Department of Economics

Testing for Neglected Nonlinearity in Long Memory Models

George Kapetanios

Working Paper No. 473 November 2002 ISSN 1473-0278

Queen Mary
University of London

Testing for Neglected Nonlinearity in Long Memory Models

George Kapetanios*
Queen Mary , University of London

November 2002

Abstract

Interest in the interface of nonstationarity and nonlinearity has been increasing in the econometric literature. The motivation for this development maybe be traced to the perceived possibility that processes following nonlinear models maybe mistakenly taken to be unit root or long-memory nonstationary. This paper considers the possibility that processes may exhibit both long memory and nonlinearity. We test against the possibility that the process u_t in the model $(1-L)^d y_t = u_t$ is nonlinear. We do not assume a particular parametric form for the nonlinear process but construct a pure significance test. Clearly, such a test could be straightforwardly constructed if d were known. Unfortunately, if a linear model is assumed while estimating d the power of the test will be reduced. We propose new more powerful tests for this problem. We present Monte Carlo evidence on the performance of the new tests and apply them to Yen real exchange rates.

JEL Classification: C22, C12, F31.
Key Words: Long Memory, Nonlinearity, Neural Networks, Real Exchange Rates.

*Department of Economics, Queen Mary, University of London, Mile End Road, London E1 4NS. email: G.Kapetanios@qmul.ac.uk

1 Introduction

Interest in the interface of nonstationarity and nonlinearity has been increasing in the econometric literature. The motivation for this development maybe be traced to the perceived possibility that processes following nonlinear models maybe mistakenly taken to be unit root nonstationary. Previous work in this area includes Enders and Granger (1998), Caner and Hansen (2001), Kapetanios, Snell, and Shin (2002) and Kapetanios and Shin (2002a).

A related stand of the literature on the investigation of highly persistent processes looks at the presence of long memory in the data. Long memory and nonlinearity have rarely been jointly analysed. Exceptions include Davidson and Sibbertsen (2002), Diebold and Inoue (2001), van Dijk, Frances, and Paap (2002) and Kapetanios and Shin (2002b) Within this small set of papers two strands are apparent. One strand considers long-memory and nonlinearity as alternative representations which maybe confused and tries to investigate their similarities and differences. Diebold and Inoue (2001) juxtapose the covariance structures of long memory and Markov switching models. Davidson and Sibbertsen (2002) discusses one class of nonlinear models which have a similar covariance structure to long memory models. Kapetanios and Shin (2002b) suggest a formal test for distinguishing between nonstationary long memory and nonlinear geometrically ergodic models in small samples. On the other hand van Dijk, Frances, and Paap (2002) investigate the possibility that the nature of the process driving the long memory process is nonlinear. They apply such a model to US unemployment data with interesting results.

This paper is in the spirit of the second strand. We test against the possibility that the process u_t in the model $(1-L)^d y_t = u_t$ is nonlinear. We do not assume a particular parametric form for the nonlinear process but construct a pure significance test. Clearly, such a test could be straightforwardly constructed if d were known. Unfortunately, if a linear model is assumed while estimating d the power of the test will be reduced. We therefore suggest a model based on a neural network approximation to estimate d prior to applying standard linearity tests. We apply the new tests to Yen real exchange rates. We find significant evidence of neglected nonlinearity in a number of series.

The structure of the paper is as follows: Section 2 presents the framework of the analysis. Sections 3 and 4 discusses the tests and their implementation. Section 5 presents the Monte Carlo study. Section 6 presents the empirical application. Finally, Section 7 concludes. The Appendix contains the proof of the main theorem of the paper.

2 Nonlinear long memory models

We consider the following general fractionally integrated model

$$(1-L)^d y_t = u_t, \ -1/2 \geq d < 3/2, \ t = 1, \ldots, T \tag{1}$$

u_t is an $I(0)$ process. We define I(0) processes, following De Jong and Davidson (2000), to be processes whose partial sums converge weakly to Brownian motion. The model for y_t can be written as an infinite moving average in terms of u_t

$$y_t = \sum_{i=0}^{\infty} a_i^d u_{t-i} \tag{2}$$

where $a_i^d = \frac{\Gamma(d+1)}{\Gamma(i+1)\Gamma(d-i+1)}(-1)^i$. It can equivalently be written as an infinite autoregression given by

$$y_t = \sum_{i=1}^{\infty} b_i^d y_{t-i} + u_t \tag{3}$$

where $b_i^d = -\frac{\Gamma(i-d)}{\Gamma(i+1)\Gamma(-d)}$. A standard specification for the weakly dependent process u_t is that u_t follows an ARMA(p,q) process i.e.

$$A(L)u_t = B(L)\epsilon_t$$

where ϵ_t is an i.i.d. process with finite variance. This gives rise to the well known ARFIMA(p,d,q) model. However, a straightforward extension which has recently been considered is that u_t follows a geometrically ergodic nonlinear process. van Dijk, Frances, and Paap (2002)) have specified u_t to follow an ESTAR process given by

$$u_t = \alpha_0 + \sum_{i=1}^{p_l} \alpha_i u_{t-i} + \sum_{i=1}^{p_n} \beta_i \left[1 - e^{-\gamma_1 (u_{t-d} - \gamma_0)^2} \right] u_{t-i} + \epsilon_t$$

This model has been applied to the investigation of macroeconomic series and evidence from US data indicated the presence of nonlinearity of ESTAR form. However, there is no need to restrict to this form of nonlinearity and u_t can be modelled in terms of other nonlinear econometric models that have been suggested in the literature such as, for example, threshold autoregressive models or bilinear models. So the general form of the model we consider is that

$$u_t = F(u_{t-1} \dots u_{t-p}) + \epsilon_t \tag{4}$$

Clearly the validity of any tests for linearity based on a particular nonlinear model is conditional on the choice of the model and therefore a pure significance test for neglected nonlinearity may be more appropriate compared to tests based on particular nonlinear models. A wide variety of pure significance tests exist. Tests based on neural networks have been consistently found to have good power properties and are therefore favorites in the linearity testing literature.

However, an important complication arises compared to standard linearity testing. This is that the long memory parameter d needs to be known or consistently estimated prior to the application of the nonlinearity test on u_t obtained by fractional differencing of the original series y_t. Standard parametric methods of estimating d are based on ARFIMA models. Clearly this estimation strategy is not appropriate since under the nonlinearity hypothesis the ARFIMA model is misspecified. Fitting an ARFIMA model to determine d will lead to an inconsistent estimate of d under the alternative hypothesis of neglected nonlinearity and by construction to a test for nonlinearity which is likely to be less powerful than one based on the true value of d. It should be made clear that this issue is related primarily to the power of the test. Under the null hypothesis d will be estimated consistently through an ARFIMA model and therefore the test will be correctly sized. Of course, an alternative which we do not pursue here is nonparametric estimation of the long memory parameter using, e.g., frequency domain based techniques. However the spirit of our analysis is parametric in the long memory dimension of the problem. The test we suggest is clearly a first step to a parametric analysis of the neglected nonlinearity with a model belonging to the class of nonlinear models used to investigate weakly dependent stationary processes such as TAR or STAR models.

The solution we suggest is to fit a neural network type model to u_t and estimate this model. Then, the estimate of d, thus obtained is used to fractionally difference y_t and to obtain an estimate of u_t. This estimate is then tested for nonlinearity using standard neural network tests described in the next section. Alternatively, the significance of the neural network model used to estimate d could be tested to determine the presence of nonlinearity.

3 Neural network models and tests

We consider two different but related neural network tests for neglected nonlinearity.

3.1 The Lee, White, and Granger (1993) test

The null hypothesis of this test in our framework is that the conditional mean of u_t given lags of u_t is a linear function of these lags or

$$P(E(u_t|u_{t-1}\ldots u_{t-p}) = \delta_0 + \sum_{i=1}^{p} \delta_i u_{t-p}) = 1 \tag{5}$$

The Lee, White, and Granger (1993) (henceforth ANN) test specifies that $F(.)$ in (4) is given by $\sum_{j=1}^{q} \phi(\sum_{i=1}^{p} \gamma_{ij} u_{t-i})$ where $\phi(\lambda)$ is the logistic function, given by $[1 + \exp(-\lambda)]^{-1}$. The coefficients γ_{ij} are randomly generated from a uniform distribution over $[\gamma_l, \ \gamma_h]$. For given q, the constructed regressors $\phi(\sum_{i=1}^{p} \gamma_{ij} u_{t-i})$, $j = 1, \ldots, q$ may suffer from multicollinearity. We follow Lee, White, and Granger (1993) and suggest that $\tilde{q}$ largest principle components of the constructed regressors excluding the largest one be used as regressors in

$$u_t = \beta_0 + \sum_{j=1}^{\tilde{q}} \beta_j \phi(\sum_{i=1}^{p} \gamma_{ij} u_{t-i}) + \epsilon_t \tag{6}$$

We then perform a Wald test of the joint significance of the constructed regressors. This test tests the null hypothesis that $\beta_0 = \beta_1 = \ldots = \beta_{\tilde{q}} = 0$. This takes the form

$$\frac{1}{\hat{\sigma}^2} \hat{\boldsymbol{\beta}}' (R'(W'W)^{-1}R)^{-1} \hat{\boldsymbol{\beta}} \tag{7}$$

where W is the matrix of regressors of (6) and a constant, R is the selector matrix, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_q)'$ and $\hat{\sigma}^2$ is the estimated variance of the residuals in (6).

3.2 The Taylor expansion test

An alternative test that is motivated by the logistic neural network is the test proposed by Teräsvirta, Lin, and Granger (1993) and used by Blake and Kapetanios (2003). That test approximates by a Taylor expansion the logistic neural network and subsequently substitutes this expansion in the models and tests for its significance. Teräsvirta, Lin, and Granger (1993) suggest the use of the third order Taylor expansion. In our framework, the model for u_t then takes the form

$$u_t = \beta_0 + \sum_{i=1}^{3} \sum_{i=1}^{p} \beta_{i,j} u_{t-i}^{j} + \sum_{i=1}^{p-1} \sum_{j=i+1}^{p} \gamma_{1,i,j} u_{t-i} u_{t-j} + \sum_{s=0}^{1} \sum_{i=1}^{p-1} \sum_{j=i+1}^{p} \gamma_{2,s,i,j}^{2} u_{t-i}^{2-s} u_{t-j}^{s+1} + \epsilon_t \tag{8}$$

Clearly, this is just one Taylor expansion that can be used for approximating the unknown function. We consider also expansions of order 2 and 4 giving rise to the following models

$$u_t = \beta_0 + \sum_{i=1}^{2} \sum_{i=1}^{p} \beta_{i,j} u_{t-i}^{j} + \sum_{i=1}^{p-1} \sum_{j=i+1}^{p} \gamma_{1,i,j} u_{t-i} u_{t-j} + \epsilon_t \tag{9}$$

$$u_t = \beta_0 + \sum_{i=1}^{4} \sum_{i=1}^{p} \beta_{i,j} u_{t-i}^{j} + \sum_{i=1}^{p-1} \sum_{j=i+1}^{p} \gamma_{1,i,j} u_{t-i} u_{t-j} + \sum_{i=1}^{p-1} \sum_{j=i+1}^{p} \gamma_{1,i,j} u_{t-i}^{2} u_{t-j}^{2} + \sum_{s=0}^{1} \sum_{i=1}^{p-1} \sum_{j=i+1}^{p} \gamma_{2,s,i,j}^{2} u_{t-i}^{2-s} u_{t-j}^{s+1} + \sum_{s=0}^{1} \sum_{i=1}^{p-1} \sum_{j=i+1}^{p} \gamma_{2,s,i,j}^{2} u_{t-i}^{3-2s} u_{t-j}^{2s+1} + \epsilon_t \tag{10}$$

The null hypothesis of the γ coefficients being zero is tested using a Wald test. We will refer to the models underlying these tests as the TLG_i models $i = 2, 3, 4$.

3.3 The neural network model

As explained in the previous section we need to take account of the possibility of nonlinearity in u_t when estimating d. The neural network specifications on which both tests, presented above, rely can be used as a model for estimating d. We suggest the use of the Taylor expansion. The reason for this choice is that estimation of the logistic neural network model used by the White test involves nonlinear least squares and is computationally expensive. The model based on the Taylor expansion of the logistic neural network on the other hand can be estimated by OLS.

4 Implementation of the tests

The first step in the implementation of the test is the estimation of d. Following the discussion in the previous section we use the TLG_i, $i = 2, 3, 4$ to estimate d where we use the infinite AR representation of y_t in terms of u_t

to write u_t. We therefore numerically minimise the sum of squared residuals of the following models:

$$y_t = \beta_0+\sum_{l=0}^{t-p} b_l^d y_{t-l}+\sum_{i=1}^{2}\sum_{i=1}^{p} \beta_{i,j}[y_t-\sum_{l=0}^{t-p} b_l^d y_{t-l}]^j+\sum_{i=1}^{p-1}\sum_{j=i+1}^{p} \gamma_{1,i,j}[y_t-\sum_{i=0}^{t-p} b_i y_{t-i}][y_t-\sum_{i=0}^{t-p} b_i y_{t-i}]+\epsilon_t \tag{11}$$

$$y_t = \beta_0+\sum_{l=0}^{t-p} b_l^d y_{t-l}+\sum_{i=1}^{3}\sum_{i=1}^{p} \beta_{i,j}[y_t-\sum_{l=0}^{t-p} b_l^d y_{t-l}]^j+\sum_{i=1}^{p-1}\sum_{j=i+1}^{p} \gamma_{1,i,j}[y_t-\sum_{l=0}^{t-p} b_l^d y_{t-l}][y_t-\sum_{l=0}^{t-p} b_l^d y_{t-l}]+ \tag{12}$$

$$\sum_{s=0}^{1}\sum_{i=1}^{p-1}\sum_{j=i+1}^{p} \gamma_{2,s,i,j}^2[y_t - \sum_{l=0}^{t-p} b_l^d y_{t-l}]^{2-s}[y_t - \sum_{l=0}^{t-p} b_l^d y_{t-l}]^{s+1} + \epsilon_t$$

$$y_t = \beta_0+\sum_{l=0}^{t-p} b_l^d y_{t-l}+\sum_{i=1}^{4}\sum_{i=1}^{p} \beta_{i,j}[y_t-\sum_{l=0}^{t-p} b_l^d y_{t-l}]^j+\sum_{i=1}^{p-1}\sum_{j=i+1}^{p} \gamma_{1,i,j}[y_t-\sum_{l=0}^{t-p} b_l^d y_{t-l}][y_t-\sum_{l=0}^{t-p} b_l^d y_{t-l}]+ \tag{13}$$

$$\sum_{i=1}^{p-1}\sum_{j=i+1}^{p} \gamma_{1,i,j}[y_t-\sum_{l=0}^{t-p} b_l^d y_{t-l}]^2[y_t-\sum_{l=0}^{t-p} b_l^d y_{t-l}]^2+\sum_{s=0}^{1}\sum_{i=1}^{p-1}\sum_{j=i+1}^{p} \gamma_{2,s,i,j}^2[y_t-\sum_{l=0}^{t-p} b_l^d y_{t-l}]^{2-s}[y_t-\sum_{l=0}^{t-p} b_l^d y_{t-l}]^{s-}$$

$$\sum_{s=0}^{1}\sum_{i=1}^{p-1}\sum_{j=i+1}^{p} \gamma_{2,s,i,j}^2[y_t - \sum_{l=0}^{t-p} b_l^d y_{t-l}]^{3-2s}[y_t - \sum_{l=0}^{t-p} b_l^d y_{t-l}]^{2s+1} + \epsilon_t$$

The lag order of the models p may be determined by an information criterion or chosen a priori. Once d has been determined, the two nonlinearity tests of the previous section maybe applied, using the inifinite AR representation of y_t to obtain u_t. Under the null hypothesis of no nonlinearity and conditional on knowing or estimating consistently the lag order p for the tests we have the following theorem

Theorem 1 *Under the null hypothesis of linearity given by (5) and given lag order p the asymptotic distribution of the ANN and TLG tests does not change when the tests are based on* $\hat{u}_t = y_t - \sum_{l=0}^{t-p} b_l^{\hat{d}} y_{t-l}$ *rather than* $\hat{u}_t = y_t - \sum_{l=0}^{t-p} b_l^d y_{t-l}$

For a proof see the Appendix. Since the asymptotic distribution of the test statistic for known d is simply a χ^2 we have that the test we propose which uses an estimate of d is χ^2 as well.

5 Monte Carlo study

We carry out a Monte Carlo study to investigate the size and power properties of the new tests we propose. The Monte Carlo experiment considers neglected nonlinearity of the ESTAR form. This is the form of nonlinearity investigated by van Dijk, Frances, and Paap (2002) in their analysis of US unemployment data. We look at two size experiments where the model generating the data is an $ARFIMA(0, 0.6, 0)$ and an $ARFIMA(1, 0.6, 0)$ with AR coefficient 0.8 respectively. We also consider 8 power experiments where the alternative nonlinear hypothesis is an fractionally integrated model with $d = 0.6$ and u_t follows an ESTAR model. The precise specification of the ESTAR models is given below for experiments 3-10

- Exp. 3 $\alpha_0 = 0$, $\gamma_0 = 0$, $\alpha_1 = 0.8$, $\beta_1 = -1.5$ $\gamma_1 = 0.01$
- Exp. 4 $\alpha_0 = 0$, $\gamma_0 = 0$, $\alpha_1 = 0.8$, $\beta_1 = -1$ $\gamma_1 = 0.01$
- Exp. 5 $\alpha_0 = 0$, $\gamma_0 = 0$, $\alpha_1 = 0.8$, $\beta_1 = -1.5$ $\gamma_1 = 0.05$
- Exp. 6 $\alpha_0 = 0$, $\gamma_0 = 0$, $\alpha_1 = 0.8$, $\beta_1 = -1$ $\gamma_1 = 0.05$
- Exp. 7 $\alpha_0 = 0$, $\gamma_0 = 0$, $\alpha_1 = 1.3$, $\beta_1 = -1.5$ $\gamma_1 = 0.01$
- Exp. 8 $\alpha_0 = 0$, $\gamma_0 = 0$, $\alpha_1 = 1.3$, $\beta_1 = -1$ $\gamma_1 = 0.01$
- Exp. 9 $\alpha_0 = 0$, $\gamma_0 = 0$, $\alpha_1 = 1.3$, $\beta_1 = -1.5$ $\gamma_1 = 0.05$
- Exp. 10 $\alpha_0 = 0$, $\gamma_0 = 0$, $\alpha_1 = 1.3$, $\beta_1 = -1$ $\gamma_1 = 0.05$

All experiments represent geometrically ergodic processes for u_t. The last four experiments allow for the corridor regime of the nonlinear process (i.e.

the regime closer to the mean of the process) to be locally explosive as the polynomial of the AR part of the specification at the mean has a root which is less than one. Such processes have been found to be of use for modelling macroeconomic series such as US GDP by Kapetanios (2003). We apply the White and the TLG tests to the process $\hat{u}_t = u_t(\hat{d})$ where $\hat{d}$ has been obtained from estimation of the model TLG_i, $i = 1, 2, 3, 4$, TLG_1 refers to a linear model. We also estimate the linear model both in the time and the frequency domain[1]. The White test is denoted by W_i^s, $s = t, f$, $i = 0, 1, 2, 3, 4$, where the subscript i refers to the TLG model used to estimate d and the superscript s refers to estimation in the time or frequency domain. $i = 0$ indicates that the true value of d has been used. The TLG test is denoted by $TLG_{j,i}^s$, $s = t, f$, $i = 0, 1, 2, 3, 4$, $j = 3$, where the subscript i refers to the TLG model used to estimate d, the superscript s refers to estimation in the time or frequency domain and finally the subscript j refers to the order of the Taylor expansion used to test the null hypothesis of linearity. We use $j = 3$ following suggestions by Teräsvirta, Lin, and Granger (1993). Of course different j could be envisaged following e.g. Blake and Kapetanios (2003) but no significant difference in performance is observed and so for simplicity we concentrate on $j = 3$. $i = 0$ indicates that the true value of d has been used. The error term ϵ_t is generated from a $N(0, 1)$ throughout. We present results for samples of size $T = 100, 150, 200$. We present both rejection probabilities and average estimates of d. The tests based on TLG_3 and TLG_4 overreject under the null hypothesis. To correct for this when presenting power results we use empirical critical values. These are obtained by using the 95% quantile of the empirical distribution of the test statistic under the null hypothesis represented as experiment 1 in the Monte Carlo study. For similar treatments in other statistical testing contexts see e.g. Elliott, Rothenberg, and Stock (1996). For the White test we follow Lee, White, and Granger (1993) and set $q = 10$, $\tilde{q} = 2$, $\gamma_h = 2$ and $\gamma_l = -2$. Also

[1]For details on estimation in the frequency domain see Harvey (1989).

$p = 1$ throughout the Monte Carlo study. Results are presented in Tables 1 to 9.

The results are revealing. Firstly, concerning the estimate of d we see that although in a majority of the experiments the estimator of d is relatively well behaved there are experiments where the estimators based on linear models are biased especially when the estimation is in the frequency domain. The estimates based on the models which approximate the nonlinear model are always well behaved. However, the most interesting results concern the rejection frequencies of the tests. The tests based on estimates of d using linear models are uniformly less powerful that the tests based on estimates of d from models approximating the nonlinear models. In some cases the advantage can reach 20%. More interestingly the proposed tests have comparable power to tests based on the true value of d. Such tests clearly possess the property of reaching the upper bound in terms of power given specific linearity tests. Therefore, the suggested procedures provide a clear advantage compared to standard methods.

6 Empirical Application

In this section we apply the new tests to investigate the presence of neglected nonlinearity in Yen real exchange rates. Our choice of data set reflects previous work in this area by Cheung and Lai (2001) who investigated the presence of long memory in Yen real exchange rates aiming to explain the puzzle of the inability to reject the null hypothesis of unit root nonstationarity using standard unit root tests.

We construct bilateral yen real exchange rate against the *i-th* currency at time t $(q_{i,t})$ as $q_{i,t} = s_{i,t} + p_{J,t} - p^{*}_{i,t}$, where $s_{i,t}$ is the corresponding nominal exchange rate (*i-th* currency per yen), $p_{J,t}$ the price level in Japan, and $p^{*}_{i,t}$ the price level of the *i-th* country. Thus, a rise in $q_{i,t}$ implies a real yen appreciation against the *i-th* currency. The price levels are consumer price indices and all variable are in logs. All data are from the International Monetary Fund's *International Financial Statistics* in CD-ROM. The data are not seasonally adjusted. All data are quarterly, spanning from 1960Q1 to 2000Q4 and the bilateral nominal exchange rates against the currencies other than the US dollar are cross-rates computed using the US dollar rates. We consider a very large sample of countries in an attempt to make the empirical analysis more comprehensive.

The countries we consider are: USA, Germany, France, Italy, UK, Canada, Austria, Belgium, Portugal, Spain, Sweden, Australia, Korea, Malaysia, Indonesia, Thailand, Philippines and Sri Lanka.

We estimate the long memory parameter, d, using both a linear ARFI model and a model using a neural network approximation where we use a third order Taylor expansion. The model is estimated by minimising the conditional sum of squares. The use of this algorithm enables straightforward estimation for the neural network approximation model. In both cases the lag order is chosen using the Bayesian information criterion. We then apply both the White and TLG tests[2] using the estimate of d obtained both from the linear and nonlinear models. For the White test we follow Lee, White, and Granger (1993) and set $q = 10$, $\tilde{q} = 2$, $\gamma_h = 2$ and $\gamma_l = -2$. Results are presented in Table 10. Note that the probability values given are from the χ^2 tables. We have not used the empirical critical values obtained from the Monte Carlo study. Nevertheless, it is easy to check (and has been checked) that all rejections obtained for countries using an estimate of d obtained from the neural network model would not be reversed if the empirical critical values had been used as all the rejections are strong.

[2]A third order Taylor expansion is used for the TLG test.

We see that evidence for nonlinearity is widespread in the dataset we consider. There are seven countries (out of eighteen) for which both the White and TLG tests reject the null hypothesis of no neglected nonlinearity at the 5% significance level when $\hat{d}$ from the neural network approximation model is used. When $\hat{d}$, estimated from the ARFI model, is used both tests reject for four countries. The White and TLG tests seem in general to reach similar conclusions. There is a number of instances where the estimates of d from the linear and nonlinear model differ substantially. Looking at the countries for which rejection of the null hypothesis is obtained some interesting results arise. There seems to be little evidence for nonlinearity in the series relating to European countries. The null hypothesis is rejected only for Austria. On the other hand there is evidence for nonlinearity in the US/Yen real exchange rate and perhaps more interestingly there is evidence for nonlinearity in four out of the six Asian countries considered and Australia.

7 Conclusion

Recent work in the literature has initiated the investigation of the interplay between long-memory and nonlinearity. Work has mostly concentrated on the possibility that long memory and nonlinearity may be observationally related, e.g. in terms of covariance structures. Alternatively, a combination of nonlinearity and long memory may provide a more satisfactory model for macroeconomic series such as unemployment as discussed in van Dijk, Frances, and Paap (2002).

This paper follows the second strand of the literature and proposes tests based on neural networks for neglected nonlinearity in long memory models. We find that using a linear model to estimate the long memory parameter d prior to applying linearity tests leads to a significant loss of power and we therefore suggest estimation of d using an approximate neural network model which is capable of picking up arbitrary forms of nonlinearity. We find that this strategy entails no loss of power compared to the case of known d and we therefore recommend this approach.

An empirical application to Yen real exchange rates shows that evidence for neglected nonlinearity may be widespread in series previously analysed using linear long memory models. Almost half of the series investigated produced evidence of neglected nonlinearity.

Appendix: Proof of Theorem 1

Define the following:

$$\boldsymbol{u}(d) = (u_1(d), \ldots, u_T(d))'$$

$$\boldsymbol{v}_t(d) = (u_{t-1}(d), u_{t-1}(d), \ldots, u_{t-p}(d))'$$

$$\boldsymbol{v}(d) = (\boldsymbol{v}_1(d), \ldots, \boldsymbol{v}_T(d))'$$

Let $\boldsymbol{z}_t(d)$ be the set of cross product regressors used to test the null hypothesis of neglected nonlinearity. Then,

$$\boldsymbol{z}(d) = (\boldsymbol{z}_1(d), \ldots, \boldsymbol{z}_T(d))'$$

$$\boldsymbol{M}_v(d) = I - \boldsymbol{v}(d)(\boldsymbol{v}(d)'\boldsymbol{v}(d))^{-1}\boldsymbol{v}(d)'$$

Then, the Wald test of the null hypothesis is given by

$$W(d) = 1/\hat{\sigma}^2\boldsymbol{u}(d)'\boldsymbol{M}_v(d)\boldsymbol{z}(d)(\boldsymbol{z}(d)'\boldsymbol{M}_v(d)\boldsymbol{z}(d))^{-1}\boldsymbol{z}(d)'\boldsymbol{M}_v(d)\boldsymbol{u}(d) =$$

$$1/\hat{\sigma}^2(d)T\left[1/T(\boldsymbol{u}(d)'\boldsymbol{M}_v(d)\boldsymbol{z}(d))\right]\left[1/T(\boldsymbol{z}(d)'\boldsymbol{M}_v(d)\boldsymbol{z}(d))\right]^{-1}\left[1/T(\boldsymbol{z}(d)'\boldsymbol{M}_v(d)\boldsymbol{u}(d))\right]$$

Denote the true value of d by d^0. Then the theorem is proven if we show that

$$W(d^0) - W(\hat{d}) = o_p(1) \tag{14}$$

This follows if we show that

$$\hat{\sigma}^2 - \sigma^2 = o_p(1) \tag{15}$$

$$1/T(\boldsymbol{u}(d^0)'\boldsymbol{M}_v(d^0)\boldsymbol{z}(d^0)) - 1/T(\boldsymbol{u}(\hat{d})'\boldsymbol{M}_v(\hat{d})\boldsymbol{z}(\hat{d})) = o_p(1) \qquad (16)$$

and

$$1/T(\boldsymbol{z}(d^0)'\boldsymbol{M}_v(d^0)\boldsymbol{z}(d^0)) - 1/T(\boldsymbol{z}(\hat{d})'\boldsymbol{M}_v(\hat{d})\boldsymbol{z}(\hat{d})) = o_p(1) \qquad (17)$$

and $1/T(\boldsymbol{z}(d^0)'\boldsymbol{M}_v(d^0)\boldsymbol{z}(d^0))$ has a positive definite probability limit. The last statement is assumed to hold by assumption. Estimation of any of the models TLG_i, $i = 2, 3, 4$ can be shown straightforwardly to lead to an $\sqrt{T}$-consistent estimator of d or $d^0 - \hat{d} = O_p(T^{-1/2})$ under the null hypothesis. This implies that (15) holds. We show that (16) holds. (17) can be shown to hold similarly. We have

$$1/T(\boldsymbol{u}(d^0)'\boldsymbol{M}_v(d^0)\boldsymbol{z}(d^0)) - 1/T(\boldsymbol{u}(\hat{d})'\boldsymbol{M}_v(\hat{d})\boldsymbol{z}(\hat{d})) = 1/T\left[(\boldsymbol{u}(d^0)' - \boldsymbol{u}(\hat{d})')\boldsymbol{M}_v(d^0)\boldsymbol{z}(d^0)\right] \qquad (18)$$

$$+1/T\left[\boldsymbol{u}(\hat{d})'(\boldsymbol{M}_v(d^0) - \boldsymbol{M}_v(\hat{d}))\boldsymbol{z}(d^0)\right] + 1/T\left[\boldsymbol{u}(\hat{d})'\boldsymbol{M}_v(\hat{d})(\boldsymbol{z}(d^0) - \boldsymbol{z}(\hat{d}))\right]$$

Now examine the first term of (18)

$$1/T\left[(\boldsymbol{u}(d^0)' - \boldsymbol{u}(\hat{d})')\boldsymbol{M}_v(d^0)\boldsymbol{z}(d^0)\right] \leq ||\boldsymbol{M}_v(d^0)||||1/T\left[(\boldsymbol{u}(d^0)' - \boldsymbol{u}(\hat{d})')\boldsymbol{z}(d^0)\right]||$$

where $||A||$ denotes matrix norm ($\equiv tr(A'A)$). But

$$1/T\left[(\boldsymbol{u}(d^0)' - \boldsymbol{u}(\hat{d})')\boldsymbol{z}(d^0)\right] = 1/T\sum_{i=1}^{T} z_t'(d^0)(u_t(d^0) - u_t(\hat{d}))$$

Now

$$u_t(d) = y_t - \sum_{i=1}^{t} b_l(d)y_{t-l}$$

By the $\sqrt{T}$-consistency of $\hat{d}$ we have $b_l(d^0) - b_l(\hat{d}) = O(T^{-1/2})$ Also $b_l \sim l^{-d-1}$ for large l. Therefore

$$u_t(d^0) - u_t(\hat{d}) = O_p(T^{-1/2})$$

Thus

$$||1/T\sum_{i=1}^{T} z_t'(d^0)(u_t(d^0) - u_t(\hat{d}))|| \leq max_t||z_t'(d^0)(u_t(d^0) - u_t(\hat{d}))|| = o_p(1)$$

We work similarly through the rest of the terms of (18).

References

Blake, A. P., and G. Kapetanios (2003): "A Radial Basis Function Artificial Neural Network Test for Neglected Nonlinearity," *Forthcoming in Econometrics Journal.*

Caner, M., and B. E. Hansen (2001): "Threshold Autoregression with a Unit Root," *Econometrica*, 69, 1555–1596.

Cheung, Y. W., and K. S. Lai (2001): "Long Memory and Nonlinear Mean Reversion in Japanese Yen-Based Real Exchange Rates," *Journal of International Money and Finance*, 20, 115–132.

Davidson, J., and Sibbertsen (2002): "Generating Schemes for Long Memory Processes: Regimes, Aggregation and Linearity," *Mimeo, Cardiff University.*

De Jong, R. M., and J. Davidson (2000): "The Functional Central Limit Theorem and Convergence to Stochastic Integrals I: The Weakly Dependent Process," *Econometric Theory*, 16, 621–642.

Diebold, F. X., and A. Inoue (2001): "Long Mwmory and Regime Switching," *Journal of Econometrics*, 105, 131–159.

Elliott, G., T. J. Rothenberg, and J. H. Stock (1996): "Efficient Tests of the Unit Root Hypothesis," *Econometrica*, 64, 813–836.

Enders, W., and C. W. J. Granger (1998): "Unit Root Test and Asymmetric Adjustment with an Example Using the Term Structure of Interest Rates," *Journal of Business and Economic Statistics*, 16, 304–311.

Harvey, A. C. (1989): *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge University Press.

Kapetanios, G. (2003): "Threshold Models for Trended Time Series," Forthcoming in Empirical Economics.

Kapetanios, G., and Y. Shin (2002a): "Testing for a Unit Root against Threshold Nonlinearity," Unpublished Manuscript, University of Edinburgh.

——— (2002b): "Testing for nonstationary long memory against nonlinear ergodic models," Unpublished Manuscript, University of Edinburgh.

Kapetanios, G., A. Snell, and Y. Shin (2002): "Testing for a Unit Root in the Nonlinear STAR Framework," Journal of Econometrics, Forthcoming.

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.

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.

van Dijk, D., P. H. Frances, and R. Paap (2002): "A Nonlinear Long Memory Model with Application to US Unemployment," *Journal of Econometrics*, 110(2), 135–165.

Table 1: TLG tests, T=100

Exp	$TLG^f_{3,1}$	$TLG^t_{3,0}$	$TLG^t_{3,1}$	$TLG^t_{3,2}$	$TLG^t_{3,3}$	$TLG^t_{3,4}$
Exp 1	0.054	0.050	0.056	0.092	0.120	0.110
Exp 2	0.036	0.042	0.034	0.060	0.104	0.092
Exp 3	0.052	0.096	0.066	0.044	0.062	0.074
Exp 4	0.060	0.092	0.086	0.048	0.074	0.082
Exp 5	0.308	0.420	0.324	0.264	0.352	0.366
Exp 6	0.180	0.250	0.206	0.136	0.194	0.214
Exp 7	0.106	0.616	0.418	0.458	0.504	0.522
Exp 8	0.066	0.274	0.190	0.182	0.200	0.220
Exp 9	0.308	0.968	0.786	0.776	0.938	0.938
Exp 10	0.180	0.916	0.682	0.700	0.860	0.860

Table 2: TLG tests, T=150

Exp	$TLG^f_{3,1}$	$TLG^t_{3,0}$	$TLG^t_{3,1}$	$TLG^t_{3,2}$	$TLG^t_{3,3}$	$TLG^t_{3,4}$
Exp 1	0.036	0.040	0.042	0.048	0.070	0.074
Exp 2	0.040	0.030	0.036	0.068	0.086	0.092
Exp 3	0.110	0.154	0.138	0.168	0.176	0.172
Exp 4	0.078	0.110	0.104	0.116	0.134	0.130
Exp 5	0.462	0.598	0.490	0.536	0.618	0.630
Exp 6	0.280	0.392	0.300	0.352	0.416	0.402
Exp 7	0.114	0.812	0.546	0.696	0.734	0.736
Exp 8	0.042	0.512	0.356	0.448	0.418	0.424
Exp 9	0.342	1.000	0.848	0.898	0.996	0.996
Exp 10	0.180	0.994	0.804	0.920	0.984	0.988

Table 3: TLG tests, T=200

Exp	$TLG^f_{3,1}$	$TLG^t_{3,0}$	$TLG^t_{3,1}$	$TLG^t_{3,2}$	$TLG^t_{3,3}$	$TLG^t_{3,4}$
Exp 1	0.040	0.034	0.042	0.052	0.076	0.078
Exp 2	0.040	0.038	0.032	0.060	0.088	0.082
Exp 3	0.116	0.198	0.150	0.174	0.216	0.214
Exp 4	0.086	0.100	0.090	0.112	0.120	0.116
Exp 5	0.568	0.714	0.594	0.620	0.714	0.704
Exp 6	0.352	0.490	0.378	0.414	0.474	0.466
Exp 7	0.166	0.922	0.642	0.804	0.860	0.852
Exp 8	0.052	0.608	0.392	0.472	0.468	0.470
Exp 9	0.394	1.000	0.868	0.920	1.000	0.998
Exp 10	0.230	1.000	0.842	0.922	0.994	0.992

Table 4: White tests, T=100

Exp	W_1^f	W_0^t	W_1^t	W_2^t	W_3^t	W_4^t
Exp 1	0.050	0.050	0.056	0.090	0.122	0.108
Exp 2	0.038	0.042	0.032	0.062	0.094	0.090
Exp 3	0.048	0.100	0.064	0.046	0.062	0.076
Exp 4	0.056	0.092	0.082	0.048	0.068	0.080
Exp 5	0.308	0.414	0.326	0.262	0.346	0.354
Exp 6	0.178	0.248	0.204	0.142	0.192	0.202
Exp 7	0.104	0.610	0.420	0.452	0.510	0.514
Exp 8	0.060	0.278	0.192	0.178	0.192	0.218
Exp 9	0.306	0.958	0.780	0.764	0.928	0.930
Exp 10	0.184	0.904	0.676	0.700	0.862	0.852

Table 5: White tests, T=150

Exp	W_1^f	W_0^t	W_1^t	W_2^t	W_3^t	W_4^t
Exp 1	0.042	0.038	0.040	0.052	0.074	0.074
Exp 2	0.044	0.032	0.040	0.064	0.082	0.092
Exp 3	0.108	0.150	0.130	0.166	0.180	0.156
Exp 4	0.080	0.112	0.096	0.114	0.120	0.110
Exp 5	0.458	0.604	0.488	0.530	0.594	0.622
Exp 6	0.280	0.388	0.298	0.346	0.380	0.382
Exp 7	0.114	0.810	0.542	0.690	0.716	0.726
Exp 8	0.040	0.516	0.362	0.448	0.408	0.424
Exp 9	0.344	1.000	0.846	0.900	0.992	0.988
Exp 10	0.184	0.990	0.804	0.914	0.982	0.984

Table 6: White tests, T=200

Exp	W_1^f	W_0^t	W_1^t	W_2^t	W_3^t	W_4^t
Exp 1	0.040	0.036	0.042	0.052	0.082	0.080
Exp 2	0.042	0.040	0.034	0.060	0.086	0.092
Exp 3	0.118	0.192	0.146	0.172	0.208	0.210
Exp 4	0.086	0.100	0.084	0.110	0.114	0.108
Exp 5	0.566	0.712	0.590	0.612	0.692	0.710
Exp 6	0.342	0.496	0.370	0.410	0.468	0.470
Exp 7	0.166	0.920	0.640	0.806	0.858	0.850
Exp 8	0.052	0.608	0.396	0.472	0.460	0.474
Exp 9	0.390	0.990	0.864	0.914	0.996	0.998
Exp 10	0.226	0.998	0.836	0.920	0.992	0.992

Table 7: Mean estimate of d, T=100

Exp	f_1	t_1	t_2	t_3	t_4
Exp 1	0.643	0.435	0.428	0.427	0.437
Exp 2	0.993	0.647	0.664	0.658	0.672
Exp 3	0.892	0.608	0.623	0.612	0.632
Exp 4	0.906	0.598	0.614	0.604	0.619
Exp 5	0.740	0.534	0.550	0.567	0.573
Exp 6	0.793	0.556	0.573	0.585	0.591
Exp 7	0.540	0.528	0.563	0.567	0.576
Exp 8	0.243	0.529	0.531	0.543	0.548
Exp 9	1.217	0.546	0.614	0.600	0.606
Exp 10	1.254	0.533	0.588	0.592	0.604

Table 8: Mean estimate of d, T=150

Exp	f_1	t_1	t_2	t_3	t_4
Exp 1	0.648	0.476	0.477	0.482	0.477
Exp 2	0.930	0.638	0.662	0.651	0.667
Exp 3	0.825	0.589	0.608	0.599	0.607
Exp 4	0.830	0.603	0.618	0.619	0.628
Exp 5	0.698	0.527	0.555	0.555	0.561
EXp 6	0.742	0.559	0.581	0.574	0.583
Exp 7	0.544	0.489	0.530	0.558	0.560
Exp 8	0.171	0.530	0.530	0.536	0.542
Exp 9	1.197	0.477	0.538	0.595	0.599
Exp 10	1.242	0.513	0.550	0.592	0.599

Table 9: Mean estimate of d, T=200

Exp	f_1	t_1	t_2	t_3	t_4
Exp 1	0.650	0.529	0.530	0.516	0.512
Exp 2	0.902	0.621	0.634	0.632	0.647
Exp 3	0.783	0.591	0.602	0.595	0.604
Exp 4	0.779	0.602	0.615	0.608	0.616
Exp 5	0.687	0.529	0.541	0.565	0.566
Exp 6	0.720	0.540	0.559	0.564	0.566
Exp 7	0.631	0.490	0.514	0.570	0.567
Exp 8	0.214	0.509	0.513	0.525	0.528
Exp 9	1.199	0.458	0.534	0.590	0.596
Exp 10	1.242	0.469	0.519	0.590	0.595

Table 10: Results of empirical application. Probability values of neglected nonlinearity tests and estimated long memory parameters.

Country	W_3^t	$TLG_{3,3}^t$	d[a]	W_1^t	$TLG_{3,1}^t$	d[b]
US	0.004	0.004	0.392	0.062	0.058	0.202
Germany	0.844	0.845	0.175	0.832	0.835	0.166
France	0.742	0.739	0.166	0.738	0.741	0.193
Italy	0.145	0.143	0.022	0.075	0.045	0.156
UK	0.211	0.189	0.236	0.573	0.594	0.096
Canada	0.166	0.164	0.453	0.230	0.238	0.311
Austria	0.002	0.002	0.435	0.003	0.003	0.385
Belgium	0.599	0.604	0.099	0.651	0.649	0.135
Portugal	0.110	0.111	0.074	0.865	0.858	-0.128
Spain	0.814	0.794	0.018	0.804	0.796	-0.020
Sweden	0.114	0.094	0.105	0.235	0.248	0.260
Australia	0.004	0.006	0.123	0.008	0.009	0.280
Korea	0.013	0.015	0.069	0.012	0.012	0.129
Malaysia	0.000	0.001	0.251	0.673	0.666	0.500
Indonesia	0.134	0.135	0.223	0.134	0.134	0.233
Thailand	0.000	0.000	0.215	0.000	0.000	0.310
Philippines	0.025	0.027	0.095	0.796	0.773	0.321
Sri Lanka	0.100	0.101	0.156	0.112	0.112	0.171

[a]Neural Network
[b]Linear ARFI

Queen Mary
University of London

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

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