Threshold Autoregressions with a Unit Root[1]

Mehmet Caner

Koc University[2]

and

Bruce E. Hansen

Boston College[3]

August, 1997

[2]Department of Economics, Cayir Cad. No. 5, Istinye 80860, Istanbul, Turkey. mcaner@ku.edu.tr.

[3]Department of Economics, Chestnut Hill, MA 02167 USA. bruce.hansen@bc.edu

Abstract

This paper develops an asymptotic theory of inference for a two-regime threshold autoregressive (TAR) model with an autoregressive root which is local-to-unity. We find that the asymptotic null distribution of the Wald test for a threshold is non-standard and mildly dependent on the local-to-unity coefficient. We also study the asymptotic null distribution of the Wald test for an autoregressive unit root, and find that it is non-standard and dependent on the presence of a threshold effect. These tests and distribution theory allow for the joint consideration of non-linearity (thresholds) and non-stationarity (unit roots).

Our limit theory is based on a new set of tools which combines unit root asymptotics with empirical process methods. We work with a particular two-parameter empirical processes which converges weakly to a two-parameter Brownian motion. Our limit distributions involve stochastic integrals with respect to this two-parameter process. This theory is entirely new and may find applications in other contexts.

We illustrate the methods with an application to the U.S. monthly unemployment rate. We find strong evidence of a threshold effect. The point estimates suggest that in about 80% of the observations, the regression function is close to a driftless I(1) process, and in the other 20% of the observations, the regression function is mean-reverting with an unconditional mean of 5%. While the conventional ADF test for a unit root is quite insignificant, our TAR unit root test is arguably significant, with an asymptotic p-value of 3.5%, suggesting that the unemployment rate follows a stationary TAR process.

1 Introduction

The threshold autoregressive (TAR) model was introduced by Tong (1978) and has since become quite popular in non-linear time series. See Tong (1983,1990) for reviews. A sampling theory of inference has been quite slow to develop, however. Chan (1993) provided the first major contribution, showing that the least squares estimate of the threshold is super-consistent and the slope estimates have conventional asymptotic distributions. More recently, Hansen (1996) found the asymptotic distribution of the likelihood ratio test for a threshold, and Hansen (1997) found the asymptotic distribution of the threshold estimate. Chan and Tsay (1997) analyzed the related continuous TAR model, and found the asymptotic distribution of the parameter estimates in this model.

In all of the papers listed above, an important maintained assumption is that the data are stationary and ergodic. Tsay (1997) attempts an analysis for non-ergodic data, but his results require that the threshold is known *a priori*, which is not relevant for applications. Balke and Fomby (1997) introduce a multivariate model of threshold cointegration, but offer no rigorous distribution theory. An interesting simulation experiment is reported by Pippenger and Goering (1993) who document that the power of the Dickey-Fuller (1979) unit root test falls dramatically within one class of TAR models.

We extend this literature by examining a two-regime threshold autoregression allowing for an autoregressive root which is local-to-unity. We study the Wald test for a threshold, and find that its asymptotic null distribution is non-standard. This is partially due to the presence of a parameter which is not identified under the null (see Davies (1987), Andrews and Ploberger (1994) and Hansen (1996)), and partially due to the assumption of a near non-stationary autoregression. The asymptotic null distribution appears to mildly depend on the local-to-unity parameter but is otherwise free of nuisance parameters. To our knowledge ours is the first theory which allows for non-stationarity while testing for threshold effects.

We also examine the Wald test for a unit root in our threshold autoregressive model. We find that the asymptotic null distribution depends on whether or not there is a true threshold effect, and the critical values for the case where there is no threshold effect may be used as an asymptotic conservative bound. To our knowledge this is the first unit root test which allows for non-linearity in the regression function.

Our distribution theory is based on a new set of asymptotic tools utilizing a double-indexed empirical process which converges weakly to a two-parameter Brownian motion. We tabulate appropriate critical values and calculate approximating p-value functions based on a chi-square approximation. Monte Carlo simulations are provided to illustrate the finite

sample size and power of the tests. The simulations show that our unit root test has much improved power over the conventional ADF unit root test (Said and Dickey, 1984) when the true process is non-linear.

The methods are illustrated by an application to the monthly U.S. unemployment rate among adult males. We provide very strong evidence that the autoregression has a threshold non-linearity. This non-linearity takes the following form. When the unemployment rate has been relatively constant or increasing, the process behaves similarly to a driftless random walk. On the other hand, when the unemployment rate has experienced a recent sharp decrease, the process attempts to mean-revert towards 5%.

There is a substantial literature documenting non-linearities and threshold effects in the U.S. unemployment rate. A partial list includes Rothman (1991), Chen and Lee (1995), Montgomery, Zarnowitz, Tsay and Tiao (1996), Altissimo and Violante (1996), Chan and Tsay (1997), Hansen (1997) and Tsay (1997). This literature is connected to a broader literature studying non-linearities in the business cycle, which includes contributions by Neftci (1984), Hamilton (1989), Beaudry and Koop (1993), Potter (1995), and Galbraith (1996).

This paper is organized as follows. Section 2 presents the specific TAR model we study. Section 3 introduces a new set of asymptotic tools which are useful for the study of threshold processes with possible unit roots. Section 4 presents the distribution theory for the threshold test. Critical values and a Monte Carlo simulation of size and power are presented. Section 5 presents the distribution theory for the unit root test, including critical values and a simulation study. Section 6 is the empirical application to the U.S. unemployment rate. The mathematical proofs are presented in the Appendix.

A GAUSS program which replicates the empirical work is available from the webpage http://www2.bc.edu/~hansenb/.

2   TAR Model

The model is the following threshold autoregression (TAR) :

$$\Delta y_t = \theta_1' x_{t-1} 1_{\{Z_{t-1} < \lambda\}} + \theta_2' x_{t-1} 1_{\{Z_{t-1} \geq \lambda\}} + a(L)\Delta y_t + e_t, \tag{1}$$

$t = 1, ..., T$, where $x_{t-1} = (1 \ t \ y_{t-1})'$, $1_{\{\cdot\}}$ is the indicator function, $a(L) = a_1 L + \cdots + a_k L^k$ is a k'th order lag polynomial, and Z_{t-1} is any predetermined stationary variable with continuous marginal distribution function $F(\cdot)$. For our empirical analysis, we set $Z_t = y_t - y_{t-m}$ for

some $m \geq 1$.. Note that (1) specifies that the intercept, trend, and the slope coefficient on y_{t-1} may switch between regimes, but the coefficients on the lagged Δy_t variables do not switch. The threshold λ is unknown. It takes on values in the interval $\lambda \in \Lambda = [\lambda_1, \lambda_2]$ where λ_1 and λ_2 are picked so that $P(Z_t \leq \lambda_1) = \pi_1 > 0$ and $P(Z_t \leq \lambda_2) = \pi_2 < 1$. For the tables, simulations, and empirical work which follows, we set $\pi_1 = .15$ and $\pi_2 = .85$.

For some of our analysis, it will be convenient to separately discuss the components of θ_1 and θ_2. Partition these vectors as

$$\theta_1 = \begin{pmatrix} \mu_1 \\ \beta_1 \\ \rho_1 \end{pmatrix}, \qquad \theta_2 = \begin{pmatrix} \mu_2 \\ \beta_2 \\ \rho_2 \end{pmatrix}.$$

Thus (μ_1, μ_2) are the intercepts, (β_1, β_2) the trend slopes, and (ρ_1, ρ_2) are the slope coefficients on y_{t-1} in the two regimes.

The TAR model (1) is estimated by least squares. For each $\lambda \in \Lambda$, (1) is estimated by OLS:

$$\Delta y_t = \hat{\theta}_1(\lambda)' x_{t-1} 1_{\{Z_{t-1} < \lambda\}} + \hat{\theta}_2(\lambda)' x_{t-1} 1_{\{Z_{t-1} \geq \lambda\}} + \hat{a}_\lambda(L) \Delta y_t + \hat{e}_t(\lambda). \tag{2}$$

Let

$$\hat{\sigma}^2(\lambda) = T^{-1} \sum_1^T \hat{e}_t(\lambda)^2$$

be the OLS estimate of σ^2 for fixed λ. The least-squares estimate of the threshold λ is found by minimizing $\sigma^2(\lambda)$:

$$\hat{\lambda} = \operatorname*{argmin}_{\lambda \in \Lambda} \hat{\sigma}^2(\lambda).$$

The least-squares estimates of the other parameters are then found by plugging in the point estimate $\hat{\lambda}$, i.e. $\hat{\mu}_1 = \hat{\mu}_1(\hat{\lambda})$, etc. We can write the estimated model as

$$\Delta y_t = \hat{\theta}_1' x_{t-1} 1_{\{Z_{t-1} < \hat{\lambda}\}} + \hat{\theta}_2' x_{t-1} 1_{\{Z_{t-1} \geq \hat{\lambda}\}} + \hat{a}(L) \Delta y_t + \hat{e}_t. \tag{3}$$

The estimates (3) can be used to conduct inference concerning the parameters of (1) using standard Wald statistics. We are particularly interested in restrictions concerning the presence of a threshold and the presence of a unit root. First, the threshold effect disappears under the joint hypothesis

$$H_0 : \theta_1 = \theta_2. \tag{4}$$

An appropriate test of (4) is the standard Wald statistic W_T for this restriction from (3). To establish notation, let $W_T(\lambda)$ denote the Wald statistic of the hypothesis (4) for fixed λ

from regression (2). It is useful to note that because the parameter λ does not enter the model under H_0, we have the relationship

$$W_T = W_T(\hat{\lambda}) = \sup_{\lambda \in \Lambda} W_T(\lambda).$$

The other hypothesis of major interest is the presence of a unit root in the autoregressive structure. A unit root in y_{t-1} occurs in (1) when

$$H_0 : \rho_1 = \rho_2 = 0. \tag{5}$$

The appropriate test of (5) is the standard Wald statistic R_T from (3). To fix notation, let $R_T(\lambda)$ be the standard Wald statistic for hypothesis (5) for fixed λ, then $R_T = R_T(\hat{\lambda})$. The statistic R_T may be viewed as a two-parameter generalization of the standard Dickey-Fuller statistic.

From the estimates (3) we have proposed two Wald tests — W_T and R_T — which test restrictions on the coefficients implying the absence of threshold effects and presence of a unit root, respectively. While the statistics are standard, their sampling distributions are non-standard. We explore large-sample approximations in the following sections.

3 Unit Root Asymptotics for Threshold Processes

The sampling distributions for our proposed statistics require some new asymptotic tools. The problems and solutions are easiest to illustrate in a simplified setting. Take the model

$$\Delta y_t = \rho y_{t-1} 1_{\{Z_{t-1} < \lambda\}} + e_t, \tag{6}$$

where the restriction $\rho = 0$ implies that y_t is a random walk with no threshold effect. The test of this hypothesis is the maximal pointwise squared t-statistic for ρ, where the maximum is taken over all values of λ. Let $\hat{\rho}(\lambda)$ denote the point estimate of ρ, considered as a function of λ. It is helpful to note that $1_{\{Z_{t-1} < \lambda\}} = 1_{\{G(Z_{t-1}) < G(\lambda)\}}$ for any monotonically increasing function $G(\cdot)$. Since Z_{t-1} has a continuous distribution F, we have in particular that $1_{\{Z_{t-1} < \lambda\}} = 1_{\{U_{t-1} < u\}}$ where $U_t = F(Z_t)$ and $u = F(\lambda)$. Thus $\hat{\rho}(\lambda) = \hat{\rho}^*(u)$, where

$$
\begin{aligned}
T\hat{\rho}^*(u) &= \frac{\dfrac{1}{T}\displaystyle\sum_{t=2}^{T} 1_{\{U_{t-1}<u\}} y_{t-1}\Delta y_t}{\dfrac{1}{T^2}\displaystyle\sum_{t=2}^{T} 1_{\{U_{t-1}<u\}} y_{t-1}^2} \\
&= \frac{N_T(u)}{D_T(u)},
\end{aligned}
\tag{7}
$$

say. We will focus our attention on (7). We need the following restrictions.

Assumption 1 (e_t, Z_t) *is strictly stationary and ergodic and adapted to the sigma-field* $\Im_t$. *In addition,*

$$E\left(e_t \mid \Im_{t-1}\right) = 0 \tag{8}$$

$$E\left(e_t^2 \mid \Im_{t-1}\right) = \sigma^2, \tag{9}$$

and for some $\gamma > 1$,

$$E\left(e_t^{2\gamma} \mid \Im_{t-1}\right) \le B < \infty. \tag{10}$$

Conditions (8) and (9) specify that the error is a conditionally homoskedastic martingale difference sequence. Condition (10) bounds the extent of heterogeneity in the conditional distribution of e_t.

Under the null hypothesis, y_t is random walk, so $T^{-1/2}y_{[Ts]} \Rightarrow \sigma W(s)$, a Brownian motion. Since $N_T(u)$ and $D_T(u)$ are a function of y_{t-1}, one might think that they could be studied using the asymptotic techniques introduced by Phillips (1987). A major difficulty arises, however, due to the presence of the indicator functions $1_{\{U_{t-1}<u\}}$.

First consider the denominator $D_T(u)$. Its asymptotic distribution can be found from a generalization of the fact that integrated and stationary processes are asymptotically uncorrelated.

Theorem 1 *If the array* X_{Tt} *satisfies* $X_{T[Ts]} \Rightarrow X(s)$ *on* $s \in [0,1]$, *and* $X(s)$ *is continuous almost surely, then*

$$\frac{1}{T}\sum_{t=1}^{T} 1_{\{U_{t-1}<u\}} X_{Tt} \Rightarrow u \int_0^1 X(s)ds.$$

on $u \in [0,1]$.

Setting $X_{Tt} = T^{-1/2}y_t$, Theorem 1 implies that

$$D_T(u) \Rightarrow u\sigma^2 \int_0^1 W(s)^2 ds$$

on $u \in [0,1]$. We see that $D_T(u)$ is asymptotically a (random) linear function of its argument u.

We now turn to the numerator. To develop a sampling theory, define

$$e_t(u) = 1_{\{U_{t-1}<u\}} e_t.$$

7

Note that for each u, $e_t(u)$ is a is a strictly stationary and ergodic martingale difference sequence with variance

$$
\begin{aligned}
Ee_t(u)^2 &= E\left(1_{\{U_{t-1}<u\}}e_t^2\right) \\
&= E\left(1_{\{U_{t-1}<u\}}\right)\sigma^2 \\
&= u\sigma^2.
\end{aligned}
$$

The second equality follows since e_t is conditionally homoskedastic, and the third equality holds since the marginal distribution of U_{t-1} is $U[0,1]$.

Under the null hypothesis $\rho = 0$,

$$
N_T(u) = \frac{1}{T}\sum_{t=2}^{T} y_{t-1}e_t(u),
$$

which takes a form to which we can apply the theory of convergence to stochastic integrals (e.g. Kurtz and Protter (1991)). Define the partial-sum process

$$
Y_t(u) = \sum_{i=1}^{t} e_i(u)
$$

and scaled array

$$
\begin{aligned}
Y_T(s,u) &= \frac{1}{\sqrt{T}}Y_{[Ts]}(u) \\
&= \frac{1}{\sqrt{T}}\sum_{t=1}^{[Ts]} e_t(u).
\end{aligned}
$$

Note that $EY_T(s,u) = 0$ and

$$
\begin{aligned}
EY_T(s,u)^2 &= \frac{1}{T}\sum_{i=1}^{[Ts]} E\left(e_i 1_{\{U_{i-1}<u\}}\right)^2 \\
&\simeq su\sigma^2
\end{aligned}
$$

which is linear in both s and u. It is helpful to recall that the variance of a Brownian motion is linear in its argument. The double-linearity of the variance of $Y_T(s,u)$ suggests the need for a double-argument generalization of Brownian motion.

Definition 1 $W(s,u)$ *a two-parameter Brownian motion on* $[0,1]^2$ *if it is a zero-mean Gaussian process with covariance kernel*

$$
E\left(W(s_1,u_1)W(s_2,u_2)\right) = (s_1 \wedge s_2)(u_1 \wedge u_2).
$$

8

Indeed, we can show that $Y_T(s, u)$ can be approximated in distribution by a two-parameter Brownian motion.

Theorem 2 *Under Assumption 1,*

$$Y_T(s, u) \Rightarrow \sigma W(s, u) \tag{11}$$

on $(s, u) \in [0, 1]^2$ *as* $T \to \infty$.

It may be helpful to think of Theorem 2 as a two-parameter generalization of the usual functional limit theorem.

To use Theorem 2 to establish the asymptotic distribution of $N_T(u)$, we need to consider $N_T(u)$ as a stochastic integral with respect to a two-parameter process, which we define as integration over the first argument, holding the second argument constant. Specifically,

$$\int_0^1 X(s) dZ(s, u) \equiv \operatorname*{plim}_{N \to \infty} \sum_{j=1}^{N} X\left(\frac{j-1}{N}\right) \left(Z\left(\frac{j}{N}, u\right) - Z\left(\frac{j-1}{N}, u\right)\right),$$

which is a stochastic process in u. Using this notation, we have

$$N_T(u) = \int_0^1 Y_T(s) dY_T(s, u)$$

where $Y_T(s) = Y_T(s, 1) = T^{-1/2} y_{[Ts]}$. We will need a somewhat more general class of integrating functions. Let X_{Tt} be any $\Im_{t-1}-$ adapted process such that $X_T(s) = X_{T[Ts]} \Rightarrow X(s)$ on $s \in [0, 1]$ (jointly with (11)) and $X(s)$ is continuous almost surely.

Theorem 3 *Under Assumption 1,*

$$\int_0^1 X_T(s) dY_T(s, u) \Rightarrow \sigma \int_0^1 X(s) dW(s, u)$$

on $u \in [0, 1]$ *as* $T \to \infty$.

This result is a natural extension of the theory of weak convergence to stochastic integrals to the case of integration with respect to a two-parameter process. Theorems 2 and 3 will serve as the building blocks for the subsequent theory developed in this paper.

Theorem 3 thus implies that $N_T(u) \Rightarrow \sigma^2 \int_0^1 W(s) dW(s, u)$, where $W(s) = W(s, 1)$ is a standard Brownian motion. Theorems 1 and 3 combine to yield

$$T\hat{\rho}^*(u) = \frac{N_T(u)}{D_T(u)} \Rightarrow \frac{\int_0^1 W(s) dW(s, u)}{u \int_0^1 W(s)^2 ds}. \tag{12}$$

9

on $u \in [0, 1]$. (12) gives the asymptotic distribution for $\hat{\rho}^*(u)$ considered as a function of u under the null hypothesis $\rho = 0$.

While model (1) is more complicated than (6), the basic methods of analysis developed in this section can be generalized appropriately. These results are presented in the following sections.

4   Testing for a Threshold Effect

In Section 2 we introduced the test statistic W_T as the natural Wald statistic for the test of the hypothesis (4) of no threshold effect within model (1). Under this hypothesis the parameter λ is not identified. The asymptotic distribution of W_T for stationary data has been investigated by Davies (1987), Andrews and Ploberger (1994) and Hansen (1996). Our concern is with data which is either non-stationary or near-non-stationary. We incorporate this condition by specifying the largest root as local-to-unity. Let $\rho = \rho_1 = \rho_2$ be the coefficient on y_{t-1} which is common to the two regimes under (4). For fixed c, we specify that this coefficient is local-to-zero:

$$\rho = -c/T. \tag{13}$$

Under (13), the presence of a linear trend in the model (1) induces a quadratic trend in the reduced form for y_t. To prevent this possibility we assume that the true values of the trend coefficients is zero:

$$\beta_1 = \beta_2 = 0. \tag{14}$$

This is conventional in the analysis of autoregressions with unit roots or near unit roots.

Let $W(s, u)$ be a two-parameter Brownian motion as defined in the previous section. Let $W(s) = W(s, 1)$ be a standard Brownian motion derived from $W(s, u)$ and let $W_c(s)$ be the Ornstein-Uhlenbeck process that is the solution to the stochastic differential equation

$$dW_c(s) = -cW_c(s) + dW(s).$$

Set $X_c(s) = (1 \ s \ W_c(s))'$.

Theorem 4 *Under the null of no threshold $H_0 : \theta_1 = \theta_2$ and Assumption 1,*

$$W_T \Rightarrow T_c = \sup_{\pi_1 \le u \le \pi_2} S_c(u),$$

where

$$S_c(u) = \frac{(J_c(u) - uJ_c(1))' \left(\int_0^1 X_c(s)X_c(s)'ds \right)^{-1} (J_c(u) - uJ_c(1))}{u(1-u)}, \tag{15}$$

10

$$J_c(u) = \int_0^1 X_c(s)dW(s,u),$$

$$\pi_1 = F(\lambda_1), \text{ and } \pi_2 = F(\lambda_2).$$

Theorem 4 gives the large sample distribution of the conventional Wald statistic for a threshold for the non-stationary autoregression (1). It is noticeably different from the distribution found by Hansen (1996). Unlike the stationary case, the limit distribution T_c is relatively free of nuisance parameters, only depending on the local-to-unity parameter c. (The limits π_1 and π_2 are not considered nuisance parameters since they are under the control of the econometrician.)

Table 1: Asymptotic Critical Values for Threshold Test

$c =$	-10	**0**	15	30	45
1%	17.23	**17.14**	17.15	17.29	17.44
5%	13.30	**13.30**	13.28	13.38	13.39
10%	11.51	**11.51**	11.54	11.56	11.65
20%	9.63	**9.62**	9.62	9.63	9.74

Calculated from 50,000 simulations using samples of size 10,000.

Estimates of the critical values for T_c are given in Table 1. These critical values were calculated by simulation, using 50,000 replications of samples of size $T = 10,000$ using Gaussian innovations. The estimates indicate that the critical values are not very sensitive to c, and they appear to be minimized at $c = 0$. Our concrete recommendation is to use the critical values corresponding to $c = 0$. These are the correct critical values in the leading case where the series actually has a unit root. If the leading root is non-unity, but reasonably close, there will be only a slight distortion in the asymptotic approximation.

It is often convenient to report p-values for observed tests. Hansen (1997) shows how to use chi-square distributions to fit non-standard asymptotic distributions. Using his methods, we obtained the approximation

$$-2.0001 + 1.096 \cdot T_0 \approx \chi^2(6). \tag{16}$$

to the distribution of T_c for the case $c = 0$. This means that the statistic

$$W_T^* = -2.0001 + 1.096 \cdot W_T$$

11

has an asymptotic distribution which is approximately $\chi^2(6)$. The approximation is extremely good. If (16) is used to calculate asymptotic p-values, the maximal error is less than 0.2% (relative to simulated p-values).

4.1 A Monte Carlo Experiment

In order to examine the size and power of the proposed test a small sample study is conducted. The model used is equation (1) with $k = 1$ and $z_{t-1} = \Delta y_{t-1}$:

$$
\begin{aligned}
\Delta y_t &= (\mu_1 + \beta_1 t + \rho_1 y_{t-1})1_{\{\Delta y_{t-1} < \lambda\}} \\
&+ (\mu_2 + \beta_2 t + \rho_2 y_{t-1})1_{\{\Delta y_{t-1} \geq \lambda\}} \\
&+ a_1 \Delta y_{t-1} + e_t,
\end{aligned}
$$

and e_t iid $N(0,1)$. The sample size we use is $T = 100$. We fixed $\beta_1 = \beta_2 = 0$ for all exercises. We examine nominal 5% size tests using the asymptotic critical values for $c = 0$. All calculations are empirical rejection frequencies from 1000 monte carlo replications.

We first examined the size of the test. Under the null hypothesis of no threshold, data is generated by the process

$$
\Delta y_t = \mu + \rho y_{t-1} + a_1 \Delta y_{t-1} + e_t. \tag{17}
$$

We set $\mu = 0$ and varied ρ among -.15, -.10, -.05 and .02. (The negative values correspond to stationary processes and the positive value implies a mildly explosive process.) We also varied the second AR parameter a_1 among the values -.7, -.5, -.2, .2, .5, and .7 to assess sensitivity to additional serial correlation. The results are presented in Table 2.

Table 2: Size of Threshold Test

$\rho =$	$-.15$	-.10	-.05	0	.02
$a_1 = -.7$	9.4	9.5	9.7	8.6	10.3
$a_1 = -.5$	9.1	8.5	8.8	9.0	8.8
$a_1 = -.2$	8.2	8.2	8.3	8.1	7.3
$a_1 = 0$	7.0	6.6	6.8	6.7	6.3
$a_1 = .2$	7.7	8.9	8.9	9.0	10.6
$a_1 = .5$	9.1	8.7	8.5	9.5	18.5
$a_1 = .7$	9.1	10.0	10.1	12.0	36.5

Note: $T = 100$. Nominal size 5%. Rejection rates from 1000 replications.

12

When $a_1 = 0$, the actual size of the test is reasonable, ranging from 6% to 7% depending on ρ. For $a \neq 0$, however, the rejection rates are larger. For $\rho \leq 0$, the actual size ranges from 8% to 12%. When $\rho > 0$ and a is large, the size can be quite poor. This parameter configuration is probably not meaningful for most economic applications, however. We conclude that the size of the test is reasonable for the non-explosive region $\rho \leq 0$.

Second, we explored the power of the test against local alternatives. We set the threshold at $\lambda = 0$ and the serial correlation parameter $a_1 = 0$ for this study. The alternative includes both $\mu_1 \neq \mu_2$ and $\rho_1 \neq \rho_2$ which we examined separately.

To study the impact of a switching intercept, we considered the symmetric case $\mu_1 = -\mu_2 = \Delta\mu$ and varied $\Delta\mu$ among .2, .5 and 1.0. We also varied the nuisance parameter $\rho = \rho_1 = \rho_2$ among 0, -.05 and -.10 to assess sensitivity. The results are presented in Table 3. We see that the power of the test is increasing in $\Delta\mu$ as expected, and that the power is not sensitive to the nuisance parameter ρ.

Table 3: Power of Threshold Test Against Intercept Switch

$\rho =$	0	-.05	-.10
$\Delta\mu = .2$	23	19	20
$\Delta\mu = .5$	93	95	91
$\Delta\mu = 1.0$	100	100	100

Note: $T = 100$. Nominal size 5%. Rejection rates from 1000 replications.

To study the impact of a switching ρ coefficient we define $\Delta\rho = \rho_2 - \rho_1$ as the size of the threshold effect. We let $\Delta\rho$ range over $\{-.05, -.10, -.15\}$, ρ_1 among $\{0, -.10\}$ and $\mu = \mu_1 = \mu_2$ among $\{0, .5, 1.0\}$. The results are presented in Table 4. As expected, the power of the tests is increasing in $|\Delta\rho|$. What is somewhat surprising is the strong dependence of power on the nuisance parameters, with power increasing in $\mu > 0$, and much higher for $\rho_1 = 0$ than $\rho_1 = -.10$. Our best guess is that small ρ_1 and large μ induces trends (stochastic and/or deterministic) into the stochastic process which increase the signal to noise ratio, hence decreasing standard errors and increasing power.

Table 4: Power of Threshold Test Against Switch in ρ

	$\rho_1 = 0$			$\rho_1 = -.10$		
$\mu =$	0	0.5	1.0	0	0.5	1.0
$\Delta\rho = -.05$	14	57	99	10	10	24
$\Delta\rho = -.10$	21	73	100	13	18	58
$\Delta\rho = -.15$	29	79	100	15	27	79

Note: $T = 100$. Nominal size 5%. Rejection rates from 1000 replications.

5   Testing for a Unit Root

In section 4 we introduced R_T as the appropriate test statistic for a unit root. We now derive large sample approximations to the distribution of R_T under the null hypothesis of a unit root (5). A difficulty arises in specifying the threshold effect, as the null of a unit root is compatible with either the existence or non-existence of a threshold effect. Since $\rho_1 = \rho_2$ under the null, a threshold effect occurs when $\mu_1 \neq \mu_2$ but there is no threshold effect when $\mu_1 = \mu_2$. It turns out that the asymptotic distribution of R_T is different in these two cases. Since the truth is typically unknown we consider both cases.

Let $r(s) = (1 \ s)'$. Let $W(s) = W(s, 1)$ (a standard Brownian motion) and let

$$W^*(s) = W(s) - \int_0^1 W(a)r(a)'da \left(\int_0^1 r(a)r(a)'da \right)^{-1} r(s)$$

be detrended $W(s)$. Let $S_0(u)$ be defined in (15) with $c = 0$.

Theorem 5 *Suppose that $\mu_1 = \mu_2$. Under $H_0 : \rho_1 = \rho_2 = 0$,*

$$R_T \Rightarrow R^* = R(u^*)$$

where

$$u^* = \operatorname*{argmax}_{u \in [\pi_1, \pi_2]} S_0(u),$$

$$R(u) = \left[\frac{\int_0^1 W^*(s)dW_1(s, u)'}{\left(\int_0^1 W^*(s)^2 ds \right)^{1/2}} \right]^2 + \left[\frac{\int_0^1 W^*(s)dW_2(s, u)}{\left(\int_0^1 W^*(s)^2 ds \right)^{1/2}} \right]^2,$$

$$W_1(s, u) = \frac{W(s, u)}{\sqrt{u}},$$

14

and

$$W_2(s, u) = \frac{W(s, 1) - W(s, u)}{\sqrt{1 - u}}.$$

The limiting maximizer u^* is random since the threshold is not identified under the conditions of Theorem 5. The processes $W_1(s, u)$ and $W_2(s, u)$ are scaled so that for fixed u they are standard (and independent) Brownian motions. What is important for inference is that the limiting distribution in Theorem 5 is free of nuisance parameters.

Theorem 6 *Suppose that $\mu_1 \neq \mu_2$. Under $H_0 : \rho_1 = \rho_2 = 0$, for some $a \in [0, 1)$,*

$$\begin{aligned} R_T &\Rightarrow R_* = \chi_1^2 + \left(aN(0, 1) + \left(1 - a^2\right)^{1/2} DF \right)^2 \\ &\leq \chi_1^2 + DF^2 \end{aligned}$$

where

$$DF^2 = \frac{\left(\int_0^1 W^* dW\right)^2}{\int_0^1 W^{*2}}$$

is the square of the conventional detrended Dickey-Fuller t-distribution, and χ_1^2 is an independent chi-square random variable with one degree of freedom.

Theorem 6 shows that the limiting distribution of the unit root test R_T takes a mixture form, but can be bounded by the sum of the squared Dickey-Fuller and chi-square distributions, which is free of nuisance parameters.

Theorems 5 and 6 together give asymptotic approximations to the null distribution of the TAR unit root test R_T under differing assumptions concerning the threshold. The source of the differences lies in the identification of the threshold parameter λ. When there is no threshold effect, then λ is not identified, so $\hat{\lambda}$ remains random in large sample. Thus $R_T(\hat{\lambda})$ inherits randomness both from the random function $R_T(\lambda)$ and the random argument $\hat{\lambda}$. In contrast, when there is a threshold effect ($\lambda_1 \neq \lambda_2$) then λ is identified and $\hat{\lambda}$ will be close to λ_0 in large samples. Hence the randomness in $R_T(\hat{\lambda})$ will be due mostly to the random function $R_T(\lambda)$. This heuristic analysis suggests that R^* will be "more random" than R_*, in the sense that R^* should have large critical values. Table 5 reports critical values for both R^* and R_*, and it is indeed the case that the critical values of R^* exceed those of R_*.

Using the methods of Hansen (1997), chi-square approximations were fit to these two distributions. We found that

$$1.006 + 1.105 \cdot R_* \approx \chi_8^2$$

15

and

$$1.221 + 1.001 \cdot R^* \approx \chi_9^2.$$

were excellent approximations.

Table 5: Asymptotic Critical Values for Unit Root Test

	R_*	R^*
1%	17.29	20.37
5%	13.12	15.69
10%	11.17	13.43
20%	9.07	11.02

Calculated from 50,000 simulations using samples of size 10,000.

5.1 Finite Sample Study

Using Monte Carlo methods, we now examine the finite sample performance of the unit root test R_T in the context of an AR(2) model, and contrast its performance with the conventional Augmented Dickey-Fuller (ADF) t-test.

We first study the size of nominal 5% tests. The data is simulated under the null from model (1) with $k = 1$ and $m = 1$, setting $\rho_1 = \rho_2 = 0$ and $\beta_1 = \beta_2 = 0$. Thus the data is generated by

$$\Delta y_t = \mu_1 1_{\{\Delta y_{t-1} < \lambda\}} + \mu_2 1_{\{\Delta y_{t-1} \geq \lambda\}} + a_1 \Delta y_{t-1} + e_t.$$

We used samples of size $T = 100$ and generate e_t as iid $N(0,1)$. The tests should be a function of (μ_1, μ_2) only through the difference $\Delta\mu = \mu_1 - \mu_2$, so we set $\mu_1 = -\mu_2$ and varied $\Delta\mu$ among $\{0, .4, 1, 2\}$. We also varied a_1 among $\{-.5, -.2, 0, .2, .5\}$, and set $\lambda = 0$ for simplicity. Table 6 contains the results, reporting rejection frequencies from 1000 Monte Carlo replications. As discussed in the previous section, there are two sets of critical values: the liberal critical values R_* and the conservative critical values R^*, so we report rejection rates using both critical values.

It is clear from Table 6 that the R_* critical value is a poor approximation for all cases examined, as the rejection rates are uniformly above the nominal size. The R^* critical values produces conservative rejection rates for large $\Delta\mu$ and a_1, yet liberal rates for small $\Delta\mu$. The rejection rates are quite sensitive to the autoregressive parameter a_1. Based on the simulation results, we recommend the R^* critical values for the R_T test.

Table 6: Size of Unit Root Tests

	R_T test								ADF test			
	R_* critical values				R^* critical values							
$\Delta\mu =$	0	.4	1	2	0	.4	1	2	0	.4	1	2
$a_1 = -.5$	17.6	15.5	12.8	9.0	9.7	7.6	7.5	5.9	6.5	5.7	4.6	2.7
$a_1 = -.2$	15.3	18.4	9.6	9.1	9.6	8.7	5.8	2.5	7.8	5.4	6.5	3.1
$a_1 = 0$	20.0	16.8	10.5	7.9	11.5	8.4	6.0	3.0	6.6	6.3	6.1	3.6
$a_1 = .2$	18.3	16.0	10.8	7.3	10.9	8.5	5.7	2.8	6.1	7.6	6.1	4.5
$a_1 = .5$	23.8	19.7	14.3	7.2	12.9	11.0	7.9	3.7	6.2	7.5	7.5	6.2

Note: $T = 100$. Nominal size 5%. Rejection rates from 1000 replications.

We next explore the power of the tests against some local alternatives. The model is the same as before except that we fix the serial correlation parameter at $a_1 = 0$ and do not impose $\rho_1 = \rho_2 = 0$. We consider two sample sizes, $T = 100$ and $T = 200$. We report size-adjusted power (rejection rates based on finite sample critical values) to control for the size distortions found in Table 9.

We consider two experiments. In the first, we restrict $\rho_1 = \rho_2 = \rho$, and vary ρ among $\{-.05, -.15, -.25\}$. This is the setting which should be the most favorable to the ADF test, as there is no difference in the serial correlation coefficients between the two regimes. The results are presented in Table 7. When $\Delta\mu = 0$ there is no threshold effect and the ADF test has considerably more power than the R_T test. As $\Delta\mu$ is increased, however, the power of the R_T test remains roughly invariant, while the power of the ADF test falls to zero, even for the larger sample size. Apparently, the conventional ADF test does not have the ability to discriminate a unit root from a stationary root when there is a strong threshold effect. These results are similar to those reported by Pippenger and Goering (1993), who find that the power of the Dickey-Fuller unit root test falls in a similar TAR model.

For our second power experiment, we allow for a threshold effect in the serial correlation coefficient, setting $\rho_1 = 0$ and letting ρ_2 vary among $\{-.05, -.15, -.25\}$. The results are presented in Table 8, and are similar to those from the first experiment. The obvious conclusion from the simulation exercises if the true process is a TAR, the standard ADF statistic (from a linear model) will not be able to distinguish a non-stationary process from a stationary process. Other experiments (not reported) using other choices for m, ρ_1 and Z_t reached similar conclusions.

Table 7: Power of R_T and ADF unit root tests
$$\rho_1 = \rho_2 = \rho < 0$$

	R_T			ADF		
	$\rho = -.05$	$\rho = -.15$	$\rho = -.25$	$\rho = -.05$	$\rho = -.15$	$\rho = -.25$
			$T = 100$			
$\Delta\mu = 0$	7	29	56	14	60	94
$\Delta\mu = 0.4$	7	21	49	2	10	34
$\Delta\mu = 1$	7	20	47	0	1	3
$\Delta\mu = 2$	8	24	56	0	0	0
			$T = 200$			
$\Delta\mu = 0$	27	81	99	53	98	100
$\Delta\mu = 0.4$	13	66	99	6	56	97
$\Delta\mu = 1$	13	66	99	0	6	43
$\Delta\mu = 2$	14	76	99	0	0	0

Note: Nominal size 5%. Rejection rates from 1000 replications.

Table 8: Power of R_T and ADF unit root tests,
$$\rho_1 = 0, \ \rho_2 < 0$$

	R_T			ADF		
	$\rho_2 = -.05$	$\rho_2 = -.15$	$\rho_2 = -.25$	$\rho_2 = -.05$	$\rho_2 = -.15$	$\rho_2 = -.25$
			$T = 100$			
$\Delta\mu = 0$	2	17	39	8	22	43
$\Delta\mu = 0.4$	6	17	41	2	3	5
$\Delta\mu = 1$	8	17	41	0	0	1
$\Delta\mu = 2$	5	18	43	0	0	0
			$T = 200$			
$\Delta\mu = 0$	18	78	96	31	56	73
$\Delta\mu = 0.4$	5	43	90	2	12	41
$\Delta\mu = 1$	10	61	94	0	1	3
$\Delta\mu = 2$	14	63	95	0	0	0

Note: Nominal size 5%. Rejection rates from 1000 replications.

6 U.S. Unemployment Rate

Our application is to the U.S. unemployment rate among adult males, monthly from 1956 through April 1997. To establish a baseline, we first fit a linear model with $k = 12$ lagged differences. The OLS estimates and standard errors are reported in Table 9. Note that the point estimate for ρ is $\hat{\rho} = -0.016$. Its t-statistic (the ADF test for a unit root) is insignificant at -2.408. This leads to the standard conclusion that the linear representation for the unemployment rate has a unit root.

Table 9: OLS Estimates, Linear Model

Regressor	Estimate	s.e.
Constant	0.069	(0.030)
t	0.00005	(0.00007)
y_{t-1}	-0.016	(0.007)
Δy_{t-1}	0.072	(0.045)
Δy_{t-2}	0.252	(0.045)
Δy_{t-3}	0.152	(0.046)
Δy_{t-4}	0.121	(0.047)
Δy_{t-5}	0.042	(0.047)
Δy_{t-6}	-0.022	(0.047)
Δy_{t-7}	-0.032	(0.047)
Δy_{t-8}	-0.020	(0.047)
Δy_{t-9}	0.005	(0.047)
Δy_{t-10}	0.042	(0.047)
Δy_{t-11}	0.073	(0.045)
Δy_{t-12}	-0.184	(0.046)

ADF= -2.408

The next question to ask is if there is any evidence to support a threshold model of the form (1). Setting $Z_t = y_t - y_{t-m}$ we need to select an appropriate value for m. In Table 10, we report the values of the threshold test statistics W_T for each fixed m from 1 through 12. We see that 10 of the 12 W_T statistics exceed the asymptotic 1% critical value of 17.1 from Table 1. The asymptotic p-values for the Wald statistics are also reported, calculated using the approximating p-value function. Since the Monte Carlo simulation of

section 4 indicated that there may be some size distortions in finite samples, we also report a bootstrap p-value for each test statistic. This p-value is calculated as follows. 1000 samples were generated from an AR(13) with the estimated coefficients from Table 9, using the initial conditions in the data and the empirical distribution of the OLS residuals to generate the errors. These samples satisfy the null of no threshold, so are valid bootstrap replications of the null distribution. For each sample and each m, the W_T statistic was calculated from the simulated data, and the percentage which exceeded the actual W_T of Table 10 is the simulation estimate of the bootstrap p-value. These are also reported in Table 10. While the bootstrap p-values are slightly higher than the asymptotic p-values, they do not change any conclusions.

Table 10: Threshold Tests, Fixed m

m	W_T	Asymptotic p-value	Bootstrap P-Value
1	15.0	0.025	0.028
2	33.6	0.000	0.000
3	18.5	0.006	0.013
4	20.5	0.002	0.004
5	20.5	0.002	0.004
6	13.3	0.049	0.063
7	21.6	0.001	0.003
8	20.9	0.002	0.005
9	30.0	0.000	0.000
10	27.3	0.000	0.000
11	22.7	0.001	0.002
12	23.9	0.000	0.002

Since the W_T test reject the null of no threshold for practically any choice of m, it appears obvious that we can reject the linear AR model in favor of the TAR model. As a general rule, however, this testing methodology is subject to the criticism that it conditions on m, while m is generally unknown. We can address this criticism by making the selection of m endogenous. The least squares estimate of m is the value which minimizes the residual variance. Since the Wald test W_T is a monotonic function of the residual variance, this is equivalent to selecting m as the value which maximizes W_T. This estimate is $\hat{m} = 2$, yielding a threshold test statistic of $W_T = 33.6$. The reported asymptotic p-value of 0.000 assumes that m is known and fixed. Our theory does not explicitly allow for estimated m. It is easy, however, to incorporate into the calculation of bootstrap p-values. For each of the 1000

bootstrap samples described above, the largest W_T across the 12 values of m was stored. The percentage of these statistics which exceeded the observed value of $W_T = 33.6$ is the bootstrap p-value. We found that this p-value is still 0.000, implying that it is extremely unlikely that the linear AR model (1) could generate a test statistic this large. These results are summarized in Table 11. We conclude that there is very strong evidence for a TAR model.

Table 11: Threshold Tests, Estimated m

$\hat{m}$	W_T	Asymptotic p-value	Bootstrap P-Value
2	33.6	0.000	0.000

Table 12: LS Estimates, TAR Model, $m = 2$
$$\hat{\lambda} = -0.240$$

	Regressor	Estimate	s.e.
$(Z_{t-1} < \hat{\lambda})$	Constant	0.243	(0.071)
	t	0.00050	(.00015)
	y_{t-1}	-0.050	(0.015)
$(Z_{t-1} > \hat{\lambda})$	Constant	0.062	(0.032)
	t	-0.00000	(0.00008)
	y_{t-1}	-0.017	(0.007)
$a(L)$	Δy_{t-1}	0.178	(0.051)
	Δy_{t-2}	0.333	(0.050)
	Δy_{t-3}	0.107	(0.046)
	Δy_{t-4}	0.109	(0.046)
	Δy_{t-5}	0.051	(0.046)
	Δy_{t-6}	-0.037	(0.046)
	Δy_{t-7}	-0.017	(0.046)
	Δy_{t-8}	-0.009	(0.046)
	Δy_{t-9}	0.028	(0.046)
	Δy_{t-10}	0.038	(0.045)
	Δy_{t-11}	0.078	(0.045)
	Δy_{t-12}	-0.178	(0.045)

The point estimates and standard errors from the model are reported in Table 12. The point estimate of the threshold $\hat{\lambda}$ is -0.24. Thus the TAR splits the regression function depending on whether the variable $Z_{t-1} = y_{t-1} - y_{t-3}$ lies above or below -0.24. The first regime is when $Z_{t-1} < -0.24$, which occurs when the unemployment rate has fallen by more

than .24 points (e.g. from 5.64 to 5.40) over a two month period. Approximately 20% of the observations fall in this category. The second regime is when $Z_{t-1} > -0.24$, which occurs when the unemployment rate has fallen by less than .24 points, has stayed constant, or has risen, over a two month period. Approximately 80% of the observations fall in this regime. Comparing the coefficients in the first regime (falling unemployment) versus those in the second (constant or rising unemployment), we see the following differences. The second regime behaves essentially as a driftless random walk, as the constant and trend estimates are insignificant, and the coefficient on y_{t-1} is near zero. In contrast, the coefficients in the first regime imply mean-reversion. It is perhaps helpful to observe that an AR process with the "regime 1" parameters has an unconditional mean of $-\mu_1/\rho_1 = 4.86$ (if we set $\beta_1 = 0$), so regime 1 may be viewed as having a tendency to revert to an unemployment rate of 5%. Since 80% of the observations fall in the random walk regime 2, it is not surprising that the linear representation "looks" like a unit root process.

We can of course rigorously address the unit root question using the R_T test. Table 13 reports the R_T test statistic, the asymptotic p-values calculated using both the liberal lower bound and the conservative upper bound, and two bootstrap p-values. For both bootstrap p-values the same 1000 samples were generated. These samples were generated from TAR processes with parameters taken from Table 12 except that we set $\rho_1 = \rho_2 = 0$ and $\beta_1 = \beta_2 = 0$ to enforce the null hypothesis of a unit root. For the "Fixed m" bootstrap, the R_T statistic was calculated setting $m = 2$. For the "Estimated m" bootstrap, first $\hat{m}$ was calculated by least squares, then R_T was calculated for this $\hat{m}$. In both cases, the bootstrap p-value is the percentage of the 1000 samples for which the simulated R_T exceeded the actual value of $R_T = 16.7$.

Table 10 shows that the if $\hat{m} = 2$ is considered as fixed, the conservative and bootstrap p-values agree at approximately 3.5%, suggesting that we can reject the null of a unit root in favor of a stationary process. If we view the delay parameter as estimated, the bootstrap p-value for R_T is 8.9%, which provides less strong evidence against the unit root hypothesis.

Table 13: Unit Root Test and P-Values, $m = 2$

	Asymptotic P-Values		Bootstrap P-Values	
R_T	Liberal	Conservative	Fixed m	Estimated m
16.7	0.012	0.035	0.033	0.089

7 Conclusion

This paper developed a new asymptotic theory for threshold autoregressive models with a possible unit root. The joint application of the two tests – for a threshold and for a unit root – allow a researcher to distinguish non-linear from non-stationary processes. We illustrate the methods with an application to the U.S. unemployment rate, and find evidence to support the hypothesis that the process is a stationary TAR.

Several extensions of our methods could be explored in future research. For example, it would be useful to allow for the coefficients on the lagged Δy_t to switch between regimes, and it would be interesting to allow for multiple (more than two) regimes. Of particular interest are multivariate extensions, such as the threshold cointegration model of Balke and Fomby (1997).

8 Appendix

Throughout the appendix, we simplify notation by setting $1_t(u) = 1_{\{U_t < u\}}$. Observe that since $U_t = F(Z_t) \sim U[0,1]$, $E1_t(u) = u$.

Proof of Theorem 1: Let $v_t(u) = 1_t(u) - u$ so that $Ev_t(u) = 0$. Since

$$\frac{1}{T}\sum_{t=1}^{T} X_{Tt}1_t(u) = \frac{1}{T}\sum_{t=1}^{T} X_{Tt}v_t(u) + u\frac{1}{T}\sum_{t=1}^{T} X_{Tt}$$

and

$$\frac{1}{T}\sum_{t=1}^{T} X_{Tt} \Rightarrow \int_0^1 X(s)ds,$$

it is sufficient to show that

$$\sup_{0\leq u\leq 1}\left|\frac{1}{T}\sum_{t=1}^{T} X_{Tt}v_t(u)\right| \to_p 0. \tag{A.1}$$

Fix $\varepsilon > 0$. Since $X(s)$ is continuous almost surely, there is some $\delta > 0$ such that

$$P\left(2\sup_{|s-s'|\leq\delta} |X(s) - X(s')| \leq \varepsilon\right) \geq 1 - \varepsilon. \tag{A.2}$$

Set $N = [1/\delta]$, and for $k = 0, ..., N-1$ set $t_k = [kT\delta] + 1$ and $t_k^* = t_{k+1} - 1$. Then

$$E \sum_{k=0}^{N-1} \sup_{0 \le u \le 1} \left| \frac{1}{T} \sum_{t=t_k}^{t_k^*} v_t(u) \right| = E \sup_{0 \le u \le 1} \left| \frac{1}{T\delta} \sum_{t=t_k}^{t_k^*} v_t(u) \right| \to 0$$

by the uniform weak law of large numbers. Hence

$$
\begin{aligned}
\sup_{0 \le u \le 1} \left| \frac{1}{T} \sum_{t=1}^{T} X_{Tt} v_t(u) \right| &= \sup_{0 \le u \le 1} \left| \frac{1}{T} \sum_{k=0}^{N-1} \sum_{t=t_k}^{t_k^*} X_{Tt} v_t(u) \right| \\
&\le \frac{1}{T} \sum_{k=0}^{N-1} |X_{Tt_k}| \sup_{0 \le u \le 1} \left| \sum_{t=t_k}^{t_k^*} v_t(u) \right| + \frac{1}{T} \sum_{k=0}^{N-1} \sum_{t=t_k}^{t_k^*} |X_{Tt} - X_{Tt_k}| \sup_{0 \le u \le 1} |v_t(u)| \\
&\le \sup_{1 \le t \le T} |X_{Tt}| \sum_{k=0}^{N-1} \sup_{0 \le u \le 1} \left| \frac{1}{T} \sum_{t=t_k}^{t_k^*} v_t(u) \right| + 2 \sup_{|t-t'| \le T\delta} |X_{Tt} - X_{Tt'}| \\
&\Rightarrow 2 \sup_{|s-s'| \le \delta} |X(s) - X(s')| \\
&\le \varepsilon.
\end{aligned}
$$

where the last inequality is (A.2). This establishes (A.1) as needed.   □

Proof of Theorem 2:

As noted in the text, for all u, $\{e_t(u), \Im_t\}$ is a strictly stationary and ergodic MDS with variance $E e_t(u)^2 = \sigma^2 u$. Thus by the MDS central limit theorem, for any (s, u),

$$Y_T(s, u) = \frac{1}{\sqrt{T}} \sum_{t=1}^{[Ts]} e_t(u) \to_d N\left(0, su\sigma^2\right).$$

Furthermore, the asymptotic covariance kernel is determined by

$$
\begin{aligned}
E\left(Y_T(s_1, u_1) Y_T(s_2, u_2)\right) &= \frac{1}{T} \sum_{t=1}^{[Ts_1] \wedge [Ts_2]} E\left(e_t^2 1_{t-1}(u_1) 1_{t-1}(u_2)\right) \\
&= \sigma^2 (s_1 \wedge s_2)(u_1 \wedge u_2).
\end{aligned}
$$

This yields the convergence of the finite dimensional distributions.

The stochastic equicontinuity of $Y_T(s, u)$ over $[0, 1]^2$ follows from the stochastic equicontinuity of

$$Y_T^*(s, u) = \frac{1}{\sqrt{T}} \sum_{t=1}^{[Ts]} e_t \left(1_{t-1}(u) - u\right).$$

This is established by Bai (1996, Theorem A.1) using a different set of dependence assumptions and moment bounds. A careful reading of his proof shows that these moment and dependence conditions are only used to prove inequality (A.3) below (his Lemma A.1). Thus (A.3) is sufficient to establish that $Y_T(s, u)$ is stochastically equicontinuous.

For any $0 \leq s_1 < s_2 \leq 1$ and $0 \leq u_1 < u_2 \leq 1$, let

$$Y_T^{**} = Y_T^*(s_2, u_2) - Y_T^*(s_1, u_2) - Y_T^*(s_2, u_1) + Y_T^*(s_1, u_1).$$

We need to show that for some $K < \infty$,

$$E\left|Y_T^{**}\right|^{2\gamma} \leq K\left[(s_2 - s_1)^\gamma (u_2 - u_1)^\gamma + \frac{(s_2 - s_1)(u_2 - u_1)}{T^{\gamma-1}} + o(1)\right] \qquad (A.3)$$

Direct calculation shows that

$$Y_T^{**} = T^{-1/2} \sum_{t=[Ts_1]}^{[Ts_2]} e_t \eta_{t-1}.$$

where $\eta_t = 1_{\{u_1 \leq U_t < u_2\}} - (u_2 - u_1)$. Note that $E\eta_t = 0$, $E|\eta_t|^{2\gamma} \leq u_2 - u_1$ and $E\eta_t^2 \leq u_2 - u_1$. Furthermore, η_{t-1} is $\Im_{t-1}$−measurable, so by (10), $E\left(e_t^2 \eta_{t-1}^2 \mid \Im_{t-1}\right) \leq B^{1/\gamma} \eta_{t-1}^2$ and $E\left(|e_t|^{2\gamma} |\eta_{t-1}|^{2\gamma}\right) \leq B(u_2 - u_1)$.

The fact that $e_t \eta_{t-1}$ is a martingale difference sequence allows the application of Rosenthal's inequality (Hall and Heyde (1980, p. 23)): For some $M < \infty$,

$$
\begin{aligned}
E\left|T^{-1/2} \sum_{t=[Ts_1]}^{[Ts_2]} e_t \eta_{t-1}\right|^{2\gamma} &\leq M\left[E\left|\frac{1}{T} \sum_{t=[Ts_1]}^{[Ts_2]} E\left(e_t^2 \eta_{t-1}^2 \mid \Im_{t-1}\right)\right|^\gamma + \frac{1}{T^\gamma} \sum_{t=[Ts_1]}^{[Ts_2]} E\left(|e_t|^{2\gamma} |\eta_{t-1}|^{2\gamma}\right)\right] \\
&= K\left[E\left|\frac{1}{T} \sum_{t=[Ts_1]}^{[Ts_2]} \eta_{t-1}^2\right|^\gamma + \frac{(s_2 - s_1)(u_2 - u_1)}{T^{\gamma-1}}\right]
\end{aligned}
$$

where we set $K = MB$. Since $\eta_t^2 - E\eta_{t-1}^2$ is bounded, its sample average converges in L^γ to zero, hence

$$
\begin{aligned}
E\left|\frac{1}{T} \sum_{t=[Ts_1]}^{[Ts_2]} \eta_{t-1}^2\right|^\gamma &\leq \left(\frac{1}{T} \sum_{t=[Ts_1]}^{[Ts_2]} E\eta_{t-1}^2 + \left[E\left|\frac{1}{T} \sum_{t=[Ts_1]}^{[Ts_2]} \left(\eta_{t-1}^2 - E\eta_{t-1}^2\right)\right|^\gamma\right]^{1/\gamma}\right)^\gamma \\
&\leq \left((s_2 - s_1)(u_2 - u_1) + o(1)\right)^\gamma.
\end{aligned}
$$

These bounds establish (A.3) as needed. □

Proof of Theorem 3:

For all u, $Y_T(s, u)$ is a martingale with square integrable innovations $T^{-1/2}e_t(u)$. For fixed u, Theorem 2 above and Theorem 2.2 in Kurtz and Protter (1991) yield the stated result. Technically, Kurtz and Protter (1991) only allow fixed u, while we need uniformity over $u \in [0, 1]$. A careful reading of their proof shows that uniformity over u holds if their equation (1.13) holds uniformly in u, which can be verified using the bounded convergence theorem. $\square$

Proof of Theorem 4. To simplify the derivations, we will assume that no lags of Δy_t appear in the regression. The inclusion of the intercept and trend variables in the regression means that the test statistic is invariant to the actual values of μ as well as σ^2. We can thus without loss of generality set $\mu = 0$ and $\sigma = 1$. Under H_0, we see that y_t is generated by the stochastic process

$$y_t = (1 - \frac{c}{T})y_{t-1} + e_t$$

so $T^{-1/2}y_{[Ts]} \Rightarrow W_c(s)$. Letting

$$X_{Tt} = \begin{pmatrix} 1 \\ \frac{t}{T} \\ \frac{1}{\sqrt{T}}y_{t-1} \end{pmatrix},$$

we find that

$$X_{T[Ts]} = \begin{pmatrix} 1 \\ \frac{[Ts]}{T} \\ \frac{1}{\sqrt{T}}y_{[Ts]-1} \end{pmatrix} \Rightarrow \begin{pmatrix} 1 \\ s \\ W_c(s) \end{pmatrix} = X_c(s). \tag{A.4}$$

Due to the equality $1_{\{Z_{t-1} < \lambda\}} = 1_{\{U_{t-1} < F(\lambda)\}}$, the change-of-variables $s = F(\lambda)$ allows us to re-write the test statistic as

$$W_T = \sup_{u \in [\pi_1, \pi_2]} W_T^*(u), \tag{A.5}$$

where $W_T^*(u)$ is the Wald statistic for the equality of $\theta_1 = \theta_2$ in the regression

$$\Delta y_t = X_{Tt}'\hat{\theta}_1(u)1_{\{U_{t-1} < u\}} + X_{Tt}'\hat{\theta}_2(u)1_{\{U_{t-1} \geq u\}} + \hat{e}_t(u). \tag{A.6}$$

Standard algebra shows that

$$W_T^*(u) = \frac{S_T^*(u)'M_T^*(u)^{-1}S_T^*(u)}{\hat{\sigma}^2(u)} \tag{A.7}$$

26

where

$$M_T(u) = \frac{1}{T}\sum_{t=1}^{T} X_{Tt}X_{Tt}'1_{t-1}(u),$$

$$M_T^*(u) = M_T(u) - M_T(u)M_T(1)^{-1}M_T(u)$$

$$S_T^*(u) = S_T(u) - M_T(u)M_T(1)^{-1}S_T(1),$$

and

$$S_T(u) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T} X_{Tt}e_t(u).$$

Theorem 1 and (A.4) imply

$$M_T(u) \Rightarrow u\int_0^1 X_c(s)X_c(s)'ds = uM_c,$$

say, and hence

$$M_T^*(u) \Rightarrow uM_c - (uM_c)\,M_c^{-1}\,(uM_c) = u(1-u)M_c. \qquad (A.8)$$

Theorem 3 and (A.4) show that

$$S_T(u) \Rightarrow \int_0^1 X_c(s)dW(s,u) = J_c(u)$$

and thus

$$S_T^*(u) \Rightarrow J_c(u) - uM_cM_c^{-1}J_c(1) = J_c(u) - uJ_c(1). \qquad (A.9)$$

Equations (A.7), (A.8), and (A.9) together show

$$W_T^*(u) \Rightarrow S_c(u). \qquad (A.10)$$

(A.5) and (A.10) yield

$$W_T \Rightarrow \sup_{u\in[\pi_1,\pi_2]} S_c(u)$$

which is the stated result.   □

Proof of Theorem 5.

To simplify the derivations, we will assume that no lags of Δy_t appear in the regression. Since the regression includes a trend and is studentized, the test statistic is invariant to the intercept $\mu = \mu_1 = \mu_2$ and the variance σ^2, so we set $\mu = 0$ and $\sigma^2 = 1$ to ease exposition.

We reparameterize the model as in (A.6) letting $u = F(\lambda)$. Standard algebraic results for linear regression show that

$$R_T(u) = \hat{t}_1(u)^2 + \hat{t}_2(u)^2$$

27

where $\hat{t}_1(u)$ and $\hat{t}_2(u)$ are the OLS t-statistics for $\hat{\rho}_1(u)$ and $\hat{\rho}_2(u)$ in equation (A.6), respectively. Letting

$$y_{t-1}^*(u) = y_{t-1} - \sum_{j=1}^T y_{j-1} r_j' 1_{j-1}(u) \left(\sum_{j=1}^T r_j r_j' 1_{j-1}(u) \right)^{-1} r_t,$$

we can write $\hat{t}_1(u)^2$ explicitly as

$$\hat{t}_1(u)^2 = \frac{N_T(u)^2}{\hat{\sigma}^2(u) D_T(u)}, \qquad (A.11)$$

where

$$N_T(u) = \frac{1}{T} \sum_{t=1}^T y_{t-1}^*(u) \Delta y_t 1_{t-1}(u)$$

and

$$D_T(u) = \frac{1}{T^2} \sum_{t=1}^T y_{t-1}^*(u)^2 1_{t-1}(u).$$

Our approach to finding the limit distribution of $R_T = R_T(\hat{u})$ is to find the limit distribution of $R_T(u)$ considered as a function of u, find the limit distribution of $\hat{u}$, and combine these two results with the continuous mapping theorem.

We first examine the denominator $D_T(u)$. Since under the null hypothesis (5), $c_1 = c_2 = 0$, it follows that

$$T^{-1/2} y_{[Ts]} \Rightarrow W(s).$$

Letting $r_{Tt} = (1 \quad t/T)'$ so that $r_{T[Ts]} \Rightarrow r(s)$, two applications of Theorem 1 yield

$$
\begin{aligned}
T^{-1/2} y_{[Ts]}^*(u) &= T^{-1/2} y_{[Ts]} - \sum_{j=1}^T y_{j-1} r_{Tj}' 1_{j-1}(u) \left(\sum_{j=1}^T r_{Tj} r_{Tj}' 1_{j-1}(u) \right)^{-1} r_{T[Ts]} \qquad (A.12) \\
&\Rightarrow W(s) - u \int_0^1 W(a) r(a)' da \left(u \int_0^1 r(a) r(a)' da \right)^{-1} r(s) \\
&= W^*(s).
\end{aligned}
$$

Theorem 1 and (A.12) yield

$$D_T(u) \Rightarrow u \int_0^1 W^*(s)^2 ds. \qquad (A.13)$$

We next turn to the numerator $N_T(u)$. Since $\mu_1 = \mu_2$, $N_T(u) = N_T^*(u)$, where

$$N_T^*(u) = \frac{1}{T} \sum_{t=1}^T y_{t-1}^*(u) e_t 1_{t-1}(u).$$

28

From Theorem 3 and (A.12) we find

$$N_T^*(u) \Rightarrow \int_0^1 W^*(s)dW(s,u). \qquad (A.14)$$

(A.11), (A.13) and (A.14) together yield

$$\hat{t}_1(u)^2 \Rightarrow \frac{\left[\int_0^1 W^*(s)dW(s,u)\right]^2}{u \int_0^1 W^*(s)^2 ds},$$

where $W_1(s,u) = u^{-1/2}W(s,u)$. Using similar arguments, we can show that

$$\hat{t}_2(u)^2 \Rightarrow \frac{\left[\int_0^1 W^*(s)\left(dW(s,1) - dW(s,u)\right)\right]^2}{(1-u) \int_0^1 W^*(s)^2 ds}.$$

where $W_2(s,u) = (1-u)^{-1/2}\left(W(s,1) - W(s,u)\right).$

Using the definitions $W_1(s,u) = u^{-1/2}W(s,u)$ and $W_2(s,u) = (1-u)^{-1/2}\left(W(s,1) - W(s,u)\right),$ we conclude that

$$R_T(u) \Rightarrow \frac{\left[\int_0^1 W^*(s)dW_1(s,u)\right]^2 + \left[\int_0^1 W^*(s)dW_2(s,u)\right]^2}{\int_0^1 W^*(s)^2 ds} = R(u), \qquad (A.15)$$

yielding the limit distribution of the function $R_T(u)$ considered as a function of u.

In our re-parameterized space the threshold estimator is defined as

$$\hat{u} = \underset{u \in [\pi_1, \pi_2]}{\operatorname{argmax}} W_T^*(u).$$

In equation (A.10) we showed that $W_T^*(u) \Rightarrow S_0(u)$ (since $c = 0$ under the null). This limit process $S_0(u)$ is continuous in u and has a unique maximum in $[\pi_1, \pi_2]$ with probability one. This allows the application of Theorem 2.7 of Kim and Pollard (1990), hence

$$\hat{u} \Rightarrow \underset{u \in [\pi_1, \pi_2]}{\operatorname{argmax}} S_0(u) = u^*. \qquad (A.16)$$

Equation (A.15) and (A.16) combine for the desired result:

$$R_T = R_T(\hat{u}) \Rightarrow R(u^*) = R^*.$$

Proof of Theorem 6.

As in the previous proof we use the reparameterized model (A.6) where $u = F(\lambda)$, $\hat{u} = F(\hat{\lambda})$ and $u_0 = F(\lambda_0)$. As in the previous proof, to simplify the derivations, we will

assume that no lags of Δy_t appear in the regression. Since the test statistic is invariant to the variance σ^2, we set $\sigma^2 = 1$.

Let $1_{t-1} = 1_{t-1}(u_0)$ and $\mu = \mu_1 E 1_{t-1} + \mu_2(1 - E 1_{t-1})$. Under H_0,

$$\Delta y_t = \mu_1 1_{t-1} + \mu_2(1 - 1_{t-1}) + e_t = \mu + \xi_t.$$

The sequence ξ_t is zero-mean and strictly stationary. We have the joint convergence over $(s, u) \in [0, 1]^2$:

$$\left(\begin{array}{c} \frac{1}{\sqrt{T}} \sum_{t=1}^{[Ts]} e_t 1_{t-1}(u) \\ \frac{1}{\sqrt{T}} \sum_{t=1}^{[Ts]} \xi_t \end{array} \right) \Rightarrow \left(\begin{array}{c} W_a(s, u) \\ \sigma_\xi W_b(s) \end{array} \right)$$

where $W_a(s, u)$ is a two-parameter Brownian motion, W_b is a standard Brownian motion, and the two processes are correlated. Since the regression includes a time trend, the test statistic is invariant to the parameter μ so we set $\mu = 0$. In this case we have

$$\frac{1}{\sqrt{T}} y_{[Ts]} \Rightarrow \sigma_\xi W_b(s)$$

and

$$\frac{1}{\sqrt{T}} y_{[Ts]}^*(u) \Rightarrow \sigma_\xi W_b^*(s)$$

where $W_b^*(s)$ is a detrended version of $W_b(s)$.

Let $D_T(u)$, $N_T(u)$, and $N_T^*(u)$ be defined in the previous proof. Theorems 1 and 3 yield

$$D_T(u) \Rightarrow u \sigma_\xi^2 \int_0^1 W_b^*(s)^2 ds$$

and

$$N_T^*(u) \Rightarrow \sigma_\xi \int_0^1 W_b^*(s) dW_a(s, u).$$

We need to show that $N_T(u)$ is asymptotically equivalent to $N_T^*(u)$. For $u < u_0$, $N_T(u) = N_T^*(u)$. For $u > u_0$,

$$\begin{aligned} \Delta y_t &= \mu_1 1_{\{U_{t-1} < u_0\}} + \mu_2 1_{\{U_{t-1} \geq u_0\}} + e_t \\ &= \mu_2 + (\mu_1 - \mu_2) 1_{\{U_{t-1} < u_0\}} + e_t \\ &= \mu_2 + (\mu_1 - \mu_2) 1_{\{U_{t-1} < u\}} - (\mu_1 - \mu_2) 1_{\{u_0 \leq U_{t-1} < u\}} + e_t. \end{aligned}$$

Linear projection shows that

$$N_T(u) = N_T^*(u) - (\mu_1 - \mu_2) A_T(u)$$

where

$$A_T(u) = \frac{1}{T} \sum_{t=1}^{T} y_{t-1}^*(u) 1_{\{u_0 \leq U_{t-1} < u\}}.$$

Observe that

$$
\begin{aligned}
|A_T(u)| &\leq \max_{t \leq T} T^{-1/2} |y_t^*(u)| \cdot \frac{1}{T^{1/2}} \sum_{t=1}^{T} 1_{\{u_0 \leq U_{t-1} < u\}} \\
&= O_p(1) \cdot T^{1/2}(u - u_0).
\end{aligned}
$$

Since $\mu_1 \neq \mu_2$, the threshold u_0 is identified and $T(\hat{u} - u_0) = O_p(1)$ (see Hansen, 1997). It follows that $N_T(\hat{u}) = N_T^*(\hat{u}) + o_p(1)$ as desired.

We find that as $\hat{u} \to_p u_0$,

$$
\begin{aligned}
\hat{t}_1(\hat{u})^2 &= \frac{N_T^*(\hat{u})^2}{\sigma^2 D_T(\hat{u})} + o_p(1) \\
&\Rightarrow \frac{\left[\sigma_\xi \int_0^1 W_b^*(s) dW_a(s, u_0) \right]^2}{u_0 \sigma_\xi^2 \int_0^1 W_b^*(s)^2 ds} \\
&= \frac{\left[\int_0^1 W_b^*(s) dW_1(s) \right]^2}{\int_0^1 W_b^*(s)^2 ds},
\end{aligned}
$$

where we have defined

$$
W_1(s) = \frac{W_a(s, u_0)}{\sqrt{u_0}}.
$$

Similar arguments for $\hat{t}_2(\hat{u})^2$ allow us to find that

$$
\hat{t}_2(\hat{u})^2 \Rightarrow \frac{\left[\int_0^1 W_b^*(s) dW_2(s) \right]^2}{\int_0^1 W_b^*(s)^2 ds}
$$

where

$$
W_2(s) = \frac{W_a(s, 1) - W_a(s, u_0)}{\sqrt{1 - u_0}}.
$$

It is important to observe that $W_1(s)$ and $W_2(s)$ are mutually independent standard Brownian motions which are both correlated with $W_b(s)$. We have shown that

$$
\begin{aligned}
R_T &= \hat{t}_1(\hat{u})^2 + \hat{t}_2(\hat{u})^2 \\
&\Rightarrow \frac{\left[\int_0^1 W_b^*(s) dW_1(s) \right]^2 + \left[\int_0^1 W_b^*(s) dW_2(s) \right]^2}{\int_0^1 W_b^*(s)^2 ds} \\
&\equiv R_*
\end{aligned}
$$

We now simplify the limiting distribution. Let

$$
W_0(s) = \begin{pmatrix} W_1(s) \\ W_2(s) \end{pmatrix} \sim BM(I_2).
$$

31

For any H such that $H'H = I_2$, we have that

$$\begin{pmatrix} W_1^H(s) \\ W_2^H(s) \end{pmatrix} = H'W_0(s) \sim BM(I_2).$$

We can pick H so that W_1^H and W_b are uncorrelated. Then by projection for some $a \in [0,1)$,

$$W_2^H(s) = aW_{2 \cdot b}(s) + (1 - a^2)^{1/2}W_b(s)$$

where $W_{2 \cdot b}$ is independent of W_b (and also of W_1^H). We thus have

$$
\begin{aligned}
R_* &= \frac{\left[\int_0^1 W_b^*(s)dW_0(s)'\right]\left[\int_0^1 dW_0(s)W_b^*(s)\right]}{\int_0^1 W_b^*(s)^2 ds} \\
&= \frac{\left[\int_0^1 W_b^*(s)dW_0^H(s)'\right]\left[\int_0^1 dW_0^H(s)W_b^*(s)\right]}{\int_0^1 W_b^*(s)^2 ds} \\
&= \frac{\left[\int_0^1 W_b^*(s)dW_1^H(s)\right]^2 + \left[\int_0^1 W_b^*(s)dW_2^H(s)\right]^2}{\int_0^1 W_b^*(s)^2 ds} \\
&= \left[\frac{\int_0^1 W_b^*(s)dW_1^H(s)}{\left(\int_0^1 W_b^*(s)^2 ds\right)^{1/2}}\right]^2 + \left[a\frac{\int_0^1 W_b^*(s)dW_{2 \cdot b}}{\left(\int_0^1 W_b^*(s)^2 ds\right)^{1/2}} + (1 - a^2)^{1/2}\frac{\int_0^1 W_b^*(s)dW_b(s)}{\left(\int_0^1 W_b^*(s)^2 ds\right)^{1/2}}\right]^2 \\
&\sim \chi_1^2 + \left[aN(0,1) + (1-a^2)^{1/2}DF\right]^2.
\end{aligned}
$$

The final equality in distribution holds since W_1^H, $W_{2 \cdot b}$ and W_b are mutually independent.
$\square$

References

[1] Altissimo, F. and G. Violante (1996): "Persistence and nonlinearity in U.S. gnp and unemployment: An endogenous delay threshold VAR,", University of Pennsylvania.

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

[3] Bai, J. (1996): "Testing for parameter constancy in linear regressions: An empirical distribution function approach," *Econometrica*, 64, 597-622.

[4] Balke, N.S. and T.B. Fomby (1997): "Threshold cointegration," *International Economic Review*, 38, 627-645.

[5] Beaudry, P. and G. Koop (1993): "Do recessions permanently change output?" *Journal of Monetary Economics*, 31, 149-164.

[6] 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.

[7] Chan, K.S. and R. S. Tsay (1997): "Limiting properties of the least squares estimator of a continuous threshold autoregressive model," *Biometrika*, forthcoming.

[8] Chen, C.W.S. and J.C. Lee (1995): "'Bayesian inference of threshold autoregressive models," *Journal of Time Series Analysis*, 16, 483-492.

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

[10] Dickey, D.A. and W.A. Fuller (1979): "Distribution of the estimators for autoregressive time series with a unit root," *Journal of the American Statistical Association*, 74, 427-431.

[11] Galbraith, J.W. (1996): " Credit Rationing and Threshold Effects In The Relation Between Money And Output," *Journal of Applied Econometrics*, 11, 419-429.

[12] Hall, P. and C.C. Heyde (1980): *Martingale Limit Theory and Its Application.* New York: Academic Press.

[13] Hamilton, J. (1989): "A new approach to the economic analysis of nonstationary time series and the business cycle," *Econometrica*, 5, 357-384.

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

[15] Hansen, B.E. (1997): "Approximate asymptotic p-values for structural change tests," *Journal of Business and Economic Statistics*, 15, 60-67.

[16] Hansen, B.E. (1997): "Inference in TAR models," *Studies in Nonlinear Dynamics and Econometrics*, 1(2), 119-131, forthcoming.

[17] Kim, J. and D. Pollard (1990): "Cube root asymptotics," *The Annals of Statistics,* 18, 191-219.

[18] Kurtz, T.G. and P. Protter (1991): "Weak limit theorems for stochastic integrals and stochastic differential equations," *The Annals of Probability*, 19, 1035-1070.

[19] Montgomery, A.L., V. Zarnowitz, R.S. Tsay, and G.C. Tiao (1996): "Nonlinearity in modeling and forecasting the U.S. unemployment rate," University of Chicago.

[20] Neftci, S. (1984): "Are economic time series asymmetric over the business cycle?" *Journal of Political Economy*, 92, 307-328.

[21] Phillips, P.C.B. (1987): "Time series regression with a unit root," *Econometrica*, 55, 277-302.

[22] Pippenger, M.K. and G.E. Goering (1995): "A note on the empirical power of unit root tests under threshold processes," *Oxford Bulletin of Economics and Statistics*, 55, 473-481.

[23] Potter, S.M. (1995): "A nonlinear approach to U.S. GNP," *Journal of Applied Econometrics*, 2, 109-125.

[24] Rothman, P. (1991): "Further evidence on the asymmetric behavior of unemployment rates over the business cycle," *Journal of Macroeconomics*, 13, 291-298.

[25] Said, S.E. and D.A. Dickey (1984): "Testing for unit roots in autoregressive-moving average models of unknown order," *Biometrika*, 71, 599-608.

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

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

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

[29] Tsay, R.S. (1997): "Unit-root tests with threshold innovations," University of Chicago.