The University of Manchester

Economics Discussion Paper Series EDP-0724

Weighted smooth transition regressions

Ralf Becker
Denise Osborn

November 2007

Economics
School of Social Sciences
The University of Manchester
Manchester M13 9PL

Weighted Smooth Transition Regressions

Ralf Becker and Denise R Osborn*
Economics, School of Social Sciences, University of Manchester

June 28, 2007

Abstract

A new procedure is developed for modelling and testing nonlinearity of a smooth transition form, allowing the possibility that the transition variable is a weighted function of lagged observations. This is achieved through use of a beta function and requires specification of only the maximum permissable lag. Nonlinearity testing uses a search over the beta function parameters, with inference explicitly reconising these are unidentified under the null hypothesis. A wild boostrap procedure is recommended to allow for heteroscedasticity of unknown form, with a Monte Carlo study showing this to perform well even for a homoscedastic DGP. Estimation issues are also discussed. An application to the yield curve as a predictor of quarterly UK growth illustrates the usefulness of the procedure for modelling data of mixed frequencies.

Corresponding author
Ralf Becker
Economics, School of Social Sciences
University of Manchester
Manchester M13 9PL, UK

email `ralf.becker@manchester.ac.uk`
Ph +44 (0)161 275 4807.

Keywords: Nonlinear regression; nonlinearity testing; smooth transition; regime-switching; wild bootstrap

*The authors gratefully acknowledge comments from Heather Anderson, Dick van Dijk and participants at SNDE2007 conference that have helped in the development of this paper.

1 Introduction

Nonlinear models play an increasingly important role in the analysis of observed economic and financial time series. The state-dependence allowed by these models is an attractive feature, since it is often plausible that the nature of economic responses may depend on underlying conditions, such as the state of the business cycle, the monetary policy stance of the central bank or conditions in financial markets. Indeed, the popular class of nonlinear threshold models exploits this state-dependence through the use of regimes, with the regime applying in any specific time period defined by the value of the (so-called) transition variable in relation to one or more thresholds. A popular specification from this class is the smooth transition regression (STR) model, for which Teräsvirta (1994) provides a coherent modelling strategy in the univariate smooth transition autoregression ($STAR$) context which is generalized in Teräsvirta (1998) to the STR case. The many examples of recent applications of STR models include Anderson and Vahid (2001), Fok, van Dijk and Franses (2005), Sensier, Osborn and Öcal (2002) and Taylor, Peel and Sarno (2001), while van Dijk, Teräsvirta and Franses (2002) review recent developments.

A crucial issue in applying threshold models, including those of the STR type, is the specification of the transition variable whose value determines the regime applying in the current period. In practice, the form employed is almost invariably either a single lag (often referred to as the delay) or a simple transformation of lags (such as using an annual difference in a model for quarterly or monthly fluctuations), with many STR applications following the recommendation of Teräsvirta (1994) to select the delay based on linearity test statistics computed for a range of lags. On the other hand, rather than selecting a single transition variable lag prior to estimation, a number of potential STR models may be estimated and selection between them deferred to a later stage of the analysis (van Dijk *et al.* 2002). Nevertheless, the retention of models based on a number of candidate lags indicates that each potentially contains some information about the regimes and hence more general specifications for the transition variable might be appropriate. Medeiros and Viega (2003, 2005) allow the possibility that the transition variable may be an unknown linear function of multiple lags. However, the resulting procedure is fairly complicated and the "holes" [gaps] that may result in the lags that enter the transition function could lead to regime-switching behaviour that is implausible in contexts like the business cycle.

By employing a beta function over potential lags, our approach simplifies STR model specification because the only transition function lag that needs to be specified is the maximum lag that can enter this function. The use of the beta function delivers a transition variable that is a weighted function of past observations, which has the attractive implication that the current regime is defined as a smooth function of these observations over time. Hence we refer to the model as a $WSTR$ (weighted STR) specification. Although the $WSTR$ model requires estimation of one additional parameter compared to procedures that estimate the (single lag) delay through a search procedure, we believe that this cost is minimal in relation to the added flexibility it delivers[1]. Our approach is also more parsiminous than that of Medeiros and Veiga (2003, 2005), while avoiding their model specification procedure and the potentially implausible regime-switching behaviour the model may imply. Although the Medeiros and Veiga specification allows different signs on lagged values in the transition function whereas ours does not, our approach might be preferred in the many situations in economics or finance where it is natural to consider a transition variable in terms of a weighted average of lagged values.

Our $WSTR$ specification is developed from the mixed data frequency $MIDAS$ approach of Ghysels, Santa-Clara and Valkanov (2005, 2006), which has also been used recently by Galvão (2006) in a STR context. However, Galvão focuses on the use of high frequency data for forecasting a lower frequency variable, whereas our context is the more general one of STR model specification and nonlinearity testing. Nevertheless, the $WSTR$ model is applicable in a mixed frequency context,

[1] This view is supported by the evaluation of forecasts produced from $WSTR$ models in Becker and Osborn (2007).

as illustrated in our application of Section 5 that examines the relationship of quarterly UK GDP growth to monthly yield curve information.

Nonlinearity testing prior to STR modelling is also a problematic issue. Although the smallest p-value obtained from a search over potential transition variables cannot be readily interpreted in terms of a test of the linearity null hypothesis, such p-values are widely reported. The approach of Luukkonen, Saikkonen and Teräsvirta (1988) provides an asymptotically valid test, but this is often claimed to suffer from lack of power due to the large number of additional coefficients that typically enter the test. Overparameterization issues can also affect the related V23 test of Teräsvirta, Lin and Granger (1993), which they propose as a test against a neural network model and is also suggested by Medeiros and Veiga (2005) in their flexible lag $STAR$ context. Accounting for heteroscedasticity is a further issue for empirical modellers, which is important not only when analyzing financial data but also for macroeconomic time series (see, for example, Sensier and van Dijk, 2004). To date, however, accounting for heteroscedasticity when testing for STR nonlinearity has been problematic, since Lundberg and Teräsvirta (1998) find that robustification can remove most of the power of the test. Consequently, van Dijk *et al.* (2002), for example, recommend that heteroscedasticity-robust nonlinearity tests should not be applied although Becker and Hurn (2007) demonstrate that appropriate bootstrapping techniques can deliver reliable inference. Nevertheless, the widespread failure to consider the oversizing of nonlinearity tests due to heteroscedasticity indicates that many estimated nonlinear models may be spurious.

Based on our $WSTR$ model, we propose a test for the presence of possible nonlinearity through a search over a plausible set of beta function parameters, with the consequences of searching explicitly recognised using the procedures of Hansen (1996). However, rather than following Hansen (1996) by allowing for heteroscedasticity through robust covariance estimation, we advocate the use of the wild boostrap to account for possible heteroscedasticity of unknown form Our results indicate that the wild bootstrap approach performs very well when testing for the presence of nonlinearity, delivering reliable finite sample size and power comparable to that achieved by tests that assume homoscedasticity when the true data generating process is homoscedastic.

The structure of the paper is as follows. Section 2 discusses the $WSTR$ model, with Section 3 then developing our nonlinearity test. Properties of the $WSTR$ test are examined in Section 4 through a Monte Carlo analysis and Section 5 examines an empirical application to the relationship between quarterly UK output (GDP) growth and monthly yield curve information. A concluding section completes the paper, with model estimation issues discussed in an appendix.

2 The $WSTR$ Model

This section briefly reviews STR models, before outlining our proposed $WSTR$ generalisation and the associated weighting functions we propose. Estimation issues are also briefly addressed, with further details in the Appendix.

2.1 STR and $WSTR$ models

For a given transition variable s_t, the STR model may be written as:

$$\begin{aligned} y_t &= \alpha_0 + \mathbf{x}_t \boldsymbol{\alpha}_1 + f(s_t)(\beta_0 + \mathbf{x}_t \boldsymbol{\beta}_1) + \varepsilon_t \\ &= \widetilde{\alpha}_{0t} + \mathbf{x}_t \widetilde{\boldsymbol{\alpha}}_{1t} + \varepsilon_t \end{aligned} \tag{1}$$

where $\mathbf{x}_t$ is a $(1 \times n)$ vector of explanatory variables (typically including p lagged values of y_t), $\boldsymbol{\alpha}_1$ and $\boldsymbol{\beta}_1$ are $(n \times 1)$ parameter vectors and $f(s_t)$ is a smooth function of its scalar argument s_t and ε_t is a zero mean independent process, which is usually assumed to have constant variance σ^2. By defining $\widetilde{\alpha}_{0t} = \alpha_0 + f(s_t)\beta_0$ and $\widetilde{\boldsymbol{\alpha}}_{1t} = \boldsymbol{\alpha}_1 + f(s_t)\boldsymbol{\beta}_1$ the second line of (1) emphasizes that $\widetilde{\alpha}_{0t}$ and $\widetilde{\boldsymbol{\alpha}}_{1t}$ change through time as a function of the transition variable s_t. The logistic transition function

$$f(s_t) = [1 + \exp\{-\gamma(s_t - c)\}]^{-1} \qquad \gamma > 0 \tag{2}$$

is frequently used[2], since this can represent two regimes measured by the value of s_t in relation to the threshold c, with $f(s_t) < 0.5$ for $s_t < c$ and $f(s_t) > 0.5$ for $s_t > c$. As the slope $\gamma \to \infty$, then (2) approaches the threshold model with binary regimes defined by $s_t \lessgtr c$; see, for example, Teräsvirta (1998).

In the univariate $STAR$ case, $\mathbf{x}_t = (y_{t-1}, y_{t-2}, ..., y_{t-p})$ and $s_t = y_{t-k}$. It is frequently assumed that the delay k satisfies $1 \leq k \leq p$, but in general k is unknown. Teräsvirta (1994) proposes specifying this parameter through a sequence of tests for the null hypothesis of a linear specification against the alternative of a logistic $STAR$ ($LSTAR$) model for each value in the pre-specified range for k. The lag which produces the strongest rejection of the null is then used as k. This principle is easily extended to the STR case, in which $\mathbf{x}_t$ contains relevant lags of additional variables and the search for the transition variable s_t may extend over lags of more than one variable.

Although $f(s_t)$ is (in general) a smooth function of s_t, it is not necessarily a smooth function over time. Hence when quarterly or monthly data are used, a single lag may be too noisy to adequately capture regimes which have a duration of (say) a year or more, such as those associated with the business cycle. This has sometimes led researchers to specify s_t in the $LSTR$ specification of (2) as a multi-period growth rate, as in van Dijk, Franses and Paap (2002) or Sensier, Osborn and Öcal (2002). Although this smoothing of observations may lead to regimes that are more easily interpretable in terms of the business cycle, the defintion of s_t is then essentially *ad hoc*.

Therefore, consider the straightforward generalization of (1) that allows the transition variable to be a $(q \times 1)$ vector of lagged values $\mathbf{s}_t = (z_{t-1}, z_{t-2}, ..., z_{t-q})$ on an observed variable z_t and the logistic transition function (2) is generalized to

$$f(\mathbf{s}_t) = [1 + \exp\{-\gamma(\mathbf{s}_t\boldsymbol{\delta} - c)\}]^{-1} \qquad \gamma > 0 \tag{3}$$

[2]Much of our discussion can also be applied to other forms of the transition function, such as the widely-used exponential STR specification. However, we focus on the logistic case for expositional purposes.

and $\boldsymbol{\delta} = (\delta_1, \delta_2, ..., \delta_q)$ is a $(q \times 1)$ vector of parameters on which sufficient restrictions are applied in order to ensure that the parameters are identified. The $LSTR$ special case of (1) and (2) with $s_t = z_{t-k}$ can be represented in this more general form with

$$\boldsymbol{\delta} = \mathbf{e}_k, \tag{4}$$

where $\mathbf{e}_k$ is the kth column of a q dimensional identity matrix. However, any linear transformation is permitted through an appropriate definition of $\boldsymbol{\delta}$ in (3); for example, a transition variable that is the average of the first four lags of z_t implies $\boldsymbol{\delta} = (0.25, 0.25, 0.25, 0.25, 0, ..., 0)$.

Medeiros and Viega (2003, 2005) adopt the framework of (3) as the starting point for their univariate $NCSTAR$ (neuro coefficient $STAR$) model, assuming no *a priori* restrictions on $\boldsymbol{\delta}$ beyond the maximum lag order p and the identification conditions $||\delta_i|| = 1$ and $\delta_i > 0$ for a specific i. However, in practice not all elements of $\mathbf{s}_t = \mathbf{x}_t = (y_{t-1}, y_{t-2}, ..., y_{t-p})'$ are necessarily included in the estimated transition function. In the spirit of Teräsvirta (1994), Medeiros and Veiga (2003, 2005) propose that the relevant lags should be selected by applying linearity tests for all possible subsets of the elements of $\mathbf{s}_t$ and choosing the subset which produces the strongest rejection (according to the p-value). Then model estimation includes the implied nonzero coefficients of $\boldsymbol{\delta}$ entering (3), in addition to γ and c. Although they discuss a univariate model, the generalization is straightforward to the case where the transition variable $\mathbf{s}_t$ is weakly exogenous. Their approach adds substantial flexibility to the transition function compared to the more commonly used single lag case, but it does not ensure that the resulting estimated $f(\mathbf{s}_t)$ is a smooth function of lagged observations, and hence may lead to implausible implied regime changes for y_t. Further, since estimation of the transition function parameters γ and c in (1) has sometimes proved to be rather difficult in practice for macroeconomic data, it is likely that only a very small number of parameters could be estimated in $\boldsymbol{\delta}$ in (3) with such data.

Our approach is also based on the model defined by (1) and (3), but we prefer to impose restrictions on the values of the elements of $\boldsymbol{\delta}$ at the outset of the analysis. The conditions $\sum_{i=1}^{q} \delta_i = 1$ with $\delta_i \geqslant 0$ $(i = 1, ..., q)$ ensure identification of the parameters, while also giving the interpretation of $\boldsymbol{\delta}$ as a weighting function of the elements of $\mathbf{s}_t$. The recent development of the $MIDAS$ methodology to deal with data sampled at different frequencies has produced a resurgence of interest in parsimonious weighting functions[3] and Ghysels *et al.* (2005, 2006) propose the weighting function

$$\delta_i(\kappa_1, \kappa_2) = \frac{g(i\,|q; \kappa_1, \kappa_2)}{\sum_{j=1}^{q} g(j\,|q; \kappa_1, \kappa_2)}, \qquad i = 1, 2, ..., q \tag{5}$$

where $g(i\,|q; \kappa_1, \kappa_2)$ is the density function of the beta distribution used to calculate the ith weight δ_i, q is the maximum lag length considered and κ_1, κ_2 are parameters to be estimated. The weights (5) computed from the beta distribution are well suited for this purpose, as they can take a range of plausible shapes, as discussed in the next subsection. Although specification of a maximum lag is required, this is common to all procedures in the realistic case where the delay in the STR model is unknown.

An important advantage of our $WSTR$ model based on (5) is that, by introducing one extra parameter (κ_1 and κ_2 as compared to the selection of a single lag k), very flexible and parsiminous weighting functions are obtained. Further, this avoids the step required in Teräsvirta (1994) for specification of the single delay k or in Medeiros and Veiga (2003, 2005) for the selection of the subset of variables whose coefficients are to be estimated in (3).

2.2 Weight distributions

As noted above, the usual STR modelling strategy imposes a weight vector $\boldsymbol{\delta}$ in (3) that assigns all weight to one lag and it is therefore important that an apparently more general function is capable of reproducing this case. As a density function for a continuous random variable, the use of a beta distribution in (5) cannot place unit weight on a single lag. Nevertheless, depending on the parameter values κ_1 and κ_2, the weights derived from the distribution may be concentrated

[3] An early example of a parsimonious weighting function is the Koyck lag.

around a single lag, and hence approximate a STR data generating process (DGP) with single-lag transition function of the form $f(z_{t-k})$.

Some applications choose the transition variable as an average of past values or, when the observed variable is nonstationary, an m-period change when the dependent variable is the first difference (as in Skalin and Teräsvirta, 2002, Teräsvirta *et al.*, 2004, or Sensier *et al.*, 2002)[4]. This implies equal weights over the lags included in $\mathbf{s}_t$ and is reproduced by the beta distribution with $\kappa_1 = \kappa_2 = 1$.

The shape of some possible weight distributions deriving from (5) is illustrated in Figures 1 and 2 for the case of quarterly data with $q = 4$ and $q = 8$ respectively; these shapes are obtained using the parameters for the beta distribution as shown in Table 1. The upper panel of each figure shows five weight distributions with weights distributed over all lags, either equally or with modal weight at lag one or two. The weight distributions displayed in the lower panel in each case mimic the restrictions in a traditional STR model where $\boldsymbol{\delta}$ has the form of a single delay at one of $k = 1, 2, 3, 4$. In these latter cases about 90% of the weight is attached to a single lag, hence providing a good approximation in practice.

The shapes in Figures 1 and 2 are, of course, only for illustrative purposes. Different values of the beta function parameters κ_1 and κ_2 give rise to different shapes; for example, the modal weight could occur at a longer lag than in the cases illustrated. Nevertheless, these figures indicate that the functions represented by (5) can capture the features of weight distributions likely to apply in economic applications. In addition to this flexibility, $WSTR$ models have the substantial advantage over other STR specifications that the relevant lags are selected endogenously, constrained merely by the weighting functions allowed by the beta distribution and the maximum lag specified by the researcher.

[4] Note that the m-period change is merely a rescaled version of a simple average over m one-period changes.

2.3 Estimation

For a given transition variable $s_t = z_{t-k}$, the parameter vector $\theta = \left(\alpha_0,\ \boldsymbol{\alpha}_1',\ \beta_0,\ \boldsymbol{\beta}_1',\ \gamma,\ c\right)'$ of the STR model in (1) and (2) is estimated by nonlinear least squares. However, since OLS can be utilized conditional on γ and c (provided $\gamma \neq 0$), the parameter vector $\boldsymbol{\theta}_1 = (\alpha_0, \boldsymbol{\alpha}_1', \beta_0, \boldsymbol{\beta}_1')'$ can be concentrated out of the nonlinear criterion function. Therefore, nonlinear optimisation only needs to be undertaken with respect to γ and c, with estimates of the remaining coefficients then recovered by OLS conditional on the parameters of (2). A procedure of this type is also proposed by Medeiros and Veiga (2005) for their $NCSTAR$ model.

We also advocate estimating the $WSTR$ model by optimising the criterion function (the residual sum of squares) concentrated with respect to $\boldsymbol{\theta}_1$, so that nonlinear optimisation is performed only over the elements of $\boldsymbol{\theta}_2 = (\ \gamma,\ c,\ \kappa_1,\ \kappa_2)'$ which are used in (5) to define the elements of the weighting vector $\boldsymbol{\delta}$ in (3). The required starting values for $\boldsymbol{\theta}_2$ can be obtained as a by-product of the nonlinear testing procedure outlined below or from an initial $LSTR$ model grid search. Further discussion can be found in the appendix.

3 Nonlinearity Testing

As nonlinear models are more difficult to estimate and use than linear ones, it is widely recognised that appropriate tests should be performed to establish the presence of nonlinearity of the form to be modelled prior to estimation. This section considers testing for nonlinearity of the $WSTR$ form, as a by-product also discussing appropriate testing in the context of the conventional single lag STR special case. We first discuss general issues concerned with the application of tests based on Taylor series approximations, before considering our approach that employs a range of plausible weight functions. Heteroscedastic-consistent tests are discussed in the third subsection.

3.1 Taylor series approximations

The testing procedure for smooth transition models based on Taylor series approximations is laid out in Teräsvirta (1994) and Luukkonen, Saikkonen and Teräsvirta (1988). Although the model may be univariate, our discussion of testing for STR nonlinearity in (1) considers the bivariate case with $\mathbf{x}_t = (y_{t-1}, ..., y_{t-p}, z_{t-1}, ..., z_{t-r})'$ and $n = p + r$, since this is sufficient to illustrate more general models. Further, we consider a leading indicator context and assume the transition variable is a linear combination of the elements of $\mathbf{s}_t = (z_{t-1},\ z_{t-2},\ ...,\ z_{t-q})$; the modifications required for $\mathbf{s}_t = (y_{t-1},\ y_{t-2},\ ...,\ y_{t-q})$, including the univariate case, are straightforward. Note that we do not require $r = q$, so that the maximum lag q potentially entering the transition function can differ from that for z_t in (1).

The logistic smooth transition function of (3) requires definition of the vectors $\boldsymbol{\delta}$ and $\mathbf{s}_t$. For the purpose of derivation, it is convenient to define the scalar $s_t^* = \gamma\,(\mathbf{s}_t\boldsymbol{\delta} - c)$ and, without loss of generality, centre the logistic function as

$$f(s_t^*) = [1 + \{\exp(-s_t^*)\}]^{-1} - 0.5 \tag{6}$$

so that $f(0) = 0$. In their derivation, Luukkonen *et al.* (1988) do not specify a particular functional form for $f(s_t^*)$ but rather formulate a number of conditions this function needs to fulfill when it is monotonically increasing[5].

The null hypothesis of linearity can be represented by $H_0 : \gamma = 0$ in (3), with the process following a nonlinear path if $H_A : \gamma \neq 0$. Standard distributions for common tests (like Wald, LR and LM) do not apply as under this null hypothesis the parameters $\boldsymbol{\delta}$ and c are unidentified. The ingenious contribution by Luukkonen *et al.* (1988) is to replace the transition function by a Taylor Series approximation around $f\,(0)$. A third order approximation has frequently been used in the

[5] Although this rules out the possibility of an exponential STR specification, analogous results apply for this case; see, for example, Teräsvirta (1994).

literature, namely

$$f_{T3}\left(s_t^*\right) = f\left(0\right) + f'\left(0\right) s_t^* + \frac{1}{2} f''\left(0\right) \left(s_t^*\right)^2 + \frac{1}{6} f'''\left(0\right) \left(s_t^*\right)^3.$$

The conditions imposed on the functional form of $f\left(s_t^*\right)$ ensure that $f\left(0\right) = f''\left(0\right) = 0$ and consequently

$$f_{T3}\left(s_t^*\right) = f'\left(0\right) s_t^* + \frac{1}{6} f'''\left(0\right) \left(s_t^*\right)^3.$$

Replacing $f\left(s_t\right)$ in (1) with $f_{T3}\left(s_t^*\right)$ yields the approximation

$$y_t = \alpha_0 + \mathbf{x}_t \boldsymbol{\alpha}_1 + \left(f'\left(0\right) s_t^* + \frac{1}{6} f'''\left(0\right) \left(s_t^*\right)^3\right) \left(\beta_0 + \mathbf{x}_t \boldsymbol{\beta}_1\right) + \varepsilon_t \tag{7}$$

where it should be noted that ε_t now also includes the approximation error due to $f_{T3}\left(s_t^*\right) \neq f\left(s_t^*\right)$ and the values of the parameters in $(\alpha_0,\, \boldsymbol{\alpha}_1^{'})'$ in this last expression differ from those in (1) due to the centering in (6)[6].

Now, substituting $s_t^* = \gamma\left(\mathbf{s}_t \boldsymbol{\delta} - c\right)$ in (7), then rearranging and collecting terms leads to

$$y_t = \phi_0 + \mathbf{x}_t \boldsymbol{\phi}_1 + \phi_2 (\mathbf{s}_t \boldsymbol{\delta}) \mathbf{x}_t \boldsymbol{\beta}_1 + \phi_3 \left(\mathbf{s}_t \boldsymbol{\delta}\right)^2 \mathbf{x}_t \boldsymbol{\beta}_1 + \phi_4 \left(\mathbf{s}_t \boldsymbol{\delta}\right)^3 \mathbf{x}_t \boldsymbol{\beta}_1 + \varepsilon_t \tag{8}$$

where the scalar parameters ϕ_i $(i = 1, 2, 3, 4)$ are functions of $\alpha_0, \boldsymbol{\alpha}_1, \beta_0, \gamma$ and $\boldsymbol{\delta}$, in addition to scalar factors from the Taylor expansion and from the derivatives of the transition function at $s_t^* = 0$[7]. Further, it is easy to demonstrate that γ enters ϕ_0 and $\boldsymbol{\phi}_1$ additively and ϕ_i for $i = 2, 3, 4$ multiplicatively. This implies that under $H_0 : \gamma = 0$ all $\phi_i = 0$ for $i = 2, 3, 4$, which indicates that the null hypothesis can potentially be tested by testing the restriction $\phi_i = 0$ for $i = 2, 3, 4$ in an auxiliary regression. However, specifying the appropriate auxiliary regression requires investigation of the nature of the terms $\left(\mathbf{s}_t \boldsymbol{\delta}\right)^2$ and $\left(\mathbf{s}_t \boldsymbol{\delta}\right)^3$ in (8). Further, since the vector $\boldsymbol{\beta}_1$ is unknown, the tests apply to $\phi_i \boldsymbol{\beta}_1 = 0$ for $i = 2, 3, 4$.

Many researchers repeatedly apply this nonlinearity test for different possible delay parameters $k = 1, ..., q$, in (4). The value of k which triggers the strongest rejection of the null hypothesis

[6] For notational ease we do not introduce different notations to reflect these changes.

[7] Note that terms in $(\mathbf{s}_t \boldsymbol{\delta})$, $(\mathbf{s}_t \boldsymbol{\delta})^2$, $(\mathbf{s}_t \boldsymbol{\delta})^3$ contribute to $\mathbf{x}_t \boldsymbol{\phi}_1$, $(\mathbf{s}_t \boldsymbol{\delta}) \mathbf{x}_t \boldsymbol{\phi}_1$, $(\mathbf{s}_t \boldsymbol{\delta})^2 \mathbf{x}_t \boldsymbol{\phi}_1$ respectively.

is selected as the delay used when estimating (1). However, if applied as a test for the existence of nonlinearity, this procedure suffers from the multiple testing problem. In other words, if the overall null hypothesis of linearity is rejected when (at least) one individual test for some delay k is rejected, the true level of significance for the overall test may be considerably higher than the nominal level[8]. In fact, Luukkonen *et al.* (1988) develop a joint test which does not suffer from this problem and is also sufficiently general to cover the $WSTR$ model introduced here.

Luukkonen *et al.* (1988) assume the specific form for the vector $\boldsymbol{\delta}$ as in (4), so that scalar $s_t = z_{t-k}$, leading to the STR special case of (1). With the maximum value for the delay specified to be q, this assumption simplifies higher order powers of $(\mathbf{s}_t\boldsymbol{\delta})$ to

$$(\mathbf{s}_t\boldsymbol{\delta})^j = \begin{cases} \sum_{i=1}^{q} \delta_i z_{t-i}^j & \text{for unknown delay parameter } k \\ z_{t-k}^j & \text{for known delay parameter } k. \end{cases} \tag{9}$$

Therefore, when the transition variable is z_{t-k} with delay k known, and eliminating redundant lags, it is easy to see that an auxiliary regression for testing against STR nonlinearity has the form

$$y_t = \phi_0 + \mathbf{x}_t\boldsymbol{\phi}_1 \ + (\mathbf{x}_t z_{t-k})'\boldsymbol{\phi}_2 + (\mathbf{x}_t z_{t-k}^2)'\boldsymbol{\phi}_3 + (\mathbf{x}_t z_{t-k}^3)'\boldsymbol{\phi}_4 + \varepsilon_t \tag{10}$$

where $\boldsymbol{\phi}_2, \boldsymbol{\phi}_3, \boldsymbol{\phi}_4$ are each $(n\times 1)$ vectors. The test statistic for the $3n$ restrictions $\boldsymbol{\phi}_2 = \boldsymbol{\phi}_3 = \boldsymbol{\phi}_4 = \mathbf{0}$ can be computed using a conventional asymptotic χ^2 distribution, or an F distribution which may better take account of the finite sample size.

When the restriction (4) applies but the delay parameter k is unknown except for the maximum permitted lag q, the z_{t-k}^j $(j = 1, 2, 3)$ terms in (10) are replaced by sums, leading to the auxilliary regression of the Luukkonen *et al.* (1988) test which is

$$y_t = \phi_0 + \mathbf{x}_t\boldsymbol{\phi}_1 + \sum_{k=1}^{q}(\mathbf{x}_t z_{t-k})\boldsymbol{\phi}_{2,k} + \sum_{k=1}^{q}(\mathbf{x}_t z_{t-k}^2)\boldsymbol{\phi}_{3,k} + \sum_{k=1}^{q}(\mathbf{x}_t z_{t-k}^3)\boldsymbol{\phi}_{4,k} + \varepsilon_t \tag{11}$$

When $q = r$, the number of null hypothesis restrictions $\boldsymbol{\phi}_{2,k} = \boldsymbol{\phi}_{3,k} = \boldsymbol{\phi}_{4,k} = 0$ is $3qn - [q(q-1)/2]$,

[8] Teräsvirta (1994) is well aware that such a procedure results in a test statistic with a non-standard distribution. Rather, he discusses this strategy only as a tool for model selection. Nevertheless, practitioners often apply this strategy in a testing context and Section 4 investigates its statistical properties in this context.

and it is evident that the number of restrictions can be large relative to the sample sizes often available[9].

If the transition variable is an unknown linear combination of lags of z_t, then $(\mathbf{s}_t\boldsymbol{\delta})^j$ in (9) produces all possible cross products of order j between the lags z_{t-1}, $z_{t-2}, ..., z_{t-q}$, namely

$$(\mathbf{s}_t\boldsymbol{\delta})^2 = \sum_{k=1}^{q}\sum_{k'=k}^{q} \lambda_{i,i'} z_{t-k}\, z_{t-k'} \tag{12}$$

$$(\mathbf{s}_t\boldsymbol{\delta})^3 = \sum_{k=1}^{q}\sum_{k'=k}^{q}\sum_{k''=k'}^{q} \lambda_{k,k',k''} z_{t-k}\, z_{t-k'}\, z_{t-k''} \tag{13}$$

where the coefficients $\lambda_{k,k'}$ and $\lambda_{k,k',k''}$ are functions of the δ_i. For $q = r = 4$, for instance, this generates 10 distinct terms of order 2. As in (8), these terms are multiplied with $\mathbf{x}_t\boldsymbol{\beta}_1$ so that the inflation of cross-product continues and it becomes apparent that without restricting the parameter vector $\boldsymbol{\delta}$, no sensible testing strategy based on auxiliary regressions and full third-order Taylor approximations appears to be feasible, except for very small values of q.

One way to mitigate this problem is to employ an "economy" version of the test, which is suggested by Luukkonen *et al.* (1988) with the single-lag restriction of (4) and by Medeiros and Veiga (2005) when these restrictions are not imposed. This economy version then employs the terms of (12) and (13) arising from the Taylor series approximation of $f(\mathbf{s}_t)\beta_0$, but includes terms from only a first order Taylor series approximation in the expansion of $f(\mathbf{s}_t)\mathbf{x}_t\boldsymbol{\beta}_1$ in (1). Thus, with no restrictions on $\boldsymbol{\delta}$, the economy version amounts to using the auxiliary regression

$$y_t = \phi_0 + \mathbf{x}_t\boldsymbol{\phi}_1 + \sum_{k=1}^{q}(\mathbf{x}_t z_{t-k})\boldsymbol{\phi}_{2,k} + \sum_{k=1}^{q}\sum_{k'=k}^{q}\sum_{k''=k'}^{q} \phi_{3,k,k',k''} z_{t-k}\, z_{t-k'}\, z_{t-k''} + \varepsilon_t. \tag{14}$$

with the null hypothesis involving a test of $\boldsymbol{\phi}_{2,k} = \mathbf{0}, \phi_{3,k,k',k''} = 0$. For the univariate case $p = 2$ considered in the Monte Carlo analysis of Medeiros and Veiga (2005), this economy version implies testing only 7 restrictions, but (even in a univariate model) the number mushrooms to 30 should $p = q = 4$ be contemplated and to 156 for $p = q = 8$! Although not mentioned by Medeiros and

[9]Note that $q(q-1)/2$ terms in $\sum_{k=1}^{q}(\mathbf{x}_t z_{t-k})'\boldsymbol{\phi}_{2,k}$ are then redundant as terms involving $z_{t-k}z_{t-k'}$ $(k \neq k')$ appear twice in this sum. Also note that, in both (10) and (11), different numbers of restrictions apply when $q \neq r$ or when $\mathbf{s}_t = (y_{t-1}, ..., y_{t-q})$.

Veiga, the test arising from (10) is the same as that referred to by Teräsvirta, Lin and Granger (1993) as the V23 test and recommended by them as a test against nonlinearity of the neural network type[10].

Again in the neural network context, Lee, White and Granger (1993) devise a linearity test where they solve the unidentified parameter problem by effectively taking repeated random draws for possible $\boldsymbol{\delta}$, applying the test for each of these realisations and then obtaining significance levels by means of Bonferroni or Hochberg bounds[11]. However, this test is not widely applied for economic data, apparently for two reasons. Firstly, the parameter space of possible $\boldsymbol{\delta}$ may not be well defined and a large number of random draws may be required to have a good chance of covering all directions against which the test should have power. Secondly, there is little evidence on the empirical performance of the significance bounds using sample sizes typically available in economics and the suspicion is that they are rather conservative. Consequently this strategy is rarely applied in economics and neural network tests are usually conducted through the V23 test.

From this discussion it should be obvious that for most practical purposes, especially when the possible maximum lag q of the transition function is not very small, restrictions are desirable for the vector $\boldsymbol{\delta}$. Fortunately the $WSTR$ model introduced above provides such a set of restrictions, while being much more flexible than (4) at the cost of only one more parameter. Rather than applying a Taylor series approximation to the transition function (3), our proposal is to make explicit use of the restrictions implied by the model.

3.2 Inference using weight functions

As evident from (12) and (13), to develop a test for $WSTR$ nonlinearity based on a Taylor series approximation, the parameters κ_1 and κ_2 of the beta function (5) need to be specified. We believe

[10] The one layer neural network alternative considered there is a special case of the weighted $STAR$ model considered here with $\boldsymbol{\beta}_1 = 0$.

[11] If the test is performed m times then the Bonferroni bound is used in the following manner. Order the obtained p-values in increasing order, $p_{(1)}, p_{(2)}, ..., p_{(m)}$. Assume that a significance level of α is to be applied, then the H_0 is rejected if the smallest of the m p-values, $p_{(1)}$, is smaller than α/m. When the Hochberg bound is applied then the null hypothesis is rejected if for any $j = 1, ..., m$, $p_{(j)} < \alpha/(m - j + 1)$.

that a range of combinations for κ_1 and κ_2 in (5) can be used to capture the features of weight distributions relevant for modelling the responses of macroeconomic variables in a $WSTR$ context. Indeed, examples of such weight distributions have already been discussed in relation to Figures 1 and 2. Our proposal is that a set of plausible weight functions be defined *a priori* for a specific case and a test for nonlinearity based on a Taylor series expansion be applied in relation to each pair of κ_1 and κ_2 values. The overall nonlinearity hypothesis test is then conducted using a bootstrap procedure. By initially defining plausible pairs of κ_1, κ_2, we avoid the overparameterization inherent in the Luukkonen *et al.* (1988) and the V23 tests. On the other hand, our procedure does not involve estimation of the nonlinear $WSTR$ model until after rejection of linearity, thereby avoiding the heavy computational burden of conducting bootstrap inference by estimating nonlinear models when the linearity null hypothesis is true, as in Galvão (2006).

Given values for the beta function parameters, say κ_{1i}, κ_{2i}, yield a vector $\boldsymbol{\delta}_i$, from which the scalar value for the transition variable $s_{t,i} = \mathbf{s}_t\boldsymbol{\delta}_i$ can be calculated for each time period t. Then replacing $\mathbf{s}_t\boldsymbol{\delta}$ in (8) with $s_{t,i}$ yields an auxiliary regression of the form:

$$y_t = \phi_0 + \mathbf{x}_t\boldsymbol{\phi}_1 + (s_{t,i}\mathbf{x}_t)\boldsymbol{\phi}_2 + (s_{t,i}^2\mathbf{x}_t)\boldsymbol{\phi}_3 + (s_{t,i}^3\mathbf{x}_t)\boldsymbol{\phi}_4 + \varepsilon_t \tag{15}$$

where each $\boldsymbol{\phi}_j$ $(j = 2, 3, 4)$ is $(n\times1)$. In deriving (15) we continue to assume that $\mathbf{x}_t = (y_{t-1}, ..., y_{t-p}, z_{t-1}, ..., z_{t-r})$ with $\mathbf{s}_t = (z_{t-1},\ z_{t-2},\ ...,\ z_{t-q})$ or $\mathbf{s}_t = (y_{t-1},\ y_{t-2},\ ...,\ y_{t-q})$.

If one is willing to specify a discrete set Δ', a subset of R^q which includes the potential $\boldsymbol{\delta}$ to be considered, the test statistic can be calculated for each element in Δ'. Utilizing weight distributions constrained according to (5) reduces the dimensionality of the weight set to 2, namely the length of the vector $\boldsymbol{\kappa}$. While this still leaves the possibility of an infinite set on R_+^2 with all $\boldsymbol{\kappa} = (\kappa_1, \kappa_2)' \in \Delta$, it is argued above that for many economic applications it may be reasonable to restrict the potential weight vectors to a relatively small set which capture the characteristics *a priori* considered to be plausible for the problem at hand.

Under the null hypothesis of linearity $\boldsymbol{\kappa} = (\kappa_1, \kappa_2)$ are nuisance parameters. It has long been recognised (Davies, 1987, Andrews and Ploberger, 1994, and Hansen, 1996, provide seminal contributions) that the distributions of test statistics which depend on unidentified nuisance parameters are nonstandard. Let $LM(\boldsymbol{\kappa})$ be the lagrange multiplier test statistic[12] for H_0 given $\boldsymbol{\kappa}$. Three procedures to translate $LM(\boldsymbol{\kappa}_i)$, for $\boldsymbol{\kappa}_i \in \Delta'$ and $i = 1, ..., m$, into a single test statistic have been proposed in the literature (Davies, 1987, Andrews and Ploberger, 1994), namely

$$LM^{\max} = \sup_{\boldsymbol{\kappa}_i \in \Delta'} LM(\boldsymbol{\kappa}_i) \tag{16}$$

$$LM^{\exp} = \ln\left(m^{-1}\sum_{i=1}^{m}\exp\left(\frac{LM(\boldsymbol{\kappa}_i)}{2}\right)\right) \tag{17}$$

$$LM^{ave} = m^{-1}\sum_{i=1}^{m} LM(\boldsymbol{\kappa}_i). \tag{18}$$

While $LM(\boldsymbol{\kappa}_i)$ for fixed $\boldsymbol{\kappa}_i$ is asymptotically χ^2 distributed under H_0, none of $LM^{\max}$, $LM^{\exp}$ or LM^{ave} follow a standard distribution. In particular the distribution of these test statistics depends on $E(LM(\boldsymbol{\kappa}_i), LM(\boldsymbol{\kappa}_j))$, $i \neq j$, which prevents the tabulation of critical values, except in limited specific cases (see Andrews, 1993).

Therefore, distributions for the test statistics under the null hypothesis have to be simulated for the specific problem under study. Hansen (1996) proposes such a procedure and applies it to a special case of the STR model described above. With a null hypothesis of linearity, the alternative model considered in Hansen is the self-exciting threshold autoregressive model, which arises when $\boldsymbol{\delta}$ is defined as in (4) and γ in (2) is such that the transition function acts like a step function, γ^{step}. The remaining unidentified parameters are the threshold c and the delay k of z_t that defines the transition variable. Hansen proposes specification of a set Γ from which $\tau = (c, k)$ arise. For any given τ, $f(\mathbf{s}_t) = I[z_{t-k} > c] = I_t$, where $I[\cdot]$ is the indicator function and the auxilliary test regression is

$$y_t = \alpha_0 + \mathbf{x}_t\boldsymbol{\alpha}_1 + \beta_0^* I_t + I_t\mathbf{x}_t\boldsymbol{\beta}_1^* + \varepsilon_t \tag{19}$$

[12] This procedure is universally valid, not merely for the specific case of an LM test.

from which the null hypothesis $\boldsymbol{\beta}^*= \mathbf{0}$, with $\boldsymbol{\beta}^*= \left(\beta_0^* \; \boldsymbol{\beta}_1^{*\prime}\right)'$, is tested. Let $\mathbf{x}_t(\tau) = (1, \mathbf{x}_t, I_t, I_t\mathbf{x}_t)$ and let capitalised matrices $\mathbf{Y}$ and $\mathbf{X}(\tau)$ represent the stacked matrices of observations y_t and $\mathbf{x}_t(\tau)$ respectively, with $M(\tau) = \mathbf{X}(\tau)' \mathbf{X}(\tau)/T$[13]. Define the $n_\alpha = n+1$ and $n_\beta = 2(n+1)$ dimensional parameter vectors $\boldsymbol{\alpha} = (\alpha_0 \; \boldsymbol{\alpha}_1')'$ and $\boldsymbol{\theta} = \left(\boldsymbol{\alpha}' \; \boldsymbol{\beta}^{*\prime}\right)'$, with $\widehat{\boldsymbol{\theta}}(\tau) = \left(\mathbf{X}(\tau)' \mathbf{X}(\tau)\right)^{-1} \mathbf{X}(\tau)' \mathbf{Y}$. Also let $R = (\mathbf{0} \; \mathbf{I})'$ with $\mathbf{0}$ being a $(n_\beta \times n_\alpha)$ matrix of zeros and $\mathbf{I}$ a $(n_\beta \times n_\beta)$ dimensional identity matrix. Further let $\widetilde{\mathbf{e}}$ be the estimated residual vector for the model imposing the null hypothesis, with the score function evaluated at the null hypothesis being $\widetilde{w}_t(\tau) = \mathbf{x}_t(\tau)\,\widetilde{e}_t$. Assuming homoscedasticity, the LM test of $H_0 : \boldsymbol{\beta}^*= \mathbf{0}$ is then calculated according to

$$\begin{aligned} LM_T(\tau) &= T\widehat{\boldsymbol{\theta}}(\tau)' R \left[R'\sigma_{\widetilde{e}}^2 M(\tau)^{-1} R\right]^{-1} R'\widehat{\boldsymbol{\theta}}(\tau) \\ &= T\widehat{\boldsymbol{\theta}}(\tau)' R\, \widetilde{\Omega}^{-1}\, R'\widehat{\boldsymbol{\theta}}(\tau) \qquad (20) \end{aligned}$$

where the T subscript indicates that the test statistic relates to a particular sample of size T and $\widetilde{\Omega}$ is defined as the term in square brackets. These tests are easily modified for the present problem by recognising that $\mathbf{x}_t(\tau)$ consists of $\mathbf{x}_t$ in the auxiliary test regression (15), $\tau = \boldsymbol{\kappa}$ and the parameter vector $\boldsymbol{\beta}^*$ contains all the coefficients appearing in $\boldsymbol{\phi}_2, \boldsymbol{\phi}_3$ and $\boldsymbol{\phi}_4$ in (15). The maximum, exponential and average statistics are computed over the set of all $\boldsymbol{\kappa} = (\kappa_1, \kappa_2)' \in \Delta'$, namely the specific weight functions considered, with inference then conducted using p-values computed as in Hansen (1996).

Hansen (1996) describes a methodology to generate draws from the asymptotic distribution of $LM^{\max}$, $LM^{\exp}$ and LM^{ave} respectively, enabling hypothesis tests of the computed statistics to be undertaken. Heuristically the procedure can be described as follows. Under the null hypothesis $\sqrt{T}R'\widehat{\boldsymbol{\theta}}(\tau)$ is a $(n_\beta \times 1)$ vector which is asymptotically normally distributed with mean 0 and covariance Ω. Under the assumptions presented in Hansen, this covariance is consistently estimated by $\widetilde{\Omega}$. Recalling the standard regression result $R'\widehat{\boldsymbol{\theta}}(\tau) = R'\boldsymbol{\theta}_0(\tau) + R'\left(\mathbf{X}(\tau)'\mathbf{X}(\tau)\right)^{-1} \mathbf{X}(\tau)' \boldsymbol{\varepsilon}$ where $\boldsymbol{\theta}_0(\tau)$ is the true value of the vector $\boldsymbol{\theta}$ and as under the null hypothesis $R'\boldsymbol{\theta}_0(\tau) = \boldsymbol{\beta}^* = 0$, it

[13] The notation is as close as possible to that in Hansen (1996).

then follows that $R'\widehat{\boldsymbol{\theta}}(\tau) = R'\left(\mathbf{X}(\tau)'\mathbf{X}(\tau)\right)^{-1}\mathbf{X}(\tau)'\boldsymbol{\varepsilon}$. Consequently, under the null hypothesis,

$$\sqrt{T}R'\widehat{\boldsymbol{\theta}}(\tau) = R'M(\tau)^{-1}\left(\mathbf{X}(\tau)'\boldsymbol{\varepsilon}/\sqrt{T}\right). \tag{21}$$

The procedure proposed by Hansen involves making random draws of $LM(\tau)$ by resampling the term in parentheses in (21) by means of $\mathbf{W}^j(\tau) = \mathbf{X}(\tau)'\mathbf{z}^j/\sqrt{T}$, where $\mathbf{z}^j$ is a $(T \times 1)$ vector of standard normally distributed random variables, yielding the following realisation j from the asymptotic null distribution of the LM test statistic:

$$LM_j(\tau) = W^j(\tau)'M(\tau)^{-1}R\,\widetilde{\Omega}^{-1}\,R'M(\tau)^{-1}W^j(\tau). \tag{22}$$

For a particular $\mathbf{z}^j$ these statistics can be calculated for all $\tau_i \in \Gamma$ and the maximum, exponential or average test statistics for given T can then be created according to (16) to (18). By repeating this procedure for J draws of $\mathbf{z}^j$, J draws from the asymptotic distribution are generated, with approximations to the p-values of the maximum test statistic obtained by means of

$$\widehat{p}^{\max} = J^{-1}\sum_{j=1}^{J} I\left[LM_j^{\max} > LM_T^{\max}\right].$$

Clearly the same principle applies to $p-$values for the $LM^{\exp}$ and LM^{ave} test statistics.

It is obvious that this procedure can also be applied as an alternative to the Luukkonen *et al.* (1988) test if $\boldsymbol{\delta}$ is restricted to the single lag form of (4) and the lag k is unknown. In this case, the $LM^{\max}, LM^{\exp}$ and/or LM^{ave} statistics are computed over the potential lags $1, 2, ..., q$ rather than over the set $\boldsymbol{\kappa} = (\kappa_1, \kappa_2)' \in \Delta'$.

3.3 Heteroscedasticity-consistent tests

In the presence of possible heteroscedasticity, the LM statistic (20) for testing $H_0 : \boldsymbol{\beta}^* = \mathbf{0}$ for a specific I_t in (19) becomes

$$\begin{aligned} LM_{T,hc}(\tau) &= T\widehat{\boldsymbol{\theta}}(\tau)'R\left[R'M(\tau)^{-1}\widetilde{V}(\tau)M(\tau)^{-1}R\right]^{-1}R'\widehat{\boldsymbol{\theta}}(\tau) \\ &= T\widehat{\boldsymbol{\theta}}(\tau)'R\,\widetilde{\Omega}_{hc}^{-1}\,R'\widehat{\boldsymbol{\theta}}(\tau), \end{aligned} \tag{23}$$

where $\widetilde{\Omega}_{hc} = R'M(\tau)^{-1}\widetilde{V}(\tau)M(\tau)^{-1}R$ is a consistent estimator of the covariance Ω_{hc} of $\sqrt{T}R'\widehat{\boldsymbol{\theta}}(\tau)$ in (21), with $\widetilde{V}(\tau) = \widetilde{\mathbf{W}}(\tau)'\widetilde{\mathbf{W}}(\tau)/T$ in which $\widetilde{\mathbf{W}}(\tau)$ is obtained by stacking $\widetilde{w}_t(\tau)$ and $\widetilde{w}_t(\tau) = \mathbf{x}_t(\tau)\widetilde{e}_t$ is the score vector evaluated under the null hypothesis. In our case, the maximum, exponential and average statistics from (23) are computed over the set of all $\boldsymbol{\kappa} = (\kappa_1, \kappa_2)' \in \Delta'$ to give $LM_{hc}^{\max}$, $LM_{hc}^{\exp}$ and LM_{hc}^{ave}.

In order to conduct asymptotically valid tests that replicate the heteroscedasticity of unknown form in the DGP, Hansen (1996) proposes resampling the term in brackets in (21) using $\mathbf{W}_{hc}^{j}(\tau) = \widetilde{\mathbf{W}}(\tau)'\mathbf{z}^{j}/\sqrt{T}$, where $\mathbf{z}^{j}$ is again a $(T \times 1)$ vector of standard normally distributed random variables. This leads to the realisation j from the asymptotic null distribution of (23) computed as

$$LM_{j,hc}(\tau) = W_{hc}^{j}(\tau)'M(\tau)^{-1}R\,\widetilde{\Omega}_{hc}^{-1}\,R'M(\tau)^{-1}W_{hc}^{j}(\tau).$$

Hence, using the resulting distribution for $LM_{T,hc}^{\max}, LM_{T,hc}^{\exp}$ and $LM_{T,hc}^{ave}$, approximate p-values for $LM_{hc}^{\max}$, $LM_{hc}^{\exp}$ and LM_{hc}^{ave} can be obtained.

In effect, the Hansen (1996) approach accounts for heteroscedasticity by considering the vector $\mathbf{x}_t(\tau)\widetilde{e}_t$, namely the interaction between the values of the regressors in (19) and the resulting residual, which is then randomized by multiplication by an *iid* standard normal variable. Asymptotically this randomization preserves the observed heteroscedasticity, but will not exactly reproduce the heteroscedastic pattern of the observed data in any given random draw j. Following the recent bootstrapping literature (e.g. Gonçalves and Kilian, 2004), an alternative procedure to generate random draws from the null distribution is to use a fixed design wild bootstrap procedure. Our proposed procedure replaces $\widetilde{w}_{t,hc}^{j}(\tau) = \mathbf{x}_t(\tau)\widetilde{e}_t z_t^{j}$ where z_t^{j} is an independent standard normal variate, with $\widetilde{w}_{t,wb}^{j}(\tau) = \mathbf{x}_t(\tau)\widetilde{e}_t\eta_t^{j}$ where η_t^{j} $(j = 1, ..., J)$ are generated as independent draws from the Rademacher distribution such that

$$\eta_t^{j} = \begin{cases} +1 \text{ with probability } 0.5 \\ -1 \text{ with probability } 0.5 \end{cases}. \tag{24}$$

Thus, by using the observed residual (computed under the null hypothesis) but randomizing its sign, the fixed design wild bootstrap exactly replicates the heteroscedasticity observed for each t in the finite sample under test. Since $\mathbf{x}_t(\tau)$ is held fixed over replications, any covariance between the regressors and the heteroscedasticity is maintained. The remainder of the procedure is as above and we refer to the test results obtained from this procedure, for a sample of size T, as $LM_{T,wb}^{\max}$, $LM_{T,wb}^{\exp}$ and $LM_{T,wb}^{ave}$.

Although originally developed in the context of a standard regression model with possibly heteroscedastic disturbances (see Liu, 1988, and Mammen, 1993), Gonçalves and Kilian (2004) establish (under certain assumptions) the validity of the wild bootstrap for testing for autocorrelation in a dynamic model allowing for heteroscedasticity of unknown form. Becker and Hurn (2007) demonstrate that a wild bootstrap procedure may be applied to the V23 test conditional on the heteroskedastic innovation process meeting strict moment existence conditions. However, the empirical properties appear to be satisfactory even when these formal conditions are not met. As the V23 test also includes higher order terms as regressors, it is anticipated that similar conditions on the transition variable $\mathbf{s}_t$ will be required here. In particular, if $\mathbf{s}_t$ is defined in terms of lagged dependent variables, moment conditions will have to be imposed on the error process itself. Becker and Hurn (2007) show how, in the case of heteroskedasticity of the GARCH type, the existence of moments depends on the GARCH parameters and the conditional innovation process.

Different choices for the distribution of η_t^j have been investigated in the literature. However, the Monte Carlo analysis of Godfrey and Orme (2002) finds the randomization scheme in (24) using the Rademacher distribution to perform well for a variety of regression misspecification tests, which leads us to use it in our nonlinearity testing context.

4 Empirical properties of the proposed test

This section examines the finite sample size and power properties of our proposed nonlinearity test procedure, also comparing the latter properties to other tests used to detect nonlinearity of the STR type.

4.1 Size

To investigate the empirical size with homoscedastic error terms, the univariate AR(1) process $y_t = 0.4\ y_{t-1} + \varepsilon_t$ with $\varepsilon_t \sim N(0,1)$, is simulated. The same AR(1) process is also used in a heteroskedastic setting, where $\varepsilon_t \sim N\left(0, \sigma_t^2\right)$ with $\sigma_t^2 = 1$ for $t = 1, ..., \frac{T}{2} - 1$ and $\sigma_t^2 = 2$ for $t = \frac{T}{2}, ..., T$. Although it is common to employ GARCH processes to capture conditional heteroscedasticity in financial variables, volatility changes for macroeconomic variables appear to be characterized by occasional abrupt shifts; see Sensier and van Dijk (2004). With this in mind, we adopt a simple break form of heteroscedasticity. The $WSTR$ nonlinear test is applied, utilizing the auxiliary regression (15). In all cases, the standard and heteroscedasticity-robust versions of the LM statistics, (20) and (23), are computed. Significance is evaluated as described above where we use hc and wb versions to evaluate the significance of the test statistic in (23) and J is set to 400. In all cases, $10,000$ replications are used to obtain the empirical characteristics of the test for sample sizes $T = 200,\ 500$ and $1,000$.

Three different versions of the test are implemented, with these differing in the dimensions allowed for the linear part $\mathbf{x}_t\boldsymbol{\phi}_1$ in (15) and the number of lags considered for the transition vector, $\mathbf{s}_t$. In this univariate context we define $\mathbf{x}_t = (y_{t-1}, ..., y_{t-p})$ and $\mathbf{s}_t = (y_{t-1}, ..., y_{t-q})$, so that the same maximum lag does not necessarily apply in the linear part and the transition vector. Two cases consider equal dimensions, with $p = q = 4$ and $p = q = 8$ used to replicate the (often arbitrary) values used in applications with quarterly data. In addition a third case utilises $p = 4$ lags in the linear part of the equation, but allows up to eight lags of y_t in the transition variable

($q = 8$). The potential weight distributions included in the set Δ' are those displayed in Figures 1 and 2 for 4 and 8 lags respectively in the transition function, and the corresponding values for κ_1 and κ_2 are shown in Table 1. In each case these consist of nine different weight distributions, some of which put almost all weight on one particular lag to allow the possibility that the nonlinearity follows the restricted DGP assumed by Teräsvirta (1994), while the remaining weight vectors put substantial weight on recent observations and declining weight on observations with longer lags. As already noted, a researcher can change these sets of weights and the number of different weight structures utilized without any affecting the general procedure.

Size results are displayed in Table 2 for nominal significance levels of 1, 5 and 10 percent. To avoid unnecessary repetition, the results for tests based on the exponential version of the LM test (17) are not shown, since these are always very similar to those obtained using the average and, more particularly, the max versions of the statistics. It is easily seen from Panel (a) that the empirical size for the tests assuming homoscedasticity are fairly accurate when applied to the homoscedastic AR(1) DGP, being only slightly conservative. On the other hand, the heteroscedasticity robust versions of the Hansen (1996) procedure are very conservative, and this characteristic is also observed by Hansen (1996). The sizes of both versions of this test generally improve with the sample size and also generally improve when fewer parameters are estimated. Thus, although the same number of parameters are subject to testing irrespective of the assumed linear AR order (p) or the maximum order of transition variable (q), the empirical size results are very similar for the cases of $q = 4$ and 8 when a common assumed AR order $p = 4$ is employed. Nevertheless, for a realistic sample size with monthly data of $T = 500$, the empirical size for the Hansen (1996) heteroscedastic-consistent procedure is around 0.02 in all cases when a nominal 0.05 level is used, with this deteriorating to around 0.015 with $p = 8$.

Although it also allows for the possibility of heteroscedasticity that is not present in this DGP, it is striking that the use of the wild bootstrap procedure delivers very good size for all cases in

Table 2, irrespective of the sample size. Indeed, the finite sample size using this procedure is more reliable than that delivered by the use of standard Hansen (1996) homoscedastic draws.

When the true DGP is heteroscedastic, panel (b) of Table 2 shows that tests assuming homoscedasticity are badly oversized. For example, in all cases the use of a nominal significance level of 0.05 results in an empirical size of around 0.40 or more, and exceeding 0.60 when $p = 8$ is employed. Of course, this result does not inherently depend on the use of the maximal, average or exponential statistics (16), (18) or (17), but is rather a function of the failure to account for heteroscedasticity in the computation of the underlying LM statistic (20). Therefore, rejections of linearity using tests based on (20) should be treated with extreme caution when heteroscedasticity may be present. Due to the invalid asymptotic inference being employed, there is no evidence that larger sample sizes yields better empirical size when such statistics are used with a heteroscedastic DGP in panel (b) of Table 2.

In common with the homoscedastic DGP, the use of the heteroscedasticity consistent Hansen (1996) approach results in conservative empirical size, although this is even more marked here than in the homoscedastic DGP case. There is, however, a substantial improvement in size when moving from $T = 200$ to $T = 500$ sample observations. In the latter case, and as a rule of thumb, the use of a nominal significance level of 0.10 yields an empirical size around 0.05. Once again, however, much more reliable size results are obtained when wild bootstrap inference is employed. Indeed, although some oversizing is observed in these tests for the heteroscedastic DGP in panel (b), this is very modest compared with the size distortions observed in the other two cases.

4.2 Comparison with related tests

We next compare our $WSTR$ nonlinearity test to two related procedures, namely the minimum p-value of the sequence of tests proposed in Teräsvirta (1994) for model specification and the overall test developed by Luukkonen *et al.* (1988). Although all three procedures deal with the parameters

γ and c through a third order Taylor series expansion of $f(\mathbf{s}_t)$ in (3), they differ in the handling of $\boldsymbol{\delta}$. More specifically, the Teräsvirta and Luukkonen *et al.* tests restrict $\boldsymbol{\delta}$ to the single lag form in (4), whereas the $WSTR$ procedure allows $\boldsymbol{\delta}$ to follow the more flexible form in (5). As in the previous subsection, the DGP is univariate.

As indicated in the discussion of Section 3, if the transition variable $s_t = y_{t-k}$ applies with k known, the auxiliary regression (10) is relevant and the corresponding LM test statistic is asymptotically χ^2 distributed with $3p$ degrees of freedom. Although proposed by Teräsvirta (1994) only for model specification purposes, the literature abounds with examples of applying a strategy of allowing k to vary from 1 to q and selecting the value of k that leads to the strongest rejection of the null hypothesis, with the corresponding minimum p-value treated as applying in the context of a test of nonlinearity. When used as a nonlinearity test, we refer to this as $p-\min$. It is, however, clear that the distribution of this test statistic will be nonstandard.

The Luukkonen *et al.* (1988) auxiliary regression of (11) reflects the treatment of the delay parameter as unknown while maintaining the restriction on $\boldsymbol{\delta}$ as in (4). In the univariate case examined here, the test statistic is asymptotically χ^2 distributed with $3qp - [p(p-1)/2]$ degrees of freedom, which can imply a large increase in restrictions tested compared to the Teräsvirta test and it is consequently often conjectured that the Luukkonen *et al.* test may suffer from low power. The Luukkonen *et al.* (1988) test is denoted LST in the table

Our comparison assumes $p = q = 4$ in all cases. Hence each of the individual $LSTAR$ and $WSTAR$ auxiliary regressions of (10) and (15) respectively has 12 restrictions to be tested while the Luukkonen *et al.* (1988) test in (15) has 42 such restrictions, but requires only a single regression. The $WSTAR$ test is implemented with two different weight sets. The first of these is the set shown in Table 1 and displayed in Figure 1. The second set of weights is a restricted version of this set, with only those weight distributions which effectively give unit weight to one particular lag (using the weights in the bottom panel of Figure 1). This latter case, in combination with the maximal mapping in equation (16), emulates the $p - \min$ test for $LSTAR$ nonlinearity, but should ensure that this procedure is correctly sized. The procedure is labelled LM^{STAR} in the table.

The first DGP used is the linear homoscedastic AR(1) process used in Section 4.1, which facilitates the evaluation of the severity of size distortion when a multiple testing strategy is employed. The second DGP is a nonlinear $LSTAR$ model as in (1) and (3) with parameters $\alpha_0 = \beta_0 = 0$, $\boldsymbol{\alpha}_1 = (0.6\ 0\ 0\ 0)'$, $\boldsymbol{\beta}_1 = (-0.4\ 0\ 0\ 0)'$, $\gamma = 20$ and $c = 0$, which provides information on power. Two different sets of parameters are used for $\boldsymbol{\delta}$, namely the conventional $LSTAR$ specification: $\boldsymbol{\delta}_1 = (1\ 0\ 0\ 0)'$ and the $WSTAR$ weights $\boldsymbol{\delta}_2 = (1/3\ 1/3\ 1/3\ 0)'$. The former corresponds to the single-lag alternative on which the Teräsvirta (1994) and Luukkonen *et al.* (1988) tests are based, whereas the latter uses an average of lagged observations to form the transition variable. It should, however, be noted that although the latter weight distribution is an average, this average is over three not four lags. Consequently none of the weight distributions included in Δ' exactly corresponds to $\boldsymbol{\delta}_2$ and hence we are investigating the power of our test procedure when the weight distribution in the DGP is not precisely replicated in the weight functions considered[14]. All DGPs used in these comparisons are homoscedastic, with decisions for all procedures except $WSTAR$ being based on the conventional χ^2 distribution; $10,000$ replications are employed in each case.

The results for the AR(1) in Table 3 provide a size comparison for the Luukkonen *et al.* test (denoted LST) and the $p - \min$ test in comparison to the procedures whose size has already been discussed in relation to Table 2. (Note that the LM^{STAR} test is a special case of $LM^{\max}$.) The LST test is undersized, particularly at a sample size of $T = 200$, with this undersizing (not surprisingly) becoming less severe as the sample size increases. On the other hand, the $p - \min$ procedure is always badly oversized, for example having an empirical size of around 15% at a nominal 5% significance level. This overrejection under the null is, of course, a consequence of using a multiple testing procedure without taking this into account when computing the test statistic. However, this

[14] Clearly, the power obtained in the simulations could be artifically increased by choosing a weight vector for the DGP that exactly matches one of the elements in Δ'.

strategy is often employed as a nonlinearity test prior to building STR models, and the size results for the linear AR(1) in Table 3 emphasize that apparent significance according to this statistic is not a reliable indicator of the presence of such nonlinearity in the DGP.

Turning to the power results for the $LSTAR$ and $WSTAR$ DGPs, the LST test always has lower power than the homoscedastic and wild bootstrap versions of the LM procedure[15], due primarily to the large number of parameter restrictions which need to be tested in LST. For example, in the $LSTAR$ specification with all weight at lag 1, the LST test has power of 0.18 with $T = 500$ and a 5 percent significance level, whereas the LM^{STAR} and $WSTAR$ procedures in the corresponding case each have power around 0.25, except when Hanson (1996) heteroscedastic-consistent draws are used for the latter. Hence the higher power shown by these other tests can be explained by their more economical use of degrees of freedom in the individual regressions. While the $p-\min$ test has the largest nominal power it should be noted that this is merely the flip side of the test's size distortion.

When the weight distribution is of the $WSTAR$ form, it is interesting that the $LSTAR$ test LM^{STAR} has very similar power to $LM^{\max}$ and $LM^{\max}_{wb}$ when $T = 200$, although the power for the latter is increased when the average form of the statistic is used. Despite the power being relatively modest for all these tests with this sample size, this finding indicates that the single-lag $LSTAR$ specification retains power even against this more general model. Indeed, this remains true when the sample size increases to $T = 500$ or 1000. Although not included in our comparison, this implies than an average form of the LM test based on single-lag weights may perform similarly to LM^{ave} and LM^{ave}_{wb} in this context. It may also be noted that all tests gain power when applied to the $WSTAR$ DGP rather than the $LSTAR$ one, nevertheless the advantages of the $WSTAR$ approach over the Luukkonen *et al.* test is even more marked when the DGP has this more general

[15] Results were also obtained for the *exp* form of all LM statistics. However, as for size (discussed in the preceding subsection), these results were qualitatively very similar to those obtained using the max and *ave* versions. Indeed, the empirical size and power for the exp form was typically intermediate between the max and *ave* results shown.

form.

Given the undersizing exhibited by the Hansen (1996) heteroscedasticity-consistent procedure, it is to be anticipated that this approach will have lower nominal power than the $LM^{\max}$ and LM^{ave} tests that have more reliable size, and this is indeed borne out. However, it appears there is no loss of power in using the wild bootstrap form allowing for possible heteroscedasticity, even when no such heteroscedasticity is present in the DGP.

Although considered here only for $p = 4$, it is anticipated that the performance of the $WSTAR$ nonlinearity test will dominate the Luukkonen *et al.* (1988) test even more when higher potential lag orders are considered, especially for realistic sample sizes in macroeconomics, such as $T = 200$ or $T = 500$. It is also very reassuring that robustification against heteroscedasticity using the wild bootstrap form of the LM statistics does not lead to a deterioration of power. Indeed, this enables us to overturn the recommendation of van Dijk *et al.* (2002, p.160) that robust procedures not be used for nonlinearity testing. Indeed, our recommendation is to always apply the wild bootstrap form of the test.

5 Empirical Application

This section applies the $WSTR$ testing and modelling approach above to examine the leading indicator properties of the yield curve for quarterly output (GDP) growth in the UK. There is an extensive literature on this issue, particularly in relation to the US economy; see, among many examples, Estrella and Hardouvelia (1991), Hamilton and Kim (2002), Stock and Watson (2003), while international evidence is examined in Davis and Fagan (1997) and Bernard and Gerlach (1998). Although much of the literature on this topic employs linear models, Galvão (2006) uses higher frequency data on financial variables in a smooth transition $MIDAS$ specification for US GDP growth , while Anderson, Athanasopoulos and Vahid (2007) examine nonlinear interactions between quarterly GDP growth and the yield curve for the G7 countries.

In a similar way to Galvão (2006), our $WSTR$ specification enables us to examine the leading indicator properties of the yield curve for GDP growth, without restricting yield curve data to match the quarterly frequency for which GDP data are available. This represents a generalisation of the $LSTR$ models examined in Anderson *et al.* (2007). We study the leading indicator properties of the yield curve for the UK and, in order to be able to compare our results to those in Anderson *et al.* (2007), we employ the same data period as in their study. To be specific, we employ monthly data on the slope of the slope of the yield curve, constructed as end of month values of ten year bond returns less the three month Treasury Bill yield, from January 1960 to December 1999, alongside seasonally adjusted quarterly GDP growth (computed as the first difference of the logarithm) for the same period.

If a researcher simply uses four quarterly lags on each of GDP growth and the yield curve (with quarterly data for the latter constructed by taking the third month of the quarter), the application of the usual nonlinearity test to (8) for known transition lag k produces strong evidence of nonlinearity by the usual χ^2 p-values, with those for the second to fourth lags of GDP being 0.0072, 0.0013 and 0.031 respectively and 0.024 and 0.00094 for the first and second lags, respectively, of the yield curve. Anderson *et al.* (2007) similarly find evidence of STR nonlinearity for this relationship with lags of either variable being the possible transition. However, as discussed in Sections 3 and 4 above, these results are unreliable as a test for nonlinearity since the lag is unknown and heteroscedasticity may be present. When our $WSTR$ testing procedure is applied, using the quarterly weights shown in Table 1 and 1,000 bootstrap replications, significant results are still obtained for a GDP transition assuming homoscedasticity, but this result appears to be spurious since the wild bootstrap tests yield $LM_{wb}^{\max} = 0.422$, $LM_{wb}^{ave} = 0.243$ and $LM_{wb}^{\exp} = 0.396$ as the p-values. For a yield curve transition, on the other hand, the corresponding results are $LM_{wb}^{\max} = 0.064$, $LM_{wb}^{ave} = 0.017$ and $LM_{wb}^{\exp} = 0.062$, providing compelling evidence that the yield curve is the more appropriate choice as transition variable[16].

The selection of the yield curve as the transition variable implies that the use of higher frequency data may provide more information on the appropriate transition. However, the inclusion of unrestricted monthly lags in the linear part of the equation implies a highly parameterized model. To obtain a more parsiminous specification, a preliminary linear analysis was undertaken using quarterly GDP and monthly yield curve lags to a maximum of one year (that is, 4 and 12 lags respectively). A dummy variable was also included for 1973Q1, which experienced abnormal quarter to quarter growth of 5 percent and led to an outlier residual (greater than 3.5 standard deviations) in all preliminary linear and nonlinear models. Selection of the maximum lag order by AIC leads to the inclusion of three GDP lags and two yield curve lags[17]. Using an obvious notation for variables, and with robust t-statistics in parentheses, the resulting estimated linear model is

$$\widehat{GDP}_t = \underset{(2.87)}{0.45} + \underset{(18.2)}{4.42} DUM73_t - \underset{(-1.02)}{0.099} GDP_{t-1} + \underset{(0.85)}{0.076} GDP_{t-2} + \underset{(1.09)}{0.085} GDP_{t-3}$$

$$- \underset{(-0.47)}{0.080} YC^m_{t-1/3} + \underset{(0.92)}{0.164} YC^m_{t-2/3}, \quad \widehat{\sigma} = 0.975 \tag{25}$$

where the superscript m indicates monthly data while the subscript $t - 1/3$ indicates a lag of one third of a quarter, that is one month. While the GDP and yield curve slope lags prove insignificant at this stage it is important to not prune the model further until after allowing for the inclusion of nonlinear terms. Application of the $WSTR$ nonlinearity test with the variables in this model and maximum lag of the yield curve transition variable of 12 months yields strong evidence of nonlinearity, with p-values of $LM^{\max}_{wb} = 0.049$, $LM^{ave}_{wb} = 0.006$, and $LM^{\exp}_{wb} = 0.032$. The gamma function parameters used in computing these tests, shown in Table 1, capture a variety of plausible shapes for this monthly case.

[16] In fact, Anderson *et al.* (2007) select GDP as the transition variable, although this is based on a number of nonlinearity tests rather than the conventional STR specification test alone.

[17] Lag selection for a linear model as the basis of a potentially nonlinear model is not straightforward, as nonlinear dependence may not show in linear dependency measures. The following ad-hoc procedure was applied. With a maximum lag of 4 quarterly GDP lags and 12 monthly lags for the yield curve slope the AIC criterion was used to find the 10 best linear models. The maximum GDP and yield curve slope lag amongst these models is 3 and 2 respectively.

To move to a parsiminous model, the $WSTR$ transition function is estimated and, conditional on this $\widehat{f}(\widehat{s}_t)$, OLS is applied to the nonlinear model of (1) and AIC is used to select individual lagged variables from the set (1, $DUM73_t$ $\mathbf{x}_t$, $\widehat{f}(\widehat{s}_t)$, $\widehat{f}(\widehat{s}_t)\mathbf{x}_t$)[18]. Nonlinear least squares estimation of the resulting model yields (robust t-statistics[19])

$$\widehat{GDP}_t = \underset{(2.04)}{0.25} + \underset{(9.94)}{3.97} DUM73_t - \underset{(-1.99)}{0.13}\, GDP_{t-2} + \underset{(6.73)}{0.39} GDP_{t-2} + \underset{(1.67)}{0.12} GDP_{t-3} + \underset{(2.82)}{0.15} YC^m_{t-1/3}$$

$$+\widehat{f}(\widehat{s}_t)\left\{\underset{(2.37)}{0.77} - \underset{(-6.56)}{0.72}\, GDP_{t-2} - \underset{(-1.33)}{0.45}\, YC^m_{t-1/3} + \underset{(0.63)}{0.23} YC^m_{t-2/3}\right\} \qquad \widehat{\sigma} = 0.885 \quad (26)$$

where

$$\begin{aligned} \widehat{f}(\widehat{s}_t) &= \{1 + \exp -2379\,[\widehat{s}_t - 1.50]\}^{-1} \\ \widehat{s}_t &= \sum_{i=1}^{12} \delta_i\,[18.51, \quad 53.05]\; YC^m_{t-i/3} \end{aligned} \qquad (27)$$

As shown in Figure 3 (middle panel), the estimated weight function $\widehat{\boldsymbol{\delta}}(\widehat{\kappa}_1, \widehat{\kappa}_2)$ implies that nonzero weights apply to the yield curve slope at lags of two to five months, with around 90% of the total weight applying at lags three and four. Such weights will not be well approximated by the use of either quarterly yield curve data or by a conventional single lag $LSTR$ specification, emphasizing the usefulness of the flexibility provided by our $WSTR$ approach.

As indicated by $\widehat{\gamma} = 2379$ in (27), the estimated model implies an abrupt transition between regimes, with these regimes effectively defined in terms of a threshold of 1.5 percentage points for the yield curve slope. The time series properties of the transition function shown in Figure 3 (bottom panel) indicate that the upper yield curve regime is predominant during much of the 1970s, and also in the periods 1982-1985 and 1993-1997. Whereas the yield curve slope has a positive impact on GDP growth in the lower regime (that is, when longterm interest rates are less than 1.5 percentage points above short-term ones), the estimates of (26) imply that this effect disappears

[18] A similar general to specific procedure is used by Sensier *et al.* (2002) for the specification of $LSTR$. models.

[19] The parameter estimates shown are the results of full nonlinear least squares, the t-statistics, however, are conditional on $\widehat{\boldsymbol{\theta}}_2$. Obtaining a full variance-covariance matrix for all parameter estimates is notoriously difficult in smooth transition models. In the present case the estimated full Hessian matrix proved to be singular.

in the upper regime. Nevertheless, the upper yield curve regime is associated with higher GDP growth. For example, in a lower regime steady state and assuming an interest rate differential of 0.17 percentage points (the average yield curve slope conditional on being in the lower regime), the implied mean growth is 0.44% per quarter, whereas an upper regime steady state with a differential of 2.45 percentage points implies a mean growth of 0.63% per quarter[20]. Nevertheless, it should be noted that these mean calculations depend not only on the assumed yield curve slope, but also on the implausible assumption of a steady-state within a given regime. Perhaps a more important difference between regimes occurs in the dynamics of growth, with persistence (as measured by the sum of the autoregressive coefficients) being 0.38 in the lower regime and -0.34 in the higher regime. Thus, shocks to growth will have persistent effects over time when the interest rate differential is below 1.5, whereas GDP growth exhibits no persistence when the differential exceeds this value.

6 Conclusions

Establishing the nature any nonlinearity in the relationships between economic or financial variables is an important issue for empirical modelling, especially in the presence of possible heteroscedasticity. The smooth transition class of models promoted by Teräsvirta (1994, 1998) and his co-authors (for example, van Dijk *et al.* 2002) provide an intuitively attractive and plausible form for such nonlinearity, but the statistical basis for employing these models in practice is often weak due to the failure to take account of the application of multiple tests when the appropriate lag of the transition variable generating the nonlinearity is unknown and/or when heteroskedasticity may be present. Further, the nonlinearity permitted in practice is restricted in that the regimes are (almost always) assumed to be triggered by the value of a single lagged observation.

Inspired by the $MIDAS$ methodology of Ghysels *et al.* (2005, 2006), the present paper gen-

[20] These mean values are computed using the estimated coefficients of (26), without any possible further restrictions imposed.

eralizes the smooth transition class of models by defining the transition function in terms of a weighted average of lags of the transition variable. Our approach removes the need to specify the appropriate individual lag of the transition variable in the nonlinear modelling process, while also enabling a wide variety of weight functions to be considered. Further, using the approach of Hansen (1996), we develop a testing procedure that delivers a correctly sized hypothesis test for the presence of nonlinearity in the realistic situation that the appropriate weight function (or individual lag) is unknown. Although Hansen also shows how an asymptotically valid test can be undertaken when heteroscedasticity may be present, we propose the use of the fixed design wild bootstrap (Gonçalves and Kilian, 2004) in this context, showing that it delivers well-sized inference in finite samples. Indeed, our wild bootstrap test not only has has good size, but also has power comparable to that shown by the homoscedastic-robust Hansen (1996) test when heteroscedasticity is absent. Therefore, we recommend that this version of the test should always be used in practice, whether heteroscedasticity is anticipated to be an issue or not.

Because both testing and modelling are based on flexible but parsiminous weighted functions of the lagged values of the transition variable, our approach can be applied in the mixed frequency context considered by Ghysels *et al.* (2005, 2006). Our application examines monthly values of the yield curve slope as a leading indicator of quarterly GDP growth in the UK. In contrast to the ambiguous results of Anderson *et al.* (2007) in terms of which variable is the appropriate transition variable for this relationship, when heteroscedastity is permitted through the wild bootstrap version of our test, the results clearly point to this being the yield curve. Further, our estimated $WSTR$ model places substantial weight on two (monthly) lags of this variable, so that values three and four months prior to the current quarter are important in generating the transition between regimes.

Galvão (2006) has independently proposed a similar approach to ours in the context of prediction using the smooth transition regression model with mixed frequency data. Although the focus of the present paper is different, as we deal with modelling and (more particularly) nonlinearity testing, our companion paper Becker and Osborn (2007) examines $WSTR$ models in a forecasting context. Indeed, our results there indicate not only that the $WSTR$ model is to be preferred over conventional STR models for forecasting purposes even when all data are of the same frequency, but also that no significant loss of accuracy occurs when a $WTSR$ specification is used to forecast a linear data generating process. Based on the results of the present paper and also Becker and Osborn (2007), we believe that further development and application of the $WSTR$ approach is warrented..

References

Anderson, H.M., G. Athanasopoulos and F. Vahid (2007) "Nonlinear autoregressive leading indicator models of output in G-7 countries", *Journal of Applied Econometrics*, **22**, 63-87.

Anderson, H.M. and F. Vahid (2001) "Predicting the probability of a recession with nonlinear autoregressive leading-indicator models", *Macroeconomic Dynamics*, **5**, 482-505.

Andrews D.W.K. (1993) "Tests for parameter instability and structural change with unknown change point. Optimal tests when a nuisance parameter is present only under the alternative" *Econometrica*, **61**, 821-856.

Andrews D.W.K. and Ploberger W. (1994) "Optimal tests when a nuisance parameter is present only under the alternative" *Econometrica*, **62**, 1383-1414.

Becker R. and A.S. Hurn (2006) "Testing for nonlinearity in mean in the presence of heteroskedasticity", Working Paper No. 8, National Centre for Econometric Research, Australia.

Becker R and D.R. Osborn (2007) "Forecasting with weighted smooth transition regressions", mimeo, Economics, University of Manchester.

Bernard H. and S. Gerlach (1998) “Does the term structure predict recessions: The international evidence” *Journal of Finance and Economics*, **3**, 195-215.

Davies R.B. (1987) “Hypothesis testing when a nuisance parameter is only identified under the alternative” *Biometrika*, **47**, 33-43.

Davis E.P. and Fagan G (1997) “Are financial spreads useful indicators of future inrlation and output growth in EU countries?” *Journal of Applied Econometrics*, **12**, 701-714.

Estrella A and Hardouvelis G.A. (1991) “The term structure as a predictor of real economic activity” *Journal of Finance*, **46**, 555-576.

Fok D., van Dijk D. and Franses P.H. (2005) “A multi-level STAR model for US manufacturing sectors”, *Journal of Applied Econometrics*, **20**, 811-827.

Galvão A.B. (2006) “Changes in predictive ability with mixed frequency data”, mimeo, Department of Economics, Queen Mary, University of London, December 2006.

Ghysels E., Santa-Clara P. and Valkanov R. (2005) “There is a risk-return tradeoff after all”, *Journal of Financial Economics*, **76**, 509-548.

Ghysels E., Santa-Clara P. and Valkanov R. (2006) “Predicting volatility: Getting the most out of return data sampled at different frequencies”, *Journal of Econometrics*, **131**, 59-95.

Godfrey, L.G and Orme C.D. (2004) “Controlling the finite sample significance levels of heteroskedasticity-robust tests of several linear restrictions on regression coefficients”, *Economics Letters*, **82**, 281-287.

Gonçalves S. and Kilian L. (2004) “Bootstrapping autoregressions with conditional heteroskedasticity of unknown form”, *Journal of Econometrics*, **123**, 89-120.

Hamilton J.D. (1994) *Time Series Analysis*, Princeton University Press: Princeton.

Hamilton J.D. and Kim D.H. (2002) "A re-examination of the predictability of economic activity using the yield spread", *Journal of Money, Credit and Banking*, **34**, 340-360.

Hansen B. (1996) "Inference when a nuisance parameter is not identified under the null hypothesis" *Econometrica*, **64**, 413-430.

Lee T., White H. and Granger C.W.J., (1993) "Testing for neglected nonlinearity in time series models", *Journal of Econometrics*, **56**, 269-290.

Lundbergh S. and Teräsvirta T. (1998) "Modelling high-frequency time series with STAR-GARCH models", SSE/EFI working paper number 291, Stockholm School of Economics.

Luukkonen R. Saikkonen P. and Teräsvirta T. (1988) "Testing linearity against smooth transition autoregressive models" *Biometrika*, **75**, 491-499.

Medeiros M.C. and Veiga A. (2003). "Diagnostic checking in a flexible nonlinear time series model", *Journal of Time Series Aanalysis*, **24**, 461-482.

Medeiros M.C. and Veiga A. (2005). "A flexible coefficient smooth transition time series model", *IEEE Transactions on Neural Networks*, 16, 97-113.

Sensier M., Osborn D.R. and Öcal N. (2002) "Asymmetric interest rate effects for the UK real economy" *Oxford Bulletin of Economics and Statistics*, **64**, 315-339.

Sensier M. and van Dijk D. (2004) "Testing for volatility changes in US macroeconomic time series" *Review of Economics and Statistics*, **86**, 833-839.

Skalin J. and Teräsvirta T. (2002) "Modeling asymmetries and moving equilibria in unemployment rates" *Macroeconomic Dynamics*, **6**, 202-241.

Stock J and Watson M.W. (2003) "Forecasting output and inflation: The role of asset prices" *Journal of Economic Literature*, **91**, 788-829.

Taylor, M.P., Peel D.A. and Sarno L. (2001) "Nonlinear mean-reversion in real exchange rates: Towards a solution to the purchasing power parity puzzles", *International Economic Review*, **42**, 1015-1042.

Teräsvirta T. (1994) "Specification, estimation and evaluation of smooth transition autoregressive models" *Journal of the American Statistical Association*, **78**, 81-89.

Teräsvirta T. (1998) "Modelling economic relationshipe with smooth transition regressions", in Ullah A and Giles D.E.A. (eds.) *Handbook of Applied Economic Statistics*, pp.507-552. Marcel Dekker, New York.

Teräsvirta T., van Dijk D. and Medeiros M.C. (2004) "Linear models, smooth transition autoregressions, and neural networks for forecasting macroeconomic time series: A re-examination" *International Journal of Forecasting*, forthcoming.

Teräsvirta T., Lin C.F. and Granger C.W.J. (1993) "Power of the neural network linearity test" *Journal of Time Series Analysis*, **14**, 209-220.

van Dijk D., Franses P.H. and Paap R. (2002) "A nonlinear long memory model, with qn application to US unemployment", *Journal of Econometrics*, **110**, 135-165.

van Dijk, D., Teräsvirta T. and Franses P.H. (2002) "Smooth transition autoregressive models - A survey of recent developments", *Econometric Reviews*, **14**, 1-47.

A Estimation

In this appendix a number of issues arising from estimating $LSTR$ and $WSTR$ models are discussed. The reported results use starting values for nonlinear least squares estimation of $\theta_{wSTAR} = \left(\alpha_0 \; \boldsymbol{\alpha}_1' \; \beta_0 \; \boldsymbol{\beta}_1' \; \gamma \; c \; \kappa_1 \kappa_2\right)'$ for the $WSTR$ model in equations (**??**) and (5) assuming $\boldsymbol{\delta} = \mathbf{e}_k$,that is, starting from the more restricted STR form. The STR starting values, denoted $\theta_{STAR,0}$, for $\theta_{STAR} = \left(\alpha_0\ \boldsymbol{\alpha}_1'\ \beta_0\ \boldsymbol{\beta}_1'\ \gamma\ c\ k\right)'$ are obtained from a grid search over $\gamma \in [\underline{\gamma}, \overline{\gamma}]$, $c \in [\underline{c}, \overline{c}]$ and $k \in [1, p]$. This vector is then translated into a vector $\theta_{WSTR,0}$ corresponding to the more general $WSTR$ model by mapping the $LSTR$ delay parameter k_0 into κ_{10} and κ_{20}. This requires a parameter combination for $(\kappa_1\ \kappa_2)'$ that gives a weight of 1 for k and 0 for all $i \neq k \in [1, p]$. However, as the resulting distribution has to fit only a finite number of points, it is possible that a range of parameter combinations can fit this simple distribution.

In particular, a number of parameter combinations for $(\kappa_1\ \kappa_2)'$ will give almost unit weight on lag k and little weight to other lags. Therefore, if the nonlinear optimisation is commenced from such values, problems may be encountered because the objective function is extremely flat in the neighbourhood of the starting values. For this reason it is convenient to alter the starting values for κ_{10} and κ_{20} such that the resulting weight distribution gives nonzero weight to more than one lag. In practice, we achieve this by premultiplying the starting values for κ_{10} and κ_{20} by a factor smaller than one, which increases the variance of the weight distribution.

Thie above outlines an initialization of the optimisation procedure beginning from an $LSTR$ grid search, and this procedure is adopted in all results presented in the paper. An alternative procedure might be to commence the optimisation from the weight combination $(\kappa_1\ \kappa_2)'$ that give rises to the $LM^{\max}$ (or $LM_{wb}^{\max}$ in the case of possible heteroscedasticity) in the nonlinearity test.

Assuming that $\varepsilon_t \sim N\left(0, \sigma^2\right)$, the likelihood function can be maximised by minimising the sum of squared residuals. The constrained maximum likelihood procedure of GAUSS 6.0 is then used to minimise this function over the parameter vector $\boldsymbol{\theta} = \left(\alpha_0\ \boldsymbol{\alpha}_1'\ \beta_0\ \boldsymbol{\beta}_1'\ \gamma\ c\ \kappa_1 \kappa_2\right)'$ for the $WSTR$ or $\boldsymbol{\theta} = \left(\alpha_0\ \boldsymbol{\alpha}_1'\ \beta_0\ \boldsymbol{\beta}_1'\ \gamma\ c\ k\right)'$ for the standard STR. The parameter vector can be decomposed into $\theta = \left(\theta_1'\ \theta_2'\right)'$, where $\boldsymbol{\theta}_1 = \left(\alpha_0\ \boldsymbol{\alpha}_1'\ \beta_0\ \boldsymbol{\beta}_1'\right)'$ and $\boldsymbol{\theta}_2 = (\ \gamma\ c\ \kappa_1 \kappa_2)'$ or $\theta_2 = (\ \gamma\ c\ k)'$. This is convenient as, given any estimate for parameter vector θ_2, namely $\widehat{\theta}_2$, θ_1 has an analytical representation and therefore can be concentrated out of the criterion function. Consequently the nonlinear optimisation algorithm merely needs to search over the relevant parameter space for θ_2. Since $\widehat{\theta}_2$ yields $\widehat{f}(\widehat{\mathbf{s}})$, $\widehat{\theta}_1(\widehat{\theta}_2)$ is merely the standard OLS estimate obtained by regressing y_t on $[1\ \mathbf{x}_t\ \widehat{f}(\widehat{\mathbf{s}})\ \widehat{f}(\widehat{\mathbf{s}})\mathbf{x}_t]'$. As the information matrix is not block-diagonal in θ_1 and θ_2, however, the standard OLS $Var\left(\widehat{\theta}_1\right)$ obtained from the latter linear regression is not correct.

The variance-covariance matrix for the estimated parameters can be estimated by means of gradient and Hessian estimates obtained through the unconcentrated likelihood functions (see for example Hamilton, 1994).

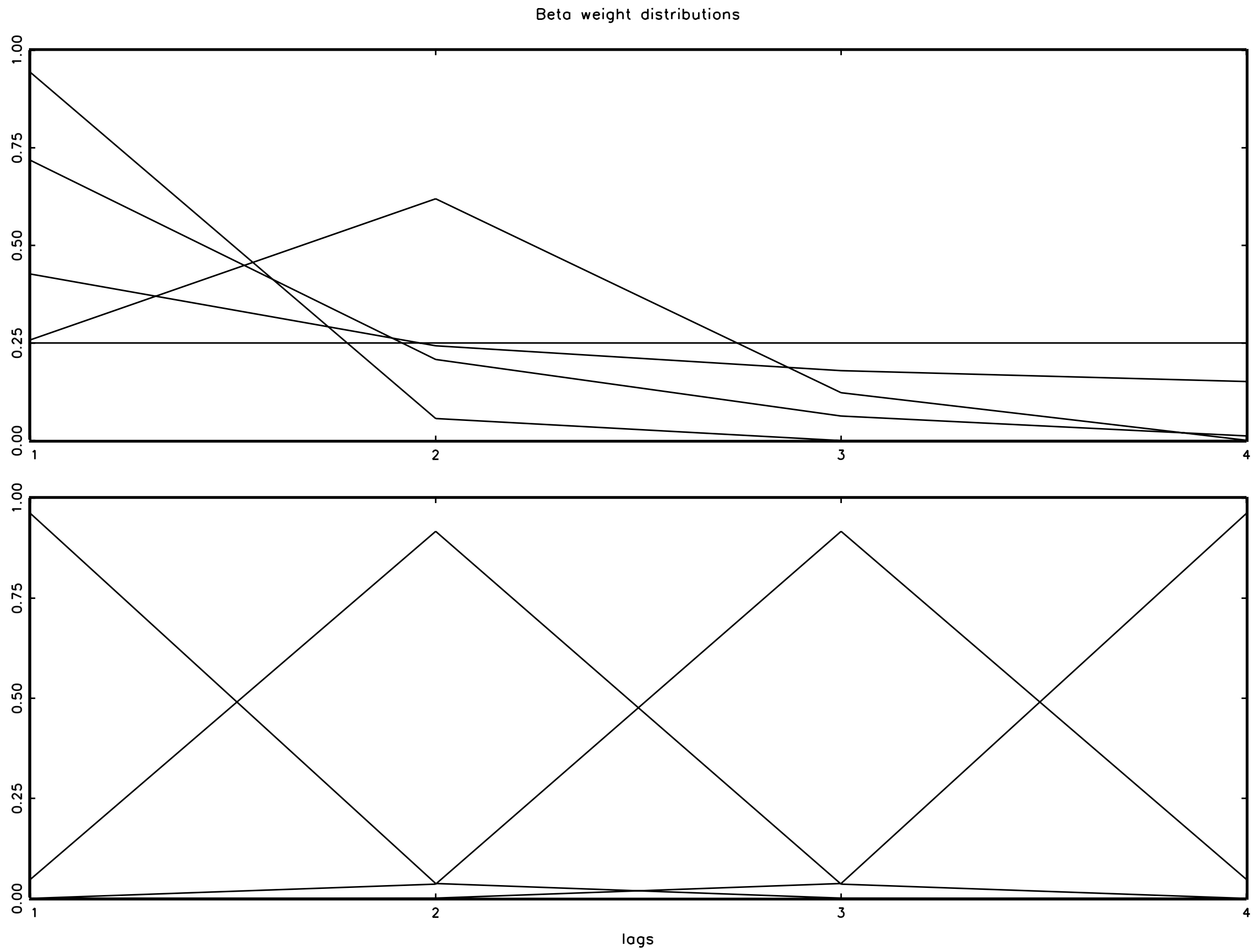

Figure 2: Nine weight distributions based on beta weights for a maximum lag of 4. The top panel displays weight distributions that give substantial weight to more than one lag. The bottom panel shows weight distributions that put almost all weight on one lag.

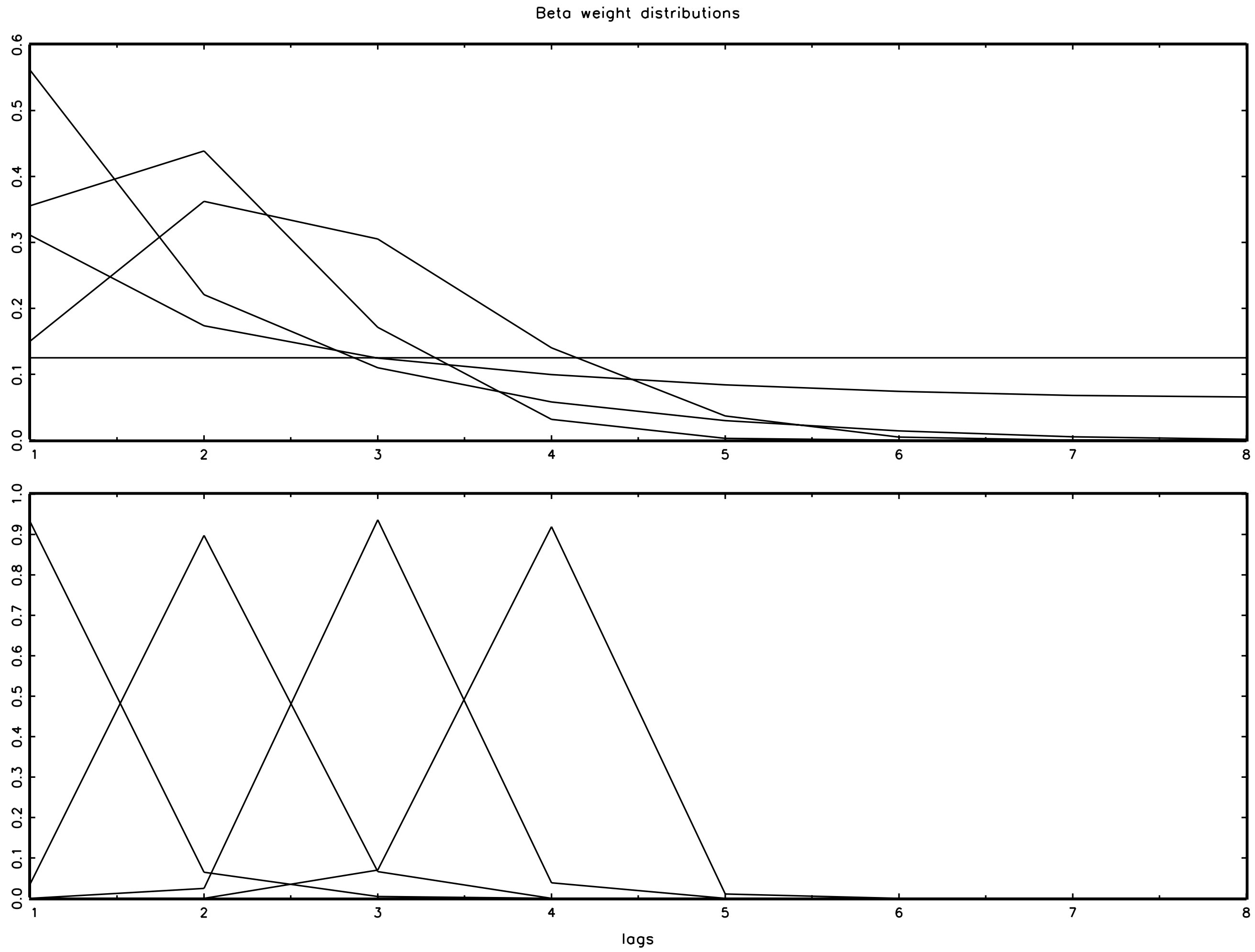

Figure 2: Nine weight distributions based on beta weights for a maximum lag of 8. The top panel displays weight distributions that give substantial weight to more than one lag. The bottom panel shows weight distributions that put almost all weight to one lag.

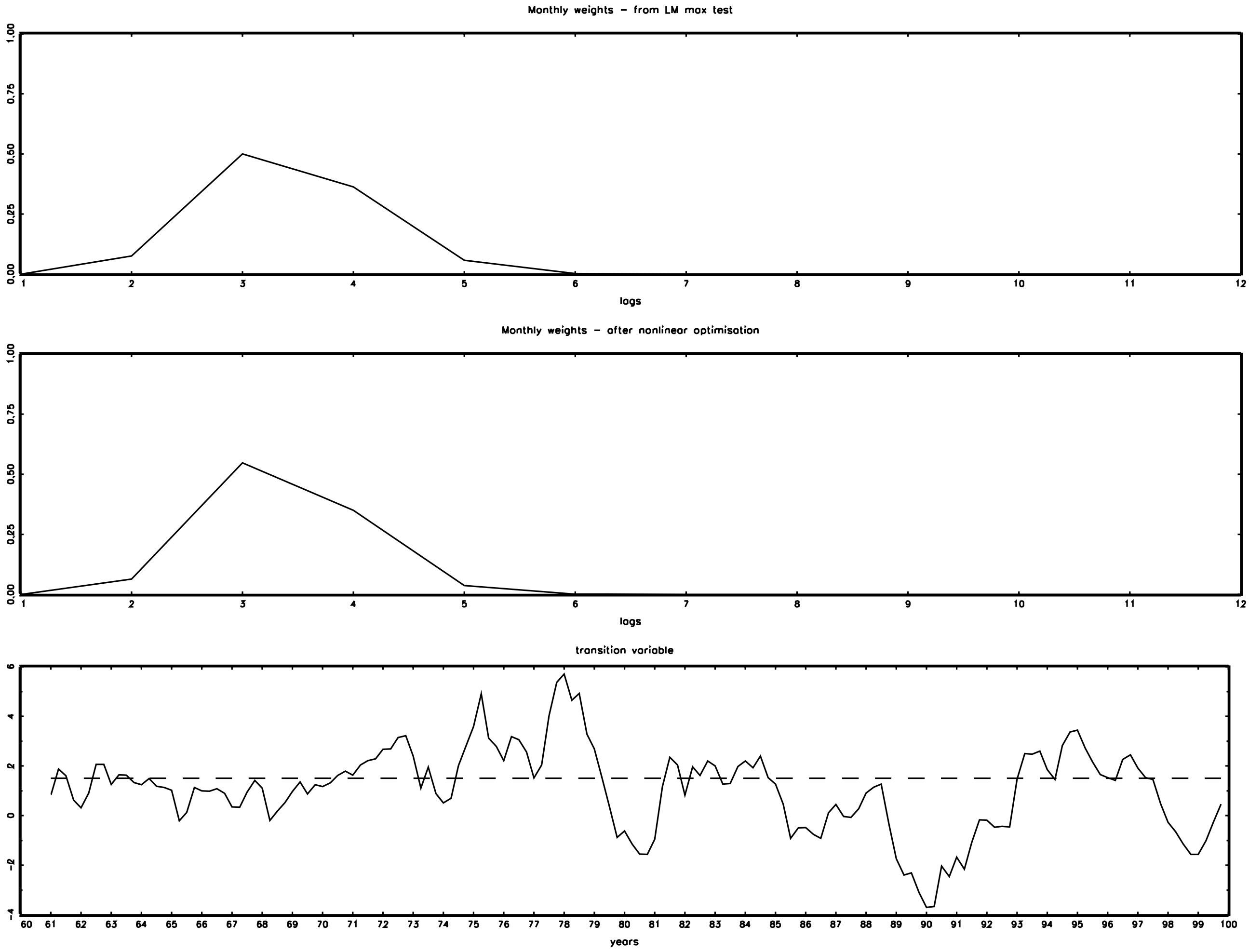

Figure 3: Top Panel: Optimal transition variable weights obtained from the testing procedure. Middle Panel: Optimal transition variable weights obtained from nonlinear least squares estimation. Bottom Panel: Transition variable and estimates threshold (dashed line).

$q=4$		$q=8$		$q=12$	
κ_1	κ_2	κ_1	κ_2	κ_1	κ_2
0.04	3.00	0.04	3.00	1.00	1.00
4.00	18.0	4.00	15.0	8.00	45.0
6.00	10.0	4.00	10.0	16.0	45.0
0.14	0.89	0.14	0.89	16.0	25.0
1.00	1.00	1.00	1.00	15.0	15.0
0.04	10.0	0.04	16.0	65.0	65.0
14.0	22.0	17.0	60.0	25.0	16.0
22.0	14.0	40.0	80.0	45.0	16.0
10.0	0.04	60.0	80.0	45.0	8.00

Table 1: Parameters used to calculate the weight parameters in the wSTAR testing procedure. Beta distributions in the left column are for a maximum lag length of $p_{max} = 4$ and in the right column for $p_{max} = 8$.

		$p=4$			$p=8$			$p=4$		
		$q=4$			$q=8$			$q=8$		
T	Test	1%	5%	10%	1%	5%	10%	1%	5%	10%
(a) Homoscedastic DGP										
200	$LM^{\max}$	0.009	0.041	0.086	0.007	0.048	0.099	0.008	0.044	0.093
	LM^{ave}	0.007	0.038	0.087	0.006	0.043	0.098	0.008	0.042	0.088
	$LM_{hc}^{\max}$	0.002	0.015	0.041	0.001	0.007	0.025	0.001	0.014	0.040
	LM_{hc}^{ave}	0.002	0.016	0.046	0.000	0.005	0.020	0.002	0.014	0.044
	$LM_{wb}^{\max}$	0.012	0.048	0.094	0.011	0.048	0.099	0.012	0.049	0.099
	LM_{wb}^{ave}	0.011	0.049	0.095	0.010	0.051	0.104	0.012	0.050	0.097
500	$LM^{\max}$	0.010	0.046	0.093	0.008	0.046	0.097	0.010	0.042	0.088
	LM^{ave}	0.010	0.041	0.089	0.009	0.043	0.090	0.009	0.040	0.082
	$LM_{hc}^{\max}$	0.002	0.021	0.058	0.002	0.015	0.043	0.003	0.019	0.054
	LM_{hc}^{ave}	0.004	0.025	0.064	0.001	0.016	0.046	0.003	0.022	0.055
	$LM_{wb}^{\max}$	0.010	0.050	0.100	0.011	0.050	0.097	0.011	0.047	0.097
	LM_{wb}^{ave}	0.012	0.051	0.102	0.012	0.050	0.096	0.011	0.045	0.095
1000	$LM^{\max}$	0.010	0.046	0.091	0.012	0.049	0.096	0.010	0.044	0.091
	LM^{ave}	0.011	0.045	0.091	0.011	0.045	0.095	0.009	0.045	0.088
	$LM_{hc}^{\max}$	0.003	0.026	0.063	0.003	0.023	0.055	0.004	0.028	0.064
	LM_{hc}^{ave}	0.005	0.033	0.073	0.004	0.024	0.061	0.006	0.032	0.069
	$LM_{wb}^{\max}$	0.010	0.049	0.097	0.013	0.052	0.101	0.013	0.050	0.102
	LM_{wb}^{ave}	0.013	0.051	0.096	0.012	0.049	0.101	0.013	0.051	0.098
(b) Heteroscedastic DGP										
200	$LM^{\max}$	0.165	0.379	0.522	0.348	0.631	0.768	0.183	0.401	0.544
	LM^{ave}	0.159	0.370	0.510	0.368	0.651	0.784	0.171	0.386	0.526
	$LM_{hc}^{\max}$	0.001	0.009	0.033	0.000	0.005	0.019	0.000	0.008	0.031
	LM_{hc}^{ave}	0.001	0.008	0.035	0.000	0.003	0.015	0.000	0.008	0.031
	$LM_{wb}^{\max}$	0.013	0.053	0.102	0.016	0.064	0.124	0.013	0.056	0.107
	LM_{wb}^{ave}	0.014	0.055	0.110	0.016	0.068	0.132	0.012	0.055	0.106
500	$LM^{\max}$	0.204	0.414	0.557	0.408	0.678	0.796	0.211	0.428	0.572
	LM^{ave}	0.195	0.401	0.540	0.438	0.698	0.812	0.199	0.414	0.557
	$LM_{hc}^{\max}$	0.002	0.016	0.050	0.001	0.012	0.036	0.003	0.017	0.048
	LM_{hc}^{ave}	0.002	0.018	0.054	0.001	0.010	0.034	0.001	0.020	0.053
	$LM_{wb}^{\max}$	0.012	0.057	0.106	0.015	0.057	0.106	0.013	0.057	0.108
	LM_{wb}^{ave}	0.012	0.053	0.104	0.014	0.059	0.118	0.013	0.054	0.106
1000	$LM^{\max}$	0.219	0.428	0.569	0.424	0.678	0.798	0.222	0.445	0.581
	LM^{ave}	0.211	0.423	0.556	0.454	0.700	0.808	0.213	0.430	0.572
	$LM_{hc}^{\max}$	0.003	0.023	0.058	0.003	0.018	0.048	0.003	0.022	0.059
	LM_{hc}^{ave}	0.004	0.026	0.069	0.002	0.022	0.057	0.004	0.026	0.068
	$LM_{wb}^{\max}$	0.013	0.055	0.107	0.014	0.060	0.113	0.012	0.055	0.112
	LM_{wb}^{ave}	0.014	0.057	0.112	0.016	0.060	0.116	0.011	0.056	0.114

Table 2: Empirical size of the $WSTR$ test based on the empirical p-values calculated according to Hansen's methodology assuming homoskedastic (no subscript) and heteroskedastic (subscript hc) disturbances, and using a wild bootstrap (subscript wb). q is the number of lags used to calculated the weighted transition variable. p is the lag length used for the linear part. All DGPs are AR(1) processes with coefficient 0.4; the disturbance variance in the heteroskedastic DGP of panel (b) doubles at the mid-point of the sample period.

		$AR(1)$			$LSTAR$ $\boldsymbol{\delta} = (1\ 0\ 0\ 0)'$			$WSTAR$ $\boldsymbol{\delta} = (1/3\ 1/3\ 1/3\ 0)'$		
T	Test	1%	5%	10%	1%	5%	10%	1%	5%	10%
200	$LM^{\max}$	0.008	0.041	0.084	0.020	0.084	0.161	0.041	0.148	0.258
	LM^{ave}	0.006	0.039	0.081	0.018	0.086	0.166	0.051	0.178	0.294
	$LM_{hc}^{\max}$	0.001	0.012	0.037	0.004	0.028	0.076	0.008	0.063	0.137
	LM_{hc}^{ave}	0.001	0.012	0.041	0.003	0.034	0.095	0.010	0.085	0.186
	$LM_{wb}^{\max}$	0.008	0.044	0.094	0.022	0.086	0.163	0.049	0.154	0.257
	LM_{wb}^{ave}	0.009	0.047	0.097	0.026	0.105	0.187	0.055	0.169	0.275
	LM^{STAR}	0.010	0.042	0.089	0.023	0.087	0.160	0.044	0.142	0.240
	$p-\min$	0.028	0.152	0.301	0.056	0.248	0.427	0.104	0.342	0.531
	LST	0.005	0.036	0.081	0.010	0.066	0.142	0.017	0.091	0.186
500	$LM^{\max}$	0.007	0.044	0.086	0.094	0.247	0.368	0.254	0.479	0.609
	LM^{ave}	0.008	0.044	0.090	0.087	0.243	0.368	0.294	0.531	0.659
	$LM_{hc}^{\max}$	0.003	0.022	0.058	0.040	0.153	0.270	0.150	0.368	0.523
	LM_{hc}^{ave}	0.003	0.026	0.064	0.041	0.171	0.297	0.194	0.448	0.610
	$LM_{wb}^{\max}$	0.012	0.052	0.099	0.097	0.251	0.371	0.277	0.505	0.635
	LM_{wb}^{ave}	0.012	0.051	0.101	0.108	0.272	0.391	0.338	0.574	0.690
	LM^{LSTAR}	0.011	0.045	0.093	0.109	0.249	0.359	0.239	0.448	0.581
	$p-\min$	0.031	0.156	0.296	0.201	0.468	0.639	0.387	0.683	0.817
	LST	0.006	0.042	0.091	0.054	0.183	0.296	0.114	0.311	0.449
1000	$LM^{\max}$	0.009	0.045	0.092	0.376	0.593	0.706	0.718	0.880	0.932
	LM^{ave}	0.008	0.046	0.092	0.305	0.551	0.679	0.754	0.904	0.948
	$LM_{hc}^{\max}$	0.003	0.025	0.065	0.267	0.507	0.638	0.638	0.840	0.909
	LM_{hc}^{ave}	0.004	0.031	0.078	0.228	0.490	0.633	0.693	0.881	0.941
	$LM_{wb}^{\max}$	0.010	0.050	0.101	0.381	0.598	0.709	0.747	0.892	0.939
	LM_{wb}^{ave}	0.009	0.053	0.106	0.334	0.570	0.695	0.788	0.917	0.957
	LM^{LSTAR}	0.009	0.048	0.094	0.382	0.606	0.713	0.688	0.859	0.919
	$p-\min$	0.033	0.166	0.313	0.542	0.784	0.883	0.813	0.950	0.981
	LST	0.008	0.045	0.095	0.210	0.444	0.579	0.452	0.696	0.810

Table 3: Empirical size and power comparisons. $WSTAR$ tests are LM^{max} and LM^{ave} (subscripts hc and wb indicate use of Hansen (1996) heteroskedasticity-consistent and wild bootstrap draws, respectively). The LM^{LSTAR} test is the LM^{max} test with the potential weight distributions restricted to approximate those admissable under an $LSTAR$ specification. $p-min$ treats the minimum $p-value$ in the auxiliary regression for known lag k as a test for nonlinearity while LST indicates the Luukkonen et al. (LST) test. In all cases p=q=4 in the test regression.