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

Christian M. Dahl*

Department of Economics

University of Aarhus

Denmark

July 1998

(Very preliminary version, comments are welcome)

Abstract

A new approach recently suggested by Hamilton for flexible parametric inference in nonlinear models is examined through simulation studies.

* Comments and suggestions from James D. Hamilton are gratefully acknowledged. Futhermore, I would like to thank participants at the Econometrics Workshop at UCSD for helpful comments. Any errors are my own.

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 and the so-called V23 test. With respect to size and power properties our 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 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 least square estimator can be substantial.

1. Introduction

Hamilton (1998) 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 on the size and

power properties of the of the new Lagrange multiplier test for neglected nonlinearity suggested by Hamilton (1998). We compare the test with other popular tests for neglected nonlinearity, tests which - as it is the case for Hamilton's test - are not based on any knowledge of the functional form under the alternative. Secondly we conduct a Monte Carlo experiment on the convergence properties of Hamilton's estimator of the conditional mean function $\mu(x_t)$. We report results on how it performs by applying it to a wide range of the most common nonlinear models in the literature, and provide some statistics on how big the improvement is relative to the linear estimator. 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. 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 $\mu(x_t)$ and derive the Lagrange multiplier test statistic for neglected nonlinearity. Section 3 discusses some of the most popular alternative tests for neglected nonlinearity. Section 4 describes the simulation design for the Monte Carlo experiment on size and power and in section 5 the results are reported and discussed. In section 6 the experiment on the convergence properties of Hamilton's estimator are conducted and finally section 7 contains concluding remarks.

2. Hamilton's approach to flexible nonlinear inference

Consider the model

$$y_t = \mu(x_t) + \epsilon_t \tag{2.1}$$

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 (1998) suggests representing $\mu(x_t)$ as[1]

$$\mu(x_t) = \alpha_0 + \alpha_1' x_t + \lambda m(g \odot x_t) \tag{2.2}$$

[1] Here g is a $k \times 1$ vector of parameters and $\odot$ denotes element-by-element multiplication.

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

$$\begin{aligned} m(z) &\sim N(0,1) \\ E(m(z)'m(w)) &= H_k(h) \end{aligned} \tag{2.3}$$

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

$$G_k(h,r) = \int_h^r (r^2 - z^2)^{\frac{k}{2}} dz \tag{2.4}$$

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

$$H_k(h) = \begin{cases} G_{k-1}(h,1)/G_{k-1}(0,1) & \text{if } h \leq 1 \\ 0 & \text{if } h > 1 \end{cases} \tag{2.5}$$

Closed form expressions for $H_k(h)$ for $k = \{1, .., 5\}$ are provided by Hamilton (1998) and are reprinted in the appendix of this paper. Since we cannot observe $m(z)$ - for any choice of z - directly 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 (1998) 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

$$\begin{aligned} y &= (y_1, y_2, .., y_T) \\ X &= \begin{bmatrix} 1 & x_1' \\ 1 & x_2' \\ . & . \\ . & . \\ 1 & x_T' \end{bmatrix} \\ \beta &= (\alpha_0, \alpha') \\ \epsilon &= (\epsilon_1, \epsilon_2, .., \epsilon_T) \end{aligned} \tag{2.6}$$

and shows that the concentrated log likelihood function can be written as

$$\begin{aligned}\eta(y, X; \lambda, g, \sigma) &= -\frac{T}{2}\ln(2\pi) - \frac{1}{2}\ln|P_0 + \sigma^2 I_T| \\ &\quad -\frac{1}{2}[y - X\widehat{\beta}_T(\lambda, g, \sigma)]'(P_0 + \sigma^2 I_T)^{-1}[y - X\widehat{\beta}_T(\lambda, g, \sigma)] \qquad (2.7)\\ \widehat{\beta}_T(\lambda, g, \sigma) &= [X'(P_0 + \sigma^2 I_T)^{-1}X]^{-1}[X'(P_0 + \sigma^2 I_T)^{-1}y] \qquad (2.8)\end{aligned}$$

where I_T is the identity matrix of dimension $(T \times T)$ and the $\{t, s\}$ entry of the matrix $\mathrm{P_0}$ - denoted $\mathrm{P_0}(t, s)$ - is equal to

$$\begin{aligned}P_0(t, s) &= \begin{cases} \lambda^2 H_k(h_{ts}) & \text{if } h_{ts} \leq 1 \\ 0 & \text{if } h_{ts} > 1 \end{cases} \qquad (2.9)\\ h_{ts} &= \frac{1}{2}[(\widetilde{x}_t - \widetilde{x}_s)'(\widetilde{x}_t - \widetilde{x}_s)]^{\frac{1}{2}} \\ \widetilde{x}_t &= g \odot x_t\end{aligned}$$

The concentrated likelihood function can be maximized with respect to λ, g, σ using standard maximization algorithms such as $BFGS$ or $Newton - Raphson$. Once the estimates of (λ, g, σ) has been obtained, $\widehat{\beta}_T$ is given. In the case of continuous valued variables and deterministic regressors Hamilton (1998) proves that if the true relation given by (2.2) is indeed linear then under some regularity conditions the estimator of $\mu(x_t)$ given by the t'th row of $\widehat{\xi}_T = X\widehat{\beta}_T + \widehat{P}_0(\widehat{P}_0 +$

$\widehat{\sigma}_T^2 I_T)^{-1}[y - X\widehat{\beta}_T]$ - denoted $\widehat{\xi}_T(x_t)$ - is still a consistent estimator of the conditional mean, implying that $\widehat{\beta}_T$ is a consistent estimator. Furthermore, Hamilton (1998) 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

$$T^{-1}\sum_{t=1}^{T}\{\mu(x_t) - \widehat{\xi}_T(x_t)\}^2 \to 0 \tag{2.10}$$

Testing for neglected nonlinearity in this setup amounts to testing whether $\lambda = 0$. When λ equals zero, g is not identified by the model under the null. Hamilton (1998) suggests solving this problem by fixing g_i to be proportional to the standard deviation of the i'th row in x_t. Under this assumption the Lagrange multiplier statistic for neglected nonlinearity becomes

$$v^2 = \frac{[\widehat{\epsilon}' H \widehat{\epsilon} - \widetilde{\sigma}^2 tr(MHM)]^2}{\widetilde{\sigma}^4[2tr\{[MHM - (T-k)^{-1}Mtr(MHM)]^2\}]} \tag{2.11}$$

where

$$\begin{aligned} \widehat{\epsilon} &= My \\ \widetilde{\sigma}^2 &= (T-k)^{-1}\widehat{\epsilon}'\widehat{\epsilon} \end{aligned} \tag{2.12}$$

$$M = I_T - X(X'X)^{-1}X'$$

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

$$\begin{aligned} H(t,s) &= \begin{cases} H_k(h_{ts}) \text{ if } h_{ts} \leq 1 \\ 0 \quad \text{if } h_{ts} > 1 \end{cases} \\ h_{ts} &= \frac{1}{2}[k^{-1}\sum_{i=1}^{k}\frac{(x_{i,t}-x_{i,s})^2}{s_i^2}]^{\frac{1}{2}} \\ s_i^2 &= T^{-1}\sum_{t=1}^{T}(x_{i,t}-T^{-1}\sum_{t=1}^{T}x_{i,t})^2 \end{aligned} \tag{2.13}$$

As mentioned earlier the closed form expressions for H_k, $k = \{1,..,5\}$ is given in the appendix. The Lagrange multiplier statistic 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 ($Reset2$, Ramsey 1969) two tests based on the "duals" of Volterra expansions e.g Priestley (1980), denoted the $Tsay1$ test (Tsay 1986) and the $V23$ test (Terasvirta *et al.* 1993) respectively, the Neural Network test ($Neural1$,Lee *et al.* 1993) and finally a particular version of White's information matrix test ($White3$, White 1987,1992) aimed at detecting dynamic misspecification.

3.1. The Reset2, Tsay1 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

$$y_t = x_t'\beta + u_t \tag{3.1}$$

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 $\widehat{v}_t = \widehat{u}_t - x_t'\widehat{\alpha}_1 - M_t'\widehat{\alpha}_2$ and

the residual sum of squares $SSR = \sum_{t=1}^{T} \widehat{v}_t^2$. Finally in the third step compute the F-statistic given by

$$F = \frac{(SSR_0 - SSR)/m}{SSR/(T-p-1-m)} \sim F(m, T-p-1-m) \tag{3.2}$$

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

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 sends 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) takes 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 $\theta = \{\theta_0, \theta_1, .., \theta_k\}$ and assuming that the output also contains white noise the total network output can be written as

$$y_t = x_t'\theta + \sum_{j=1}^{q} \beta_j (1 + \exp(-x_t'\gamma_j))^{-1} + \epsilon_t \tag{3.3}$$

where $\beta_1, .., \beta_q$ are hidden-units-to-output weights and $\epsilon_t \sim nid(0, \sigma^2)$. When the null hypothesis of linearity is true i.e. $\mathrm{H_0} : \Pr[E(y_t|X_t) = x_t'\theta^*] = 1$ for some choices of θ^* and $X_t = \{x_1', x_2', .., x_t'\}$ the optimal network weights β_j are zero for $j = 1, .., q$. The neural network test for neglected nonlinearity can therefore be interpreted as testing the hypothesis $\mathrm{H_0} : \beta_1 = \beta_2 = .. = \beta_q = 0$ 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 unadjusted squared multiple correlation coefficient R^2. The LM-test statistic and its asymptotic distribution is given by

$$T * R^2 \rightarrow \chi^2(q) \tag{3.4}$$

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 statistic is given by

$$T * R_{pc}^2 \rightarrow \chi^2(q^*) \tag{3.5}$$

where R_{pc}^2 is the unadjusted 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

$$l_t(x_t, \theta, \sigma) = -\frac{1}{2}\log(2\pi) - \log(\sigma) - \frac{1}{2}u_t^2 \tag{3.6}$$

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

$$s_t(x_t, \theta, \sigma) = \sigma^{-1}(u_t, u_t x_t', u_t^2 - 1)' \tag{3.7}$$

Evaluating the conditional score at the quasi maximum likelihood estimators of the correctly specified model under $\mathrm{H_0}$ gives $\widehat{s}_t = s_t(x_t, \widehat{\theta}, \widehat{\sigma})$. The information

matrix test is based on forming the indicator $\widehat{m}_t = S * vec(\widehat{s}_t\widehat{s}_t')$ where S is a selection matrix. In particular we obtain the test statistic denoted "White3" in Lee *et al.*(1993) by regressing $\widehat{u}_t = \widehat{\sigma}^{-1}(y_t - x_t'\widehat{\theta})$ on x_t and $\widehat{k}_t$ - where $\widehat{k}_t$ is defined as to satisfy $\widehat{m}_t = \widehat{k}_t\widehat{u}_t'$ - and calculate the unadjusted squared multiple correlation coefficient R^2 from this regression. The test statistic and its asymptotic distribution is then given by

$$T * R^2 \rightarrow \chi^2(q) \tag{3.8}$$

where q denotes the dimension of m_t.

4. The design of the Monte Carlo experiment

We consider three blocks of models. All the chosen models have been used in previous studies on the testing of linearity and are included to allow for comparisons with these studies. The models included in *block* 1 and two of the bivariate models was originally used by Lee *et al.* (1993). The models of *block* 2 have been more extensively used, 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 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 Andersen (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 4.1. We also consider two bivariate representations. In order to simplify 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 are also shown in table 4.1.

[*Table* 4.1]

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 4.2. *Model*

1 is a $MA(2)$ 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. *Model* 3, 5 and 6 are the truly nonlinear models in *block* 2. *Model* 3 is a nonlinear $MA(2)$ model. *Model* 5 and *model* 6 belongs to the family of bilinear models. *Model* 5 is a bilinear autoregressive model, while *model* 6 is a bilinear autoregressive moving average model.

[*Table* 4.2]

Terasvirta *et al.* (1993) argues 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 test. By the appropriate LM type or LM type test Terasvirta *et al.* (1993) refer to a LM-test designed particular to test linearity against a specific nonlinear alternative. Now, the simulation design in Lee *et al.* (1993) is only concerned with evaluating linearity test where the alternative don't

have to be specified. However the critique raised by Terasvirta *et al.* (1993) is still relevant in the sense that the choice of testor-model-mix may undeliberate have been favoring 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 don't need to be explicitly specified. The first model in *block* 3 is the logistic smooth transition autoregressive model ($LSTAR$). It's 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 denotes univariate and bivariate neural network models respectively.

[*Table* 4.3]

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 4.4. For the neural network test $q^* = 2$ for the *block* 1 models and the bivariate models. When applied to the models in *block* $2+3$ $q^* = 3$.

[*Table* 4.4]

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 Lee *et al.*(1993). The results from a simulation of the critical values at a 5% level are shown in table 5.1. 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 5.1 seems good - in the sense that the simulated size seems to correspond well with the nominal/asymptotic size - when simulations

are based on the $AR(1)$ model.

[*Table* 5.1]

The results on size properties does 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, and in general this result holds for the other test statistics considered as well, see table 5.2.

[*Table* 5.2]

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. 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 in *block* 1 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 an $AR(1)$ 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 5.3 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 relative low spread in size, ranging from 3.3 to 5.0

[*Table* 5.3]

Based on the simulated critical values reported in table 5.1, table 5.4 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 has good power properties.

[*Table* 5.4]

Table 5.5 shows the power properties of the test based on *block* 2 models and the simulated critical values reported in table 5.2. By inspection of table 5.5 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 appears to be good compared to the $Neural1$ test. Also the $Tsay1$, the $White3$ and the $V23$ tests seems 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 causes the size of the neural network test, the Tsay, White, Reset and V23 test to be incorrect. By inspection of table 5.5 and the results based on model 2 it becomes clear that this particular feature

also seems to be shared by Hamilton's test.

[*Table* 5.5]

Considering the results on power of the various test statistics when applied to the models of *block* 3, Hamilton's tests again seems to perform relatively well. In the case of the neural network models the neural network test and the $V23$ test turns 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 test considered has very good power. If the nonlinearity is of the $ESTAR$ type only Hamilton's test, $White3$, $Reset2$ and the $V23$ test has satisfactory power. Against the $ESTAR$ type of nonlinearity the neural network test has very low power. The results on the *block* 3 models again seems to confirm that Hamilton's test have 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* 5.6]

Table 5.7 and table 5.8 shows 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 do 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.

[$Table$ 5.7]

[$Table$ 5.8]

6. Convergence properties of Hamilton's estimator

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

$$\begin{aligned} C^1_{T,N} &= N^{-1}\sum_{n=1}^{N}[T^{-1}\sum_{t=1}^{T}\{\mu(x_{t,n}) - \widehat{\alpha}'_{T,n}x_{t,n}\}^2] \\ C^2_{T,N} &= N^{-1}\sum_{n=1}^{N}[T^{-1}\sum_{t=1}^{T}\{\mu(x_{t,n}) - \widehat{\xi}_T(x_{t,n})\}^2] \end{aligned} \tag{6.1}$$

$$C^3_{T,N} = N^{-1}\sum_{n=1}^{N}[T^{-1}\sum_{t=1}^{T}\{\mu(x_{t,n}) - \xi_T(x_{t,n})|_{\lambda=\sigma,g=2(kV(x_n))^{-1/2}}\}^2]$$

$$C^4_{T,N} = N^{-1}\sum_{n=1}^{N}[T^{-1}\sum_{t=1}^{T}\{\mu(x_{t,n}) - \widehat{\omega}'_{T,n}\widetilde{x}_{t,n}\}^2]$$

Common for all four measures is that $\mu(x_{t,n})$ denote 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 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 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 (2.7) and equation (2.8) with respect to λ, g and σ using the CML procedure and the $BFGS$ algorithm in $GAUSS$[2]. If Hamilton's estimator improves on the linear 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. In order to measure how much Hamilton's estimator

[2] With a maximum of 10 iterations.

improves on a linear estimator in finite samples we consider two statistics, denoted $G^i_{T,N}$ and γ^j respectively. $G^i_{T,N}$ is defined as

$$G^i_{T,N} = 1 - \frac{C^i_{T,N}}{C^1_{T,N}} \ , \quad i = 2, 3 \tag{6.2}$$

and is a measure of the percentage improvement in fit of Hamilton's estimator(s) relative to the linear estimator. As suggested by Ngern (1996) we calculate a measure of the rate of convergence of $C^n_{T,N}$ by assuming that

$$C^j_{T,N} = \kappa_{n,N} T^{-\gamma^j}, j = \{1, .., 4\} \tag{6.3}$$

where $\kappa_{n,N}$ is a constant independent of T, and γ^j for $j = \{1, .., 4\}$ is the measure of convergence. $C^3_{T,N}$ 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 equals 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. Finally we calculate $C^4_{T,N}$ defined as the average mean square error between $\mu(x_{t,n})$ and a 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 $\widetilde{x}_{t,n}$

is defined in table 6.1 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 $C^4_{T,N}$ 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 $C^i_{T,N}$, $i = 1,..,4$ 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 $\{\widehat{\alpha}_{T,n}, \widehat{\omega}_{T,n}, \widehat{\xi}_{T,n}, \xi_{T,n}\}$.

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 $C^i_{T,N}$, $i = 1,..,4$ based on $\mu(x_{t,n})$, $\widehat{\alpha}'_{T,n}x_{t,n}$, $\widehat{\omega}'_{T,n}\widetilde{x}_{t,n}$, $\widehat{\xi}_T(x_{t,n})$ and $\xi_T(x_{t,n})|_{\lambda=\sigma,g=2(kV(x_n))^{-1/2}}$ 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 turns out to be very high by this

approach.

[*Table* 6.1]

Table 6.2, table 6.3 and table 6.4 report the estimates of $C^1_{T,N}$, $C^2_{T,N}$, $C^3_{T,N}$ and $C^4_{T,N}$ together with the statistics $G^i_{T,N}$, γ^j for $i = 2, 3$ and $j = 1..4$ based on a "small" Monte Carlo experiment with $N = 100$. Looking at the Monte Carlo results for the linear AR model it is evident that C^1_T and C^2_T numerically are approximately of the same size. This implies that the rate of convergence of C^1_T and C^2_T is almost identical. The result based on the $AR(2)$ model, i.e. *model* 4 in table 6.3, seems to confirm this result although the average mean squared errors from the linear regression tends to be a bit lower than the average mean squared errors obtained from the flexible nonlinear estimator. However the convergence of the latter seems to be almost as high. This suggest 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 when the true relation is linear.

[*Table* 6.2]

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 are 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 $T = 50$. However, the rate of convergence of this estimator seems in general to be somewhat lower. Comparison of C_T^2 and C_T^4 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 suggests very clearly that applying Hamilton's estimator

to nonlinear multivariate models seems to be a particular fruitful approach.

[*Table* 6.3]

By inspecting the results on the *block* 2 models reported in Table 6.3, we observe that for all of the models that are linear in mean C_T^2 tends to be about 10-20 pct. higher than C_T^1 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_T^2 tends to be somewhat higher than the convergence rate for C_T^1. The percentage improvement of fit arising from applying Hamilton's estimator to *model* 3 - the $NMA(2)$ 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 one. 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 becomes more pronounced as the sample size increases. Furthermore, these features becomes more pronounced for parameterizations where the model is close to violating at least one of four moment restrictions that establishing invertability and stationarity conditions. In this light Brunner and Hess (1995) recommend *extreme caution* when dealing with the bilinear models.

[*Table* 6.4]

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 range 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.

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 is good the power properties of Hamilton's test are in most cases 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.

References

[1] R. Ashley, D. Patterson, and M. Hinich. A diagnostic test for nonlinear serial dependence in time series fitting errors. *Journal of Time Series Analysis*,

7:165–178, 1986.

[2] W.S. Chan and H. Tong. On tests for nonlinearity in time series models. *Journal of Forecasting*, 5:217–228, 1986.

[3] V. Haggan and T. Ozaki. Modelling nonlinear random vibrations using an amplitude dependent autoregressive time series model. *Biometrica*, 68:189–196, 1981.

[4] James D. Hamilton. A parametric maximum likelihood approach to flexible nonlinear inference. *Unpublished Manuscript*, 1998.

[5] D.M. Keenan. A tukey nonadditivity test for time series nonlinearity. *Biometrica*, 72:39–44, 1985.

[6] Tae-Hwy Lee, Halbert White, and Clive W.J. Granger. Testing for neglected nonlinearity in time series models. *Journal of Econometrics*, 56:269–290, 1993.

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

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

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

[10] T. Terasvirta. Specification, estimation and evaluation of smooth transition autoregressive models. *Discussion paper 90-39, Department of Economics University of California, San Diego*, 90, 1990.

[11] T. Terasvirta, C.F. Lin, and C.W.J. Granger. Power of the neural network linearity test. *Journal of Time Series Analysis*, 14:209–220, 1993.

[12] H. Tong. *Threshold models in nonlinear time series analysis.* Springer-Verlag, New York, NY, 1983.

[13] R.S. Tsay. Nonlinearity test for time series. *Biometrica*, 73:461–466, 1986.

[14] G. Wahba. 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, 1978.

[15] William E. Wecker and Craig F. Ansley. The signal extraction approach to nonlinear regression and spline smoothing. *Journal of the American Statistical Association*, 78:81–89, 1983.

[16] Halbert White. Specification testing in dynamic models. In T.F. Bewley, editor, *Advances in econometrics, Fifth world congress*, pages 1–58. Cambridge University Press, Cambridge, 1987.

[17] Halbert White. An additional hidden unit test for neglected nonlinearity in multilayer feedforward networks. In *Proceedings of the international joint conference on neural networks*, pages 451–455, Washington, DC, 1989. IEEE Press, New York, NY.

[18] Halbert White. *Estimation, Inference and specification analysis*. Cambridge University Press, New York, NY, 1992.

8. Tables

Table 4.1
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 4.2

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 4.3
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 4.4. Definitions of the Reset, Tsay and V23 test used in the simulation study

	p^*	p	k	M_t							Distribution
				y_{t-1}^2	$y_{t-1}y_{t-2}$	y_{t-2}^2	y_{t-1}^3	$y_{t-1}^2y_{t-2}$	$y_{t-1}y_{t-2}^2$	y_{t-2}^3	
block 1											
Reset	1	1	5								F(1,T-3)
Tsay	.	1	.	×							F(1,T-3)
V23	.	1	.	×			×				F(2,T-4)
block 2+3											
Reset	1	2	5								F(1,T-4)
Tsay	.	2	.	×	×	×					F(3,T-5)
V23	.	2	.	×	×	×	×	×	×	×	F(7,T-10)

Note:

Table 5.1
Critical values (5%) based on the AR model in block 1

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]

Note: 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 statistic 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.

Table 5.2
Critical values (5%) based on model 4 in block 2

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]

Note: 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 statistic 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..

Table 5.3
Size of test and similarity

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

Note: (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.

Table 5.4
Power vs. sample size for block 1 and bivariate models

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

Note: Power using 5% simulated critical values reported in table 5.1 is shown.
Replications = 1000, sample size = 50, 100, 200.

Table 5.5
Power vs. sample size for block 2 models

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

Note: Power using 5% simulated critical values reported in table 5.2 is shown. Replications = 1000, sample size = 50, 100, 200.

Table 5.6
Power vs. sample size for block 3 models

	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

Note: Power using 5% simulated critical values reported in table 5.2 is shown.

Table 5.7
Power vs. sample size and noise for bivariate model (SQ)

	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

Note: Power using 5% simulated critical values reported in table 5.1 is shown. The signal-to-noise ratio equals 700% for $\sigma = 1$, 28% for $\sigma = 5$ and 2% for $\sigma = 20$. 1000 replications. Sample size = 50, 100, 200.

Table 5.8
Power vs. sample size and noise for bivariate model (EXP)

	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

Note: Power using 5% simulated critical values reported in table 5.1 is shown. The signal-to-noise ratio equals 216% for $\sigma = 1$, 8.6% for $\sigma = 5$ and 0.5% for $\sigma = 20$. Number of replications equals 1000. Sample size = 50, 100, 200.

Table 6.1
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	$1, \epsilon_{t-1,n}, \epsilon_{t-2,n}, \epsilon_{t-1,n}\epsilon_{t-2,n}, \epsilon^2_{t-2,}$
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	$1, y_{t-1,n}, y_{t-2,n}, y_{t-1,n}\epsilon_{t-1,n}, \epsilon_{t-1,}$
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	$1,$ $[1+\exp(-100(y_{t-1}-0.8y_{t-2}))]^-$ $[1+\exp(-100(y_{t-1}+0.8y_{t-2}))]^-$
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 6.2
Convergence properties of block 1 and bivariate models

	$C^1_{T,N}$	γ^1	$C^2_{T,N}$	γ^2	G^2_T	$C^3_{T,N}$	γ^3	G^3_T	$C^4_{T,N}$	γ^4
					AR					
T=50	0.052		0.053		-0.025	0.117		-1.238	0.052	
T=100	0.029	0.844	0.030	0.844	-0.024	0.088	0.398	-2.047	0.029	0.844
T=200	0.015	0.986	0.016	0.888	-0.107	0.059	0.579	-3.023	0.015	0.986
					BL					
T=50	2.439		2.642		-0.083	2.650		-0.087	0.095	
T=100	2.202	0.272	2.235	0.241	-0.106	2.219	0.256	-0.099	0.027	1.828
T=200	1.851	0.251	1.968	0.184	-0.063	2.023	0.133	-0.093	0.018	0.585
					TAR					
T=50	0.276		0.181		0.345	0.157		0.430	0.059	
T=100	0.258	0.097	0.112	0.685	0.564	0.116	0.442	0.551	0.029	1.028
T=200	0.245	0.076	0.073	0.614	0.701	0.082	0.505	0.666	0.013	1.158
					SGN					
T=50	0.368		0.213		0.422	0.203		0.450	0.047	
T=100	0.344	0.099	0.145	0.552	0.577	0.143	0.505	0.585	0.021	1.171
T=200	0.333	0.048	0.100	0.537	0.700	0.095	0.583	0.713	0.012	0.807
					NAR					
T=50	0.057		0.061		-0.084	0.121		-1.129	0.046	
T=100	0.040	0.500	0.042	0.541	-0.054	0.084	0.526	-1.091	0.022	1.081
T=200	0.030	0.420	0.029	0.558	0.046	0.059	0.505	-0.980	0.011	1.00
					SQ					
T=50	5.362		2.044		0.619	2.374		0.557	0.045	
T=100	5.644	-0.074	1.216	0.749	0.785	1.578	0.589	0.720	0.027	0.729
T=200	5.251	0.104	0.680	0.839	0.871	0.916	0.785	0.826	0.012	0.826
					EXP					
T=50	13.220		8.506		0.357	9.250		0.300	0.118	
T=100	9.858	0.423	4.755	0.839	0.518	5.369	0.785	0.455	0.035	1.751
T=200	9.502	0.053	2.882	0.722	0.697	3.743	0.520	0.606	0.013	1.429

Note: $C^1_{T,N}$, $C^2_{T,N}$, $C^3_{T,N}$ and $C^4_{T,N}$ are calculated according to the their definitions stated above where $\{\widehat{\alpha}_{T,n}, \widehat{\zeta}_T(.)\}$ and $\{\widehat{\omega}_{T,n}\}$ are estimated conditional 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. This is done in order to avoid overfitting. The number of replications equals N=100.

Table 6.3
Convergence properties of block 2 models

	C_T^1	γ^1	C_T^2	γ^2	G_T^2	C_T^3	γ^3	G_T^3	C_T^4	γ^4
					Model1					
T=50	0.113		0.126		-0.114	0.209		-0.855	0.069	
T=100	0.062	0.866	0.067	0.906	-0.084	0.152	0.460	-1.458	0.031	1.144
T=200	0.047	0.387	0.048	0.484	-0.011	0.120	0.337	-1.538	0.017	0.867
					Model2					
T=50	0.393		0.445		-0.132	0.515		0-.309	0.334	
T=100	0.352	0.160	0.374	0.250	-0.063	0.461	0.160	-0.309	0.291	0.199
T=200	0.321	0.132	0.337	0.152	-0.048	0.415	0.153	-0.290	0.259	0.166
					Model3					
T=50	0.384		0.377		0.018	0.386		-0.004	0.104	
T=100	0.386	0.007	0.344	0.132	0.109	0.351	0.134	0.090	0.061	0.783
T=200	0.331	0.223	0.241	0.516	0.272	0.269	0.382	0.186	0.026	1.213
					Model4					
T=50	0.061		0.074		-0.208	0.156		-1.555	0.061	
T=100	0.039	0.634	0.044	0.729	-0.131	0.125	0.321	-2.172	0.039	0.634
T=200	0.016	1.300	0.019	1.204	-0.206	0.092	0.437	-4.828	0.016	1.300
					Model5					
T=50	1.343		1.271		0.054	1.175		0.125	0.129	
T=100	1.018	0.400	0.820	0.633	0.195	0.773	0.604	0.240	0.046	1.505
T=200	1.001	0.024	0.746	0.136	0.254	0.698	0.147	0.302	0.023	1.011
					Model6					
T=50	2.773		2.877		-0.038	2.614		0.057	0.144	
T=100	2.728	0.024	2.715	0.084	0.005	2.499	0.065	0.084	0.060	1.259
T=200	2.321	0.233	2.190	0.310	0.056	2.035	0.296	0.123	0.025	1.256

Note: See table 6.2.

Table 6.4
Convergence properties of block 3 models

	C_T^1	γ^1	C_T^2	γ^2	G_T^2	C_T^3	γ^3	G_T^3	C_T^4	γ^4
					LSTAR					
T=50	1.754		1.627		0.072	1.271		0.275	0.384	
T=100	1.518	0.208	1.018	0.676	0.329	0.828	0.618	0.454	0.173	1.147
T=200	1.488	0.068	0.590	0.786	0.592	0.553	0.582	0.618	0.064	1.44
					ESTAR					
T=50	0.231		0.236		-0.022	0.227		0.015	0.055	
T=100	0.203	0.184	0.172	0.454	0.153	0.158	0.528	0.224	0.034	0.688
T=200	0.183	0.150	0.100	0.777	0.451	0.113	0.487	0.384	0.017	1.028
					NN					
T=50	0.182		0.151		0.168	0.128		0.296	0.018	
T=100	0.172	0.080	0.103	0.555	0.402	0.095	0.425	0.446	0.007	1.325
T=200	0.167	0.040	0.076	0.432	0.544	0.073	0.371	0.561	0.004	0.871
					BN					
T=50	0.220		0.182		0.173	0.157		0.287	0.027	
T=100	0.159	0.468	0.104	0.809	0.347	0.096	0.712	0.398	0.009	1.524
T=200	0.151	0.067	0.070	0.570	0.536	0.069	0.485	0.545	0.005	0.951

Note: See table 6.2.

Appendix A.

Table A.1
Closed form expressions for $H_k(h)$

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	$1-\frac{3}{2}h+\frac{1}{2}h^3-\frac{3h}{8}(1-h^2)^2$

Note: $H_k(h)$ equals unity when $h=0$, and $H_k(h)$ equals zero when $h \geq 1$.