DEPARTMENT OF ECONOMICS

Working Paper

AN INVESTIGATION OF TESTS FOR LINEARITY AND THE ACCURACY OF FLEXIBLE NONLINEAR INFERENCE

Christian M. Dahl

Working Paper No. 1999-8
Centre for Non-linear Modelling in Economics

ISSN 1396-2426

UNIVERSITY OF AARHUS • DENMARK

CENTRE FOR NON-LINEAR MODELLING IN ECONOMICS
DEPARTMENT OF ECONOMICS - UNIVERSITY OF AARHUS - DK - 8000 AARHUS C - DENMARK
☎ +45 89 42 11 33 - TELEFAX +45 86 13 63 34

WORKING PAPER

AN INVESTIGATION OF TESTS FOR LINEARITY AND THE ACCURACY OF FLEXIBLE NONLINEAR INFERENCE

Christian M. Dahl

Working Paper No. 1999-8

DEPARTMENT OF ECONOMICS
SCHOOL OF ECONOMICS AND MANAGEMENT - UNIVERSITY OF AARHUS - BUILDING 350
8000 AARHUS C - DENMARK ☎ +45 89 42 11 33 - TELEFAX +45 86 13 63 34

An investigation of tests for linearity and the accuracy of flexible nonlinear inference

Christian M. Dahl
CNLME, Department of Economics, University of Aarhus

May 4, 1999

Abstract

A new approach recently suggested by Hamilton for flexible parametric inference in nonlinear models is examined through simulation studies. Hamilton suggests a new test for neglected nonlinearity and we compare it with the neural network test, Tsay's test, White's dynamic misspecification test, Ramsey's Reset test, the so-called V23 test, and the nonparametric BDS test. With respect to size and power properties, the results on the relative performance of Hamilton's test are very encouraging. In particular, we find that against almost all the nonlinear alternatives where the size and power properties of the popular neural network test are good the size and power properties of Hamilton's new test are even better. Secondly, we examine the convergence properties of Hamilton's estimator of the conditional mean function. Our findings suggest that in the case of a true linear relationship, the costs of using the flexible nonlinear approach in terms of efficiency and speed of convergence are minor. We also show that for many nonlinear models the percentage improvement in fit relative to the linear least squared estimator can be substantial. Finally, we present evidence showing that in finite samples the flexible regression approach suggested by Hamilton clearly outperforms the neural network regression approach in terms of accuracy.

- JEL *Classification* : C15,C20
- Keywords: Flexible nonlinear inference; Tests for linearity; Power and size comparison; Convergence in small samples

1 Introduction

Hamilton (1999) suggests a new method of estimating models of the form $y_t = \mu(x_t) + \epsilon_t$, where the functional form of $\mu(x_t)$ is unknown. As a by-product a new Lagrange multiplier test for neglected nonlinearity is suggested. The aim of the paper is twofold. First, we conduct a Monte Carlo experiment analyzing the size and power properties of the new Lagrange multiplier test for neglected nonlinearity suggested by Hamilton (1999). We compare the test with other popular tests for neglected nonlinearity, tests which - similar to Hamilton's test - are not based on any knowledge of the functional form under the alternative. Secondly, we conduct a Monte Carlo experiment examining the convergence properties of Hamilton's estimator of the conditional mean function $\mu(x_t)$. We report results on the performance of the estimator by applying it to a wide range of the most common nonlinear models in the literature, and measures on how big the improvement is relative to the linear estimator are provided. The measures of accuracy are also compared with those obtained from the competing flexible neural network approach. Finally, we investigate the convergence properties of Hamilton's estimator when the true model is linear, in order to determine a potential loss in efficiency and convergence speed. Again a direct comparison to the neural network approach is made. The paper is organized as follows. Section 2 gives a brief introduction to Hamilton's approach to nonlinear inference. We show how to obtain a consistent estimate of the conditional mean function and derive the Lagrange multiplier test statistics for neglected nonlinearity. Section 3 discusses some of the most popular alternative tests for neglected nonlinearity which are available in the literature. Section 4 describes the simulation design for the Monte Carlo experiment and in section 5 the results are reported and discussed. In section 6 the experiment on the convergence properties of Hamilton's estimator is examined and compared with the linear least squared estimator and the estimator based on the neural network approach. Finally, section 7 contains some concluding remarks.

2 Hamilton's approach to flexible nonlinear inference

Consider the model

$$y_t = \mu(x_t) + \epsilon_t \tag{1}$$

where ϵ_t is a sequence of $NI(0, \sigma^2)$ error terms and $\mu(x_t)$ is a function of a $k \times 1$ vector x_t. In most cases a parametric form for $\mu(x_t)$ can be obtained directly from economic theory. However, in the more troublesome cases where economic theory does not give any clear guidance on how to specify the functional form of $\mu(x_t)$, or in situations where the complexity of the data requires more than a simple deterministic model - such as low order polynomials - more general or flexible approaches to represent $\mu(x_t)$ are needed. Building on the ideas of Wahba (1978) and Wecker and Ansley (1983), who viewed $\mu(x_t)$ as a realization of a Brownian motion, Hamilton (1999) suggests representing $\mu(x_t)$ as[1]

$$\mu(x_t) = \alpha_0 + \alpha_1' x_t + \lambda m(g \odot x_t) \tag{2}$$

[1] Here g is a $k \times 1$ vector of parameters and $\odot$ denotes the Hadamard or elementwise product of matrices.

where $m(z)$ - for any choice of z - represents a realization from a random field with the following statistical properties

$$\begin{aligned} m(z) &\sim N(0,1) \\ E(m(z)'m(w)) &= H_k(h) \end{aligned} \tag{3}$$

and where h is defined as $h \equiv \frac{1}{2}[(z-w)'(z-w)]^{\frac{1}{2}}$. The realization of $m(.)$ is viewed as having been settled previous to $\{x_1, .., x_{T,}\epsilon_1, .., \epsilon_T\}$ and is therefore considered to be independent of $\{x_1, .., x_{T,}\epsilon_1, .., \epsilon_T\}$. If we define a variable $G_k(h, r)$ as

$$G_k(h,r) = \int_h^r (r^2 - z^2)^{\frac{k}{2}} dz \tag{4}$$

it is possible to write $H_k(h)$ as

$$H_k(h) = \begin{cases} G_{k-1}(h,1)/G_{k-1}(0,1) & \text{if } h \leq 1 \\ 0 & \text{if } h > 1 \end{cases} \tag{5}$$

Closed form expressions for $H_k(h)$ for $k = \{1, .., 5\}$ are provided by Hamilton (1999) and are reprinted in the appendix of this paper. Since we cannot observe $m(z)$ directly - for any choice of z - we are not able to observe the functional form of $\mu(x_t)$. The objective is to draw inference about the unknown parameters of the model summarized by $\varphi = \{\alpha_0, \alpha_1, \lambda, g, \sigma\}$ by observing the realizations of y_t and x_t only. Using some basic conditioning rules for multivariate normals and treating $\mu(x_t)$ as unobservable, Hamilton (1999) shows how to obtain a maximum likelihood estimate of φ based on a recursive algorithm very similar to the recursive algorithm of the Kalman filter used to obtain the maximum likelihood estimates of state space models. However, in order to cut down the amount of computations Hamilton introduces an equivalent method of calculating the maximum likelihood estimates. He reformulates the model in a more compact form and applies GLS. In particular, he defines

$$\begin{aligned} y &= (y_1, y_2, .., y_T) \\ X &= \begin{bmatrix} 1 & x_1' \\ 1 & x_2' \\ . & . \\ . & . \\ 1 & x_T' \end{bmatrix} \\ \beta &= (\alpha_0, \alpha') \\ \epsilon &= (\epsilon_1, \epsilon_2, .., \epsilon_T) \end{aligned} \tag{6}$$

and shows that the parameters of the linear part of the model consisting of β and σ^2 can be concentrated out of the likelihood function. By defining $\zeta \equiv \frac{\lambda}{\sigma}$ and $W(X; g, \zeta) = \zeta^2 H + I_T$ the concentrated log likelihood function can be written as

$$\eta(y,X;g,\zeta) = -\frac{T}{2}\ln(2\pi) - \frac{T}{2}\ln\widehat{\sigma}^2_T(g,\zeta) - \frac{1}{2}\ln|W(X;g,\zeta)| - \frac{T}{2} \quad (7)$$

$$\widehat{\beta}_T(g,\zeta) = [X'W(X;g,\zeta)^{-1}X]^{-1}[X'W(X;g,\zeta)^{-1}y] \quad (8)$$

$$\widehat{\sigma}^2_T(g,\zeta) = \frac{1}{T}[y - X\widehat{\beta}_T(g,\zeta)]'W(X;g,\zeta)^{-1}[y - X\widehat{\beta}_T(g,\zeta)] \quad (9)$$

where I_T is the identity matrix of dimension $(T \times T)$ and the $\{t,s\}$ entry of the matrix H - denoted $H(t,s)$ - is equal to

$$H(t,s) = \begin{cases} H_k(h_{ts}) & \text{if } h_{ts} \le 1 \\ 0 & \text{if } h_{ts} > 1 \end{cases} \quad (10)$$

$$h_{ts} = \frac{1}{2}[(\widetilde{x}_t - \widetilde{x}_s)'(\widetilde{x}_t - \widetilde{x}_s)]^{\frac{1}{2}}$$

$$\widetilde{x}_t = g \odot x_t$$

The concentrated likelihood function is maximized with respect to (g,ζ) using standard maximization algorithms[2]. Once the estimates of (g,ζ) have been obtained, $\widehat{\beta}_T$ and $\widehat{\sigma}^2_T$ is given. In the case of continuous valued variables and deterministic regressors Hamilton (1999) shows that if the true relation given by (2) is indeed linear then under some regularity conditions the estimator of $\mu(x_t)$ defined as $\widehat{\xi}_T(x_t)$ and given by the t'th row of

$$\widehat{\xi}_T = X\widehat{\beta}_T + \widehat{P}_0(\widehat{P}_0 + \widehat{\sigma}^2_T I_T)^{-1}[y - X\widehat{\beta}_T] \quad (11)$$

is still a consistent estimator of the conditional mean, implying that $\widehat{\beta}_T$ is a consistent estimator. Here the $\{t,s\}$ entry of the matrix P_0 - denoted $P_0(t,s)$ - is defined as

$$P_0(t,s) = \begin{cases} \lambda^2 H_k(h_{ts}) & h_{ts} \le 1 \\ 0 & h_{ts} > 1 \end{cases} \quad (12)$$

$$h_{ts} = \frac{1}{2}[(\widetilde{x}_t - \widetilde{x}_s)'(\widetilde{x}_t - \widetilde{x}_s)]^{\frac{1}{2}} \quad (13)$$

$$\widetilde{x}_t = g \odot x_t \quad (14)$$

Furthermore, Hamilton (1999) proves that his algorithm will provide a consistent estimator of the conditional mean $\mu(x_t)$, for a very general class of nonlinear models, that is

$$T^{-1}\sum_{t=1}^{T}\{\mu(x_t) - \widehat{\xi}_T(x_t)\}^2 \to 0 \quad (15)$$

Since we are going to evaluate the forecast accuracy of the model out-of-sample and equation (11) only works for cases where the conditional mean function is

[2]The *BFGS* and *Newton − Raphson* algorithms included in the CML-routine in GAUSS turned out to work well.

evaluated at points observed in the sample, we have to do some small modifications. To be more specific we seek to calculate an estimate of $\mu(x^*)$ where $x^* = \{x_1^*, x_2^*, .., x_k^*\}$ do not belong to the sample used to obtain the maximum likelihood estimate of $\mu(x_t)$. If we let $P_0^*(t)$ denote the covariance between $\mu(x_t)$ and $\mu(x^*)$ we can obtain an estimate of $\mu(x^*)$ - denoted $\widehat{\xi}_T^*(x_t^*)$ - as

$$\widehat{\xi}_T^*(x_t^*) = X\widehat{\beta}_T + \widehat{P}_0^{*\prime}(\widehat{P}_0 + \widehat{\sigma}_T^2 I_T)^{-1}[y - X\widehat{\beta}_T] \tag{16}$$

where

$$P_0^* = \{P_0^*(t), t = 1, 2, ...T\} \tag{17}$$

$$P_0^*(t) = \begin{cases} \lambda^2 H_k(h_t^*) & \text{if } h_{ts} \leq 1 \\ 0 & \text{if } h_{ts} > 1 \end{cases} \tag{18}$$

$$h_t^* = \frac{1}{2}[(\widetilde{x}_t - \widetilde{x}^*)'(\widetilde{x}_t - \widetilde{x}^*)]^{\frac{1}{2}} \tag{19}$$

$$\widetilde{x}_t = g \odot x_t \tag{20}$$

$$\widetilde{x}^* = g \odot x^* \tag{21}$$

for $t = 1, .., T$. In equation (16) $\widehat{P}_0$ and $\widehat{P}_0^*$ denote P_0 and P_0^* evaluated at the maximum likelihood estimates of λ and g[3].

Testing for neglected nonlinearity in this setup amounts to testing the null $H_0 : \lambda = 0$. When λ equals zero, g is not identified by the model under the null. Hamilton (1999) suggests solving the nuisance parameter problem by fixing g_i to be proportional to the standard deviation of the i'th row in x_t when computing the statistics. Under this assumption the Lagrange multiplier statistics for neglected nonlinearity becomes

$$v^2 = \frac{[\widehat{\epsilon}' H \widehat{\epsilon} - \widetilde{\sigma}^2 tr(MHM)]^2}{\widetilde{\sigma}^4 [2tr\{[MHM - (T-k)^{-1} M tr(MHM)]^2\}]} \tag{22}$$

where

$$\begin{aligned} \widehat{\epsilon} &= My \\ \widetilde{\sigma}^2 &= (T-k)^{-1}\widehat{\epsilon}'\widehat{\epsilon} \\ M &= I_T - X(X'X)^{-1}X' \end{aligned} \tag{23}$$

and the (t, s) element of the matrix H with dimension $(T \times T)$ is given by

[3]The maximum likelihood estimate of λ is obtained as $\widehat{\lambda}_T = \sqrt{\widehat{\zeta}_T^2 \widehat{\sigma}_T^2}$, whereas the maximum likelihood estimate of g is obtained directly from the maximization of $\eta(y, X; g, \zeta)$.

$$
\begin{aligned}
H(t,s) &= \begin{cases} H_k(h_{ts}) & \text{if } h_{ts} \leq 1 \\ 0 & \text{if } h_{ts} > 1 \end{cases} & (24) \\
h_{ts} &= \frac{1}{2}[k^{-1}\sum_{i=1}^{k}\frac{(x_{i,t}-x_{i,s})^2}{s_i^2}]^{\frac{1}{2}} \\
s_i^2 &= T^{-1}\sum_{t=1}^{T}(x_{i,t}-T^{-1}\sum_{t=1}^{T}x_{i,t})^2
\end{aligned}
$$

for $t = 1, .., T$ and $s = 1, .., T$. As mentioned earlier the closed form expressions for H_k, are fairly easy to compute. H_k for $k = \{1, .., 5\}$ is given in the appendix. The Lagrange multiplier statistics v^2 is asymptotically $\chi^2(1)$ distributed. We will evaluate this new test by comparing it to some of the most powerful tests for neglected nonlinearity reported in the literature. These tests will be introduced in the following section.

3 Alternative tests for neglected nonlinearity

In this section we briefly discuss some of the most popular tests for neglected nonlinearity. The tests presented are all selected because of their relatively good performance with respect to size and power properties already reported in the literature. This collection of test statistics will include the Regression Error Specification Test called the Reset2 due to Ramsey (1969) , two tests based on the "duals" of Volterra expansions e.g Priestley (1980), denoted the Tsay1 test according to Tsay (1986) and the V23 test suggested by Terasvirta *et al.* (1993) respectively, the neural network test denoted Neural1 as in Lee *et al.* (1993), and a particular version of White's (1987,1992) information matrix test - for short White3 - aimed at detecting dynamic misspecification. Finally we consider a nonparametric test for serial dependence suggested by Brock, Dechert and Scheinkman (1987) denoted the BDS that has been used rather often as a test for neglected nonlinearity, particularly in the financial literature.

3.1 The Reset, Tsay and V23 tests

The Reset test, Tsay's test and the V23 test can all be conducted within the following framework. Consider the linear model

$$y_t = x_t'\beta + u_t \tag{25}$$

where y_t is the series of interest and where we consider $x_t = \{1, y_t, .., y_{t-p}\}$ to be the relevant variables used to explain y_t. The first step consists of regressing y_t on x_t in order to obtain an estimate of β and to calculate the residuals $\widehat{u}_t = y_t - f_t$ and sum of squared residuals $SSR_0 = \sum_{t=1}^{T}\widehat{u}_t^2$, where $f_t = x_t'\widehat{\beta}$. In the second step regress $\widehat{u}_t$ on x_t and on m auxiliary regressors given by the vector M_t (to be defined later) and compute the residuals from this regression

$\widehat{v}_t = \widehat{u}_t - x_t'\widehat{\phi}_1 - M_t'\widehat{\phi}_2$ and the residual sum of squares $SSR = \sum_{t=1}^{T} \widehat{v}_t^2$. Finally, in the third step compute the F-statistics given by

$$F = \frac{(SSR_0 - SSR)/m}{SSR/(T-p-1-m)} \sim F(m, T-p-1-m) \tag{26}$$

Under the linearity hypothesis the $F-$statistics above is approximately F-distributed with m and $T-p-1-m$ degrees of freedom. The Reset2 test defines $M_t = \{f_t^2, .., f_t^l\}$ and $m = l-1$. Because f_t^i, $i = 1, .., l$ tends to be highly correlated with x_t and with themselves the test is conducted using the $p^* < l-1$ largest principal components of $f_t^2, .., f_t^l$ not collinear with x_t. Tsay (1986) suggests using $M_t = vech(\widetilde{x}_t\widetilde{x}_t')$ for $\widetilde{x}_t = \{x_{1t}, .., x_{pt}\}$ in forming the Tsay1 test, while Terasvirta *et al.* (1993) suggests $M = vec(S * (vech(\widetilde{x}_t\widetilde{x}_t') \odot x_t'))$ (where S is a selection matrix removing the identical entries in $vech(\widetilde{x}_t\widetilde{x}_t') \odot x_t'$ and $\odot$ denotes element-by-element multiplication) when forming the V23 statistics.

3.2 The Neural network test

The neural network test for neglected nonlinearity as suggested by White (1989) and Lee *et al.* (1993) is based on a single hidden layer feedforward network model. In this type of network k input units send signals x_{it} to so-called "hidden" units across weighted connections γ_{ij} for $i = 1, .., k$ and $j = 1, .., q$. There are in total q hidden units each observing the weighted sum of the k input signals, that is, hidden unit j observes $x_t'\gamma_j$ where $x_t = \{1, x_{1t}, .., x_{kt}\}$ and $\gamma_j = \{\gamma_{0j}, \gamma_{1j}, .., \gamma_{kj}\}$. The hidden unit j then outputs a signal $\psi_j(x_t'\gamma_j)$ where ψ_j denotes the "activation" or "squashing" function commonly assumed to be bounded and monotonically increasing. White (1989) and Lee *et al.* (1993) take the activation function to be of the logistic distribution and to be identical for all hidden units, i.e. $\psi_j(x_t'\gamma_j) = \psi(x_t'\gamma_j) = (1 + \exp(-x_t'\gamma_j))^{-1}$ for $j = 1, .., q$. Augmenting the single hidden layer network by direct links from the input units to a single output with weights β and assuming that the output also contains white noise, the total network output can be written as

$$y_t = \rho(x_t, \varkappa) + \epsilon_t \tag{27}$$

where

$$\begin{aligned} \rho(x_t, \varkappa) &= x_t'\beta + \sum_{j=1}^{q} \theta_j (1 + \exp(-x_t'\gamma_j))^{-1} \\ \kappa &= \{\beta, \theta_1, .., \theta_q, \gamma_1, .., \gamma_q\} \end{aligned} \tag{28}$$

The hidden-units-to-output weights are given by $\theta_1, .., \theta_q$ and the noise term distributed according to $\epsilon_t \sim nid(0, \sigma^2)$. When the null hypothesis of linearity is true, i.e. $\mathrm{H}_0 : \Pr[E(y_t|X_t) = x_t'\beta^*] = 1$ for some choices of β^* and $X_t = \{x_1', x_2', .., x_t'\}$, the optimal network weights θ_j are zero for $j = 1, .., q$. The neural network test for neglected nonlinearity can therefore be interpreted as testing the hypothesis $\mathrm{H}_0 : \theta_1 = \theta_2 = .. = \theta_q = 0$ for particular choices of q and

γ_j. As in Lee *et al.*(1993) we set q equal to 10 and draw the direction vectors γ_j independently from a uniform distribution on the interval [-2:2]. The test is then carried out by regressing $\widehat{\epsilon}_{(T\times 1)} = y_T - X_T(X_T'X_T)^{-1}(X_T'y_T)$ on $1_{(T\times 1)}$ and $\Psi_{(T\times q)} = \{\psi(X_T\overline{\gamma}_1)_{(T\times 1)}, .., \psi(X_T\overline{\gamma}_q)_{(T\times 1)}\}'$ where $y_T = \{y_1, y_2, .., y_T\}$ and calculate the uncentered squared multiple correlation coefficient R^2. The LM-test statistics and its asymptotic distribution are given by

$$T * R^2 \to \chi^2(q) \tag{29}$$

Because the observed components of Ψ_t typically are highly correlated Lee *et al.* (1993) recommend using a small number of principal components instead of the q original variables. Using the $q^* < q$ principal components of Ψ_t - denoted Ψ_t^* - not collinear with x_t an equivalent test statistics is given by

$$T * R_{pc}^2 \to \chi^2(q^*) \tag{30}$$

where R_{pc}^2 is the uncentered squared multiple correlation coefficient from a linear regression of $\widehat{\epsilon}_{(T\times 1)}$ on $1_{(T\times 1)}$ and $\Psi^*_{(T\times q*)}$.

3.3 White's dynamic information matrix test

White's dynamic misspecification tests are based on the idea that if a model is correctly specified then there usually exists a number of consistent estimators for the parameters of interest. In particular, if a model is well specified then the information matrix equality will hold under very general conditions. In other words, a test based on the information matrix equality will have power because of the failure of the equality in the case of a misspecified model. The version of White's dynamic misspecification test considered in this paper will be based on the covariance of the conditional score functions. For a Gaussian linear model the log likelihood function can be written as

$$l_t(x_t, \theta, \sigma) = -\frac{1}{2}\log(2\pi) - \log(\sigma) - \frac{1}{2}u_t^2 \tag{31}$$

where $u_t = \sigma^{-1}(y_t - x_t'\beta)$. The conditional score function is then given by

$$s_t^{\nabla}(x_t, \beta, \sigma) = \sigma^{-1}(u_t, u_t x_t', u_t^2 - 1)' \tag{32}$$

Evaluating the conditional score at the quasi maximum likelihood estimators of the correctly specified model under H_0 gives $\widehat{s}_t^{\nabla} = s_t^{\nabla}(x_t, \widehat{\beta}, \widehat{\sigma})$. The information matrix test is based on forming the indicator $\widehat{m}_t^{\nabla} = S^{\nabla} * vec(\widehat{s}_t^{\nabla}\widehat{s}_t^{\nabla\prime})$ where S^{∇} is a selection matrix. In particular we obtain the test statistics denoted White3 in Lee *et al.*(1993) by regressing $\widehat{u}_t = \widehat{\sigma}^{-1}(y_t - x_t'\widehat{\beta})$ on x_t and $\widehat{c}_t$ - where $\widehat{c}_t$ is defined to satisfy $\widehat{m}_t^{\nabla} = \widehat{c}_t\widehat{u}_t'$ - and calculate the uncentered squared multiple correlation coefficient R^2 from this regression. The test statistics and its asymptotic distribution are then given by

$$T * R^2 \to \chi^2(q) \tag{33}$$

where q denotes the dimension of m_t^{∇}.

3.4 The Brock-Dechert-Scheinkman test

Finally, we consider a nonparametric test for neglected nonlinearity suggested by Brock, Dechert and Scheinkman (1987) denoted the BDS test. The test is relevant to include because it has become very popular, particularly in the finance literature. The BDS test is essentially a test for serial dependence based on the correlation integral of the scalar series u_t. The approach begins by organizing the data u_t into a sequence of n-histories denoted u_t^n defined by

$$u_t^n = \{u_{t-n+1}, .., u_t\} \tag{34}$$

where the parameter n is known as the embedding dimension. Next, define the correlation integral $CI_{n,T}$ of u_t as

$$\begin{aligned} CI_{n,T}(\delta) &= \frac{2 * \sum_{t=1}^{T} \sum_{s=1}^{T} I_\delta(u_t^n, u_s^n)}{T_n(T_n - 1)} \\ I_\delta(u_t, u_s^n) &= \begin{cases} 1 & if\ max_{i=0,..,n-1}|u_{t-i} - u_{s-i}| < \delta \\ 0 & otherwise \end{cases} \\ T_n &= T - (n-1) \end{aligned} \tag{35}$$

Hereafter the BDS statistics can be calculated as

$$J_{n,T}(\delta) = T^{-1/2} \frac{CI_{n,T}(\delta) - CI_{1,T}(\delta)^n}{\widehat{\sigma}_{n,T}(\delta)} \tag{36}$$

where $\widehat{\sigma}_{n,T}(\delta)$ is an estimator of the asymptotic standard deviation of $CI_{n,T}(\delta) - CI_{1,T}(\delta)^n$. The BDS statistics is asymptotically standard normal distributed under the null where u_t is IID, that is

$$J_{n,T}(\delta)^2 \sim \chi^2(1) \tag{37}$$

Notice that when applying the BDS test as a test for neglected nonlinearity, u_t will be the residuals estimated from the linear model.

4 The design of the Monte Carlo experiments

We examine the size and power properties of the tests for neglected nonlinearity by considering three blocks of linear and nonlinear dynamic models. All the chosen models have been used in previous studies on the testing of linearity. The use of "benchmark" models makes it much to draw comparisons to earlier studies and it furthermore prevents the results reported from being to design specific in the sence that the general design of the Monte Carlo experiment actually is predetermined. The models included in block-1 and the two bivariate models shown in table 1 were originally used by Lee *et al.* (1993). The models of block-2 have been more extensively studied, in particular by Keenan (1985), Tsay(1986), Ashley, Pattersen and Hinich (1986), Chan and Tong (1986) , and Lee *et al.* (1993). Finally, all of the models in block-3 have been studied by

Terasvirta *et al.* (1993). The five models contained in block-1 are all characterized by being simple dynamic univariate models, where the dynamic part is represented by one lag of the endogenous variable only. The models are all stationary. The models included are the autoregressive model (AR), the bilinear model (BL) of Granger and Anderson (1978), the threshold autoregressive model (TAR) of Tong (1983), the sign autoregressive model (SGN), and the nonlinear autoregressive model (NAR). The exact parameterization of the models is given in table 1. We also consider two bivariate representations. For simplifying reasons, we do not impose any dynamic structure on the bivariate models. We consider a squared relation which we denote SQ, and we consider an exponential relation, denoted EXP. We consider these two bivariate models for 3 different values of σ. Varying σ, keeping the other parameters fixed, alters the signal-to-noise ratio. We investigate how this affects the size and power properties of the various test for neglected nonlinearity. The parameterization of the bivariate models is also shown in table 1.

[*Table* 1]

The models in block-2 are characterized by having a much richer dynamic structure compared to the models in block-1. The models are presented in table 2. model-1 is an MA(2) representation and model-2 is a heteroskedastic MA(2), due to the last term on the right hand side. These two models together with model-4 - an AR(2) model - are all linear models. They are included primarily to evaluate the nominal size of the nonlinearity tests and their ability to distinguish between dynamic misspecification and misspecification due to nonlinearity in the conditional mean. model-3, model-5 and model-6 are the truly nonlinear models in block-2. model-3 is a nonlinear MA(2) model. model-5 and model-6 belong to the family of bilinear models. model-5 is a bilinear autoregressive model, while model-6 is a bilinear autoregressive moving average model.

[*Table* 2]

Terasvirta *et al.* (1993) argue that the main reason for the neural network test to perform so very well compared to a wide range of other linearity tests in the simulation studies by Lee *et al.* (1993) is because they did not include the appropriate LM or LM type tests. By the appropriate LM type or LM type tests Terasvirta *et al.* (1993) refer to tests particularly designed to test linearity against a fully specified nonlinear alternative. Now, the simulation design in Lee *et al.* (1993) is only concerned with evaluating linearity tests where the alternative does not have to be fully specified. However the critique raised by Terasvirta *et al.* (1993) is still relevant in the sense that the choice of testor-model-mix may undeliberately favor the neural network test. The models in block-3 are included in order to reduce this possible source of bias. Still we will restrict ourselves only to consider the general class of tests for linearity for which the nonlinear alternatives do not need to be explicitly specified. The first model in block-3 is the logistic smooth transition autoregressive model (LSTAR). Its properties are discussed in details in Terasvirta (1990). The second model is a special case of the exponential smooth transition autoregressive model (ESTAR). By the parameterization chosen, the model reduces to the exponential autoregressive model of Haggan and Ozaki (1981). The NN and BN models denote univariate and bivariate neural network models, respectively.

$[Table\ 3]$

Throughout $\epsilon_t \sim N(0,1)$ is a white noise series. The information set in the block1 models and bivariate models contains $\{y_{t-1}\}$ and $\{x_t\}$, respectively. The information set for the models contained in the block-2 and block-3 - except the BN model - equals $\{y_{t-1}, y_{t-2}\}$. For the BN model the information set contains $\{y_{t-1}, x_t\}$. The exact parameterization of the Reset2, Tsay1 and V23 test is summarized in table 4. For the neural network test $q^* = 2$ for the block-1 models and the bivariate models. When applied to the models in block-2+3, $q^* = 3$. In constructing the BDS test statistics the embedding dimension equals $n = 2$ while the measure of closeness equals $\delta = (0.8)^6$ for all the models under consideration[4].

$[Table\ 4]$

5 Results on size and power properties

In order to make a comparison with previous studies of size and power properties as straightforward as possible, the setup in this section follows the general design outlined in Lee *et al.*(1993). The results from a simulation of the critical values at a 5% level are shown in table 5. The simulations are based on data being generated from the AR model in block-1. From inspection of the critical values generated by Hamilton's Lagrange multiplier test it appears to have quite good size properties in the sense that the simulated values based on finite samples are very close to the critical values based on the asymptotic distribution. In general, the size properties of the test for neglected nonlinearity in table 5 seem good (in the sense that the simulated size seems to correspond well with the nominal/asymptotic size) when simulations are based on the AR model.

$[Table\ 5]$

The results on size properties do not change very much, when the simulation of the critical values at a 5% level is based on model-4 in block-2. Again the simulated critical values of Hamilton's test are close to the asymptotic values,

[4]The simulation results on the BDS test statistics are all based on a light modified version of Pedro de Lima's Gauss-code `bds.prc`. The code can be downloaded directly from Pedro de Lima's homepage.

and in general this result holds for the other test statistics considered as well, see table 6. Furthermore, one should notice the rather high percentage of rejections of true linear models by the BDS statistics even at sample size $T = 200$. This result stresses the importance of size correcting the nonparametric BDS statistics when applied to "small" samples.

$[Table\ 6]$

Next we analyze the sensitivity of the simulated critical values at a 5% significance level, when the autoregressive coefficient of the AR model in block-1 is changing. The reason for this exercise is mainly to examine how sensitive the tests are to increasing persistence in the time series. In particular, we simulate a set of critical values based on the models $y_t = \rho y_{t-1} + \epsilon_t$, $\rho = \{-0.9, -0.6, -0.3, 0, 0.3, 0.6, 0.9\}$. After simulating these critical values we generate data from the AR model and count the number of times (in percentage) the test rejects the null of linearity based on the simulated set of critical values. As an example notice that based on the simulated 5% critical value generated from the AR(1) model with $\rho = 0.9$, Hamilton's test rejects the hypothesis of linearity in 5% of all cases, when the data actually comes from the AR model of block-1. Ideally the rejection percentages in table 7 should equal 5% in all cases. The 95% critical value around 5% is $CI_{0.95} = \{3.6; 6.4\}$. Regarding Hamilton's test, the simulated sizes are all inside the confidence interval except in the case where $\rho = -0.9$. However, Hamilton's test has a satisfactory, relatively low spread in size, ranging from 3.3 to 5.0

$[Table\ 7]$

Based on the simulated critical values reported in table 5, table 8 shows the results on power of the tests using the models in block-1 and the bivariate models with $\sigma = 1$, and sample size varying from $T = 50$ over $T = 100$ to $T = 200$. It becomes evident that Hamilton's test has very strong power against the TAR, SGN, SQ and EXP alternatives. For these four nonlinear models the power of Hamilton's test is at least as high or even higher than the power of the Neural1 test. Hamilton's test has low power against the NAR model, but this is a common feature shared by all of the tests. Hamilton's test has also low power against the bilinear alternative. Here only White's test and the otherwise disappointing BDS test have good power properties.

$[Table\ 8]$

table 9 shows the power properties of the test based on block-2 models and the simulated critical values reported in table 6. By inspection of table 9 we notice that for the true nonlinear models, the power of Hamilton's test is almost as good as the power of the neural network test in case of model 3 and at least as good or better in the case of model 5 and model 6. Furthermore, looking at the rejection frequencies for the linear models it seems evident that the size properties of Hamilton's test and the Neural1 test are almost identical. This implies that also the size-corrected power properties of Hamilton's test appear to be good compared to the Neural1 test. Also the Tsay1, the White3 and the V23 tests seem to have a little more power against model-3 relative to Hamilton's test. However, their size properties in the case of model-1 and model-2 are not as good as the size properties of Hamilton's test. This might indicate that the size-corrected power for these tests may be somewhat lower than the rejection frequencies actually reported. However, as pointed out by Lee *et al*. (1993) and Granger and Terasvirta (1993) ARCH effects cause the size of the neural network test, the Tsay1, White3, Reset2 and V23 test to be incorrect. By inspection of table 9 and the results based on model 2 it becomes clear that this particular feature also seems to be shared by Hamilton's test. Again the power properties of the size corrected BDS statistics seem inferior. However it does appear to be relatively robust to ARCH effects which on the other hand is a bit surprising because asymptotically the statistics is not able to distinguish between nonlinearities in the mean and in the variance!

[*Table* 9]

Considering the results on power of the various test statistics when applied to the models of block-3, Hamilton's test again seems to perform relatively well. In the case of the neural network models, the neural network test and the V23 test turn out to be the appropriate LM-test statistics - apart from a missing constant, see Terasvirta *et al*. (1993) . As expected, their power properties are very good when applied to the neural network models. However, the power properties of Hamilton test are just as good. With respect to the LSTAR model all the tests considered have very good power. If the nonlinearity is of the ESTAR type, only Hamilton's test, White3, Reset2 and the V23 test have satisfactory power. Against the ESTAR type of nonlinearity the neural network test has very low power. Also based on the models in block-3 the power properties of the BDS statistics appear to be very poor[5]. The results on the block-3 models again seem to confirm that Hamilton's test has better power or at least as good power properties as the neural network test even in the case where the true nonlinear model is in fact of the neural network type.

[*Table* 10]

table 11 and table 12 show the power of the nonlinearity tests against the SQ alternative and the EXP alternative respectively, when the signal-to-noise ratio decreases. The results from these tables suggest that Hamilton tests perform almost as well as the Neural1, Tsay1 and the Reset2 test, when the signal-to-noise ratio falls. The White3 test performs very poorly in terms of power when the signal-to-noise ratio decreases in these bivariate models.

[*Table* 11]

[5] The power of the BDS statistics seems suspiciously low. However, increasing the numerical value of σ^2 in the models of block-3,- making the series look much more nonlinear, turned out to improve a lot upon the power properties of the BDS test as it should.

[*Table* 12]

6 Convergence properties of Hamilton's estimator

In the following we will analyze and compare five different measures of convergence. These measures are defined as

$$\begin{aligned}
C^1_{T,N} &= N^{-1}\sum_{n=1}^{N}[T^{-1}\sum_{t=1}^{T}\{\mu(x_{t,n})-\widehat{\alpha}'_{T,n}x_{t,n}\}^2] \quad (38)\\
C^2_{T,N} &= N^{-1}\sum_{n=1}^{N}[T^{-1}\sum_{t=1}^{T}\{\mu(x_{t,n})-\widehat{\xi}_T(x_{t,n})\}^2]\\
C^3_{T,N} &= N^{-1}\sum_{n=1}^{N}[T^{-1}\sum_{t=1}^{T}\{\mu(x_{t,n})-\xi_T(x_{t,n})|_{\lambda=\sigma,g=2(kV(x_n))^{-1/2}}\}^2]\\
C^4_{T,N} &= N^{-1}\sum_{n=1}^{N}[T^{-1}\sum_{t=1}^{T}\{\mu(x_{t,n})-\rho(x_{t,n},\widehat{\kappa}_{T,n})\}^2]\\
C^5_{T,N} &= N^{-1}\sum_{n=1}^{N}[T^{-1}\sum_{t=1}^{T}\{\mu(x_{t,n})-\widehat{\omega}'_{T,n}\widetilde{x}_{t,n}\}^2]
\end{aligned}$$

Common for all five measures is that $\mu(x_{t,n})$ denotes a realization of the true functional form conditional on $x_{t,n}$, N denotes the number of replications in the Monte Carlo experiment, and T equals the number of observations in each sample. $C^1_{T,N}$ is the average mean square error between $\mu(x_{t,n})$ and a linear ordinary least square estimator of $\mu(x_{t,n})$. The linear estimator is given by $\widehat{\alpha}'_{T,n}x_{t,n}$ where $\widehat{\alpha}_{T,n}$ is obtained from an ordinary least square regression of $y_{t,n}$ on $x_{t,n}$ for $t=\{1,..T\}$. $C^2_{T,N}$ is the mean square error between $\mu(x_{t,n})$ and Hamilton's maximum likelihood based estimator of $\mu(x_{t,n})$ - denoted $\widehat{\xi}_T(x_{t,n})$ - averaged over N replications. $\widehat{\xi}_T(x_{t,n})$ is obtained by maximizing the profile likelihood function stated in equation (**??**) and equation (8) with respect to (g,ζ). If Hamilton's maximum likelihood based estimator improves upon the linear least square estimator, one would expect $C^2_{T,N} < C^1_{T,N}$. If the estimator is consistent, the $C^2_{T,N}$ should converge to zero when the sample size increases whereas $C^1_{T,N}$ should converge to a positive constant if the true model is nonlinear. $C^3_{T,N}$ is the average mean square error between $\mu(x_{t,n})$ and a generic version of Hamilton's estimator of $\mu(x_{t,n})$ - denoted $\xi_T(x_{t,n})|_{\lambda=\sigma,g=2(kV(x_n))^{-1/2}}$ - where λ and g equal σ and $2(kV(x_n))^{-1/2}$ respectively. We refer to $\xi_T(x_{t,n})|_{\lambda=\sigma,g=2(kV(x_n))^{-1/2}}$ as generic because it is obtained without involving any kind of estimation. It would be natural to compare the convergence properties of Hamilton's estimator not only with the linear estimator but also with competing flexible regression methods. A natural choice for comparison would be the neural network regression model that has gained increasingly popularity in applied time series analysis lately, see Swanson and White (1997) and Stock and Watson (1998) and the references herein. Consequently, we let $C^4_{T,N}$ denote the average mean squared error between $\mu(x_{t,n})$ and the estimator of the conditional mean function in the neural network model specified in equation (28) and denoted $\rho(x_{t,n}, \widehat{\kappa}_{T,n})$. To be more specific the estimated parameters of $\rho(x_{t,n}, \kappa_{T,n})$ are obtained by the method of nonlinear least square, ie. as a solution to problem

$$\min_{\kappa_{T,n}} \sum_{t=1}^{T} (y_{t,n} - \rho(x_{t,n}, \kappa_{T,n}))^2$$

The appropriate number of hidden units in the Neural Network model is in general unknown even when the true DGP is known, hence q has to be estimated. Following the approach suggested in a series of papers by Swanson and White, e.g. Swanson and White (1995,1997a,1997b), the number of hidden units q is determined in a forward stepwise manner by first adding the linear part and then by adding one hidden unit at a time until the BIC model selection criterion can no longer be improved upon. To stress the fact that q is actually estimated simultaneously with $\kappa_{T,n}$, we denoted the vector of estimated parameters $\kappa_{T,n}$ as a function of the estimated value of q, i.e. as $\widehat{\kappa}_{T,n}(\widehat{q})$. Finally, we compute $C^5_{T,N}$ defined as the average mean square error between $\mu(x_{t,n})$ and the least square estimator of $\mu(x_{t,n})$ based on knowledge of the true functional form i.e. the true nonlinear regressors. $\widehat{\omega}_{T,n}$ is obtained as the least square estimator from a regression of $y_{t,n}$ on $\widetilde{x}_{t,n}$ for $t = \{1, ..T\}$ where the regressors $\widetilde{x}_{t,n}$ are defined explicitely in table 13 for all of the models under consideration. In all models except model-2 $\widehat{\omega}_{T,n}$ is a consistent and efficient estimator, implying that $\widehat{\omega}'_{T,n}\widetilde{x}_{t,n}$ will be a consistent (and efficient) estimator of $\mu(x_{t,n})$. For that reason $C^5_{T,N}$ has the interpretation of being an estimate of the lower bound on the average mean squared error between $\mu(x_{t,n})$ and any possible estimate of $\mu(x_{t,n})$ conditional on available information up to time $t-1$. Finally it is worth mentioning that $C^i_{T,N}$, $i = 1, .., 5$ are all out-of-sample measures. This will become evident from the following description of the simulation design:

1. For every $n = 1, .., N$ draw the sequence $\{y^*_{t,n}, x^*_{t,n}, \widetilde{x}^*_{t,n}\}_{t=1}^{T}$ from the model under consideration. Based on these realizations obtain the various estimates given by $\{\widehat{\alpha}_{T,n}, \widehat{\omega}_{T,n}, \widehat{\xi}_{T,n}, \xi_{T,n}, \widehat{\kappa}_{T,n}(\widehat{q})\}$.

2. For every $n = 1, .., N$ draw a whole new sequence $\{y_{t,n}, x_{t,n}, \widetilde{x}_{t,n}\}_{t=1}^{T}$ from the model under consideration. Compute $C^i_{T,N}$, $i = 1, .., 5$ based on $\mu(x_{t,n})$, $\widehat{\alpha}'_{T,n}x_{t,n}$, $\widehat{\omega}'_{T,n}\widetilde{x}_{t,n}$, $\widehat{\xi}_T(x_{t,n})$, $\xi_T(x_{t,n})|_{\lambda=\sigma, g=2(kV(x_n))^{-1/2}}$ and $\rho(x_{t,n}, \widehat{\kappa}_{T,n}(\widehat{q}))$ for $n = 1, .., N$ and $t = 1, .., T$.

By this approach it is possible to avoid the effects of overfitting in-sample, a common feature often associated with flexible nonlinear modelling. In fact, the costs of capturing spurious nonlinear patterns turn out to be very high by this approach.

In order to get an impression of how accurate the nonlinear estimators actually are relative to the linear least square estimator in finite samples we consider the measure $G^i_{T,N}$ defined as

$$G^i_{T,N} = 1 - \frac{C^i_{T,N}}{C^1_{T,N}} \ , \quad i = 2,3,4 \tag{39}$$

where $G^i_{T,N}$ is a measure of the percentage improvement in fit of the nonlinear estimator(s) relative to the linear estimator.

[$Table$ 13]

Now turning to the outcome of the simulations, table 14, table 15, and table 16 report the estimates of $C^1_{T,N}$, $C^2_{T,N}$,$C^3_{T,N}$, $C^4_{T,N}$,and $C^5_{T,N}$ together with the improvement in fit measures $G^i_{T,N}$,for $i = 2,3,4$ based on a Monte Carlo experiment with $N = 100$. Looking at the Monte Carlo results for the linear AR model it is evident that $C^1_{T,N}$ and $C^2_{T,N}$ numerically are approximately of the same size. This implies that the rate of convergence of $C^1_{T,N}$ and $C^2_{T,N}$ is almost identical. The result based on the $AR(2)$ model, i.e. model-4 in table 15, seems to confirm this result although the average mean squared errors from the linear regression tend to be a bit lower than the average mean squared errors obtained from the flexible nonlinear estimator. However, the speed of convergence[6] of the latter seems to be equally high. This suggests that when the true model is truly linear and the linear model is correctly specified in terms of $\widehat{\alpha}'_{T,n}x_{t,n}$, the rate of convergence of $\widehat{\xi}_T(x_{t,n})$ to $\mu(x_{t,n})$ in finite samples is almost as fast as that of $\widehat{\alpha}'_{T,n}x_{t,n}$, that is, little is lost by forming a general nonlinear inference according to Hamilton's method when the true relation is linear.

[$Table$ 14]

In general the convergence of $\widehat{\xi}_T(x_{t,n})$ appears to be good for all the models in block-1, except for the BL model. The measure of the percentage improvement in fit of Hamilton's estimator relatively to the linear estimator in block-1 is largest in the case of nonlinearity of the TAR and SGN type. Here the improvement of Hamilton's estimator is about 70 pct., when $T = 200$. The improvement of fit is very modest in the case of the NAR model, which may seem a little disappointing. Looking at the AR model we see that loss of fit in the cases where $T = 50$ or $T = 100$ seems very minor while it tends to increase a little when T grows to 200 . It is also worth noticing that in the case of the TAR and SGN alternative the percentage improvement in fit of the generic estimator is larger than the improvement in the maximum likelihood based estimator when

[6]As suggested by Ngern (1996) a simple measure of the rate of convergence of $C^n_{T,N}$ can be obtained by assuming that $C^j_{T,N} = \vartheta_{n,N}T^{-\delta^j}, j = \{1,..,4\}$where $\vartheta_{n,N}$ is a constant independent of T, and δ^j for $j = \{1,..,5\}$ is the measure of the rate/speed of convergence

$T = 50$. However, the rate of convergence of this estimator seems in general to be somewhat lower. Comparison of $C^2_{T,N}$ and $C^5_{T,N}$ clearly illustrates that although Hamilton's estimator improves a lot upon the estimation of $\mu(x_{t,n})$ relatively to the linear estimator, it is still lacking some efficiency, as one would expect. The results on the bivariate models show the overall best percentage improvement in fit measures amounting to nearly 90 pct. when the true model is the SQ model, and nearly 60 pct. in the case of an EXP model. Also the convergence measures are very good. These result suggest very clearly that applying Hamilton's estimator to nonlinear multivariate models seems to be a particularly fruitful approach. The flexible estimator of the conditional mean function based on the neural network approach also turns out to contribute to a significant improvement in fit over the linear least square estimator in case of nonlinearities, particularly in the TAR, SGN, SQ and EXP case. However the improvement in fit of the neural network approach in case of the nonlinearities reported in table 1 never exceeds that of Hamilton's approach. Furthermore, in the case of the true model being linear the loss of efficiency of the neural network approach relatively to the least square estimator can be substantial. In case of the AR model in table 14 and model-4 in table 15 the loss amounts to about 50 pct.

[*Table* 15]

By inspecting the results on the block-2 models reported in table 15, we observe that for all the models that are linear in mean, $C^2_{T,N}$ tends to be about 10-20 pct. higher than $C^1_{T,N}$ for $T = 50$. However, in the case of model-1 and model-2 the efficiency loss of the flexible approach seems to be reduced rather quickly as the sample size increases because the convergence rate $C^2_{T,N}$ tends to be somewhat higher than the convergence rate for $\mathrm{C}^1_{T,N}$. The percentage improvement of fit arising from applying Hamilton's estimator to model-3 - the nonlinear MA(2) model - is rising from about zero pct. when $T = 50$ to approximately 30 pct. when $T = 200$. As it is the case for some of the models in block-1, the generic estimator again seems to perform almost as well as the maximum likelihood based estimator when it comes to the percentage improvement in fit in situations where the true underlying model is nonlinear. Looking at the results from the bilinear models 5 and 6 when $T = 200$ we notice that the improvement in fit amounts to about 25 pct. and 5 pct. respectively. In particular, in the case of model-6 the rate of convergence of Hamilton's estimator appears rather low as it was the case for the bilinear model in block-1. One reason for the poor convergence results observed with respect to two of the bilinear models could be caused by some very undesirable properties featuring this family of models as pointed out by Brunner and Hess (1995). They show that the expected likelihood function associated with the bilinear models in some cases will exhibit bimodality, with the true optimum characterized by a long narrow spike that become more pronounced as the sample size increases. Furthermore, these features become even more pronounced for parameterizations where the model is close to violating at least one of four moment restrictions that establish invertability and stationarity conditions. In this light Brunner and Hess (1995) recommend *extreme caution* when dealing with the bilinear models. By comparison with the estimates obtained from the neural network approach the picture from table 14 seems to be confirmed in the sense that the neural network model again contributes to an improvement in fit over the linear estimator in case of nonlinearities, but that it does not appear to be nearly as efficient as Hamilton's approach in finite samples.

[*Table* 16]

The results based on the regime switching and the neural network models of block 3 are indeed encouraging. The improvement in fit of the flexible nonlinear estimator ranges from 45 pct. in the case of the true model being an ESTAR model up to 60 pct. for the LSTAR model in the case of 200 observations. In addition, the speed of convergence of the flexible nonlinear estimator is much higher than the speed of convergence of the linear estimator promising an even higher improvement in fit when the sample size increases. The performance of the neural network approach seems rather disappointing in the case of the LSTAR and ESTAR models. However, in the case of the data being generated from a true neural network model, as in the NN and BN models the approach is able to outperform Hamilton's flexible regression approach when the sample size is of a limited size.

7 Conclusion

We find that the new test for neglected nonlinearity proposed by Hamilton performs well in finite samples. In general, it has good size and power properties when compared to existing tests. In particular, our findings indicate that against nonlinear alternatives where the power properties of the neural network test are good, the power properties of Hamilton's test in most cases are even better. Looking at the properties of Hamilton's nonlinear estimator our main finding is that even in situations where the true model is linear, the costs of using the flexible nonlinear approach are limited in terms of efficiency and speed of convergence. We have also found that for many nonlinear models the percentage improvement in fit of Hamilton's estimator relative to the least square estimator can be substantial. Comparing Hamilton's approach with the neural network regression models approach, we find that based on all the nonlinear models considered - apart from the true neural network models - there can be an substantial increase in accuracy and efficiency by using Hamilton's flexible regression model approach instead of the neural network approach.

Acknowledgements: Comments and suggestions from James D. Hamilton, Niels Haldrup and Svend Hylleberg are gratefully acknowledged. Furthermore, I would like to thank participants at the Econometrics Workshop at UCSD for helpful comments. Any errors are my own.

References

Ashley, R., Patterson, D. & Hinich, M. (1986), 'A diagnostic test for nonlinear serial dependence in time series fitting errors', *Journal of Time Series Analysis* **7**, 165–178.

Brock, W., Deckert, W. & Scheinkman, J. (1987), A test for independence based on the correlation dimension. Unpublished manuscript, University of Wisconsin at Madison, University of Houston and University of Chicago.

Brunner, A. D. & Hess, G. D. (1995), 'Potential problems in estimating bilinear time-series models', *Journal of Economic Dynamics and Control* **19**, 663–681.

Chan, W. S. & Tong, H. (1986), 'On tests for nonlinearity in time series models', *Journal of Forecasting* **5**, 217–228.

Granger, C. W. J. & Andersen, A. (1978), *An introduction to bilinear time series models*, Vandenhoech und Ruprecht, Gottingen.

Haggan, V. & Ozaki, T. (1981), 'Modelling nonlinear random vibrations using an amplitude dependent autoregressive time series model', *Biometrika* **68**, 189–196.

Hamilton, J. D. (1999), 'A parametric approach to flexible nonlinear inference'. Unpublished manuscript, Department of Economics, University of California, San Diego.

Keenan, D. M. (1985), 'A tukey nonadditivity test for time series nonlinearity', *Biometrica* **72**, 39–44.

Lee, T.-H., White, H. & Granger, C. W. J. (1993), 'Testing for neglected nonlinearity in time series models', *Journal of Econometrics* **56**, 269–290.

Ngern, A. (1996), 'Implementation of recursive nonparametric kernel estimation and a monte carlo study on its finite sample properties', *Econometric Reviews* **15**, 69–80.

Priestley, M. B. (1980), 'State dependent models: a general approach to nonlinear time series analysis', *Journal of Time Series Analysis* **1**, 47–71.

Ramsey, J. B. (1969), 'Tests for specification errors in classical linear least square regression analysis', *Journal of the Royal Statistical Society* **Series B, 31**, 350–371.

Stock, J. H. & Watson, M. W. (1998), 'A comparison of linear and nonlinear univariate models for forecasting macroeconomic time series', *NBER, Working Paper* (6607).

Swanson, N. R. & White, H. (1995), 'A model-selection approach to assessing the information in the term structure using linear models and artificial neural networks', *Journal of Business and Economic statistics* **13**(3), 265–275.

Swanson, N. R. & White, H. (1997*a*), 'Forecasting economic time series using flexible versus fixed specification and linear versus nonlinear econometric models', *International Journal of Forecasting* (13), 439–461.

Swanson, N. R. & White, H. (1997*b*), 'A model selection approach to real-time macroeconomic forecasting using linear models and artificial neural networks', *The Review of Economics and Statistics* pp. 541–550.

Teraesvirta, T. (1990), 'Specification, estimation and evaluation of smooth transition autoregressive models'. Discussion paper 90-39, Department of Economics, University of California, San Diego.

Teraesvirta, T., Lin, C. F. & Granger, C. W. J. (1993), 'Power of the neural network linearity test', *Journal of Time Series Analysis* **14**, 209–220.

Tong, H. (1983), *Threshold models in nonlinear time series analysis*, Springer-Verlag, New York, NY.

Tsay, R. S. (1986), 'Nonlinearity test for time series', *Biometrica* **73**, 461–466.

Wahba, G. (1978), 'Improper priors, spline smoothing and the problem of guarding against model errors in regression', *Journal of the Royal Statistical Society* **Series B, 40**, 364–372.

Wecker, W. E. & Ansley, C. F. (1983), 'The signal extraction approach to nonlinear regression and spline smoothing', *Journal of the American Statistical Association* **78**, 81–89.

White, H. (1987), Specification testing in dynamic models, *in* T. Bewley, ed., 'Advances in econometrics, Fifth world congress', Cambridge University Press, Cambridge, pp. 1–58.

White, H. (1989), An additional hidden unit test for neglected nonlinearity in multilayer feedforward networks, *in* 'Proceedings of the international joint conference on neural networks', IEEE Press, New York, NY, Washington, DC, pp. 451–455.

White, H. (1992), *Estimation, Inference and Specification Analysis*, Cambridge University Press, New York, NY.

8 Tables

Table 1: Block 1 models and bivariate models

AR	$y_t = 0.6y_{t-1} + \epsilon_t$
BL	$y_t = 0.7y_{t-1}\epsilon_{t-2} + \epsilon_t$
TAR	$y_t = 0.9y_{t-1}\partial_{(\lvert y_{t-1}\rvert \leq 1)} - 0.3y_{t-1}\partial_{(\lvert y_{t-1}\rvert > 1)} + \epsilon_t$
SGN	$y_t = \partial_{(y_{t-1>1})} - \partial_{(y_{t-1<1})} + \epsilon_t$
NAR	$y_t = (0.7\lvert y_{t-1}\rvert)/(\lvert y_{t-1}\rvert + 2) + \epsilon_t$
SQ	$y_t = x_t^2 + e_t$ $x_t = 0.6x_{t-1} + \epsilon_t$ $e_t \sim N(0, \sigma^2),\ \sigma^2 = 1, 25, 400$
EXP	$y_t = \exp(x_t) + e_t$ $x_t = 0.6x_{t-1} + \epsilon_t$ $e_t \sim N(0, \sigma^2),\ \sigma^2 = 1, 25, 400$

Table 2: Block 2 models

Model1	$y_t = \epsilon_t - 0.4\epsilon_{t-1} + 0.3\epsilon_{t-2}$
Model2	$y_t = \epsilon_t - 0.4\epsilon_{t-1} + 0.3\epsilon_{t-2} + 0.5\epsilon_t\epsilon_{t-2}$
Model3	$y_t = \epsilon_t - 0.3\epsilon_{t-1} + 0.2\epsilon_{t-2} + 0.4\epsilon_{t-1}\epsilon_{t-2} - 0.25\epsilon_{t-2}^2$
Model4	$y_t = 0.4y_{t-1} - 0.3y_{t-2} + \epsilon_t$
Model5	$y_t = 0.4y_{t-1} - 0.3y_{t-2} + 0.5y_{t-1}\epsilon_{t-1} + \epsilon_t$
Model6	$y_t = 0.4y_{t-1} - 0.3y_{t-2} + 0.5y_{t-1}\epsilon_{t-1} + 0.8\epsilon_{t-1} + \epsilon_t$

Table 3: Block 3 models

LSTAR	$y_t = 1.8y_{t-1} - 1.06y_{t-2} + (0.02 - 0.9y_{t-1} + 0.795y_{t-2})F(y_{t-1}) + v_t$ $F(y_{t-1}) = [1 + \exp(-100(y_{t-1} - 0.02))]^{-1}$ $v_t \sim N(0, \sigma^2),\ \sigma^2 = 0.02^2$
ESTAR	$y_t = 1.8y_{t-1} - 1.06y_{t-2} + (-0.9y_{t-1} + 0.795y_{t-2})F(y_{t-1}) + v_t$ $F(y_{t-1}) = 1 - \exp(-4000(y_{t-1})^2)$ $v_t \sim N(0, \sigma^2),\ \sigma^2 = 0.01^2$
NN	$y_t = -1 + [1 + \exp(-100(y_{t-1} - 0.8y_{t-2}))]^{-1}$ $+[1 + \exp(-100(y_{t-1} + 0.8y_{t-2}))]^{-1} + v_t$ $v_t \sim N(0, \sigma^2),\ \sigma^2 = 0.05^2$
BN	$y_t = -1 + [1 + \exp(-100(y_{t-1} - x_t))]^{-1} + [1 + \exp(-100(y_{t-1} + x_t))]^{-1} + v_t$ $x_t = 0.8x_{t-1} + u_t$ $v_t \sim N(0, \sigma^2),\ \sigma^2 = 0.05^2, u_t \sim N(0, \sigma_u^2),\ \sigma_u^2 = 0.05^2$

Table 4: Definitions of the Reset, Tsay and V23 test used in the simulation study. M_t=$\{y_{t-1}^2, y_{t-1}y_{t-2}, y_{t-2}^2, y_{t-1}^3, y_{t-1}^2 y_{t-2}, y_{t-1}y_{t-2}^2, y_{t-2}^3\}$

	$\{p^*, p, k\}$	M_t	Dist.
block 1			
Reset2	{1,1,5}		F(1,T-3)
Tsay1	{.,1,.}	$\{\times, ., ., ., ., ., .\}$	F(1,T-3)
V23	{.,1,.}	$\{\times, ., ., \times, ., ., .\}$	F(2,T-4)
block 2+3			
Reset2	{1,2,5}		F(1,T-4)
Tsay1	{.,2,.}	$\{\times, \times, \times, ., ., ., .\}$	F(3,T-5)
V23	{.,2,.}	$\{\times, \times, \times, \times, \times, \times, \times\}$	F(7,T-10)

Table 5: Critical values (5%) based on the AR model in block 1. The first number equals the simulated 5% critical value. The number in parantheses in the second row is the asymptotic 5% critical value. The number in brackets denotes the "asymptotic" size of the statistics when based on the simulated 5% critical values (equals the area under the asymptotic distribution to the right of the simulated 5% critical value). The results are based on 10000 replications.

Test	T=50	T=100	T=200
HAMILTON	3.35	3.49	3.69
	(3.84)	(3.84)	(3.84)
	[0.067]	[0.062]	[0.055]
NEURAL1	5.40	5.48	5.66
	(5.99)	(5.99)	(5.99)
	[0.067]	[0.065]	[0.059]
TSAY1	3.10	3.39	3.65
	(4.05)	(3.94)	(3.84)
	[0.085]	[0.069]	[0.058]
WHITE3	9.32	9.39	9.16
	(9.49)	(9.49)	(9.49)
	[0.054]	[0.052]	[0.057]
RESET2	3.41	3.42	3.53
	(4.05)	(3.94)	(3.84)
	[0.071]	[0.067]	[0.062]
V23	2.73	2.75	2.84
	(3.20)	(3.09)	(3.04)
	[0.076]	[0.069]	[0.061]
BDS	8.346	5.709	4.881
	(3.84)	(3.84)	(3.84)
	[0.004]	[0.017]	[0.027]

Table 6: Critical values (5%) based on model 4 in block 2. The first number equals the simulated 5% critical value based on model 4. The number in parantheses in the second row is the asymptotic 5% critical value. The number in brackets denotes the "asymptotic" size of the statistics when based on the simulated 5% critical values (equals the area under the asymptotic distribution to the right of the simulated 5% critical value). The results are based on 10000 replications.

Test	T=50	T=100	T=200
HAMILTON	3.47	3.41	3.58
	(3.84)	(3.84)	(3.84)
	[0.062]	[0.065]	[0.058]
NEURAL1	7.59	7.58	7.75
	(7.81)	(7.81)	(7.81)
	[0.055]	[0.056]	[0.051]
TSAY1	2.65	2.56	2.64
	(2.81)	(2.70)	(2.60)
	[0.060]	[0.060]	[0.051]
WHITE3	15.23	15.31	15.35
	(15.51)	(15.51)	(15.51)
	[0.055]	[0.053]	[0.053]
RESET2	3.92	3.76	3.88
	(4.06)	(3.94)	(3.84)
	[0.054]	[0.055]	[0.050]
V23	2.14	2.01	2.03
	(2.25)	(2.12)	(2.01)
	[0.061]	[0.062]	[0.053]
BDS	9.939	6.208	4.993
	(3.84)	(3.84)	(3.84)
	[0.002]	[0.013]	[0.025]

Table 7: Size of tests and similarity. (1) Each column shows the size (%) for AR(1) $y_t = 0.6y_{t-1} + \epsilon_t$, using the 5% critical values simulated with $y_t = \phi y_{t-1} + \epsilon_t$, $\phi = -0.9, -0.6, -0.3, 0, 0.3, 0.6, 0.9$. The last column shows the size (%) for the AR(1) using 5% asymptotic critical values. (2) 95% confidence interval of the observed size is {3.6;6.4}. (3) Sample size = 100, replications = 1000.

Test	-0.9	-0.6	-0.3	0.0	0.3	0.6	0.9	Asymp.
HAMILTON	3.3	4.3	4.8	3.9	3.8	4.3	5.0	4.2
NEURAL1	3.7	5.5	4.7	4.4	4.4	4.5	4.4	4.6
TSAY1	1.9	3.1	2.9	3.5	4.3	4.7	8.0	3.4
WHITE3	4.8	6.0	5.1	5.8	5.3	4.6	5.1	4.2
RESET2	3.9	3.6	2.9	2.9	3.5	4.6	2.2	3.5
V23	3.7	4.2	4.4	3.7	4.3	4.3	3.3	3.7
BDS	4.1	4.8	5.2	3.9	5.8	4.3	5.2	12.5

Table 8: Power versus sample size for block 1 and bivariate models. Power based on 5% simulated critical values reported in table 5 is shown. Replications = 1000, sample size = 50, 100, 200.

Test	AR	BL	TAR	SGN	NAR	SQ	EXP
			HAMILTON				
T=50	5.2	12.7	63.3	75.5	8.4	100.0	96.6
T=100	4.3	19.4	93.1	98.1	11.9	100.0	99.8
T=200	3.8	24.4	99.8	100.0	22.4	100.0	100.0
			NEURAL1				
T=50	5.8	31.0	34.2	53.3	9.9	100.0	98.9
T=100	4.5	45.6	52.7	81.1	12.0	100.0	99.9
T=200	4.8	58.7	80.5	98.0	18.4	100.0	100.0
			TSAY1				
T=50	6.1	23.9	9.5	19.6	12.6	100.0	98.8
T=100	4.7	33.7	6.1	17.8	15.4	100.0	99.9
T=200	5.1	40.8	5.8	17.4	21.6	100.0	100.0
			WHITE3				
T=50	5.1	78.4	6.4	32.8	7.2	41.3	32.4
T=100	4.6	97.1	5.2	58.6	7.6	74.7	55.5
T=200	6.5	99.6	8.0	87.0	14.1	97.1	88.0
			RESET2				
T=50	6.0	24.2	28.5	13.7	8.6	86.2	72.8
T=100	4.6	33.7	48.4	10.9	12.8	95.2	77.0
T=200	5.3	42.2	71.9	12.7	18.4	99.1	80.3
			V23				
T=50	6.6	32.2	35.3	55.0	10.0	100.0	99.1
T=100	4.3	44.4	52.6	82.2	13.1	100.0	100.0
T=200	4.6	58.9	77.6	98.6	17.4	100.0	100.0
			BDS				
T=50	4.8	39.0	9.1	8.7	5.1	12.2	14.8
T=100	4.5	79.3	14.9	6.5	5.9	26.4	26.8
T=200	4.8	96.9	22.6	5.9	6.1	51.8	48.3

Table 9: Power vs. sample size for block 2 models. Power based on 5% simulated critical values reported in table 6 is shown. Replications = 1000, sample size = 50, 100, 200.

Test	Model1	Model2	Model3	Model4	Model5	Model6
			HAMILTON			
T=50	5.2	7.0	25.2	4.8	51.8	37.0
T=100	5.2	8.4	53.7	5.5	91.7	74.9
T=200	3.8	13.0	87.4	3.9	100.0	98.9
			NEURAL1			
T=50	4.4	9.2	49.9	5.5	61.0	50.8
T=100	5.9	14.4	79.5	5.8	79.1	68.7
T=200	5.3	15.8	97.2	5.2	90.5	83.6
			TSAY1			
T=50	5.1	9.8	53.5	5.0	76.2	56.9
T=100	7.0	15.9	85.5	4.9	95.0	77.8
T=200	5.9	19.0	99.3	5.5	98.7	90.8
			WHITE3			
T=50	5.2	13.2	29.8	3.9	81.4	71.7
T=100	14.2	18.6	58.1	5.8	99.4	93.8
T=200	24.7	30.8	90.7	5.2	100.0	99.5
			RESET2			
T=50	6.1	8.8	14.9	3.7	21.4	40.5
T=100	6.0	10.5	21.4	6.4	37.4	52.3
T=200	4.4	11.9	29.0	5.5	60.3	65.6
			V23			
T=50	4.5	13.8	43.5	4.2	82.9	72.9
T=100	6.0	19.8	77.8	6.4	99.0	93.0
T=200	5.3	21.2	97.8	4.7	100.0	99.4
			BDS			
T=50	4.4	2.1	7.6	4.0	29.4	31.9
T=100	5.4	3.0	15.8	5.0	77.8	72.4
T=200	4.5	4.4	25.7	4.4	98.6	96.0

Table 10: Power vs. sample size for block 3 models. Power based on 5% simulated critical values reported in table 6 is shown. Replications = 1000, sample size = 50, 100, 200.

	LSTAR	ESTAR	NN	BN
		HAMILTON		
T=50	33.8	20.7	61.3	61.0
T=100	71.0	47.2	96.6	96.5
T=200	98.7	82.2	100.0	100.0
		NEURAL1		
T=50	56.6	9.2	45.3	45.2
T=100	85.7	11.2	76.0	71.7
T=200	95.6	13.5	93.6	88.3
		TSAY1		
T=50	65.0	9.4	9.6	17.6
T=100	93.8	9.7	11.5	15.9
T=200	100.1	8.9	13.6	15.4
		WHITE3		
T=50	33.2	22.8	13.2	18.2
T=100	69.3	45.7	21.2	30.3
T=200	95.9	74.2	41.3	60.9
		RESET2		
T=50	33.1	44.0	5.4	11.4
T=100	42.4	76.5	6.8	7.7
T=200	63.1	93.5	6.4	8.2
		V23		
T=50	63.1	32.4	47.0	49.0
T=100	93.0	65.2	88.0	86.7
T=200	100.0	92.8	99.5	99.7
		BDS		
T=50	0.0	0.0	1.9	2.1
T=100	0.0	0.0	4.8	4.0
T=200	0.0	0.0	5.6	6.4

Table 11: Power vs. sample size and noise for bivariate model (SQ). Power based on 5% simulated critical values reported in table 5 is shown. The signal-to-noise ratio equals 700% for σ = 1, 28% for σ= 5 and 2% for σ= 20. 1000 replications. Sample size = 50, 100, 200.

	σ =1	σ =5	σ =20
	HAMILTON		
T=50	100.0	40.7	7.3
T=100	100.0	75.4	9.2
T=200	100.0	96.7	14.7
	NEURAL1		
T=50	100.0	62.9	12.1
T=100	100.0	90.2	16.1
T=200	100.0	99.6	27.0
	TSAY1		
T=50	100.0	75.1	14.1
T=100	100.0	94.7	19.5
T=200	100.0	99.8	33.8
	WHITE3		
T=50	41.3	8.3	4.2
T=100	74.7	14.3	4.4
T=200	97.1	27.9	5.9
	RESET2		
T=50	86.2	40.9	37.1
T=100	95.2	64.5	62.9
T=200	99.1	86.1	85.3
	V23		
T=50	100.0	63.7	11.0
T=100	100.0	92.3	17.6
T=200	100.0	99.7	26.5
	BDS		
T=50	12.2	38.3	81.4
T=100	26.4	46.2	73.9
T=200	51.8	42.3	74.2

Table 12: Power vs. sample size and noise for bivariate model (EXP). Power based on 5% simulated critical values reported in table 5 is shown. The signal-to-noise ratio equals 216% for σ = 1, 8.6% for σ= 5 and 0.5% for σ= 20. 1000 replications. Sample size = 50, 100, 200.

	σ =1	σ =5	σ =20
	HAMILTON		
T=50	96.6	37.5	8.8
T=100	99.8	60.9	10.4
T=200	100.0	87.8	15.7
	NEURAL1		
T=50	98.9	56.2	14.5
T=100	99.9	81.5	22.8
T=200	100.0	97.9	35.4
	TSAY1		
T=50	99.1	59.2	15.1
T=100	100.0	82.7	22.5
T=200	100.0	97.6	35.3
	WHITE3		
T=50	32.4	10.3	5.7
T=100	55.5	15.1	5.7
T=200	88.0	28.4	7.3
	RESET2		
T=50	72.8	28.7	10.8
T=100	77.0	40.8	15.1
T=200	80.3	49.6	21.9
	V23		
T=50	99.1	55.1	15.6
T=100	100.0	83.2	23.8
T=200	100.0	98.5	36.7
	BDS		
T=50	14.8	37.5	68.7
T=100	26.8	43.7	64.9
T=200	48.3	40.5	70.3

Table 13: The regressors included in $\widetilde{x}_{t,n}$ for the various models under consideration.

Model	Regressors, $\widetilde{x}_{t,n}$	Model	Regressors, $\widetilde{x}_{t,n}$
AR	$1, y_{t-1,n}$	Model1	$1, \epsilon_{t-1,n}, \epsilon_{t-2,n}$
BL	$1, y_{t-1,n}\epsilon_{t-2,n}$	Model2	$1, \epsilon_{t-1,n}, \epsilon_{t-2,n}$
TAR	$1, y_{t-1,n}\partial_{(\lvert y_{t-1,n}\rvert \leq 1)}, y_{t-1,n}\partial_{(\lvert y_{t-1,n}\rvert > 1)}$	Model3	$1, \epsilon_{t-1,n}, \epsilon_{t-2,n}, \epsilon_{t-1,n}\epsilon_{t-2,n}, \epsilon^2_{t-2,n}$
SGN	$1, [\partial_{(y_{t-1,n>1})} - \partial_{(y_{t-1,n<1})}]$	Model4	$1, y_{t-1,n}, y_{t-2,n}$
NAR	$1, (0.7\lvert y_{t-1,n}\rvert)/(\lvert y_{t-1,n}\rvert + 2)$	Model5	$1, y_{t-1,n}, y_{t-2,n}, y_{t-1,n}\epsilon_{t-1,n}$
SQ	$1, x^2_{t,n}$	Model6	$1, y_{t-1,n}, y_{t-2,n}, y_{t-1,n}\epsilon_{t-1,n}, \epsilon_{t-1,n}$
EXP	$1, \exp(x_{t,n})$		
LSTAR	$1, F(y_{t-1,n})$, $y_{t-1,n}, y_{t-1,n} * F(y_{t-1,n})$, $y_{t-2,n}, y_{t-2,n} * F(y_{t-1,n})$	NN	1, $[1+\exp(-100(y_{t-1} - 0.8y_{t-2}))]^{-1}$, $[1+\exp(-100(y_{t-1} + 0.8y_{t-2}))]^{-1}$
ESTAR	$1, F(y_{t-1,n})$, $y_{t-1,n}, y_{t-1,n} * F(y_{t-1,n})$, $y_{t-2,n}, y_{t-2,n} * F(y_{t-1,n})$	BN	1, $[1+\exp(-100(y_{t-1} - x_t))]^{-1}$, $[1+\exp(-100(y_{t-1} + x_t))]^{-1}$

Table 14: Convergence properties of block 1 and bivariate models. $C^1_{T,N}$, $C^2_{T,N}$, $C^3_{T,N}$, $C^4_{T,N}$ and $C^5_{T,N}$ are calculated according to the their definitions stated above where $\{\widehat{\alpha}_{T,n}, \widehat{\zeta}_T(.), \rho(., \widehat{\kappa}_{T,n}(\widehat{q}))\}$ and $\{\widehat{\omega}_{T,n}\}$ are estimated conditionally on a pair of $\{y^*_{t,n}, x^*_{t,n}\}$ and $\{y^*_{t,n}, \widetilde{x}^*_{t,n}\}$ that differs nummerically from $\{y_{t,n}, x_{t,n}\}$ and $\{y_{t,n}, \widetilde{x}_{t,n}\}$ but is generated from same underlying process/model. The number of replications equals N=100.

	$C^1_{T,N}$	$C^2_{T,N}$	$G^2_{T,N}$	$C^3_{T,N}$	$G^3_{T,N}$	$C^4_{T,N}$	$G^4_{T,N}$	$C^5_{T,N}$
				AR				
T=50	0.052	0.053	-0.025	0.117	-1.238	0.075	-0.451	0.052
T=100	0.029	0.030	-0.024	0.088	-2.047	0.043	-0.498	0.029
T=200	0.015	0.016	-0.107	0.059	-3.023	0.022	-0.490	0.015
				BL				
T=50	2.439	2.642	-0.083	2.650	-0.087	2.568	-0.053	0.095
T=100	2.202	2.235	-0.106	2.219	-0.099	2.275	-0.033	0.027
T=200	1.851	1.968	-0.063	2.023	-0.093	1.903	-0.028	0.018
				TAR				
T=50	0.276	0.181	0.345	0.157	0.430	0.262	0.052	0.059
T=100	0.258	0.112	0.564	0.116	0.551	0.251	0.027	0.029
T=200	0.245	0.073	0.701	0.082	0.666	0.205	0.161	0.013
				SGN				
T=50	0.368	0.213	0.422	0.203	0.450	0.307	0.166	0.047
T=100	0.344	0.145	0.577	0.143	0.585	0.246	0.286	0.021
T=200	0.333	0.100	0.700	0.095	0.713	0.219	0.342	0.012
				NAR				
T=50	0.057	0.061	-0.084	0.121	-1.129	0.072	-0.258	0.046
T=100	0.040	0.042	-0.054	0.084	-1.091	0.046	-0.157	0.022
T=200	0.030	0.029	0.046	0.059	-0.980	0.035	-0.161	0.011
				SQ				
T=50	5.362	2.044	0.619	2.374	0.557	3.463	0.354	0.045
T=100	5.644	1.216	0.785	1.578	0.720	3.086	0.453	0.027
T=200	5.251	0.680	0.871	0.916	0.826	2.892	0.449	0.012
				EXP				
T=50	13.220	8.506	0.357	9.250	0.300	9.271	0.299	0.118
T=100	9.858	4.755	0.518	5.369	0.455	6.667	0.324	0.035
T=200	9.502	2.882	0.697	3.743	0.606	6.384	0.328	0.013

Table 15: Convergence properties of block 2 models. $C^1_{T,N}$, $C^2_{T,N}$, $C^3_{T,N}$, $C^4_{T,N}$ and $C^5_{T,N}$ are calculated according to the their definitions stated above where $\{\widehat{\alpha}_{T,n}, \widehat{\zeta}_T(.), \rho(., \widehat{\kappa}_{T,n}(\widehat{q}))\}$ and $\{\widehat{\omega}_{T,n}\}$ are estimated conditionally on a pair of $\{y^*_{t,n}, x^*_{t,n}\}$ and $\{y^*_{t,n}, \widetilde{x}^*_{t,n}\}$ that differs nummerically from $\{y_{t,n}, x_{t,n}\}$ and $\{y_{t,n}, \widetilde{x}_{t,n}\}$ but is generated from same underlying process/model. The number of replications equals N=100.

	$C^1_{T,N}$	$C^2_{T,N}$	$G^2_{T,N}$	$C^3_{T,N}$	$G^3_{T,N}$	$C^4_{T,N}$	$G^4_{T,N}$	$C^5_{T,N}$
				Model1				
T=50	0.113	0.126	-0.114	0.209	-0.855	0.159	-0.408	0.069
T=100	0.062	0.067	-0.084	0.152	-1.458	0.092	-0.491	0.031
T=200	0.047	0.048	-0.011	0.120	-1.538	0.059	-0.260	0.017
				Model2				
T=50	0.393	0.445	-0.132	0.515	0.309	0.478	-0.215	0.334
T=100	0.352	0.374	-0.063	0.461	-0.309	0.386	-0.097	0.291
T=200	0.321	0.337	-0.048	0.415	-0.290	0.339	-0.058	0.259
				Model3				
T=50	0.384	0.377	0.018	0.386	-0.004	0.399	-0.040	0.104
T=100	0.386	0.344	0.109	0.351	0.090	0.378	0.022	0.061
T=200	0.331	0.241	0.272	0.269	0.186	0.283	0.144	0.026
				Model4				
T=50	0.061	0.074	-0.208	0.156	-1.555	0.089	-0.464	0.061
T=100	0.039	0.044	-0.131	0.125	-2.172	0.069	-0.774	0.039
T=200	0.016	0.019	-0.206	0.092	-4.828	0.023	-0.415	0.016
				Model5				
T=50	1.343	1.271	0.054	1.175	0.125	1.295	0.036	0.129
T=100	1.018	0.820	0.195	0.773	0.240	0.869	0.147	0.046
T=200	1.001	0.746	0.254	0.698	0.302	0.881	0.120	0.023
				Model6				
T=50	2.773	2.877	-0.038	2.614	0.057	2.857	-0.030	0.144
T=100	2.728	2.715	0.005	2.499	0.084	2.849	-0.045	0.060
T=200	2.321	2.190	0.056	2.035	0.123	2.259	0.027	0.025

Table 16: Convergence properties of block 3 models. $C^1_{T,N}$, $C^2_{T,N}$, $C^3_{T,N}$, $C^4_{T,N}$ and $C^5_{T,N}$ are calculated according to the their definitions stated above where $\{\widehat{\alpha}_{T,n}, \widehat{\zeta}_T(.), \rho(., \widehat{\kappa}_{T,n}(\widehat{q}))\}$ and $\{\widehat{\omega}_{T,n}\}$ are estimated conditionally on a pair of $\{y^*_{t,n}, x^*_{t,n}\}$ and $\{y^*_{t,n}, \widetilde{x}^*_{t,n}\}$ that differs nummerically from $\{y_{t,n}, x_{t,n}\}$ and $\{y_{t,n}, \widetilde{x}_{t,n}\}$ but is generated from same underlying process/model. The number of replications equals N=100.

	$C^1_{T,N}$	$C^2_{T,N}$	$G^2_{T,N}$	$C^3_{T,N}$	$G^3_{T,N}$	$C^4_{T,N}$	$G^4_{T,N}$	$C^5_{T,N}$
				LSTAR				
T=50	1.754	1.627	0.072	1.271	0.275	2.128	-0.213	0.384
T=100	1.518	1.018	0.329	0.828	0.454	1.503	0.010	0.173
T=200	1.488	0.590	0.592	0.553	0.618	1.311	0.119	0.064
				ESTAR				
T=50	0.231	0.236	-0.022	0.227	0.015	0.311	-0.345	0.055
T=100	0.203	0.172	0.153	0.158	0.224	0.235	-0.159	0.034
T=200	0.183	0.100	0.451	0.113	0.384	0.176	0.036	0.017
				NN				
T=50	0.182	0.151	0.168	0.128	0.296	0.126	0.308	0.018
T=100	0.172	0.103	0.402	0.095	0.446	0.091	0.470	0.007
T=200	0.167	0.076	0.544	0.073	0.561	0.071	0.575	0.004
				BN				
T=50	0.220	0.182	0.173	0.157	0.287	0.149	0.323	0.027
T=100	0.159	0.104	0.347	0.096	0.398	0.094	0.411	0.009
T=200	0.151	0.070	0.536	0.069	0.545	0.069	0.542	0.005

9 Appendix A.

Table 17: Closed form expressions for $H_k(h)$. $H_k(h)$ equals unity when $h = 0$, and equals zero when $h \geq 1$

k	$H_k(h)$
1	$1 - h$
2	$1 - \frac{2}{\pi}[h(1-h^2)^{\frac{1}{2}} + \sin^{-1}(h)]$
3	$1 - (\frac{3h}{2}) + (\frac{h^3}{2})$
4	$1 - \frac{2}{\pi}[\frac{2}{3}h(1-h^2)^{\frac{3}{2}} + h(1-h^2)^{\frac{1}{2}} + \sin^{-1}(h)]$
5	$1 - \frac{3}{2}h + \frac{1}{2}h^3 - \frac{3h}{8}(1-h^2)^2$

Working Paper

1998-15	Henning Bunzel, Bent Jesper Christensen, Niels Haldrup, Svend Hylleberg, Viggo Høst, Peter Jensen og Allan Würtz: Udviklingslinier i Økonometrien.
1998-16	Peter Skott: Wage Formation and the (Non-)Existence of the NAIRU.
1998-17	Svend Hylleberg and Per Baltzer Overgaard: Competition Policy with a Coasian Prior?
1998-18	Lykke Eg Andersen and Osvaldo Nina: Micro-Credit and Group Lending: The Collateral Effect.
1998-19	Torben M. Andersen and Svend Hylleberg: Sources of Persistence in Employment Adjustment - Denmark 1974-1993.
1999-1	Bo Sandemann Rasmussen: Balancing the Budget: Long-Run Adverse Effects of Progressive Taxation.
1999-2	Palle Schelde Andersen, Mark Klau, and Ebbe Yndgaard: Higher Profits and Lower Capital Prices: Is Factor Allocation Optimal?
1999-3	Niels Haldrup and Michael Jansson: Spurious Regression, Cointegration, and Near Cointegration: A Unifying Approach.
1999-4	Christian M. Dahl and Svend Hylleberg: Specifying Nonlinear Econometric Models by Flexible Regression Models and Relative Forecast Performance.
1999-5	David L. Kelly and Jamsheed Shorish: Stability of Functional Rational Expectations Equilibria.
1999-6	Torben M. Andersen and Eric Toulemonde: Adapting Prices or Quantities in the Presence of Adjustment Costs?
1999-7	Torben M. Andersen and Eric Toulemonde: Imperfectively Competitive Labour Markets and the Productivity Puzzle.
1999-8	Christian M. Dahl: An Investigation of Tests for Linearity and the Accurracy of Flexible Nonlinear Inference.

CENTRE FOR NON-LINEAR MODELLING IN ECONOMICS

DEPARTMENT OF ECONOMICS - UNIVERSITY OF AARHUS - DK - 8000 AARHUS C - DENMARK
☎ +45 89 42 11 33 - TELEFAX +45 86 13 63 34

Working papers, issued by the Centre for Non-linear Modelling in Economics:

1996-4	Morten O. Ravn and Martin Sola: A Reconsideration of the Empirical Evidence on the Asymmetric Effects of Money-Supply shocks: Positive vs. Negative or Big vs. Small?
1996-13	Robert F. Engle and Svend Hylleberg: Common Seasonal Features: Global Unemployment.
1996-14	Svend Hylleberg and Adrian R. Pagan: Seasonal Integration and the Evolving Seasonals Model.
1997-1	Tom Engsted, Jesus Gonzalo and Niels Haldrup: Testing for Multi-cointegration.
1997-7	Luca Fanelli: Estimating Multi-Equational LQAC Models with I(1) Variables: a VAR Approach.
1997-12	Niels Haldrup: A Review of the Econometric Analysis of I(2) Variables.
1997-14	Viggo Høst: Better Confidence Intervals for the Popoulation Mean by Using Trimmed Means and the Iterated Bootstrap?
1997-17	N.E. Savin and Allan H. Würtz: The Effect of Nuisance Parameters on Size and Power; LM Tests in Logit Models.
1997-18	Tom Engsted and Niels Haldrup: Multicointegration in Stock-Flow Models.
1998-6	Svend Hylleberg and Rikke Willemoes Jørgensen: A Note on the Estimation of Markup Pricing in Manufacturing.
1998-15	Henning Bunzel, Bent Jesper Christensen, Niels Haldrup, Svend Hylleberg, Viggo Høst, Peter Jensen og Allan Würtz: Udviklingslinier i Økonometrien.
1999-3	Niels Haldrup and Michael Jansson: Spurious Regression, Cointegration, and Near Cointegration: A Unifying Approach.
1999-4	Christian M. Dahl and Svend Hylleberg: Specifying Nonlinear Econometric Models by Flexible Regression Models and Relative Forecast Performance.
1999-8	Christian M. Dahl: An Investigation of Tests for Linearity and the Accuracy of Flexible Nonlinear Inference.