give an asymptotic distribution that is free of nuisance parameters. We therefore do not explore the asymptotic distribution discussed above.

To overcome these issues for both tests we use the bootstrap. In what follows we explain the bootstrap method adopted and justify its use. Following many others in the literature, we use the sieve bootstrap. This essentially involves fitting a long autoregression to the estimated residuals $\hat{\epsilon}_t$, $\mathbf{v}_t = \Delta \mathbf{x}_t$ and, if present, to the stationary regressors. Define $\hat{\mathbf{q}}_t = (\mathbf{v}'_t, \mathbf{w}'_t, \hat{\epsilon}_t)'$. The vector autoregression then takes the form:

$$\hat{\mathbf{q}}_t = \sum_{i=1}^{p_T} \mathbf{Q}_i \hat{\mathbf{q}}_{t-i} + \epsilon_t.$$

We assume that the true model is given by:

$$\mathbf{q}_t = \sum_{i=1}^{\infty} \mathbf{Q}_i \mathbf{q}_{t-i} + \epsilon_t$$

with the property that:

Assumption 4. $\sum_{k=0}^{\infty} \mathbf{Q}_k$ is nonsingular and $\sum_{k=0}^{\infty} k^b \|\mathbf{Q}_k\| < \infty$, $b > 1$.

Given this, p_T can be chosen by an information criterion which will guarantee that $p_T \rightarrow \infty$ at a rate of $\ln T$ as discussed in, e.g., Ng and Perron (1995); Chang and Park (2003). Note that we do not impose any of the restrictions made earlier on the cross correlation structure of the errors. This does not affect the validity of our analysis. Once estimates of \mathbf{Q}_i are obtained then the estimated residuals $\hat{\epsilon}_t$ may be resampled as usual and combined with the estimated VAR model to provide bootstrap samples for \mathbf{q}_t denoted by \mathbf{q}_t^* . Cumulating \mathbf{v}_t^* gives bootstrap samples for \mathbf{x}_t which are used in $y_t^* = \hat{\beta}\mathbf{x}_t^* + \hat{\gamma}\mathbf{w}_t^* + \epsilon_t^*$ to construct bootstrap samples for y_t .

We have the following theorem concerning the validity of the bootstrap:

Theorem 3. *Under assumptions 1-4, the sieve bootstrap can estimate consistently the asymptotic distribution of the TLG and, assuming knowledge of m , c_i and τ , of the RBF test statistic.*

Proof. It is proven in Theorem 3.3 of Chang and Park (2003) following the work of Einmahl (1987) that:

$$\begin{pmatrix} y_t^* \\ \mathbf{x}_t^* \end{pmatrix} \equiv \begin{pmatrix} W_{1,T}^*(r) \\ \mathbf{W}_{1,T}^*(r) \end{pmatrix} \xrightarrow{d^*} \begin{pmatrix} W_1(r) \\ \mathbf{W}_1(r) \end{pmatrix} \quad (15)$$

in probability, where the notation d^* denotes that the convergence in distribution is conditional on the observed data realisation with respect to the conditional probability measure. The statement ‘in probability’ made above then reflects that the distributional result in (15) occurs with probability approaching one (in probability) with respect to the probability measure underlying the observed data. This result is referred to as the bootstrap invariance principle.

This result has been established for the case $\gamma = 0$. But it readily extends to our case once we define $\tilde{\epsilon}_t = \hat{\gamma}'\mathbf{w}_t + \epsilon_t$. The $T^{1/2}$ consistency of $\hat{\gamma}$, under

the null hypothesis, established in Theorem 7 of Chang, Park, and Phillips (2001) leads to the following result for $\tilde{\epsilon}_t^*$:

$$E^*|\tilde{\epsilon}_t^*|^a = O_p(1)$$

for some $a > 2$, via Lemma A.1 of Chang and Park (2003). This then straightforwardly leads to the following result by equation (1.3) of Einmahl (1987):

$$P\left(\sup_{0 \leq r \leq 1} |\mathbf{W}_T(r) - \mathbf{W}(r)| > 0\right) = o(1)$$

where $\mathbf{W} = (W_1, \mathbf{W}'_1)'$, proving the bootstrap invariance principle in our case.

By our assumptions this invariance principle satisfies Assumption 2.1 of Park and Phillips (2001) leading to the required result of the validity of the bootstrap in nonlinear regressions in the context of the analysis of Park and Phillips (2001). The analysis of TLG again follows straightforwardly given the bootstrap invariance principle, the continuous mapping theorem and Theorem 2.1 of Kurtz and Protter (1991). \square

Nevertheless, we have made clear that the neural network framework is not strictly speaking covered by the results of Park and Phillips (2001). Therefore, we have not proved rigorously that the application of the bootstrap is valid for our nonlinear regressions. There are two missing links related to the behaviour of the nonlinear functions of multiple integrated regressors. This behaviour (as explained by Park and Phillips (2001)) will depend crucially on the dimension of the regressor space because the spatial behaviour of a multivariate Brownian motion is dependent on its dimension. By the bootstrap invariance principle, however, this behaviour will be captured by the bootstrap. Further, the presence of any identifiable nuisance parameters will be taken into account by the bootstrap if they are consistently estimated. Finally, by the continuity of the parameter estimates c_i and τ with respect to the Brownian motions involved, we conjecture that the

bootstrap will accommodate this departure from the standard framework as well.

To summarise our results we have a full asymptotic theory for (i) the TLG test with a single x_t and (ii) the RBF test when estimated by NLS for a single x_t . Nuisance parameters appear in the asymptotic distribution of the TLG test when two or more x_t 's are considered. Further, the asymptotic distribution of the RBF test cannot be obtained when the centres and radii are determined by ad hoc methods rather than by a formal estimation procedure. Nevertheless, we should point out that these ad hoc methods enable a much faster and more robust application of the test without the need for iterative methods and, further, a number of studies have shown that they provide the basis for very powerful tests (see Blake and Kapetanios, 2000, 2003a,b). For all the above reasons we advocate the use of the bootstrap which we show to be valid for the TLG test and for the RBF test when estimated by NLS. We conjecture the validity of the bootstrap for the RBF test when implemented using ad hoc methods.

4 Monte Carlo experiments

In this section we carry out a Monte Carlo investigation of the properties of our proposed procedures. We concentrate on the following nonlinear cointegrating model:

$$y_t = \boldsymbol{\beta}'\mathbf{x}_t + f(\boldsymbol{\theta}, \mathbf{x}_t) + \epsilon_t \quad (16)$$

where $\epsilon_t \sim N(0, \sigma^2)$. We consider three possible functional forms for $f(\cdot, \cdot)$. These are:

$$f_1(\boldsymbol{\theta}, \mathbf{x}_t) = \sum_{j=1}^m \theta_j e^{-\sum_{i=1}^m \theta_{m+i} x_{i,t}^2} x_{j,t}, \quad \theta_i > 0, \quad i = m+1, \dots, 2m$$

$$f_2(\boldsymbol{\theta}, \mathbf{x}_t) = \sum_{j=1}^m \theta_j \frac{1}{1 + e^{-\sum_{i=1}^m \theta_{m+i} x_{i,t}}} x_{j,t}, \quad \theta_i > 0, \quad i = m+1, \dots, 2m$$

$$f_3(\boldsymbol{\theta}, \mathbf{x}_t) = \sum_{j=1}^m \theta_j I\left(\sum_{i=1}^m \theta_{m+i} x_{i,t} > 0\right) x_{j,t}.$$

These three functions provide an approximate span of the space of functions that have been considered most often in the empirical analysis of nonlinearity in econometrics. The first function is integrable and therefore the system will have a different behaviour when all x 's are close to zero compared to the case when one or more x 's are large. The second function is a logistic-type function and results in a system which behaves differently for large and large negative x 's. Finally, the third function is similar to the logistic but provides abrupt adjustment.

We consider the following values for the parameters. m is set to either 1 or 2. The case $m = 1$ can be dealt with using the existing asymptotic theory (abstracting from the issue of how c_i , and τ are chosen). $m = 2$ provides information on whether the bootstrap can deal with multivariate cointegrating relationships. We set $\beta_i = \theta_{m+i} = 1$, $i = 1, \dots, m$ and $\sigma^2 = 1$ throughout. $\theta_i = b$, $i = 1, \dots, m$ with $b = 0, 1, 5$ or 10 . This parameter controls the magnitude of the nonlinear effect. The value $b = 0$ is used to obtain the empirical size of the test. We consider samples of $T = 100, 200$ and 400 . We set the number of bootstrap replications to 99. The number of Monte Carlo replications is set to 1000. All tests are carried out at the 95% significance level. All random walks have zero as an initial value. Tables 1 and 2 provide estimated rejection frequencies.

The results make very interesting reading. Given the discussion in the previous section, the asymptotic RBF test is significantly overrejecting as expected. The degree of overrejection increases with the number of observations indicating that χ^2 asymptotics are not appropriate for this test. On the other hand, the size of the bootstrap tests is well behaved for all sample sizes considered. This result extends to multivariate models ($m = 2$). Indeed, the sizes are as well behaved in this case as in the case $m = 1$. This supports our argument that the bootstrap can deal both with multivariate models and

Table 1: Empirical Rejection Frequencies: $m = 1$

Obs		RBF (boot)			RBF (asym)			TLG		
		100	200	400	100	200	400	100	200	400
$b = 0$	f_1	0.072	0.043	0.050	0.204	0.192	0.238	0.089	0.070	0.060
	f_2	0.038	0.064	0.044	0.153	0.212	0.247	0.052	0.067	0.053
	f_3	0.050	0.050	0.043	0.179	0.204	0.253	0.058	0.066	0.056
$b = 1$	f_1	0.068	0.060	0.048	0.221	0.234	0.265	0.080	0.076	0.072
	f_2	0.634	0.685	0.783	0.731	0.781	0.866	0.657	0.715	0.801
	f_3	0.614	0.725	0.773	0.707	0.795	0.836	0.636	0.751	0.780
$b = 5$	f_1	0.503	0.400	0.358	0.683	0.620	0.647	0.461	0.349	0.232
	f_2	0.935	0.939	0.950	0.962	0.970	0.977	0.948	0.941	0.951
	f_3	0.803	0.844	0.879	0.844	0.876	0.909	0.821	0.855	0.881
$b = 10$	f_1	0.775	0.689	0.693	0.865	0.835	0.854	0.619	0.472	0.382
	f_2	0.983	0.991	0.991	0.993	0.997	0.996	0.984	0.990	0.982
	f_3	0.854	0.886	0.925	0.883	0.912	0.942	0.861	0.892	0.923

Table 2: Empirical Rejection Frequencies: $m = 2$

Obs		RBF (boot)			RBF (asym)			TLG		
		100	200	400	100	200	400	100	200	400
$b = 0$	f_1	0.044	0.044	0.047	0.136	0.144	0.178	0.084	0.059	0.058
	f_2	0.057	0.049	0.051	0.145	0.175	0.171	0.077	0.057	0.059
	f_3	0.052	0.048	0.039	0.150	0.159	0.156	0.065	0.064	0.045
$b = 1$	f_1	0.061	0.059	0.042	0.144	0.151	0.169	0.079	0.060	0.050
	f_2	0.506	0.637	0.735	0.591	0.723	0.791	0.602	0.710	0.779
	f_3	0.546	0.676	0.779	0.629	0.762	0.833	0.649	0.737	0.820
$b = 5$	f_1	0.103	0.083	0.065	0.225	0.222	0.232	0.151	0.125	0.091
	f_2	0.737	0.799	0.862	0.822	0.865	0.910	0.866	0.886	0.914
	f_3	0.771	0.814	0.878	0.819	0.866	0.915	0.835	0.854	0.899
$b = 10$	f_1	0.225	0.157	0.136	0.371	0.325	0.306	0.279	0.200	0.151
	f_2	0.801	0.855	0.912	0.897	0.932	0.958	0.945	0.953	0.960
	f_3	0.810	0.841	0.876	0.862	0.883	0.916	0.860	0.878	0.900

with the data dependence of the parameter values used for c_i and τ .

Moving on to the power of the tests, we observe some interesting patterns. Overall, the TLG test appears more powerful. It sometimes has higher rejection probabilities compared to the asymptotic RBF test whose empirical size is wrong, as we know. So from a first look this might be the preferred test of nonlinearity in cointegrating relationships.

However, this conclusion is undermined by the results obtained for f_1 . There we see that for the sample sizes we consider although the TLG test is more powerful, its power is reduced as the sample size grows. A slight reduction also appears for the RBF test. We have carried out further experiments with the asymptotic versions of these tests. In particular, for $m = 1$ the TLG test has a nuisance parameter free distribution discussed in the previous section. For the RBF test, the asymptotic distribution will depend on the ad hoc method used to obtain c_i and τ . For the particular method we use we have obtained the critical values under the null hypothesis and used these for our work. To evaluate the power patterns we have increased the sample size, up to 2000 observations. Such an increase, makes the evaluation of the bootstrap through Monte Carlo computationally expensive. In any case, our work illustrates the relevant points without resorting to the bootstrap. The results are surprising. The TLG test keeps losing power as the sample size increases. The RBF has a dip in power but eventually the power increases as the sample size grows. This pattern for RBF is discernible, for example, in Table 1 for $m = 1$ and $b = 10$.

Why is this happening for f_1 ? Note that the TLG test is not actually based on a neural network per se but on a Taylor approximation of a logistic neural network. Therefore, the universal approximation properties of neural networks do not extend straightforwardly to the TLG formulation. The explanation of what is happening is relatively simple. The nonlinearity for f_1 appears only for a small part of the state space, since f_1 is integrable. In particular, the nonlinearity is activated for $O_p(\sqrt{T})$ observations only,

i.e. those for which the random walk in x_t resides in the area of the state space that activates f_1 . For all the other observations—which form the vast majority of the sample—there is no nonlinearity and the coefficients of x_t^i are zero. Given the predominance of such observations, it is not surprising that the TLG method has no power asymptotically. On the other hand the RBF test is based on an integrable functional form (which in our case coincides, as well, with f_1). Therefore, the constructed variables are zero for large x_t , unlike TLG. Therefore the coefficients of the constructed variables are consistently estimated and different from zero. The low power of the test is easily explained by the fact that the constructed variables are activated only for a short time in the sample. In other words as we mentioned above there are only $O_p(\sqrt{T})$ effective observations with which to detect nonlinearity. Formally, this is reflected by the rate of convergence obtained for coefficients of integrable functions which is $T^{1/4} = \sqrt{\sqrt{T}}$, as discussed in Park and Phillips (2001). Finally, we note that the above provides a further reason to prefer of the RBF over the logistic function (which is not integrable) to construct neural network neglected nonlinearity tests in this context.

It seems apparent that the TLG test is more powerful in small samples than the RBF test. But it is not consistent for the class of nonlinearities generated by integrable functions. Even so, in small samples TLG may be more powerful than RBF even for such functions as seen from the Monte Carlo. In practise, we suggest that either TLG is used, or a combination of the two tests perhaps formalised through the use of the Bonferonni inequality.³ More specifically, we know that for some events in a probability space, E_1, \dots, E_n , the probability of their union is smaller than or equal to the sum of the individual probabilities of each event, i.e. $P(E_1 \cup \dots \cup E_n) \leq P(E_1) + \dots + P(E_n)$. This implies that $P(\cap_{i=1}^n E_i) \geq 1 - \sum_{i=1}^n P(\bar{E}_i)$ where \bar{E}_i denotes the complement of E_i . If we define a new procedure which rejects the null of linearity if either the RBF test or the TLG test rejects the null then we can use the

³See Blake and Kapetanios (2003a) for an application of this to unit root testing.

Bonferroni inequality to put an upper bound on the Type I error of the new procedure. We will not investigate this procedure further.

5 Conclusion

In the past two decades, considerable effort has been placed on the investigation of nonstationary and nonlinear processes. Nevertheless the interplay of nonlinearity and nonstationarity has not been widely explored. In the context of investigating long run economic relationships there has been no empirical work on the exploration of the possibility that such relationships may have nonlinear characteristics despite plenty of evidence in favour of nonlinearity in the adjustment process towards these long run relationships. On a theoretical level, work by Phillips and his co-authors in a series of papers headed by Park and Phillips (2001) has made such an investigation feasible.

In this paper we propose two new tests for the presence of neglected nonlinearity in long run relationships. We investigate a number of issues concerning the asymptotic properties of these test and conclude that the use of the bootstrap provides the necessary flexibility for applying this test to a wide variety of settings. We therefore provide a theoretical justification for the use of bootstrap in this context. A Monte Carlo study clearly shows the usefulness of the new tests and the trade-offs involved in choosing one or the other. Further research is needed mainly at an empirical level to illustrate the relevance of this work in practice.

References

- AKAIKE, H. (1976): “Canonical Correlation Analysis of Time Series and the Use of an Information Criterion,” in *System Identification*, ed. by R. Mehra, and D. Lainiotis. Academic Press.
- BISHOP, C. M. (1995): *Neural Networks for Pattern Recognition*. Oxford University Press.
- BLAKE, A. P., AND G. KAPETANIOS (2000): “A Radial Basis Function Artificial Neural Network Test for ARCH,” *Economics Letters*, 69, 15–23.
- (2003a): “Pure Significance Tests of the Unit Root Hypothesis Against Nonlinear Alternatives,” *Journal of Time Series Analysis*, 24(3), 253–267.
- (2003b): “A Radial Basis Function Artificial Neural Network Test for Neglected Nonlinearity,” *The Econometrics Journal*, 6(2), 357–373.
- CAMPBELL, J. Y., A. W. LO, AND A. C. MACKINLAY (1997): *The Econometrics of Financial Markets*. Princeton University Press.
- CHANG, Y., AND J. Y. PARK (2003): “Bootstrapping Cointegrating Regressions,” *Mimeo, Rice University*.
- CHANG, Y., J. Y. PARK, AND P. C. B. PHILLIPS (2001): “Nonlinear Econometric Models with Cointegrated and Deterministically Trending Regressors,” *The Econometrics Journal*, 4, 1–36.
- CYBENKO, G. (1989): “Approximation by Superpositions of a Sigmoidal Function,” *Mathematics of Control, Signals and Systems*, 2, 304–314.
- DAVIES, R. B. (1977): “Hypothesis Testing when a Nuisance Parameter is Present Only under the Alternative,” *Biometrika*, 64(2), 247–254.
- EINMAHL, U. (1987): “A Useful Estimate in the Multidimensional Invariance Principle,” *Probability Theory and Related Fields*, 76, 81–101.
- HORNIK, K., M. STINCHCOMBE, AND H. WHITE (1989): “Multi-Layer Feedforward Networks and Universal Approximators,” *Neural Network*, 2, 359–366.

- KURTZ, T. G., AND P. PROTTER (1991): “Weak Limit Theorems for Stochastic Integrals and Stochastic Differential Equations,” *Annals of Probability*, 19(3), 1035–1070.
- LEE, T. H., H. WHITE, AND C. W. J. GRANGER (1993): “Testing for Neglected Nonlinearity in Time Series Models: A Comparison of Neural Network Methods and Alternative Tests,” *Journal of Econometrics*, 56, 269–290.
- NG, S., AND P. PERRON (1995): “Unit Root Tests in ARMA Models with Data Dependent Methods for the Truncation Lag,” *Journal of the American Statistical Association*, 429, 268–281.
- ORR, M. J. (1995): “Regularisation in the Selection of Radial Basis Function Centers,” *Neural Computation*, 7(3), 606–623.
- PARK, J. Y., AND P. C. B. PHILLIPS (1999): “Asymptotics for Nonlinear Transformations of Integrated Time Series,” *Econometric Theory*, 15, 269–298.
- (2001): “Nonlinear Regressions with Integrated Time Series,” *Econometrica*, 69, 117–161.
- SCHWARZ, G. (1978): “Estimating the Dimension of a Model,” *Annals of Statistics*, 6, 461–464.
- STINCHCOMBE, M., AND H. WHITE (1989): “Universal Approximation Using Feedforward Networks with Non-Sigmoid Hidden Layer Activation Functions,” *Proceedings of the International Joint Conference on Neural Networks*, 1.
- TERÄSVIRTA, T., C. F. LIN, AND C. W. J. GRANGER (1993): “Power of the Neural Network Linearity Test,” *Journal of Time Series Analysis*, 14, 209–220.

**This working paper has been produced by
the Department of Economics at
Queen Mary, University of London**

**Copyright © 2004 Andrew P. Blake and George Kapetanios
All rights reserved.**

**Department of Economics
Queen Mary, University of London
Mile End Road
London E1 4NS
Tel: +44 (0)20 7882 5096 or Fax: +44 (0)20 8983 3580
Email: j.conner@qmul.ac.uk
Website: www.econ.qmul.ac.uk/papers/wp.htm**