Inference in TAR Models

Bruce E. Hansen*
Boston College†

May 1996
Revised: December 1996
Forthcoming in *Studies in Nonlinear Dynamics and Econometrics*, 1997

Abstract

A distribution theory is developed for least squares estimates of the threshold in threshold autoregressive (TAR) models. We find that if we let the threshold effect (the difference in slopes between the two regimes) get small as the sample size increases, then the asymptotic distribution of the threshold estimator is free of nuisance parameters (up to scale). Similarly, the likelihood ratio statistic for testing hypotheses concerning the unknown threshold is asymptotically free of nuisance parameters. These asymptotic distributions are non-standard, but are available in closed form so critical values are readily available. To illustrate this theory, we report an application to the U.S. unemployment rate. We find statistically significant threshold effects.

*This research was supported by a grant from the National Science Foundation and a Alfred P. Sloan Foundation Research Fellowship. I would like to thank a referee for helpful comments.

†Department of Economics, Boston College, Chestnut Hill MA 02167-3806. bruce.hansen@bc.edu. Homepage: http://www2.bc.edu/~hansenb/

1 Introduction

Threshold Autoregressive (TAR) models are quite popular in the nonlinear time series literature. This popularity is due to the fact that they are relatively simple to specify, estimate, and interpret, at least in comparison with many other nonlinear time series models. Despite this popularity, there is only a small literature studying the sampling properties of the estimators and test statistics associated with TAR models. Our goal in this paper is to propose a distribution theory for the estimate of the threshold which can be used to form asymptotic confidence intervals for the model parameters.

The idea of approximating a general nonlinear autoregressive structure by a threshold autoregression with a small number of regimes is probably due to Tong. See Tong (1983) for an early review of this approach and Tong (1990) for a more mature view. If the discontinuity of the threshold is replaced by a smooth transition function, the TAR model can be generalized to the smooth transition autoregressive (STAR) model. See, for example, Chan and Tong (1986), Granger and Teräsvirta (1993) and Teräsvirta, Tjostheim and Granger (1994). Two difficult statistical issues arise in connection with these models. First, conventional tests of the null of a linear autoregressive model against the TAR alternative have non-standard distributions, as the threshold parameter is not identified under the null of linearity. This problem was pointed out by Davies (1977, 1987); see also Andrews and Ploberger (1994) and Andrews (1994). To circumvent this problem, Luukkonen, Saikkonen and Teräsvirta (1988) proposed an LM test for a Taylor series approximation to the regression function under the STAR alternative. Chan (1990a) (see also Chan (1991) and Chan and Tong (1990)) found an empirical process representation for the asymptotic distribution of the likelihood ratio test. Hansen (1996a) showed that a bootstrap method replicates this asymptotic distribution.

The second difficult statistical issue associated with TAR models is the sampling distribution of the threshold estimate. Chan (1993) showed that the LS estimator is rate$-n$ consistent, and found an empirical process representation for the limiting distribution. Since the latter depends on a host of nuisance parameters it is not useful as a basis for forming confidence intervals for the unknown threshold. In contrast, our theory develops an alternative approximation to the sampling distribution of the threshold estimator based the empirical process results of Hansen (1996b) who studied general threshold models. Translated into the TAR context, our results show that if we let the threshold effect (the difference between the regression slopes in the two regimes) diminish as the sample size diverges, then we can approximate the sampling distribution of the threshold estimate by an asymptotic distribution which is free of nuisance parameters (other than scale). Similarly, we obtain the limiting distribution of the likelihood ratio statistic for tests on hypotheses concerning the threshold, which we find is completely free of nuisance parameters. The latter gives a computationally convenient way to construct confidence intervals for the threshold: simply plot the likelihood ratio as a function of the threshold, draw in the critical value associated with the desired confidence level, and mark off the values of the threshold whose likelihood ratio fall below the critical value.

This is the first statistical technique which allows confidence interval construction for threshold estimates in TAR models. The theory of Chan (1993) has been used only to justify the super-consistency of threshold estimates, and it is unclear if his theory could be used to construct confidence intervals.

Our theory is partially derived from an analogous theory for the sampling distribution of the estimate of changepoints. For the latter see Picard (1985), Yao (1987), Dümbgen (1991) and Bai (1996).

We are also interested in approximations to the sampling distributions of the other regression parameter estimates. Since sampling error in the estimated threshold is likely to affect the sampling distribution of the regression estimates in finite samples, we propose a simple procedure to form confidence intervals which appears to produce superior finite sample approximations than the conventional approach.

To make our recommendations concrete, we walk through a simple empirical exercise concerning the U.S. unemployment rate. We find strong evidence for a TAR model using average unemployment changes as the threshold variable, and estimate the threshold to be near zero, meaning that the autoregressive structure changes in expansions (declining unemployment) relative to contractions (increasing unemployment).

The remainder of the paper is organized as follows. The next section introduces the model and estimation methods. Tests of the null of no threshold effect are reviewed. Section 3 describes the main asymptotic theory for the threshold estimator. Section 4 is concerned with confidence interval construction. We introduce methods to form confidence intervals for the threshold parameter, the regression parameters, and discuss corrections in the presence of heteroskedasticity. Section 5 contains the unemployment rate application. The final section contains a brief conclusion, and the proof of the theorem is contained in an appendix.

A GAUSS program which replicates the empirical work reported in this paper is available on request from the author or can be downloaded from his WEB homepage.

2 Preliminaries

2.1 Model

The observed data is $(y_1, ..., y_n)$ with initial conditions $(y_0, y_{-1}, ..., y_{-p+1})$. A two-regime threshold autoregressive (TAR) model takes the form

$$y_t = (\alpha_0 + \alpha_1 y_{t-1} + \cdots + \alpha_p y_{t-p}) \, 1\left(q_{t-1} \leq \gamma\right) + (\beta_0 + \beta_1 y_{t-1} + \cdots + \beta_p y_{t-p}) \, 1\left(q_{t-1} > \gamma\right) + e_t \tag{1}$$

where $1(\cdot)$ denotes the indicator function, and $q_{t-1} = q(y_{t-1}, ..., y_{t-p})$ is a known function of the data. The autoregressive order is $p \geq 1$ and the threshold parameter is γ. The parameters α_j are the autoregressive slopes when $q_{t-1} \leq \gamma$, and β_j are the slopes when $q_{t-1} > \gamma$. The error e_t is assumed to be a martingale difference sequence with respect to the past history of y_t. In principle, we would like to allow e_t to be conditionally heteroskedastic, but for the formal theory we will assume that e_t is iid $(0, \sigma^2)$.

Two alternative representations of (**??**) will be useful in our exposition. Let

$$x_t = \left(\begin{array}{cccc} 1 & y_{t-1} & \cdots & y_{t-p} \end{array} \right)'$$

and

$$x_t(\gamma) = \left(\begin{array}{cc} x_t' 1\left(q_{t-1} \leq \gamma\right) & x_t' 1\left(q_{t-1} > \gamma\right) \end{array} \right)'$$

so that (**??**) can be written as either

$$y_t = x_t'\alpha 1\left(q_{t-1} \le \gamma\right) + x_t'\beta 1\left(q_{t-1} > \gamma\right) + e_t \tag{2}$$

or

$$y_t = x_t(\gamma)'\theta + e_t, \tag{3}$$

where $\theta = (\alpha' \ \beta')'$.

2.2 Estimation

The parameters of interest are θ and γ. Since equation (**??**) is a regression equation (albeit non-linear in parameters) an appropriate estimation method is least squares (LS). Under the auxiliary assumption that e_t is iid $N(0, \sigma^2)$, LS is equivalent to maximum likelihood estimation. Since the regression equation is non-linear and discontinuous, the easiest method to obtain the LS estimates is to use sequential conditional LS. For a given value of γ, the LS estimate of θ is

$$\hat{\theta}(\gamma) = \left(\sum_{t=1}^{n} x_t(\gamma)x_t(\gamma)'\right)^{-1}\left(\sum_{t=1}^{n} x_t(\gamma)y_t\right)$$

with residuals $\hat{e}_t(\gamma) = y_t - x_t(\gamma)'\hat{\theta}(\gamma)$ and residual variance

$$\hat{\sigma}_n^2(\gamma) = \frac{1}{n}\sum_{t=1}^{n} \hat{e}_t(\gamma)^2. \tag{4}$$

The LS estimate of γ is the value which minimizes (**??**):

$$\hat{\gamma} = \underset{\gamma \in \Gamma}{\operatorname{argmin}}\, \hat{\sigma}_n^2(\gamma), \tag{5}$$

where $\Gamma = [\underline{\gamma}, \overline{\gamma}]$.

The minimization problem (**??**) can be solved by direct search. Observe that the residual variance $\hat{\sigma}_n^2(\gamma)$ only takes on at most n distinct values as γ is varied, and these values correspond to $\hat{\sigma}_n^2(q_{t-1})$, $t = 1, ..., n$. Thus to find the LS estimates (**??**), we employ the following algorithm. Run OLS regressions of the form (**??**) setting $\gamma = q_{t-1}$ for each $q_{t-1} \in \Gamma$. (This amounts to slightly less than n regressions.) For each regression, calculate the residual variance $\hat{\sigma}_n^2(\gamma)$. Pick the value of γ which corresponds to the smallest variance. This can be expressed as

$$\hat{\gamma} = \underset{q_{t-1} \in \Gamma}{\operatorname{argmin}}\, \hat{\sigma}_n^2(q_{t-1}). \tag{6}$$

The LS estimates of θ is then found as $\hat{\theta} = \hat{\theta}(\hat{\gamma})$. Similarly the LS residuals are $\hat{e}_t = y_t - x_t(\hat{\gamma})'\hat{\theta}$ with sample variance $\hat{\sigma}_n^2 = \hat{\sigma}_n^2(\hat{\gamma})$.

2.3 Estimating the Delay Parameter

In the SETAR model, the threshold variable is $q_{t-1} = y_{t-d}$ for some integer $d \in [1, \overline{d}]$. The integer d is called the delay lag, and typically is unknown so must be estimated. The least squares principle allows d to be estimated along with the other parameters. The estimation problem (**??**) is augmented to include a search over d, so instead of n regressions the search method requires approximately $n\overline{d}$ regressions. Since the parameter space for d is discrete the LS estimate $\hat{d}$ is super-consistent and for the purpose of inference on the other parameters we can act as if d is known with certainty. This is the approach taken in the applications which follow.

2.4 Testing for Threshold Autoregression

An important question is whether the TAR model (**??**) is statistically significant relative to a linear AR(p). The relevant null hypothesis is $H_0 : \alpha = \beta$. As is well known, this testing problem is tainted by the difficulty that the threshold γ is not identified under H_0. We review in this section the testing methodology suggested by Hansen (1996a).

If the errors are iid, from the theory of Davies (1977, 1987) and Andrews-Ploberger (1994), a test with near-optimal power against alternatives distant from the null hypothesis is the standard F statistic

$$F_n = n\left(\frac{\tilde{\sigma}_n^2 - \hat{\sigma}_n^2}{\hat{\sigma}_n^2}\right),$$

where

$$\tilde{\sigma}_n^2 = \frac{1}{n}\sum_{t=1}^{n}\left(y_t - x_t'\tilde{\alpha}\right)^2,$$

and

$$\tilde{\alpha} = \left(\sum_{t=1}^{n} x_t x_t'\right)^{-1} \left(\sum_{t=1}^{n} x_t y_t\right)$$

is the OLS estimate of α under the assumption that $\alpha = \beta$. Since F_n is a monotonic function of $\hat{\sigma}_n^2$, it is easy to see that

$$F_n = \sup_{\gamma \in \Gamma} F_n(\gamma)$$

where

$$F_n(\gamma) = n\left(\frac{\tilde{\sigma}_n^2 - \hat{\sigma}_n^2(\gamma)}{\hat{\sigma}_n^2(\gamma)}\right)$$

is the pointwise F statistic against the alternative $H_1 : \alpha \neq \beta$ when γ is known.

Since γ is not identified, the asymptotic distribution of F_n is not chi-square. Hansen (1996a) shows that the asymptotic distribution may be approximated by the following bootstrap procedure. Let u_t^*, $t = 1, ..., n$ be iid $N(0,1)$ random draws and set $y_t^* = u_t^*$. Using the observations x_t, $t = 1, ..., n$, regress y_t^* on x_t to obtain the residual variance $\tilde{\sigma}_n^{*2}$, on $x_t(\gamma)$ to obtain the residual variance $\hat{\sigma}_n^{*2}(\gamma)$, and form $F_n^*(\gamma) = n\left(\tilde{\sigma}_n^{*2} - \hat{\sigma}_n^{*2}(\gamma)\right)/\hat{\sigma}_n^{*2}(\gamma)$ and $F_n^* = \sup_{\gamma\in\Gamma} F_n^*(\gamma)$. Hansen (1996a) shows that the distribution of F_n^* converges weakly in probability to the null distribution of F_n under local alternatives for β, so that repeated (bootstrap) draws from F_n^* may be used to approximate the asymptotic null distribution of F_n. The bootstrap approximation to the asymptotic p-value of the test is formed by counting the percentage of bootstrap samples for which F_n^* exceeds the observed F_n.

If e_t is conditionally heteroskedastic, it is necessary to replace the F-statistic $F_n(\gamma)$ with a heteroskedasticity-consistent Wald or Lagrange Multiplier test. For example, setting $R = \begin{pmatrix} I & -I \end{pmatrix}$, $M_n(\gamma) = \sum x_t(\gamma)x_t(\gamma)'$ and $V_n(\gamma) = \sum x_t(\gamma)x_t(\gamma)'\hat{e}_t^2$, then the pointwise Wald statistic is

$$W_n(\gamma) = \left(R\hat{\theta}(\gamma)\right)' \left[R\left(M_n(\gamma)^{-1}V_n(\gamma)M_n(\gamma)^{-1}\right)R'\right]^{-1} R\hat{\theta}(\gamma)$$

and the appropriate test of H_0 is

$$W_n = \sup_{\gamma\in\Gamma} W_n(\gamma).$$

To obtain critical values, bootstrap the data as before, but instead set $y_t^* = \hat{e}_t u_t^*$. Hansen (1996a) shows that this procedure produces the asymptotically correct null distribution for this class of models.

3 Asymptotic Distribution

We will explicitly derive our distribution theory for the self-exciting threshold autoregressive model (SETAR) which is the special case where $q_{t-1} = y_{t-d}$ for some integer $d \in [1, p]$. This is not essential to the main theory, but is helpful in focusing our derivations.

Assumption 1 *For some $\delta > 0$,*

1. *e_t is iid, $E(e_t) = 0$, $E(e_t^2) = \sigma^2 < \infty$, $E|e_t|^{2+\delta} < \infty$, and e_t has a density function $f(\cdot)$ which is continuous and positive everywhere on R;*

2. $\sum_{j=1}^{p} |\alpha_j| < 1$, $\sum_{j=1}^{p} |\beta_j| < 1$;

3. *One of the following inequalities holds: Either $(\alpha_0 - \beta_0) + (\alpha_d - \beta_d)\gamma \neq 0$ or $\alpha_j \neq \beta_j$ for some $j \neq 0, d$.*

Assumption 1.1 is standard. Assumption 1.2 is sufficient to ensure that y_t is geometrically ergodic, which is necessary for our theory. Assumption 1.3 rules out a degenerate case. Let

$$D = E\left(x_t x_t' \mid q_{t-1} = \gamma_0\right), \tag{7}$$

$$\lambda_n = n(\alpha - \beta)' D(\alpha - \beta) f(\gamma_0)$$

and

$$LR_n(\gamma) = n\left(\frac{\hat{\sigma}_n^2(\gamma) - \hat{\sigma}_n^2(\hat{\gamma})}{\hat{\sigma}_n^2(\hat{\gamma})}\right).$$

Note that $LR_n(\gamma_0)$ is the likelihood ratio (or F) statistic to test the hypothesis $H_0 : \gamma = \gamma_0$. The following result is proved in the appendix.

Theorem 1 *If $\lambda_n \to \infty$ yet $\lambda_n / n \to 0$ as $n \to \infty$, then*

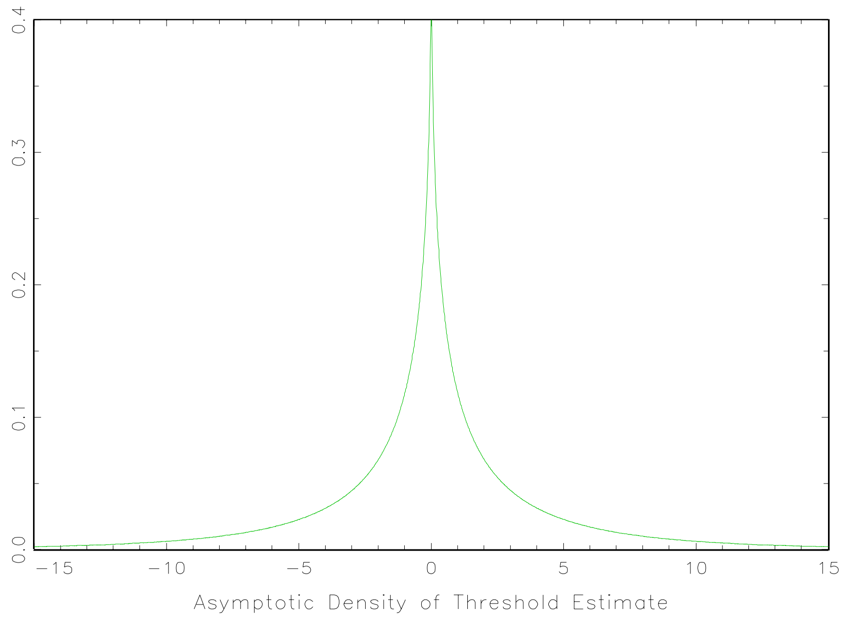
1. $\lambda_n(\hat{\gamma} - \gamma_0) \to_d \sigma^2 T$,

2. $LR_n(\gamma_0) \to_d \xi$,

where

$$T = \underset{s \in R}{\operatorname{argmax}} \left[W(s) - \frac{1}{2}|s|\right],$$

Figure 1: Asymptotic Density of Threshold Estimator



$$\xi = \max_{s \in R} \left[2W(s) - |s| \right],$$

and

$$W(\nu) = \begin{cases} W_1(-\nu), & \nu < 0 \\ 0 & \nu = 0 \\ W_2(\nu) & \nu > 0 \end{cases},$$

and $W_1(\nu)$ and $W_2(\nu)$ are two independent standard Brownian motions on $[0, \infty)$.

The distribution functions for T and ξ are available in closed form. First, for $x \geq 0$

$$P(T \leq x) = 1 + \sqrt{\frac{x}{2\pi}} \exp\left(-\frac{x}{8}\right) + \frac{3}{2} \exp(x) \Phi\left(-\frac{3\sqrt{x}}{2}\right) - \left(\frac{x+5}{2}\right) \Phi\left(-\frac{\sqrt{x}}{2}\right),$$

while for $x < 0$, $P(T \leq x) = 1 - P(T \leq -x)$. The density function of this distribution is plotted in Figure 1. Second,

$$P(\xi \leq x) = \left(1 - e^{-x/2}\right)^2.$$

Selected values of $P(|T| \leq x)$ and $P(\xi \leq x)$ can be found in Table 1.

Table 1: Asymptotic Critical Values

	.80	.85	.90	.925	.95	.975	.99
$P(\lvert T\rvert \leq x)$	4.70	5.89	7.69	9.04	11.04	14.66	19.77
$P(\xi \leq x)$	4.50	5.10	5.94	6.53	7.35	8.75	10.59

4 Confidence Intervals and Testing

4.1 Threshold Parameter

To construct asymptotically valid confidence intervals for γ, Hansen (1996b) recommends inverting the likelihood ratio statistic $LR_n(\gamma)$. Let $c_\xi(\beta)$ be the β-level critical value for ξ from the second row of Table 1. Set

$$\hat{\Gamma} = \{\gamma : LR_n(\gamma) \leq c_\xi(\beta)\}. \tag{8}$$

Theorem 1, part 2, shows that $P(\gamma_0 \in \hat{\Gamma}) \to \beta$, so $\hat{\Gamma}$ is an asymptotically valid β-level confidence set for γ. A graphical method to find $\hat{\Gamma}$ is to plot the likelihood ratio $LR_n(\gamma)$ against γ and draw a flat line at $c_\xi(\beta)$. (Note that the likelihood ratio is identically zero at $\gamma = \hat{\gamma}$.) Equivalently, one may plot the residual variance $\hat{\sigma}_n^2(\gamma)$ against γ, and draw a flat line at $\hat{\sigma}_n^2 \left(1 + c_\xi(\beta)/n\right)$.

The fact that the region $\hat{\Gamma}$ may be disjoint may be unsatisfactory in practice. A more conservative procedure is to define the convexified region $\hat{\Gamma}^c = [\hat{\gamma}_1, \hat{\gamma}_2]$ where $\hat{\gamma}_1 = \min_\gamma \hat{\Gamma}$ and $\hat{\gamma}_2 = \max_\gamma \hat{\Gamma}$.

To investigate the accuracy of our asymptotic approximations in finite samples, we report a simple Monte Carlo experiment. The model is a SETAR of the form (**??**) with $p = 1$, $q_{t-1} = y_{t-1}$ and e_i iid $N(0, 1)$. We fixed $\alpha_0 = 0$, $\beta_1 = 0$, $\gamma = 0$, and varied α_1 among $-.3$, $.3$, 0 and .6 (to assess sensitivity to serial correlation), β_0 from .1 to .6 (to assess sensitivity to the strength of the threshold effect), and n from 50 to 1000 (to assess sensitivity to sample size). The results were similar for the four values of α_1, so we report only the results for $\alpha_1 = .6$. 1000 replications were made for each parameterization. We report in Table 2 the rejection frequencies of a nominal 10% size test of $H_0 : \gamma = 0$. The first six columns report rejection rates using the likelihood ratio region $\hat{\Gamma}$. The final six columns report rejection rates

Table 2: Confidence Interval Converage for γ at 10% Level

	$\hat{\Gamma}$						$\hat{\Gamma}^c$					
$\beta_0 =$	.1	.2	.3	.4	.5	.6	.1	.2	.3	.4	.5	.6
$n = 50$	.15	.14	.13	.17	.14	.15	.08	.09	.07	.11	.10	.11
$n = 100$	.22	.20	.21	.19	.15	.16	.09	.08	.08	.08	.07	.09
$n = 250$	.29	.24	.21	.20	.17	.13	.08	.07	.09	.08	.08	.07
$n = 500$	.35	.31	.20	.16	.12	.11	.08	.09	.08	.07	.07	.07
$n = 1000$	.38	.24	.24	.11	.09	.08	.10	.09	.08	.06	.06	.05

using the convexified region $\hat{\Gamma}^c$.

The rejection rates for the likelihood ratio test are generally liberal, implying that the confidence region $\hat{\Gamma}$ will have true coverage rates which are less than the nominal levels. The rejection rates appear to decrease as the threshold effect β_0 increases (except at the smallest sample size), but the size distortion does not uniformly diminish as the sample size increases, indeed it is increasing in n for the smallest value of β_0. This does not contradict our asymptotic distribution theory, for the latter is based on a delicate argument that the threshold effect β_0 decreases as n get large. To see this in Table 2 for $\hat{\Gamma}$, note that for $n \geq 250$ the rejection rate appears to be decreasing monotonically as β_0 increases. Thus there will be a unique $\beta_0(n)$ which yield (exactly) the correct size.

A better approximation appears to be achieved by the convexified region $\hat{\Gamma}^c$. The rejection rates are generally close to the nominal, and only somewhat conservative when both β_0 and n are large. These results suggest that $\hat{\Gamma}^c$ may be successfully used to construct confidence intervals for the threshold parameter γ.

4.2 Slope Parameters

Standard asymptotic theory shows that if γ_0 is known, then

$$\sqrt{n}\left(\hat{\theta}(\gamma_0) - \theta_0\right) \rightarrow_d N(0, \Psi(\gamma_0)) \tag{9}$$

where

$$\Psi(\gamma) = \left(E\left(x_i(\gamma)x_i(\gamma)'\right)\right)^{-1}\sigma^2.$$

Let z_β denote the β-level critical value for the normal distribution and $\hat{s}(\gamma) = \sqrt{\hat{\Psi}(\gamma)/n}$ denote a standard error for $\hat{\theta}(\gamma)$. Let

$$\hat{\Theta}(\gamma) = \hat{\theta}(\gamma) \pm z_\beta \hat{s}(\gamma) \tag{10}$$

be the $\beta-$level confidence interval for θ, conditional on γ fixed. When γ_0 is known, the region $\hat{\Theta}(\gamma_0)$ is the natural $\beta-$level confidence region for θ.

Since $\hat{\gamma}$ is consistent for γ_0 at a fast rate, it is possible to show that the first-order asymptotic approximation to the distribution of $\hat{\theta}$ (when γ is estimated) is identical to that given in (**??**). Thus we can act as if $\hat{\gamma} = \gamma_0$ and use $\hat{\Theta}(\hat{\gamma})$ as an asymptotically valid confidence interval for θ. One might be skeptical that this approach will yield good finite sample approximations in practice. In a small samples, γ might not be estimated very precisely, and this sampling error will contaminate the distribution of $\hat{\theta}$. It appears desirable to use a sampling approach which takes this uncertainty into account, and one such suggestion is made in Hansen (1996b). For some $\phi < 1$, construct an ϕ-level confidence interval for γ (as discussed in the previous section) and for each γ in this interval calculate a confidence interval for θ, and take the union of all these sets. Formally, let $\hat{\Gamma}(\phi)$ denote a confidence interval for γ with asymptotic coverage ϕ. For each $\gamma \in \hat{\Gamma}(\phi)$, construct the pointwise confidence region $\hat{\Theta}(\gamma)$ as in (**??**) and set

$$\hat{\Theta}_\phi = \bigcup_{\gamma \in \hat{\Gamma}(\phi)} \hat{\Theta}(\gamma).$$

By construction, $\hat{\Theta}_\phi$ increases with ϕ in the sense that $\hat{\Theta}_{\phi_1} \subset \hat{\Theta}_{\phi_2}$ if $\phi_1 < \phi_2$. Note that the smallest member of this class is $\hat{\Theta}_0 = \hat{\Theta}(\hat{\gamma})$, the confidence interval formed by ignoring the sampling variation in $\hat{\gamma}$, so $\hat{\Theta}_\phi$ is by construction more conservative (larger) than $\hat{\Theta}(\hat{\gamma})$ if $\phi > 0$.

To assess the accuracy of these confidence regions, we report a simple Monte Carlo experiment using the same simulation design as in the previous section. We constructed 95% confidence regions for β_0 using the conventional region $\hat{\Theta}(\hat{\gamma}) = \hat{\Theta}_0$, and using the conservative regions $\hat{\Theta}_\phi$ for $\phi = .5$, .8, and .95. For the latter, we used the likelihood ratio region[1] $\hat{\Gamma}$ from (**??**) for γ. Table 4 reports the frequencies that the true value of β_0 fell outside

[1] Alternatively, the region $\hat{\Gamma}^c$ could be used.

Table 3: Confidence Interval Converage for β_0 at 5% Level

	$\hat{\Theta}_0$			$\hat{\Theta}_{.5}$			$\hat{\Theta}_{.8}$			$\hat{\Theta}_{.95}$		
$\beta_0 =$	.2	.4	.6	.2	.4	.6	.2	.4	.6	.2	.4	.6
$n = 50$	.41	.39	.34	.20	.18	.17	.08	.07	.08	.02	.03	.03
$n = 100$	.52	.43	.34	.23	.18	.14	.08	.07	.06	.03	.02	.02
$n = 250$	.55	.37	.22	.22	.13	.10	.08	.05	.05	.02	.01	.01
$n = 500$	.50	.25	.09	.19	.11	.04	.08	.05	.02	.02	.02	.01
$n = 1000$	.40	.11	.04	.13	.05	.03	.06	.02	.02	.02	.02	.01

of these confidence regions. To simplify the table we only report the results for $\beta_0 = .2$, .4, .6 and $\alpha_1 = .6$.

The coverage probabilities for the conventional region $\hat{\Theta}_0$ are quite poor, except when the sample is very large and the threshold effect is quite strong. The conservative regions do much better, with the region $\hat{\Theta}_{.8}$ appearing to strike a reasonable balance between under- and over-rejection. It produces a confidence region which is slightly too liberal when the threshold effect is very small or the sample size is small, and somewhat too conservative when the threshold effect and the sample size are large. Thus our recommendation is to use the region $\hat{\Theta}_{.8}$ to construct confidence regions for the regression slope parameters.

4.3 Heteroskedastic Errors

If the error e_t is not iid but a heteroskedastic martingale difference, Assumption 1 does not hold. Hansen (1996b) shows that if the data y_t satisfy the technical requirement of absolute regularity (β-mixing), then the basic results go through. Can we make this extension for TAR processes? The difficulty is verifying the technical requirement of absolute regularity. It appears nearly impossible to verify such requirements under heteroskedasticity so we cannot formally state a theorem. Yet it seems likely that this requirement is only an artifact of the proof technique, so we present the results for heteroskedastic processes anyway.

The key assumption needed to extend the theory is that while e_t can be conditionally heteroskedastic, the conditional heteroskedasticity cannot be regime-dependent. Specifically, the conditional expectation $E\left(e_t^2 \mid q_{t-1} = \gamma\right)$ must be continuous at γ_0. If this condition is violated (for example, if $E\left(e_t^2 \mid q_{t-1} \leq \gamma\right) = \sigma_1^2$ and $E\left(e_t^2 \mid q_{t-1} > \gamma\right) = \sigma_2^2$ with $\sigma_1^2 \neq \sigma_2^2$) then different methods will be necessary than those outlined below.

With heteroskedastic errors, the asymptotic distributions depend on the new nuisance parameter

$$\eta^2 = \frac{(\alpha-\beta)' V (\alpha-\beta)}{(\alpha-\beta)' D (\alpha-\beta)},$$

where D is defined in (**??**) and

$$V = E\left(x_t x_t' e_t^2 \mid q_{t-1} = \gamma_0\right).$$

Note that in the homoskedastic case $E\left(e_t^2 \mid q_{t-1}\right) = \sigma^2$, then $V = D\sigma^2$ and hence $\eta^2 = \sigma^2$. We find that Theorem 1 is modified as follows. Result 1 is replaced by

$$\lambda_n(\hat{\gamma} - \gamma_0) \to_d \eta^2 T$$

and result 2 is replaced by

$$LR_n(\gamma_0) \to_d \frac{\eta^2}{\sigma^2}\xi.$$

Since the second result is used to construct confidence intervals for γ (and hence θ), we can modify the approach as follows. Given an estimate $\hat{\eta}$ of η (to be discussed shortly), define the modified likelihood ratio sequence

$$\begin{aligned} LR_n^*(\gamma) &= \frac{\hat{\sigma}_n^2}{\hat{\eta}^2} LR_n(\gamma) \\ &= n\left(\frac{\hat{\sigma}_n^2(\gamma) - \hat{\sigma}_n^2}{\hat{\eta}^2}\right) \end{aligned}$$

and the modified likelihood ratio confidence region

$$\hat{\Gamma}^* = \left\{\gamma : LR_n^*(\gamma) \leq c_\xi(\beta)\right\}.$$

The region $\hat{\Gamma}^*$ is an asymptotically valid β-level confidence region for γ.

To construct confidence regions for the slope parameters θ we proceed as before. Rather than using $\hat{\Gamma}(\phi)$ to construct a preliminary ϕ-level confidence interval for γ, we use $\hat{\Gamma}^*(\phi)$. To construct the pointwise confidence regions $\hat{\Theta}(\gamma)$ for θ, it is also necessary to use a heteroskedasticity-consistent covariance matrix as in White (1980). Otherwise, the procedures are the same.

It remains to discuss the estimation of the nuisance parameter η. Let

$$r_{1t} = \left((\alpha - \beta)' x_t\right)^2,$$

$$r_{2t} = \left((\alpha - \beta)' x_t\right)^2 e_t^2,$$

$$g_1(\gamma) = E\left(r_{1i} \mid q_{t-1} = \gamma\right),$$

and

$$g_2(\gamma) = E\left(r_{2i} \mid q_{t-1} = \gamma\right).$$

Then

$$\eta^2 = \frac{g_2(\gamma_0)}{g_1(\gamma_0)},$$

and we see that this nuisance parameter equals the ratio of two conditional expectations, evaluated at the single point γ_0. Since these depend on unknown parameters, we can use $\hat{r}_{1t} = \left(\left(\hat{\alpha} - \hat{\beta}\right)' x_t\right)^2$, $\hat{r}_{2t} = \left(\left(\hat{\alpha} - \hat{\beta}\right)' x_t\right)^2 \hat{e}_t^2$ and $\hat{\gamma}$ in place of the true values.

To estimate the functions g_1 and g_2, either polynomial or kernel regression is appropriate. A polynomial regression fits by OLS an equation such as

$$\hat{r}_{1t} = \hat{\mu}_0 + \hat{\mu}_1 q_{t-1} + \hat{\mu}_2 q_{t-1}^2 + \hat{\varepsilon}_i$$

from which we set $\hat{g}_1(\hat{\gamma}) = \hat{\mu}_0 + \hat{\mu}_1\hat{\gamma} + \hat{\mu}_2\hat{\gamma}^2$. Similarly $\hat{g}_2(\hat{\gamma})$ is found by a regression of $\hat{r}_{2t}$ on q_{t-1} and q_{t-1}^2. Then the estimate of η^2 is

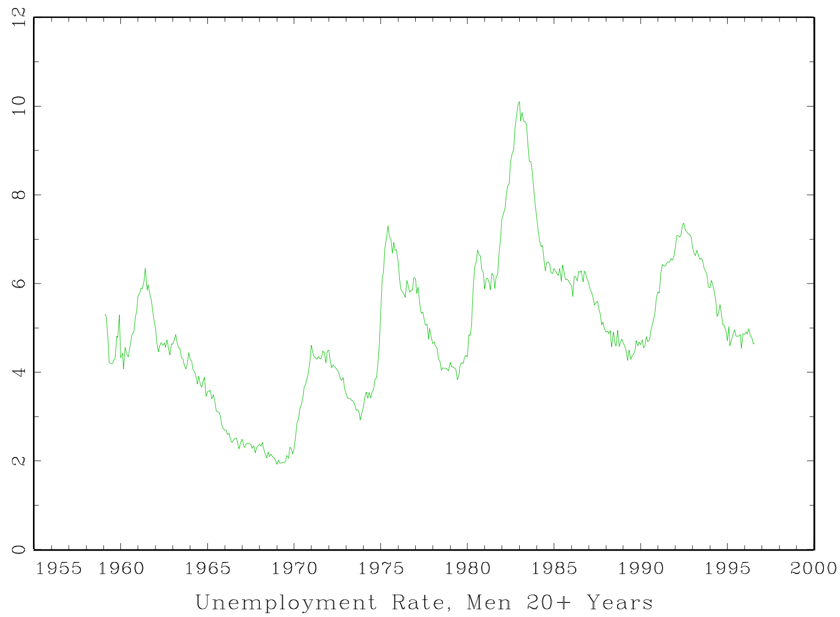
$$\hat{\eta}^2 = \frac{\hat{g}_2(\hat{\gamma})}{\hat{g}_1(\hat{\gamma})}.$$

The kernel estimate of η^2 is

$$\hat{\eta}^2 = \frac{\sum_{t=1}^{n} K\left(\frac{\hat{\gamma} - q_{t-1}}{h}\right) \hat{r}_{2t}}{\sum_{t=1}^{n} K\left(\frac{\hat{\gamma} - q_{t-1}}{h}\right) \hat{r}_{1t}}$$

where $K(x)$ is a kernel function such as the Epanechikov: $K(x) = (3/4)(1 - x^2)1\left(|x| \leq 1\right)$ and h is a bandwidth.

Figure 2: Unemployment Rate: Men 20+ Years



5 U.S. Unemployment Rate

In this section, we explore the presence of non-linearities in the business cycle through the use of a threshold autoregressive model for U.S. unemployment. We measure unemployment among males age 20 and over, using the ratio of the Citibase files LHMU and LHMC. The sample is monthly from 1959.1 through 1996.7, and is plotted in Figure 2. Standard unit root tests, such as the augmented Dickey-Fuller, suggest that the unemployment rate may have an autoregressive unit root, so we work with the first-differenced series Δy_t, to ensure stationarity. We set $p = 12$ as this appears to the minimum necessary to adequately describe the short-run dynamics.

We consider two choices for the threshold variable q_{t-1}. The first is a standard delay lag Δy_{t-d} for some $d \leq 12$. The second is a long difference

$$y^*_{t-d} = y_{t-1} - y_{t-d} \tag{11}$$

for some $d \leq 12$, which measures the recent trend in the unemployment rate. Table 4 reports the model sum of squared errors (SSE) from the various models, and the bootstrap-

Table 4: TAR Models for the Unemployment Rate

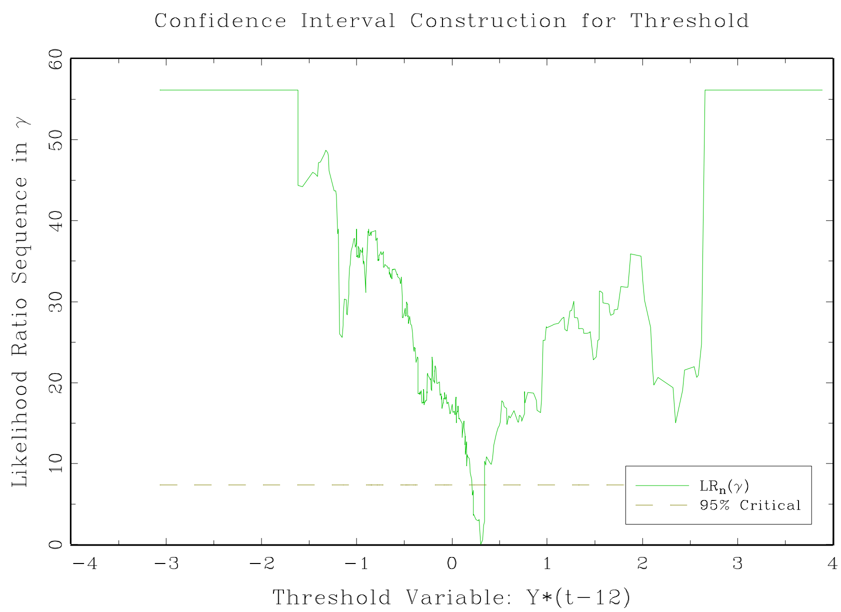
$q_t = \Delta y_{t-d}$												
$d =$	1	2	3	4	5	6	7	8	9	10	11	12
SSE	12.1	12.4	12.2	12.6	12.4	12.4	12.3	12.4	12.1	12.4	12.4	12.5
P-Value	.053	.13	.203	.294	.269	.128	.398	.149	.002	.041	.377	.866
$q_t = y_{t-1} - y_{t-d}$												
$d =$		2	3	4	5	6	7	8	9	10	11	12
SSE		11.8	12.0	11.9	11.8	11.9	11.9	11.9	11.9	11.8	12.0	11.7
P-Value		.020	.010	.141	.004	.000	.042	.007	.001	.000	.000	.000

calculated asymptotic p-value (using 1000 replications) for the test of the null of linearity against the particular threshold model. For the latter test we use a Wald statistic robust to heteroskedasticity as suggested by White (1980). (There is evidence of residual heteroskedasticity in all of the models we estimated.) For these and the other calculations, Γ was selected a priori to contain 70% of the observations, trimming the bottom and top 15% quantiles of the threshold variable to ensure that the model is well identified for all thresholds in Γ. See Andrews (1993) and Hansen (1996a) for discussion of this point.

The least-squares principle suggests selecting $\hat{d}$ through the minimization of the sum of squared errors. It is clear from Table 4 that the model using the long difference $y_{t-1} - y_{t-d}$ for the threshold fits better than using a simple lag value Δy_{t-d}. The smallest squared error is found by setting $\hat{d} = 12$. This model is highly statistically significant. Among our 1000 bootstrap replications, there was no simulated test statistic which exceeded the sample value, suggesting that the TAR model with threshold variable $q_{t-1} = y_{t-1} - y_{t-12}$ is significant at literally any significance level. The latter result is robust to the choice of d, as setting $q_{t-1} = y_{t-1} - y_{t-d}$ for any $d \geq 5$ yields a p-value less than 1%.

Setting $\hat{d} = 12$, the LS estimate of the threshold is $\hat{\gamma} = 0.302$, with a 95% asymptotic confidence interval [0.213, 0.340]. The latter was calculated using the convexified likelihood ratio approach, adjusting the likelihood ratio for residual heteroskedasticity using a kernel estimator for the nuisance parameters with a bandwidth selected by the plug-in method to minimize asymptotic mean-squared error. A plot of the adjusted likelihood ratio $LR_n^*(\gamma)$ is

Figure 3: Confidence Interval Construction for Threshold

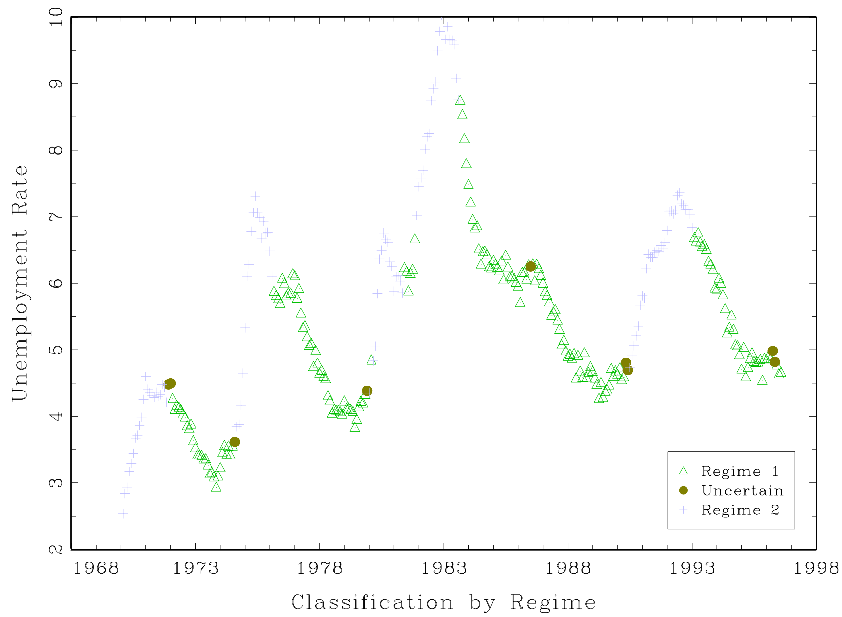


displayed in Figure 3. The values of γ where the likelihood ratio lies beneath the dotted line yield the confidence region. We can read from this graph that the threshold estimate is quite precise, and the confidence interval is quite tight.

The estimate $\hat{\gamma} = .3$ means that the TAR model splits the regression function into two regimes depending on whether the unemployment rate has been rising over the past 12 months more than 0.3% (i.e., a change in the unemployment rate from 5.6 to 5.9). Of the 438 observations in the fitted sample, 314 observations lie in "regime 1" where $y_{t-1} - y_{t-12} < .3$, and 124 lie in "regime 2" where $y_{t-1} - y_{t-12} > .3$. Heuristically, we can of regime 2 as corresponding to economic contractions.

From these point estimates, we can look back at the historical sample to examine how the TAR model splits the observations into regimes. In Figure 4 we plot the unemployment rate over the period 1960-1996, coded whether the observation falls in the estimated regime 1 (crosses) or regime 2 (triangles). To assess the precision of the estimate of γ, we code observations for which $y_{t-1} - y_{t-12}$ falls in the 95% confidence interval [0.213, 0.340] as "uncertain" (solid circles). From the plot, we see how upswings in the unemployment rate

Figure 4: Classification By Regime



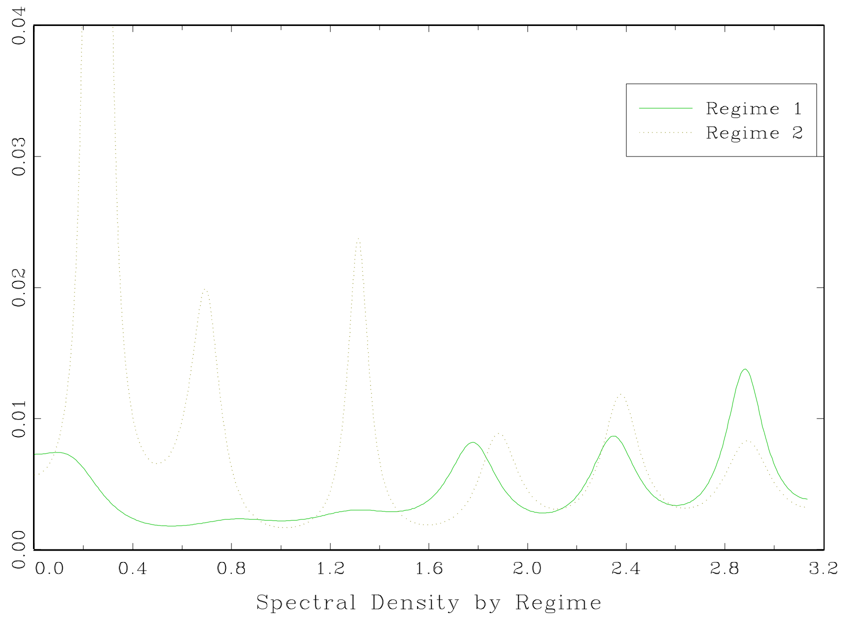
are categorized into regime 2, and downswings into regime 1. What seems quite surprising is how few observations fall in the uncertain category. Interestingly, two of the these uncertain observations appeared this year (in March and April, 1996)

Table 5: TAR Estimates for Unemployment Rate

				$y_{t-1} - y_{t-12} \leq 0.302$			
Variable	Intercept	y_{t-1}	y_{t-2}	y_{t-3}	y_{t-4}	y_{t-5}	y_{t-6}
$\hat{\alpha}$	−.018	-.186	.084	.132	.165	.070	.267
s.e.	(.012)	(.062)	(.065)	(.069)	(.056)	(.065)	(.065)
95% Conf.	[−.043, .010]	[−.309, −.035]	[−.048, .214]	[−.008, .275]	[.047, .290]	[−.065, .204]	[−.107, .162]
Variable		y_{t-7}	y_{t-8}	y_{t-9}	y_{t-10}	y_{t-11}	y_{t-12}
$\hat{\alpha}$		.062	.044	-.031	-.057	.091	-.136
s.e.		(.062)	(.055)	(.059)	(.060)	(.059)	(.058)
95% Conf.		[−.075, .194]	[−.063, .169]	[−.159, .093]	[−.177, .077]	[−.031, .208]	[−.254, −.015]
				$y_{t-1} - y_{t-12} > 0.302$			
Variable	Intercept	y_{t-1}	y_{t-2}	y_{t-3}	y_{t-4}	y_{t-5}	y_{t-6}
$\hat{\beta}$	.086	.241	.241	.123	-.026	-.020	-.084
s.e.	(.032)	(.101)	(.080)	(.090)	(.085)	(.085)	(.084)
95% Conf.	[.013, .151]	[.006, .441]	[.085, .414]	[−.053, .318]	[−.197, .158]	[−.199, .160]	[−.272, .090]
Variable		y_{t-7}	y_{t-8}	y_{t-9}	y_{t-10}	y_{t-11}	y_{t-12}
$\hat{\beta}$		-.151	-.035	.092	.103	-.114	-.412
s.e.		(.071)	(.78)	(.089)	(.085)	(.078)	(.085)
95% Conf.		[−.361, .004]	[−.202, .136]	[−.087, .276]	[−.064, .314]	[−.267, .056]	[−.608, −.217]

Table 5 reports the parameter estimates for the TAR model. We report parameter es-

Figure 5: Spectral Density by Regime



timates, heteroskedasticity-consistent standard errors, and the conservative 95% confidence regions calculated from an 80% first-step confidence region for γ. The most noticeable parameter shifts between the two regimes occurs in the constant and the autoregressive coefficients at lags 1, 2, and 12. In regime 1 (constant or decreasing unemployment), the AR(1) coefficient is slightly negative, the AR(2) coefficient is near zero, and the intercept is near zero. The implication is that the unemployment rate will be close to a random walk with slight negative serial correlation and a slight negative drift. On the other hand, in regime 2 (rising unemployment), the intercept and the AR(1) and AR(2) coefficients are all positive, implying that unemployment rate changes will be serially correlated with a positive drift.

It is difficult to assess the dynamics implicit in point estimates from an autoregression. One method is to plot the corresponding spectral density function. In Figure 5 we plot the spectral density functions corresponding to the autoregressive coefficients from the two regimes as reported in Table 5. These are not actually "spectral densities", but are intended to convey information about the dynamic properties in the two regimes. We find that in regime 1, Δy_t has a nearly flat spectral shape, while in regime 2, there is a large peak corresponding to the business cycle. Interestingly, both regimes display nearly identical higher-frequency spectral shape and power. This suggests that the differences between the two regimes pertains to the low frequencies, and a useful subject for future research is how to incorporate this restriction into estimation and testing procedures.

6 Conclusion

This paper has developed new methods of inference for threshold autoregressive models. We have shown how to test for threshold effects, estimate the threshold parameter, construct asymptotic confidence intervals for the threshold parameter. We have used these confidence intervals to improve the confidence interval construction for the regression slope parameters. An application to the U.S. unemployment rate illustrated how these techniques may be used in practical applications.

References

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

[2] Andrews, D.W.K. (1994), "Andrews, D.W.K. (1993), "Empirical process methods in econometrics," *Handbook of Econometrics, Vol IV*, 2248-2296, R.F. Engle and D.L. McFadden, eds., Elsevier Science, Amsterdam.

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

[4] Bai, J. (1996): "Estimation of a change point in multiple regression models," *Review of Economics and Statistics*, forthcoming.

[5] Chan, K.S. (1990a): "Testing for threshold autoregression," *The Annals of Statistics* 18, 1886-1894.

[6] Chan, K.S. (1990b): "Deterministic stability, stochastic stability, and ergodicity," Appendix to H. Tong, *Non-Linear Time Series: A Dynamical System Approach*, Oxford University Press, Oxford.

[7] Chan, K.S. (1991): "Percentage points of likelihood ratio tests for threshold autoregression," *Journal of the Royal Statistical Society, Series B*, 53, 691-696.

[8] 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.

[9] Chan, K.S. and H. Tong (1986): "On estimating thresholds in autoregressive models," *Journal of Time Series Analysis*, 7, 179-194.

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

[11] Davies, R.B. (1977): "Hypothesis testing when a nuisance parameter is present only under the alternative," *Biometrika*, 64, 247-254.

[12] Davies, R.B. (1987): "Hypothesis testing when a nuisance parameter is present only under the alternative," *Biometrika*, 74, 33-43.

[13] Dumbgen, L. (1991): "The asymptotic behavior of some nonparametric change point estimators," *The Annals of Statistics*, 19, 1471-1495.

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

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

[16] Hansen, B.E. (1996b): "Sample splitting and threshold estimation," working paper, Boston College.

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

[18] Picard, D. (1985): "Testing and estimating change-points in time series," *Advances in Applied Probability*, 17, 841-867.

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

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

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

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

[23] Yao, Y.-C. (1987): "Approximating the distribution of the ML estimate of the change-point in a sequence of independent r.v.'s," *The Annals of Statistics*, 3, 1321-1328.

7 Appendix: Proof of Theorem 1

We simply need to verify the following conditions, which allows us to invoke Theorems 1 and 2 of Hansen (1996b). For some $s > 1$ and $\delta > 0$,

1. y_t is strictly stationary with β-mixing coefficients β_m satisfying $\beta_m^{(s-1)/2s} = O\left(m^{-(1+\delta)}\right)$;
2. $E\left(e_t \mid F_{t-1}\right) = 0$;
3. $E\left|y_t\right|^{2s} < \infty$ and $E\left|e_t\right|^{2s} < \infty$;
4. $f(\gamma)$, $D(\gamma)$, and $D_s(\gamma) = E\left(\left(x_t' x_t\right)^s \mid y_{t-d} = \gamma\right)$ are continuous at $\gamma = \gamma_0$;
5. $f(\gamma_0) > 0$
6. $(\alpha - \beta)' D (\alpha - \beta) > 0$;
7. $P\left(y_{t-d} \in \Gamma\right) < 1$.

Chan (1990b) gives conditions for the strict stationarity of TAR processes. In the discussion following Theorem A1.11 (p. 464) he shows that under Assumption 1, parts 1 and 2, our TAR process y_t is strictly stationary and geometrically ergodic. The latter condition implies absolute regularity with exponentially declining coefficients, so Condition 1 is satisfied.

Condition 2 is satisfied since e_t is iid and mean zero. Condition 3 follows directly from the linear structure of y_t, Minkowski's inequality, and the assumption of finite $2+\delta$ moments for e_t. Condition 4 holds since e_t is iid with a continuous density. Condition 5 holds by the assumption that $f(\gamma)$ is everywhere positive. Condition 6 is guaranteed by Assumption 1, part 3. Since the support e_t is the entire real line, similarly the support of y_t is the entire real line. Condition 7 follows as Γ is a proper subset of R. $\square$