Diagnostic Checking in a Flexible Nonlinear Time Series Model

Marcelo C. Medeiros

Dept. of Electrical Engineering, Catholic University of Rio de Janeiro

Dept. of Economic Statistics, Stockholm School of Economics

Alvaro Veiga

Dept. of Electrical Engineering, Catholic University of Rio de Janeiro

June 7, 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 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 constant variance of the error term against the hypothesis that the variance smoothly changes between regimes. The small sample behaviour of the proposed tests is evaluated throw 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.

JEL Classification Codes: C22, C51

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.

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, Tong and Lim 1980, Tong 1983, Tong 1990), the Smooth Transition AutoRegressive (STAR) model (Chan and Tong 1986, Granger and Teräsvirta 1993, Teräsvirta 1994), and the AutoRegressive Artificial Neural Network (AR-ANN) model (Kuan and White 1994, Zhang, Patuwo and Hu 1998, Leisch, Trapletti and Hornik 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. A modelling strategy for this family of models, following Teräsvirta, Lin and Granger (1993), Teräsvirta and Lin (1993), Eitrheim and Teräsvirta (1996), and Rech, Teräsvirta and Tschernig (1999), was developed in Medeiros and Veiga (2000a). 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 diagnostic check 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. 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

The flexible nonlinear NCSTAR model has the following form.

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

where $\beta_1 \leq \ldots \leq \beta_h$ and $\omega_{1i} > 0$, $i = 1, \ldots, h$. These restrictions are to ensure the identifiability of the model. $\{\varepsilon_t\}$ is a sequence of normally independently distributed random variables with zero mean and variance σ^2. 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 $F(\boldsymbol{\omega}_i'\mathbf{x}_t - \beta_i)$ 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. 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.

3 Diagnostic Checking

Estimation of (1) has been discussed in Medeiros and Veiga (2000a). After the model has been estimated it has to be evaluated. We propose three diagnostic checking 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 the test developed in Breusch and Pagan (1979) (see also Breusch and Pagan (1980) and Godfrey (1988, pages 123–136)).

In order to derive the tests we make the general assumption that nonlinear least-squares estimated of the parameter vector $\boldsymbol{\Psi}$ in (1) is consistent and asymptotically normal.

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 $\boldsymbol{\alpha}_i$ 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 Lundbergh, Teräsvirta and van Dijk (2000).

To develop the test rewrite model (1) as

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

where

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

and

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

The parameter vectors $\boldsymbol{\Psi}$ and $\tilde{\boldsymbol{\Psi}}$ are defined as $\boldsymbol{\Psi} = [\boldsymbol{\alpha}', \boldsymbol{\lambda}_1', \ldots, \boldsymbol{\lambda}_h, \boldsymbol{\omega}_1, \ldots, \boldsymbol{\omega}_h, \beta_1, \ldots, \beta_h]'$, and $\tilde{\boldsymbol{\Psi}} = [\breve{\boldsymbol{\alpha}}_1', \ldots, \breve{\boldsymbol{\alpha}}_B', \breve{\boldsymbol{\lambda}}_{11}', \ldots, \breve{\boldsymbol{\lambda}}_{1B}', \ldots, \breve{\boldsymbol{\lambda}}_{h1}', \ldots, \breve{\boldsymbol{\lambda}}_{hB}', \zeta_1, \ldots, \zeta_B, \eta_1, \ldots, \eta_B]'$. In order to guarantee the identifiability of the model, we must impose the additional restrictions: $\eta_1 \leq \eta_2 \leq \ldots \eta_B$ and $\zeta_j > 0$, $j = 1, \ldots, B$. The parameters ζ_j are responsible for the smoothness of the changing in the autoregressive parameters. When $\zeta_j \to \infty$, equations (3) and (4) represent a model with B structural breaks. Combining (3) and (4) with (2), we have the following model

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

3.1.1 Testing $B = 0$ Against $B = 1$

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

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

The null hypothesis hypothesis of parameter constancy is

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

Note that model (6) is only identified under the alternative $\zeta \neq 0$. 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 $F\left(\zeta(t-\eta)\right)$ into a first-order Taylor expansion around $\zeta = 0$, given by

$$t_1 = \frac{1}{4}\zeta(t-\eta) + R(t;\zeta,\eta), \tag{8}$$

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

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

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

The null hypothesis becomes

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

Under H_0, $\varepsilon_t^* = \varepsilon_t$. The local approximation to the normal log likelihood function in a neighborhood of

H_0 for observation t and ignoring $R(t;\zeta,\eta)$ is

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

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

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

where $\hat{\mathbf{h}}_t = \nabla G(\mathbf{z}_t, \mathbf{x}_t; \hat{\boldsymbol{\Psi}})$ and $\hat{\boldsymbol{\nu}}_t = \left[t\mathbf{z}_t', t\mathbf{z}_t' \hat{F}(\boldsymbol{\omega}_1' \mathbf{x}_t - \beta_1), \ldots, t\mathbf{z}_t' \hat{F}(\boldsymbol{\omega}_h' \mathbf{x}_t - \beta_h) \right]'$.

The test can be carried out in stages as follows:

1. Estimate model (1) assuming 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 $\nabla G(\mathbf{z}_t, \mathbf{x}_t; \hat{\boldsymbol{\Psi}})$, and compute the residual sum of squares $SSR_0 = \left(\frac{1}{T}\right) \sum_{t=1}^{T} \tilde{\varepsilon}_t^2$.

2. Regress $\tilde{\varepsilon}_t$ on $\hat{\mathbf{h}}_t$ and $\hat{\boldsymbol{\nu}}_t$. Compute the residual sum of squares $SSR_1 = \left(\frac{1}{T}\right) \sum_{t=1}^{T} \hat{v}_t^2$.

3. Compute the χ^2 statistic

$$LM_{\chi^2}^{pc} = T \frac{SSR_0 - SSR_1}{SSR_0}, \tag{19}$$

or the F version of the test

$$LM_F^{pc} = \frac{(SSR_0 - SSR_1)/m}{SSR_1/(T - n - m)}, \tag{20}$$

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, $LM_{\chi^2}^{pc}$ is asymptotically distributed as a χ^2 with m degrees of freedom and LM_F^{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 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

3.1.2 Testing for $B > 1$

In a practical situation it should be interesting to estimate the parameters of model (5). In order to do that, we should determine the value of B. If the null hypothesis defined by (10) is rejected, we should estimate a model with $B = 1$ and test for $B = 2$. We proceed in that way until the first acceptance of the null hypothesis. Consider the following model

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

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

$$\mathrm{H}_0 : \zeta_2 = 0. \tag{22}$$

We assume that under this null hypothesis all the parameters are consistently estimated and that the estimators are asymptotically normal. Note that again (21) is only identified under the alternative. Thus, we should proceed as before and expand $F\left(\zeta_2(t-\eta_2)\right)$ into a first order Taylor expansion around $\zeta_2 = 0$. After rearranging terms, the resulting model is

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

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

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 $\hat{\mathbf{h}}_t = \nabla \tilde{G}(\mathbf{z}_t, \mathbf{x}_t; \hat{\boldsymbol{\Psi}}, \hat{\tilde{\boldsymbol{\Psi}}})$, where

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

Defining the residuals estimated under the null as $\hat{\varepsilon}_t = y_t - \tilde{G}(\mathbf{z}_t, \mathbf{x}_t; \hat{\boldsymbol{\Psi}}, \hat{\tilde{\boldsymbol{\Psi}}})$, and following the same steps as before, the test can be carried out in stages as follows.

1. Estimate model (5) with $B = 1$ and compute the residuals $\hat{\varepsilon}_t$. Orthogonalize the residuals by regressing them on $\nabla \tilde{G}(\mathbf{z}_t, \mathbf{x}_t; \hat{\boldsymbol{\Psi}}, \hat{\tilde{\boldsymbol{\Psi}}})$, and compute the residual sum of squares $SSR_0 = \left(\frac{1}{T}\right) \sum_{t=1}^{T} \tilde{\varepsilon}_t^2$.
2. Regress $\tilde{\varepsilon}_t$ on $\hat{\mathbf{h}}_t$ and $\hat{\boldsymbol{\nu}}_t$. Compute the residual sum of squares $SSR_1 = \left(\frac{1}{T}\right) \sum_{t=1}^{T} \hat{v}_t^2$.
3. Compute the χ^2 statistic

$$LM_{\chi^2}^{pc} = T \frac{SSR_0 - SSR_1}{SSR_0}, \tag{25}$$

or the F version of the test

$$LM_F^{pc} = \frac{(SSR_0 - SSR_1)/m}{SSR_1/(T-n-m)}. \tag{26}$$

Under H_0, $LM_{\chi^2}^{pc}$ is approximately distributed as a χ^2 with m degrees of freedom and LM_F^{pc} has approximately an F distribution with m and $T-n-m$ degrees of freedom. Again, to avoid numerical problems, we can omit the terms that depend on the derivatives of the logistic function from the regression in step 2 without affecting the test statistic.

3.2 Test of Serial Independence

Consider the following nonlinear model with autocorrelated errors:

$$\begin{aligned} y_t &= G(\mathbf{z}_t, \mathbf{x}_t; \boldsymbol{\Psi}) + \varepsilon_t = \boldsymbol{\alpha}' \mathbf{z}_t + \sum_{i=1}^{h} \boldsymbol{\lambda}_i' \mathbf{z}_t F(\boldsymbol{\omega}_i' \mathbf{x}_t - \beta_i) + \varepsilon_t, \\ \varepsilon_t &= \boldsymbol{\pi}' \boldsymbol{\nu}_t + u_t, \end{aligned} \tag{27}$$

where $\boldsymbol{\pi}' = [\pi_1, \dots, \pi_r]$ is a parameter vector, $\boldsymbol{\nu_t}' = [\varepsilon_{t-1}, \dots, \varepsilon_{t-r}]$, and $u_t \sim \text{NID}(0, \sigma^2)$. 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 (27) can be consistently estimated by nonlinear least squares.

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

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

$$
\begin{aligned}
l_t = &-\frac{1}{2}\ln(2\pi) - \frac{1}{2}\ln\sigma^2 \\
&- \frac{1}{2\sigma^2}\left\{ y_t - \sum_{j=1}^{r} \pi_j y_{t-j} - G(\mathbf{z}_t, \mathbf{x}_t; \boldsymbol{\Psi}) + \sum_{j=1}^{r} \pi_j G(\mathbf{z}_{t-j}, \mathbf{x}_{t-j}; \boldsymbol{\Psi}) \right\}^2 .
\end{aligned}
\tag{28}
$$

The information matrix related to (28) is block diagonal such that the element corresponding to the second derivative of (28) forms its own block. The variance σ^2 can thus be treated as a fixed constant in (28) when deriving the test statistic. The first partial derivatives of the normal log-likelihood with respect to $\boldsymbol{\pi}$ and $\boldsymbol{\Psi}$ are

$$
\begin{aligned}
\frac{\partial l_t}{\partial \pi_j} &= \left(\frac{u_t}{\sigma^2}\right)\left\{ y_{t-j} - G(\mathbf{z}_{t-j}, \mathbf{x}_{t-j}; \boldsymbol{\Psi}) \right\}, j = 1, \dots, r \\
\frac{\partial l_t}{\partial \boldsymbol{\Psi}} &= -\left(\frac{u_t}{\sigma^2}\right)\left\{ \frac{\partial G(\mathbf{z}_t, \mathbf{x}_t; \boldsymbol{\Psi})}{\partial \boldsymbol{\Psi}} - \sum_{j=1}^{r} \pi_j \frac{\partial G(\mathbf{z}_{t-j}, \mathbf{x}_{t-j}; \boldsymbol{\Psi})}{\partial \boldsymbol{\Psi}} \right\} .
\end{aligned}
\tag{29}
$$

Under the null hypothesis, the consistent estimators of (29) are

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

where $\hat{\boldsymbol{\nu}}_t' = [\hat{\varepsilon}_{t-1}, \dots, \hat{\varepsilon}_{t-r}]$, $\hat{\varepsilon}_{t-j} = y_{t-j} - G(\mathbf{z}_{t-j}, \mathbf{x}_{t-j}; \hat{\boldsymbol{\Psi}})$, $j = 1, \dots, r$, $\hat{\mathbf{h}}_t = \nabla G(\mathbf{z}_t, \mathbf{x}_t; \hat{\boldsymbol{\Psi}})$, and $\hat{\sigma}^2 = (1/T)\sum_{t=1}^{T} \hat{\varepsilon}_t$. The LM statistic is (18) with $\hat{\mathbf{h}}_t$ and $\hat{\boldsymbol{\nu}}_t$ defined as above.

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

The test can be performed in three stages as follows:

1. Estimate model (1) under the assumption of uncorrelated errors and compute the residual $\hat{\varepsilon}_t$. Orthogonalize the residuals by regressing $\hat{\varepsilon}_t$ on $\hat{\mathbf{h}}_t$, and compute the residual sum of squares $SSR_0 =$

$\left(\frac{1}{T}\right)\sum_{t=1}^{T}\tilde{\varepsilon}_t^2$.

2. Regress $\tilde{\varepsilon}_t$ on $\hat{\boldsymbol{\nu}}_t$ and $\hat{\mathbf{h}}_t$. Compute the residual sum of squares SSR_1.

3. Compute the χ^2 statistic

$$LM_{\chi^2}^{si} = T\frac{SSR_0 - SSR_1}{SSR_0}, \tag{30}$$

   or the F version of the test

$$LM_F^{si} = \frac{(SSR_0 - SSR_1)/r}{SSR_1/(T-n-r)}, \tag{31}$$

   where n is the dimension of $\hat{\mathbf{h}}_t$.

Under H$_0$, LM_F^{si} has approximately an F distribution with r and $T-n-r$ degrees of freedom.

3.3 Test of Homoscedasticity Against Smoothly Changing Variance

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

$$\sigma_t^2 = \sigma^2 + \sum_{i=1}^{h}\sigma_i^2 F(\boldsymbol{\omega}_{\sigma,i}'\mathbf{x}_t - \beta_{\sigma,i}), \tag{32}$$

where $\beta_{\sigma,1} \leq \ldots \leq \beta_{\sigma,h}$, 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 $\beta_{\sigma,i}$, $i = 1,\ldots,h$. To circumvent this problem, we rewrite equation (32) as

$$\sigma_t^2 = \exp\left(G_\sigma(\mathbf{x}_t;\boldsymbol{\Psi}_\sigma)\right) = \exp\left(\varsigma + \sum_{i=1}^{h}\varsigma_i F(\boldsymbol{\omega}_{\sigma,i}'\mathbf{x}_t - \beta_{\sigma,i})\right), \tag{33}$$

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 (33) as

$$\sigma_t^2 = \exp\left(\varsigma + \varsigma_1 F(\gamma_\sigma(\tilde{\boldsymbol{\omega}}'_\sigma \mathbf{x}_t - c_\sigma))\right), \tag{34}$$

where $\gamma_\sigma = \|\boldsymbol{\omega}_\sigma\|$, $\tilde{\boldsymbol{\omega}}_\sigma = \boldsymbol{\omega}_\sigma/\gamma_\sigma$, and $c_\sigma = \beta_\sigma/\gamma_\sigma$.

The null hypothesis hypothesis of parameter constancy is

$$\mathrm{H}_0 : \gamma_\sigma = 0. \tag{35}$$

Note that model (34) is only identified under the alternative $\gamma_\sigma \neq 0$. To solve this 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

$$t_1 = \frac{1}{4}\gamma_\sigma \left(\sum_{i=1}^{q} \tilde{\omega}_{\sigma,i} x_{i,t} - c_\sigma \right) + R(\mathbf{x}_t; \gamma_\sigma, \tilde{\boldsymbol{\omega}}_\sigma, c_\sigma), \tag{36}$$

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 (34) by (36), and ignoring $R(\mathbf{x}_t; \gamma_\sigma, \tilde{\boldsymbol{\omega}}_\sigma, c_\sigma)$, we get

$$\sigma_t^2 = \exp\left(\rho + \sum_{i=1}^{q} \rho_i x_{i,t}\right), \tag{37}$$

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

The null hypothesis becomes

$$\mathrm{H}_0 : \rho_1 = \rho_2 = \cdots = \rho_q = 0. \tag{38}$$

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

$$l_t = -\frac{1}{2}\ln(2\pi) - \frac{1}{2}\left(\rho + \sum_{i=1}^{q} \rho_i x_{i,t}\right) - \frac{\varepsilon_t^2}{2\exp\left(\rho + \sum_{i=1}^{q} \rho_i x_{i,t}\right)}. \tag{39}$$

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

$$\frac{\partial l_t}{\partial \rho} = -\frac{1}{2} + \frac{\varepsilon_t^2}{2\exp\left(\rho + \sum_{i=1}^{q} \rho_i x_{i,t}\right)}, \tag{40}$$

$$\frac{\partial l_t}{\partial \rho_i} = -\frac{x_i}{2} + \frac{\varepsilon_t^2 x_i}{2\exp\left(\rho + \sum_{i=1}^{q} \rho_i x_{i,t}\right)}. \tag{41}$$

Under the null hypothesis, the consistent estimators of (40) and (41) 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 $\hat{\sigma}^2 = (1/T)\sum_{t=1}^{T} \hat{\varepsilon}_t^2$. The LM statistic can be written as

$$LM = \frac{1}{2}\left\{\sum_{t=1}^{T}\left(\frac{\hat{\varepsilon}_t^2}{\hat{\sigma}^2} - 1\right)\tilde{\mathbf{x}}_t\right\}'\left\{\sum_{t=1}^{T}\tilde{\mathbf{x}}_t\tilde{\mathbf{x}}_t'\right\}^{-1}\left\{\sum_{t=1}^{T}\left(\frac{\hat{\varepsilon}_t^2}{\hat{\sigma}^2} - 1\right)\tilde{\mathbf{x}}_t\right\}, \tag{42}$$

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 homocedasticity and compute the residuals $\hat{\varepsilon}_t$. Orthogonalize the residuals by regressing them on $\nabla G(\mathbf{z}_t, \mathbf{x}_t; \hat{\mathbf{\Psi}})$, and compute $SSR_0 = \left(\frac{1}{T}\right)\sum_{t=1}^{T}\left(\frac{\tilde{\varepsilon}_t^2}{\hat{\sigma}_{\tilde{\varepsilon}}^2} - 1\right)^2$, where $\hat{\sigma}_{\tilde{\varepsilon}}^2$ is the unconditional variance of $\tilde{\varepsilon}_t$.

2. Regress $\left(\frac{\tilde{\varepsilon}_t^2}{\hat{\sigma}_{\tilde{\varepsilon}}^2} - 1\right)$ on $\tilde{\mathbf{x}}_t$. Compute the residual sum of squares $SSR_1 = \left(\frac{1}{T}\right)\sum_{t=1}^{T}\hat{\nu}_t^2$.

3. Compute the χ^2 statistic

$$LM_{\chi^2}^{\sigma} = T\frac{SSR_0 - SSR_1}{SSR_0}, \tag{43}$$

or the F version of the test

$$LM_F^{\sigma} = \frac{(SSR_0 - SSR_1)/q}{SSR_1/(T-1-q)}, \tag{44}$$

where T is the number of observations.

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

3.3.1 Estimation

If the null hypothesis is rejected we can estimate the parameters of model (33). The estimation algorithm is an extension of the three-phase procedure proposed in Medeiros and Veiga (2000a) and the algorithm in Medeiros and Veiga (2000a). 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 (2000a), assuming that the error variance is fixed.

2. Test the null hypothesis of homoscedasticity. If H_0 is rejected, consider that the conditional mean is correctly specified and estimate the parameters of model (33) by minimizing

$$L_T(\boldsymbol{\Psi}_{\sigma}) = \frac{1}{2}\sum_{t=1}^{T}\left\{\ln(2\pi) + \ln\left(G_{\sigma}(\mathbf{x}_t;\boldsymbol{\Psi}_{\sigma})\right) + \frac{\hat{\varepsilon}_t^2}{G_{\sigma}(\mathbf{x}_t;\boldsymbol{\Psi}_{\sigma})}\right\}. \tag{45}$$

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

$$L_T(\boldsymbol{\Psi},\boldsymbol{\Psi}_{\sigma}) = \frac{1}{2}\sum_{t=1}^{T}\left\{\ln(2\pi) + \ln\left(G_{\sigma}(\mathbf{x}_t;\boldsymbol{\Psi}_{\sigma})\right) + \frac{\left[y_t - G(\mathbf{z}_t,\mathbf{x}_t;\boldsymbol{\Psi})\right]^2}{G_{\sigma}(\mathbf{x}_t;\boldsymbol{\Psi}_{\sigma})}\right\}, \tag{46}$$

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. In order not to estimate a nonlinear model from a time series where there is no 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 experiment.

The simulated models are as follows.

- Model I:

$$\begin{aligned} y_t = {} & 0.5 + 0.8y_{t-1} - 0.2y_{t-2} \\ & + (1.5 + 0.6y_{t-1} - 0.3y_{t-2})F(8.49(0.7071y_{t-1} - 0.7071y_{t-2} + 1.0607)) \\ & + (-0.5 - 1.2y_{t-1} + 0.7y_{t-2})F(8.49(0.7071y_{t-1} - 0.7071y_{t-2} - 1.0607)) \\ & + \varepsilon_t - \rho\varepsilon_{t-1}, \quad \varepsilon_t \sim \text{NID}(0, \sigma_t^2). \end{aligned} \tag{47}$$

- Model II:

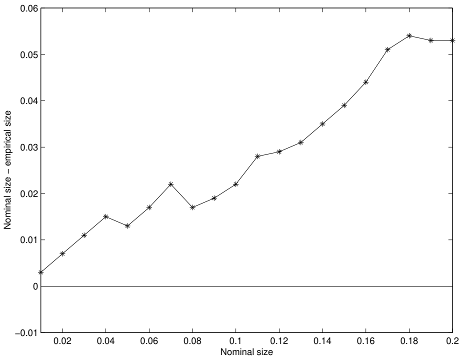
$$y_t = \begin{cases} 0.5 + 0.8y_{t-1} - 0.2y_{t-2} + (1.5 + 0.6y_{t-1} - 0.3y_{t-2})F_1(\cdot) \\ -(0.5 + 1.2y_{t-1} - 0.7y_{t-2})F_2(\cdot) + \varepsilon_t, & \text{if } t \leq 50 \\ -0.8y_{t-1} + (1.2y_{t-1} - 0.7y_{t-2})F_1(\cdot) - (0.6y_{t-1} - 0.3y_{t-2})F_2(\cdot) + \varepsilon_t, & \text{otherwise.} \end{cases} \tag{48}$$

- Model III:

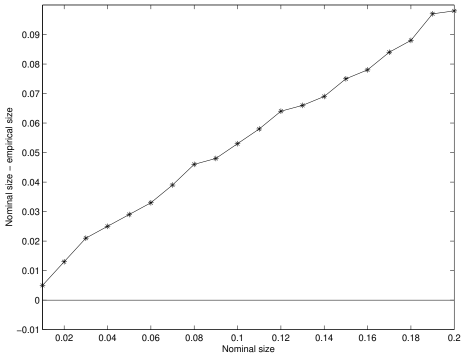
$$y_t = \begin{cases} 0.5 + 0.8y_{t-1} - 0.2y_{t-2} + (1.5 + 0.6y_{t-1} - 0.3y_{t-2})F_1(\cdot) \\ -(0.5 + 1.2y_{t-1} - 0.7y_{t-2})F_2(\cdot) + \varepsilon_t, & \text{if } t \leq 30 \\ -0.8y_{t-1} + (1.2y_{t-1} - 0.7y_{t-2})F_1(\cdot) \\ -(0.6y_{t-1} - 0.3y_{t-2})F_2(\cdot) + \varepsilon_t, & \text{if } 30 < t \leq 60 \\ 3.0 + 0.8y_{t-1} + (0.1y_{t-1} - 0.3y_{t-2})F_1(\cdot) - \\ (0.5 + 1.2y_{t-1} - 0.7y_{t-2})F_2(\cdot) + \varepsilon_t, & \text{otherwise.} \end{cases} \tag{49}$$

In models (48) and (49), $\varepsilon_t \sim \text{NID}(0, 1^2)$, $F_1(\cdot) = F(8.49(0.7071y_{t-1} - 0.7071y_{t-2} + 1.0607))$, and $F_2(\cdot) = F(8.49(0.7071y_{t-1} - 0.7071y_{t-2} - 1.0607))$.

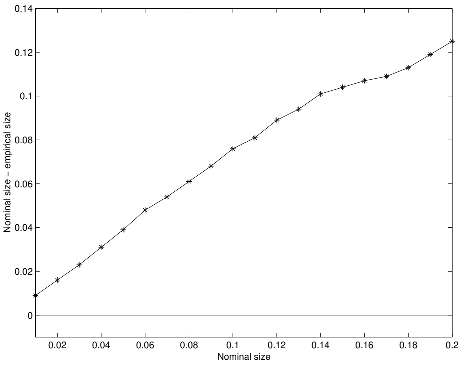
To evaluate the size and power of the tests, we assume that the number of hidden units (h) and the



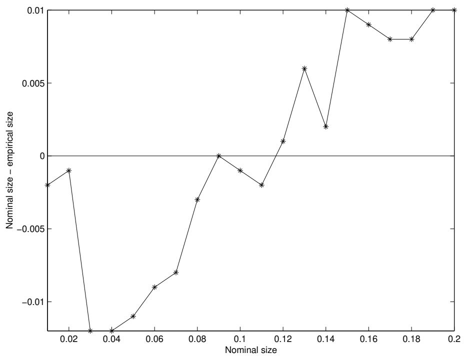
(a)



(b)



(c)



(d)

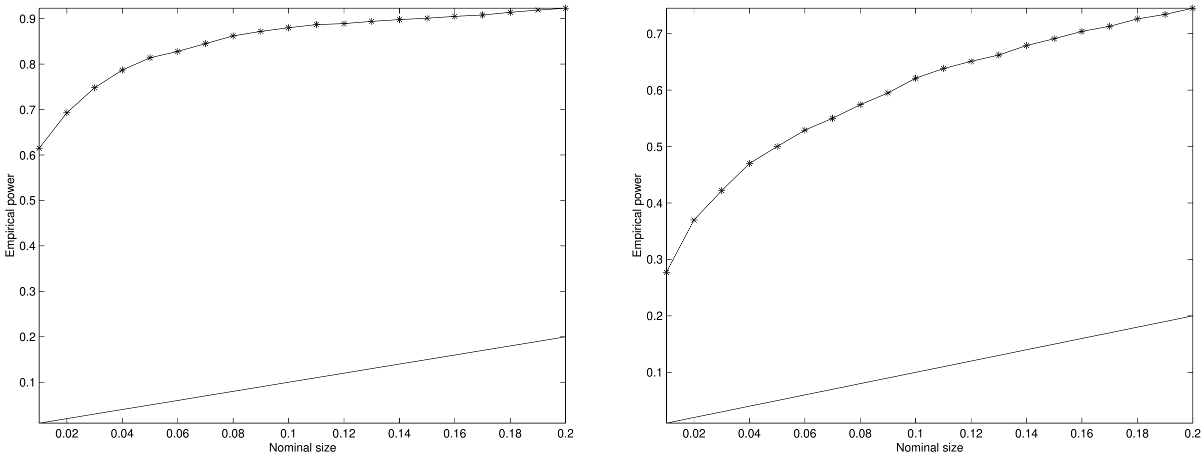
Figure 1: Size discrepancy plot of the parameter constancy test at sample size of 100 observations based on 1000 replications of (a) model (47) with $\rho = 0$ and $\sigma_t^2 = 1$, (b) model (47) with $\rho = 0.2$ and $\sigma_t^2 = 1$, (c) model (47) with $\rho = 0.4$ and $\sigma_t^2 = 1$, (d) model (47) with $\rho = 0$ and σ_t^2 given by (50).

elements of $\mathbf{z}_t$ and $\mathbf{x}_t$ in (1) are correctly specified. In size simulations we generated 1000 time series from model (47) with $\rho = 0$ and $\sigma_t^2 = 1$. Each replication has 100 observations.

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 (48) and (49). Figure 2 shows the power-size curve. The test has good power against models with structural breaks.



(a) (b)

Figure 2: Power-size plot of the parameter constancy test at sample size of 100 observations based on 1000 replications of (a) model (48), (b) model (49).

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.

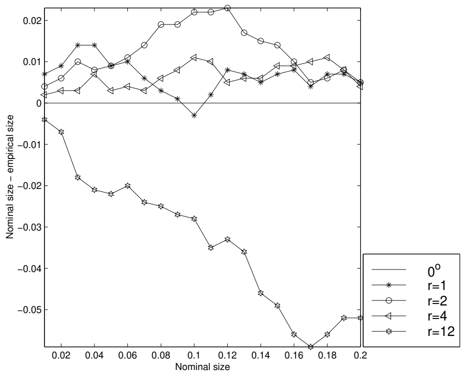
In power simulations of the serial independence test, we generated the data from model (47) with $\rho = 0.2, 0.4$ and $\sigma_t^2 = 1$. Power-size plots are shown in Figure 4. The power of the test increases, as it should, when we increase the value of ρ.

4.3 Test of Homoscedasticity

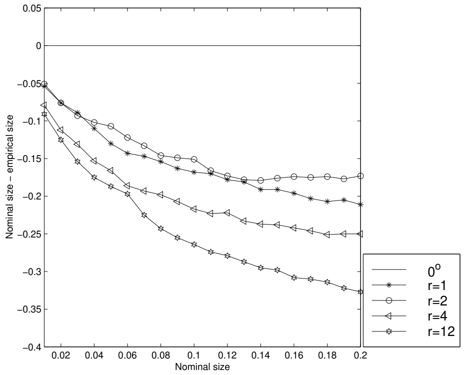
The results of the size simulations are shown in Figure 5. We observe that the empirical size of the test is close to the nominal one. However, the test has non-trivial power against time-varying parameters. In power simulations of the test, we generated the data from model (47) with $\rho = 0$ and

$$\begin{aligned}\sigma_t^2 = \exp(&-0.6931 + 0.6931F(8.49(0.7071y_{t-1} - 0.7071y_{t-2} + 1.0607)) \\ &+ 0.6931F(8.49(0.7071y_{t-1} - 0.7071y_{t-2} - 1.0607))).\end{aligned} \tag{50}$$

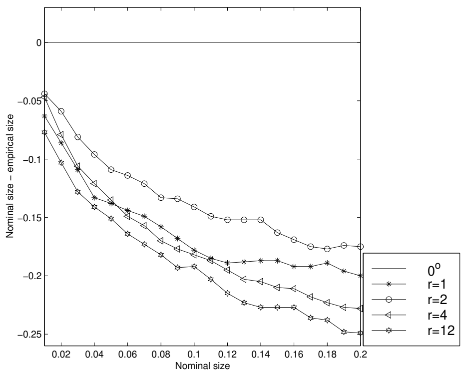
Results are shown in Figure 6.



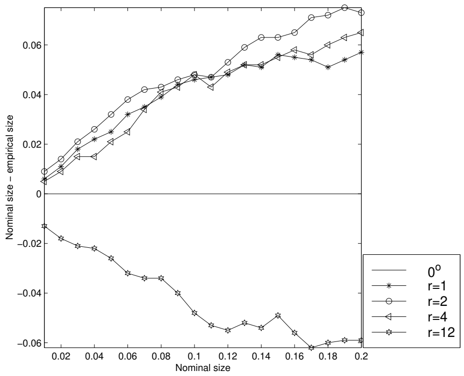
(a)



(b)

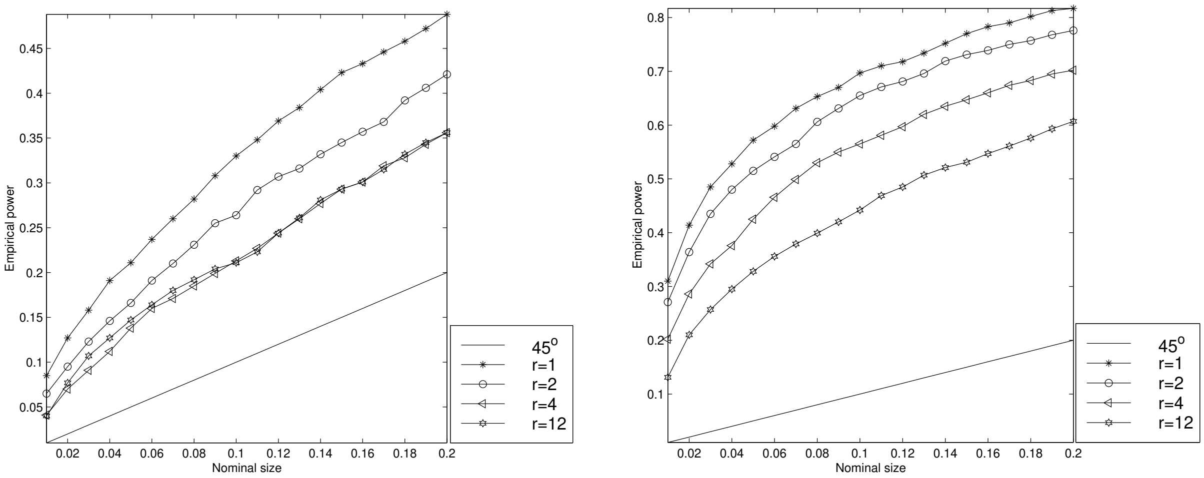


(c)



(d)

Figure 3: Size discrepancy plot of the serial independence test at sample size of 100 observations based on 1000 replications of (a) model (47) with $\rho = 0$ and $\sigma_t^2 = 1$, (b) model (48), (c) model (49), (d) model (47) with $\rho = 0$ and σ_t^2 given by (50).

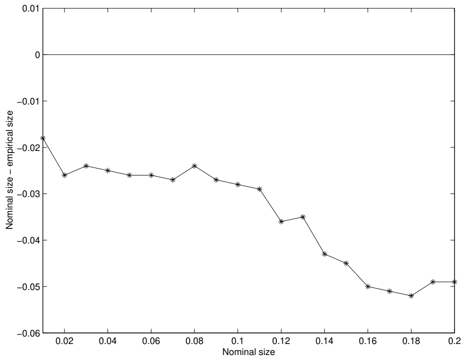


(a) (b)

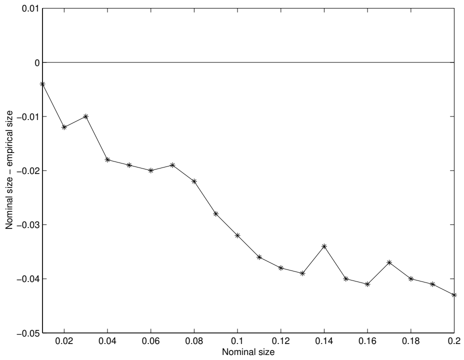
Figure 4: Power-size curve of the test of serial independence at sample size of 100 observations based on 1000 replications of (a) model (47) with $\rho = 0.2$ and $\sigma_t^2 = 1$, (a) model (47) with $\rho = 0.4$ and $\sigma_t^2 = 1$.

5 Conclusions

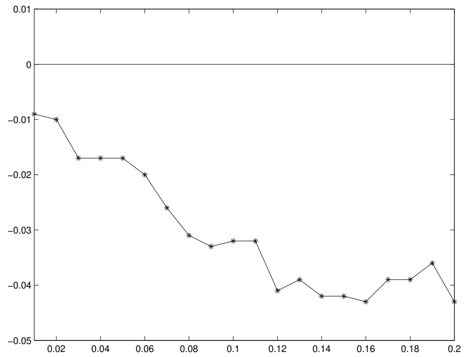
In this paper we consider a sequence of misspecification tests for a flexible nonlinear time series model, called the Neuro-Coefficient Smooth Transition AutoRegressive (NCSTAR) model. They are LM-type tests for testing the hypotheses of parameter constancy, serial independence, and homoscedasticity. 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.



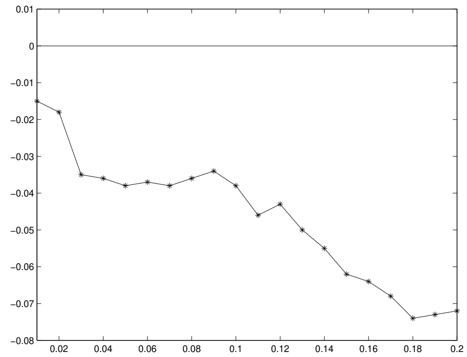
(a)



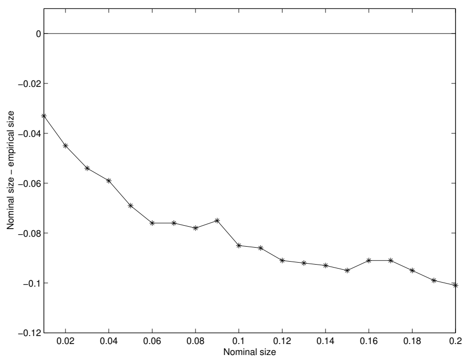
(b)



(c)



(d)



(e)

Figure 5: Size discrepancy plot of the heteroscedasticity test at sample size of 100 observations based on 1000 replications of (a) model (47) with $\rho = 0$ and $\sigma_t^2 = 1$, (b) model (47) with $\rho = 0.2$ and $\sigma_t^2 = 1$, (c) model (48) with $\rho = 0.4$ and $\sigma_t^2 = 1$, (d) model (48), (e) model (49).

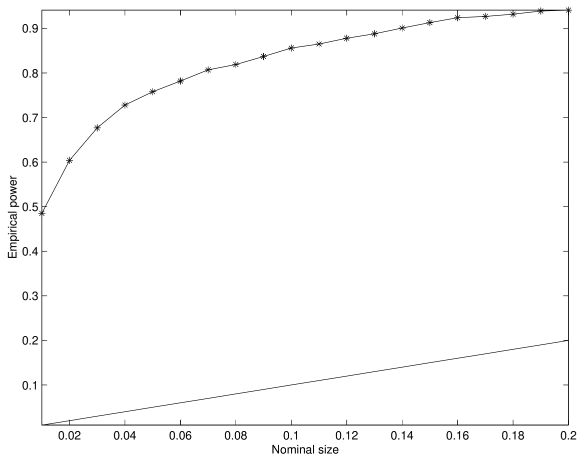


Figure 6: Power-size plot of the heteroscedasticity test at sample size of 100 observations based on 1000 replications of (47) with error variance given by (50).

References

Breusch, T. S. and Pagan, A. R. (1979). A simple test for heteroscedasticity and random coefficent variation, *Econometrica* **47**: 1287–1294.

Breusch, T. S. and Pagan, A. R. (1980). The lagrange multiplier test and its applications to model specification in econometrics, *Review of Economic Studies* **47**: 239–253.

Chan, K. S. and Tong, H. (1986). On estimating thresholds in autoregressive models, *Journal of Time Series Analysis* **7**: 179–190.

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*, second edn, Cambridge University Press, New York, NY.

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

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* M. S. Kearns, S. A. Solla and D. A. Cohn (eds), *Advances in Neural Information Processing Systems*, Vol. 11, 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.

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

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

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.

Teräsvirta, T. (1994). Specification, estimation, and evaluation of smooth transition autoregressive models, *Journal of the American Statistical Association* **89**: 208–218.

Teräsvirta, T. 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.

Teräsvirta, T., Lin, C. F. and Granger, C. W. J. (1993). Power of the neural network linearity test, *Journal of Time Series Analysis* **14**: 309–323.

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

Tong, H. (1983). *Threshold Models in Non-linear Time Series Analysis*, Springer-Verlag, Heidelberg.

Tong, H. (1990). *Non-linear Time Series: A Dynamical Systems Approach*, Oxford University Press, Oxford.

Tong, H. and Lim, K. (1980). Threshold autoregression, limit cycles and cyclical data (with discussion), *Journal of the Royal Statistical Society, Series B* **42**: 245–292.

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.

Appendix

Rewrite equation (39) as

$$l_t = -\frac{1}{2}\ln(2\pi) - \frac{1}{2}\boldsymbol{\varrho}'\tilde{\mathbf{x}}_t - \frac{\varepsilon_t^2}{2\exp(\boldsymbol{\varrho}'\tilde{\mathbf{x}}_t)}, \tag{51}$$

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

$$LM = T\overline{\mathbf{q}}_T(\varrho)'|_{\mathrm{H}_0}\mathbf{I}(\varrho)^{-1}|_{\mathrm{H}_0}\overline{\mathbf{q}}_T(\varrho)|_{\mathrm{H}_0}, \tag{52}$$

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

It is straightforward to show that

$$\overline{\mathbf{q}}_T(\boldsymbol{\varrho}) = \frac{1}{T}\sum_{t=1}^{T}\frac{1}{2}\left(\frac{\varepsilon_t^2}{\sigma_t^2} - 1\right)\tilde{\mathbf{x}}_t. \tag{53}$$

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

$$\mathbf{I}(\boldsymbol{\varrho}) = -E\left(\frac{1}{T}\sum_{t=1}^{T}\frac{\partial^2 l_t}{\partial\boldsymbol{\varrho}\partial\boldsymbol{\varrho}'}\right), \tag{54}$$

where

$$\frac{\partial^2 l_t}{\partial\boldsymbol{\varrho}\partial\boldsymbol{\varrho}'} = -\frac{1}{2}\frac{\varepsilon_t^2}{\exp(\boldsymbol{\varrho}'\tilde{\mathbf{x}}_t)}\tilde{\mathbf{x}}_t\tilde{\mathbf{x}}_t'. \tag{55}$$

Combining (54) with (55), the population information matrix becomes

$$\mathbf{I}(\boldsymbol{\varrho}) = \frac{1}{2T}E\left(\sum_{t=1}^{T}\frac{\varepsilon_t^2}{\exp(\boldsymbol{\varrho}'\tilde{\mathbf{x}}_t)}\tilde{\mathbf{x}}_t\tilde{\mathbf{x}}_t'\right). \tag{56}$$

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

$$\hat{\tilde{\mathbf{q}}}_T(\boldsymbol{\varrho})|_{\mathrm{H}_0} = \frac{1}{2T}\sum_{t=1}^{T}\left(\frac{\hat{\varepsilon}_t^2}{\hat{\sigma}^2} - 1\right)\tilde{\mathbf{x}}_t, \tag{57}$$

and

$$\hat{\mathbf{I}}(\boldsymbol{\varrho})|_{\mathrm{H}_0} = \frac{1}{2T}\sum_{t=1}^{T}\tilde{\mathbf{x}}_t\tilde{\mathbf{x}}_t', \tag{58}$$

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

$$\begin{aligned} LM &= T\left\{\frac{1}{2T}\sum_{t=1}^{T}\left(\frac{\hat{\varepsilon}_t^2}{\hat{\sigma}} - 1\right)\tilde{\mathbf{x}}_t\right\}'\left\{\frac{1}{2T}\sum_{t=1}^{T}\tilde{\mathbf{x}}_t\tilde{\mathbf{x}}_t',\right\}^{-1}\left\{\frac{1}{2T}\sum_{t=1}^{T}\left(\frac{\hat{\varepsilon}_t^2}{\hat{\sigma}} - 1\right)\tilde{\mathbf{x}}_t\right\} \\ &= \frac{1}{2}\left\{\sum_{t=1}^{T}\left(\frac{\hat{\varepsilon}_t^2}{\hat{\sigma}} - 1\right)\tilde{\mathbf{x}}_t\right\}'\left\{\sum_{t=1}^{T}\tilde{\mathbf{x}}_t\tilde{\mathbf{x}}_t',\right\}^{-1}\left\{\sum_{t=1}^{T}\left(\frac{\hat{\varepsilon}_t^2}{\hat{\sigma}} - 1\right)\tilde{\mathbf{x}}_t\right\}. \end{aligned} \tag{59}$$