TESTING FOR LINEARITY

Bruce E. Hansen

University of Wisconsin

Abstract. The problem of testing for linearity and the number of regimes in the context of self-exciting threshold autoregressive (SETAR) models is reviewed. We describe least-squares methods of estimation and inference. The primary complication is that the testing problem is non-standard, due to the presence of parameters which are only defined under the alternative, so the asymptotic distribution of the test statistics is non-standard. Simulation methods to calculate asymptotic and bootstrap distributions are presented. As the sampling distributions are quite sensitive to conditional heteroskedasticity in the error, careful modeling of the conditional variance is necessary for accurate inference on the conditional mean. We illustrate these methods with two applications — annual sunspot means and monthly U.S. industrial production. We find that annual sunspots and monthly industrial production are SETAR(2) processes.

Keywords. SETAR models; Thresholds; Non-standard asymptotic theory; Bootstrap

1. Introduction

If a researcher proposes a non-linear time series model, the question will invariably arise: Is the non-linear specification superior to a linear model? The statistical analog is: Can you reject the hypothesis of linearity in favor of the non-linear model? This question is quite central to the analysis of self-exciting threshold autoregressive (SETAR) models. More generally, we are interested in determining the number of thresholds or regimes in a SETAR model, and hypothesis tests are useful tools in this determination.

In this paper we describe the least-squares (LS) approach to estimation and inference in SETAR models. LS methods are conceptually and computationally straightforward. As the SETAR models are nested, testing is based on classic F statistics which are computationally straightforward to calculate. Inference is complicated, however, as the asymptotic distributions of the tests are non-standard due to the presence of nuisance parameters which are only identified under the alternative hypothesis. As a result, simulation-based methods are necessary for correct inference. Luckily, with the advancement in computing power, simulation-based inference is relatively easy to implement on modern personal computers.

The author's web page is: www.ssc.wisc.edu/˜bhansen

0950-0804/99/05 0551–26 JOURNAL OF ECONOMIC SURVEYS Vol. 13, No. 5

The SETAR model, which is a particular class of piecewise linear autoregressions, is attributed to Tong (1978). For detailed reviews see Tong (1983, 1990). The subject of testing for non-linearity in the context of threshold models has been studied by Tsay (1989), Chan (1990, 1991), Chan and Tong (1990), Hansen (1996, 1997), and Caner and Hansen (1998). The testing problem is algebraically quite similar to the issue of testing for structural change of unknown timing, which dates back to Quandt (1960). Tests for single structural change in stationary models has been studied by Andrews (1993) and Andrews and Ploberger (1994), and in the context of non-stationary models by Hansen (1992, 1999a), Seo (1998), and Kuo (1998). Tests for multiple structural change have been studied by Bai (1997) and Bai and Perron (1998), and similar techniques have been applied to threshold models by Hansen (1999b). Both testing problems fall in the class of tests in the presence of unidentified nuisance parameters, which have been studied by Davies (1977, 1987), Andrews and Ploberger (1994), Hansen (1996), and Stinchcomb and White (1998).

The class of SETAR models is a restriction on the class of smooth transition autoregression (STAR) models, considered by Chan and Tong (1986), Luukkonen, *et al*. (1988), and extensively reviewed in Granger and Terasvirta (1993) and Terasvirta, *et al*. (1994). The testing issues which apply to SETAR models also apply to STAR models, and the methods discussed in this paper could easily be extended to cover STAR models as well.

SETAR and STAR models may be viewed as parsimonious approximations to general nonlinear autoregressions. While linear autoregressions dominate the empirical modeling of time series, there is no compelling a priori reason to presume that the true dynamic structure is linear. The primary argument for linearity is simplicity (estimation, interpretation, forecasting), yet current research is showing that the analysis of SETAR and STAR models is reasonably straightforward. Furthermore, there is no compelling theoretical reason to focus exclusively on linear models. Models derived from first-principles (utility and production functions) will only have linear dynamics under narrow functional form restrictions. Non-linearities becomes especially important in the presence of asymmetric costs of adjustment, irreversibilities, transactions costs, liquidity constraints, and other forms of rigidities.

Non-linear autoregressions been used in several economic applications, including: Industrial production (Terasvirta and Anderson (1992), Granger and Terasvirta (1993); GNP (Granger, Terasvirta, and Anderson (1993), Potter (1995), Hansen (1996), Galbraith (1996), Koop and Potter (1999)): Unemployment (Rothman (1991), Burgess (1992), Hansen (1997), Montgomery, *et al*. (1998), Caner and Hansen (1998); Stock volatilities (Cao and Tsay (1992). See also Brock and Potter (1993) for a review.

The organization of this paper is as follows. In Section 2 we introduce the SETAR model, and discuss the general principle of least-squares estimation and testing within the class of SETAR models. In Section 3 we introduce two time-series which will serve to illustrate the methods for the remainder of the paper. Section 4 discusses estimation methods. Explicit methods to estimate one-regime, two-regime, and three-regime SETAR models are presented. Section 5 discusses testing the SETAR(1) model against the SETAR(2) model. Asymptotic and bootstrap approximations are described, allowing both for homoskedastic and for heteroskedastic errors. Part of the purpose of this section is to show how inference can be sensitive to the assumptions and methods employed. Section 6 discusses testing SETAR(1) against SETAR(3), and Section 7 discusses testing SETAR(2) against SETAR(3). A conclusion follows.

GAUSS programs which replicate the empirical work can be downloaded from the author's webpage.

2. SETAR model classes

Let Y_t be a univariate time series and let $X_{t-1} = (1\, Y_{t-1} Y_{t-2} \ldots Y_{t-p})'$, a $k \times 1$ vector with $k = 1 + p$. A SETAR(m) model[1] takes the form

$$Y_t = \alpha_1' X_{t-1} I_{1t}(\gamma, d) + \cdots + \alpha_m' X_{t-1} I_{mt}(\gamma, d) + e_t, \tag{1}$$

where $\gamma = (\gamma_1, \ldots, \gamma_{m-1})$ with $\gamma_1 < \gamma_2 < \cdots < \gamma_{m-1}$, and $I_{jt}(\gamma, d) = I(\gamma_{j-1} < Y_{t-d} \leqslant \gamma_j)$, where $I(\cdot)$ is the indicator function[2] and we use the convention $\gamma_0 = -\infty$ and $\gamma_m = \infty$. The parameters γ_j are called the thresholds, and d is called the delay parameter. The latter may be any strictly positive integer less than some upper bound $\bar{d}$, where typically $\bar{d} = p$.

The error e_t in (1) is a uniformly square integrable martingale difference sequence, hence

$$E(e_t \mid \Im_{t-1}) = 0, \tag{2}$$

where $\Im_t$ denotes the natural filtration,[3] and $\sigma^2 = Ee_t^2 < \infty$.

A SETAR(m) model has m 'regimes', where the jth regime occurs when $I_{jt}(\gamma, d) = 1$. Our interest in this paper is the determination of the number of regimes m. The class of SETAR(m) models is strictly nested, with $m = 1$ being the most restrictive. Hence it is conceptually convenient to consider the sequence of SETAR(m) models as a sequence of nested hypotheses, which lends itself readily to hypothesis testing.

The class SETAR(1) is the class of (linear) autoregressions, which can be written as

$$Y_t = \alpha_1' X_{t-1} + e_t. \tag{3}$$

Thus testing for linearity (within the SETAR class of models) is a test of the null hypothesis of SETAR(1) against the alternative of SETAR(m) for some $m > 1$. Similarly, we can test the null hypothesis of the SETAR(2) model

$$Y_t = \alpha_1' X_{t-1} I_{1t}(\gamma, d) + \alpha_2' X_{t-1} I_{2t}(\gamma, d) + e_t, \tag{4}$$

against the alternative of a SETAR(m) for some $m > 2$.

Note that we are implicitly assuming that there are no additional constraints placed on the vectors α_j, while in some applications it may be desirable to impose constraints (such as exclusion restrictions). We will not consider such constraints in our analysis, but the following should be noted. If the same constraints are imposed on all vectors α_j, then there are no complications. If different constraints are imposed on the different vectors α_j, then the SETAR(m) classes are no longer strictly nested, so the testing problem concerns non-nested hypotheses which are more delicate to handle.

The parameters of (1) may be collected as $\theta = (\alpha_1, \alpha_2, ..., \alpha_m, \gamma, d)$. Under assumption (2) the appropriate estimation method is least-squares (LS). The LS estimator $\hat{\theta}$ solves the minimization problem

$$\hat{\theta} = \underset{\theta}{\text{argmin}} \sum_{t=1}^{n} (Y_t - \alpha_1' X_{t-1} I_{1t}(\gamma, d) - \cdots - \alpha_m' X_{t-1} I_{mt}(\gamma, d))^2. \tag{5}$$

We discuss computational solutions to this problem in Section 4. Collect the LS residuals into the $n \times 1$ vector $\hat{e}_m$. Then the sum of squared residuals is $S_m = e_m' \hat{e}_m$, a natural by-product of LS estimation.

The natural LS test of the hypothesis of SETAR(j) against SETAR(k) $(k > j)$ is to reject for large values of

$$F_{jk} = n \left(\frac{S_j - S_k}{S_k} \right). \tag{6}$$

This is the likelihood ratio test when the errors e_t are independent $N(0, \sigma^2)$. It is also the conventional F (or Wald) test, and is equivalent to the conventional Lagrange multiplier (or score) test. These observations suggest that this test is likely to have excellent power relative to alternative tests.

One important testing issue is that it is necessary to restrict the thresholds γ_j so that each regime contains a minimal number of observations. Let

$$n_j = \sum_{t=1}^{n} I_{jt}(\gamma, d),$$

be the number of sample observations in the jth regime. The asymptotic theory suggests that we should constrain the thresholds so that as $n \to \infty$, $n_j/n \geqslant \tau$ for some $\tau > 0$. While there is no clear choice for τ, a reasonable value (which we use in our applications) is $\tau = 0.1$.

3. Data

We illustrate our methods with applications to two univariate time series. The first is annual sunspot means for the time period 1700–1988. The numbers are well known, ours are taken from Appendix 3 of Tong (1990). We follow Ghaddar and Tong (1981) and make a square-root transformation $N_t = 2(\sqrt{1 + N_t^*} - 1)$, where N_t^* denotes the raw sunspot series. The series N_t is displayed in Figure 1.

Many authors have analyzed this series. In particular, Tong and Lim (1980) estimated a constrained SETAR(2) with $p = 11$ for the period 1700–1920, and Ghaddar and Tong (1981) fit a similar specification for the period 1700–1979. We follow their lead and set $p = 11$ for our applications.

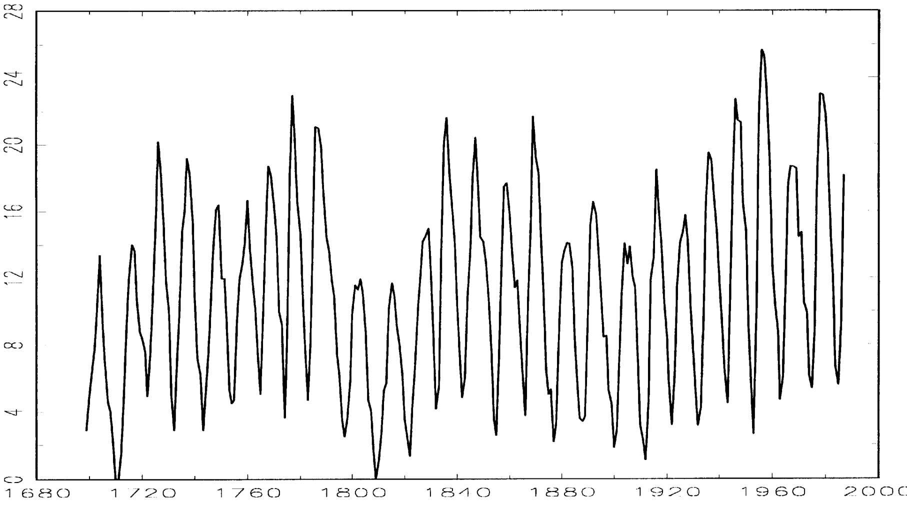

Figure 1. Annual sunspot means, 1700–1988.

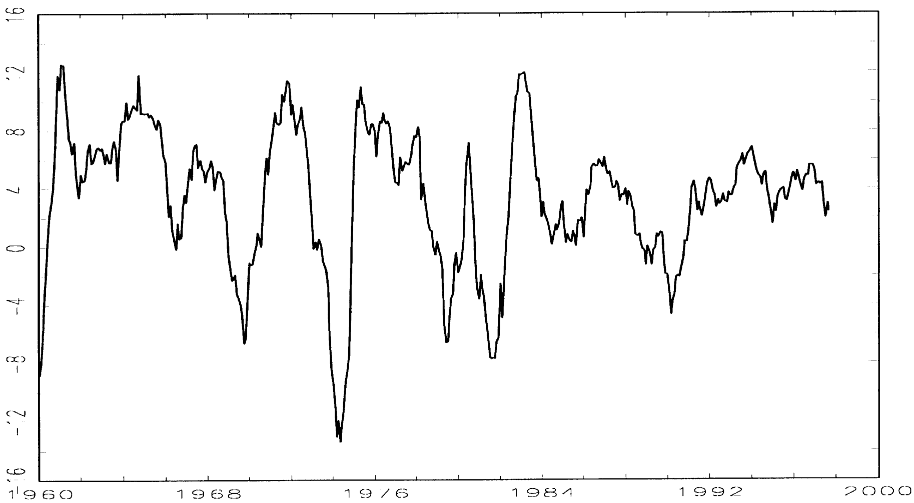

Figure 2. U.S. Monthly industrial production annual growth, 1960–1998.

The second application is to U.S. monthly industrial production for the period 1960.01 through 1998.09. Smooth SETAR models have been used to model quarterly industrial production by Terasvirta and Anderson (1992). See also Section 9.1 of Granger and Terasvirta (1993). SETAR models have been fit to a similar series (U.S. GNP) by Potter (1995) and Hansen (1996). We transform the series to approximate stationarity by taking annualized growth rates setting

$Q_t = 100 * (\ln(IP_t) - \ln(IP_{t-12}))$, where IP_t denotes the raw industrial production series. The series Q_t is displayed in Figure 2. Some experimentation with a baseline SETAR(1) model suggested that $p = 16$ is sufficient to reduce the errors to white noise.

4. Estimation

4.1. *SETAR(1) model*

The SETAR(1) is the linear autoregression (3). The solution to the LS problem (5) is ordinary least squares (OLS):

$$\hat{\alpha}_1 = (X'X)^{-1}(X'Y),$$

where X is the $n \times k$ matrix whose ith row is X'_{t-1} and Y is the $n \times 1$ vector whose ith element is y_t. The residual vector is $\hat{e}_1 = Y - X\hat{\alpha}_1$ and the sum of squared errors is $S_1 = \hat{e}'_1\hat{e}_1$. Table 1 reports the OLS estimates of SETAR(1) models fit to our two data series. The reported standard errors (here and throughout the paper) are computed as in White (1980) to be robust to heteroskedasticity of unknown form.

As in many testing contexts, the sampling distribution of the test statistics F_{12} and F_{13} will depend upon the conditional variance properties of the error e_t. To

Table 1. Least squares estimates of SETAR(1) models.

	Annual sunspots		Industrial production	
	$\hat{\alpha}_1$	s.e.	$\hat{\alpha}_1$	s.e.
Const.	1.39	(0.45)	0.22	(0.07)
y_{t-1}	1.22	(0.07)	1.22	(0.06)
y_{t-2}	−0.48	(0.12)	−0.13	(0.08)
y_{t-3}	−0.15	(0.12)	0.01	(0.07)
y_{t-4}	0.27	(0.10)	−0.05	(0.06)
y_{t-5}	−0.24	(0.10)	−0.14	(0.06)
y_{t-6}	0.01	(0.09)	−0.04	(0.06)
y_{t-7}	0.16	(0.09)	0.13	(0.06)
y_{t-8}	−0.21	(0.10)	0.03	(0.07)
y_{t-9}	0.30	(0.10)	−0.02	(0.07)
y_{t-10}	0.02	(0.10)	0.03	(0.06)
y_{t-11}	−0.02	(0.06)	−0.04	(0.07)
y_{t-12}			−0.49	(0.08)
y_{t-13}			0.57	(0.09)
y_{t-14}			−0.13	(0.08)
y_{t-15}			0.15	(0.07)
y_{t-16}			−0.16	(0.05)
S_1	1135		362	
n	278		437	
$\hat{\sigma}_1^2$	4.08		0.817	

assess the presence of conditional heteroskedasticity, we regressed the squared residual on squares of the regressors, and tested the joint significance of the regressors using a Wald statistic. For the sunspot series N_t, the statistic was 58, which is highly significant with a p-value near zero (based on the $\chi^2(11)$ distribution). For the industrial production series Q_t, the statistic was 15, with an asymptotic p-value (based on the $\chi^2(16)$ distribution) of 0.52. This difference will turn out to be important in the inference procedures we discuss later.

4.2. *SETAR(2) model*

In the SETAR(2) model, $\gamma = \gamma_1$, so we let $I_{1t}(\gamma, d) = I(Y_{t-d} \leqslant \gamma)$ and $I_{2t}(\gamma, d) = I(\gamma < Y_{t-d})$. Let $\alpha = (\alpha_1' \alpha_2')'$ and

$$X_{t-1}(\gamma, d) = \begin{pmatrix} X_{t-1} I_{1t}(\gamma, d) \\ X_{t-1} I_{2t}(\gamma, d) \end{pmatrix}.$$

Let $X(\gamma, d)$ be the $n \times 2k$ matrix whose ith row is $X_{t-1}(\gamma, d)'$.

Observe that the minimization problem (5)

$$S_2 = \min_{d, \gamma, \alpha} (Y - X(\gamma, d)\alpha)'(Y - X(\gamma, d)\alpha)$$

can be solved sequentially through concentration. That is, for given (d, γ), minimization over α is an OLS regression of Y on $X(\gamma, d)$. We can write the solution as

$$\hat{\alpha}(\gamma, d) = (X(\gamma, d)'X(\gamma, d))^{-1}(X(\gamma, d)'Y). \tag{7}$$

Let

$$S_2(\gamma, d) = (Y - X(\gamma, d)\hat{\alpha}(\gamma, d))'(Y - X(\gamma, d)\hat{\alpha}(\gamma, d))$$

be the residual sum of squared errors for given (d, γ). Then

$$(\hat{\gamma}_1, \hat{d}) = \underset{\gamma, d}{\operatorname{argmin}} \; S_2(\gamma, d). \tag{8}$$

Once the solution to (8) is found, we find $\hat{\alpha}$ through (7), vis $\hat{\alpha} = \hat{\alpha}(\hat{\gamma}_1, \hat{d})$, and then obtain $S_2 = S_2(\hat{\gamma}_1, \hat{d})$ and $F_{12} = n((S_1 - S_2)/S_2)$ as natural by-products.

Now observe that

$$S_2(\gamma, d) = Y'(I - X(\gamma, d)(X(\gamma, d)'X(\gamma, d))^{-1}X(\gamma, d)')Y \tag{9}$$

involves $X(\gamma, d)$ only through a projection, so the result is invariant to linear reparameterizations of $X(\gamma, d)$, and in particular, we can redefine $X(\gamma, d) = [X \quad X(\gamma, d)]$, where $X(\gamma, d)$ is the matrix whose ith row is $X_{t-1} I_{1t}(\gamma, d)$. Noting the identity $Y = X\hat{\alpha}_1 + \hat{e}_1$, and since X lies in the space spanned by $X(\gamma, d)$, this means that we can replace Y in (9) by $\hat{e}_1$. Since $X'\hat{e}_1 = 0$, standard partitioned matrix inversion calculations and the fact that $X_1(\gamma, d)'X = X_1(\gamma, d)'X_1(\gamma, d)$ show that

$$\begin{aligned} S_2(\gamma, d) &= \hat{e}_1'\hat{e}_1 - \hat{e}_1'X_1(\gamma, d)M_n^*(\gamma, d)^{-1}X_1(\gamma, d)'\hat{e}_1 \\ &= S_1 - f_2(\gamma, d) \end{aligned}$$

where

$$M_n^*(\gamma, d) = X_1(\gamma, d)'X_1(\gamma, d) - (X_1(\gamma, d)'X_1(\gamma, d))(X'X)^{-1}(X_1(\gamma, d)'X_1(\gamma, d)) \tag{10}$$

and

$$f_2(\gamma, d) = \hat{e}_1'X_1(\gamma, d)M_n^*(\gamma, d)^{-1}X_1(\gamma, d)'\hat{e}_1. \tag{11}$$

We thus see that the minimization (8) can be equivalently achieved through maximization of $f_2(\gamma, d)$. It is also interesting to observe that we can rewrite the linearity test statistic as

$$F_{12} = n\left(\frac{f_2(\hat{\gamma}_1, \hat{d})}{S_1 - f_2(\hat{\gamma}_1, \hat{d})}\right). \tag{12}$$

The maximization (11) is best solved through a grid search, noting that the argument d is discrete and that the function $f_2(\gamma, d)$ is typically a highly erratic function of γ. Since the parameter γ only arises through the indicator functions $I(Y_{t-d} \leqslant \gamma)$, there is no loss in restricting the search to the observed values of Y_{t-d}. The requirement that $n_1 \geqslant n\tau$ and $n_2 \geqslant n\tau$ further restricts the search to values of Y_{t-d} lying between the τth and $(1-\tau)$th quantiles. For the bootstrap methods we discuss latter, when n and p are large a full grid search can prove too costly for all but the most patient researchers. A close approximation can be achieved by restricting the search to N values of γ lying on a grid between the τth and $(1-\tau)$th quantiles of Y_{t-d}. If $\bar{d} = p$, then a joint search over (d, γ) will require pN function evaluations. For the empirical work we report here, we set $N = 100$. Since we set $p = 11$ in the sunspot application (and $p = 16$ in the industrial production application), this means that the maximization (11) requires a grid search over 1100 (respectively 1600) pairs of (γ, d). While this may seem like an intensive search, it only takes a few seconds on a personal computer. For example, a GAUSS program running on a 400 Mhz Pentium II computes the SETAR(2) model for the sunspot series in 2.5 seconds, and for the industrial production series in 7.6 seconds.

We report in Table 2 our estimates of the SETAR(2) models for our two data sets. For the sunspot series N_t, we find $\hat{d} = 2$ and $\hat{\gamma}_1 = 7.42$. For the industrial production series Q_t, we find $\hat{d} = 6$ and $\hat{\gamma}_1 = 0.226$. For both series, we find the F_{12} statistic for the test of SETAR(1) against SETAR(2) equals 70. The sampling distribution of F_{23}, the test for SETAR(2) versus SETAR(3), will depend on whether the SETAR(2) errors e_t are conditionally heteroskedastic. We assessed this through an OLS regression of the squared LS residual on the squares of the

Table 2. Least squares estimates of SETAR(2) models.

	Annual sunspots				Industrial production			
	$N_{t-2} \leqslant 7.4$		$N_{t-2} > 7.4$		$Q_{t-6} \leqslant 0.23$		$Q_{t-6} > 0.23$	
	$\hat{\alpha}_1$	s.e.	$\hat{\alpha}_2$	s.e.	$\hat{\alpha}_1$	s.e.	$\hat{\alpha}_2$	s.e.
Const.	−0.58	(0.90)	2.32	(0.55)	−0.07	(0.13)	0.10	(0.11)
Y_{t-1}	1.22	(0.10)	0.95	(0.08)	1.34	(0.11)	1.04	(0.06)
Y_{t-2}	−0.97	(0.26)	−0.03	(0.11)	−0.37	(0.18)	0.03	(0.08)
Y_{t-3}	0.49	(0.29)	−0.48	(0.10)	−0.01	(0.13)	0.07	(0.07)
Y_{t-4}	−0.19	(0.26)	0.32	(0.09)	−0.07	(0.13)	−0.05	(0.07)
Y_{t-5}	−0.14	(0.28)	−0.21	(0.08)	0.06	(0.12)	0.21	(0.07)
Y_{t-6}	0.12	(0.26)	−0.04	(0.08)	−0.22	(0.11)	0.04	(0.07)
Y_{t-7}	0.13	(0.21)	0.18	(0.08)	0.16	(0.12)	0.16	(0.07)
Y_{t-8}	−0.22	(0.23)	−0.22	(0.09)	0.13	(0.11)	−0.02	(0.06)
Y_{t-9}	0.46	(0.26)	0.19	(0.09)	−0.11	(0.11)	0.02	(0.07)
Y_{t-10}	−0.07	(0.20)	−0.02	(0.09)	0.03	(0.14)	0.03	(0.06)
Y_{t-11}	−0.07	(0.12)	0.12	(0.07)	0.04	(0.19)	−0.04	(0.06)
Y_{t-12}					−0.88	(0.19)	−0.37	(0.09)
Y_{t-13}					0.95	(0.12)	0.40	(0.09)
Y_{t-14}					−0.32	(0.16)	−0.07	(0.08)
Y_{t-15}					0.25	(0.16)	0.16	(0.07)
Y_{t-16}					−0.25	(0.12)	−0.15	(0.05)
n_j	86		192		96		341	
S_2	907				312			
F_{12}	70				70			

lagged dependent variable, and on dummy variables indicating the regime. These results are reported in Table 3. The F statistic for the exclusion of all variables other than an intercept (F_{hetero}) is highly significant for the sunspot series, but not for the industrial production series.

4.3. *SETAR(3) model*

The SETAR(3) model is

$$Y_t = \alpha_1' X_{t-1} I_{1t}(\gamma, d) + \alpha_2' X_{t-1} I_{2t}(\gamma, d) + \alpha_3' X_{t-1} I_{3t}(\gamma, d) + e_t, \qquad (13)$$

where $\gamma = (\gamma_1, \gamma_2)$. In principle, this model can be estimated using the same techniques described in the previous section, namely conditional on (γ, d), the parameters $(\alpha_1, \alpha_2, \alpha_3)$ may be estimated by OLS, and then a grid search over (γ, d) yields the LS estimates. The difficulty is that if N points are evaluated at each of γ_1 and γ_2, then this search involves $p \times N^2$ OLS regressions. While such estimation is feasible, it does not lend itself easily to bootstrap evaluation of the test statistics. (Estimation would take about 12 minutes for the industrial production series, and 1000 bootstrap replications would take about 200 hours.)

Table 3. Least squares estimates of SETAR(2) conditional variance.

	Annual sunspots		Industrial production	
I_{1t}	3.765	(0.880)	0.761	(143)
I_{2t}	1.858	(0.698)	0.504	(0.084)
y_{t-1}^2	0.001	(0.005)	−0.007	(0.006)
y_{t-2}^2	−0.000	(0.006)	0.017	(0.008)
y_{t-3}^2	−0.005	(0.007)	−0.011	(0.006)
y_{t-4}^2	0.008	(0.008)	0.004	(0.005)
y_{t-5}^2	−0.011	(0.009)	−0.004	(0.005)
y_{t-6}^2	0.022	(0.010)	0.004	(0.005)
y_{t-7}^2	−0.028	(0.011)	0.000	(0.006)
y_{t-8}^2	0.027	(0.013)	−0.003	(0.005)
y_{t-9}^2	−0.005	(0.009)	−0.002	(0.005)
y_{t-10}^2	−0.005	(0.008)	0.014	(0.006)
y_{t-11}^2	0.000	(0.004)	−0.011	(0.007)
y_{t-12}^2			−0.000	(0.009)
y_{t-13}^2			0.003	(0.007)
y_{t-14}^2			0.001	(0.007)
y_{t-15}^2			−0.008	(0.006)
y_{t-16}^2			0.008	(0.004)
F_{hetero}	48.1		20.8	

Fortunately, a computational short-cut was proposed by Bai (1997) and Bai and Perron (1998) in the change-point literature. Arguments analogous to those suggested by these authors show that if the true model is (13), but the (misspecified) SETAR(2) model (4) is actually estimated, the least-squares estimate $\hat{d}$ will be consistent for d and $\hat{\gamma}_1$ will be consistent for one of the pair (γ_1, γ_2). They show further that if $\gamma = (\gamma_1, \gamma_2)$ is estimated by least-squares on (13), enforcing the constraint that $d = \hat{d}$ and that one element of γ equals $\hat{\gamma}_1$, then the second-stage estimate $\hat{\gamma}_2$ will be consistent for the remaining element of the pair (γ_1, γ_2). Thus this two-step method yields consistent estimation of $\hat{d}$ and $\hat{\gamma} = (\hat{\gamma}_1, \hat{\gamma}_2)$. Furthermore, Bai (1997) shows that these estimates can be made asymptoticaly efficient, in the sense of having the same asymptotic distribution as estimates obtained from joint estimation of (13), if this method is iterated at least once. That is, $\gamma = (\gamma_1, \gamma_2)$ is estimated by least-squares on (13), enforcing the constraint that $d = \hat{d}$ and that one element of γ equals $\hat{\gamma}_2$, yielding a refined estimate $\hat{\gamma}_1$. Further iteration does not affect the asymptotic distribution, but may yield finite-sample improvements.

This 'one-step-at-a-time' approach yields enormous computational savings. Rather than pN^2 function evaluations, it involves approximately $pN + 2N$ function evaluations, which is only a minor increase over the requirements for estimation of the SETAR(2) model.

It is important to impose the requirement that all three regimes have at least $n\tau$ observations. In addition to the restrictions imposed on the search discussed in

Table 4. Least squares estimates of SETAR(3) models.

	Annual sunspots						Industrial production					
	$N_{t-2} \leqslant 5.3$		$5.3 < N_{t-2} \leqslant 8.0$		$N_{t-2} > 8.0$		$Q_{t-6} \leqslant -2.5$		$-2.5 < Q_{t-6} \leqslant 0.35$		$Q_{t-6} > 0.35$	
	$\hat{\alpha}_1$	s.e.	$\hat{\alpha}_2$	s.e.	$\hat{\alpha}_3$	s.e.	$\hat{\alpha}_1$	s.e.	$\hat{\alpha}_2$	s.e.	$\hat{\alpha}_3$	s.e.
Const.	0.037	(1.14)	9.08	(2.99)	2.31	(0.55)	0.28	(0.29)	−0.06	(0.20)	0.11	(0.11)
y_{t-1}	1.57	(0.12)	1.09	(0.10)	0.91	(0.07)	1.03	(0.16)	1.53	(0.12)	1.04	(0.06)
y_{t-2}	−1.18	(0.33)	−1.10	(0.39)	−0.02	(0.11)	−0.13	(0.22)	−0.52	(0.18)	0.04	(0.08)
y_{t-3}	0.62	(0.37)	−0.11	(0.36)	−0.44	(0.09)	−0.03	(0.17)	−0.02	(0.14)	0.08	(0.07)
y_{t-4}	−0.57	(0.28)	0.00	(0.29)	0.27	(0.08)	0.15	(0.19)	−0.31	(0.18)	−0.05	(0.07)
y_{t-5}	0.36	(0.24)	−1.11	(0.38)	−0.17	(0.08)	0.03	(0.16)	0.31	(0.17)	−0.21	(0.07)
y_{t-6}	−0.31	(0.26)	0.70	(0.41)	−0.05	(0.08)	−0.28	(0.17)	−0.36	(0.22)	−0.05	(0.07)
y_{t-7}	0.43	(0.22)	0.11	(0.28)	0.16	(0.08)	0.11	(0.13)	0.34	(0.21)	0.16	(0.07)
y_{t-8}	−0.30	(0.24)	0.58	(0.25)	−0.21	(0.09)	0.14	(0.16)	0.09	(0.17)	−0.03	(0.06)
y_{t-9}	0.30	(0.26)	−0.48	(0.27)	0.17	(0.09)	−0.37	(0.19)	0.03	(0.11)	0.03	(0.07)
y_{t-10}	−0.02	(0.24)	0.33	(0.26)	0.03	(0.09)	0.24	(0.23)	−0.12	(0.15)	0.03	(0.06)
y_{t-11}	−0.02	(0.13)	−0.32	(0.15)	0.12	(0.06)	0.11	(0.28)	−0.10	(0.18)	−0.04	(0.06)
y_{t-12}							−0.96	(0.26)	−0.65	(0.20)	−0.36	(0.09)
y_{t-13}							0.74	(0.31)	0.97	(0.13)	0.39	(0.09)
y_{t-14}							−0.17	(0.31)	−0.34	(0.16)	−0.07	(0.08)
y_{t-15}							0.27	(0.31)	0.25	(0.14)	0.15	(0.07)
y_{t-16}							−0.16	(0.18)	−0.30	(0.10)	−0.15	(0.05)
n_j	58		36		184		46		53		338	
S_2	769						294					
F_{13}	132						101					
F_{23}	50						27					

Section 4.2, we need to impose the requirement in the second- and third-stage searches that at least $n\tau$ observations lie in the regime where $\gamma_1 \leqslant Y_{t-d} \leqslant \gamma_2$ (or $\gamma_2 \leqslant Y_{t-d} \leqslant \gamma_1$ if $\gamma_2 < \gamma_1$).

In Table 4 we report the least-squares estimates of the SETAR(3) models for our two time-series. For the sunspot series, the two thresholds are 5.32 and 8.04. We find that the F_{13} statistic for the test of SETAR(1) against SETAR(3) is 132 and the F_{23} statistic for the test of SETAR(2) against SETAR(3) is 50.

For the industrial production series, the two thresholds are -2.53 and 0.348. The F_{13} statistic is 101 and the F_{23} statistic is 27.

5. Testing SETAR(1) against SETAR(2)

5.1. *Homoskedasticity*

While the standard theory of hypothesis testing suggests that a good test of the SETAR(1) model against the SETAR(2) alternative is to reject for large values of the statistic F_{12}, the test cannot be implemented unless we know the distribution of F_{12} under the null hypothesis, as this is the only way to control the Type I error of the test.

We start by imposing the assumption of conditional homoskedasticity

$$E(e_t^2 \mid \Im_{t-1}) = \sigma^2 \tag{14}$$

and consider the general case of conditional heteroskedasticity in the next section.

In most testing contexts, test statistics such as F_{12} can be expected to have an asymptotic $\chi^2(k)$ distribution under (14). In the present context, however, this is not the case. This can perhaps best be seen by examining the form of the statistic F_{12} as defined in (11) and (12) Let

$$F_{12}(\gamma, d) = n\left(\frac{f_2(\gamma, d)}{S_1 - f_2(\gamma, d)}\right),$$

which is a monotonically increasing function of $f_2(\gamma, d)$. Since $F_{12} = F_{12}(\hat{\gamma}_1, \hat{d})$ and $(\hat{\gamma}_1, \hat{d})$ maximize $f_2(\gamma, d)$, it follows that

$$F_{12} = \max_{\gamma, d} F_{12}(\gamma, d). \tag{15}$$

Now $F_{12}(\gamma, d)$ is a fairly conventional test statistic. It is equivalent to the test for the exclusion of $X_1(\gamma, d)$ (with (γ, d) fixed) from a regression of Y on X and $X_1(\gamma, d)$. If the data are weakly stationary and satisfy standard regularity conditions, we can show that for any fixed (γ, d), $F_{12}(\gamma, d)$ has an asymptotic $\chi^2(k)$ distribution. Now the problem is that the maximization (15) involves not just a single value of (γ, d), but a very large number of values. Our proposed implementation involves maximization over pN values of (γ, d) (which is 1100 for the sunspot application), so we are taking the maximum of pN distinct asymptotic chi-square random variables. Thus the distribution of F_{12} is distinctly greater than the $\chi^2(k)$. Thus if F_{12}

is not significant when compared to the $\chi^2(k)$, it will be certainly not significant when compared to the correct asymptotic distribution. In most applications this will not be a helpful bound, however, as typically the observed value of F_{12} will be very 'significant' when compared to the $\chi^2(k)$ distribution.

Thus it is not helpful to think of the statistic $F_{12}(\gamma, d)$ for fixed values of (γ, d). Instead, we need to think of $F_{12}(\gamma, d)$ as a random function of the arguments (γ, d), and view F_{12} (as defined in (15)) as the random maximum of this random function. To develop an asymptotic distribution theory for this statistic, we therefore need an asymptotic theory appropriate for random functions, which is known as empirical process theory. A good review can be found in Andrews (1994). For the stationary SETAR model under (14), Hansen (1996) has shown that the asymptotic distribution of the empirical process $F_{12}(\gamma, d)$ is

$$F_{12}(\gamma, d) \Rightarrow T(\gamma, d)$$

where

$$\begin{aligned} T(\gamma, d) &= G(\gamma, d)' M^*(\gamma, d)^{-1} G(\gamma, d) \\ M^*(\gamma, d) &= M(\gamma, d) - M(\gamma, d) M^{-1} M(\gamma, d), \\ M &= E(X_{t-1} X'_{t-1}) \\ M(\gamma, d) &= E(X_{t-1} X'_{t-1} I_{1t}(\gamma, d)) \end{aligned} \tag{16}$$

and $G(\gamma, d)$ is a mean-zero Gaussian process with covariance kernel

$$E[G(\gamma, d) G(\gamma', d')'] = E(X_{t-1} X'_{t-1} I_{1t}(\gamma, d) I_{1t}(\gamma', d')) - M(\gamma, d) M^{-1} M(\gamma', d'). \tag{17}$$

That is, for fixed (γ, d) the distribution of $G(\gamma, d)$ is multivariate normal with covariance matrix $M^*(\gamma, d)$, and all pairs $(G(\gamma, d), G(\gamma', d'))$ are jointly normal with covariance given in (17). Thus $G(\gamma, d)$ is a random function with arguments (γ, d).

Note that for fixed (γ, d) the random variable $T(\gamma, d)$ is $\chi^2(k)$. Thus for fixed (γ, d), $F_{12}(\gamma, d) \rightarrow_d \chi^2(k)$. The statistic F_{12} is the maximum of this random function, so coverges in distribution to the maximum of this random limit function, or

$$F_{12} \underset{d}{\rightarrow} T = \max_{\gamma, d} T(\gamma, d).$$

While for fixed (γ, d), $T(\gamma, d)$ is $\chi^2(k)$, the distribution of T is less easy to characterize. Its distribution depends to a great extent on the degree of dependence between the random variables $T(\gamma, d)$ for distinct values of (γ, d), which is determined through the covariance functional (17), and thus by moments of the regressors X_{t-1} and the threshold variables Y_{t-d}. Since the distribution of T depends upon these moments (which are application-specific), the distribution T cannot be tabulated for general use. Rather, critical values and p-values must be calculated for each and every application.

Hansen (1996) describes an algorithm to calculate the asymptotic distribution.[4] It can be described as follows. In the formula for the asymptotic distribution T given in (16), replace all population moments by sample counterparts (e.g., replace M by $M_n = n^{-1}X'X$) to define a random variable T_n. Since the sample moments are consistent estimates of the population moments, T_n is the asymptotic approximation of interest. An *exact* draw from the asymptotic distribution T_n can be made by letting u denote a random $N(0, I_n)$ vector, $\hat{u} = u - X(X'X)^{-1}X'u$, and then setting

$$T_n = \max_{\gamma, d} \ \hat{u}'(\gamma, d) M_n^*(\gamma, d)^{-1} X_1(\gamma, d)'\hat{u} \tag{18}$$

where $M_n^*(\gamma, d)$ is defined in (10). This is similar (and asymptotically equivalent), to a bootstrap replication where u is treated as the dependent variable, and the regressors X_{t-1} and threshold variables Y_{t-d} are held fixed at their sample values. To calculate the distribution of T_n, a large number (we use 2000) of independent draws are made from (18). Then critical values may be calculated from the quantiles of these draws, or better yet, a p-value may be calculated by counting the percentage of the draws which exceed the observed F_{12}.

This is similar to a bootstrap, but it should not be confused with a bootstrap distribution. The distribution T_n is the asymptotic distribution of the test statistic, and simply the fact that a simulation is used to compute the p-value does not make it more accurate than any other asymptotic approximation. The main advantage of the calculation of this asymptotic distribution is that it is computationally less costly than a bootstrap calculation. Since most of the computational work in implementing (18) comes through the matrix inversion of

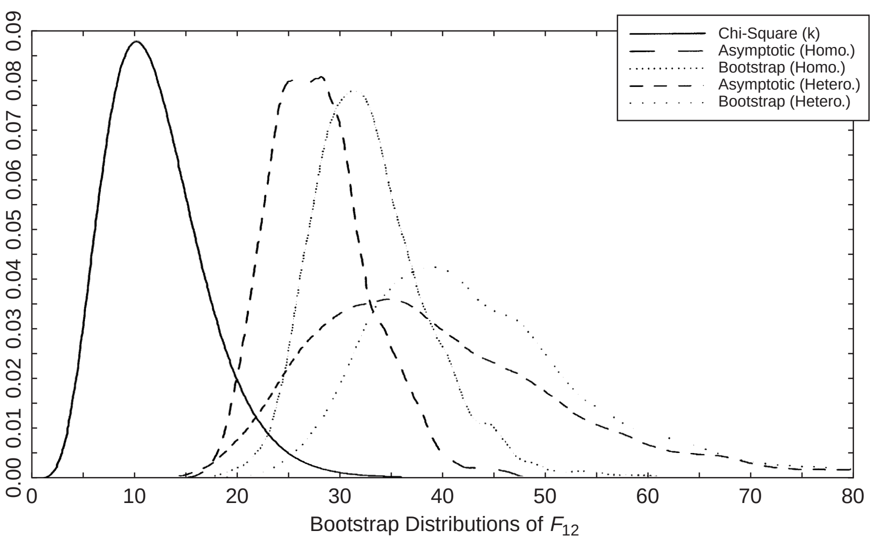

Figure 3. Sunspot series asymptotic and bootstrap distributions of F_{12}.

$M_n^*(\gamma, d)$, and these are constant across draws of T_n, computational savings may be made if the matrices $M_n^*(\gamma, d)^{-1}$ are stored, and not re-calculated for each draw. The computational savings are such that this algorithm takes only one-quarter the time of the bootstrap method to be described next.

The asymptotic distributions (18) were calculated using 2000 independent draws for each of our two time-series. Estimates[5] of the density functions are plotted (long dashes) in Figures 3 and 4, and labeled 'Asymptotic (Homo.)'. The $\chi^2(k)$ density is also plotted (solid line) for reference. It is clear that the $\chi^2(k)$ distribution is highly misleading relative to the asymptotic distribution. Still, the observed value of the test statistic F_{12} is 70 for each application, which is in the far right tail of the asymptotic distribution, so the observed value appears to be highly significant. In Table 5 we report the asymptotic p-value, which is 0.000 in both applications, since none of the 2000 simulations exceeded the test statistic in the observed sample.

Since a fair amount of computation is involved in calculating the asymptotic p-value, one might ask: Why not make a little extra effort and calculate the p-values using a bootstrap approximation? There is a considerable body of statistical theory (e.g. Hall (1992), Shao and Tu (1995), Davison and Hinkley (1997)) that the bootstrap is a better approximation to finite sample distributions than first-order asymptotic theory. Under certain technical conditions (such as the existence of an Edgeworth expansion), the bootstrap distribution of an asymptotically pivotal statistic achieves a higher rate of convergence to the sampling distribution than the first-order asymptotic approximation. These conditions have not been verified for the SETAR model (and may in fact not hold) so it is unclear if the boostrap will achieve an accelerated rate of convergence.

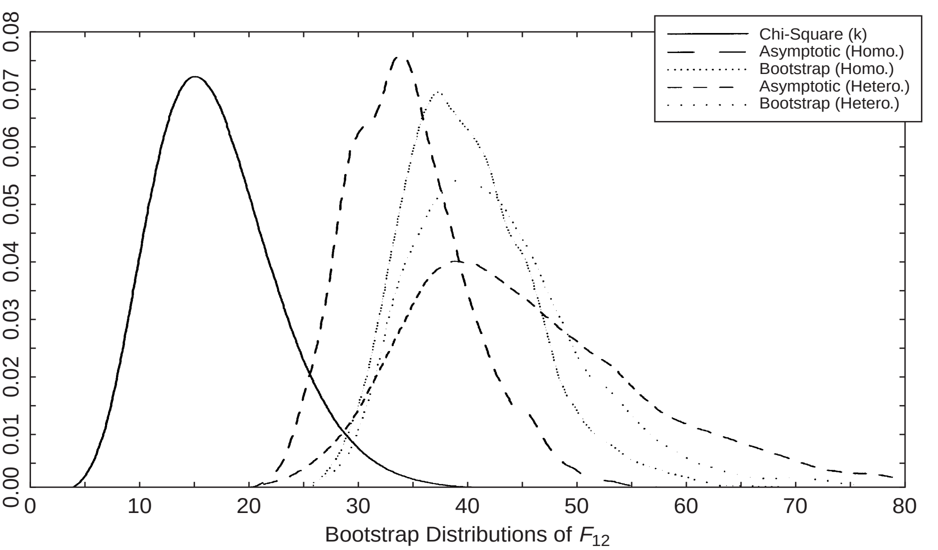

Figure 4. Industrial production asymptotic and bootstrap distributions of F_{12}.

Table 5. Asymptotic and bootstrap tests of SETAR(1) against SETAR(2).

		Homoskedastic p-values		Heteroskedastic p-values	
	F_{12}	Asymptotic	Bootstrap	Asymptotic	Bootstrap
Sunspots	70	0.000	0.000	0.030	0.031
Industrial production	70	0.000	0.001	0.047	0.010

An argument in favor of the bootstrap is that it appears to work globally in the parameter space. While the asymptotic approximation of Hansen (1996) outlined above requires that the process Y_t be stationary, excluding unit roots or near unit roots, Caner and Hansen (1998) show that the bootstrap achieves a good approximation even if there is a unit root or near unit root.

While there is no Monte Carlo study comparing the bootstrap and asymptotic approximations in the context of testing for SETAR models, Diebold and Chen (1996) present an analogous study of the Andrews structural change test in the AR(1) model. They find that the bootstrap yields an excellent approximation, a certain improvement over the asymptotic distribution.

There is no free lunch, and the downside to any bootstrap implementation is that it requires taking a position on the distribution of the errors e_t. This requires the imposition of more structure on the model than might be desirable. For example, the bootstrap algorithm that we now describe imposes the condition that the errors e_t are independent of $\Im_{t-1}$, which is considerably stronger than the martingale difference assumption (2) and the homoskedasticity condition (14).

An appropriate bootstrap distribution will calculate the distribution of the statistic F_{12} under the assumption that the data satisfy the SETAR(1) hypothesis and the parameters are calibrated to match the observed data. The natural method to do this is to use the SETAR(1) estimates and add an auxiliary assumption on the errors e_t. The assumption we make is that the errors e_t are independent over time, and estimate the distribution of the errors by the empirical distribution of the SETAR(1) residuals $\hat{e}_t$. The bootstrap distribution also depends on how the initial conditions $(y_0, y_{-1}, y_{-2}, \ldots, y_{-p+1})$ are modeled. We take the simple approach of conditioning on the observed values, so hold these values fixed in repeated samples.

Thus the algorithm is as follows. Generate a random sample e_t^*, $t = 1, \ldots, n$ by sampling (with replacement) from the OLS residuals from the SETAR(1) model. Then using the fixed initial conditions $(y_0, y_{-1}, y_{-2}, \ldots, y_{-p+1})$, recursively generate a sample y_t^*, $t = 1, \ldots, n$ using the SETAR(1) model (3) with the parameter $\hat{\alpha}_1$ taken from the SETAR(1) estimates. On this simulated series y_t^*, calculate the statistic F_{12}^* using the same methods as to calculate F_{12} on the actual series. Repeat this a large number of times. (We make 2000 replications.) The bootstrap p-value is the percentage of simulated F_{12}^* which exceed the observed F_{12}.

Estimated density functions for this bootstrap distribution of F_{12}^* are displayed using dotted lines (closely spaced) in Figures 3 and 4, labeled 'Bootstrap (Homo.)'. In both applications, the bootstrap distribution takes a similar shape to the asymptotic distribution, but is noticeably shifted to the right and has a thicker right tail. There is no simple explanation for this phenomenon, but it suggests that for many potential values of F_{12}, the asymptotic distribution and bootstrap distribution would give contrary results. In our applications, the value of F_{12} is sufficiently high that there is no meaningful difference between the asymptotic and bootstrap p-values reported in Table 5.

5.2. *Heteroskedasticity*

The previous section evaluated the distribution of F_{12} under homoskedasticity of the error term. This might seem innocuous, but is actually quite powerful. As we discussed in Section 4.1, there is strong evidence that this assumption is violated for at least the sunspot series. We now turn to sampling approximations which do not impose (14).

Hansen (1996) has shown how to calculate the asymptotic distribution for the case of stationary data with possibly heteroskedastic error terms. It is identical to (16), except that $G(\gamma, d)$ is a mean-zero Gaussian process with covariance kernel

$$\begin{aligned} E[G(\gamma, d)G(\gamma', d')'] = E\left(X_{t-1}X'_{t-1}I_{1t}(\gamma, d)I_{1t}(\gamma', d')\frac{e_t^2}{\sigma^2}\right) \\ - M(\gamma, d)M^{-1}E\left(X_{t-1}X'_{t-1}I_{1t}(\gamma', d')\frac{e_t^2}{\sigma^2}\right) \\ - E\left(X_{t-1}X'_{t-1}I_{1t}(\gamma, d)\frac{e_t^2}{\sigma^2}\right) M^{-1}M(\gamma', d') \\ - M(\gamma, d)M^{-1}E\left(X_{t-1}X'_{t-1}\frac{e_t^2}{\sigma^2}\right) M^{-1}M(\gamma', d'). \quad (19) \end{aligned}$$

Let T^H denote the distribution (16) under this alternative covariance kernel. There is no clear relationship between T and T^H, so it is not clear what is the bias if one is calculated instead of the other.

Hansen (1996) has provided an algorithm which allows the calculation of this asymptotic distribution. Let T_n^H denote the distribution (16) with covariance kernel (19), where population moments are replaced by sample moments. For example,

$$E\left(X_{t-1}X'_{t-1}I_{1t}(\gamma, d)I_{1t}(\gamma', d')\frac{e_t^2}{\sigma^2}\right)$$

is estimated by

$$\frac{1}{n}\sum_{t=1}^{n}\left(X_{t-1}X'_{t-1}I_{1t}(\gamma,d)I_{1t}(\gamma',d')\frac{\hat{e}_t^2}{\hat{\sigma}^2}\right).$$

Then an exact draw from T_n^H can be made by letting u denote a random $N(0, I_n)$ vector, setting $\eta = u \odot \hat{e}/\hat{\sigma}$ (element-by-element multiplication), $\hat{\eta} = \eta - X(X'X)^{-1}X'\eta$, and then setting T_n^H as in (18). Computationally this is not more complicated that the calculation of T_n. The advantages of this asymptotic approximation are that it is easy to implement, and is asymptotically robust to heteroskedasticity of unknown form.

Estimated densities of the distributions of T_n^H are plotted in short dashes in Figures 3 and 4 for the two time-series, labeled 'Asymptotic (Hetero.)'. In both cases, there is a striking distinction between the two asymptotic distributions, with that for the null of linear/heteroskedastic being dramatically shifted to the right relative to that for linear/homoskedasticity. Not surprisingly, the p-values are quite different as well. For the sunspot series, the p-value is 0.03, and for the industrial production series it is 0.047. What is evident from this calculation is that the allowance for heteroskedasticity dramatically moderates the evidence in favor of the SETAR(2) model.

We found in our analysis of the homoskedastic model that there was a large distinction between the asymptotic and bootstrap distributions, and there is good reason to expect this distinction to be even larger in the heteroskedastic case. It is therefore desirable to calculate a bootstrap distribution of F_{12} allowing for the possibility of general heteroskedasticity. The difficulty is that there is not a well-accepted bootstrap method which is appropriate in the present context. Block resampling schemes are inappropriate because they do not impose the null hypothesis. On the other hand, any model-based bootstrap will require a parametric model for the conditional variance, and the validity of the bootstrap method will depend upon the validity of the selected conditional variance functional. While this calls for careful selection of an empirically-determined conditional variance function, the presumption must be that the results will not be overly sensitive to misspecification of the conditional variance.

For our conditional variance function, we specify that $\sigma_{t-1}^2 = E(e_t^2 \mid \Im_{t-1})$ is a linear function in the squares of the regressors. Hence, let Z_{t-1} be the $k \times 1$ vector of the squared regressors (e.g., $Z_{t-1} = X_{t-1} \odot X_{t-1}$), so that $\sigma_{t-1}^2 = Z'_{t-1}\beta$ for some vector β. Then $e_t^2 = Z'_{t-1}\beta + \xi_t$ with $E(\xi_t \mid \Im_{t-1}) = 0$, so β can be estimated by OLS regression of $\hat{e}_t^2$ on Z_{t-1}, where $\hat{e}_t$ is the OLS residual from the SETAR(1) model. We calculate the fitted values $\hat{\sigma}_{t-1}^2 = Z'_{t-1}\hat{\beta}$ and the rescaled residuals $\hat{\varepsilon}_t = e_t/\hat{\sigma}_{t-1}$ (with the convention that $\hat{\varepsilon}_t = 0$ if $\hat{\sigma}_{t-1}^2 \leqslant 0$).

Our heteroskedastic bootstrap method assumes that the rescaled errors $\varepsilon_t = e_t/\sigma_{t-1}$ are independent over time, and works similarly to the homoskedastic bootstrap in Section 5.1 except how the errors e_t^* are generated. We fix the initial conditions $X_0^* = (y_0, y_{-1}, y_{-2}, \ldots, y_{-p+1})$ and now describe the recursion $X_{t-1}^* \to y_t^*$.

Let $\sigma_{t-1}^{*2} = \max(Z_{t-1}^{*\prime}\hat{\beta}, 0)$ where $Z_{t-1}^{*} = X_{t-1}^{*} \odot X_{t-1}^{*}$. Let ε_t^* be an independent draw from the empirical distribution of $\{\hat{\varepsilon}_t\}$, set $e_t^* = \sigma_{t-1}^* \varepsilon_t^*$, and $y_t^* = \hat{\alpha}_1' X_{t-1}^* + e_t^*$. This recursion creates simulated time-series y_t^* with the desired conditional mean and variance functions. On this sample, we calculate the test statistic F_{12}^*, and repeat a large number of times to find the bootstrap distribution.

Estimated densities of the bootstrap distributions are displayed using dotted lines in Figures 3 and 4, with the label 'Bootstrap (Hetero.)'. In both cases, the heteroskedastic bootstrap distributions are considerably different from both the asymptotic distributions and the homoskedastic bootstrap distribution. In both cases, the heteroskedastic bootstrap distribution is shifted more to the right than the homoskedastic bootstrap. It is interesting to observe, however, that for the sunspot series the heteroskedastic bootstrap is more shifted to the right than the asymptotic distribution allowing heteroskedasticity, while in the industrial production application these rankings are reversed.

The heteroskedastic bootstrap p-values are reported in Table 5. In both applications, the F_{12} test appears to be statistically significant.

In summary, examining the displays in Figures 3 and 4, we can make the following general recommendations. In the presence of conditional heteroskedasticity, distributions calculated under the assumption of homoskedasticity can be quite misleading. Since it is unknown whether or not there is meaningful conditional heteroskedasticity (tests are helpful but not decisive), this suggests that the preferred distributions are those which allow for conditional heteroskedasticity. There can be large discrepancies, however, between asymptotic and bootstrap approximations, suggesting that inference be made in practice using carefully selected bootstrap distributions which account for the error heteroskedasticity.

For example, in the sunspot example, there is very strong evidence for conditional heteroskedasticity. Thus the more appropriate p-value is the heteroskedastic bootstrap, which is 0.031. This is marginally significant, leading us to lean towards rejecting the SETAR(1) model in favor of the SETAR(2) model, but reserving some hesitations. In the industrial production example, there is no strong evidence for heteroskedasticity, so it is less clear whether we should prefer the homoskedastic bootstrap (p-value of 0.001) or the heteroskedastic bootstrap (p-value of 0.010). Since both are highly significant, we feel safe in concluding that the evidence allows us to reject the SETAR(1) model in favor of the SETAR(2) for this series.

6. Testing SETAR(1) against SETAR(3)

As discussed in Section 2, the natural test for SETAR(1) against SETAR(3) is to reject for large values of $F_{13} = n(S_1 - S_3)/S_3$. As discussed in the previous section, the statistic has a non-standard asymptotic distribution under the SETAR(1) hypothesis, so conventional critical values (such as the $\chi^2(2k)$) are not appropriate. An asymptotic approximation similar to (16) can be developed, and can be calculated using methods similar to those described in Section 4. As we argued in Section 4, however, the asymptotic distributions appear to be quite different from bootstrap distributions, and we expect the latter to provide better approximations. In addition, there is no obvious short-cut which enables faster computation of the asymptotic distribution relative to the bootstrap distribution.[6] Since there are no clear computational advantages, it appears advisable to simply focus on bootstrap distributions.

There are no additional complications in calculating the bootstrap distribution of F_{13} relative to calculating that of F_{12}. Both are calculated under the same null hypothesis (the SETAR(1) model) so the same technique is used. For either the homoskedastic (or heteroskedastic) bootstrap, simulated time-series are generated as described in Section 4.1 (or Section 4.2) and the F_{13} statistic is calculated on this simulated data. Through repeated replication (we use 2000), the bootstrap distribution is uncovered.

We display the bootstrap distributions (both homoskedastic and heteroskedastic) for the sunspot series in Figure 5. The conventional $\chi^2(2k)$ is plotted also for reference. The distributions, as expected, are noticeably different from the chi-square, and are also noticeably different from one another, with the heteroskedastic bootstrap distribution shifted out more to the right. The bootstrap p-values are presented in Table 6, and are both highly significant. We are able to easily reject the hypothesis of the SETAR(1) in favor of the SETAR(3).

In Figure 6 we display the bootstrap distributions for the industrial production data, and report the bootstrap p-values in Table 6. The evidence suggests the rejection of the SETAR(1) model, but the rejection is not as strong as the rejection from the previous section.

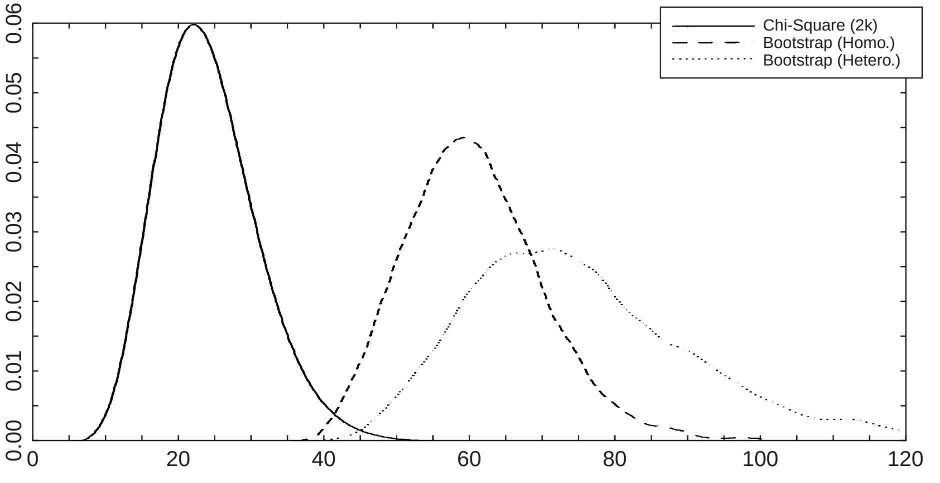

Figure 5. Sunspot series asymptotic and bootstrap distributions of F_{13}.

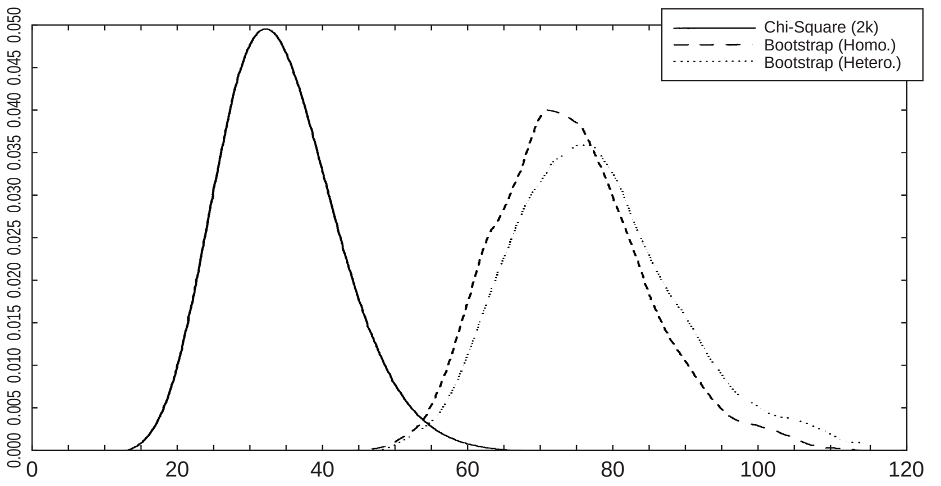

Figure 6. Industrial production asymptotic and bootstrap distributions of F_{13}.

Table 6. Bootstrap tests of SETAR(1) against SETAR(3).

		Bootstrap p-values	
	F_{13}	Homoskedastic	Heteroskedastic
Sunspots	132	0.000	0.004
Industrial production	101	0.013	0.046

7. Testing SETAR(2) against SETAR(3)

The F_{13} test does not allow the discrimination between the SETAR(2) and SETAR(3) models, and therefore is not a sufficient tool for model selection. We now consider the F_{23} test, which directly allows a comparison between these models. As in the previous section, we use bootstrap methods to evaluate the sampling distribution. We do this with some caution, because there has not yet been a demonstration that a bootstrap procedure can properly approximate the sampling distribution of F_{23} under the SETAR(2) null hypothesis. The problem is that under the null hypothesis, the model is a non-linear SETAR(2) model, and one of the parameter estimates, $\hat{\gamma}_1$, has a non-standard asymptotic distribution (see Chan (1993)).

Despite these concerns, there is no reason to expect the bootstrap to fail to achieve the correct first-order asymptotic distribution, so we proceed and describe bootstrap methods of inference. To calculate the bootstrap distribution of F_{23} under the SETAR(2) hypothesis, we need to generate simulated data from the SETAR(2) model. Given such simulated data, we can calculate the F_{23} statistic, and then repeat this procedure a large number of times to generate the bootstrap distribution.

The key feature is to generate simulated data from the SETAR(2). We use the SETAR(2) parameter estimates from Section 4.2. We first consider the homoskedastic bootstrap, which treats the errors e_t as independent draws. We implement this assumption by drawing the bootstrap errors from the empirical distribution of the SETAR(2) least-squares residuals. Then the simulated series is created according to the SETAR(2) model defined in (4).

We next consider forms of the heteroskedastic bootstrap. We consider two forms. The first assumes that the conditional heteroskedasticity is limited to a regime effect, namely, that $E(e_t^2 \mid \Im_{t-1}) = \sigma_1^2 I_{1t} + \sigma_2^2 I_{2t}$. (This is an assumption commonly made in SETAR applications.) We implement this bootstrap by first dividing the SETAR(2) residuals $\hat{e}_t$ into two groups: the n_1 errors $\hat{e}_{1t}$ for which $I_{1t} = 1$, and the n_2 errors $\hat{e}_{2t}$ for which $I_{2t} = 1$. Then when simulating the distribution of y_t^* given $\Im_{t-1}^*$, if $I_{1t}^* = 1$, we draw e_t^* randomly from $\{\hat{e}_{1t}\}$, and if $I_{2t}^* = 1$, we draw e_t^* randomly from $\{\hat{e}_{2t}\}$. We call this the Regime Heteroskedastic Bootstrap.

The second form of the heteroskedastic bootstrap we consider uses the functional form estimated for the conditional variance as reported in Table 3. This is a model of the conditional variance which has regime indicators, and is linearly a function of the squares of the regressors. Simulation from this process is similar to that described in Section 5.2. We call this procedure the General Heteroskedastic Bootstrap, since it allows for heteroskedasticity of general form.

We display in Figures 7 and 8 estimated densities of the bootstrap distributions for our two time-series applications. The $\chi^2(k)$ is also plotted for reference. We find that the bootstrap distributions for the sunspot series are quite sensitive the the specification of the error process, with an increasingly 'fat' distribution as the

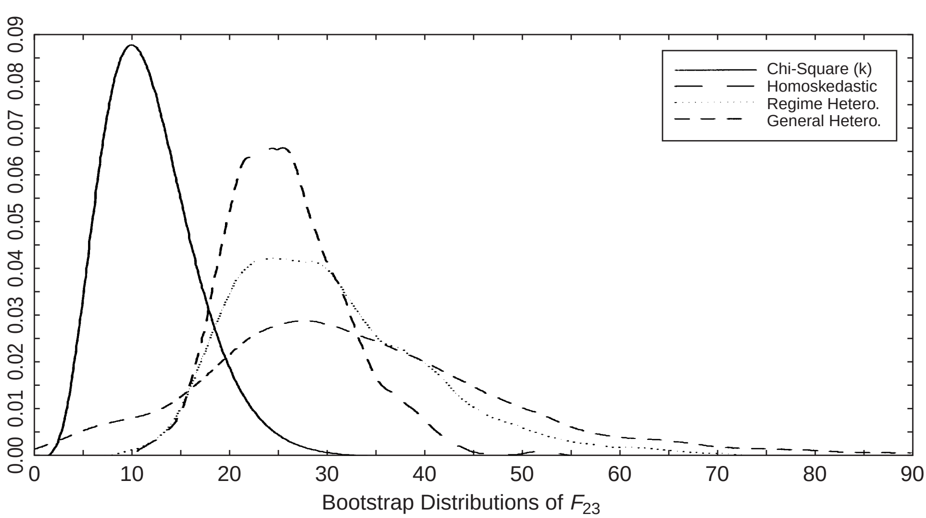

Figure 7. Sunspot series asymptotic and bootstrap distributions of F_{23}.

degree of heteroskedasticity is increased. In contrast, for the industrial production series, the bootstrap distributions are relatively insensitive to the heteroskedastic specification. This difference is likely due to our finding that the sunspot series exhibits a strong degree of heteroskedasticity, but not the industrial production series. The important message is that it is necessary to be careful about the modeling of the error process when pursuing bootstrap inference.

The bootstrap p-values for the F_{23} statistics are given in Table 7. Considering the sunspot series, the test would appear to be significant if the errors were (mistakenly) assumed to be homoskedastic, and marginally significant if the Regime Heteroskedastic Bootstrap were applied. The p-value rises to 12%, however, if we allow a general process for the error heteroskedasticity. Since this type of heteroskedasticity seems quite likely in this data, we conclude that we cannot reject the hypothesis of the SETAR(2) model against the SETAR(3) model. This suggests that an appropriate model for the sunspot series is the SETAR(2) model.

Table 7. Bootstrap tests of SETAR(2) against SETAR(3).

		Bootstrap p-values		
	F_{23}	Homoskedastic	Regime heteroskedastic	General heteroskedastic
Sunspots	50	0.001	0.044	0.126
Industrial production	27	0.828	0.808	0.856

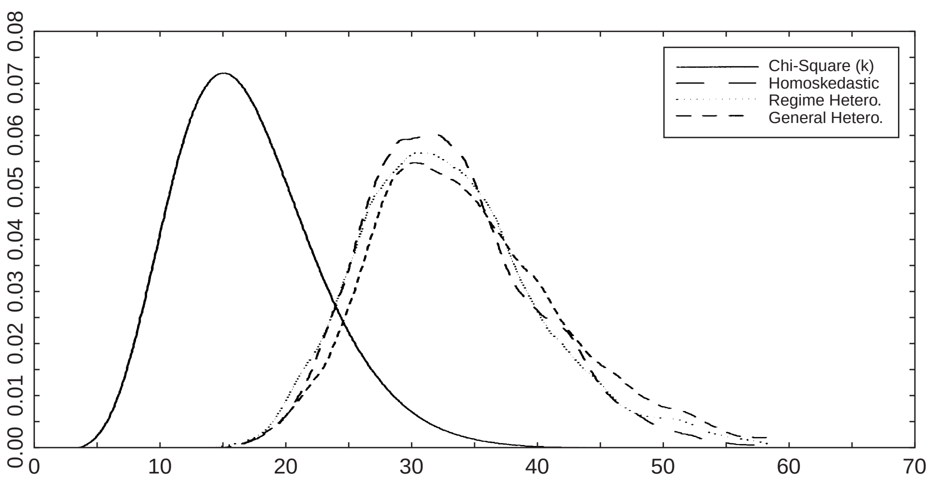

Figure 8. Industrial production asymptotic and bootstrap distributions of F_{13}.

The p-values for the industrial production series are similar across the error modeling choices, and are all far from significant, leading to the conclusion that we cannot reject the SETAR(2) model in favor of the SETAR(3) model. As for the sunspot series, we find that the evidence supports a SETAR(2) model for the industrial production series.

8. Conclusion

We have presented the theory of least-squares inference for the number of regimes in SETAR models. Least-squares estimation and test construction is conceptually and computationally straightforward. Evaluation of test significance is complicated, however, by the fact that the asymptotic distributions are non-standard and non-similar, precluding tabulation. While it is possible to calculate the asymptotic distribution in any application, it seems most prudent to report bootstrap p-values. Such bootstrap p-values can be sensitive to how the bootstrap data is generated, our suggestion is to pay careful attention to the specification of the conditional variance. A naive bootstrap which assumes independent errors can yield inaccurate inferences.

The procedures described in this paper are not very difficult to program, and the computation requirements appear quite reasonable for applications.

We illustrated these methods with two applications, comparing SETAR(1), SETAR(2), and SETAR(3) specifications. Our tests led to the conclusion that annual sunspots and monthly U.S. industrial production are SETAR(2) processes.

While we only explicitly examine tests between SETAR(1), SETAR(2), and SETAR(3) models, the methods extend to higher-order SETAR models as well. The main caution to consider, however, is that we expect the accuracy of the bootstrap approximations to deteriorate when higher-order SETAR models are tested, due to the more complicated forms of nonlinearity.

Acknowledgements

Prepared for a special issue of the *Journal of Economic Surveys* on nonlinear modeling. This research was supported by a grant from the National Science Foundation. I thank Nikolay Gospodinov for finding an error in a previous draft, and Simon Potter for helpful comments which improved the presentation.

Notes

1. Tong (1990) calls this a SETAR($m; p, ..., p$).
2. $I(a) = 1$ if a is true, else $I(a) = 0$.
3. $\Im_t$ equals the Borel sigma-field $\sigma(Y_t, Y_{t-1}, Y_{t-2} ...)$.
4. This algorithm applies as well to the Andrews–Ploberger (1994) exponentially weighted and averaged test statistics.
5. The density estimates were calculated using an Epanechnikov kernel with the Silverman (1986) rule-of-thumb bandwidth. See Hardle and Linton (1994) for a description of non-parametric density estimation.

6. The main computational savings from calculating the asymptotic distribution, rather than the bootstrap, is that the moments matrices $M_n^*(\gamma, d)^{-1}$ can be stored. While the list of such matrices is relatively small for estimation of the SETAR(2), it is quite large for the SETAR(3) model, making programming and memory requirements quite prohibitive.

References

Andrews, D. W. K. (1993) Tests for parameter instability and structural change with unknown change point. *Econometrica*, 61, 821–856.

Andrews, D. W. K. (1994) Empirical process methods in econometrics. *Handbook of Econometrics, Vol IV*, 2248–2296, R. F. Engle and D. L. McFadden, eds., Amsterdam: Elsevier Science.

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

Bai, J. (1997) Estimating multiple breaks one at a time. *Econometric Theory*, 13, 315–352.

Bai, J. and Perron, P. (1998) Estimating and testing linear models with multiple structural changes. *Econometrica*, 66, 47–78.

Brock, W. A. and Potter, S. (1993) *Nonlinear Time Series and Macroeconometrics*. In G. S. Maddala, C. R. Rao, H. D. Vinod (eds), *Handbook of Statistics*. Amsterdam: North-Holland.

Burgess, S. M. (1992) Nonlinear dynamics in a structural model of employment. *Journal of Applied Econometrics*, 7, S101–S118.

Caner, M. and Hansen, B. E. (1998) *Threshold autoregression with a near unit root*. Madison: Department of Economics, University of Wisconsin, Madison.

Cao, C. Q. and Tsay, R. S. (1992) Nonlinear time-series analysis of stock volatilities. *Journal of Applied Econometrics*, 7, S165–S186.

Chan, K. S. (1990) Testing for threshold autoregression. *The Annals of Statistics*, 18, 1886–1894.

Chan, K. S. (1991) Percentage points of likelihood ratio tests for threshold autoregresion. *Journal of the Royal Statistical Society, Series B*, 53, 691–696.

Chan, K. S. (1993) Consistency and limiting distribution of the least squares estimator of a threshold autoregressive model. *The Annals of Statistics*, 21, 520–533.

Chan, K. S. and Tong, H. (1986) On estimating thresholds in autoregressive models. *Journal of Time Series Analysis*, 7, 179–194.

Chan, K. S. and Tong, H. (1990) On likelihood ratio tests for threshold autoregression. *Journal of the Royal Statistical Society B*, 52, 469–476.

Davison, A. C. and Hinkley, D. V. (1997) *Bootstrap Methods and their Application*. Cambridge University Press.

Davies, R. B. (1977) Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika*, 64, 247–254.

Davies, R. B. (1987) Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika*, 74, 33–43.

Diebold, F. X. and Chen, C. (1996) Testing structural stability with endogenous break point: A size comparison of analytic and bootstrap procedures. *Journal of Econometrics*, 70, 221–241.

Galbraith, J. W. (1996) Credit rationing and threshold effects in the relation between money and output. *Journal of Applied Econometrics*, 11, 419–429.

Ghaddar, D. K. and Tong, H. (1981) Data transformation and self-exciting threshold autoregression. *Journal of the Royal Statistical Society*, C30, 238–248.

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

Hall, P. (1992) *The Bootstrap and Edgeworth Expansion*, New York: Springer-Verlag.

Hansen, B. E. (1992) Testing for parameter instability in regressions with I(1) processes. *Journal of Business and Economic Statistics*, 10, 321–55.

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

Hansen, B. E. (1997) Inference in TAR models. *Studies in Nonlinear Dynamics and Econometrics*, 1(2), 1–14.

Hansen, B. E. (1999a) Testing for structural change in conditional models. *Journal of Econometrics*, forthcoming.

Hansen, B. E. (1999b) Threshold effects in non-dynamic panels: Estimation, testing and inference. *Journal of Econometrics*, forthcoming.

Hardle, W. and O. Linton (1994) *Applied Nonparametric Methods*. In R. F. Engle and D. L. McFadden, (eds), *Handbook of Econometrics, Vol. IV*, pp. 2295–2339. Amsterdam: Elsevier Science.

Koop, G. and Potter, S. M. (1999) Bayes factors and nonlinearity: evidence from economic time series. *Journal of Econometrics*, 88, 251–282.

Kuo, B. S. (1998) Test for partial parameter instability in regressions with I(1) processes. *Journal of Econometrics*, 86, 337–368.

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

Montgomery, A. L., Zarnowitz, V., Tsay, R. S. and Tiao, G. C. (1998) Forecasting the U.S. unemployment rate. *Journal of the American Statistical Association*, 93, 478–493.

Potter, S. M. (1995) A nonlinear approach to U.S. GNP. *Journal of Applied Econometrics*, 2, 109–125.

Quandt, R. (1960) Tests of the hypothesis that a linear regression obeys two separate regimes. *Journal of the American Statistical Association*, 55, 324–330.

Rothman, P. (1991) Further evidence on the asymmetric behavior of unemployment rates over the business cycle. *Journal of Macroeconomics*, 13, 291–298.

Seo, B. (1998) Tests for structural change in cointegrated systems. *Econometric Theory*, 14, 222–259.

Shao, J. and Tu, D. (1995) *The Jackknife and Bootstrap*. NY: Springer.

Silverman, B. W. (1986) *Density Estimation for Statistics and Data Analysis*. London: Chapman and Hall.

Stinchcomb, M. B. and White, H. (1998) Consistent specification testing with nuisance parameters present only under the alternative. *Econometric Theory*, 14, 295–325.

Terasvirta, T. and Anderson, H. M. (1992) Characterizing nonlinearities in business cycles using smooth transition autoregressive models. *Journal of Applied Econometrics*, 7, S119–S139.

Terasvirta, T., Tjostheim, D. and Granger, C. W. J. (1994) Aspects of modelling nonlinear time series. *Handbook of Econometrics, Vol IV*, 2917–2957, R. F. Engle and D. L. McFadden, (eds), Amsterdam: Elsevier Science.

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

Tong, H. (1983) *Threshold Models in Non-linear Time Series Analysis. Lecture Notes in Statistics*, 21, Berlin: Springer.

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

Tong, H. and Lim, K. S. (1980) Threshold autoregression, limit cycles and cyclical data (with discussion). *Journal of the Royal Statistical Society*, B42, 245–292.

Tsay, R. S. (1989) Testing and modeling threshold autoregressive processes. *Journal of the American Statistical Association*, 84, 231–240.

White, H. (1980) A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, 48, 817–838.