Department of Economics

Testing for Structural Breaks in Nonlinear Dynamic Models
Using Artificial Neural Network Approximations

George Kapetanios

Queen Mary
University of London

Testing for structural breaks in nonlinear dynamic models using artificial neural network approximations

George Kapetanios[*]
Queen Mary, University of London

November 2002

Abstract

In this paper we suggest a number of statistical tests based on neural network models, that are designed to be powerful against structural breaks in otherwise stationary time series processes while allowing for a variety of nonlinear specifications for the dynamic model underlying them. It is clear that in the presence of nonlinearity standard tests of structural breaks for linear models may not have the expected performance under the null hypothesis of no breaks because the model is misspecified. We therefore proceed by approximating the conditional expectation of the dependent variable through a neural network. Then, the residual from this approximation is tested using standard residual based structural break tests. We investigate the asymptoptic behaviour of residual based structural break tests in nonlinear regression models. Monte Carlo evidence suggests that the new tests are powerful against a variety of structural breaks while allowing for stationary nonlinearities.

---

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

1 Introduction

In the statistical literature considerable work has been devoted to the development of theoretical results and methods for the detection of structural breaks, defined as sudden changes in the parameters of otherwise stationary dynamic models. Structural breaks present a serious challenge both for theoretical and applied statistics. Their presence leads to biases in estimation and breakdown in forecasting. Virtually all the work in the area has concentrated on linear models. Nevertheless, nonlinear dynamic models have been receiving increasing attention in the literature recently. One exception to this is the paper by Delgado and Hidalgo (2000) where methods for estimating break dates in nonlinear models have been proposed. However, no tests for the detection of breaks in nonlinear models have been provided. It is clear that once a particular nonlinear model has been selected, then methods that have been developed for linear models can be readily modified to be applied on nonlinear models. However, model selection between altenative nonlinear models is notoriously difficult. Further, one may not wish to commit to one particular model but simply test whether a dynamic model describing a given time series has undergone a structural change while allowing for robustness of the test to the possibility of nonlinearities in the dynamic model.

In this paper we provide a number of tests that are designed to be powerful against structural breaks while allowing for a variety of nonlinear specifications for the dynamic model. It is clear that in the presence of nonlinearity standard tests of structural breaks for linear models may not have the expected performance under the null hypothesis of no breaks because the model is misspecified. Of course, many forms of nonlinearity may be accomodated by appropriately extending the currently available procedures. For example, when a linear model is fitted to a series which follows an ergodic nonlinear process the residual series will be weakly dependent and nonparametric methods may be used to modify the residual based structural break tests to

account for this weak dependence.

This paper adopts a different approach. Rather than modify a structural break test, we choose to approximate the conditional expectation of the dependent variable by a neural network. Then, the residual from this approximation is tested using standard residual based structural break tests. The property that allows the approximation to work is the universal approximator property of neural networks. Testing for structural breaks in an unspecified nonlinear model has not attracted particular attention in the literature and therefore no comparison with any available procedure is possible. However, we expect our suggested method to work better than any nonparametric correction to standard structural break tests especially in small samples. The reason is that any correction will have an asymptotic justification and given the wide variety of alternative nonlinear models the small sample performance of such a correction is likely to be very variable. As a by-product of our analysis we provide a number of results. Firstly we provide a set of conditions under which information criteria may be used to specify a neural network model. Secondly, we examine the asymptotic behaviour of the cumulative sum of nonlinear least squares residuals both under the null hypothesis of no break and under local alternative hypotheses.

The paper is structured as follows: Section 2 discusses the theoretical aspects of the proposed methodology. Section 3 discuses the alternative specifications for the particular testing procedures we propose. Section 4 presents Monte Carlo evidence on the performance of the procedures. Finally, section 5 concludes. Proofs of the main theorems are in the Appendix.

2　Theoretical Considerations

Let the model underlying the stochastic process y_t be given by

$$y_t = f_t(y_{t-1}, \ldots, y_{t-k}, \boldsymbol{x}_{1,t}) + \epsilon_t \tag{1}$$

To simplify notation we introduce $\boldsymbol{x}_t = (y_{t-1}, \ldots, y_{t-k}, \boldsymbol{x}'_{1,t})'$. We specify the null hypothesis of no structural break to be given by

$$H_0 : f_t(.) = f(.) \quad \forall t$$

Under the null hypothesis we approximate the true unknown model by a neural network model. The neural network model takes the form

$$y_t = \alpha + \sum_{i=1}^{R} \beta_i g(\boldsymbol{x}_t, \boldsymbol{\delta}_i) + \epsilon_t \tag{2}$$

There exist large classes of functions, $g(.,.)$ for which such an approximation holds. We propose nonlinear least squares (NLLS) as a general method of estimating the neural network. The objective function of NLLS is given by

$$Q_{T,R} = Q_{T,R}(\boldsymbol{\gamma}) = \sum_{t=1}^{T} \left(y_t - \alpha - \sum_{i=1}^{R} \beta_i g(\boldsymbol{x}_t, \boldsymbol{\delta}_i) \right)^2 \equiv \sum_{t=1}^{T} G_R(\boldsymbol{x}_t, \boldsymbol{\gamma})^2 \equiv \sum_{t=1}^{T} q_t$$

where $\boldsymbol{\gamma} = (\alpha, \boldsymbol{\beta}, \boldsymbol{\delta}'_1, \ldots, \boldsymbol{\delta}'_R)'$, and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_R)'$. In what follows $Q_{T,R}$ and $G_R(.,.)$ may be abbreviated to Q_T and $G(.,.)$ respectively if denoting dependence on R is not of crucial importance to the argument. Below we make a number of assumptions that will be used in the Theorems below.

Assumption 1 y_t and $\boldsymbol{x}_{1,t}$ are L_2-NED processes of size $-1/2$ on processes $\{(\epsilon_t, \boldsymbol{v}'_t)'\}$ and $\{\boldsymbol{v}_t\}$ with finite fourth moments for all t, where ϵ_t and $\boldsymbol{v}_t$ are independent of each other, have finite second moments, are α-mixing of size $-r/r(r-2)$, $r > 2$ and have continuous densities with strictly positive support over the relevant Cartesian space.

Assumption 2 The disturbances ϵ_t are stationary and ergodic, with

$$E(\epsilon_t | \mathcal{X}_t) = 0 \quad E(\epsilon_t^2 | \mathcal{X}_t) = \sigma^2$$

where $\mathcal{X}_t = \sigma(y_{t-s}, \boldsymbol{x}_{t-s+1} | s \geq 1)$

Assumption 3 *For the neural network model $\boldsymbol{\delta}_i \neq \boldsymbol{\delta}_j$, $i \neq j$, $i, j = 1, \ldots, R$.*

Assumption 4 *The neural network parameter space $\Gamma = \Gamma_{\alpha} \times \Gamma_{\boldsymbol{\beta}} \times \Gamma_{\boldsymbol{\delta}}^R$ is compact. For each R and sample size T, the vector of parameters $\boldsymbol{\gamma}_{T,R}^*$ that minimises the expectation of $Q_{T,R}$ lies in the interior of Γ and is unique.*

Assumption 5 *$g(.,.)$ belongs to C^2 in its first argument and to C^3 in its second argument*

Assumption 6 *The following uniform Lipschitz condition is satisfied for (i) $g(.,\boldsymbol{\delta})$, $\boldsymbol{\delta} \in \Gamma_{\boldsymbol{\delta}}$, (ii) each element of the first and second derivatives of (i) with respect to its second argument, (iii) the supremum and infimum of $g(.,\boldsymbol{\delta})$ and its second derivative with respect to its second argument over open balls $B(\boldsymbol{\delta}, \rho)$ around $\boldsymbol{\delta}$ and of radius ρ for all sufficiently small ρ.*

$$|f(\boldsymbol{x}_1, \boldsymbol{\delta}) - f(\boldsymbol{x}_2, \boldsymbol{\delta})| \leq B_f \sum_{i=1}^{v} |x_1^i - x_2^i|, \quad \forall \boldsymbol{x}_1, \boldsymbol{x}_2$$

$$f \in \{g, \frac{\partial g}{\partial \boldsymbol{\delta}}, \frac{\partial^2 g}{\partial \boldsymbol{\delta} \partial \boldsymbol{\delta}'}, sup_{B(\boldsymbol{\delta}, \rho)} g, inf_{B(\boldsymbol{\delta}, \rho)} g, sup_{B(\boldsymbol{\delta}, \rho)} \frac{\partial^2 g}{\partial \boldsymbol{\delta} \partial \boldsymbol{\delta}'}, inf_{B(\boldsymbol{\delta}, \rho)} \frac{\partial^2 g}{\partial \boldsymbol{\delta} \partial \boldsymbol{\delta}'}\}$$

Note that if the Lipschitz condition is satisfied for the functions in (iii) in Assumption 5 then it is satisfied for the function in (i) and the second set of functions in (ii).

Assumption 7

$$\overline{\lim_{T \to \infty}} 1/T \sum_{t=1}^{T} E\left[sup_{\boldsymbol{\gamma} \in \Gamma} \left|\left|\frac{\partial q_t}{\partial \boldsymbol{\gamma}}\right|\right|\right] < \infty$$

and

$$\overline{\lim_{T \to \infty}} 1/T \sum_{t=1}^{T} E\left[sup_{\boldsymbol{\gamma} \in \Gamma} \left|\left|\frac{\partial^3 q_t}{(\partial \boldsymbol{\gamma})^3}\right|\right|\right] < \infty$$

for all finite R.

Assumption 8 *Every element of $\frac{\partial q_t}{\partial \gamma}$ and $\frac{\partial^2 q_t}{\partial \gamma \partial \gamma'}$ have finite first and second moments for all t, all γ in Γ and all finite R.*

Assumption 9

$$plim_{T \to \infty} 1/T \sum_{t=1}^{T} \frac{\partial G_R(\boldsymbol{x}_t, \boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}} \frac{\partial G_R(\boldsymbol{x}_t, \boldsymbol{\gamma})'}{\partial \boldsymbol{\gamma}}$$

converges to a finite nonsingular matrix in probability for $\boldsymbol{\gamma}^0$ and for all finite R.

Assumption 10

$$plim_{T \to \infty} 1/T \sum_{t=1}^{T} G_R(\boldsymbol{x}_t, \boldsymbol{\gamma}_1) \left. \frac{\partial^2 G_R(\boldsymbol{x}_t, \boldsymbol{\gamma})}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\gamma}'} \right|_{\boldsymbol{\gamma}_2}$$

coverges to a finite matrix in probability for all $\boldsymbol{\gamma}_1$ and $\boldsymbol{\gamma}_2$ in an open neighborhood of $\boldsymbol{\gamma}^0$ for all finite R.

In order to state precicely the approximation properties of a neural network we need the following two definitions

Definition 1 *A function $g : \mathbb{R} \to [0,1]$ is a sigmoidal function if it is non-decreasing, $\lim_{x \to \infty} g(x) = 1$ and $\lim_{x \to -\infty} g(x) = 0$*

Definition 2 *A set of functions $\mathcal{S} = \{g | g : \mathbb{R}^r \to \mathbb{R}\}$ is said to approximate a function $f : \mathbb{R}^r \to \mathbb{R}$ in the supremum norm if for every $\varepsilon > 0$ there exists $g \in \mathcal{S}$ such that, for every compact set $K \subset \mathbb{R}^r$, $sup_{\boldsymbol{x} \in K} |f(\boldsymbol{x}) - g(\boldsymbol{x})| < \varepsilon$.*

Theorem 1 *If (i) $f(.)$ is a continuous function, (ii) $g(., \boldsymbol{\delta})$ can be written as $g_1(\boldsymbol{\delta}^{1'}\boldsymbol{x}_t + \delta^2, \boldsymbol{\delta}^3))$, where $\boldsymbol{\delta} = (\boldsymbol{\delta}^{1'}, \delta^2, \boldsymbol{\delta}^{3'})'$ and (iii) $g_1(., \boldsymbol{\delta}^3)$ is (a) sigmoidal or (b) belonging to the space of L_p-bounded functions, for some $p \geq 1$, with non-zero expectation with respect to Lebesque measure, then the set of neural network specifications $\sum_{i=1}^{R} \beta_i g(., \boldsymbol{\delta}_i)$ can approximate $f(.)$ in the supremum norm.*

Below we give a Lemma indicating that if a set of functions $\mathcal{S}$ approximates a function g in the supremum norm, then it approximates the same function in the Kullback-Leibler metric under some weak conditions, i.e. for every $\varepsilon > 0$ there exists $f \in \mathcal{S}$ such that $\int_{\mathbb{R}^r} (\ln f - \ln g) f d\boldsymbol{x} < \varepsilon$

Lemma 1 *For exponentially declining functions, f, g, i.e. $|f(\boldsymbol{x})| < a_1 e^{\boldsymbol{a}_2' \boldsymbol{x}}$ as $\boldsymbol{x} \to -\infty$ and $|f(\boldsymbol{x})| < a_3 e^{-\boldsymbol{a}_4' \boldsymbol{x}}$ as $\boldsymbol{x} \to \infty$ for some $a_1, \boldsymbol{a}_2, a_3, a_4 > 0$, approximation in the supremum norm implies approximation in the Kullback-Leibler metric.*

This lemma indicates that the approximation properties of neural networks imply equivalent approximation properties in the Kullback-Leibler metric used to distinguish models with information criteria which we will use below. Note that since the relevant functions in the Kullback-Leibler case are densities the exponentially declining assumption is not very strict. Note also that in the neural network case the functions involved are conditional means (regressors) while in the Kullback-Lebler metric case they are densities. However, under a continuity assumption about the density of ϵ_t and the fact that for every $\varepsilon > 0$ there exists $\delta > 0$ such that if $\sup_{x \in K} |f - g| < \delta$ then $\sup_{\boldsymbol{x} \in K} |h \circ f - h \circ g| < \varepsilon$, under continuity of h, the two approximation concepts are directly related.

In the context of the neural network specification theorem 1 guarantees that a finite number of hidden units for the neural network will be sufficient to provide an adequate approximation. However it says nothing about the number of hidden units R which needs to be determined empirically. We suggest the use of information criteria to pick the number of hidden units. For each R and sample size T we associate a penalty function $z_T(R)$ with each neural network specification. Then the chosen number of hidden units is the one for which $Q_{T,R} - z_T(R)$ is maximised. Note that Theorem (1) implies that, for given g there exists R^0 such that (i) there exists $\varepsilon > 0$ for which there is no model in the set of neural network models with $R < R^0$

7

for which $\sup_x|f(x) - \sum_{i=1}^{R}\beta_i g(\boldsymbol{x}_t, \boldsymbol{\delta}_i)| < \varepsilon$ and (ii) for all $\varepsilon > 0$ the set of neural network models with $R \geq R^0$ approximates $f(.)$ in the supremum norm. The theorem says nothing about the uniqueness of the approximation for given R. Uniqueness is needed for Theorem (2) and therefore assumed in Assumption (4). We refer to R^0 as the true number of hidden units.

Theorem 2 *Under assumptions 1-10, and assuming that $R_0 > R^0$ where R_0 is the maximum number of hidden units searched by the information criterion search, the number of hidden units needed to approximate sufficiently close the unknown function $f(.)$ can be estimated consistently if for $R^1 < R^2$, $\lim_{T\to\infty} z_T(R^2) - z_T(R^1) \to \infty$ and $z_T(R^2) - z_T(R^1) = o_p(T)$.*

By the above theorem we can assume R known for the neural network specification in what follows. Then, we examine the asymptotic behaviour of the normalised sum of the NLLS residuals. The test statistic is given by

$$B^{(T)}(z) \equiv \frac{1}{\hat{\sigma}\sqrt{T}} \sum_{t=1}^{[Tz]} \hat{\epsilon}_t^{(T)}$$

We prove the following theorem.

Theorem 3 *Under assumptions 1-10, the normalised sum of the NLLS residuals converge to a standard Brownian bridge under the null hypothesis H_0.*

We now explore the local power of the testing procedure for the neural network specification. We do not discuss the mapping between the unknown function $f(.)$, its parameters and the neural network specification under the alternative hypotheses. We assume the following form for the local alternative hypotheses

$$H_T : \tilde{\boldsymbol{\gamma}}_{t,T} = \tilde{\boldsymbol{\gamma}}^0 + 1/\sqrt{T}\boldsymbol{h}(t/T) \tag{3}$$

where $\boldsymbol{h}(t/T)$ is an arbitrary vector function defined on the interval $[0, 1]$ and $\tilde{\boldsymbol{\gamma}}$ is a $1-1$ reparametrisation of $\boldsymbol{\gamma}$ defined in the proof of Theorem 3 in the Appendix. We make the following additional assumptions:

8

Assumption 11 $\boldsymbol{h}(t/T)$ *is a uniform limit of functions that are constant on intervals*

Assumption 12 *For given* $\tilde{\boldsymbol{\gamma}}^0$ *the process* y_t *is geometrically ergodic*

Assumption 11 is imposed because it can be relatively easily verified for a wide class of nonlinear processes and implies strict stationarity which is the important condition needed for the local power properties of dynamic models. Under stationarity of the process, the Wald decomposition implies the existence of an infinite MA representation. Such a representation implies the existence of a mapping from $\tilde{\boldsymbol{\gamma}}$ to the parameters of the MA representation. Denote this mapping by Λ. Then we have the following two assumptions

Assumption 13 *The coefficients of the infinite MA representation follow* $c_i = l_i u^i$ *for a sequence of finite constants* l_i *and some* $u \in (0, 1)$.

Assumption 14 *The uniform Lipschitz condition defined in Assumption 6 holds for* Λ.

This is a relatively high level assumption which may seem difficult to verify in general. However, it can be easily verified in simple cases such as, for example, linear models. We prove the following theorem

Theorem 4 *Under assumptions 1-14 and the local alternatives* H_T *in (3)*

$$B^{(T)}(z) \Rightarrow B(z) + 1/\sigma \left[\int_0^z \boldsymbol{c}' \boldsymbol{h}(u) du - \boldsymbol{c}' z \int_0^1 \boldsymbol{h}(u) du \right]$$

The above analysis can be extended in some ways which we will not formalise but merely indicate. Firstly, we note that as long as a constant is included in the nonlinear regression, the set of neural network models considered, need not include the model which approximates the unknown function to the desired degree of closeness. In such a case the true disturbance simply

becomes a weakly dependent, zero mean, process (which can be described by the L_2-NED class of processes) which is uncorrelated with the conditional mean of the chosen neural network model. Then, it can easily be seen that the distribution of the test statistic still has the same form albeit care needs to be taken on the estimation of the variance used to normalise the partial sum of residuals given that covariances of the residual process may need to be taken into account. A kernel based estimate of the asymptotic long run variance may be of use. Of course, in this case the local power properties may change.

3 Neural Network Specifications

We need to choose suitable functions $g(\boldsymbol{x}_t, \boldsymbol{\delta})$. As we have seen in the theoretical discussion we formally need NLLS estimation for the specification of the neural network. We have also seen that the behaviour of the residual CUSUM test under the null hypothesis, will not be affected by wrong specification of the neural network model as long as proper care is taken with the estimate of the long run variance of the residual process. So essentially the computationally expensive NLLS procedure may be dispensed with and alternative specification methods for the neural network may be used. Such methods may not guarantee consistent estimation of the neural network parameters. We discuss such methods in this section.

In the context of testing for neglected nonlinearity, Lee, White, and Granger (1993) choose the logistic function. This is a monotonic function, with output bounded between 0 and 1. It is sigmoidal and therefore fulfills the conditions required for approximability of any continuous function. Another class of functions which satisfy approximability conditions and further can be used to construct an easily estimated neural network is the radial basis function (RBF) class. A RBF is a function which is monotonic about some center. Let us start by specifying neural networks which use this function.

Define q centers by $\boldsymbol{c}_j$ and a radius vector $\boldsymbol{\tau}$. We interest ourselves only in those functions that are monotonically decreasing about $\boldsymbol{c}_j$. The Gaussian RBF is then

$$\psi([\boldsymbol{c}_j' \quad \boldsymbol{\tau}']', \boldsymbol{x}_t) = \exp\left(-\frac{\|\boldsymbol{x}_t - \boldsymbol{c}_j\|^2}{\boldsymbol{\tau}^2}\right). \tag{4}$$

By the monotonicity property, each RBF has maximum activation (of unity) when the input vector coincides with the jth center independent of $\boldsymbol{\tau}$. Conversely, if the input vector is far enough away for the center the activation is zero, controlled by $\boldsymbol{\tau}$. Other functional forms, such as the multiquadratic, have the same properties and can be used instead. See Campbell, Lo, and MacKinlay (1997) for an introduction to artificial neural networks in general, which covers RBF networks. Bishop (1995) gives a more thorough account. This function being an exponential function satisfies necessary conditions needed for approximability of any general continuous function.

We need to determine the centers $\boldsymbol{c}$ and radii $\boldsymbol{\tau}$ for each RBF, and the number, q, of 'hidden units' used. We use data-based procedures for both. The radii are fixed first. It is common practice in the ANN literature to use a fixed multiple of the maximum change from period t to period $t+1$, $t = 1, \ldots, T$ of each input as the radius for that input (see Orr (1995)). We fix the radii at twice this for all possible centers and hence RBFs, appropriate for time-series data. We then allow there to be T *potential* RBFs by using all the observations themselves as possible centers. Following Orr (1995), we add RBFs to the regression in order of maximum reduction in the unexplained variance. We successively add RBFs until we minimize an information criterion. Conditional on a particular set of RBFs it is clear that estimation of the neural networks involves least squares estimation of the coefficients β in (2).

We move on to discuss the logistic neural network. For this the function, $g(\cdot, \cdot)$ in (2) is given by $\phi(\boldsymbol{\delta}'\boldsymbol{x}_t)$ where $\phi(\lambda)$ is the logistic function, given by

$[1 + \exp(-\lambda)]^{-1}$. Clearly estimation of the coefficients $\boldsymbol{\delta}_j$ is a complicated affair involving nonlinear least squares. To avoid this we follow Lee, White, and Granger (1993) who in the context of nonlinearity testing have obtained the actual coefficients, $\boldsymbol{\delta}_j$ by randomly generating each element of the vector from a uniform distribution over $[\delta_l, \delta_h]$. Then, the estimation problem for the rest of the neural network reduces to a standard least squares application. For given R, the constructed regressors $\phi(\boldsymbol{\delta}_j' \boldsymbol{x}_t)$, $j = 1, \ldots, R$, may suffer from multicollinearity. Lee, White, and Granger (1993) suggest that $\tilde{R}$ largest principle components of the constructed regressors excluding the largest one be used as extra regressors in (2). We modify the above procedure by Lee, White, and Granger (1993) by allowing for model selection between the hidden units. This is done as follows: We generate a large number of hidden units and choose those that explain most of the variation in the dependent variable y_t. Again an information criterion is used to decide the number of hidden units chosen.

Teräsvirta, Lin, and Granger (1993) have in the context of nonlinearity testing, suggested the use of a polynomial approximation to the logistic neural network. The authors suggest that a procedure which tests for the significance of the squares, cubes and cross products of the original regressors should be powerful against a wide variety of departures from linearity. As an example, in a model with two regressors, $x_{1,t}$ and $x_{2,t}$, the joint significance of the following terms is tested: $x_{1,t}^2$, $x_{2,t}^2$, $x_{1,t}^3$, $x_{2,t}^3$, $x_{1,t}^2 x_{2,t}$ and $x_{1,t} x_{2,t}^2$. We use their procedure as an alternative to a neural network for our problem. The above may be considered as a third order Taylor approximation to the logistic neural network. We further extend their procedure to allow for data dependent selection of the desired order of the approximation through an information criterion up to a fourth order approximation.

The second stage of the test involves testing the residuals, from the least

12

squares regression involved in fitting the neural network or its polynomial approximation, for structural breaks. We have theoretically motivated the validity of the NLLS residual CUSUM test. In the Monte Carlo section we use the recursive and NLLS residual CUSUM tests. Under the proposed specifications for the neural network models the estimation problem reduces to a linear least squares problem and therefore we refer to the residual based structural break tests as OLS CUSUM and recursive residual CUSUM tests.

4   Monte Carlo

In this section we present some Monte Carlo evidence on the performance of the new procedures we propose for testing against the presence of structural breaks. Let us first decribe the data generation processes we consider. We consider three classes of nonlinear autoreressive models and linear autoregressive models. The nonlinear models are self-exciting threshold autoregressive (SETAR) models, logistic smooth autoregressive (LSTAR) models and exponential smooth autoregressive (ESTAR) models. They have the following forms. The m-regime SETAR model is given by:

$$y_t = \phi_0^{J_t} + \sum_{i=1}^{p} \phi_i^{J_t} y_{t-i} + \sigma^{J_t} \epsilon_t \tag{5}$$

where $J_t = i$ if $y_{t-d} \in \mathcal{A}_i$, $\cup_{i=1}^{m} \mathcal{A}_i = \mathbb{R}$, $\mathcal{A}_i \cap \mathcal{A}_j = \emptyset$, $\forall i \neq j$ and ϵ is a zero mean, unit variance i.i.d. sequence with finite fourth moments. We see that, $\mathcal{A}_i$, $i = 1, \ldots, m$ are defined by partitioning the real line into segments. We restrict the partition by assuming that $\mathcal{A}_i$, $\forall i$, is compact. The parameters (threshold) controlling this partition are denoted by $r_1, \ldots, r_{m-1}$. The model is denoted by $SETAR(\phi_0^1, \ldots \phi_p^1; \ldots; \phi_0^m, \ldots \phi_p^m; \sigma^1, \ldots, \sigma^p; r_1, \ldots, r_{m-1}; m; d; p)$. The class of STAR models we will consider is given by:

$$y_t = \phi_0 + \psi_0 F(y_{t-d}) + \sum_{i=1}^{p} (\phi_i + \psi_i F(y_{t-d})) y_{t-i} + \sigma \epsilon_t \tag{6}$$

13

where $F(\cdot)$ is some smooth function. Usually, two forms for $F(\cdot)$ are used. One is the exponential function, $F(\lambda) = 1 - \exp(-\lambda^2)$ giving rise to an ESTAR model and the other is the logistic function $F(\lambda) = [1 + \exp(-\lambda)]^{-1}$ giving rise to an LSTAR model. We will denote the LSTAR and ESTAR models by $LSTAR(\phi_0, \ldots, \phi_p; \psi_0, \ldots, \psi_p; \sigma; d; p)$ and $ESTAR(\phi_0, \ldots, \phi_p; \psi_0, \ldots, \psi_p; \sigma; d; p)$ respectively. A linear autoregressive model will be denoted by $AR(\phi_0, \ldots, \phi_p; \sigma; p)$ where the notation is obvious. For the experiments concerning the null hypothesis we consider five cases each using one of the nonlinear (or in the first case linear) autoregressive models. These are given by AR(0,0.5;1;1), SETAR(0,0.5;0,0.5;0,-0.5;1,1,1;-1,0;3,1,1), SETAR(0,0.95;0,1.5;0,0.5;1,1,1;-1,0;3,1,1), LSTAR(0,0.8;0,-1.2;1,1,1) and ESTAR(0,0.8;0,-1.2;1,1,1) respectively and will be referred to as experiments 1-5. We also consider 12 experiments to determine the power properties of the proposed tests. The first 9 experiments consider breaks in nonlinear dynamic models. We use the following notation to denote the data generating process. A process denoted $\{M1; \alpha; M2\}$ follows the model $M1$ for the first α proportion of the sample and model $M2$ for the rest of the sample. The nine experiments referred to as experiments 6-14 are given by

$\{SETAR(0, 0.8; 0, 0.8; 0, -0.8; 1, 1, 1; -1, 0; 3, 1, 1); 0.25; SETAR(0, 0.2; 0, 0.2; 0, -0.2; 1, 1, 1; -1, 0; 3, 1, 1)\}$

$\{SETAR(0, 0.8; 0, 0.8; 0, -0.8; 1, 1, 1; -1, 0; 3, 1, 1); 0.5; SETAR(0, 0.2; 0, 0.2; 0, -0.2; 1, 1, 1; -1, 0; 3, 1, 1)\}$

$\{SETAR(0, 0.8; 0, 0.8; 0, -0.8; 1, 1, 1; -1, 0; 3, 1, 1); 0.75; SETAR(0, 0.2; 0, 0.2; 0, -0.2; 1, 1, 1; -1, 0; 3, 1, 1)\}$

$\{SETAR(0, 0.8; 0, 0.8; 0, -0.8; 1, 1, 1; -1, 0; 3, 1, 1); 0.25; ESTAR(0, 0.2; 0, -0.4; 1; 1; 1)\}$

$\{SETAR(0, 0.8; 0, 0.8; 0, -0.8; 1, 1, 1; -1, 0; 3, 1, 1); 0.5; ESTAR(0, 0.2; 0, -0.4; 1; 1; 1)\}$

$\{SETAR(0, 0.8; 0, 0.8; 0, -0.8; 1, 1, 1; -1, 0; 3, 1, 1); 0.75; ESTAR(0, 0.2; 0, -0.4; 1; 1; 1)\}$

$\{SETAR(0, 0.8; 0, 0.8; 0, -0.8; 1, 1, 1; -1, 0; 3, 1, 1); 0.25; LSTAR(0, 0.2; 0, -0.4; 1; 1; 1)\}$

$\{SETAR(0, 0.8; 0, 0.8; 0, -0.8; 1, 1, 1; -1, 0; 3, 1, 1); 0.5; LSTAR(0, 0.2; 0, -0.4; 1; 1; 1)\}$

$$\{SETAR(0, 0.8; 0, 0.8; 0, -0.8; 1, 1, 1; -1, 0; 3, 1, 1); 0.75; LSTAR(0, 0.2; 0, -0.4; 1; 1; 1)\}$$

Finally, experiments 13-15 consider linear processes and are given by

$$\{AR(0.5, 0.8; 1; 1); 0.25; AR(0.5, 0.2; 1; 1)\}$$

$$\{AR(0.5, 0.8; 1; 1); 0.5; AR(0.5, 0.2; 1; 1)\}$$

$$\{AR(0.5, 0.8; 1; 1); 0.75; AR(0.5, 0.2; 1; 1)\}$$

Note that in all of the above cases the structural break entails a shift in the unconditional mean of the process, as required by the theory we developed in the previous section for the tests to have local power. The nonlinear models have no constant but asymmetry coming from the fact that the regimes of SETAR models are asymmetric with respect to zero, leads to nonzero means for the nonlinear processes.

We consider sample sizes of T=100, 200. For each sample, 20 initial observations are dropped to minimise dependence on initial conditions which are set to zero. All errors are standard normal pseudo-random variables.

We now discuss in detail the tests that we use. There are three dimensions on the testing procedures we consider. The name of each testing procedure will reflect that. The dimensions are:

- The neural network or approximation to a neural network used. These are the RBF, logistic and the polynomial approximation to the logistic neural network. The first three letters of the procedure name reflect the neural netork used. RBF stands for the RBF network, LOG for the logistic and PAP for the polynomial approximation.

- The choice of model selection criterion carried out to determine the hidden units or order of the polynomial approximation. The fourth

and fifth letter of the procedure name reflects that. In particular, for the Akaike criterion the letters AC are used, for the BIC criterion the letters BC are used and for the Hannan-Quinn criterion the letters HC are used. If a polynomial approximation is used and no search is carried out over the order of the approximation, a number appears indicating the order of the approximation (e.g. 03). Note that the Akaike criterion is not consistent in the sense of Theorem 2 whereas BIC and HQ are.

- The structural break test that is being used on the residuals of the neural network. If the recursive residual CUSUM test is used the letter RC appear at the end of the name of the procedure. Otherwise, for OLS residuals the letters OC appear.

Throughout, x_t is set to the first lag of y_t. We further consider two tests of structural breaks when a linear AR(1) model has been fitted to the data. We do this to compare the properties of these tests when a nonlinear model underlies the data. These tests are the recursive and OLS residual CUSUM test denoted by RC and OC respectively. The number of hidden units on which model selection is carried out is equal to the number of obervations for the RBF network, equal to the $[T/10]$ for the logistic neural network, where [.] denotes integer part. For the logistic neural network, if no information criterion is used, the number of hidden units is set to $[T/10]$ whereas the number of principle components of those hidden units is set to $[T/50]$. The support of the uniformly generated random numbers which are generated for the coefficients γ is $[-2, 2]$. For the polynomial approximation, the minimum order is 2 and the maximum 4.

Tables 1-5 present the rejection probabilities for the tests for all experiments. Several conclusions emerge. We first comment on the estimated rejection probabilites under the five null hypotheses. In general the tests based on OLS residuals underreject. For example for samples of 100 observations the actual size is close to 2.5 % for a nominal size of 5%, for a number

of tests. On the other hand, the tests based on recursive residuals tend to overreject under the null hypothesis. However, as far as the nonlinear recursive residual tests are concerned the overrejection is not very pronounced.

The linear recursive residual tests overreject more significantly. In particular, for experiment 3, we see that the linear recursive residual CUSUM test overrejects quite significantly having an estimated rejection probability of 13.6% at 200 observations. Experiment 3 is of a special nature. The nonlinear process in that experiment, although globally ergodic, has both asymmetric outer regimes and explosive roots in the middle regime. The process is geometrically ergodic because both outer regimes have stable roots. Such rather extreme nonlinearities have the potential to lead to strange behaviour for the linear tests, in small samples, whereas the nonlinear tests seem able to handle them. This is one reason for preferring the nonlinear tests. We note that in results not shown here[1] we observe all tests to have good size properties for larger samples (1000 observations).

Moving on to power properties, a number of features emerge. The method of approximating the function does not seem to make a significant difference. The main source of difference in performance seems to come from the choice of the structural break test used once the approximation has been estimated. This feature is apparent in the size properties of the test as well. Having said that, we observe a slightly weaker performance for the polynomial approximation based neural network test in a few cases for the recursive residual based CUSUM test.

The tests based on OLS residuals do significantly better compared to the recursive residual tests when the break is at the middle or end of the sample. The recursive residual tests do relatively better for breaks which are

---

[1]But available upon request.

nearer the start of the sample. Nevertheless, the OLS residual based tests still outperform the recursive residual based tests for such breaks as well, in general. The performance of the tests when a linear model with a break is the true model is quite good. The tests are slightly less powerful against that alternative than against the nonlinear alternatives. They have particularly low power for linear processes and breaks near the end of the sample.

Comparing the nonlinear tests against the standard linear tests we see that the nonlinear recursive residual based tests outperform their linear counterparts for all nonlinear alternative hypotheses and have roughly equal power to them for the linear alternative hypotheses. The OLS residual based linear tests do slightly better than the nonlinear residual based tests for breaks at the start and middle of the sample. The linear tests have lower power than the nonlinear ones for breaks towards the end of the sample. Of course, this case is of added interest when investigation of real time issues is undertaken. The nonlinear tests do equally well with the linear ones for breaks at the end of the sample for linear processes.

5 Conclusion

Despite the widespread use of nonlinear models in recent econometric work little attention has focused on the detection of structural breaks in models that may contain nonlinearity. In this paper we have provided a number of new tests for detecting structural breaks in processes which follow nonlinear dynamic time series models of unknown functional form. We have used neural networks to construct approximations to the unknown functional form of the model and subsequently we have proposed the use of standard structural break tests for detecting structural breaks. The validity of fitting neural network models with information criteria has been shown and the asymptotic behaviour of structural break tests based on NLLS residuals has been established. The use of alternative approximation methods may be considered but

18

in this paper we concetrate on neural networks due to the relative computational ease of their use.

An extensive Monte Carlo study has provided evidence on the performance of the tests and several conclusions have emerged. Firstly the choice of the neural network approximation is not of vital importance. The choice of the structural break test to use is of greater importance. Evidence suggests that OLS residual based tests are better behaved under the null hypothesis and more powerful in a number of alternatives hypotheses. The power of these tests under linear alternatives seems to be acceptable. Standard linear tests have been shown to be useful in detecting breaks in nonlinear models as well but in some cases, where the nonlinearity of the process is pronounced, some of these tests may have bad size properties. The combination of linear and nonlinear tests for breaks, possibly through the use of Bonferroni inequalities, might be of interest and could be investigated in future research.

References

AMEMIYA, T. (1985): *Advanced Econometrics*. Princeton University Press.

ANDREWS, D. W. K. (1987): "Consistency in Nonlinear Econometric Models: A Generic Uniform Law of Large Numbers," *Econometrica*, 55(6), 1465–1471.

BISHOP, C. M. (1995): *Neural Networks for Pattern Recognition*. Oxford University Press.

CAMPBELL, J. Y., A. W. LO, AND A. C. MACKINLAY (1997): *The Econometrics of Financial Markets*. Princeton University Press.

CYBENKO, G. (1989): "Approximation by Superpositions of a Sigmoidal Function," *Mathematics of Control, Signals and Systems*, 2, 304–314.

DAVIDSON, J. (1994): *Stochastic Limit Theory*. Oxford University Press.

DELGADO, M. A., AND J. HIDALGO (2000): "Nonparametric Inference on Structural Breaks," *Journal of Econometrics*, 96, 113–144.

HORNIK, K., M. STINCHCOMBE, AND H. WHITE (1989): "Multilayer Feedforward Networks are Universal Approximators," *Neural Networks*, 2, 359–366.

LEE, T. H., H. WHITE, AND C. W. J. GRANGER (1993): "Testing for Neglected Nonlinearity in Time Series Models: A Comparison of Neural Network Methods and Alternative Tests," *Journal of Econometrics*, 56, 269–290.

ORR, M. J. (1995): "Regularisation in the Selection of Radial Basis Function Centers," *Neural Computation*, 7(3), 606–623.

SIN, C. Y., AND H. WHITE (1996): "Information Criteria for Selecting Possibly Misspecified Parametric Models," *Journal of Econometrics*, 71, 207–225.

STINCHCOMBE, M., AND H. WHITE (1989): "Universal Approximation Using Feedforward Networks with Non-Sigmoid Hidden Layer Activation Functions," *Proceedings of the International Joint Conference on Neural Networks*, 1.

TERÄSVIRTA, T., C. F. LIN, AND C. W. J. GRANGER (1993): "Power of the Neural Network Linearity Test," *Journal of Time Series Analysis*, 14, 209–220.

Appendix

Proof of Theorem 1

The theorem follows directly from a number of existing results. Case (a) follows from Theorem 2.3 of Hornik, Stinchcombe, and White (1989) or alternatively Cybenko (1989). Case (b) follows from Theorem 2.1 of Stinchcombe and White (1989)

QED

Proof of Lemma 1

We need to prove that $\sup_{x \in K} |f(x) - g(x)| < \varepsilon$, for all compact sets $K \subset \mathbb{R}$, implies $\int_{\mathbb{R}^r} (\ln f - \ln g) f \, dx < \varepsilon$. For $\int_K (\ln f - \ln g) f \, dx > \delta$, for some compact K and $\delta > 0$, it is necessary that $(\ln f(x) - \ln g(x)) > \varepsilon$, for some ε, x. But this is not allowed by the assumption of supremum norm approximation. Finally, we need to show that for each $\varepsilon > 0$ there exists a compact set K such that $\int_K (f - g) f \, dx - \int_{\mathbb{R}} (f - g) f \, dx < \varepsilon$. But this is easily seen to hold by the exponentially declining assumption.

QED

Proof of Theorem 2

Denote the summands of the NLLS objective function by q_t. To prove the theorem we have to prove a number of statements. These statements represent the conditions of Proposition 4.2 of Sin and White (1996) applied to the neural network specification. The statements are given below

1. Q_T belongs almost surely to C^1 over Γ

2. $E(Q_T)$ exists, is continuously differentiable and $\frac{\partial E(Q_T)}{\partial \gamma} = E(\frac{\partial Q_T}{\partial \gamma})$

3. For each R and sample size T, the vector of parameters $\boldsymbol{\gamma}^*_{T,R}$ that minimises the expectation of $Q_{T,R}$ lies in the interior of Γ

4. $\boldsymbol{\gamma}^*_{T,R}$ is unique.

5. q_t is almost surely twice-differentiable

6. The determinant of $T^{-1}\frac{\partial^2 Q_T}{\partial \gamma \partial \gamma'}$ is strictly positive in some open sphere around $\gamma^*_{T,R}$.

7. $E\left[T^{-1}\frac{\partial^2 Q_T}{\partial \gamma \partial \gamma'}\right]$ is $O_p(1)$.

8. q_t satisfies a uniform weak law of large numbers (UWLLN)

9. Each element of $\frac{\partial q_t}{\partial \gamma}$ satisfies a central limit theorem (CLT)

10. Each element of $\frac{\partial^2 q_t}{\partial \gamma \partial \gamma'}$ satisfies a uniform weak law of large numbers (UWLLN)

11. If $R^1 < R^0$, $\liminf_{t \to \infty}\left[T^{-1}E(Q_{T,R^2}) - T^{-1}E(Q_{T,R^1})\right]$

12. If $R^1 > R^0$, $Q_{T,R^2} - Q_{T,R^1} = O_p(1)$

In this context we need to clarify the definition of a CLT, a UWLLN and a pointwise law of large numbers (PWLLN). A sequence of random variables, $\{q_t\}_1^T$, is said to satisfy a CLT if (i) $E(q_t^2)$ exists and there is a sequence $\{\sigma_T\}$ of nonstochastic, finite and positive scalars such that $\sigma_T^{-1}T^{-1/2}\sum_{t=1}^{T}\left(q_t - E(q_t)\right) \Rightarrow N(0,1)$.

A sequence of random variables, $\{q_t(\boldsymbol{\gamma})\}_1^T$, is said to satisfy a UWLLN for $\boldsymbol{\gamma} \in \Gamma$ if (i) for each $\boldsymbol{\gamma} \in \Gamma$ $T^{-1}\sum_{t=1}^{T}E(q_t(\boldsymbol{\gamma}))$ exists and is continuous for all T. (ii)

$$\sup_{\boldsymbol{\gamma} \in \Gamma}\left|T^{-1}\sum_{t=1}^{T}\left[q_t(\boldsymbol{\gamma}) - E(q_t(\boldsymbol{\gamma}))\right]\right| = o_p(1)$$

A sequence of random variables, $\{q_t\}_1^T$, is said to satisfy a PWLLN if

$$T^{-1} \sum_{t=1}^{T} [q_t(\boldsymbol{\gamma}) - E(q_t(\boldsymbol{\gamma}))] = o_p(1)$$

We first prove that regularity conditions (1)-(7) above hold. Then we prove conditions (8)-(12). (1) follows from assumption 5. (2) follows from the finite second moment conditions on y_t, $\boldsymbol{x}_t$ and ϵ_t and the continuity and infinite support assumption on the processes ϵ_t and $\boldsymbol{v}_t$. (3) and (4) follows from assumption 4. (5) follows from assumption 5. (6) follows from the equality in probability limit of $T^{-1} \frac{\partial^2 Q_T}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\gamma}'}$ and $\text{plim}_{T \to \infty} 1/T \sum_{t=1}^{T} \frac{\partial G_R(\boldsymbol{x}_t, \boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}} \frac{\partial G_R(\boldsymbol{x}_t, \boldsymbol{\gamma})'}{\partial \boldsymbol{\gamma}}$ and assumption 9. The equality in probability limit follows from assumptions 8 and 10 and the analysis of Amemiya (1985, pp. 132-133). (7) follows from assumption 8.

To prove a CLT for $\frac{\partial q_t}{\partial \boldsymbol{\gamma}}$ we have the following. From assumptions $\frac{\partial q_t}{\partial \boldsymbol{\gamma}}$ has finite first and second moment. From assumption and Assumption since y_t and $\boldsymbol{x}_t$ are L_2-NED processes of size $-1/2$ and $\frac{\partial g(.,.)}{\partial \boldsymbol{\gamma}}$ satisfies the relevant Lipschitz condition then $\frac{\partial q_t}{\partial \boldsymbol{\gamma}}$ is, suitably normalised, a L_2-NED process of size $-1/2$. This implies from theorem 24.6 of Davidson (1994), and the mixing properties of ϵ_t and $\boldsymbol{v}_t$, that $\frac{\partial q_t}{\partial \boldsymbol{\gamma}}$ follows a CLT as defined above, where the normalising constants of that theorem are taken to be given by expression (24.29) of Davidson (1994).

To prove a UWLLN for q_t and $\frac{\partial^2 q_t}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\gamma}}'$ we use results from Andrews (1987). For the first part of the UWLLN to hold we need existence and continuity of the average expectation. By assumption this holds. To satisfy the second part of the definition we use the main theorem of Andrews (1987). We need the following to hold: (i) PWLLNs for (a) $\sup_{\boldsymbol{\gamma} \in B(\boldsymbol{\gamma}, \rho)} q_t(\boldsymbol{\gamma})$ (b) $\inf_{\boldsymbol{\gamma} \in B(\boldsymbol{\gamma}, \rho)} q_t(\boldsymbol{\gamma})$ (c) $\sup_{\boldsymbol{\gamma} \in B(\boldsymbol{\gamma}, \rho)} \frac{\partial q_t(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}}$ and (d) $\inf_{\boldsymbol{\gamma} \in B(\boldsymbol{\gamma}, \rho)} \frac{\partial q_t(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}}$ for all $\boldsymbol{\gamma}$ and all sufficiently small ρ, and (ii) Assumption A5 of Andrews (1987). Assumption A5 of Andrews (1987) holds by assumption. The PWLLNs are obtained as follows: By as-

sumption and Theorem of Davidson (1994) we get that the functions (a)-(d) are L_2-NED processes. Assumption gives L_2 boundedness for these functions. These two facts are used to satisfy the conditions of Theorem 19.11 of Davidson (1994) which results in the required PWLLNs. No size restrictions on the NED processes are needed.

To prove that, for $R^1 < R^0$, $\liminf_{t \to \infty}[T^{-1}E(Q_{T,R^2}) - T^{-1}E(Q_{T,R^1})]$ we have that for $\boldsymbol{x} \in \mathcal{R}$ where $\mathcal{R}$ is a set with non-zero Lebesque measure, $|G_{R^1}(\boldsymbol{x}, \boldsymbol{\gamma}) - f(\boldsymbol{x})| > \delta$ for some constant δ. Otherwise $R^1 \geq R^0$. From this and the uncorrelatedness of the conditional mean and the error sequence, the result easily follows.

To prove that, for $R^1 > R^0$, $Q_{T,R^2} - Q_{T,R^1} = O_p(1)$ we can use a first order Taylor expansion of both Q_{T,R^1} and Q_{T,R^2} around the true value of $\boldsymbol{\gamma}$, $\boldsymbol{\gamma}^0$ and since the elements of $\frac{\partial q_t}{\partial \boldsymbol{\gamma}}$ follow a CLT and $(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}^0) = O_p(T^{-1/2})$ for both Q_{T,R^1} and Q_{T,R^2} from standard NLLS analysis (or alternative see theorem 3) below), the result follows.

All the above together with the conditions given on the penalty functions in the statement of the theorem imply that the conditions of Proposition 4.2 (a) and (c) of Sin and White (1996) hold implying consistency of selection by information criteria.

QED

Proof of Theorem 3

Without loss of generality we can reparametrize the neural network model such that it is given by

$$y_t = \tilde{\alpha} + \sum_{t=1}^{n} \tilde{\beta}_i \tilde{g}(\boldsymbol{x}_t, \boldsymbol{\delta}_i) + \epsilon_t$$

where
$$\tilde{g}(\boldsymbol{x}_t, \boldsymbol{\delta}_i) = g(\boldsymbol{x}_t, \boldsymbol{\delta}_i) - 1/T \sum_{t=1}^{T} g(\boldsymbol{x}_t, \boldsymbol{\delta}_i)$$

The test statistic is given by
$$B^{(T)}(z) = \frac{1}{\hat{\sigma}\sqrt{T}} \sum_{t=1}^{[Tz]} \hat{\epsilon}_t^{(T)}$$

Now
$$\hat{\epsilon}_t^{(T)} = \epsilon_t - [\tilde{G}(\boldsymbol{x}_t, \hat{\tilde{\boldsymbol{\gamma}}}) - \tilde{G}(\boldsymbol{x}_t, \tilde{\boldsymbol{\gamma}}^0)]$$

where
$$\tilde{G}(\boldsymbol{x}_t, \tilde{\boldsymbol{\gamma}}) = \tilde{\alpha} + \sum_{t=1}^{R} \tilde{\beta}_i \tilde{g}(\boldsymbol{x}_t, \boldsymbol{\delta}_i)$$

$\tilde{\boldsymbol{\gamma}} = (\tilde{\alpha}, \tilde{\beta}_1, \ldots, \tilde{\beta}_n, \boldsymbol{\delta}_1, \ldots, \boldsymbol{\delta}_n)$. A first order Taylor expansion of $\tilde{G}(.,.)$ around $\tilde{\boldsymbol{\gamma}}^0$ gives

$$\tilde{G}(\boldsymbol{x}_t, \hat{\tilde{\boldsymbol{\gamma}}}) = \tilde{G}(\boldsymbol{x}_t, \tilde{\boldsymbol{\gamma}}^0) + \boldsymbol{G}_t^{\tilde{\gamma}}(\tilde{\boldsymbol{\gamma}}^0)(\hat{\tilde{\boldsymbol{\gamma}}} - \tilde{\boldsymbol{\gamma}}^0) + O_p((\hat{\tilde{\boldsymbol{\gamma}}} - \tilde{\boldsymbol{\gamma}}^0)'(\hat{\tilde{\boldsymbol{\gamma}}} - \tilde{\boldsymbol{\gamma}}^0))$$

or
$$\tilde{G}(\boldsymbol{x}_t, \hat{\tilde{\boldsymbol{\gamma}}}) - \tilde{G}(\boldsymbol{x}_t, \tilde{\boldsymbol{\gamma}}^0) = \boldsymbol{G}_t^{\tilde{\gamma}}(\tilde{\boldsymbol{\gamma}}^0)(\hat{\tilde{\boldsymbol{\gamma}}} - \tilde{\boldsymbol{\gamma}}^0) + O_p(T^{-1})$$

where $\boldsymbol{G}_t^{\tilde{\gamma}}(\tilde{\boldsymbol{\gamma}}^0) \equiv \left. \frac{\partial \tilde{G}(\boldsymbol{x}_t, \tilde{\gamma})}{\partial \tilde{\gamma}} \right|_{\tilde{\gamma} = \tilde{\gamma}^0}$ and $\hat{\tilde{\boldsymbol{\gamma}}}$ is the NLLS estimate of $\tilde{\boldsymbol{\gamma}}^0$. From the results of Proposition 4.1 of Sin and White (1996) which follow from the assumptions made in Theorem 2, $(\hat{\tilde{\boldsymbol{\gamma}}} - \tilde{\boldsymbol{\gamma}}^0) = O_p(T^{-1/2})$. So

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{[Tz]} \hat{\epsilon}_t^{(T)} = \frac{1}{\sqrt{T}} \sum_{t=1}^{[Tz]} \epsilon_t - \frac{1}{\sqrt{T}} \sum_{t=1}^{[Tz]} \boldsymbol{G}_t^{\tilde{\gamma}}(\tilde{\boldsymbol{\gamma}}^0)(\hat{\tilde{\boldsymbol{\gamma}}} - \tilde{\boldsymbol{\gamma}}^0)$$

where we have disregarded $o_p(T^{-1/2})$ terms. We want to prove that

$$\text{plim}_{T \to \infty} \sup_{0 \le z \le 1} \left| 1/\sqrt{T} \left(\sum_{t=1}^{[Tz]} \boldsymbol{G}_t^{\tilde{\gamma}}(\tilde{\boldsymbol{\gamma}}^0)(\hat{\tilde{\boldsymbol{\gamma}}} - \tilde{\boldsymbol{\gamma}}^0) - z \sum_{t=1}^{T} \epsilon_t \right) \right| = 0 \quad (7)$$

If this holds then the test statistic may be written as

$$1/\sqrt{T} \left(\sum_{t=1}^{[Tz]} \epsilon_t - z \sum_{t=1}^{T} \epsilon_t \right)$$

which by standard results converges to a normalised Brownian bridge. Now to prove (7) we have that

$$1/\sqrt{T}\sum_{t=1}^{[Tz]}\boldsymbol{G}_t^{\tilde{\gamma}}(\tilde{\boldsymbol{\gamma}}^0)(\hat{\tilde{\boldsymbol{\gamma}}}-\tilde{\boldsymbol{\gamma}}^0)=\left(1/T\sum_{t=1}^{Tz}\boldsymbol{G}_t^{\tilde{\gamma}}(\tilde{\boldsymbol{\gamma}}^0)\right)\sqrt{T}(\hat{\tilde{\boldsymbol{\gamma}}}-\tilde{\boldsymbol{\gamma}}^0)$$

Now by the reparametrisation of the model we know that

$$\text{plim}_{T\to\infty}1/T\sum_{t=1}^{T}\boldsymbol{G}_t^{\tilde{\gamma}}(\tilde{\boldsymbol{\gamma}}^0)=(1,0,\ldots,0)\equiv\boldsymbol{c}' \tag{8}$$

It follows that

$$\text{plim}_{T\to\infty}1/T\sum_{t=1}^{Tz}\boldsymbol{G}_t^{\tilde{\gamma}}(\tilde{\boldsymbol{\gamma}}^0)=(z,0,\ldots,0)$$

Also we have that by a first order Taylor expansion of the first derivative of the NLLS obective function

$$\left.\frac{\partial Q_T}{\partial\tilde{\gamma}}\right|_{\tilde{\gamma}=\hat{\tilde{\gamma}}}=\left.\frac{\partial Q_T}{\partial\tilde{\gamma}}\right|_{\tilde{\gamma}=\tilde{\gamma}^0}+\left.\frac{\partial^2 Q_T}{\partial\tilde{\gamma}\partial\tilde{\gamma}'}\right|_{\tilde{\gamma}=\tilde{\gamma}^0}(\hat{\tilde{\boldsymbol{\gamma}}}-\tilde{\boldsymbol{\gamma}}^0)+O_p((\hat{\tilde{\boldsymbol{\gamma}}}-\tilde{\boldsymbol{\gamma}}^0)'(\hat{\tilde{\boldsymbol{\gamma}}}-\tilde{\boldsymbol{\gamma}}^0))$$

or, by assumptions 8, 9 and 10 and by the definition and consistency of the NLLS estimator

$$\sqrt{T}(\hat{\tilde{\boldsymbol{\gamma}}}-\tilde{\boldsymbol{\gamma}}^0)=\left[1/T\sum_{t=1}^{T}\boldsymbol{G}_t^{\tilde{\gamma}}(\tilde{\boldsymbol{\gamma}}^0)'\boldsymbol{G}_t^{\tilde{\gamma}}(\tilde{\boldsymbol{\gamma}}^0)\right]^{-1}1/\sqrt{T}\sum_{t=1}^{T}\boldsymbol{G}_t^{\tilde{\gamma}}(\tilde{\boldsymbol{\gamma}}^0)'\epsilon_t=$$

$$\boldsymbol{R}_T^{-1}1/\sqrt{T}\sum_{t=1}^{T}\boldsymbol{G}_t^{\tilde{\gamma}}(\tilde{\boldsymbol{\gamma}}^0)'\epsilon_t$$

Partitioning the above expression gives

$$\sqrt{T}(\hat{\tilde{\boldsymbol{\gamma}}}-\tilde{\boldsymbol{\gamma}}^0)=1/\sqrt{T}\begin{pmatrix}1 & 0\\ 0 & \boldsymbol{R}_T^*\end{pmatrix}^{-1}\begin{pmatrix}\sum_{t=1}^{T}\epsilon_t\\ \sum_{t=1}^{T}\boldsymbol{G}_t^{*'}\epsilon_t\end{pmatrix}$$

where the starred entries indicate partitioning of the relevant matrix and vector. Multiplying out the expression and using (8), gives the required result and proves the theorem.

QED

Proof of Theorem 4

The neural network model becomes

$$y_{T,t} = \alpha_{t,T} + \sum_{i=1}^{R} \beta_{T,t,i} g(\boldsymbol{x}_t, \boldsymbol{\delta}_{T,t,i}) + \epsilon_t$$

First note that the formal analysis that follows assumes that the RHS variables do not form a triangular array like the LHS variable and are therefore exogenous. However we show that this assumption which excludes the presence of lagged dependent variables is of no consequence for the analysis and is adopted to simplify the notation. We first derive the first order asymptotic relationship between the NLLS estimator under the null hypothesis and the NLLS estimator under the local alternative hypotheses. Under the null hypothesis we have shown in Theorem 3 that

$$(\hat{\tilde{\boldsymbol{\gamma}}} - \tilde{\boldsymbol{\gamma}}^0) = \left(\sum_{t=1}^{T} \boldsymbol{G}_t^{\tilde{\gamma}}(\tilde{\boldsymbol{\gamma}}^0)' \boldsymbol{G}_t^{\tilde{\gamma}}(\tilde{\boldsymbol{\gamma}}^0) \right)^{-1} \sum_{t=1}^{T} \boldsymbol{G}_t^{\tilde{\gamma}}(\tilde{\boldsymbol{\gamma}}^0)' \epsilon_t$$

Denote the NLLS estimator under the local alternative by $\bar{\tilde{\boldsymbol{\gamma}}}$. We derive the probability limit of $\sqrt{T}(\bar{\tilde{\boldsymbol{\gamma}}} - \tilde{\boldsymbol{\gamma}})$. By a similar Taylor expansion of the first derivative of the NLLS objective function around $\tilde{\boldsymbol{\gamma}}^0$ to the one carried out in the proof of the previous theorem we have that

$$(\bar{\tilde{\boldsymbol{\gamma}}} - \tilde{\boldsymbol{\gamma}}^0) = \left(\sum_{t=1}^{T} \boldsymbol{G}_t^{\tilde{\gamma}}(\tilde{\boldsymbol{\gamma}}^0)' \boldsymbol{G}_t^{\tilde{\gamma}}(\tilde{\boldsymbol{\gamma}}^0) \right)^{-1} \sum_{t=1}^{T} \boldsymbol{G}_t^{\tilde{\gamma}}(\tilde{\boldsymbol{\gamma}}^0)' (y_t - \tilde{G}(\boldsymbol{x}_t, \tilde{\boldsymbol{\gamma}}^0)) =$$

$$\left(\sum_{t=1}^{T} \boldsymbol{G}_t^{\tilde{\gamma}}(\tilde{\boldsymbol{\gamma}}^0)' \boldsymbol{G}_t^{\tilde{\gamma}}(\tilde{\boldsymbol{\gamma}}^0) \right)^{-1} \sum_{t=1}^{T} \boldsymbol{G}_t^{\tilde{\gamma}}(\tilde{\boldsymbol{\gamma}}^0)' (\epsilon_t + \tilde{G}(\boldsymbol{x}_t, \tilde{\boldsymbol{\gamma}}_{t,T}) - \tilde{G}(\boldsymbol{x}_t, \tilde{\boldsymbol{\gamma}}^0))$$

By a first order Taylor expansion of $\tilde{G}(.,.)$ around $\tilde{\boldsymbol{\gamma}}^0$ we have that $\tilde{G}(\boldsymbol{x}_t, \tilde{\boldsymbol{\gamma}}_{t,T}) - \tilde{G}(\boldsymbol{x}_t, \tilde{\boldsymbol{\gamma}}^0)$ is equal to $1/\sqrt{T} \boldsymbol{G}_t^{\tilde{\gamma}}(\tilde{\boldsymbol{\gamma}}^0) \boldsymbol{h}(t/T)$. Thus,

$$(\bar{\tilde{\boldsymbol{\gamma}}} - \hat{\tilde{\boldsymbol{\gamma}}}) = \left(\sum_{t=1}^{T} \boldsymbol{G}_t^{\tilde{\gamma}}(\tilde{\boldsymbol{\gamma}}^0)' \boldsymbol{G}_t^{\tilde{\gamma}}(\tilde{\boldsymbol{\gamma}}^0) \right)^{-1} 1/\sqrt{T} \sum_{t=1}^{T} \boldsymbol{G}_t^{\tilde{\gamma}}(\tilde{\boldsymbol{\gamma}}^0)' \boldsymbol{G}_t^{\tilde{\gamma}}(\tilde{\boldsymbol{\gamma}}^0) \boldsymbol{h}(t/T)$$

27

Now consider the residuals of the NLLS estimation under the local alternative hypotheses. We have

$$y_{t,T} - \tilde{G}(\boldsymbol{x}_t, \bar{\tilde{\gamma}}) = \epsilon_t + \tilde{G}(\boldsymbol{x}_t, \tilde{\gamma}_{t,T}) - \tilde{G}(\boldsymbol{x}_t, \bar{\tilde{\gamma}})$$

$$= \epsilon_t + \tilde{G}(\boldsymbol{x}_t, \tilde{\gamma}^0) - \tilde{G}(\boldsymbol{x}_t, \hat{\tilde{\gamma}}) + (\tilde{G}(\boldsymbol{x}_t, \tilde{\gamma}_{t,T}) - \tilde{G}(\boldsymbol{x}_t, \tilde{\gamma}^0)) - (\tilde{G}(\boldsymbol{x}_t, \bar{\tilde{\gamma}}) - \tilde{G}(\boldsymbol{x}_t, \hat{\tilde{\gamma}}))$$

The first three terms of the above expression are exactly the same as those appearing in the same expansion under the null hypothesis. Therefore, we examine the fourth and fifth terms. We start with the fifth term. A first order Taylor expansion of $\tilde{G}(.,.)$ around $\hat{\tilde{\gamma}}$ shows that $\tilde{G}(\boldsymbol{x}_t, \bar{\tilde{\gamma}}) - \tilde{G}(\boldsymbol{x}_t, \hat{\tilde{\gamma}})$ is equal to $\boldsymbol{G}_t^{\tilde{\gamma}}(\tilde{\gamma}^0)(\bar{\tilde{\gamma}} - \hat{\tilde{\gamma}})$ From the above analysis we know that this is equal to

$$\boldsymbol{G}_t^{\tilde{\gamma}}(\tilde{\gamma}^0) \left(\sum_{t=1}^{T} \boldsymbol{G}_t^{\tilde{\gamma}}(\tilde{\gamma}^0)' \boldsymbol{G}_t^{\tilde{\gamma}}(\tilde{\gamma}^0) \right)^{-1} 1/\sqrt{T} \sum_{t=1}^{T} \boldsymbol{G}_t^{\tilde{\gamma}}(\tilde{\gamma}^0)' \boldsymbol{G}_t^{\tilde{\gamma}}(\tilde{\gamma}^0) \boldsymbol{h}(t/T)$$

We also have from above that

$$\tilde{G}(\boldsymbol{x}_t, \tilde{\gamma}_{t,T}) - \tilde{G}(\boldsymbol{x}_t, \tilde{\gamma}^0) = 1/\sqrt{T} \boldsymbol{G}_t^{\tilde{\gamma}}(\tilde{\gamma}^0) \boldsymbol{h}(t/T)$$

Combining the above results gives that

$$y_{t,T} - \tilde{G}(\boldsymbol{x}_t, \bar{\tilde{\gamma}}) = \hat{\epsilon} + 1/\sqrt{T} \boldsymbol{G}_t^{\tilde{\gamma}}(\tilde{\gamma}^0) \boldsymbol{h}(t/T) -$$

$$\boldsymbol{G}_t^{\tilde{\gamma}}(\tilde{\gamma}^0) \left(\sum_{t=1}^{T} \boldsymbol{G}_t^{\tilde{\gamma}}(\tilde{\gamma}^0)' \boldsymbol{G}_t^{\tilde{\gamma}}(\tilde{\gamma}^0) \right)^{-1} 1/\sqrt{T} \sum_{t=1}^{T} \boldsymbol{G}_t^{\tilde{\gamma}}(\tilde{\gamma}^0)' \boldsymbol{G}_t^{\tilde{\gamma}}(\tilde{\gamma}^0) \boldsymbol{h}(t/T)$$

This implies that

$$1/\sqrt{T} \sum_{t=1}^{[Tz]} (y_{t,T} - \tilde{G}(\boldsymbol{x}_t, \bar{\tilde{\gamma}})) = \frac{1}{\sqrt{T}} \sum_{t=1}^{[Tz]} \hat{u}_t + \frac{1}{T} \sum_{t=1}^{[Tz]} \boldsymbol{G}_t^{\tilde{\gamma}}(\tilde{\gamma}^0) \boldsymbol{h}(t/T) - \qquad (9)$$

$$\frac{1}{T} \sum_{t=1}^{[Tz]} \boldsymbol{G}_t^{\tilde{\gamma}}(\tilde{\gamma}^0) \left[\left(1/T \sum_{t=1}^{T} \boldsymbol{G}_t^{\tilde{\gamma}}(\tilde{\gamma}^0)' \boldsymbol{G}_t^{\tilde{\gamma}}(\tilde{\gamma}^0) \right)^{-1} 1/T \sum_{t=1}^{T} \boldsymbol{G}_t^{\tilde{\gamma}}(\tilde{\gamma}^0)' \boldsymbol{G}_t^{\tilde{\gamma}}(\tilde{\gamma}^0) \boldsymbol{h}(t/T) \right]$$

$$(10)$$

The first term tends to a Brownian bridge by Theorem 1. By the FCLT we have for the second term

$$\frac{1}{T}\sum_{t=1}^{[Tz]} \boldsymbol{G}_t^{\tilde{\gamma}}(\tilde{\boldsymbol{\gamma}}^0)\boldsymbol{h}(t/T) \Rightarrow \int_0^z \boldsymbol{c}'\boldsymbol{h}(u)du$$

uniformly in z. Finally, for the third term we firstly have

$$\left(1/T\sum_{t=1}^{T} \boldsymbol{G}_t^{\tilde{\gamma}}(\tilde{\boldsymbol{\gamma}}^0)'\boldsymbol{G}_t^{\tilde{\gamma}}(\tilde{\boldsymbol{\gamma}}^0)\right)^{-1} 1/T\sum_{t=1}^{T} \boldsymbol{G}_t^{\tilde{\gamma}}(\tilde{\boldsymbol{\gamma}}^0)'\boldsymbol{G}_t^{\tilde{\gamma}}(\tilde{\boldsymbol{\gamma}}^0)\boldsymbol{h}(t/T) \Rightarrow \boldsymbol{R}^{-1}\boldsymbol{R}\int_0^1 \boldsymbol{h}(u)du$$

Secondly $1/T\sum_{t=1}^{[Tz]} \boldsymbol{G}_t^{\tilde{\gamma}}(\tilde{\boldsymbol{\gamma}}^0) \xrightarrow{p} \boldsymbol{c}'z$. And so

$$\frac{1}{T}\sum_{t=1}^{[Tz]} \boldsymbol{G}_t^{\tilde{\gamma}}(\tilde{\boldsymbol{\gamma}}^0)\left[\left(1/T\sum_{t=1}^{T} \boldsymbol{G}_t^{\tilde{\gamma}}(\tilde{\boldsymbol{\gamma}}^0)'\boldsymbol{G}_t^{\tilde{\gamma}}(\tilde{\boldsymbol{\gamma}}^0)\right)^{-1} 1/T\sum_{t=1}^{T} \boldsymbol{G}_t^{\tilde{\gamma}}(\tilde{\boldsymbol{\gamma}}^0)'\boldsymbol{G}_t^{\tilde{\gamma}}(\tilde{\boldsymbol{\gamma}}^0)\boldsymbol{h}(t/T)\right] \Rightarrow$$

$$\boldsymbol{c}'z\int_0^1 \boldsymbol{h}(u)du$$

Combining these results leads to the stated result. The proof of the theorem is completed if we show that the conclusion of the theorem is not affected if we allow for lagged dependent variables in the RHS variables entering the neural network. For that it is sufficient to show that

$$\sum_{t=1}^{T}(y_{t,T} - y_t)^2 = O_p(1) \tag{11}$$

If this condition holds then the previous analysis is easily seen to hold when $\boldsymbol{x}_{t,T}$ which contains lagged values of $y_{t,T}$ is replaced for $\boldsymbol{x}_t$. Such a replacement leads to expression (9) with extra terms involving $\boldsymbol{x}_{t,T} - \boldsymbol{x}_t$ which are asymptotically negligible if (11) holds. For any T the models generating $y_{t,T}$ and y_t differ by the use of $\tilde{\boldsymbol{\gamma}}^0$ and $\tilde{\boldsymbol{\gamma}}_{t,T}$ respectively. By the boundedness of $\boldsymbol{h}(.)$, for some vector of constants $\boldsymbol{d}_1$ $\tilde{\boldsymbol{\gamma}}^0 - \tilde{\boldsymbol{\gamma}}_{t,T} = \boldsymbol{d}_1/\sqrt{T}$. We now examine the infinite MA representation of y_t and $y_{t,T}$. We analyse the case with no exogenous variables only, to simplify analysis. Introducing exogenous variables does not alter the essence of the argument, since the infinite MA representation would be in terms of both ϵ_t and the exogenous variables but otherwise

the analysis would be the same. We have that $\tilde{\gamma}^0 - \tilde{\gamma}_{t,T} = \boldsymbol{d}_1/\sqrt{T}$. We denote the two sets of MA coefficients by c_i and $c_{T,i}$, $i = 1, \ldots$, respectively. By assumptions 12 and 13, $|c_i - c_{T,i}| \leq d_{T,i}/\sqrt{T}$ for some sequence of constants $d_{T,i} \sim u^i$, $u \in (0,1)$. We now examine $y_t - y_{t,T}$

$$y_t - y_{t,T} = \sum_{t=1}^{T} c_i \epsilon_{t-i} - \sum_{t=1}^{T} c_{T,i} \epsilon_{t-i} = \sum_{t=1}^{T} d_{T,i}^* c_{i,T}^* \epsilon_{t-i}$$

where the sequence $c_{T,i}^* \sim u^i$, $u \in (0,1)$, and the sequence $d_{T,i}^*$ is made up of $O(T^{-1/2})$ constants. By the Markov inequality we can show that the square of the above term is $O_p(T^{-1})$ if its variance is $O(T^{-1})$. We have that the variance of $\sum_{t=1}^{T} d_{T,i}^* c_i^* \epsilon_{t-i}$ is smaller than the variance of $\max_i d_{T,i}^* \sum_{t=1}^{T} c_i^* \epsilon_{t-i}$ which by the infinite MA representation has variance which is $O(T^{-1})$. From this the required result follows upon summation.

QED

Table 1: Rejection Probabilities of structural break tests for Experiments 1-5

Test	Experiment									
	Sample Size		Sample Size		Sample Size		Sample Size		Sample Size	
	100	200	100	200	100	200	100	200	100	200
OC	0.022	0.032	0.024	0.037	0.014	0.026	0.017	0.041	0.022	0.033
RC	0.066	0.091	0.053	0.079	0.112	0.136	0.068	0.086	0.043	0.053
RBFACOC	0.019	0.028	0.024	0.037	0.005	0.020	0.013	0.031	0.018	0.031
RBFBCOC	0.024	0.049	0.024	0.035	0.008	0.023	0.017	0.038	0.022	0.032
RBFHCOC	0.022	0.037	0.024	0.036	0.006	0.022	0.015	0.034	0.020	0.030
LOGACOC	0.023	0.031	0.024	0.039	0.006	0.016	0.013	0.032	0.022	0.038
LOGBCOC	0.023	0.029	0.021	0.036	0.006	0.013	0.011	0.035	0.019	0.035
LOGHCOC	0.024	0.032	0.025	0.036	0.007	0.019	0.011	0.034	0.018	0.035
PAPACOC	0.019	0.030	0.021	0.034	0.006	0.015	0.011	0.031	0.019	0.032
PAPBCOC	0.022	0.031	0.020	0.035	0.009	0.021	0.012	0.035	0.019	0.030
PAPHCOC	0.020	0.030	0.020	0.036	0.008	0.018	0.012	0.033	0.018	0.032
PAP03OC	0.019	0.031	0.022	0.034	0.002	0.012	0.010	0.032	0.016	0.033
RBFACRC	0.049	0.050	0.051	0.065	0.041	0.043	0.054	0.058	0.043	0.044
RBFBCRC	0.066	0.073	0.054	0.071	0.052	0.050	0.061	0.077	0.047	0.065
RBFHCRC	0.060	0.064	0.052	0.070	0.050	0.046	0.058	0.075	0.047	0.056
LOGACRC	0.050	0.077	0.041	0.053	0.077	0.083	0.045	0.064	0.040	0.052
LOGBCRC	0.052	0.080	0.045	0.062	0.079	0.090	0.055	0.068	0.043	0.057
LOGHCRC	0.052	0.079	0.047	0.062	0.073	0.093	0.048	0.072	0.043	0.050
PAPACRC	0.049	0.074	0.048	0.056	0.075	0.078	0.044	0.063	0.039	0.052
PAPBCRC	0.050	0.075	0.045	0.063	0.074	0.082	0.045	0.066	0.043	0.059
PAPHCRC	0.050	0.075	0.045	0.058	0.073	0.079	0.044	0.067	0.042	0.054
PAP03RC	0.043	0.078	0.039	0.059	0.078	0.077	0.039	0.065	0.035	0.055

Table 2: Rejection Probabilities of structural break tests for Experiments 6-8

Test	Experiment					
	Sample Size		Sample Size		Sample Size	
	100	200	100	200	100	200
OC	0.324	0.754	0.479	0.901	0.116	0.426
RC	0.198	0.473	0.134	0.274	0.077	0.117
RBFACOC	0.179	0.598	0.411	0.837	0.120	0.473
RBFBCOC	0.197	0.622	0.443	0.852	0.138	0.484
RBFHCOC	0.190	0.613	0.429	0.848	0.131	0.478
LOGACOC	0.179	0.601	0.408	0.831	0.129	0.496
LOGBCOC	0.194	0.607	0.422	0.845	0.141	0.498
LOGHCOC	0.192	0.600	0.421	0.837	0.135	0.503
PAPACOC	0.176	0.591	0.401	0.829	0.122	0.495
PAPBCOC	0.185	0.603	0.425	0.851	0.137	0.497
PAPHCOC	0.183	0.597	0.414	0.838	0.128	0.496
PAP03OC	0.176	0.600	0.401	0.839	0.123	0.498
RBFACRC	0.291	0.533	0.169	0.323	0.047	0.122
RBFBCRC	0.323	0.592	0.186	0.394	0.070	0.132
RBFHCRC	0.313	0.575	0.179	0.361	0.066	0.128
LOGACRC	0.220	0.471	0.150	0.321	0.062	0.114
LOGBCRC	0.250	0.546	0.174	0.389	0.071	0.138
LOGHCRC	0.245	0.514	0.167	0.367	0.065	0.132
PAPACRC	0.231	0.523	0.162	0.371	0.060	0.132
PAPBCRC	0.248	0.540	0.174	0.389	0.072	0.138
PAPHCRC	0.239	0.530	0.168	0.376	0.065	0.134
PAP03RC	0.197	0.499	0.149	0.363	0.064	0.136

Table 3: Rejection Probabilities of structural break tests for Experiments 9-11

Test	Experiment					
	Sample Size		Sample Size		Sample Size	
	100	200	100	200	100	200
OC	0.540	0.940	0.722	0.993	0.230	0.691
RC	0.256	0.602	0.148	0.401	0.078	0.129
RBFACOC	0.363	0.871	0.657	0.981	0.276	0.762
RBFBCOC	0.392	0.882	0.679	0.986	0.287	0.768
RBFHCOC	0.380	0.878	0.671	0.984	0.283	0.767
LOGACOC	0.378	0.873	0.669	0.979	0.311	0.776
LOGBCOC	0.391	0.883	0.680	0.981	0.318	0.792
LOGHCOC	0.386	0.878	0.673	0.983	0.308	0.789
PAPACOC	0.368	0.876	0.659	0.981	0.294	0.778
PAPBCOC	0.385	0.881	0.676	0.983	0.304	0.788
PAPHCOC	0.376	0.879	0.666	0.983	0.300	0.785
PAP03OC	0.358	0.882	0.661	0.983	0.287	0.780
RBFACRC	0.405	0.730	0.243	0.546	0.079	0.140
RBFBCRC	0.452	0.808	0.283	0.618	0.090	0.168
RBFHCRC	0.438	0.789	0.275	0.599	0.080	0.154
LOGACRC	0.322	0.653	0.231	0.522	0.091	0.165
LOGBCRC	0.332	0.725	0.255	0.593	0.093	0.190
LOGHCRC	0.329	0.711	0.244	0.583	0.103	0.186
PAPACRC	0.322	0.708	0.240	0.588	0.092	0.182
PAPBCRC	0.337	0.737	0.242	0.602	0.092	0.198
PAPHCRC	0.331	0.728	0.245	0.597	0.093	0.188
PAP03RC	0.298	0.674	0.222	0.587	0.078	0.186

Table 4: Rejection Probabilities of structural break tests for Experiments 12-14

Test	Experiment					
	Sample Size		Sample Size		Sample Size	
	100	200	100	200	100	200
OC	0.418	0.852	0.584	0.949	0.145	0.556
RC	0.270	0.533	0.175	0.341	0.075	0.122
RBFACOC	0.231	0.709	0.503	0.894	0.189	0.611
RBFBCOC	0.259	0.734	0.526	0.907	0.209	0.633
RBFHCOC	0.249	0.728	0.517	0.903	0.194	0.622
LOGACOC	0.239	0.712	0.509	0.904	0.199	0.640
LOGBCOC	0.253	0.728	0.525	0.907	0.201	0.654
LOGHCOC	0.249	0.725	0.517	0.909	0.195	0.648
PAPACOC	0.233	0.716	0.489	0.896	0.179	0.643
PAPBCOC	0.245	0.727	0.519	0.909	0.187	0.656
PAPHCOC	0.238	0.720	0.499	0.902	0.177	0.644
PAP03OC	0.237	0.720	0.495	0.904	0.179	0.644
RBFACRC	0.348	0.594	0.221	0.425	0.076	0.128
RBFBCRC	0.396	0.665	0.237	0.480	0.092	0.153
RBFHCRC	0.383	0.648	0.233	0.464	0.084	0.146
LOGACRC	0.282	0.575	0.193	0.420	0.076	0.140
LOGBCRC	0.313	0.642	0.225	0.483	0.078	0.167
LOGHCRC	0.307	0.628	0.218	0.469	0.083	0.157
PAPACRC	0.292	0.607	0.211	0.467	0.082	0.155
PAPBCRC	0.314	0.640	0.223	0.485	0.080	0.162
PAPHCRC	0.303	0.627	0.214	0.475	0.079	0.165
PAP03RC	0.252	0.585	0.200	0.451	0.077	0.153

Table 5: Rejection Probabilities of structural break tests for Experiments 15-17

Test	Experiment					
	Sample Size		Sample Size		Sample Size	
	100	200	100	200	100	200
OC	0.563	0.893	0.645	0.940	0.201	0.515
RC	0.277	0.496	0.192	0.384	0.088	0.153
RBFACOC	0.187	0.570	0.384	0.726	0.157	0.406
RBFBCOC	0.218	0.585	0.445	0.735	0.192	0.412
RBFHCOC	0.204	0.568	0.408	0.729	0.169	0.408
LOGACOC	0.223	0.669	0.502	0.831	0.200	0.518
LOGBCOC	0.249	0.705	0.546	0.842	0.224	0.534
LOGHCOC	0.238	0.689	0.526	0.830	0.208	0.533
LOGNSOC	0.223	0.638	0.488	0.815	0.174	0.488
PAPACOC	0.219	0.661	0.495	0.839	0.191	0.512
PAPBCOC	0.260	0.710	0.534	0.850	0.215	0.549
PAPHCOC	0.232	0.690	0.513	0.841	0.201	0.525
PAP03OC	0.227	0.718	0.491	0.830	0.176	0.493
RBFACRC	0.346	0.428	0.206	0.294	0.065	0.095
RBFBCRC	0.442	0.608	0.293	0.408	0.095	0.123
RBFHCRC	0.399	0.528	0.244	0.344	0.075	0.109
LOGACRC	0.239	0.383	0.175	0.344	0.066	0.123
LOGBCRC	0.276	0.502	0.202	0.401	0.069	0.134
LOGHCRC	0.261	0.458	0.192	0.383	0.068	0.136
LOGNSRC	0.243	0.419	0.193	0.341	0.062	0.138
PAPACRC	0.253	0.488	0.190	0.391	0.068	0.122
PAPBCRC	0.271	0.510	0.195	0.402	0.067	0.131
PAPHCRC	0.262	0.500	0.191	0.398	0.068	0.128
PAP03RC	0.240	0.480	0.193	0.387	0.062	0.125