

Department of Economics

A Test for Serial Dependence Using Neural Networks

George Kapetanios

Working Paper No. 609

October 2007

ISSN 1473-0278



Queen Mary
University of London

A Test for Serial Dependence Using Neural Networks

G. Kapetanios*

Queen Mary, University of London.

September 6, 2007

Abstract

Testing serial dependence is central to much of time series econometrics. A number of tests that have been developed and used to explore the dependence properties of various processes. This paper builds on recent work on nonparametric tests of independence. We consider a fact that characterises serially dependent processes using a generalisation of the autocorrelation function. Using this fact we build dependence tests that make use of neural network based approximations. We derive the theoretical properties of our tests and show that they have superior power properties. Our Monte Carlo evaluation supports the theoretical findings. An application to a large dataset of stock returns illustrates the usefulness of the proposed tests.

JEL Codes: C32, C33, G12

Keywords: Independence, Neural Networks, Strict Stationarity, Bootstrap, S&P500

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

1 Introduction

Testing for and measuring serial dependence are of paramount importance to time series analysis (see e.g., Granger and Teräsvirta (1993), Robinson (1991), Tjostheim (1996)). A conventional measure of serial dependence is the autocorrelation function, which may overlook essential nonlinear features of time series that have zero autocorrelation. As Granger and Teräsvirta (1993) pointed out, there are few simple appropriate tools for analyzing nonlinear time series, although significant effort has been devoted to developing effective measures of and tests for serial dependence.

Accordingly there is a large literature on how to test independence among process and serial independence in particular, see, e.g., Hoeffding (1948), Baek and Brock (1992), Johnson and McClelland (1998), and Pinkse (1999). Tjostheim (1996) provides an excellent survey of the literature. Although there have been a number of distinct approaches to testing serial dependence there has been increasing interest recently on nonparametric entropy measures of and tests for serial dependence. This approach makes only minimal assumptions about the process under investigation avoiding, for example, normality and linearity assumptions that have been made in this context in existing literature. Prime examples of work that focuses on entropy measures are Robinson (1991), Hong and White (2005) and Fernandes and Neri (2007).

This paper provides new tests of serial dependence building on but moving away from this recent body of work. Our work is based on a fact that is essentially synonymous to the usual characterisation of serial dependence but has not been explored in the literature. In particular, letting y_t be a strictly stationary process, serial dependence is equivalent to $f_s(\cdot, \cdot) \neq g(\cdot)g(\cdot)$ for some s , where $g(\cdot)$ is the marginal density and $f_s(\cdot, \cdot)$ is the pairwise joint density for y_t and y_{t-s} , $s \in \{1, 2, \dots\}$. Of course, zero autocorrelations are not evidence of lack of serial dependence. However, we base our analysis on the following fact. For any serially dependent process, there exist some functions $h_1(\cdot)$ and $h_2(\cdot)$ such that $Cov(h_1(y_t)h_2(y_{t-s})) \neq 0$. By generalising the definition of autocorrelation in this way, we can use a simple formulation for constructing tests for serial dependence. Of course, determining $h_1(\cdot)$ and $h_2(\cdot)$ is not trivial. Neural networks with their ability to approximate arbitrarily well unknown continuous functions are useful in this respect.

We provide algorithms for constructing appropriate estimates of $h_1(\cdot)$ and $h_2(\cdot)$ based on neural network approximations. We derive the relevant asymptotic theory for our test

statistics under the null hypothesis that $f_s(\cdot, \cdot) = g(\cdot)g(\cdot)$ and also examine bootstrap approximations to the exact distribution of the test statistics. Further, we prove consistency of the test and perhaps surprisingly given the nonparametric nature of our analysis we show that the test statistic diverges at a parametric rate under any departure from the null hypothesis. This implies that our test is more powerful than other nonparametric tests which have slower rates of divergence.

The theoretical properties of our tests are reflected in their small sample properties. We compare our tests to a number of existing tests in the literature replicating exactly existing Monte Carlo studies. From this analysis we see that our tests dominate all other tests in all experiments considered that follow a number of dependence structures.

The paper is structured as follows: Section 2 provides the theoretical analysis underlying our tests and discusses their asymptotic properties. Section 3 provides the Monte Carlo analysis. Section 4 applies our tests to a large dataset of stock returns to examine the random walk hypothesis. Section 5 concludes. All proofs are relegated to an Appendix at the end of the paper.

2 Theoretical Considerations

Let y_t be a strictly stationary time series with marginal density $g(\cdot)$ and pairwise joint density $f_s(\cdot, \cdot)$ for y_t and y_{t-s} where $s \in \{1, 2, \dots\}$. An important issue in nonlinear time series analysis is determining the presence of serial dependence in y_t . A standard set of tools for this aim examines the relationship between $f_s(\cdot, \cdot)$ and $g(\cdot)$ and in particular measures the deviation between $f_s(\cdot, \cdot)$ and $g(\cdot)g(\cdot)$. Any evidence of deviation between these two quantities is taken to suggest evidence of serial dependence. Nonparametric estimates of the marginal and joint densities can be used to construct a variety of tests for serial dependence and have been the main tools used to investigate this issue.

This paper takes an alternative view in examining this issue, based on neural networks. We make the following assumptions:

Assumption 1 $f_s(\cdot, \cdot)$ and $g(\cdot)$ are continuous functions.

Assumption 2 y_t is strictly stationary.

Assumption 3 Let $f_s(\cdot, \cdot) \neq g(\cdot)g(\cdot)$. Then, y_t is a L_1 near epoque dependent (NED) process of size $-\phi$ for some $\phi > 0$, on some α -mixing process with mixing coefficients, α_ℓ , such that $\lim_{\ell \rightarrow \infty} \alpha_\ell = 0$.

Remark 1 The only noteworthy assumption is Assumption 3. We note that this assumption is extremely mild compared to assumptions made in related work such as, e.g., Assumption A.4 of Hong and White (2005).

We base our analysis on the following fact concerning dependent strictly stationary processes.

Lemma 1 Let $f_s(\cdot, \cdot) \neq g(\cdot)g(\cdot)$ on some interval with non-zero Borel measure. Let Assumptions 1-2 hold. Then, for every f_s , there exists at least one pair of continuous functions $(h_{1,f}(\cdot), h_{2,f}(\cdot))$ with $E(h_{i,f}(y_t)) = 0$, $i = 1, 2$ such that $E(h_{1,f}(y_t)h_{2,f}(y_{t-j})) \neq 0$.

Given the above Lemma, under dependence, there exist functions which when used to transform the data result in processes with non-zero correlation. Let the set of all pairs of functions referred to in Lemma 1 be given by \mathcal{H}_f . Given the bounded nature of correlation it is obvious that there exist a pair of functions in \mathcal{H}_f that maximises the absolute correlation $|Corr(h_{1,f}(y_t), h_{2,f}(y_{t-s}))|$ over \mathcal{H}_f . So the problem of testing for serial dependence is related to the problem of maximising $|Corr(h_{1,f}(y_t), h_{2,f}(y_{t-s}))|$ with respect to $(h_{1,f}(\cdot), h_{2,f}(\cdot))$.

This is of course a non-trivial optimisation. A number of alternative approaches for this problem can be envisaged. We focus on neural networks. We restate the problem as follows. Given the set of all pairs of continuous functions $(h_{1,f}(\cdot), h_{2,f}(\cdot))$ we need to determine a pair that maximises $|Corr(h_{1,f}(y_t), h_{2,f}(y_{t-s}))|$. We consider neural network approximations to continuous functions and in particular radial basis function (RBF) neural network (RBFNN) approximations for reasons that will be discussed below. The RBFNN series expansion takes the form

$$h_{i,f}(y) = \sum_{j=1}^m c_j \psi(y, t_{i,j}, \sigma_T) \quad (1)$$

where the RBF nodes, $\psi(y, t_{i,j}, \sigma_T)$, are radially symmetrical, integrable, bounded functions and $t_{i,j}$ are referred to as the centres of the RBFs. Examples include the Gaussian function of the form $\exp\left(-\left(\frac{\|y-t_{i,j}\|}{\sigma_T}\right)^2\right)$, or the multiquadratic function $\left(1 + \left(\frac{\|y-t_{i,j}\|}{\sigma_T}\right)^2\right)^{-1}$, $\sigma_T > 0$, where $\|\cdot\|$ denotes Euclidean distance. Obviously, estimation of (1) is challenging since unlike standard series expansions, there are two problems that need attention: the first is that $\psi(y, t_j, \sigma_T)$ contain unknown parameters, in particular the centres, and the second is that

the nodes are not ranked so that the choice of the nodes in the series expansion is not obvious.

We propose the following algorithm for using such approximations for our maximisation problem.

Algorithm 1 (*RBF*) *Boosting algorithm*

1. Let σ_T be some sequence such that $\sigma_T = o(1)$. We construct two initial sets of $T - s$ RBF nodes given by: $\Psi^{1,(1,\dots,T)} = \{\psi(y, y_{s+1}, \sigma_T), \psi(y, y_{s+2}, \sigma_T), \dots, \psi(y, y_T, \sigma_T)\}$ and $\Psi^{2,(1,\dots,T)} = \{\psi(y, y_1, \sigma_T), \psi(y, y_2, \sigma_T), \dots, \psi(y, y_{T-s}, \sigma_T)\}$.
2. These are ranked according to their ability to maximise $|\text{Corr}(\psi(y_t, y_{k_1}, \sigma_T), \psi(y_{t-s}, y_{k_2}, \sigma_T))|$ for $k_1 = s + 1, \dots, T$, $k_2 = 1, \dots, T - s$.
3. The pair of nodes that maximise the absolute correlation become the first pair node in the ranking of the nodes. Denote this pair of nodes by $(\psi(y_t, y_{\mathcal{S}_{1,1}}, \sigma_T), \psi(y_{t-s}, y_{\mathcal{S}_{2,1}}, \sigma_T))$. Let $\tilde{\mathcal{S}}_{1,1} = \{\mathcal{S}_{1,1}\}$ and $\tilde{\mathcal{S}}_{2,1} = \{\mathcal{S}_{2,1}\}$. Let $\Psi^{1,(1,\dots,T)/\tilde{\mathcal{S}}_{1,1}}$ and $\Psi^{2,(1,\dots,T)/\tilde{\mathcal{S}}_{2,1}}$ be the sets of nodes in $\Psi^{1,(1,\dots,T)}$ and $\Psi^{2,(1,\dots,T)}$ apart from the nodes indexed by the elements of $\tilde{\mathcal{S}}_{1,1}$ and $\tilde{\mathcal{S}}_{2,1}$.
4. Set $i = 1$ and $c_{1,1} = c_{2,1} = 1$.
5. The nodes in $\Psi^{1,(1,\dots,T)/\tilde{\mathcal{S}}_{1,i}}$ and $\Psi^{2,(1,\dots,T)/\tilde{\mathcal{S}}_{2,i}}$ are ranked according to their ability to increase

$$\left| \text{Corr} \left(\sum_{l=1}^i c_{1,l} \psi(y_t, y_{\mathcal{S}_{1,l}}, \sigma_T) + c_1^{k_1} \psi(y_t, y_{k_1}, \sigma_T), \sum_{l=1}^i c_{2,l} \psi(y_{t-s}, y_{\mathcal{S}_{2,l}}, \sigma_T) + c_2^{k_2} \psi(y_{t-s}, y_{k_2}, \sigma_T) \right) \right|$$
 where $k_1 \in \tilde{\mathcal{S}}_{1,i}$ and $k_2 \in \tilde{\mathcal{S}}_{2,i}$ and $c_1^{k_1}$ and $c_2^{k_2}$ are obtained numerically and constrained to be bounded.
6. The pair of nodes that maximise the absolute correlation becomes the $i + 1$ -th pair of nodes in the ranking of the nodes. Denote this pair of nodes by $(\psi(y_t, y_{\mathcal{S}_{1,i+1}}, \sigma_T), \psi(y_{t-s}, y_{\mathcal{S}_{2,i+1}}, \sigma_T))$ and their respective coefficients by $c_{1,i+1}$ and $c_{2,i+1}$. Let $\tilde{\mathcal{S}}_{1,i+1} = \tilde{\mathcal{S}}_{1,i} \cup \{\mathcal{S}_{1,i+1}\}$ and $\tilde{\mathcal{S}}_{2,i+1} = \tilde{\mathcal{S}}_{2,i} \cup \{\mathcal{S}_{2,i+1}\}$ where $\mathcal{S}_{1,i+1}$ and $\mathcal{S}_{2,i+1}$ denote the indices of the chosen nodes. Let $\Psi^{1,(1,\dots,T)/\tilde{\mathcal{S}}_{1,i+1}}$ and $\Psi^{2,(1,\dots,T)/\tilde{\mathcal{S}}_{2,i+1}}$ be the sets of nodes in $\Psi^{1,(1,\dots,T)}$ and $\Psi^{2,(1,\dots,T)}$ respectively apart from the nodes indexed by the elements of $\tilde{\mathcal{S}}_{1,i+1}$ and $\tilde{\mathcal{S}}_{2,i+1}$.
7. If $i = m$ for some $m = m_T$ stop and select $\sum_{l=1}^{q_1} c_{1,l} \psi(y_t, y_{\mathcal{S}_{1,l}}, \sigma_T)$ and $\sum_{l=1}^{q_2} c_{2,l} \psi(y_t, y_{\mathcal{S}_{1,l}}, \sigma_T)$, as the function approximations for $h_{1,f}(\cdot)$ and $h_{2,f}(\cdot)$, where

q_1 and q_2 are chosen so as to maximise the absolute correlation over $1, \dots, m$, else set $i = i + 1$ and go to Step 5.

A few remarks are in order.

Remark 2 *This algorithm bears close resemblance to existing algorithms in the econometrics and statistics literature that are referred to as boosting algorithms (see, e.g., Kapetanios and Blake (2007) and Buhlmann (2006)). However, its use to approximate unknown functions so as to optimise a functional as we do in this paper is to the best of our knowledge novel.*

Remark 3 *The choice for m is not discussed in Algorithm 1. Theorem 1 suggests that, as long as m is finite or $m = o(T^{1/4})$, the test resulting from Algorithm 1 is well behaved, under the null. Theorem 3 states that the test is consistent and the test statistic diverges to infinity at a parametric rate of $T^{1/2}$, under the alternative hypothesis, $f_s(\cdot, \cdot) \neq g(\cdot)g(\cdot)$, for all finite m .*

Remark 4 *The sequence σ_T is left unspecified in Algorithm 1. The work of Park and Sandberg (1991) suggests that $\sigma_T = o(1)$. Given the lack of guidance on this choice from theory, it is reasonable to consider ad hoc data-based values following the practice established by Orr (1995) for RBF neural networks. Accordingly, in practice this tuning parameter is set such that $\sigma_T = \sigma$ where $\sigma = 2 \max_t |y_t - y_{t-1}|$.*

Remark 5 *The choice of the initial set of RBF nodes given by:*

$$\Psi^{i,(1,\dots,T)} = \{\psi(y, y_1, \sigma_T), \psi(y, y_2, \sigma_T), \dots, \psi(y, y_T, \sigma_T)\}$$

may be straightforwardly generalised to $\Psi^{i,(1,\dots,p_T)}$ where p_T is chosen to reflect a subset of the observations or possibly be of a larger order than T . The theory provides no constraints as long as $p_T \rightarrow \infty$ and the candidate centres of the functions in $\Psi^{i,(1,\dots,p_T)}$ imply a partition of the support of y_t whose intervals tend uniformly to zero. The last restriction is automatically satisfied by using y_1, \dots, y_T as candidate centres.

Next we provide a result that we will use to simplify our algorithm.

Lemma 2 *Let $h_1(\cdot)$ and $h_2(\cdot)$ be continuous functions such that $E(h_1(y_t)) = E(h_2(y_t)) = 0$ and $E(h_1(y_t)h_2(y_{t-s})) \neq 0$. Then, for any positive RBF function $\psi(\cdot, \cdot, \cdot)$, there exist some constants $c_{1,i}$, $c_{2,i}$, $t_{1,i}$ and $t_{2,i}$, $i = 1, \dots, m$, such that the functions $\hat{h}_{1,m} = \sum_{l=1}^m c_{1,l}\psi(y_t, t_{1,l}, \sigma_m)$ and $\hat{h}_{2,m} = \sum_{l=1}^m c_{2,l}\psi(y_t, t_{2,l}, \sigma_m)$ approximate arbitrarily well $h_1(\cdot)$ and $h_2(\cdot)$ respectively for any sequence $\sigma_m = o(1)$ as $m \rightarrow \infty$. Further, $\lim_{m \rightarrow \infty} |E(\hat{h}_1(y_t)\hat{h}_2(y_t))| \neq 0$ implies that there exist constants $\tilde{t}_{1,i}$ and $\tilde{t}_{2,i}$, $i = 1, \dots, m$, such that $\lim_{m \rightarrow \infty} |E(\tilde{h}_1(y_t)\tilde{h}_2(y_t))| \neq 0$ where $\tilde{h}_{1,m} = \sum_{l=1}^i \psi(y_t, \tilde{t}_{1,l}, \sigma_m)$ and $\tilde{h}_{2,m} = \sum_{l=1}^i \psi(y_t, \tilde{t}_{2,l}, \sigma_m)$.*

Following the above Lemma we can modify Algorithm 1 to restrict $c_{1,i} = c_{2,i} = 1$. We refer to this Algorithm as Algorithm 2.

Theorem 1 *Let assumptions 1-2 hold. Let $\hat{\rho}$ denote the correlation obtained by applying Algorithm 1 to the observed data. Then, under the null hypothesis, and for $m = o(T^{1/4})$, $\sqrt{T}\hat{\rho}$ is asymptotically normally distributed with zero mean.*

Of course, the above result applies straightforwardly to Algorithm 2 too, making it a reasonably computationally inexpensive algorithm as it does not involve any iterative estimation. The asymptotics of Theorem 1 are not that relevant for small samples both because a test based on these asymptotics is likely to overreject in small samples and also because the variance of the statistic is not easy to obtain. As a result we resort to the bootstrap. We apply the nonparametric bootstrap for i.i.d. sequences thereby imposing the null hypothesis on the bootstrap samples. The following Theorem gives a result on the theoretical properties of the bootstrap

Theorem 2 *Let assumptions 1-2 hold. Let P_T denote the probability measure generating y_1, \dots, y_T . Let $G_T(x)$ and $\hat{G}_T(x)$ denote the exact distribution of $\sqrt{T}\hat{\rho}$ and its bootstrap approximation respectively. Then, under the null hypothesis of serial independence*

$$\lim_{T \rightarrow \infty} P_T \left[\sup_x \left| \hat{G}_T(x) - G_T(x) \right| > \varepsilon \right] = 0 \quad (2)$$

for all $\varepsilon > 0$.

We next examine the power properties of Algorithm 2. We have the following Theorem.

Theorem 3 *Let assumptions 1-3 hold. Under the alternative hypothesis of dependence given by $f_s(\cdot, \cdot) \neq g(\cdot)g(\cdot)$, and using Algorithm 2, $\sqrt{T}\hat{\rho} = O_p(T^{1/2})$, for all finite m .*

Note the simplicity, generality and optimality of this result. The test is consistent against any hypothesis that does not satisfy $f_s(\cdot, \cdot) = g(\cdot)g(\cdot)$. In this sense it is nonparametric. Notice next the parametric rate of convergence implying that this test is more powerful than the nonparametric entropy based tests proposed in the literature of testing serial dependence. As a result we choose not to consider local power settings.

Remark 6 *The test based on algorithm 2 is still relatively computationally intensive given that (i) a full search across all neural nodes takes place at each step of the Algorithm, (ii) the search is over a double grid since there are two unknown functions to be determined, and (iii) a bootstrap implementation is being considered. As a result we also consider a simplified Algorithm, referred to as Algorithm 3, in which we maximise $|\text{Corr}(h_f(y_t), h_f(y_{t-s}))|$, thereby*

searching for only one unknown function. The modifications needed to obtain this simplification from Algorithm 2 are obvious. Whereas, Algorithm 2 requires $O(T^2m)$ correlation evaluations for each bootstrap replication, Algorithm 3 only requires $O(Tm)$ evaluations.

We refer to the bootstrap tests based on Algorithms 2 and 3, for lag s as $RBF_2(s)$ and $RBF_3(s)$.

Remark 7 Both RBF tests are informative in revealing information about the lag(s) at which there exists significant serial dependence. However, for testing the null hypothesis of serial independence, it is possible that two different lag orders may give conflicting conclusions. It is thus desirable to have a portmanteau test that uses multiple lags. For this purpose, we simply follow Hong and White (2005) and suggest using

$$RBF_i = \frac{1}{\sqrt{p}} \sum_{j=1}^p RBF_i(j), \quad j = 2, 3$$

Given the results of Theorems 1 and 2 on the asymptotic normality of the $RBF_i(s)$ tests and the validity of the bootstrap, we conclude that these results extend immediately to the RBF_i tests.

3 Monte Carlo Study

In this section we consider the finite sample performance of our new test. Comparability with results of Monte Carlo studies of other serial dependence tests is very important. Therefore, we follow exactly two recent Monte Carlo studies: Hong and White (2005) and Fernandes and Neri (2007). The exact replication of their Monte Carlo setup enables us to focus on our test and take the rejection probabilities under the alternative hypotheses from existing papers thus minimising computational cost.

We now give details on the Monte Carlo experiments of Hong and White (2005) and Fernandes and Neri (2007). Starting with Hong and White (2005), the experiments considered are given by

- Exp. 1A (IID, Size):

$$y_t = \epsilon_t, \quad \epsilon_t \sim N(0, 1)$$

- Exp. 2A (AR1, Power):

$$y_t = 0.3y_{t-1} + \epsilon_t, \quad \epsilon_t \sim N(0, 1)$$

- Exp. 3A (ARCH1, Power):

$$y_t = h_t^{1/2} \epsilon_t, \quad h_t = 1 + 0.8y_{t-1}^2, \quad \epsilon_t \sim N(0, 1)$$

- Exp. 4A (TARCH, Power):

$$y_t = h_t^{1/2} \epsilon_t, \quad h_t = 0.25 + 0.6h_{t-1} + 0.5y_{t-1}^2 I_{\{\epsilon_t < 0\}} + 0.2y_{t-1}^2 I_{\{\epsilon_t \geq 0\}}, \quad \epsilon_t \sim N(0, 1)$$

- Exp. 5A (BILIN1, Power):

$$y_t = 0.8y_{t-1}\epsilon_{t-1} + \epsilon_t, \quad \epsilon_t \sim N(0, 1)$$

- Exp. 6A (NMA1, Power):

$$y_t = 0.8\epsilon_{t-1}^2 + \epsilon_t, \quad \epsilon_t \sim N(0, 1)$$

- Exp. 7A (TAR1, Power):

$$y_t = -0.5y_{t-1} I_{\{y_{t-1} < 1\}} + 0.4y_{t-1} I_{\{y_{t-1} \geq 1\}} + \epsilon_t, \quad \epsilon_t \sim N(0, 1)$$

- Exp. 8A (NAR1, Power):

$$y_t = 0.8\sqrt{|y_{t-1}|} + \epsilon_t, \quad \epsilon_t \sim N(0, 1)$$

- Exp. 9A (SIGN1, Power):

$$y_t = -I_{\{y_{t-1} < 0\}} + I_{\{y_{t-1} \geq 0\}} + 0.43\epsilon_t, \quad \epsilon_t \sim N(0, 1)$$

Throughout, we consider sample sizes of 100 observations thereby focusing on the smaller sample size considered in Hong and White (2005). Results on rejection probabilities, for the tests based on Algorithms 2 and 3, are reported in Table 1. We set $s = 1$. We also report, in Table 1, the rejection probabilities for these experiments reported for $\mathcal{T}_n(1)$ in Hong and White (2005).

Moving on to the Monte Carlo study of Fernandes and Neri (2007), the experiments of that Monte Carlo study are given by

- Exp. 1B (AR2, Power):

$$y_t = 0.1 + 0.2y_{t-1} + \epsilon_t, \quad \epsilon_t \sim N(0, 1)$$

- Exp. 2B (ARCH2, Power):

$$y_t = h_t^{1/2} \epsilon_t, \quad h_t = 0.1 + 0.2y_{t-1}^2, \quad \epsilon_t \sim N(0, 1)$$

- Exp. 3B (NMA2, Power):

$$y_t = 0.2\epsilon_{t-1}\epsilon_{t-2} + \epsilon_t, \quad \epsilon_t \sim N(0, 1)$$

- Exp. 4B (TAR2, Power):

$$y_t = (0.1 - 0.2y_{t-1})I_{\{y_{t-1} < 1\}} + (0.1 + 0.5y_{t-1})I_{\{y_{t-1} \geq 1\}} + \epsilon_t, \quad \epsilon_t \sim N(0, 1)$$

- Exp. 5B (BILIN2, Power):

$$y_t = 0.1 + 0.2y_{t-1}\epsilon_{t-1} + \epsilon_t, \quad \epsilon_t \sim N(0, 1)$$

- Exp. 6B (NAR2, Power):

$$y_t = 0.1 + 0.2\sqrt{|y_{t-1}|} + \epsilon_t, \quad \epsilon_t \sim N(0, 1)$$

- Exp. 7B (SIGN2, Power):

$$y_t = 0.1 + 0.2(-I_{\{y_{t-1} < 0\}} + I_{\{y_{t-1} \geq 0\}}) + \epsilon_t, \quad \epsilon_t \sim N(0, 1)$$

Throughout this set of experiments, we consider sample sizes of 500 observations thereby, again, focusing on the smaller sample size considered in Hong and White (2005). Results on rejection probabilities, for the tests based on Algorithms 2 and 3, are reported in Table 2. We focus solely on algorithm 3 for two reasons. Firstly, our results presented below show that Algorithm 2 is more powerful than Algorithm 3 for the Monte Carlo experiments of Hong and White (2005). Since we see that for the Monte Carlo experiments of Fernandes and Neri (2007) the test based on Algorithm 3 dominates all other tests, we feel it is redundant to incur further computational cost. This computational cost is the second reason. The test based on Algorithm 2 which searches for two rather than just one function is 25 times more expensive computationally compared to the test based on Algorithm 3, for sample sizes of 500 observations. We also report, in Table 2, the rejection probabilities for these experiments reported for the entropy based test of Hong and White (2005) in Hong and White (2005) for the values 1/2 and 4 for the tuning parameter q of their test. For all Monte Carlo experiments, the nominal significance level is 5%. We carry out 1000 Monte Carlo replications and 99 bootstrap replications throughout this section. We also set $m = \lceil T^{0.249} \rceil$, where $\lceil \cdot \rceil$

denotes integer part. Throughout the section we set $s = 1$.

Results make interesting reading. $RBF_2(1)$ and $RBF_3(1)$ are extremely well behaved under the null hypothesis. For the power experiments of the Monte Carlo study of Hong and White (2005), both $RBF_2(1)$ and $RBF_3(1)$ are more powerful than all tests considered in that paper for all experiments. $RBF_2(1)$ is substantially more powerful than $RBF_3(1)$ for a number of experiments, but as we discussed earlier this comes at a quite considerable extra computational cost. In all other cases, $RBF_2(1)$ and $RBF_3(1)$ have essentially similar power properties. Notable examples of this are the linear AR model and the ARCH model, where $RBF_2(1)$ is marginally less powerful than $RBF_3(1)$. This is of course reasonable given that the same function (identity function and square function for the two models respectively) maximises $|Corr(h_{1,f}(y_t), h_{2,f}(y_{t-s}))|$ with respect to $(h_{1,f}(\cdot), h_{2,f}(\cdot))$ for each different model. Overall, $RBF_2(1)$ is either substantially more powerful than or as powerful as $RBF_3(1)$, but at an extra computational cost.

Moving on to the Monte Carlo of Fernandes and Neri (2007) we see again that the $RBF_3(1)$ is more powerful than the entropy based test of that paper for all experiments and all values of the tuning parameter of that test, q . It is also substantially more powerful than almost all other tests and almost all other experiments examined in the Monte Carlo study of Fernandes and Neri (2007). The only exception is the BDS test of Brock, Dechert, Scheinkman, and LeBaron (1996) and experiments ARCH2 and NMA2 where the BDS test marginally outperforms $RBF_3(1)$. However, the difference between these two tests for these two cases are negligible whereas for the rest of the cases the difference in performance between these two tests is very substantial in favour of $RBF_3(1)$.

4 Empirical Application to Stock Returns

In this section, we provide an empirical application that illustrates the potential of the new test to detect the presence of serial dependence. As it is sometimes difficult to draw meaningful conclusions from the empirical analysis of a single series for the performance of a new statistical test, we consider a large dataset such as the S&P 500. It has long been hypothesized that stock prices follow a (geometric) random walk possibly with a drift. We are interested in testing this hypothesis and in identifying important lags.

Data, obtained from Datastream, are weekly returns and span the period 01/01/1993-20/01/2004 comprising 575 weekly observations. We choose to consider only companies for

which data are available throughout the period leading us to have 412 series on which to use our test. We normalise the returns series to have mean equal to zero and variance equal to one prior to applying our test. Following the Monte Carlo evidence of the previous section, we use the $RBF_3(i)$ for the empirical analysis. We carry out 149 bootstrap replications. We set $m = \lceil T^{0.249} \rceil$. We consider a variety of lags to fully capture any serial dependence in the processes. In particular we look at 1, 2, 3, 4, 12 and 52 lags, which given the weekly frequency of the data, are expected to capture monthly, quarterly and yearly seasonalities. We report probability values for all these tests in Tables 3-5.

Looking at these Tables we note that there is extensive rejection of the null hypothesis for all lags considered. The extent of the rejection is much larger than than implied by pure chance. This implies a rejection of the random walk hypothesis. In particular, we have rejection of serial independence for 58.7% of the series considered at one lag, and 34.2%, 31.7%, 25.7%, 27.6% and 21.6 % for 2, 3, 4, 12 and 52 lags respectively. Clearly, there is a decline in the extent of the evidence for serial dependence as the lag order increases, but this decline is quite slow.

5 Conclusions

Testing for serial dependence is important for time series analysis. Recent work on testing serial dependence has focused on nonparametric entropy measures. We take a different approach to that work by looking at a generalisation of the autocorrelation function. Using neural network approximations we are able to construct new tests of serial dependence. We derive the theoretical properties of our tests and show that they have superior power properties. Our Monte Carlo evaluation supports the theoretical findings. An application to a large dataset of stock returns illustrates the usefulness of the proposed tests.

A number of extensions are possible using our approach. Our tests of serial dependence should be in principle extensible, straightforwardly, to testing for dependence between different, possibly serially dependent, processes. Technical issues remain before this extension is feasible. In particular, we have relied on the independence of the process being tested under the null hypothesis. This independence simplifies greatly the analysis but cannot be relied on in the general case since dependent processes need to be analysable under the null hypothesis.

Another extension is more general. We have provided a strategy for estimating unknown functions that optimise functionals without relying on computationally intensive iterative

optimisation techniques that lack robustness. This strategy has not been shown to provide consistent estimates of the unknown functions since we did not need such a result to prove the properties of our test under the alternative hypothesis. However, such a result would be of interest in itself and would generalise results in the statistics literature on boosting and related methods (see, e.g., Kapetanios and Blake (2007), Buhlmann (2006) and Temlyakov (2000)).

Finally, we have provided a possible basis for a generalised autocorrelation type measure of dependence based on our Lemma 1 that may be of independent interest for nonlinear time series analysis.

6 Appendix

6.1 Proof of Lemma 1

Let $y = y_t$ and $x = y_{t-j}$. We have that for any pair of functions $(h_{1,f}(\cdot), h_{2,f}(\cdot))$,

$$\int \int h_{1,f}(y)h_{2,f}(x)f_s(x, y)dxdy = \tag{3}$$

$$\int \int h_{1,f}(y)h_{2,f}(x)g(x)g(y)c(G(y), G(x))dxdy = 0$$

where $c(\cdot, \cdot)$ is the copula density and $G(\cdot)$ the distribution function corresponding to $f_s(\cdot, \cdot)$ and $g(\cdot)$ respectively. We note that the copula density is given by

$$c(u, v) = \frac{\partial^2 C(u, v)}{\partial u \partial v}$$

where $C(u, v)$ is the copula function satisfying

$$f_s(\cdot, \cdot) = C(g(\cdot), g(\cdot)).$$

In the case of independence $c(u, v) = 1$. Then, (3) can be rewritten as

$$\int h_{1,f}(x) \left[\int h_{2,f}(y)g(y)c(G(y), G(x))dy \right] g(x)dx = 0$$

But, by the assumed dependence of y_t and y_{t-j} , and the fact that $c(G(y), G(x))$ is a continuous density, it follows that $c(G(y), G(x))$ is a non-constant function in both its arguments. Then, it follows that $\int h_{2,f}(y)g(y)c(G(y), G(x))dy$ cannot be zero for all possible $h_{2,f}(\cdot)$. As a result there exists some $h_{2,f}(\cdot)$ for which

$$\int h_{2,f}(y)g(y)c(G(y), G(x))dy = q(x)$$

such that $q(x)$ is nonzero for some x and also non-constant. Then, similarly, given the non-constancy of $q(x)$, there exists $h_{1,f}(\cdot)$ such that $\int h_{1,f}(x)q(x)g(x)dx$ is non-zero completing the proof of the Lemma.

6.2 Proof of Lemma 2

The first part of the Lemma follows immediately from Park and Sandberg (1991). We now prove the second part. We have

$$\left| E(\hat{h}_1(y_t)\hat{h}_1(y_t)) \right| = \left| \sum_{l=1}^m \sum_{k=1}^m c_{1,l}c_{2,k} E(\psi(y_t, t_{1,l}, \sigma_m)\psi(y_t, t_{1,l}, \sigma_m)) \right| \leq$$

$$\sum_{l=1}^m \sum_{k=1}^m |c_{1,l}c_{2,k}| |E(\psi(y_t, t_{1,l}, \sigma_m)\psi(y_t, t_{1,l}, \sigma_m))| > 0$$

But, then it immediately follows that

$$\sum_{l=1}^m \sum_{k=1}^m |E(\psi(y_t, t_{1,l}, \sigma_m) \psi(y_t, t_{1,k}, \sigma_m))| > 0$$

and by positivity of ψ that

$$\left| \sum_{l=1}^m \sum_{k=1}^m E(\psi(y_t, t_{1,l}, \sigma_m) \psi(y_t, t_{1,k}, \sigma_m)) \right| > 0$$

6.3 Proof of Theorem 1

We focus on the computationally infeasible but conceptually simpler case where the absolute correlation

$$\left| \text{Corr} \left(\sum_{l=1}^m c_{1,l} \psi(y_t, t_{1,l}, \sigma_T), \sum_{l=1}^m c_{2,l} \psi(y_{t-s}, t_{2,l}, \sigma_T) \right) \right| \quad (4)$$

is jointly maximised numerically with respect to $c_{1,l}$, $c_{2,l}$, $t_{1,l}$ and $t_{2,l}$ for $l = 1, \dots, m$. Throughout, it is assumed that $m = m_T$ but this dependence is suppressed. The above maximand is clearly equal to or larger than the absolute correlation obtained via Algorithm 1. It is clear that $c_{1,l}$, $c_{2,l}$, $t_{1,l}$ and $t_{2,l}$ are chosen so as to maximise

$$\begin{aligned} \frac{1}{T} \sum_{t=1}^T \sigma_{12}^{-1} \left[\left(\sum_{l=1}^m c_{1,l} \psi(y_t, t_{1,l}, \sigma_T) - \mu_{1,T} \right) \left(\sum_{l=1}^m c_{2,l} \psi(y_{t-s}, t_{2,l}, \sigma_T) - \mu_{2,T} \right) \right] = \quad (5) \\ \frac{1}{T} \sum_{t=1}^T D_t \end{aligned}$$

where

$$\begin{aligned} \sigma_{12} = \left[\frac{1}{T} \sum_{t=1}^T \left[\sum_{l=1}^m c_{1,l} \psi(y_t, t_{1,l}, \sigma_T) - \mu_{1,T} \right]^2 \right]^{1/2} \left[\frac{1}{T} \sum_{t=1}^T \left[\sum_{l=1}^m c_{2,l} \psi(y_t, t_{2,l}, \sigma_T) - \mu_{2,T} \right]^2 \right]^{1/2} = \quad (6) \\ A_1^{1/2} A_2^{1/2} \end{aligned}$$

By the proof of the law of large numbers given in Theorem 19.1 of Davidson (1994) it follows that $A_i \rightarrow \sigma_i^2$, $i = 1, 2$ where

$$\begin{aligned} \mu_{i,T} &= E \left(\sum_{l=1}^m c_{i,l} \psi(y_t, t_{i,l}, \sigma_T) \right) \\ \sigma_{i,T}^2 &= E \left(\left(\sum_{l=1}^m c_{i,l} \psi(y_t, t_{i,l}, \sigma_T) - \mu_{i,T} \right)^2 \right) \end{aligned}$$

$\mu_{i,T} \rightarrow \mu_i$, $\sigma_{i,T}^2 \rightarrow \sigma_i^2$ and both μ_i and σ_i^2 can be either finite or infinite. More specifically for Theorem 19.1 of Davidson (1994) it is required that

$$\text{Var} \left(\frac{1}{T} \sum_{t=1}^T \left[\sum_{l=1}^m c_{i,l} \psi(y_t, t_{i,l}, \sigma_T) - \mu_{i,T} \right]^2 \right) = o(1), \quad i = 1, 2$$

But, by independence across t ,

$$\begin{aligned} \text{Var} \left(\frac{1}{T} \sum_{t=1}^T \left[\sum_{l=1}^m c_{i,l} \psi(y_t, t_{i,l}, \sigma_T) - \mu_{i,T} \right]^2 \right) &\leq \frac{1}{T} \text{Var} \left(\left[\sum_{l=1}^m c_{i,l} \psi(y_t, t_{i,l}, \sigma_T) \right]^2 \right) \leq \\ &T^{-1} E \left(\left(\left[\sum_{l=1}^m c_{i,l} \psi(y_t, t_{i,l}, \sigma_T) \right]^2 \right)^2 \right) = \\ &\frac{1}{T} \sum_{l=1}^m \sum_{k=1}^m \sum_{r=1}^m \sum_{s=1}^m c_{i,l} c_{i,k} c_{i,r} c_{i,s} \psi(y_t, t_{i,l}, \sigma_T) \psi(y_t, t_{i,k}, \sigma_T) \psi(y_t, t_{i,r}, \sigma_T) \psi(y_t, t_{i,s}, \sigma_T) \leq \\ &\frac{1}{T} \sum_{l=1}^m \sum_{k=1}^m \sum_{r=1}^m \sum_{s=1}^m |c_{i,l} c_{i,k} c_{i,r} c_{i,s}| |\psi(y_t, t_{i,l}, \sigma_T) \psi(y_t, t_{i,k}, \sigma_T) \psi(y_t, t_{i,r}, \sigma_T) \psi(y_t, t_{i,s}, \sigma_T)| \leq \\ &T^{-1} m^4 C_\psi = o(1) \end{aligned}$$

since by boundedness of ψ , $0 < C_\psi < \infty$, and also $c_{i,j}$ are constrained to be bounded.

As a result of the independence of the process y_t , only a subset of at most $4m$ observations (equal to the number of parameter to be chosen) are relevant for the maximisation in (4). The rest of the observations are not relevant. Let us denote the set of the time indices of the observations that have been used to choose $c_{1,l}$, $c_{2,l}$, $t_{1,l}$ and $t_{2,l}$ by \mathcal{O}_m . Then,

$$\begin{aligned} &\frac{1}{T} \sum_{t=1}^T \sigma_{12}^{-1} \left[\left(\sum_{l=1}^m c_{1,l} \psi(y_t, t_{1,l}, \sigma_T) - \mu_{1,T} \right) \left(\sum_{l=1}^m c_{2,l} \psi(y_{t-s}, t_{2,l}, \sigma_T) - \mu_{2,T} \right) \right] = \\ &\frac{1}{T} \sum_{t \in \mathcal{O}_m} \sigma_{12}^{-1} \left[\left(\sum_{l=1}^m c_{1,l} \psi(y_t, t_{1,l}, \sigma_T) - \mu_{1,T} \right) \left(\sum_{l=1}^m c_{2,l} \psi(y_{t-s}, t_{2,l}, \sigma_T) - \mu_{2,T} \right) \right] + \\ &\frac{1}{T} \sum_{t \notin \mathcal{O}_m} \sigma_{12}^{-1} \left[\left(\sum_{l=1}^m c_{1,l} \psi(y_t, t_{1,l}, \sigma_T) - \mu_{1,T} \right) \left(\sum_{l=1}^m c_{2,l} \psi(y_{t-s}, t_{2,l}, \sigma_T) - \mu_{2,T} \right) \right] = \\ &B_1 + B_2 = B_1 + \frac{1}{T} \sum_{t \notin \mathcal{O}_m} B_{2,t} \end{aligned}$$

By the fact that $m = o(T^{1/4})$, $\sqrt{T}B_1 = o_p(1)$. Therefore, we focus on B_2 . But in this case, $E(B_{2,t}) = 0$. Further, the conditions of Theorem 23.18 of Davidson (1994) are easily seen to be satisfied since by an argument similar to that used above for the law of large numbers, $B_{2,t}$ are uniformly L_2 -bounded implying the result of the Theorem.

6.4 Proof of Theorem 2

In order to prove Theorem 2 we use Theorem 2.2 of Horowitz (2002) which is a restatement of a result in Mammen (1992). Given the normality result of Theorem 1 and the linearity of $\hat{\rho}$ as a function of transforms of the observations, Mammen's result immediately implies (2).

6.5 Proof of Theorem 3

Using (5) we get that

$$\frac{1}{\sqrt{T}} \sum_{t=1}^T D_t = \frac{1}{\sqrt{T}} \sum_{t=1}^T E(D_t) + \frac{1}{\sqrt{T}} \sum_{t=1}^T (D_t - E(D_t))$$

Then it follows that it is sufficient to show that

$$\lim_{m \rightarrow \infty} \text{Corr} \left(\sum_{l=1}^m \psi(y_t, y_{S_{1,l}}, \sigma_T), \sum_{l=1}^m \psi(y_{t-s}, y_{S_{2,l}}, \sigma_T) \right) \neq 0$$

By Lemma 1 and 2 and under the alternative hypothesis of temporal dependence, there exist $t_{1,l}$ and $t_{2,l}$ such that

$$\lim_{m \rightarrow \infty} \text{Corr} \left(\sum_{l=1}^m \psi(y_t, t_{1,l}, \sigma_T), \sum_{l=1}^m \psi(y_{t-s}, t_{2,l}, \sigma_T) \right) \neq 0$$

To prove this theorem we will use similar analysis to that used in Theorem 1 of Temlyakov (2000). Note that the framework considered in Temlyakov (2000) is substantially different to the one we consider here and therefore we provide a distinct analysis in what follows. Let \mathcal{F}_f be a normed space with elements given by

$$(h, g) = \left(\sum_{l=1}^{\infty} c_{1,l} \psi(y_1, t_{1,l}, \sigma), \sum_{l=1}^{\infty} c_{2,l} \psi(y_2, t_{2,l}, \sigma) \right)$$

for some constants $c_{1,l}, c_{2,l} \in \{0, 1\}$, $t_{1,l}, t_{2,l}$, $l = 1, \dots$, and norm given by

$$\|(h, g)\| = \psi_3 \left| \int \int \left(\sum_{l=1}^{\infty} \psi(y_1, t_{1,l}, \sigma) - \psi_1 \right) \left(\sum_{l=1}^{\infty} \psi(y_2, t_{2,l}, \sigma) - \psi_2 \right) f_s(y_1, y_2) dy_1 dy_2 \right|$$

where

$$\psi_i = \int \int \left(\sum_{l=1}^{\infty} \psi(y_i, t_{1,l}, \sigma) \right) f_s(y_1, y_2) dy_1 dy_2, \quad i = 1, 2$$

and

$$\begin{aligned} \psi_3 &= \psi_{3,1}^{-1/2} \psi_{3,2}^{-1/2}, \\ \psi_{3,i} &= \int \int \left(\sum_{l=1}^{\infty} \psi(y_i, t_{1,l}, \sigma) - \psi_i \right)^2 f_s(y_1, y_2) dy_1 dy_2, \quad i = 1, 2 \end{aligned}$$

It easily follows that the above norm is bounded. Define a dictionary to be a set of functions of the form

$$\mathcal{G} = \{(\psi(y_1, t_1, \sigma), \psi(y_2, t_2, \sigma)); t_1, t_2 \in \mathbb{R}\}$$

Then, we can write

$$\mathcal{F}_f = \{\oplus_{l=1}^{\infty}(\psi_{l,1}, \psi_{l,2}); (\psi_{l,1}, \psi_{l,2}) \in \mathcal{G}\}$$

where

$$\begin{aligned} \oplus_{l=1}^{\infty}(\psi_{l,1}, \psi_{l,2}) &\equiv (\psi_{1,1}, \psi_{1,2}) \oplus (\psi_{2,1}, \psi_{2,2}) \oplus \dots \equiv \\ &\left(\sum_{l=1}^{\infty} c_{1,l} \psi(y_1, t_{1,l}, \sigma), \sum_{l=1}^{\infty} c_{2,l} \psi(y_2, t_{2,l}, \sigma) \right) \end{aligned}$$

By Lemmas 1 and 2 there exist element(s) in \mathcal{H}_f that have nonzero norm. Algorithm 2 can be schematised as follows: We choose an element $(\psi_{1,1}, \psi_{1,2})$ such that $\|(\psi_{1,1}, \psi_{1,2})\|$ is maximised. Then, we proceed by iteratively choosing $(\psi_{i,1}, \psi_{i,2})$ such that

$$\|(\oplus_{i=1}^{l-1}(\psi_{i,1}, \psi_{i,2}) \oplus (\psi_{l,1}, \psi_{l,2}))\|$$

is maximised. The chosen element of \mathcal{F}_f is the one that maximises the norm over the set $\{\oplus_{l=1}^1(\psi_{l,1}, \psi_{l,2}), \dots, \oplus_{l=1}^m(\psi_{l,1}, \psi_{l,2})\}$. We first note that there can only exist element(s) in \mathcal{H}_f that have non zero norm if there exist elements in \mathcal{G} denoted by $(\psi_{1,1}, \psi_{1,2})$ such that $\|(\psi_{1,1}, \psi_{1,2})\| > 0$. The maximisations involved in Algorithm 2 imply that at least one such element will be picked as long sample moments converge to population moments, i.e., as

$$\begin{aligned} \frac{1}{T} \sum_{i=1}^T \left[\left(\sum_{l=1}^m c_{1,l} \psi(y_t, t_{1,l}, \sigma) \right) \left(\sum_{l=1}^m c_{2,l} \psi(y_{t-s}, t_{2,l}, \sigma) \right) \right] &\xrightarrow{p} \\ E \left(\left(\sum_{l=1}^m c_{1,l} \psi(y_t, t_{1,l}, \sigma) \right) \left(\sum_{l=1}^m c_{2,l} \psi(y_{t-s}, t_{2,l}, \sigma) \right) \right) & \end{aligned} \quad (7)$$

for finite m . Since $(\sum_{l=1}^m c_{1,l} \psi(y_1, t_{1,l}, \sigma)) (\sum_{l=1}^m c_{2,l} \psi(y_1, t_{2,l}, \sigma))$ is a bounded function, for finite m , (7) follows if we establish a law of large numbers for $\psi(y_t)$, where $\psi(\cdot)$ is some bounded function. To establish this law of large numbers we use Theorem 19.11 of Davidson (1994). This requires three conditions: (i) $\psi(y_t)$ is a L_1 mixingale, (ii) $\psi(y_t)$ is uniformly integrable and (iii) the mixingale coefficients of $\psi(y_t)$ tend to zero. By Assumption 3, boundedness of $\psi(y_t)$ and Theorems 17.5 and 17.13 of Davidson (1994), we obtain conditions (i) and (iii). Condition (ii) follows immediately by boundedness of $\psi(y_t)$. From the above it follows immediately that the chosen element of \mathcal{F}_f has non zero norm thus proving the theorem.

References

- BAEK, E. G., AND W. A. BROCK (1992): “A nonparametric test for independence of a multivariate time series,” *Statistica Sinica*, 2, 137–156.
- BROCK, W. A., W. D. DECHERT, J. A. SCHEINKMAN, AND B. LEBARON (1996): “A test for independence based on the correlation dimension,” *Econometric Reviews*, 15, 197–235.
- BUHLMANN, P. (2006): “Boosting for high-dimensional linear models,” *Annals of Statistics*, 34, 559–583.
- DAVIDSON, J. (1994): *Stochastic Limit Theory*. Oxford University Press.
- FERNANDES, M., AND B. NERI (2007): “Nonparametric entropy-based tests of independence between stochastic processes,” *Queen Mary University of London*.
- GRANGER, C. W. J., AND T. TERÄSVIRTA (1993): *Modelling Nonlinear Economic Relationships*. Oxford University Press.
- HOEFFDING, W. (1948): “A non-parametric test of independence,” *Annals of Mathematical Statistics*, 19, 546–557.
- HONG, Y., AND H. WHITE (2005): “Asymptotic distribution theory for nonparametric entropy measures of serial dependence,” *Econometrica*, 73, 837–901.
- HOROWITZ, J. L. (2002): “The Bootstrap,” in *Handbook in Econometrics, Vol. 5*, ed. by J. J. Heckman, and E. Leamer. Elsevier.
- JOHNSON, D., AND D. MCCLELLAND (1998): “A general dependence test and applications,” *Journal of Applied Econometrics*, 13, 627–644.
- KAPETANIOS, G., AND A. P. BLAKE (2007): “Boosting Estimation of RBF Neural Networks for Dependent Data,” *Queen Mary University of London Working Paper No. 588*.
- MAMMEN, E. (1992): “When Does Bootstrap Work? Asymptotic Results and Simulations,” *Springer, New York*.
- ORR, M. J. (1995): “Regularisation in the Selection of Radial Basis Function Centers,” *Neural Computation*, 7(3), 606–623.
- PARK, J., AND I. W. SANDBERG (1991): “Universal Approximation using Radial-Basis-Function Networks,” *Neural Computation*, 3(4), 246–257.

- PINKSE, J. (1999): “Nonparametric misspecification testing,” *University of British Columbia*.
- ROBINSON, P. M. (1991): “Consistent nonparametric entropy-based testing,” *Review of Economic Studies*, 58, 437–453.
- TEMLYAKOV, V. N. (2000): “Weak Greedy Algorithms,” *Advances in Computational Mathematics*, 12, 213–227.
- TJOSTHEIM, D. (1996): “Measures of dependence and tests for independence,” *Statistics*, 28, 249–284.

Table 1: Rejection Probabilities for Monte Carlo study of Hong and White (2005). $\mathcal{T}_n(1)$ is the entropy based test of Hong and White (2005). *best(Hong and White (2005))* refers to the best performing test in terms of rejection probabilities for a given experiment, of those tests considered and reported in Hong and White (2005).

Experiment	$RB3(1)$	$RB2(1)$	$\mathcal{T}_n(1)$	best(Hong and White (2005))
IID	0.050	0.047	-	-
AR1	0.699	0.604	0.140	0.140
ARCH1	0.887	0.875	0.376	0.612
TARCH	0.607	0.574	0.206	0.278
BILIN1	0.964	0.967	0.696	0.816
NMA1	0.411	0.607	0.340	0.348
TAR1	0.417	0.721	0.256	0.258
NAR1	0.495	0.477	0.170	0.170
SIGN1	0.625	0.616	0.608	0.608

Table 2: Rejection Probabilities for Monte Carlo study of Fernandes and Neri (2007). The second and third columns of the Table report rejection probabilities for the entropy based test of Fernandes and Neri (2007) and two values of q which is a tuning parameter for that test. *best(Fernandes and Neri (2007))* refers to the best performing test in terms of rejection probabilities for a given experiment, of those tests considered and reported in Fernandes and Neri (2007).

Experiment	$RB3(1)$	$q = 1/2$	$q = 4$	best(Fernandes and Neri (2007))
AR2	0.978	0.209	0.447	0.447
ARCH2	0.864	0.210	0.156	0.903
NMA2	0.163	0.032	0.038	0.170
TAR2	0.454	0.166	0.323	0.323
BILIN2	0.976	0.387	0.508	0.788
NAR2	0.112	0.053	0.066	0.073
SIGN2	0.882	0.241	0.392	0.392

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

**Copyright © 2007 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
Fax: +44 (0)20 8983 3580
Web: www.econ.qmul.ac.uk/papers/wp.htm**