DIAGNOSTIC CHECKING IN A FLEXIBLE NONLINEAR TIME SERIES MODEL

By Marcelo C. Medeiros and Alvaro Veiga

Department of Economics and Department of Electrical Engineering, Pontifical Catholic University of Rio de Janeiro

First Version received June 2000

Abstract. This paper considers a sequence of misspecification tests for a flexible nonlinear time series model. The model is a generalization of both the smooth transition autoregressive (STAR) and the autoregressive artificial neural network (AR-ANN) models. The tests are Lagrange multiplier (LM) type tests of parameter constancy against the alternative of smoothly changing ones, of serial independence, and of constant variance of the error term against the hypothesis that the variance changes smoothly between regimes. The small sample behaviour of the proposed tests is evaluated by a Monte-Carlo study and the results show that the tests have size close to the nominal one and a good power.

Keywords. Time series; nonlinear models; STAR models; neural networks; statistical inference; parameter constancy; serial independence; heteroscedasticity, misspecification. J.E.L: C22, C51.

1. INTRODUCTION

Over recent years, several nonlinear time series models have been proposed in the literature. Models such as the threshold autoregressive (TAR) model (Tong, 1978, 1983, 1990; Tong and Lim, 1980), the smooth transition autoregressive (STAR) model (Chan and Tong, 1986; Granger and Teräsvirta, 1993; Teräsvirtaa, 1994), and the autoregressive artificial neural network (AR-ANN) model (Kuan and White, 1994; Zhang *et al.*, 1998; Leisch *et al.*, 1999) have found a large number of successful applications.

Recently, Medeiros and Veiga (2000a) proposed a flexible nonlinear time series model, where the coefficients of a linear model are given by a single hidden layer feed-forward neural network. The model is called neuro-coefficient STAR (NCSTAR) model and has the main advantage of nesting several well-known nonlinear specifications, such as the TAR, STAR, and AR-ANN models. A modelling strategy for this family of models, following Teräsvirta *et al.* (1993), Teräsvirta and Lin (1993), Eitrheim and Teräsvirta (1996) and Rech, Teräsvirta and Tschernig (1999), was developed in Medeiros and Veiga

(2000b). However, no model evaluation procedures were yet considered in the last-mentioned paper.

This paper addresses the model evaluation issue. We present a number of diagonistic tests partially based on the work of Eitrheim and Teräsvirta (1996) and Godfrey (1988). They are Lagrange multiplier (LM) tests of parameter constancy, serial independence, and constant error variance. As the NCSTAR specification nests several well-known time series models, the tests can be directly applied to these models as well. The plan of the paper is as follows. The nonlinear model considered in this paper is presented in Section 2. The misspecification tests are discussed in Section 3. Section 4 shows a Monte-Carlo experiment. Concluding remarks are made in Section 5.

2. THE MODEL

2.1. *Mathematical formulation*

The flexible nonlinear NCSTAR model has the form

$$y_t = G(\mathbf{z}_t, \mathbf{x}_t; \boldsymbol{\Psi}) + \varepsilon_t = \boldsymbol{\alpha}'\mathbf{z}_t + \sum_{i=1}^{h} \boldsymbol{\lambda}_i'\mathbf{z}_t F(\boldsymbol{\omega}_i'\mathbf{x}_t - \beta_i) + \varepsilon_t \tag{1}$$

where $G(\mathbf{z}_t, \mathbf{x}_t; \boldsymbol{\Psi})$ is a nonlinear function of the variables $\mathbf{z}_t$ and $\mathbf{x}_t$ with the parameter vector $\boldsymbol{\Psi}$. The vector $\mathbf{z}_t$ is defined as $\mathbf{z}_t = [1, \tilde{\mathbf{z}}_t']'$, where $\tilde{\mathbf{z}}_t$ is a $p \times 1$ vector of lagged values of y_t and/or some exogenous variables. The function $F(\boldsymbol{\omega}_i'\mathbf{z}_t - \beta_i)$ is the logistic function, where $\mathbf{x}_t$ is a $q \times 1$ vector of transition variables, and $\boldsymbol{\omega}_i = [\omega_{1i}, \ldots, \omega_{qi}]'$ and β_i are real parameters. $\{\varepsilon_t\}$ is a sequence of independently normally distributed random variables with zero mean and variance σ^2. The norm of $\boldsymbol{\omega}_i$ called γ_i, is known as the *slope parameter*. In the limit, when the slope parameter approaches infinity, the logistic function becomes a step function. This model can be viewed as a linear model with time-varying coefficients. More specifically, the coefficients are given by a single hidden layer feed-forward neural network.

As pointed out in Medeiros and Veiga (2000b), model (1) is neither locally nor globally identified. There are three characteristics of the model which cause the non-identifiability. The first one is due to the symmetries in the neural network architecture. The likelihood function of the model will be unchanged if we permute the hidden units, resulting in $h!$ possibilities for each one of the coefficients of the model. The second reason is caused by the fact that $F(x) = 1 - F(-x)$, where $F(\cdot)$ is the logistic function. The third reason is the mutual dependence of the parameters $\boldsymbol{\lambda}_i, \boldsymbol{\omega}_i$ and β_i, $i = 1, \ldots, h$. If all the elements of $\boldsymbol{\lambda}_i$ equal zero, the corresponding $\boldsymbol{\omega}_i$ and β_i can assume any value without affecting the value of the likelihood function. On the other hand, if $\boldsymbol{\omega}_i = \mathbf{0}$, then $\boldsymbol{\lambda}_i$ and β_i can take any value.

To eliminate the first two sources of non-identifiability, we should restrict the parameter space imposing the following restrictions: $\beta_1 \leqslant \cdots \leqslant \beta_h$ and

$\omega_{1i} > 0, i = 1, \ldots, h$. The third one is circumvented testing for the number of hidden units in (1). The procedure is described in Medeiros and Veiga (2000b).

The NCSTAR model has the main advantage of nesting several nonlinear specifications, such as, for example:

- The SETAR model, if $\mathbf{x}_t = y_{t-d}$ and $\gamma_i \to \infty, i = 1, \ldots, h$
- The Logistic STAR (LSTAR) model, if $\mathbf{x}_t = y_{t-d}$ and $h = 1$
- The AR-ANN model, if $\mathbf{x}_t = \mathbf{z}_t$ and $\boldsymbol{\lambda}_i' = [\lambda_{0i}, 0, \ldots, 0], i = 1, \ldots, h$

2.2. *Model specification procedure*

We now briefly outline the specification procedure for the NCSTAR model developed in Medeiros and Veiga (2000b). This amounts to proceeding from a linear model to the smallest NCSTAR model and gradually towards larger ones through a sequence of LM tests. The specification phase of the modelling cycle can be summarized as follows.

1 Select the variables in $\mathbf{z}_t$.
This is done using the method proposed by Rech *et al.* (1999). They make use of a global approximation to the nonlinear model which is based on a polynomial expansion of the process. Then the variables are selected according to the value of an information criterion, such as, the AIC (Akaike, 1974) or SBIC (Schwarz, 1978).

2 Test linearity.
In the context of model (1), testing linearity has two objectives: the first is to verify if a linear model is able to adequately describe the data generating process; the second refers to the variable selection problem. The linearity test is used to determine the elements of $\mathbf{x}_t$. After selecting the elements of $\mathbf{z}_t$ with the procedure described above, we choose the elements of $\mathbf{x}_t$ by running the linearity test setting $\mathbf{x}_t$ equal to each possible subset of the elements of $\mathbf{z}_t$ and choosing the one that minimizes the p-value of the test as in Teräsvirta (1994) for the STAR case. The test is developed in the same spirit of Luukkonen *et al.* (1988), Teräsvirta *et al.* (1993), and Teräsvirta (1994), replacing the logistic function by a third-order Taylor expansion around the null hypothesis of linearity.

3 If linearity is rejected, determine the number of hidden units.
The basic idea is to start using the linearity test described above and test the linear model against the nonlinear alternative with only one hidden neuron. If the null hypothesis is rejected, then fit the model with one hidden unit test for the second one. Proceed in that way until the first acceptance of the null hypothesis. The individual tests are based on lineraizing the nonlinear contribution of the additional hidden neuron.

3. DIAGNOSTIC CHECKING

Estimation of (1) has been discussed in Medeiros and Veiga (2000b). After the model has been estimated, it has to be evaluated. We propose three misspecification tests for this purpose. The first one tests for the constancy of the parameters. The test is formulated in the same spirit as the model itself (i.e., there is a possibility of having several nonlinear functions to describe the changing parameters) and nests the special case of several structural breaks. The second one tests the assumption of no serial correlation in the errors and is an application of the results in Eitrheim and Teräsvirta (1996) and Godfrey (1988). The third one is a test of constant variance against the alternative of a smoothly changing one. The test is a special case of th test developed in Breusch and Pagan (1979); see also Breusch and Pagan (1980) and Godfrey (1988, pp. 123–36).

To derive the tests and following Eitrheim and Teräsvirta (1996), we make the general assumption that, under the null hypothesis of all the tests, the nonlinear least-squares estimate of the parameters is consistent and asymptotically normal. The necessary and sufficient conditions for this are stated in Wooldrige (1994, pp. 2653–5); see also Klimko and Nelson (1978) or Mira and Escribano (2000) for an application with smooth transition time series models.

3.1. *Test of parameter constancy*

Testing parameter constancy is an important way of checking the adequacy of linear or nonlinear models. Many parameter constancy tests are tests against unspecified alternatives or a single structural break. In this section, we present a parametric alternative to parameter constancy which allows the parameters to change smoothly as a function of time under the alternative hypothesis. In the following, we assume that the transition function has constant parameters whereas both $\boldsymbol{\alpha}$ and $\boldsymbol{\lambda}_i, i = 1, \ldots, h$, may be subject to changes over time.

Although, in this paper, we focus on diagnostic checking, the present test can be used to build up a model with time-varying parameters in the spirit of the time-varying smooth transition autoregressive (TVSTAR) model proposed by Lundberg *et al.* (2000).

To develop the test, consider a model with time-varying parameters defined as

$$y_t = \tilde{G}(\mathbf{z}_t, \mathbf{x}_t; \Psi, \tilde{\Psi}) + \varepsilon_t = \tilde{\boldsymbol{\alpha}}'(t)\mathbf{z}_t + \sum_{i=1}^{h} \left\{ \tilde{\boldsymbol{\lambda}}_i'(t)\mathbf{z}_t F(\boldsymbol{\omega}_i'\mathbf{x}_t - \beta_i) \right\} + \varepsilon_t \tag{2}$$

where

$$\tilde{\boldsymbol{\alpha}}(t) = \boldsymbol{\alpha} + \sum_{j=1}^{B} \breve{\boldsymbol{\alpha}}_j F(\zeta_j(t - \eta_j)) \tag{3}$$

and

$$\tilde{\boldsymbol{\lambda}}_i(t) = \boldsymbol{\lambda}_i + \sum_{j=1}^{B} \breve{\boldsymbol{\lambda}}_{ij} F(\zeta_j(t - \eta_j)) \tag{4}$$

$\tilde{G}(\mathbf{z}_t, \mathbf{x}_t; \Psi, \tilde{\Psi})$ is a nonlinear function of $\mathbf{z}_t$ and $\mathbf{x}_t$ with parameter vectors Ψ and $\tilde{\Psi}$ defined as

$$\Psi = [\boldsymbol{\alpha}', \boldsymbol{\lambda}_1', \ldots, \boldsymbol{\lambda}_h, \boldsymbol{\omega}_1, \ldots, \boldsymbol{\omega}_h, \beta_1, \ldots, \beta_h]'$$

and

$$\tilde{\Psi} = [\breve{\boldsymbol{\alpha}}_1', \ldots, \breve{\boldsymbol{\alpha}}_B', \breve{\boldsymbol{\lambda}}_{11}', \ldots, \breve{\boldsymbol{\lambda}}_{1B}', \ldots, \breve{\boldsymbol{\lambda}}_{h1}', \ldots, \breve{\boldsymbol{\lambda}}_{hB}', \zeta_1, \ldots, \zeta_B, \eta_1, \ldots, \eta_B]'$$

To guarantee the identifiability of the model, we must impose the additional restrictions: $\eta_1 \leqslant \eta_2 \leqslant \cdots \leqslant \eta_B$ and $\zeta_j > 0, j = 1, \ldots, B$. The parameters ζ_j are responsible for the smoothness of the changes in the autoregressive parameters. When $\zeta_j \to \infty$, (3) and (4) represent a model with B structural breaks. Combining (3) and (4) with (2), we have the model.

$$\begin{aligned} y_t = &\left\{\boldsymbol{\alpha}' + \sum_{j=1}^{B} \breve{\boldsymbol{\alpha}}_j' F(\zeta_j(t - \eta_j))\right\}\mathbf{z}_t \\ &+ \sum_{i=1}^{h}\left\{\boldsymbol{\lambda}_i' + \sum_{j=1}^{B} \breve{\boldsymbol{\lambda}}_{ij}' F(\zeta_j(t - \eta_j))\right\}\mathbf{z}_t F(\boldsymbol{\omega}_i'\mathbf{x}_t - \beta_i) + \varepsilon_t \end{aligned} \tag{5}$$

Testing *B = 0* Against *B = 1*

Consider $B = 1$, and rewrite model (5) as

$$y_t = \{\boldsymbol{\alpha}' + \breve{\boldsymbol{\alpha}}' F(\zeta(t - \eta))\}\mathbf{z}_t + \sum_{i=1}^{h}\left\{\boldsymbol{\lambda}_i' + \breve{\boldsymbol{\lambda}}_i' F(\zeta(t - \eta))\right\}\mathbf{z}_t F(\boldsymbol{\omega}_i'\mathbf{x}_t - \beta_i) + \varepsilon_t \tag{6}$$

The null hypothesis of parameter constancy is

$$\mathbf{H}_0 : \zeta = 0 \tag{7}$$

Note that model (6) is only identified under the alternative $\zeta > 0$. A consequence of this complication is that the standard asymptotic distribution theory for the likelihood ratio or other classical test statistics for testing (7) is not available. To remedy this problem, we expand $F(\zeta(t - \eta))$ into a first-order Taylor expansion around $\zeta = 0$, given by

$$T_{F,1}(\zeta(t - \eta)) = \tfrac{1}{4}\zeta(t - \eta) + R(t; \zeta, n) \tag{8}$$

where $R(t; \zeta, \eta)$ is the remainder. Replacing $F(\zeta(t - \eta))$ in (6) by (8) gives

$$y_t = (\boldsymbol{\theta}_0' + \boldsymbol{\mu}_0' t)\mathbf{z}_t + \sum_{i=1}^{h} (\boldsymbol{\theta}_i' + \boldsymbol{\mu}_i' t)\mathbf{z}_t F(\boldsymbol{\omega}_i' - \beta_i) + \varepsilon_t^* \tag{9}$$

where $\boldsymbol{\theta}_0 = \boldsymbol{\alpha} - \breve{\boldsymbol{\alpha}}\zeta\eta/4, \mu_0 = \breve{\boldsymbol{\alpha}}\zeta/4, \boldsymbol{\theta}_i = \boldsymbol{\lambda}_i - \breve{\boldsymbol{\lambda}}_i\zeta\eta/4, \boldsymbol{\mu}_i = \breve{\boldsymbol{\lambda}}\zeta/4, i = 1, \ldots, h,$ and $\varepsilon_t^* = \varepsilon_t + R(t; \zeta, \eta)$.

The null hypothesis becomes

$$\mathrm{H}_0 : \boldsymbol{\mu}_0 = \boldsymbol{\mu}_1 = \cdots = \boldsymbol{\mu}_h = 0 \tag{10}$$

Under $\mathbf{H}_0, R(t; \zeta, \eta) = 0$ and $\varepsilon_t^* = \varepsilon_t$, so that standard asymptotic theory works and $R(t; \zeta, \eta)$ can be ignored. The local approximation to the normal log likelihood function in a neighbourhood of H_0 for observation t and ignoring $R(t; \zeta, \eta)$ is

$$\begin{aligned} l_t = & -\frac{1}{2}\ln(2\pi) - \frac{1}{2}\ln\sigma^2 \\ & - \frac{1}{2\sigma^2}\left\{ y_t - (\boldsymbol{\theta}_0' + \boldsymbol{\mu}_0' t)\mathbf{z}_t - \sum_{i=1}^{h}(\boldsymbol{\theta}_i' + \boldsymbol{\mu}_i' t)\mathbf{z}_t F(\boldsymbol{\omega}_i'\mathbf{x}_t - \beta_i) \right\}^2 \end{aligned} \tag{11}$$

To derive a LM type test (assuming σ^2 constant), the consistent estimators of the partial derivatives of the log likelihood under the null are

$$\left.\frac{\partial \hat{l}_t}{\partial \boldsymbol{\theta}_0'}\right|_{\mathrm{H}_0} = \frac{1}{\hat{\sigma}^2}\hat{\varepsilon}_t\mathbf{z}_t \tag{12}$$

$$\left.\frac{\partial \hat{l}_t}{\partial \boldsymbol{\mu}_0'}\right|_{\mathrm{H}_0} = \frac{1}{\hat{\sigma}^2}\hat{\varepsilon}_t t\mathbf{z}_t \tag{13}$$

$$\left.\frac{\partial \hat{l}_t}{\partial \boldsymbol{\theta}_i'}\right|_{\mathrm{H}_0} = \frac{1}{\hat{\sigma}^2}\hat{\varepsilon}_t\mathbf{z}_t\hat{F}(\boldsymbol{\omega}_i'\mathbf{x}_t - \beta_i) \tag{14}$$

$$\left.\frac{\partial \hat{l}_t}{\partial \boldsymbol{\mu}_i'}\right|_{\mathrm{H}_0} = \frac{1}{\hat{\sigma}^2}\hat{\varepsilon}_t t\mathbf{z}_t\hat{F}(\boldsymbol{\omega}_i'\mathbf{x}_t - \beta_i) \tag{15}$$

$$\left.\frac{\partial \hat{l}_t}{\partial \boldsymbol{\omega}_i'}\right|_{\mathrm{H}_0} = \frac{1}{\hat{\sigma}^2}\hat{\varepsilon}_t\hat{\boldsymbol{\theta}}_i'\mathbf{z}_t\frac{\partial \hat{F}(\boldsymbol{\omega}_i'\mathbf{x}_t - \beta_i)}{\partial \boldsymbol{\omega}_i'} \tag{16}$$

$$\left.\frac{\partial \hat{l}_t}{\partial \beta_i}\right|_{\mathrm{H}_0} = \frac{1}{\hat{\sigma}^2}\hat{\varepsilon}_t\hat{\boldsymbol{\theta}}_i'\mathbf{z}_t\frac{\partial \hat{F}(\boldsymbol{\omega}_i'\mathbf{x}_t - \beta_i)}{\partial \beta_i} \tag{17}$$

where $i = 1, \ldots, h, \hat{\sigma}^2 = (1/T)\sum_{t=1}^{T}\hat{\varepsilon}_t^2$, and

$$\hat{\varepsilon}_t = y_t - G(\mathbf{z}_t, \mathbf{x}_t; \hat{\boldsymbol{\Psi}}) = y_t - \hat{\boldsymbol{\alpha}}'\mathbf{z}_t - \sum_{i=1}^{h}\hat{\boldsymbol{\lambda}}_i'\mathbf{z}_t\hat{F}(\boldsymbol{\omega}_i'\mathbf{x}_t - \beta_i)$$

are the residuals estimated under the null hypothesis.

The LM statistic can be written as

$$\mathrm{LM} = \frac{1}{\hat{\sigma}^2} \sum_{t=1}^{T} \hat{\varepsilon}_t \boldsymbol{v}_t' \left\{ \sum_{t=1}^{T} \hat{\boldsymbol{v}}_t \hat{\boldsymbol{v}}_t' - \sum_{t=1}^{T} \hat{\boldsymbol{v}}_t \hat{\mathbf{h}}_t' \left(\sum_{t=1}^{T} \hat{\mathbf{h}}_t \hat{\mathbf{h}}_t' \right)^{-1} \sum_{t=1}^{T} \hat{\mathbf{h}}_t \hat{\boldsymbol{v}}_t' \right\} \sum_{t=1}^{T} \hat{\boldsymbol{v}}_t \hat{\varepsilon}_t \tag{18}$$

where

$$\hat{\mathbf{h}}_t = \frac{\partial \hat{G}(\mathbf{z}_t, \mathbf{x}_t; \boldsymbol{\Psi})}{\partial \boldsymbol{\Psi}'}$$

and

$$\hat{\boldsymbol{v}}_t = \left[t\mathbf{z}_t', t\mathbf{z}_t' \hat{F}(\boldsymbol{\omega}_1' \mathbf{x}_t - \beta_1), \ldots, t\mathbf{z}_t' \hat{F}(\omega_h' \mathbf{x}_t - \beta_h) \right]'$$

The test can be carried out in stages as follows:

1 Estimate model (1) under the null hypothesis (parameter constancy) and compute the residual $\hat{\varepsilon}_t$. When the sample size is small and the model is difficult to estimate, numerical problems in applying the nonlinear least squares algorithm may lead to a solution where the residual vector is not exactly orthogonal to the gradient matrix of the nonlinear function $G(\mathbf{z}_t, \mathbf{x}_t; \hat{\boldsymbol{\Psi}})$. This has an adverse effect on the empirical size of the test. To solve this problem, we regress the residuals $\hat{\varepsilon}_t$ on $\hat{\mathbf{h}}_t$, and compute the residual sum of squares $\mathrm{SSR}_0 = \sum_{t=1}^{T} \tilde{\varepsilon}_t^2$.
2 Regress $\tilde{\varepsilon}_t$ on $\hat{\mathbf{h}}_t$ and $\hat{\boldsymbol{v}}_t$. Compute the residual sum of squares $\mathrm{SSR}_1 = \sum_{t=1}^{T} \hat{v}_t^2$.
3 Compute the χ^2 statistic

$$\mathrm{LM}_{\chi^2}^{\mathrm{pc}} = T \frac{\mathrm{SSR}_0 - \mathrm{SSR}_1}{\mathrm{SSR}_0} \tag{19}$$

or the F version of the test

$$\mathrm{LM}_F^{\mathrm{pc}} = \frac{(\mathrm{SSR}_0 - \mathrm{SSR}_1)/m}{\mathrm{SSR}_1/(T - n - m)} \tag{20}$$

where T is the number of observations, n is the number of elements of $\hat{\mathbf{h}}_t$, and $m = (h+1)(p+1)$.

Under H_0, $\mathrm{LM}_{\chi^2}^{\mathrm{pc}}$ is asymptotically distributed as a χ^2 with m degrees of freedom and $\mathrm{LM}_F^{\mathrm{pc}}$ has approximately an F distribution with m and $T - n - m$ degrees of freedom.

When applying the test, a special care should be taken. If the norm of $\hat{\boldsymbol{\omega}}_i$ is large, we may have numerical problems when carrying out the test in small samples. A solution is to omit the terms that depend on the derivatives of the logistic function from the test statistic. This can be done without significantly affecting the value of the test statistic as pointed out in Eitrheim and Teräsvirta (1996).

Testing for $B > 1$

In a practical situation, it should be interesting to estimate the parameters of model (5). To do that, we should determine the value of B. If the null hypothesis defined by (10) is rejected at a given significance level α, we should estimate a model with $B = 1$ and test for $B = 2$ at a significance level $\alpha/2$. We proceed in that way until the first acceptance of the null hypothesis, halving the significance level of the test at each step. Letting the significance level converge to zero as $B \to \infty$ keeps the dimensions of the model under control in the sense that an upper bound of the overall significance level of the sequential test is obtained through the Bonferroni upper bound.

Consider the model

$$\begin{aligned} y_t &= \{\boldsymbol{\alpha}' + \breve{\boldsymbol{\alpha}}_1' F(\zeta_1(t-\eta_1)) + \breve{\boldsymbol{\alpha}}_2' F(\zeta_2(t-\eta_2))\}\mathbf{z}_t \\ &\quad + \sum_{i=1}^{h} \Big\{ \boldsymbol{\lambda}_i' + \breve{\boldsymbol{\lambda}}_{i1}' F(\zeta_1(t-\eta_1)) + \breve{\boldsymbol{\lambda}}_{i2}' F(\zeta_2(t-\eta_2)) \Big\} \mathbf{z}_t F(\boldsymbol{\omega}_i' \mathbf{x}_t - \beta_i) + \varepsilon_t \end{aligned} \tag{21}$$

If we want to test for $B = 2$ in (21), an appropriate null hypothesis is

$$H_0 : \zeta_2 = 0 \tag{22}$$

Note that, again, (21) is only identified under the alternative. Thus, we should proceed as before and expand $F(\zeta_2(t-\eta_2))$ into a first-order Taylor expansion around $\zeta_2 = 0$. After rearranging terms, the resulting model is

$$\begin{aligned} y_t &= (\boldsymbol{\theta}_0' + \breve{\boldsymbol{\alpha}}_1' F(\zeta_1(t-\eta_1)) + \boldsymbol{\mu}_0' t)\mathbf{z}_t \\ &\quad + \sum_{i=1}^{h} \Big(\boldsymbol{\theta}_i' + \breve{\boldsymbol{\lambda}}_{i1}' F(\zeta_1(t-\eta_1)) + \boldsymbol{\mu}_i' t \Big) \mathbf{z}_t F(\boldsymbol{\omega}_i' \mathbf{x}_t - \beta_i) + \varepsilon_t^* \end{aligned} \tag{23}$$

where $\boldsymbol{\theta}_0 = \boldsymbol{\alpha} - \breve{\boldsymbol{\alpha}}_2 \zeta_2 \eta_2 / 4, \boldsymbol{\mu}_0 = \breve{\boldsymbol{\alpha}}_2 \zeta_2 / 4, \boldsymbol{\theta}_i = \boldsymbol{\lambda}_i - \breve{\boldsymbol{\lambda}}_{i2} \zeta_2 \eta / 4, \boldsymbol{\mu}_i = \breve{\boldsymbol{\lambda}}_{i2} \zeta_2 / 4, i = 1, \ldots, h$.

The null hypothesis becomes

$$\mathrm{H}_0 : \boldsymbol{\mu}_0 = \boldsymbol{\mu}_1 = \cdots = \boldsymbol{\mu}_h = 0 \tag{24}$$

The LM statistic is (18) with

$$\hat{\mathbf{h}}_t = \left[\frac{\partial \tilde{G}(\mathbf{z}_t, \mathbf{x}_t; \hat{\boldsymbol{\Psi}}, \hat{\tilde{\boldsymbol{\Psi}}})}{\partial \boldsymbol{\Psi}'} \frac{\partial \tilde{G}(\mathbf{z}_t, \mathbf{x}_t; \hat{\boldsymbol{\Psi}}, \hat{\tilde{\boldsymbol{\Psi}}})}{\partial \tilde{\boldsymbol{\Psi}}'} \right]'$$

where

$$\begin{aligned} \tilde{G}(\mathbf{z}_t, \mathbf{x}_t; \hat{\boldsymbol{\Psi}}, \hat{\tilde{\boldsymbol{\Psi}}}) &= \Big\{ \hat{\boldsymbol{\alpha}}' + \hat{\breve{\boldsymbol{\alpha}}}_1' \hat{F}(\zeta_1(t-\eta_1)) \Big\} \mathbf{z}_t \\ &\quad + \sum_{i=1}^{h} \Big\{ \hat{\boldsymbol{\lambda}}_i' + \hat{\breve{\boldsymbol{\lambda}}}_{i1}' \hat{F}(\zeta_1(t-\eta_1)) + \Big\} \mathbf{z}_t \hat{F}(\boldsymbol{\omega}_i' \mathbf{x}_t - \beta_i) \end{aligned}$$

Defining the residuals estimated under the null as $\hat{\varepsilon}_t = y_t - \tilde{G}(\mathbf{z}_t, \mathbf{x}_t; \hat{\boldsymbol{\Psi}}, \hat{\tilde{\boldsymbol{\Psi}}})$, the test can be carried out in stages as before. The only difference is the new definition of $\hat{\mathbf{h}}_t$.

3.2. *Test of serial independence*

Consider that the errors in (1) follow an rth-order autoregressive process defined as

$$\varepsilon_t = \boldsymbol{\pi}'\mathbf{v}_t + u_t \tag{25}$$

where $\boldsymbol{\pi}' = [\pi_1, \ldots, \pi_r]$ is a parameter vector, $\mathbf{v}_t' = [\varepsilon_{t-1}, \ldots, \varepsilon_{t-r}]$, and $u_t \sim$ NID $(0, \sigma^2)$. We assume that ε_t is stationary, and furthermore, that under the assumption $\varepsilon_t \sim \text{NID}(0, \sigma^2)$, i. e., $\boldsymbol{\pi} = \mathbf{0}$, $\{y_t\}$ is stationary and ergodic such that the parameters of (25) can be consistently estimated by nonlinear least squares.

The null hypothesis is formulated as $\text{H}_0 : \boldsymbol{\pi} = 0$.

The conditional normal log likelihood, given the fixed starting values has the form

$$\begin{aligned} l_t = & -\frac{1}{2}\ln(2\pi) - \frac{1}{2}\ln\sigma^2 \\ & -\frac{1}{2\sigma^2}\left\{y_t - \sum_{j=1}^{r}\pi_j y_{t-j} - G(\mathbf{z}_t, \mathbf{x}_t; \boldsymbol{\Psi}) + \sum_{j=1}^{r}\pi_j G(\mathbf{z}_{t-j}, \mathbf{x}_{t-j}; \boldsymbol{\Psi})\right\}^2 \end{aligned} \tag{26}$$

The information matrix related to (26) is block diagonal such that the element corresponding to the second derivative of (26) forms its own block. The variance σ^2 can thus be treated as a fixed constant in (26) when deriving the test statistic. The first partial derivatives of the normal log-likelihood with respect to $\boldsymbol{\pi}$ and $\boldsymbol{\Psi}$ are

$$\begin{aligned} \frac{\partial l_t}{\partial \pi_j} &= \left(\frac{u_t}{\sigma^2}\right)\{y_{t-j} - G(\mathbf{z}_{t-j}, \mathbf{x}_{t-j}; \boldsymbol{\Psi})\}, \quad j = 1, \ldots, r \\ \frac{\partial l_t}{\partial \boldsymbol{\Psi}} &= -\left(\frac{u_t}{\sigma^2}\right)\left\{\frac{\partial G(\mathbf{z}_t, \mathbf{x}_t; \boldsymbol{\Psi})}{\partial \boldsymbol{\Psi}} - \sum_{j=1}^{r}\pi_j \frac{\partial G(\mathbf{z}_{t-j}, \mathbf{x}_{t-j}; \boldsymbol{\Psi})}{\partial \boldsymbol{\Psi}}\right\} \end{aligned} \tag{27}$$

Under the null hypothesis, the consistent estimators of (27) are

$$\left.\frac{\partial \hat{l}_t}{\partial \boldsymbol{\pi}}\right|_{H_0} = \frac{1}{\hat{\sigma}^2}\hat{\varepsilon}_t\hat{\mathbf{v}}_t \quad \text{and} \quad \left.\frac{\partial \hat{l}_t}{\partial \boldsymbol{\Psi}}\right|_{\text{H}_0} = -\frac{1}{\hat{\sigma}^2}\hat{\varepsilon}_t\hat{\mathbf{h}}_t$$

where

$$\hat{\mathbf{v}}_t' = [\hat{\varepsilon}_{t-1}, \ldots, \hat{\varepsilon}_{t-r}]$$

$$\hat{\varepsilon}_{t-j} = y_{t-j} - G(\mathbf{z}_{t-j}, \mathbf{x}_{t-j}; \hat{\boldsymbol{\Psi}}) \qquad \text{for } j = 1, \ldots, r$$

$$\hat{\mathbf{h}}_t = \frac{\partial G(\mathbf{z}_t, \mathbf{x}_t; \hat{\boldsymbol{\Psi}})}{\partial \boldsymbol{\Psi}}$$

and

$$\hat{\sigma}^2 = \frac{1}{T}\sum_{t=1}^{T}\hat{\varepsilon}_t$$

The LM statistic is (18) with $\hat{\mathbf{h}}_t$ and $\hat{\mathbf{v}}_t$ defined as above.

Under the condition that the moments implied by (18) exist, the LM statistic is asymptotic distributed as a χ^2 with r degrees of freedom.

The test can be performed in three stages as shown before. The only differences are the new definition of $\hat{v}_t$ and $\hat{\mathbf{h}}_t$ at stage 2 and the degrees of freedom in the F test, r and $T - n - r$.

3.3. *Test of homoscedasticity against smoothly changing variance*

In this section, we consider a test of constant variance against the specification

$$\sigma_t^2 = \sigma^2 + \sum_{i=1}^{h} \sigma_i^2 F(\boldsymbol{\omega}'_{\sigma,i}\mathbf{x}_t - \beta_{\sigma,i}) \tag{28}$$

where $\beta_{\sigma,1} \leqslant \cdots \leqslant \beta_{\sigma,h}$, and $\omega_{\sigma,1i} > 0, i = 1, \ldots, h_\sigma$, are identifying restrictions. This formulation allows the variance to change smoothly between regimes. The idea that the error variance changes within regimes is common in the TAR literature, but, is frequently neglected in the smooth transition case. In this paper, we derive a test statistic for smoothly changing variance against a constant one.

The restrictions on the parameters to guarantee a positive variance are rather complicated and depend on the geometry of the hyperplanes defined by $\boldsymbol{\omega}_{\sigma,i}$ and $\beta_{\sigma,i} = 1, \ldots, h$. To circumvent this problem, we rewrite equation (28) as

$$\sigma_t^2 = \exp(G_\sigma(\mathbf{x}_t; \boldsymbol{\Psi}_\sigma)) = \exp\left(\varsigma + \sum_{i=1}^{h} \sigma_i F(\boldsymbol{\omega}'_{\sigma,i}\mathbf{x}_t - \beta_{\sigma,i})\right) \tag{29}$$

where $\boldsymbol{\Psi}_\sigma = [\varsigma, \varsigma_1, \ldots, \varsigma_h]'$ is a vector of real parameters.

To derive the test, consider $h = 1$. This is not a restrictive assumption because the test statistic remains unchanged if $h = 1$ or $h > 1$. Rewrite model (29) as

$$\sigma_t^2 = \exp(\varsigma + \varsigma_1 F(\gamma_\sigma(\tilde{\boldsymbol{\omega}}'_\sigma \mathbf{x}_t - c_\sigma))) \tag{30}$$

where $\|\tilde{\boldsymbol{\omega}}_\sigma\| = 1$.

The null hypothesis of constant error variance is

$$\mathrm{H}_0 : \gamma_\sigma = 0 \tag{31}$$

Note that model (30) is only identified under the alternative $\gamma_\sigma \neq 0$. To solve the problem, we expand $F(\gamma_\sigma(\tilde{\boldsymbol{\omega}}'_\sigma \mathbf{x}_t - c_\sigma))$ into a first-order Taylor expansion around $\gamma_\sigma = 0$, given by

$$T_{F,1}(\gamma_\sigma(\tilde{\boldsymbol{\omega}}'_\sigma \mathbf{x}_t - c_\sigma)) = \tfrac{1}{4}\gamma_\sigma\left(\sum_{i=1}^{q} \tilde{\omega}_{\sigma,i} x_{i,t} - c_\sigma\right) + R(\mathbf{x}_t; \gamma_\sigma, \tilde{\boldsymbol{\omega}}_\sigma, c_\sigma) \tag{32}$$

where $R(\mathbf{x}_t; \gamma_\sigma, \tilde{\boldsymbol{\omega}}_\sigma, c_\sigma)$ is the remainder. Replacing $F(\gamma_\sigma(\tilde{\boldsymbol{\omega}}'_\sigma \mathbf{x}_t - c_\sigma))$ in (30) by (32), and ignoring $R(\mathbf{x}_t; \gamma_\sigma, \tilde{\boldsymbol{\omega}}_\sigma, c_\sigma)$ gives

$$\sigma_t^2 = \exp\left(\rho + \sum_{i=1}^{q} \rho_i x_{i,t}\right) \tag{33}$$

where $\rho = \varsigma - \left(\frac{1}{4}\right)\gamma_\sigma c_\sigma \varsigma_1, \rho_i = \left(\frac{1}{4}\right)\gamma_\sigma \varsigma_1 \tilde{\omega}_{\sigma,i}, i = 1, \ldots, q$.

The null hypothesis becomes

$$\mathrm{H}_0 : \rho_1 = \rho_2 = \cdots = \rho_q = 0 \tag{34}$$

Under $\mathrm{H}_0, \exp(\rho) = \sigma^2$. The local approximation to the normal log likelihood function in a neighbourhood of H_0 for observation t is

$$l_t = -\tfrac{1}{2}\ln(2\pi) - \tfrac{1}{2}\left(\rho + \sum_{i=1}^{q} \rho_i x_{i,t}\right) - \frac{\varepsilon_t^2}{2\exp(\rho + \sum_{i=1}^{q} \rho_i x_{i,t})} \tag{35}$$

To derive a LM-type test, the partial derivatives of the log likelihood are

$$\frac{\partial l_t}{\partial \rho} = -\frac{1}{2} + \frac{\varepsilon_t^2}{2\exp(\rho + \sum_{i=1}^{q} \rho_i x_{i,t})} \tag{36}$$

$$\frac{\partial l_t}{\partial \rho_i} = -\frac{x_i}{2} + \frac{\varepsilon_t^2 x_i}{2\exp(\rho + \sum_{i=1}^{q} \rho_i x_{i,t})} \tag{37}$$

Under the null hypothesis, the consistent estimators of (36) and (37) are

$$\left.\frac{\partial \hat{l}_t}{\partial \rho}\right|_{\mathrm{H}_0} = \frac{1}{2}\left(\frac{\hat{\varepsilon}_t^2}{\hat{\sigma}^2} - 1\right)$$

$$\left.\frac{\partial \hat{l}_t}{\partial \rho_i}\right|_{\mathrm{H}_0} = \frac{x_{i,t}}{2}\left(\frac{\hat{\varepsilon}_t^2}{\hat{\sigma}^2} - 1\right)$$

where

$$\hat{\sigma}^2 = \frac{1}{T}\sum_{t=1}^{T} \hat{\varepsilon}_t^2$$

The LM statistic can be written as

$$\mathrm{LM} = \frac{1}{2}\left\{\sum_{t=1}^{T}\left(\frac{\hat{\varepsilon}_t^2}{\hat{\sigma}^2} - 1\right)\tilde{\mathbf{x}}_t\right\}' \left\{\sum_{t=1}^{T} \tilde{\mathbf{x}}_t \tilde{\mathbf{x}}_t'\right\}^{-1} \left\{\sum_{t=1}^{T}\left(\frac{\hat{\varepsilon}_t^2}{\hat{\sigma}^2} - 1\right)\tilde{\mathbf{x}}_t\right\} \tag{38}$$

where $\tilde{\mathbf{x}}_t = [1, \mathbf{x}_t]'$. For details, see the Appendix.

The test can be carried out in stages as follows:

1 Estimate model (1) assuming homoscedasticity and compute the residuals $\hat{\varepsilon}_t$. Orthogonalize the residuals by regressing them on $\partial G(\mathbf{z}_t, \mathbf{x}_t; \hat{\boldsymbol{\Psi}})/\partial \boldsymbol{\Psi}$, and compute

$$\mathrm{SSR}_0 = \sum_{t=1}^{T}\left(\frac{\tilde{\varepsilon}_t^2}{\hat{\sigma}_{\tilde{\varepsilon}}^2} - 1\right)^2$$

where $\hat{\sigma}^2_{\tilde{\varepsilon}}$ is the unconditional variance of $\tilde{\varepsilon}_t$.

2 Regress $\left(\frac{\tilde{\varepsilon}^2_t}{\hat{\sigma}^2_{\tilde{\varepsilon}}} - 1\right)$ on $\tilde{\mathbf{x}}_t$. Compute the residual sum of squares $\mathrm{SSR}_1 = \sum_{t=1}^{T} \hat{v}^2_t$.

3 Compute the χ^2 statistic

$$\mathrm{LM}^{\sigma}_{\chi^2} = T \frac{\mathrm{SSR}_0 - \mathrm{SSR}_1}{\mathrm{SSR}_0} \tag{39}$$

or the F version of the test

$$\mathrm{LM}^{\sigma}_{F} = \frac{(\mathrm{SSR}_0 - \mathrm{SSR}_1)/q}{\mathrm{SSR}_1/(T-1-q)} \tag{40}$$

where T is the number of observations.

Under H_0, $\mathrm{LM}^{\sigma}_{\chi^2}$ is approximately distributed as a χ^2 with q degrees of freedom and LM^{σ}_{F} has approximately an F distribution with q and $T-1-q$ degrees of freedom.

Estimation

If the null hypothesis is rejected, we can estimate the parameters of model (29). The estimation algorithm is an extension of the three-phase procedure proposed in Medeiros and Veiga (2000a) and the algorithm in Medeiros and Veiga (2000b). The estimation process is divided into three steps as follows.

1 Estimate the parameters of model (1) with the algorithm proposed in Medeiros and Veiga (2000b), assuming that the error variance is fixed.
2 Test the null hypothesis of homoscedasticity. If $\mathbf{H}_0$ is rejected, consider that the conditional mean is correctly specified and estimate the parameters of model (29) by minimizing

$$L_T(\Psi_\sigma) = \frac{1}{2}\sum_{t=1}^{T}\left\{\ln(2\pi) + \ln(G_\sigma(\mathbf{x}_t;\Psi_\sigma)) + \frac{\hat{\varepsilon}^2_t}{G_\sigma(\mathbf{x}_t;\Psi_\sigma)}\right\} \tag{41}$$

3 After h is determined, we estimate the full model by minimizing

$$L_T(\Psi,\Psi_\sigma) = \frac{1}{2}\sum_{t=1}^{T}\left\{\ln(2\pi) + \ln(G_\sigma(\mathbf{x}_t;\Psi_\sigma)) + \frac{\left[y_t - G(\mathbf{z}_t,\mathbf{x}_t;\Psi)\right]^2}{G_\sigma(\mathbf{x}_t;\Psi_\sigma)}\right\} \tag{42}$$

using the parameters estimated is steps 1 and 2 as initial values.

4. MONTE-CARLO EXPERIMENT

In this section, we report the results of a simulation experiment designed to study the behaviour of the proposed tests. For all the generated time series, we discarded the first 500 observations to avoid any initialization effects. So as not to

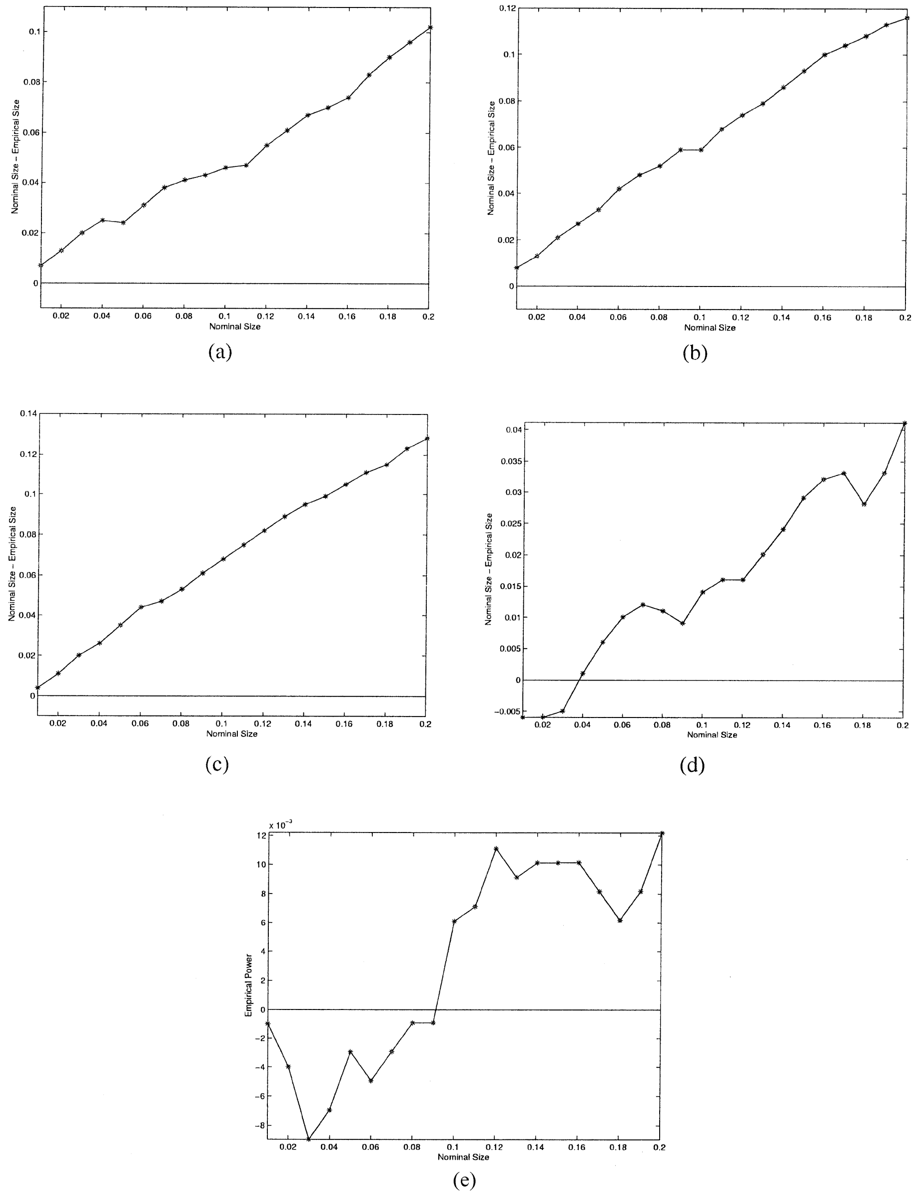


FIGURE 1. Size discrepancy plot of the parameter constancy test at sample size of 100 observations based on 1000 replications of model (43) with: (a) $\rho = 0$ and $\sigma_t^2 = 1$; (b) $\rho = 0.2$ and $\sigma_t^2 = 1$; (c) $\rho = 0.4$ and $\sigma_t^2 = 1$; (d) $\rho = 0$ and σ_t^2 given by (47); and (e) $\rho = 0, \sigma_t^2 = 1$, and estimated with $h = 1$.

estimate a nonlinear model from a time series where there is not much evidence of nonlinearity, we first test the linearity hypothesis and, if the null was not rejected at a 5% level against the NCSTAR model, we discarded the series from the

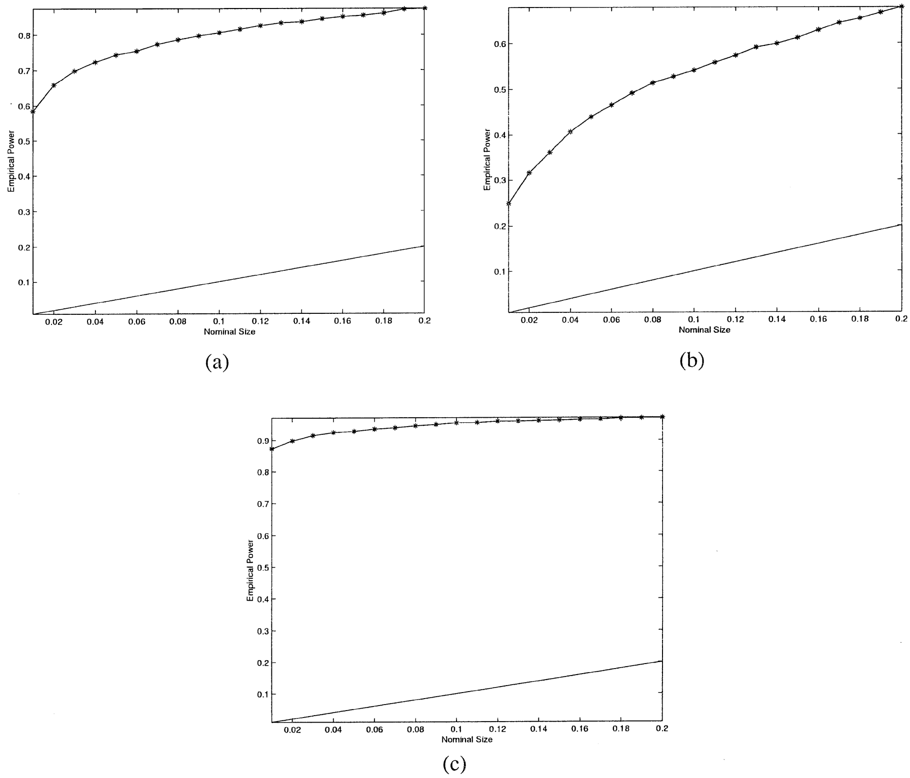


FIGURE 2. Power-size plot of the parameter constancy test at sample size of 100 observations based on 1000 replications of: (a) model (44); (b) model (45); and (c) model (46).

experiment as in Eitrheim and Teräsvirta (1996). We should also mention that the behaviour of the diagnostic tests is also investigated under alternatives other than the one for which they are derived. For example, the properties of the test of parameter constancy are also examined under processes exhibiting residual serial correlation and smoothly changing variance. Note that, strictly speaking, these are neither true size not true power experiments. We should also stress that we do not include a test of remaining nonlinearity (additional hidden unit) because it is part of the specification procedure described in Medeiros and Veiga (2000b). However, we do include a simulation study of the behaviour of the proposed tests when the models are estimated with less hidden units than necessary.

The simulated models are as follows.

- Model I

$$\begin{aligned} y_t = {} & 0.5 + 0.8y_{t-1} - 0.2y_{t-2} + (1.5 + 0.6y_{t-1} - 0.3y_{t-2})F_1(\cdot) \\ & + (-0.5 - 1.2y_{t-1} + 0.7y_{t-2})F_2(\cdot) + u_t,\ u_t = \rho u_{t-1} + \varepsilon_t \text{ and} \\ & \varepsilon_t \sim \mathrm{NID}(0, \sigma_t^2) \end{aligned} \tag{43}$$

- Model II

$$y_t = \begin{cases} 0.5 + 0.8y_{t-1} - 0.2y_{t-2} + (1.5 + 0.6y_{t-1} - 0.3y_{t-2})F_1(\cdot) & \\ -(0.5 + 1.2y_{t-1} - 0.7y_{t-2})F_2(\cdot) + \varepsilon_t & \text{if } t \leqslant 50 \\ -0.8y_{t-1} + (1.2y_{t-1} - 0.7y_{t-2})F_1(\cdot) & \\ -(0.6y_{t-1} - 0.3y_{t-2})F_2(\cdot) + \varepsilon_t & \text{otherwise} \end{cases} \tag{44}$$

- Model III

$$y_t = \begin{cases} 0.5 + 0.8y_{t-1} - 0.2y_{t-2} + (1.5 + 0.6y_{t-1} - 0.3y_{t-2})F_1(\cdot) & \\ -(0.5 + 1.2y_{t-1} - 0.7y_{t-2})F_2(\cdot) + \varepsilon_t & \text{if } t \leqslant 50 \\ -0.8y_{t-1} + (1.2y_{t-1} - 0.7y_{t-2})F_1(\cdot) & \\ -(0.6y_{t-1} - 0.3y_{t-2})F_2(\cdot) + \varepsilon_t & \text{if } 30 < t \leqslant 60 \\ 3.0 + 0.8y_{t-1} + (0.1y_{t-1} - 0.3y_{t-2})F_1(\cdot) & \\ -(0.5 + 1.2y_{t-1} - 0.7y_{t-2})F_2(\cdot) + \varepsilon_t & \text{otherwise} \end{cases} \tag{45}$$

- Model IV

$$\begin{aligned} y_t = {} & 0.5 + 0.8y_{t-1} - 0.2y_{t-2} + (-0.5 - 1.6y_{t-1} + 0.2y_{t-2})F_t(\cdot) \\ & + [-0.5 - 1.2y_{t-1} + 0.7y_{t-2} + (0.5 - 1.4y_{t-2})F_t(\cdot)]F_1(\cdot) \\ & + [3 + 0.8y_{t-1} + (0.8 - 0.8y_{t-1} - 0.1y_{t-2})F_t(\cdot)]F_1(\cdot) + \varepsilon_t \end{aligned} \tag{46}$$

In models (44)–(46), $\varepsilon_t \sim \mathrm{NID}(0, 1^2)$ and in all simulated models

$$F_1(\cdot) = F(8.49(0.7071y_{t-1} - 0.7071y_{t-2} + 1.0607))$$

and

$$F_2(\cdot) = F(8.49(0.7071y_{t-1} - 0.7071y_{t-2} - 1.0607))$$

In model (46), $F_t(\cdot) = F(0.25(t - 50))$.

To evaluate the size and power of the tests, we assume that the elements of $\mathbf{z}_t$ and $\mathbf{x}_t$ in (1) are correctly specified. In size simulations, we generated 1000 time series from model (43) with $\rho = 0$ and $\sigma_t^2 = 1$. Each replication has 100 observations. To present the results, we used size discrepancy plots and power-size curves as suggested in Davidson and MacKinnon (1998).

4.1. *Test of parameter constancy*

Results concerning size simulations are shown in Figure 1. We can see that the empirical size is close to the nominal one. However, it is interesting to notice that the test becomes rather conservative when the errors are autocorrelated.

In power simulations of the parameter constancy test, we generated data from models (44) and (45). Figure 2 shows the power-size curve. The test has good power against models with structural breaks. The power of the test increases, as expected, when the parameters change smoothly as a function of time.

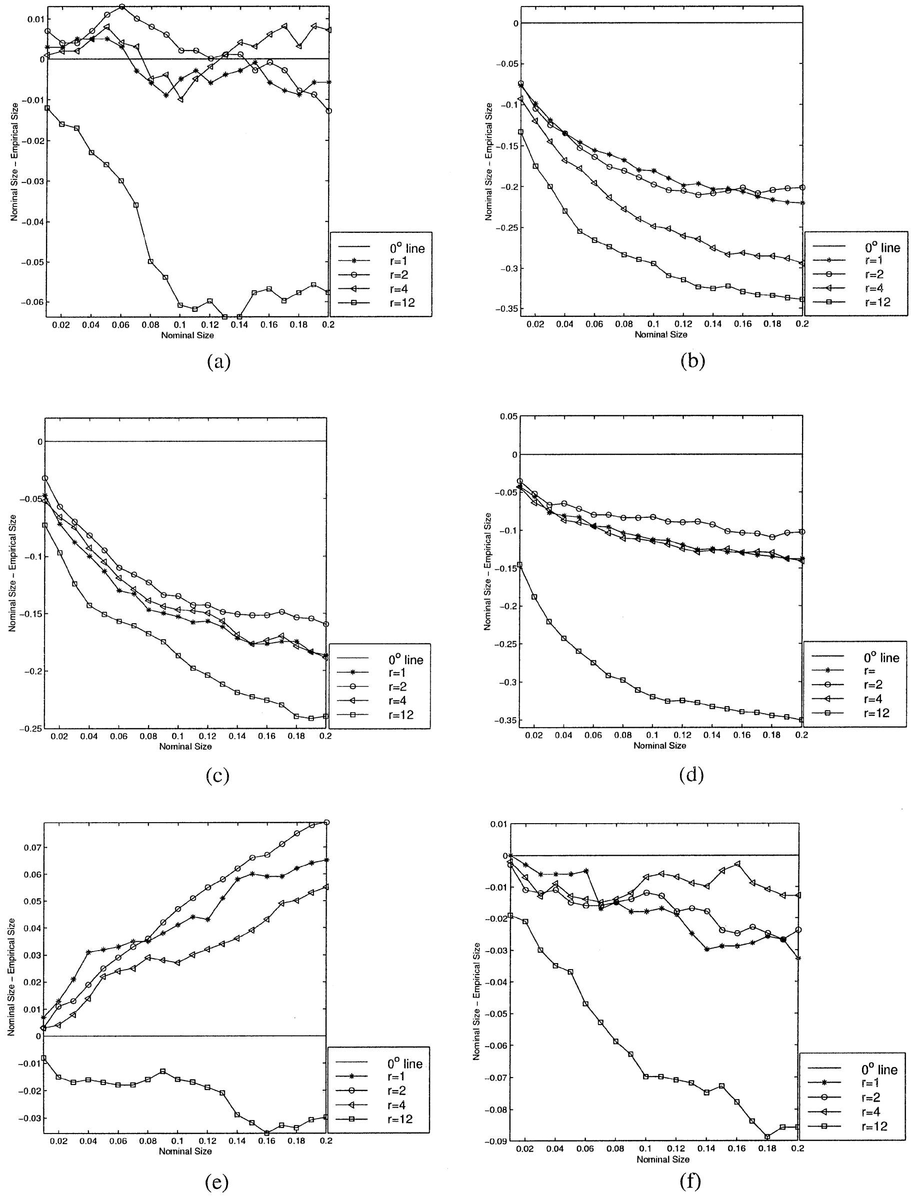
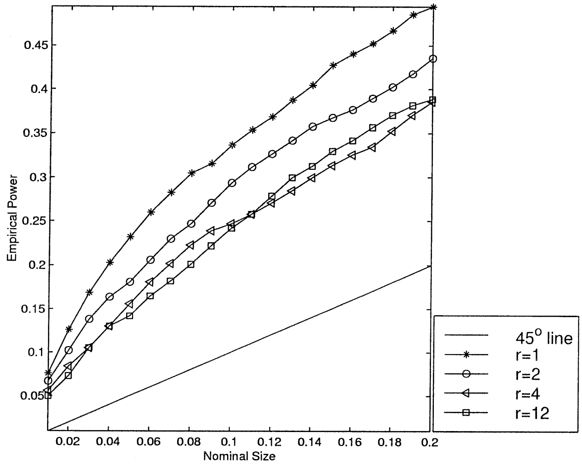
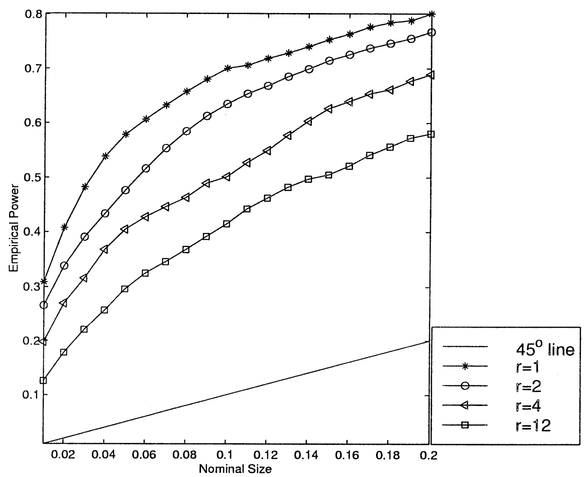


FIGURE 3. Size discrepancy plot of the serial independence test at sample size of 100 observations based on 1000 replications of: (a) model (43) with $\rho = 0$ and $\sigma_t^2 = 1$; (b) model (44); (c) model (45); (d) model (46); (e) model (43) with $\rho = 0$ and σ_t^2 given by (47); and (f) model (43) with $\rho = 0$ and $\sigma_t^2 = 1$, and estimated with $h = 1$.



(a)



(b)

FIGURE 4. Power-size curve of the test of serial independence at sample size of 100 observations based on 1000 replications of: (a) model (43) with $\rho = 0.2$ and $\sigma_t^2 = 1$; (b) model (43) with $\rho = 0.4$ and $\sigma_t^2 = 1$.

4.2. *Test of serial independence*

Figure 3 shows the results of the size simulations. The empirical size is close to the nominal one, except for the case where the model has structural breaks. Thus, the serial independence test has non-trivial power against time-varying parameters. It is interesting to mention that for $r = 12$ in (25), the test has a behaviour slightly different than the other cases. This may occur because of the small sample size (100 observations).

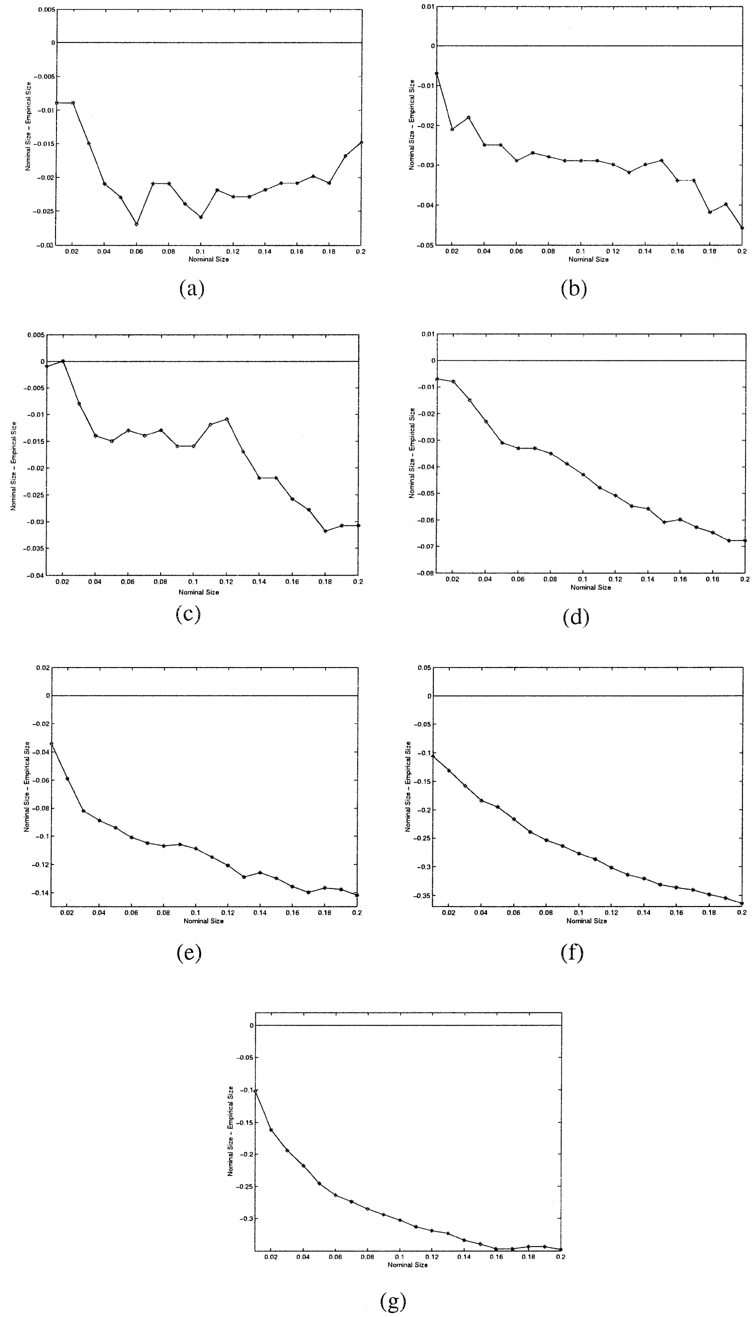


FIGURE 5. Size discrepancy plot of the heteroscedasticity test at sample size of 100 observations based on 1000 replications of: (a) model (43) with $\rho = 0$ and $\sigma_t^2 = 1$; (b) model (43) with $\rho = 0.2$ and $\sigma_t^2 = 1$; (c) model (44) with $\rho = 0.4$ and $\sigma_t^2 = 1$; (d) model (44); (e) model (45); (f) model (46); and (g) model (43) with $\rho = 0$ and $\sigma_t^2 = 1$, and estimated with $h = 1$.

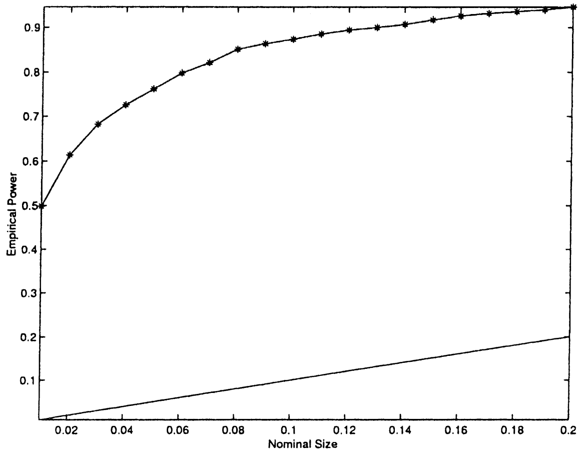


FIGURE 6. Power-size plot of the heteroscedasticity test at sample size of 100 observations based on 1000 replications of (43) with error variance given by (47).

In power simulations of the serial independence test, we generated the data from model (43) with $\rho = 0.2, 0.4$ and $\sigma_t^2 = 1$. Power-size plots are shown in Figure 4. The power of the test increases, as it should, when we increase the value of ρ.

4.3. *Test of homoscedasticity*

The results of the size simulations are shown in Figure 5. We observe that the empirical size of the test is close to the nominal one. However, the test has non-trivial power against time-varying parameters and remaining nonlinearity. In power simulations of the test, we generated the data from model (43) with $\rho = 0$ and

$$\begin{aligned}\sigma_t^2 = \exp(&-0.6931 + 0.6931F(8.49(0.7071y_{t-1} - 0.7071y_{t-2} + 1.0607)) \\ &+ 0.6931F(8.49(0.7071y_{t-1} - 0.7071y_{t-2} - 1.0607)))\end{aligned} \tag{47}$$

Results are shown in Figure 6.

5. CONCLUSIONS

In this paper, we consider a sequence of misspecification tests for a flexible nonlinear time series model, called the neuro-coefficient smooth transition autoregressive (NCSTAR) model. They are LM-type tests for testing the hypotheses of parameter constancy, serial independence, and homoscedasticity. A simulation showed that the tests are well sized and have good power in small samples. As the NCSTAR specification nests several well-known time series models, the tests can be directly applied to these models as well. These tests can be considered as a useful tool for the evaluation of estimated nonlinear models.

APPENDIX

Rewrite (35) as

$$l_t = -\frac{1}{2}\ln(2\pi) - \frac{1}{2}\boldsymbol{\varrho}'\tilde{\mathbf{x}}_t - \frac{\varepsilon_t^2}{2\exp(\boldsymbol{\varrho}'\tilde{\mathbf{x}}_t)} \tag{48}$$

where $\boldsymbol{\varrho} = [\rho, \rho_1, \ldots, \rho_q]'$. Assuming that the mean is corrected specified, the LM statistic has the general form

$$\mathrm{LM} = T\bar{\mathbf{q}}_T(\varrho)'|_{\mathrm{H}_0}\mathbf{I}(\varrho)^{-1}|_{\mathrm{H}_0}\bar{\mathbf{q}}_T(\varrho)|_{\mathrm{H}_0} \tag{49}$$

$\bar{\mathbf{q}}_T(\boldsymbol{\varrho})$ is the average score and $\mathbf{I}(\boldsymbol{\varrho})$ is the information matrix.

It is straightforward to show that

$$\bar{\mathbf{q}}_T(\boldsymbol{\varrho}) = \frac{1}{T}\sum_{t=1}^{T}\frac{1}{2}\left(\frac{\varepsilon_t^2}{\sigma_t^2} - 1\right)\tilde{\mathbf{x}}_t \tag{50}$$

The population information matrix is defined as the negative expectation of the average Hessian.

$$\mathbf{I}(\boldsymbol{\varrho}) = -E\left(\frac{1}{T}\sum_{t=1}^{T}\frac{\partial^2 l_t}{\partial\boldsymbol{\varrho}\partial\boldsymbol{\varrho}'}\right) \tag{51}$$

where

$$\frac{\partial^2 l_t}{\partial\boldsymbol{\varrho}\partial\boldsymbol{\varrho}'} = -\frac{1}{2}\frac{\varepsilon_t^2}{\exp(\boldsymbol{\varrho}'\tilde{\mathbf{x}}_t)}\tilde{\mathbf{x}}_t\tilde{\mathbf{x}}_t' \tag{52}$$

Combining (51) with (52), the population information matrix becomes

$$\mathbf{I}(\boldsymbol{\varrho}) = \frac{1}{2T}E\left(\sum_{t=1}^{T}\frac{\varepsilon_t^2}{\exp(\boldsymbol{\varrho}'\tilde{\mathbf{x}}_t)}\tilde{\mathbf{x}}_t\tilde{\mathbf{x}}_t'\right) \tag{53}$$

Under the null, the average score vector and the population information matrix can be consistently estimated as

$$\hat{\bar{\mathbf{q}}}_T(\boldsymbol{\varrho})|_{\mathrm{H}_0} = \frac{1}{2T}\sum_{t=1}^{T}\left(\frac{\hat{\varepsilon}_t^2}{\hat{\sigma}^2} - 1\right)\tilde{\mathbf{x}}_t \tag{54}$$

and

$$\hat{\mathbf{I}}(\boldsymbol{\varrho})|_{\mathrm{H}_0} = \frac{1}{2T}\sum_{t=1}^{T}\tilde{\mathbf{x}}_t\tilde{\mathbf{x}}_t' \tag{55}$$

where $\hat{\sigma}^2$ is the estimated unconditional variance of $\hat{\varepsilon}_t$ under the null hypothesis.

The LM statistic can therefore be written as

$$
\begin{aligned}
\mathrm{LM} &= T\left\{\frac{1}{2T}\sum_{t=1}^{T}\left(\frac{\hat{\varepsilon}_t^2}{\hat{\sigma}}-1\right)\tilde{\mathbf{x}}_t\right\}'\left\{\frac{1}{2T}\sum_{t=1}^{T}\tilde{\mathbf{x}}_t\tilde{\mathbf{x}}_t',\right\}^{-1}\left\{\frac{1}{2T}\sum_{t=1}^{T}\left(\frac{\hat{\varepsilon}_t^2}{\hat{\sigma}}-1\right)\tilde{\mathbf{x}}_t\right\} \\
&= \frac{1}{2}\left\{\sum_{t=1}^{T}\left(\frac{\hat{\varepsilon}_t^2}{\hat{\sigma}}-1\right)\tilde{\mathbf{x}}_t\right\}'\left\{\sum_{t=1}^{T}\tilde{\mathbf{x}}_t\tilde{\mathbf{x}}_t'\right\}^{-1}\left\{\sum_{t=1}^{T}\left(\frac{\hat{\varepsilon}_t^2}{\hat{\sigma}}-1\right)\tilde{\mathbf{x}}_t\right\}
\end{aligned} \tag{56}
$$

ACKNOWLEDGMENTS

The authors would like to thank Timo Teräsvirta and Gianluigi Rech for valuable comments and discussions, and the CNPq for the financial support. Part of this work was done while the first author was a visiting graduate student at the Department of Economic Statistics, Stockholm School of Economics, whose kind hospitality is gratefully acknowledged. This work is partly based on the first author's PhD thesis at the Department of Electrical Engineering at the Catholic University of Rio de Janeiro (PUC-Rio). We would like to thank two anomymous referees for helpful comments.

REFERENCES

Akaike, H. (1974) A new look at the statistical model identification. *IEEE Transactions on Automatic Control* 19, 716–23.

Breusch, T. S. and Pagan, A. R. (1979) A simple test for heteroscedasticity and random coefficient variation. *Econometrica* 47, 1287–94.

—— and —— (1980) The lagrange multiplier test and its applications to model specification in econometrics. *Review of Economic Studies* 47, 23–53.

Chan, K. S. and Tong, H. (1986) On estimating thresholds in autoregressive models. *Journal of Time Series Analysis* 7, 179–90.

Davidson, R. and MacKinnon, J. G. (1998) Graphical methods for investigating the size and power of hypothesis tests. *The Manchester School* 66, 1–26.

Eitrheim, Ø. and Teräsvirta, T. (1996) Testing the adequacy of smooth transition autoregressive models. *Journal of Econometrics* 74, 59–75.

Godfrey, L. G. (1988) *Misspecification Tests in Econometrics*, Vol. 16 of *Econometric Society Monographs*, 2nd edn. New York, NY: Cambridge University Press.

Granger, C. W. J. and Teräsvirta, T. (1993) *Modelling Nonlinear Economic Relationships*. Oxford: Oxford University Press.

Klimko, L. A. and Nelson, P. I. (1978) On conditional least squares estimation for stochastic process. *Annals of Statistics* 6, 629–42.

Kuan, C. M. and White, H. (1994) Artificial neural networks: An econometric perspective. *Econometric Reviews* 13, 1–91.

Leisch, F., Trapletti, A. and Hornik, K. (1999) Stationarity and stability of autoregressive neural network processes. In *Advances in Neural Information Processing Systems*. Vol. 11 (eds M. S. Kearns, S. A. Solla and D. A. Cohn). MIT Press, USA.

Lundbergh, S., Teräsvirta, T. and Van Dijk, D. (2000) Time-varying smooth transition autoregressive models. *Working Paper Series in Economics and Finance* 376, Stockholm School of Economics.

Luukkonen, R., Saikkonen, P. and Teräsvirta, T. (1988). Testing linearity against smooth transition autoregressive models. *Biometrika* 75, 491–9.

Medeiros, M. C. and Veiga, A. (2000a) A hybrid linear-neural model for time series forecasting. *IEEE Transactions on Neural Networks*: 11, 1402–1412.

—— and —— (2000b) A flexible coefficient smooth transition time series model. *Working Paper Series in Economics and Finance* 361, Stockholm School of Economics.

MIRA, S. and ESCRIBANO, A. (2000) Nonlinear time series models: Consistency and asymptotic normality of nls under new conditions. In *Nonlinear Econometric Modeling in Time Series Analysis* (eds W. A. Barnett, D. Hendry, S. Hylleberg, T. Teräsvirta, D. Tjøsthein and A. Würtz). Cambridge University Press, 119–64.

RECH, G., TERÄSVIRTA, T. and TSCHERNIG, R. (1999) A simple variable selection technique for nonlinear models. *Working Paper Series in Economics and Finance* 296, Stockholm School of Economics.

SCHWARZ, G. (1978) Estimating the dimension of a model. *Annals of Statistics* 6, 461–4.

TERÄSVIRTA, T. (1994) Specification, estimation, and evaluation of smooth transition autoregressive models. *Journal of the American Statistical Association* 89 (425), 208–18.

—— and LIN, C. -F. J. (1993) Determining the number of hidden units in a single hidden-layer neural network model. *Research report*, Bank of Norway.

——, —— and GRANGER, C. W. J. (1993) Power of the neural network linearity test. *Journal of Time Series Analysis* 14 (2), 309–23.

TONG, H. (1978) On a threshold model. In *Pattern Recognition and Signal Processing* (ed. C.H. Chen). Amsterdam: Sijthoff and Noordhoff.

—— (1983) *Threshold Models in Non-linear Time Series Analysis*, Vol. 21 of *Lecture Notes in Statistics*, Springer-Verlag, Heidelberg.

—— (1990) *Non-linear Time Series: A Dynamical Systems Approach*, Vol. 6 of *Oxford Statistical Science Series*, Oxford University Press, Oxford.

—— and LIM, K. (1980) Threshold autoregression, limit cycles and cyclical data (with discussion). *Journal of the Royal Statistical Society, Series B* 42, 245–92.

WOOLDRIDGE, J. M. (1994) Estimation and inference for dependent process, In *Handbook of Econometrics*, Vol. 4 (eds R. F. Engle and D. L. McFadden). Elsevier Science, 2639–738.

ZHANG, G., PATUWO, B. and HU, M. (1998) Forecasting with artificial neural networks: The state of the art. *International Journal of Forecasting* 14, 35–62.